COMPUTER SIMULATION OF MARKER-ASSISTED SELECTION UTILIZING LINKAGE DISEQUILIBRIUM

SARAH KEILDSON

Submitted in fulfillment of the academic
requirements of the degree of
Master of Science
In the School of Biochemistry, Genetics,
Microbiology and Plant Pathology

University of KwaZulu-Natal

Pietermaritzburg 2006

PREFACE

The experimental work described in this dissertation was conducted at the University of KwaZulu-Natal, Pietermaritzburg, under the supervision of Dr Carolyn Hancock.

The results have not been submitted in any other form to another university and, except where the work of others is acknowledged in the text, the results are of my own investigation.

.....

Sarah Keildson
December 2006

I certify that the above statement is correct.

.....

Dr Carolyn Hancock
Supervisor

ACKNOWLEDGEMENTS

I would like to thank Dr Carolyn Hancock for her guidance and support throughout the development of this thesis as well as for sharing her knowledge and time with me.

I would also like to thank Jonathan de Guisti for his useful advice regarding Microsoft Excel and his readiness to help.

Finally, I would like to thank the National Research Foundation (NRF) and University Research Fund (URF) for funding the work presented in this dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AFLP	Amplification fragment length polymorphisms
BV	Breeding value
D	Linkage disequilibrium coefficient
E	Environmental value
G	Genotypic value
$\mathbf{h^2}$	Heritability
h	Accuracy
i	Intensity of selection
MAS	Marker-assisted selection
P	Phenotypic value
QTL	Quantitative trait locus/loci
r	Recombination fraction
R	Response to selection
RAPD	Randomly amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
S	Selection differential
SNP	Single nucleotide polymorphisms
$\mathbf{V_A}$	Additive variance
$\mathbf{V_D}$	Dominance variance
$\mathbf{V_E}$	Environmental variance
$\mathbf{V_P}$	Phenotypic variance
$\boldsymbol{\sigma_A}$	Additive genetic standard deviation
$\boldsymbol{\sigma_P}$	Phenotypic standard deviation
$\boldsymbol{\mu}$	Population mean

ABSTRACT

The face of animal breeding has changed significantly over the past few decades. Traditionally, the genetic improvement of both plant and animal species focussed on the selective breeding of individuals with superior phenotypes, with no precise knowledge of the genes controlling the traits under selection. Over the past few decades, however, advances in molecular genetics have led to the identification of genetic markers associated with genes controlling economically important traits, which has enabled breeders to enhance the genetic improvement of breeding stock through linkage disequilibrium marker-assisted selection.

Since the integration of marker-assisted selection into breeding programmes has not been widely documented, it is important that breeders are able to evaluate the advantages and disadvantages of marker-assisted selection, in comparison to phenotypic selection, prior to the implementation of either selection strategy. Therefore, this investigation aimed to develop deterministic simulation models that could accurately demonstrate and compare the effects of phenotypic selection and marker-assisted selection, under the assumption of both additive gene action and complete dominance at the loci of interest.

Six computer models were developed using Microsoft Excel, namely 'Random Mating,' 'Phenotypic Selection,' 'Marker-Assisted Selection,' 'Selection with Dominance,' 'Direct Selection' and 'Indirect selection.' The 'Random Mating' model was firstly used to determine the effects of linkage disequilibrium between two loci in a randomly mating population. The 'Phenotypic Selection' and 'Marker-Assisted Selection' models focused primarily on examining and comparing the response to these two selection strategies over five generations and their consequent effect on genetic variation in a population when the QTL of interest exhibited additive gene action. In contrast, the 'Selection with Dominance' model investigated the efficiency of phenotypic selection and marker-assisted selection under the assumption of complete dominance at the QTL under selection. Finally, the 'Direct Selection and 'Indirect Selection' models were developed in order to mimic the effects of marker-

assisted selection on two cattle populations utilizing both a direct and indirect marker respectively.

The simulated results showed that, under the assumption of additive gene action, marker-assisted selection was more effective than phenotypic selection in increasing the population mean, when linkage disequilibrium was present between the marker locus and the QTL under selection and the QTL captured more than 80% of the trait variance. The response to both selection strategies was shown to decrease over five generations due to the decrease in genetic variation associated with selection. When the QTL under selection was assumed to display complete dominance, however, marker-assisted selection was markedly more effective than phenotypic selection, even when a minimal amount of linkage disequilibrium was present in the population and the QTL captured only 60% of the trait variance. The results obtained in this investigation were successful in simulating the theoretical expectations of marker-assisted selection.

The computer models developed in this investigation have potential applications in both the research and agricultural sectors. For example, the successful application of a model developed in this investigation to a practical situation that simulated marker-assisted selection, was demonstrated using data from two Holstein cattle populations. Furthermore, the computer models that have been developed may be used in education for the enhancement of students understanding of abstract genetics concepts such as linkage disequilibrium and marker-assisted selection.

CHAPTER 1
INTRODUCTION TO LINKAGE DISEQUILIBRIUM MARKER-ASSISTED SELECTION AND AIMS OF THE INVESTIGATION

1.1 OVERVIEW OF THE THEORY UNDERPINNING THE INVESTIGATION

Over the past fifty years, enormous advances in productivity have been achieved in animal species of agricultural importance (Dekkers and Hospital, 2002). This has largely been due to the genetic improvement of populations through a technique known as artificial selection (van der Werf, 2000b). Artificial selection focuses on selecting animals with desirable characteristics to reproduce, and consequently pass on the genes that code for these characteristics to their offspring (Gomez-Raya *et al.*, 2002). Theoretically, this process improves the overall genetic value of breeding populations by increasing the allele frequencies at the loci under selection (Hendrick, 2000). However, the genetic progress of many current animal breeding schemes can largely be attributed to selection on the basis of observable phenotype, phenotypic selection, without specific knowledge of the genetic architecture underlying the selected characteristics (Dekkers and Hospital, 2002).

Although improvements to economically important traits have been achieved in several species of livestock through phenotypic selection, the limitations of this breeding method are becoming evident with time (Montaldo and Meza-Herrera, 1998). Consumer demands for food quality at the lowest possible price, have forced traditional breeders to largely ignore the health traits of livestock in favour of improving productivity traits (Williams, 2005). This has resulted in the inadvertent selection of genetic defects, which now threaten the viability of livestock production in general (Xie and Xu, 1998). In addition, economically important traits are often quantitative traits, which reflect the joint action of a large number of genes confounded with environmental effects (Georges *et al.*, 1994). Consequently, the accuracy of phenotypic selection may be compromised by the effects of

environmental factors and gene interactions, both of which are difficult to quantify (Meuwissen and van Arendonk, 1992). In order to develop a sustainable industry, it is therefore necessary to address the potential problems associated with traditional phenotypic selection by exploiting new technologies that facilitate the selection of genetically superior individuals (Williams, 2005).

The limitations of phenotypic selection may be minimised by the identification of the genes that control economically important traits (van der Werf, 2000a). Knowledge of these genes would enable selection for traits based on the genotype, rather than the phenotype, of animals (Dekkers, 2004). The enormous advantage of genotypic selection is that those individuals carrying beneficial genes for several traits could be identified and mated to produce progeny that have the potential to express many desired traits simultaneously (Williams, 2005). Furthermore, while classical genetic evaluation methods assumed that economically important traits were controlled by numerous genes, each with only a small effect, recent studies have identified genes which have a major impact on certain traits (Kinghorn, 2000). Rapid advances in molecular biology have given impetus to the possibilities of exploiting these major genes in breeding programmes, for greater genetic, and therefore economic, gain (Gomez-Raya *et al.*, 2002).

Molecular genetic analysis of quantitative traits has lead to the identification of specific landmarks in the genome, known as molecular or genetic markers, which can be used to enhance genetic improvement programmes (Kinghorn, 2000). Molecular markers are easily identifiable DNA sequences that are associated with major genes in such a way that identification of a particular marker can indicate the presence of a particular gene (Dekkers and Hospital, 2002). Two broadly different types of genetic markers exist. If the marker is a causal mutation located within the economically important gene, it is referred to as a direct marker (van der Werf, 2000a). By contrast, non-functional or anonymous polymorphisms that are located close to genes of interest on a chromosome, are termed indirect markers (Kinghorn, 2000).

Whereas causative mutations give direct information about the genes of interest, the use of indirect markers for selection depends on the frequency at which the marker and the gene are inherited together in a population (Dekkers and Hospital, 2002).

When both the marker locus and the locus under selection are positioned close enough together on a chromosome, such that recombination between them is minimal, the two loci occur together at a higher frequency in a population than would be expected by chance (Falconer, 1989). In this situation, the loci are said to be in linkage disequilibrium and selection on the marker genotype will frequently result in the indirect selection for the trait of interest (Asins, 2002).

The process of selecting individuals based on their marker genotype is known as marker-assisted selection (Xie and Xu, 1998). Causal mutations rarely have major effects on economically important characteristics and as a result, marker-assisted selection often relies on linkage disequilibrium between marker loci and loci containing beneficial genes (Meuwissen and van Arendonk, 1992). Once linkage disequilibrium has been established and identified between genes controlling traits of interest and genetic markers, genotypic marker data can be used to estimate breeding values of individuals in a breeding population (Asins, 2002). Marker-assisted selection based on linkage disequilibrium may therefore enable selection to be carried out across whole populations (Williams, 2005). Thus, with the potentially large number of loci involved in the inheritance of economically important traits, linkage disequilibrium between closely linked markers and genes can be detected in population studies and used in selection (Zhang and Smith, 1992).

Marker-assisted selection has provided opportunities to enhance the response to selection of a population, particularly for traits that are difficult to improve by conventional selection (van der Werf, 2000a). While opportunities for the use of molecular information exist, their successful implementation requires a comprehensive integrated strategy that is closely aligned with business goals (Dekkers, 2004). The current attitude toward marker-assisted selection is therefore one of cautious optimism.

This investigation aimed to develop computer simulations that would demonstrate and assess the effects of linkage disequilibrium in marker-assisted selection programmes.

1.2 LINKAGE DISEQUILIBRIUM

1.2.1 Introduction

In the middle of the 19th Century, Gregor Mendel's experiments on the garden pea revealed the first genetic explanation for the inheritance of characteristics from one generation to the next (Raven and Johnson, 1999). Through crossing the F_1 seeds of parental plants that differed in two traits, Mendel noticed that the phenotypic classes in the F_2 always appeared in the same approximate ratio (Snustad and Simmons, 2000). This ratio suggested that each trait was controlled by a different gene segregating two alleles, and that the two genes were inherited independently of one another (Falconer, 1989). Using these two principles, geneticists were subsequently able to predict the genotypic proportions of the offspring of those parents whose genotypes were known (Hartl and Clark, 1997).

However, at the beginning of the twentieth century, Bateson and Punnett crossed plants that differed in two traits and found that the phenotypic distribution of the offspring generation deviated from the classical Mendelian ratios (Snustad and Simmons, 2000). Although the reason for this deviation was unknown at the time, it was later discovered that the two genes did not assort independently because they were physically linked on the same chromosome and therefore tended to be inherited together (Donovan and Welden, 2002). The discovery of linkage meant that it was no longer sufficient to talk about single-locus systems alone, but rather it became necessary to consider two linked loci simultaneously (Hendrick, 2000).

1.2.2 Use of gametic frequencies to measure extent of linkage disequilibrium

The gametic proportions within a population are strongly dependent on whether the loci are in linkage equilibrium or linkage disequilibrium (Snustad and Simmons, 2000). When two loci in a population are not linked, the loci assort independently of one another and are said to be in linkage equilibrium (Dekkers, 2004). In contrast, linkage disequilibrium occurs when the inheritance of alleles at different loci tend to co-vary, so that when a gamete contains A_1, it is more likely to contain B_1 than a

randomly chosen allele (Dekkers and Hospital, 2002). The non-random association of alleles within gametes is therefore directly related to the amount of linkage disequilibrium between the loci, making it possible to measure the extent of linkage disequilibrium present in a population by studying the gametic frequencies (Falconer, 1989).

Linkage equilibrium can be thought of as the two-locus version of the Hardy-Weinberg Principle, which states that within a randomly mating population, if allele frequencies for a single locus with two alleles are given by p and q, then the genotypic frequencies for a population in Hardy-Weinberg equilibrium will be p^2, $2pq$ and q^2 (Snustad and Simmons, 2000). Therefore, in the absence of natural selection, mutation, genetic drift and gene flow, and provided the loci are independent of one another, the genotypic frequencies at each locus will be in Hardy-Weinberg proportions, while the gametes will be in linkage equilibrium (Falconer, 1989).

For two loci, A and B, each with two alleles, A_1, A_2 and B_1, B_2, there are four possible gametes namely, A_1B_1, A_1B_2, A_2B_1, A_2B_2 (Hendrick, 2000). The frequencies of each gametic class in the population depend on the genotypic frequencies, and therefore the allele frequencies, in the parental population (Donovan and Welden, 2002). Thus, when two loci are in linkage equilibrium, the alleles at these loci are randomly associated during gamete formation and the frequency of each gametic class can be calculated by multiplying the frequencies of the alleles that make up the gamete, as shown in Figure 1.1 (Hartl and Clark, 1997). The frequencies of A_1, B_1 and A_2, B_2 in the adult population are given by p_1, q_1 and p_2, q_2 respectively.

	A_1 (p_1)	**A_2 (p_2)**
B_1 (q_1)	A_1B_1 (p_1q_1)	A_2B_1 (p_2q_1)
B_2 (q_2)	A_1B_2 (p_1q_2)	A_2B_2 (p_2q_2)

Figure 1.1 Association of alleles at the A and B loci and the expected gametic frequencies, where p_1 = freq (A_1), p_2 = freq (A_2), q_1 = freq (B_1) and q_2 = freq (B_2).

The inheritance of alleles at different loci in linkage disequilibrium is, however, not always independent of one another and therefore alleles do not always associate randomly during gamete formation (Falconer, 1989). For case of reference, the following notations are commonly used to represent the different gametic frequencies in a population where loci may be in linkage disequilibrium (Gillespie, 2004):

P_{11} = the frequency of the A_1B_1 gamete
P_{12} = the frequency of the A_1B_2 gamete
P_{21} = the frequency of the A_2B_1 gamete
P_{22} = the frequency of the A_2B_2 gamete

The examination of haplotypes (specific combinations of alleles along chromosomes) is significant because genetic polymorphisms generally exhibit complex relationships, which provide important information about the number of meiotic events between loci (Abecasis *et al.*, 2005). The extent to which alleles at different loci associate in a non-random manner during gamete formation, is dependent on the distance between the two loci and the consequent degree of recombination that occurs. The amount of recombination between two loci is measured by the recombination fraction (r), which is directly related to the proportion of recombinant gametes produced by an individual (Raven and Johnson, 1999). The possible effect of recombination on the association of alleles during gamete formation produced by double heterozygous individuals is illustrated in Figure 1.2.

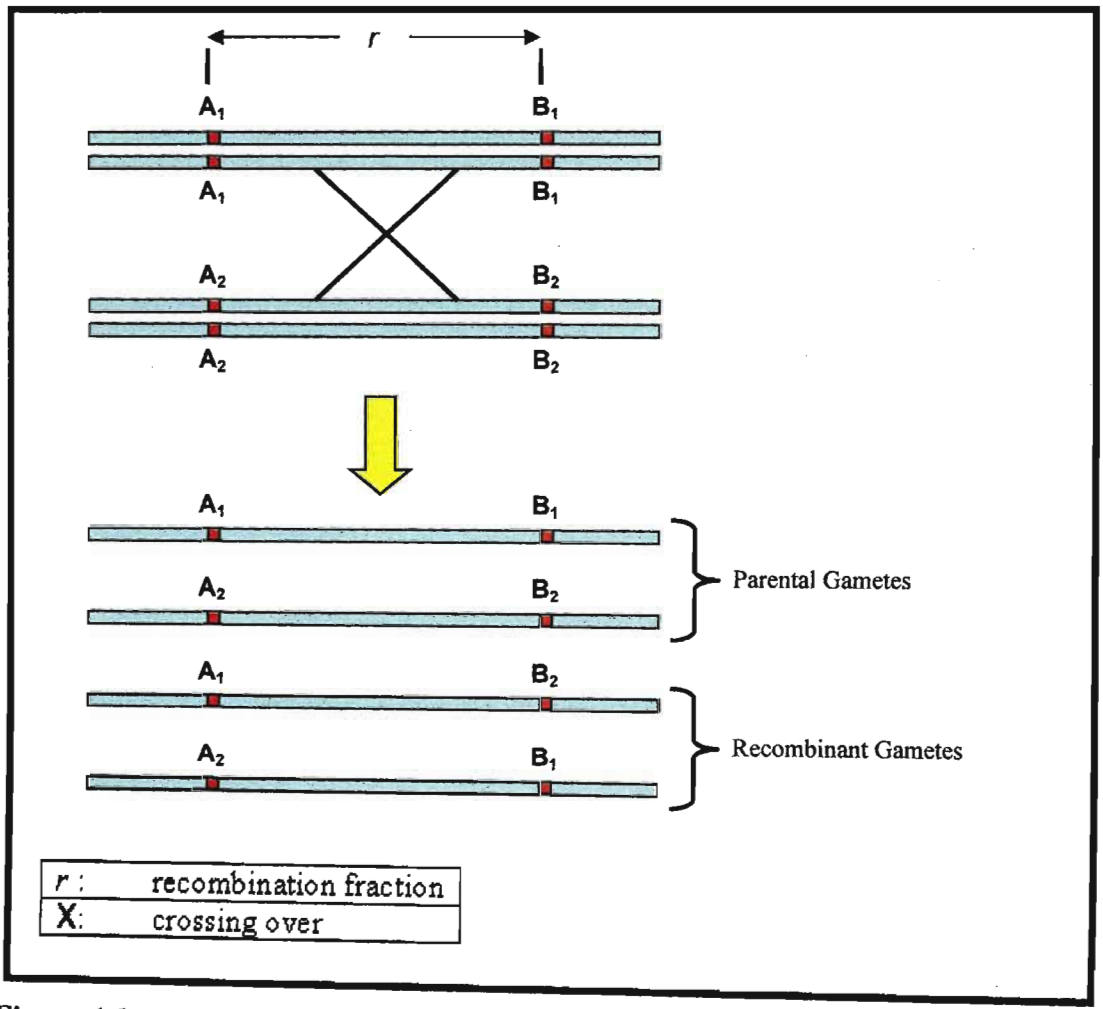


Figure 1.2 Four possible gametes produced by a double heterozygous individual.

As is evident in Figure 1.2, the parental configuration has A_1 on the same chromosome as B_1 and A_2 on the same chromosome as B_2, which is known as the coupling phase. If there is no crossing over, and therefore no recombination, between these two loci, only the two parental gametes will result. However, if there is crossing over, then recombination occurs and the two recombinant gametes will also result. The closer together the A and B loci are positioned along the chromosome, the less chance there is of crossing over occurring between them. When there is little recombination between the two loci, the majority of gametes will be of the parental rather than the recombinant type and the A and B loci will be said to be in linkage disequilibrium (Hartl and Clark, 1997).

Although the expected gametic frequencies for two loci in linkage equilibrium can be calculated by simply multiplying the frequencies of the alleles that make up the gametes (Figure 1.1), the gametic frequencies for two loci in linkage disequilibrium

are expressed as a function of the deviation from the expected gametic frequencies (***D***) (Falconer, 1989) (Table1.1).

Table 1.1 Comparison of gametic frequencies when two loci are in linkage equilibrium or linkage disequilibrium (adapted from Gillespie, 2004).

Representation of Gametic Frequencies	Gametic Frequencies at Linkage Equilibrium	Gametic frequencies at Linkage Disequilibrium
P_{11}	P_1q_1	$p_1q_1 + D$
P_{12}	P_1q_2	$p_1q_2 - D$
P_{21}	P_2q_1	$p_2q_1 - D$
P_{22}	P_2q_2	$p_2q_2 + D$

D measures the amount of linkage disequilibrium in a population and is known as the linkage disequilibrium coefficient (Hartl and Clark, 1997). It can be calculated by the formula $D = P_{11}P_{22} - P_{12}P_{21}$ which is the product of the frequencies of the coupling gametes (A_1B_1 / A_2B_2) minus the product of the frequencies of repulsion gametes (A_2B_1 / A_1B_2) (Gillespie, 2004). ***D*** ranges from 0 – 0.25, where $D = 0$ represents a population in linkage equilibrium and ***D*** increases as the association between two alleles in a population increases (Li, 1976).

Gametic frequencies are dependent on the amount of recombination between two loci and are expressed as a function of ***D***. For this reason, the level of linkage disequilibrium between two loci in a population, can be determined by analysing the gametic frequencies (Hendrick, 2000).

1.2.3 Approach to equilibrium

The approach to equilibrium of a one-locus compared to a two-locus system is vastly different (Falconer, 1989). In the case of a single locus, the Hardy-Weinberg Principle states that after one generation of random mating, allele frequencies will reach their equilibrium values (Snustad and Simmons, 2000). However, because linked loci are not independent of one another, equilibrium will not be reached in one generation, but instead it will be approached slowly as the association between the loci is broken down (Hendrick, 2000).

The approach to equilibrium for two loci in linkage disequilibrium is determined firstly by ***D***, which is the amount of linkage disequilibrium between the two loci, and secondly by r (Hartl and Clark, 1997).

Since the recombination fraction is directly proportional to the distance between two loci, if the loci are very far apart on the same chromosome, or on different chromosomes altogether, r will be at its maximum of 0.5, and each of the four types of gametes will be produced in equal proportions (Hendrick, 2000). However, if two loci in a population are in linkage disequilibrium, then the gametic frequencies of the next generation will not be equal to the gametic frequencies of the parental population, as would be in the case of linkage equilibrium (Li, 1976). For linkage disequilibrium, the gametic frequencies are dependent on both the parental genotype and r, as is shown in Table 1.2.

Table 1.2 Predicted offspring gametic frequencies in a population where the A and B loci in the parental generation are linked, with a recombination fraction, r (adapted from Hendrick, 2000).

Genotype	Frequency	Gametic Proportions			
		A_1B_1	A_1B_2	A_2B_1	A_2B_2
A_1B_1/A_1B_1	P_{11}^2	1	0	0	0
A_1B_1/A_1B_2	$2P_{11}P_{12}$	$1/2$	$1/2$	0	0
A_1B_1/A_2B_1	$2P_{11}P_{21}$	$1/2$	0	$1/2$	0
A_1B_1/A_2B_2	$2P_{11}P_{22}$	$(1-r)/2$	$r/2$	$r/2$	$(1-r)/2$
A_1B_2/A_1B_2	P_{12}^2	0	1	0	0
A_1B_2/A_2B_1	$2P_{12}P_{21}$	$r/2$	$(1-r)/2$	$(1-r)/2$	$r/2$
A_1B_2/A_2B_2	$2P_{12}P_{22}$	0	$1/2$	0	$1/2$
A_2B_1/A_2B_1	P_{21}^2	0	0	1	0
A_2B_1/A_2B_2	$2P_{21}P_{22}$	0	0	$1/2$	$1/2$
A_2B_2/A_2B_2	P_{22}^2	0	0	0	1

It is important to note that linkage does not affect the gametic output of genotypes with one or two pairs of genes in the homozygous condition (Li, 1976). For example, Table 1.2 demonstrates that A_1B_1/A_1B_1 individuals can only produce A_1B_1 gametes and A_1B_1/A_1B_2 individuals can only produce A_1B_1 and A_1B_2 gametes. The recombinant gametes will therefore always be identical to the parental gametes, regardless of the amount of recombination that exists between them (Gillespie, 2004). The only effect that linkage has is to modify the gametic output of the double

heterozygotes, A_1B_1/A_2B_2 and A_1B_2/A_2B_1, which have the potential to generate four different types of gametes as seen in Table 1.2 (Gaut and Long, 2003).

The frequencies of the gametes produced by the double heterozygote individuals, depend on whether the parental alleles were in the coupling (A_1B_1 / A_2B_2) or repulsion (A_1B_2 / A_2B_1) phase (Falconer, 1989). The overall frequency of recombinant gametes, shown in Table 1.2, is equal to r and the two recombinant gametes each occur in an equal frequency of $^{r}/_{2}$. The two parental gametes also occur in equal proportions, each with a frequency of $^{(1-r)}/_{2}$. The sum of the recombinant and parental gametes is equal to 1.

The information presented in Table 1.2 can be used to calculate the frequency of any gamete in the next generation by summing the product of the frequency and the proportion of the different gametes produced by each genotype (Li, 1976). For example, the frequency of the A_1B_1 gamete in the next generation (P_{11}') can be calculated as follows:

$$P_{11}' = P_{11}^2 + (P_{11})(P_{12}) + (P_{11})(P_{21}) + (1\text{-}r)(P_{11})(P_{22}) + (r)(P_{12})(P_{21})$$
$$= (P_{11})(P_{11} + P_{12} + P_{21} + P_{22}) - (r)((P_{11})(P_{22}) - (P_{12})(P_{21}))$$

Since $P_{11} + P_{12} + P_{21} + P_{22} = 1$ (the sum of the gametic frequencies in a population) and $(P_{11})(P_{22}) - (P_{12})(P_{21}) = \boldsymbol{D}$, the equation can be further simplified to:

$$P_{11}' = P_{11} - (r)(\boldsymbol{D})$$

Similarly,

$$P_{12}' = P_{12} - (r)(\boldsymbol{D})$$
$$P_{21}' = P_{21} - (r)(\boldsymbol{D})$$
$$P_{22}' = P_{22} - (r)(\boldsymbol{D})$$

The equations given above, demonstrate that recombination removes associations between alleles on chromosomes and is directly responsible for the approach to equilibrium of linked loci (Gillespie, 2004). If the loci are tightly linked and the recombination fraction is small as a result, the approach to equilibrium will be slow

(Hendrick, 2000). In contrast, if recombination readily occurs between loosely linked loci, equilibrium will be reached far more quickly as the linkage disequilibrium is broken down (Hartl and Clark, 1997).

Disequilibrium declines by a factor r every generation, so that the breakdown of linkage disequilibrium is measured by the equation $D_n = (1-r)^n D_0$, where D_n is the amount of disequilibrium in generation n, and D_0 is the amount of disequilibrium in the initial population (Li, 1976). Thus, as the number of generations (n) approaches infinity, the amount of linkage disequilibrium approaches 0 so that the loci will be in linkage equilibrium (Gillespie, 2004). It may therefore be concluded recombination is directly and primarily responsible for the decay in linkage disequilibrium between two loci in a population (Abecasis *et al.*, 2005).

In addition to the recombination frequency between two loci and the amount of linkage disequilibrium present in the initial population, recurrent mutations and gene conversion can also break down the amount of linkage disequilibrium in a population and thus speed up the approach to equilibrium (Hendrick, 2000). Recurrent mutations at a particular locus serve to generate new alleles and, as a consequence, the existing frequency distributions of linked loci are altered and linkage disequilibrium is broken down (Terwilliger *et al.*, 1998). Some single-nucleotide polymorphisms might have high mutation rates and therefore show little or no linkage disequilibrium with nearby markers, even in the absence of historical recombination (Abecasis *et al.*, 2005). Similarly, gene conversion, which occurs when a short stretch of one copy of a chromosome is transferred to the other copy during meiosis, can also have an effect on linkage (Ardlie *et al.*, 2001). This effect is equivalent to two very closely spaced recombination events, and can break down linkage disequilibrium in a manner similar to recombination or recurrent mutation (Hartl and Clark, 1997). However, it should be noted that in natural populations, the reduction in the magnitude of linkage disequilibrium by recombination, mutation and gene conversion is opposed by several evolutionary forces that may serve to increase $\boldsymbol{D}$ and slow down the approach to equilibrium (Gillespie, 2004).

As the approach to equilibrium is largely determined by the frequency of recombination and the amount of linkage disequilibrium present in the initial population, the effects are demonstrated by the following examples.

(a) Example illustrating the effect of recombination frequency

The example given in Table 1.3 demonstrates the effect of recombination on the approach to linkage equilibrium after three generations in a population where the amount of linkage disequilibrium in the initial population was 0.25 (absolute linkage). As is evident from the table, when the recombination fraction is 0.2, there will be approximately half the amount of linkage disequilibrium in the population after three generations relative to the amount of linkage disequilibrium in the initial population. In contrast, when the recombination fraction is 0.5, almost all the linkage disequilibrium has been broken down and linkage equilibrium is reached after three generations. This shows that the higher the recombination fraction between two loci in linkage disequilibrium, the faster the approach to linkage equilibrium.

Table 1.3 Illustration of the effect of recombination on the approach to equilibrium, where r is the recombination fraction and D_x is the amount of linkage disequilibrium in generation x.

Generation	r = 0.2	r = 0.5
1	$D_1 = 0.200$	$D_1 = 0.125$
2	$D_2 = 0.160$	$D_2 = 0.063$
3	$D_3 = 0.128$	$D_3 = 0.031$

(b) Example illustrating the effect of the amount of linkage disequilibrium present in the initial population

Another factor which influences the approach to equilibrium is the amount of linkage disequilibrium present in the initial population (Gillespie, 2004). Since linkage disequilibrium is dependent on the frequency of the double heterozygotes, the approach to equilibrium essentially depends on the frequency of the heterozygotes in the initial population (Li, 1976).

The example given in Table 1.4 demonstrates the effect of the amount of linkage disequilibrium in the initial population (D_0) on the approach to linkage equilibrium

after 3 generations in a population where the recombination fraction is 0.2. It is evident that when the amount of linkage disequilibrium in the initial population is high, the approach to equilibrium is much slower. When $D_0 = 0.25$, there is approximately half the initial amount of linkage disequilibrium present in the population after three generations. When $D_0 = 0.12$, however, linkage equilibrium is almost reached after only three generations.

Table 1.4 Illustration of the effect of the amount of linkage disequilibrium in the initial population on the approach to equilibrium, where D_0 is the initial amount of linkage disequilibrium and D_x is the amount of linkage disequilibrium in generation x.

Generation	$D_0 = 0.25$	$D_0 = 0.12$
1	$D_1 = 0.200$	$D_1 = 0.096$
2	$D_2 = 0.160$	$D_2 = 0.077$
3	$D_3 = 0.128$	$D_3 = 0.061$

1.2.4 Causes of linkage disequilibrium

If the full potential of linkage disequilibrium is to be exploited in breeding programmes and other association studies, understanding its causes is essential. There are six evolutionary forces which generate linkage disequilibrium between loci in populations, namely the distance between two loci, recurrent mutations, admixture, genetic drift, population structure and inbreeding.

Firstly, it has already been established that linkage disequilibrium is a function of whether two loci are linked. The more tightly linked the loci, the lower the recombination fraction will be between two loci and the higher the amount of linkage disequilibrium that will exist between them (Hendrick, 2000).

Recurrent mutations are the second evolutionary force that generates linkage disequilibrium. Generally speaking, single mutation events in the distant past have resulted in the different alleles that exist today (Gaut and Long, 2003). While the original mutant allele would have been confined to a specific haplotype belonging to a specific individual in a specific population, over time this allele may have spread to other haplotypes through recombination, gene-conversion or recurrent mutation

(Abecasis *et al.*, 2005). Furthermore, it may have spread to other populations through migration of its carriers and it may have changed in frequency due to natural selection or genetic drift (Gillespie, 2004). Together, these phenomena have resulted in the certain patterns of linkage disequilibrium which exist in modern populations (Falconer, 1989).

Thirdly, multi-locus models, which represent a finite population size, have shown that non-random association that develops between alleles at different loci may be as a result of admixture or the mixing of two populations which are genetically different (Gaut and Long, 2003). The amount of linkage disequilibrium generated by this admixture depends very strongly on the differences in gene frequencies between the initial populations and the recombination fraction between the two loci (Ardlie *et al.*, 2001).

Genetic drift is the fourth cause of linkage disequilibrium. Genetic drift can cause the haplotypes flanking genes of interest to be indirectly selected (Gaut and Long, 2003). As a result, those haplotypes could occur at high frequencies, or even become fixed, creating noticeable disequilibrium between the loci making up the haplotype (Abecasis *et al.*, 2005). However, the changes in allele frequency due to genetic drift, are mainly a function of population size (Gaut and Long, 2003). When the frequency of a particular haplotype is substantially smaller than the frequency of another haplotype in a population, genetic drift will cause a more dramatic variance in allele frequencies from one generation to the next with regards to the haplotype with the lowest frequency (Terwilliger *et al.*, 1998). According to Farnir *et al.* (2000), the larger the population size, the smaller the variance in allele frequency between generations. In contrast, allele frequencies change relatively quickly in smaller populations, which subsequently reach linkage disequilibrium at a faster rate (Gilean and McVeana, 2002).

Fifthly, population structure has the potential to generate unique associations between certain alleles (Hendrick, 2000). Allelic association due to population structure can occur when sub-populations become geographically isolated (Snustad and Simmons, 2000). Over time, mutation and genetic drift produce unique patterns of linkage disequilibrium in these isolated populations, to the extent that allele combinations

which are rare in some populations, can become common in isolated populations or some may be lost altogether (Ohta, 1981).

Finally, inbreeding generally leads to a more rapid generation of linkage disequilibrium than random mating (Falconer, 1980). This is because inbreeding has a tendency to decrease the proportion of the double heterozygotes in the population and, since crossing over can only generate recombinant gametes in double heterozygotes, the rate of decay of linkage disequilibrium is far lower in inbred populations (Gillespie, 2004).

Whether the linkage disequilibrium between two loci in a population is caused by their physical position on a chromosome or by the evolutionary forces that act upon them, it results in a deviation from the normal patterns of Mendelian inheritance, which can be quantified. However, the importance of understanding the causes of linkage disequilibrium and the factors that influence it, lies in its applications to selective breeding strategies.

1.2.5 Applications of linkage disequilibrium

Recently, linkage disequilibrium has received considerable attention in the agricultural sector because of its ability to be uniquely exploited in gene mapping and marker-assisted selection (Abecasis *et al.*, 2005).

Linkage disequilibrium has played an important role in increasing the mapping resolution of genes controlling both simple and complex traits (Georges *et al.*, 1995). Conventional linkage mapping of genes has been hampered by the fact that the information from linkage analysis is family specific and must therefore be re-evaluated for each family group in the population (Meuwissen *et al.*, 2002). The potential advantage of linkage disequilibrium mapping over linkage mapping, lies in its use of historical recombinants rather than the family-specific information on recombinants obtained through linkage analysis (Farnir *et al.*, 2000). Linkage disequilibrium can thus be used for the high resolution mapping of genes, because it reduces the environmental effects that hamper linkage analysis within family groups (Meuwissen *et al.*, 2002).

Markers in linkage disequilibrium with beneficial genes allow for selection of genotypes across an entire population because of the consistent association between the marker and a genotype (Williams, 2005). The implementation of marker-assisted selection based on population-wide linkage disequilibrium is both straightforward and potentially very accurate and can therefore be utilized in the livestock breeding industry to increase the response to selection with regards to economically important characteristics (Meuwissen and van Arendonk, 1992).

1.3 MARKER-ASSISTED SELECTION

1.3.1 Introduction

To date, most genetic progress for quantitative traits in livestock has been made by selection on phenotype, or on estimates of breeding values derived from phenotype, without any direct knowledge of the genes that affect the trait of interest (Babu *et al.*, 2004). Breeding programmes have therefore had to resort to information from performance data and progeny testing to make inferences about the nature of these genes (Montaldo and Meza-Herrera, 1998). Indeed, large progeny tests can be highly accurate, but are also expensive and for some traits (such as those related to meat quality and milk production), it takes several years before the benefits from progeny test results have an effect (van der Werf, 2000a). However, direct knowledge of the genes conferring merit on a trait could lead to much faster and more accurate genetic progress (Kinghorn, 2000). Genetic markers have made such knowledge available to plant and animal breeders.

1.3.2 Description of genetic markers

Well-designed experiments can be used to find genetic markers that are in linkage disequilibrium with genes of interest, so that breeders are able to identify the breeding value of an individual by the specific genetic marker it carries (van der Werf, 2000a). The use of genetic markers in breeding involves four steps. The first step is to search for genetic markers in the genome, and secondly, to establish a high resolution linkage map of these markers (Meuwissen and Goddard, 1996). Associations between markers and genes of interest can then be detected by using markers to track the

inheritance of chromosomal regions in families where the trait is segregated, and correlating this information with measurements on the individuals displaying the trait (Mauricio, 2001). Finally, these marker-gene associations can be utilized by marker-assisted selection programmes (Butterfield *et al.*, 2001). Genetic markers, however, have to satisfy a number of different requirements if they are to be utilized successfully in any kind of breeding strategy.

1.3.3 Marker requirements

Ideally, genetic markers should fulfil the following five requirements in order to be used effectively in marker-assisted selection.

1. Marker loci should be highly polymorphic (Kinghorn, 2000). The association between marker type and the alleles at loci under selection should be unique, such that a different form of the marker can potentially be identified as linked to each variant of the gene of interest (Rothwell, 1993). Since a number of different alleles can exist at each locus under selection in a population, it is necessary that many different forms of the marker locus are also present (Hartl and Clark, 1997).

2. Marker loci need to be neutral with respect to both the quantitative trait of interest and reproductive fitness (Jansen, 2001). If the markers under selection decrease the viability of a trait in any way, the response to selection may be compromised to the extent that there is an overall negative response (Williams, 2005).

3. Marker loci should be co-dominant so that all potential genotypes can be identified (Falconer, 1989). This is because, if the markers are dominant, then the homozygous dominant and the heterozygous conditions will appear to be the same and mistakes could be made in the estimation of breeding values (Jansen, 2001).

4. Marker loci should be abundant throughout the genome (Hartl and Clark, 1997). This is important since every gene of potential importance needs to lie

close enough to a genetic marker in the genome that some degree of linkage disequilibrium exists between them (Gillespie, 2004). If marker loci are scarcely distributed along chromosomes, important genes could be overlooked because they are not linked to markers (Dekkers, 2004).

5. Linkage disequilibrium has to exist at the population level between the markers and alleles at the locus controlling the trait of interest in order to accurately predict breeding values (Williams, 2005). To determine whether linkage disequilibrium exists between the genetic markers and gene of interest, the frequencies of the associations between the marker and gene need to be calculated (Hartl and Clark, 1997). As mentioned earlier, the probability of population-wide linkage disequilibrium is higher for closely linked markers and in selected populations of small effective size, but population-wide linkage disequilibrium can also be created between loosely linked markers for several generations by crossing inbred lines (Jansen, 2001).

If a sufficient number of genetic markers, which fulfill the aforementioned requirements, are identified in the genome, marker maps can be established and individuals can be selected on the basis of their marker genotypes, providing a vast array of benefits, and giving new impetus to, selective breeding programmes. The selective breeding strategy, however, will depend on the type of marker system used.

1.3.4 Types of genetic markers

The major challenge that faces molecular geneticists is to identify markers for genes that control the phenotypic variation in the traits of interest (Williams, 2005). Direct markers for quantitative traits are hard to find because the probability of a marker occurring inside the gene of interest is relatively low (Dekkers, 2004). In contrast, non-functional polymorphisms are abundant across the genome and their linkage with genes of interest can be established by evidence of empirical associations of marker genotypes with particular phenotypes (van der Werf, 2000a).

Direct markers are generally preferred to linked markers, because they can be used without trait measurement or pedigree recording (Kinghorn, 2000). Furthermore, once

the functional polymorphism is known, it is possible to predict the effect of particular alleles in all individuals in a population, without first having to determine the linkage phase (Williams, 2005). However, not withstanding the convenience of direct markers, there are currently only a few known direct genetic markers for economically important traits in animal breeding (Brascamp *et al.*, 1993). One example is the double muscling marker in cattle (giving increased muscle mass) which has been found to be the myostatin gene (Dekkers, 2004).

Although highly convenient, direct markers do pose problems. Firstly, there could be more than one mutation causing the desired genetic effect and DNA testing for only one of those mutations would not pick up all the individuals with the desired effect (van der Werf, 2000a). This could produce false negatives in diagnostic tests. Furthermore, exclusive utilization of direct markers could incorrectly identify a gene as a major gene directly affecting the trait of interest, because of its location near the true causative gene (Williams, 2005). This could produce false positive results and highlights the value of re-evaluating markers and the continuous need for trait recording for monitoring purposes (Dekkers, 2004).

Markers in linkage disequilibrium with genes of interest are more widely used in selection programmes (Kinghorn, 2000). However there is always a chance of recombination occurring between the marker and the locus of interest, which poses a real problem because one can never be sure which marker variant is associated with each gene variant in an animal (van der Werf, 2000a). Thus pedigrees and trait measurements need to be recorded every generation in order to successfully work with indirect markers (Dekkers, 2004).

The information provided by indirect genetic markers will vary depending on the type of marker system used, with each having its own advantages and disadvantages.

The earliest form of DNA markers, used to construct the first true genomic maps, were restriction fragment length polymorphisms or RFLPs (Meuwissen and van Arendonk, 1992). The RFLP technique has been used to screen for carriers of genetic defects such as bovine leukocyte adhesion deficiency (BLAD) in Holstein cattle (Williams, 2005). However this technique is time-consuming and expensive and the

reduced variability observed in domestic species due to inbreeding, makes many RFLPs sites non-informative (Montaldo and Meza-Herrera, 1998).

Other types of DNA markers, used mainly in plant breeding, but which have potential for use in animal breeding, are randomly amplified polymorphic DNA (RAPD) and amplification fragment length polymorphisms (AFLP) (Montaldo and Meza-Herrera, 1998). While analysis for RAPD markers is quick and simple, the results are sensitive to laboratory conditions (Hartl and Clark, 1997). On the other hand, AFLP markers allow for selective amplification of restriction fragments, giving rise to large numbers of useful markers, which can be located on the genome relatively quickly and reliably (Falconer, 1989). While neither of these marker systems have the high degree of variability needed for a completely successful marker system, practically they provide a quick and easy way to screen for markers because only a few different forms of the marker exist.

In contrast, minisatellites provide a high degree of variability (Kinghorn, 2000). Minisatellites are tandemly repeated sequences, typically 20 to 50 base pairs in length, which may occur at 10 to 100 different sites in the genome (Snustad and Simmons, 2000). Variations in the number of repeat sequences present at a particular locus give rise to a large number of alleles and these highly variable markers have been used to identify relationships between individuals in wild populations, to verify pedigrees in many species and in genetic mapping studies (Williams, 2005).

The development of microsatellites followed the development of minisatellites (Barnett, 2000). Like minisatellites, microsatellites are also short tandem repeat sequences of DNA, but are usually only 2 or 3 bases long (Kinghorn, 2000). Since the number of repeat units varies between individuals, a large number of microsatellite marker variants are generated for a given locus in a population (Rothwell, 1993). Microsatellites are excellent genetic markers because they are easy to detect with the polymerase chain reaction and have more variants than those from other marker systems (Barnett, 2000). Therefore microsatellite markers provide the power to determine associations between the marker variant, the contributing allele and the performance of the progeny inheriting a favourably linked gene (van der Werf, 2000a).

In recent years, single nucleotide polymorphisms (SNPs) have become an increasingly important class of molecular marker (Abecasis *et al.*, 2005). Single nucleotide polymorphisms (SNPs) are single base changes in a DNA sequence, which occur at a high frequency (approximately every 200 base pairs) in both non-coding and coding regions of the genome (Rothwell, 1993). Due to the high potential number of SNP markers, micro-array procedures have been developed for automatically scoring hundreds of SNP loci simultaneously at a low cost per sample, making them invaluable genetic markers (Montaldo and Meza-Herrera, 1998).

Knowledge of the alleles at particular genetic loci will enable the identification of individuals that carry beneficial alleles and allow for the direct selection of genetically superior animals at several loci simultaneously (van der Werf, 2000a). A strategy to select for improved performance in a number of traits could thus be developed using genetic markers, even when, at the phenotypic level, the traits may seem to be in conflict because they are negatively correlated at the genotypic level (Hospital *et al.*, 1997).

The molecular marker systems described above, have allowed for high density DNA marker maps to be constructed for a range of economically important agricultural species, thus providing the framework needed for the eventual applications of marker-assisted selection (Williams, 2005). Using these marker maps, putative genes affecting traits of interest, have been detected by testing for statistical associations between marker variants and any trait of interest (Doerge, 2002).

Various studies have shown how genetic markers have been used to map loci affecting quantitative traits of economic importance (Weller *et al.*, 1990). Using these maps, marker-assisted selection has the potential to enhance breeding efforts through the genetic improvement of populations (Asins, 2002). However, in order to fully exploit the potential of marker-assisted selection, it is therefore necessary to analyze the factors affecting genetic gain, so that these factors can be optimized for a maximum response to marker-assisted selection (van der Werf, 2000b).

1.3.5 Response to marker-assisted selection

The aim of any breeding strategy is to select for individuals in a population with the best genotype or breeding value, in order to increase the response to selection by increasing the frequency of favourable genes (Butterfield *et al.*, 2001). Genomic response to selection is the change in allele frequency at specific loci as a result of selection for a trait (Gomez-Raya, 2002). The response to selection (R) is a measurement of genetic gain from selective breeding programmes and can be calculated as follows (Falconer, 1989):

$$R = \frac{\textit{intensity of selection} \times \textit{accuracy of selection} \times \textit{additive genetic standard deviation}}{\textit{generation interval}}$$

Accordingly, R is dependent on the intensity and accuracy of selection, the additive component of the genetic standard deviation and the generation interval in a population. These four parameters therefore need to be fully understood in order for them to be manipulated in such a way that the maximum response can be achieved (Falconer and Mackay, 1996).

Intensity of selection (*i*) is a prediction of the superiority of the selected group and is expressed in terms of the phenotypic standard deviation (σ_P) of a particular population (van der Werf, 2000b). Selection intensity is a function of the proportion of individuals used as parents, such that *i* increases as the proportion of individuals selected as parents decreases (Gillespie, 2004). Decreasing the number selected individuals in a breeding population will therefore increase *i* and consequently R.

The accuracy of selection is an indication of how well the estimated genetic value reflects the true genetic value in a population (van der Werf, 2000b). Accordingly, accuracy of selection is often abbreviated as h, because it can be calculated as the square root of the heritability estimate, h^2 (Falconer and Mackay, 1996). Traits with a high h will show greater response to selection.

The response to selection can also be increased by increasing the additive component of genetic standard deviation (σ_A). The additive component of genetic standard

deviation, calculated as the square root of the additive variance (V_A), is a measure of the amount of additive genetic variation in a population (Hartl and Clark, 1997). Since variation has to be present in a population for selection to be successful, R is directly dependent on σ_A.

In contrast, R is indirectly related to generation interval (L), which is the average age of the parents when their progeny are born (Lynch and Walsh, 1998). Although marker-assisted selection does not affect the age at which individuals produce progeny, it does allow for selection to be carried out on individuals in the population at a much earlier age, thereby decreasing L and increasing R (van der Werf, 2000b).

It is important to note that these parameters are all inextricably linked and therefore cannot be viewed in isolation. For example, decreasing L in order to improve R would result in a decrease in h, because fewer progeny records would be available for genetic prediction (Lynch and Walsh, 1998). Thus, if h is significantly decreased, lessening L could have an adverse effect on R (van der Werf, 2000b).

Genetic markers, however, have the unique potential to optimize the response to selection by minimising the effect of the relationships between the parameters of interest. An improved response to selection due to the use of genetic markers, is most often as a result of increased accuracy of selection (Meuwissen and van Arendonk, 1992). As explained above, conventional schemes, such as progeny testing, also have a high accuracy of selection, but at the same time, these schemes have long generation intervals which have an adverse effect of genetic gain (Falconer, 1989). In contrast, marker-assisted selection schemes could be designed to decrease the L through early selection, while still maintaining a high h, thereby ensuring maximum genetic gain (Dekkers, 2004).

The extent to which marker-assisted selection can improve the response to selection, however, is also dependent on the amount of trait variation that may be accounted for by the marker (Meuwissen and Goddard, 1996). This is significant as economically important traits are often controlled by chromosomal segments, known as quantitative trait loci (QTL), which potentially encompass many hundreds of loci that contain genes that contribute towards the phenotypic trait of interest (Mauricio, 2001).

Although the term 'QTL' strictly applies to gene sequences which account for any proportion of the total phenotypic variation in the trait of interest, in practice it refers only to major genes with effects large enough to be detected (Dekkers, 2004). Genes which only have a minor effect on the phenotypic variance of the trait are termed polygenes and variation at these polygenes, together with polymorphisms at the QTL, determine the total genetic variation (van der Werf, 2000a). Thus, only if the genetic effects at QTL are large, can they be exploited in marker-assisted selection in the same way that single genes are, with the assumption that the additive variance of the trait of interest will largely be explained by QTL linked to marker loci (Jansen, 2001).

1.3.6 Benefits of marker-assisted selection

Under specific conditions, marker-assisted selection is especially valuable when compared to other methods of selection. The following five circumstances reveal the unique benefits of marker-assisted selection.

1. Marker-assisted selection can be extremely beneficial when the heritability of a trait is low, because the value of the information on the genes involved, increases the accuracy of the estimation of breeding values and makes selection more effective (Hospital *et al.*, 1997).

2. In instances where the trait(s) of interest cannot be measured on one sex, for example milk production in sires, marker information gives a basis to rank the animals of that sex (Zhang and Smith, 1992). This is particularly useful when deciding which males should be progeny tested (Lande and Thompson, 1990).

3. Marker information can be used to select at a juvenile stage if the trait is not measurable before sexual maturity, thus saving the breeder time and money that may have been invested in an animal with a poor breeding value (Williams, 2005).

4. Marker information can be utilized in order to bypass the difficulties associated with traits that require sacrifice, for example many carcass traits can only be measured after the animal has been killed (Soller and Beckman,

1988). The genetic information regarding such traits can potentially be measured anytime in an animal's life with the use of genetic markers, thereby avoiding destructive testing.

5. Marker-assisted selection can be useful in increasing the response to selection at loci where dominance is present (Kinghorn, 2000). When there is dominance at a particular locus, the genetic value of a heterozygote can be similar to the genetic value of a homozygote, even though their breeding values differ (Hartl and Clark, 1997). However, it is the breeding value of an individual, not the genetic value, which reflects the merit that can be transmitted to the next generation (Falconer, 1989). Since marker-assisted selection is carried out on the breeding values of individuals, the accuracy, and therefore the response to selection, will be improved (van der Werf, 2000b).

Deterministic analysis indicates that molecular marker loci can be used to substantially increase the rate of genetic gain in quantitative characters by marker-assisted selection (Lande and Thompson, 1990). However, many details regarding the optimal use of marker-assisted selection in long-term selection programmes remain to be determined (Dekkers, 2004). In this respect, computer simulations could be developed in order to accurately predict the response to marker-assisted selection over a number of generations (Montaldo and Meza-Herrera, 1998). Furthermore, computer simulations could be used to maximise the components that increase genetic gain and to consequently improve the overall response to selection.

1.4 COMPUTER SIMULATIONS

A computer simulation is a representation or model of a real system, situation or phenomenon and, in most cases, this representation is simplified or scaled down (Lunce, 2004). The purpose of a simulation model is to mimic the real system so that its behaviour can be simplified into readily understandable, functional components and studied (Partner *et al.*, 1993).

Effective computer simulations are built upon mathematical models which accurately depict the phenomena or processes being studied (Lunce, 2004). Simplifying assumptions are often used within a quantitative genetics framework, and while attitudes to these assumptions are variable, it is generally accepted that they are necessary in order to render many of the problems tractable to some form of algebraic solution (Podlich and Cooper, 1998). In designing simulation models, the involvement of statistical and mathematical equations is therefore necessary to appropriately translate the concepts of the process into a mathematical context (Newby *et al.*, 2000). Thus, simulations may be used to predict how systems might behave under certain assumptions and to evaluate the behaviour of these systems (Montaldo and Meza-Herrera, 1998).

A model's assumptions consist of both the descriptions of a physical system and relevant decision making strategies (Yin *et al.*, 2003). In order for computer simulations to respond to change in the same way that a real system would, the model's assumptions should accurately describe the physical system being modeled (Podlich and Cooper, 1998). When describing the physical system, computer simulations should therefore be designed to handle perturbations, like the type observed in real life situations, even when given simple operational rules (Chapman *et al.*, 2003). Moreover, the model should reflect the actual decision-making strategies used by the people in the system being modeled, including the limitations and errors of those strategies (Sterman, 1991).

The developers and users of models, the decision-makers using information derived from models and people affected by decisions based on such models are all concerned with whether a model and its results are accurate (Sargent, 1992). Owing to the versatility of computer simulations, different criteria have to be considered when assessing the acceptability of different simulation studies (Oren and Zeigler, 1979).

Since computer models have been developed for specific purposes, their validity should be determined with respect to the purpose or goal of the study (Tuncer, 1981). In other words, the acceptability of simulated data has to be considered with respect to real system data that has been collected under similar conditions (Oren and Ziegler, 1979). If the simulated data is acceptable, then the model used is said to be valid

(Newby *et al.*, 2000). Since it is often too costly and time consuming to determine if a model is absolutely valid, tests and evaluations should be conducted until sufficient confidence is obtained that a model can be considered valid for its intended applications (Sargent, 1992).

A major advantage of computer simulations is that they provide a method for checking our understanding of the real world and facilitating an interactive practice of real-world skills through focusing on essential elements of a real system (Newby *et al.*, 2000). Spreadsheets, for example, make it possible to concentrate on real-world problems, without becoming bogged down in the calculations, and are therefore useful in both the research and educational sectors.

Computers can manage multiple variables simultaneously, and since these variables can be manipulated in order to observe their effects on the system being modeled, computer simulations have the potential to realistically depict complex phenomena (Newby *et al.*, 2000). The power of computer simulations can therefore be to understand a system through considering alternative scenarios and seeing how the outputs are altered by changing the structure or the inputs and enabling different scenarios to be compared (Lunce, 2004). By creating a representation of the real life system on the computer, a modeler can perform experiments that are impossible, unethical, or prohibitively expensive in the real world (Sterman, 1991). The computer can also contract or expand time to allow study of phenomena that are too slow or too fast for normal observation (Newby *et al.*, 2000).

Furthermore, computer models have the ability to dynamically present information such that the changes which affect real systems can be simulated and the relevant processes can be graphically depicted (Partner *et al.*, 1993). This makes them a powerful tool for analysing, designing and interacting with complex systems as well as in education (Yin *et al.*, 2003).

In education, computer simulations can be useful in relating graphical representations to the real-world situations which they characterize, and can enable learners to manipulate these representations to work out differences between their inaccurate mental models and the formal principles pertaining to the system (Good and Berger,

1998). In this way, computers potentially allow learners to function at levels that their cognitive system would otherwise not have allowed them to reach and may aid students in their understanding of difficult concepts (Berger *et al.*, 1994). For abstract genetics concepts such as linkage disequilibrium and marker-assisted selection, computer simulations may therefore be useful in achieving learning outcomes that traditional approaches to learning may not (Tinker and Mather, 1993).

Since computers have become faster, cheaper and more widely available, computer simulations have become a useful tool for analyzing the potential applications of marker-assisted selection (Frisch *et al.*, 2000). Deterministic simulation models are able to calculate and compare the response to selection across different breeding strategies without the time and money involved in the traditional trial and error approach (Montaldo and Meza-Herrera, 1998). Furthermore, important parameters such as linkage disequilibrium coefficients and heritability estimates can be calculated automatically for each population with differing genotypic frequencies. It is therefore clear that agricultural breeding programmes involving the utilization of linkage disequilibrium for marker-assisted selection could benefit through the use of computer simulations.

1.5 AIMS OF THE INVESTIGATION

The aims of this investigation were to:

1. Simulate the effects of linkage disequilibrium between two loci in a randomly mating population.
   The associations between two loci at different recombination fractions were compared (Chapter 3, section 3.2).
2. Simulate the effects of both phenotypic selection and marker-assisted selection with the assumption of no dominance at the loci of interest. The response to phenotypic selection and marker-assisted selection, and the effects of both these selection strategies on variation and the amount of linkage disequilibrium in the population, was examined over five generations (Chapter 3, section 3.3.1 and 3.3.2).

3 Compare the efficiency of phenotypic selection and marker-assisted selection, with the assumption of no dominance at the loci of interest, in improving population mean values.

Over five generations, the response to phenotypic selection and marker-assisted selection was compared at different recombination fractions, when the percentage of trait variation accounted for by the QTL was varied during marker-assisted selection (Chapter 3, section 3.3.3).

4 Simulate and compare the effects of phenotypic selection and marker-assisted selection, under the assumption of complete dominance at the loci under investigation.

The response to one generation of phenotypic selection and marker-assisted selection was calculated and compared at different recombination fractions when the percentage of trait variation accounted for by the QTL was varied during marker-assisted selection (Chapter 3, section 3.4).

5 Mimic the effects of linkage disequilibrium marker-assisted selection, utilizing both a direct and an indirect marker, in two different Holstein cattle populations.

Data from the different Holstein populations was incorporated into the model developed to simulate marker-assisted selection and the response was measured over five generations (Chapter 3, section 3.5).

CHAPTER 2
MATERIALS AND METHODS

2.1 INTRODUCTION

2.1.1 Assumptions made in the formulation of models

Microsoft Excel 2003 was used in this investigation to develop six different computer models. The models are simplified representations of real breeding populations. The following assumptions were made in the development of all the computer models:

- The population size remained constant over time. The reason for this is that every individual was used once as a male parent and once as a female parent and one offspring was produced from each mating.
- The B locus, which was the locus under selection, was linked to a single gene marker at the A locus and each of these loci had two alleles designated A_1, A_2, B_1 and B_2. The association of the alleles along the parental chromosome is illustrated in Figure 1.2.
- The marker at the A locus was a co-dominant marker.
- Environmental conditions remained constant over time.
- No genotype-environment interaction.
- No epistasis.

2.1.2 Description of models formulated in the investigation

The name and purpose of each of the six computer models developed in this investigation, as well as the specific assumptions made in the formulation of each model, are shown in Table 2.1.

Table 2.1 The name, purpose and assumptions pertinent to each of the six computer models.

Name of the Model	Purpose of the Model	Assumptions Made in the Development of the Model
1. Random Mating	To simulate the effects of random mating and linkage disequilibrium.	• Random Mating • Additive gene action at the B locus
2. Phenotypic Selection	To simulate the effect of phenotypic selection	• Non-random mating • Additive gene action at the B locus
3. Marker-Assisted Selection	To simulate the effect of marker-assisted selection utilizing linkage disequilibrium	• Non-random mating • Additive gene action at the B locus
4. Selection with Dominance	To simulate the effects of both phenotypic selection and marker-assisted selection.	• Non random mating • Complete dominance at the B locus ($B_1 > B_2$)
5. Direct Selection	To mimic the effects of marker-assisted selection for an increase in milk yield utilizing a direct marker in a Holstein cattle population.	• Non-random mating • Additive gene action at the B locus • B_1 allele synonymous with QTL associated with an increase in milk yield
6. Indirect Selection	To mimic the effects of marker-assisted selection for an increase in milk fat percentage utilizing an indirect marker in a Holstein cattle population	• Non-random mating • Additive gene action at the B locus • B_1 allele synonymous with QTL associated with an increase in milk fat percentage

2.1.3 Motivation for use of Microsoft Excel for development of the models

Microsoft Excel was used to simulate the effects of both a two locus system in linkage disequilibrium and the response to different selection strategies. Microsoft Excel was used to build these models for a number of different reasons. Firstly, Microsoft Excel is a programme which is very accessible to most students, educational institutions and companies. Therefore, if the models developed in this investigation were utilized, the users would incur no added expense and would not have to install a specialised programme.

Secondly, many people are able to access and work in Microsoft Excel without

extensive prior instruction, as it is a relatively easy programme to utilize. However, despite its relative simplicity compared to other more complex simulations and databases, Microsoft Excel still allows for the accurate construction of genetic simulations according to mathematical formulae and assumptions. Therefore Microsoft Excel was considered ideal for the purposes of this investigation.

2.2 MODEL OF A RANDOMLY MATING POPULATION IN LINKAGE DISEQUILIBRIUM

2.2.1 Generation of population

A file was opened in Microsoft Excel and saved as 'Random Mating'. On Sheet 1, a population of 1000 individuals was generated according to the design laid out by Donovan and Welden (2002).

In order to simulate the genotypes of 1000 individuals in a population, as well as the gametes they produced, column headings for Genotype, Frequency and Tally count were created (Figure 2.1). The heading 'Genotype' was entered into cells A2-A4, which were merged, and the nine possible genotypes were entered into cells A5-A13. The heading 'Frequency' was entered into the merged cells B2-B4 and 'Tally count 0' was entered into the merged cells C2-C4. In cells A16, B16, C16 and D16, the headings 'Individual', 'Random #', 'Genotype' and 'Gamete' were entered respectively.

	A	B	C	D
1	Random Mating			
2				
3	Genotype	Frequency	Tally count	
4			0	
5	A1A1B1B1			
6	A1A1B1B2			
7	A1A1B2B2			
8	A1A2B1B1			
9	A1A2B1B2			
10	A1A2B2B2			
11	A2A2B1B1			
12	A2A2B1B2			
13	A2A2B2B2			<= This number MUST = 1
14				
15				
16	Individual	Random #	Genotype	Gamete

Figure 2.1 The range A1-D16 in Sheet 1 of the Microsoft Excel file 'Random Mating.'

It should be noted that the genotypic frequencies must be manually entered by the user into cells B5-B13 as the genotypes of each individual in the population depend on the proportions specified by these cells. However, in order to check that the sum of the genotypic frequencies in the population was always equal to one, a tally count was set up to automatically calculate the sum of the genotypic frequencies entered into the cells B5-B13. To calculate the 'Tally count' the following formulae were entered:

- In cell C4, the number 0 was entered in order to show that the frequency count began at 0.
- In cell C5, the formula **=B5** was entered
- In cell C6, the formula **=SUM($B$5:B6)** was entered and copied down to cell C13.

Finally, the heading 'This number MUST = 1' was entered into cell D13 and highlighted in red since the user needs to be aware that the running tally of the genotypic frequencies in cell C13 should always equal 1.

The following steps were taken to assign a random value to each of the 1000 individuals in the population:

- In cell A17, the number 1 was entered
- In cell A18, the formula **=A17+1** was entered and copied down to cell A1016 to make a population of 1000 individuals.
- The formula **=RAND()** was entered in cell B17 and copied down to cell 1016.

If required the F9 key could be pressed so that the spreadsheet model generates a new set of random numbers.

In order to calculate the allele frequencies and to determine the probability of finding each of the gamete types, tables were set up as shown in Figure 2.2.

	D	E	F	G	H	I
1						
2						
3		Allele frequencies				
4		Locus 1		Locus 2		
5		p_1 = A1 =		p_2 = B1 =		
6		q_1 = A2 =		q_2 = B2 =		
7						
8		Gamete probabilities				
9		A1B1	A1B2	A2B1	A2B2	
10	Expected					
11	Observed					
12						

Figure 2.2 The range D3-H11 in Sheet 1 of the Microsoft Excel file 'Random Mating.'

In order to determine the frequency of the A_1, A_2, B_1 and B_2 alleles in the population, the formulae provided in Table 2.2 were entered into the appropriate cells.

Table 2.2 The cell numbers and formulae used to calculate the different allele frequencies.

Allele	Cell	Formula entered
A_1	F5	=(COUNTIF(C17:C1016,'A1A1*')*2+COUNTIF(C17:C1016,'A1A2*'))/(2*A1016)
A_2	F6	=1-F5
B_1	H5	=(COUNTIF(C17:C1016,'B1B1*')*2+COUNTIF(C17:C1016,'B1B2*'))/(2*A1016)
B_2	H6	=1-H5

To clarify which of the gamete frequencies, shown by the lower table in Figure 2.2, were expected based on Hardy-Weinberg Equilibrium and which were observed in the population, the following headings were entered:

- In cell D10, the name 'Expected' was entered.
- In cell D11, the name 'Observed' was entered.

In order too calculate the expected and the observed gametic proportions, the following formulae were entered into the cells shown in Table 2.3.

Table 2.3 The cell numbers and formulae used to calculate the expected and observed gametic probabilities.

Gamete (Expected)	Cell	Formula
A1B1	E10	=F5*H5
A1B2	F10	=F5*H6
A2B1	G10	=F6*H5
A2B2	H10	=F6*H6
Gamete (observed)	**Cell**	**Formula**
A1B1	E11	=COUNTIF($D17:$D1016,E9)/1000
A1B2	F11	=COUNTIF($D17:$D1016,F9)/1000
A2B1	G11	=COUNTIF($D17:$D1016,G9)/1000
A2B2	H11	=COUNTIF($D17:$D1016,H9)/1000

To simulate random mating, headings were set up as shown in Figure 2.3. The formulae used to simulate random mating were entered into the columns headed 'Random Mom', 'Mom Gametes', 'Random Dad' and 'Dad Gametes. For example, a random number, taken from the range A17-A1016 (Figure 2.1, Page 32), was generated in the column headed 'Random Mom' and then the corresponding gamete, taken from the range D17-D1016 (Figure 2.1, Page 32), was generated into the

column headed 'Mom Gametes.' The formulae used to randomly combine one allele from each parent at the A and B loci were entered into the columns headed Locus A and Locus B respectively. The offspring genotypes were generated in the column headed 'Genotype'. Tables were also set up to display the observed and expected genotypic frequencies in the population and 'Linkage disequilibrium coefficient (***D***) =' was written below these tables.

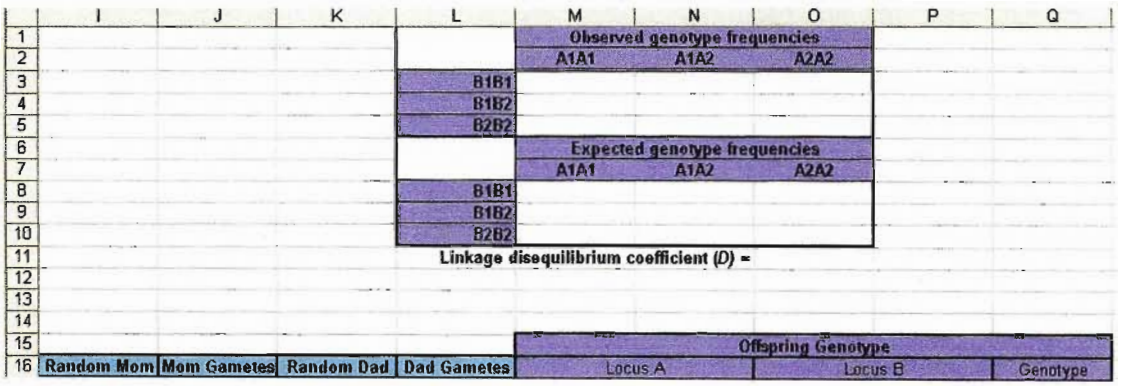


Figure 2.3 The range I1-Q16 on Sheet 1 of the 'Random Mating' computer model.

The following functions were used to select random parents in the population and determine their gametes from those that were generated in column D17-D1016:

- The formula **=ROUNDUP(RAND()*1000,0)** was entered in cell I17 and K17 and copied down to H1016 and J1016, respectively.
- In cell J17, the formula **=VLOOKUP(H17,$A$17:$D$1016,4)** was entered and copied down to J1016.
- In cell L17, the formula **=VLOOKUP(J17,$A$17:$D$1016,4)** was entered and copied down to L1016.

The gametes of the two random parents were then combined to produce an offspring genotype for both the A and B loci as follows:

- In cell M17, the formula **=LEFT(J17,2)&LEFT(L17,2)** was entered and copied down to M1016.
- In cell O17, the formula **=RIGHT(J17,2)&RIGHT(L17,2)** was entered and copied down to O1016.

A Microsoft Excel spreadsheet interprets an A_1A_2 individual as having a different genotype from an A_2A_1 individual, when they are in fact the same. Therefore the

following formulae were needed to ensure that A_1A_2 and A_2A_1 individuals would be interpreted as having the same genotype:

- In cell N17, the formula **=IF(M17='A1A2','A1A2',M17)** was entered and copied down to cell N1015.
- In cell P17, the formula **=IF(O17='A1A2','A1A2',O17)** was entered and copied down to cell P1016.

The genotypes at the A and B loci were combined to produce the offspring genotype when considering the two loci simultaneously as follows:

- In cell Q17, the formula **=N17&P17** was entered and copied down to P1016.

To calculate the observed genotypic frequencies as well as the expected genotypic frequencies (under the assumption of Hardy-Weinberg equilibrium) of the offspring generation, the following formulae were entered into the tables shown in Figure 2.3.

Table 2.4 The formulae entered into each cell to calculate the observed and expected genotypic frequencies.

Genotype (observed)	Cell	Formula Entered
A1A1B1B1	M3	=COUNTIF($M$17:$M$1016,'A1A1B1B1')/1000
A1A1B1B2	M4	=COUNTIF($M$17:$M$1016,'A1A1B1B2')/1000
A1A1B2B2	M5	=COUNTIF($M$17:$M$1016,'A1A1B2B2')/1000
A1A2B1B1	N3	=COUNTIF($M$17:$M$1016,'A1A2B1B1')/1000
A1A2B1B2	N4	=COUNTIF($M$17:$M$1016,'A1A2B1B2')/1000
A1A2B2B2	N5	=COUNTIF($M$17:$M$1016,'A1A2B2B2')/1000
A2A2B1B1	O3	=COUNTIF($M$17:$M$1016,'A2A2B1B1')/1000
A2A2B1B2	O4	=COUNTIF($M$17:$M$1016,'A2A2B1B2')/1000
A2A2B2B2	O5	=COUNTIF($M$17:$M$1016,'A2A2B2B2')/1000
Genotype (expected)	**Cell**	**Formula Entered**
A1A1B1B1	M8	=E10*E10
A1A1B1B2	M9	=E10*F10+F10*E10
A1A1B2B2	M10	=F10*F10
A1A2B1B1	N8	=E10*G10+G10*E10
A1A2B1B2	N9	=E10*H10+H10*E10+F10*G10+G10*F10
A1A2B2B2	N10	=F10*H10+H10*F10
A2A2B1B1	O8	=G10*G10
A2A2B1B2	O9	=G10*H10+H10*G10
A2A2B2B2	O10	=H10*H10

In order to determine the linkage disequilibrium coefficient (***D***), which represents the frequency of the product of the recombinant gametes (A_1B_2 and A_2B_1) minus the

frequency of the product of the parental gametes (A_1B_1 and A_2B_2), the following formula was used:

- In cell O11, the formula **=E11*H11-F11*G11** was entered.

To illustrate the association of the alleles at the A and B loci in a population, as a result of the amount of linkage disequilibrium between the two loci, a column graph was created using the following steps:

- Cells L1-O10 were selected.
- The chart wizard icon was selected and the bar graph option was used.

Once the initial population had been set up, no more of the design proposed by Donovan and Welden (2002) was used, such that all the following spreadsheet entries were entirely original.

2.2.2 Alteration of recombination fraction

To demonstrate the effects of recombination on linkage disequilibrium between two loci in a population, it was necessary to generate gametic proportions that reflected the given recombination fraction (r). To accomplish this, genotypes were randomly assigned to every individual in the population, based on the genotypic frequencies that may be manually entered in cells B5 to B13, as follows:

- In cell C17, the formula **=LOOKUP(RAND(),$C$4:$C$13,$A$5:$A$13)** was entered and copied down to C1016.

In order to simulate the effect of r on the proportion of recombinant gametes (shown in Figure 2.2, Page 33), it was necessary for each individual in the population to generate gametes based on their genotype and the recombination fraction (r). This was accomplished as follows:

- In cell D1, 'Recombination Fraction =' was entered and highlighted in pink.
- In cell D17, the following formula was entered and copied down to D1016.
  =IF(B17<($E$1/2),MID(C17,3,4),IF(B17<($E$1),LEFT(C17,2)&RIGHT(C17,2),IF(B17<($E$1+(1-E$1)/2),LEFT(C17,2)&MID(C17,5,2),MID(C17,3,2)&RIGHT(C17,2))))

The recombination fraction was entered into cell E1, which was also highlighted in pink. This figure would be changed for different populations.

2.2.3 Estimation of individual values and population statistics

In order to make this computer model applicable to an actual breeding population, it was necessary to assign breeding values (BV) to each of the individuals, as well as an environmental effect. It was also essential to set up a table that calculated population statistics such as the mean, additive variance, environmental variance, phenotypic variance and heritability.

The first step was to create the population statistics table as shown in Figure 2.4. These statistics included the population mean (μ), the additive variance (V_A), the environmental variance (V_E), the phenotypic variance (V_P) and the heritability (h^2) for generation 1 and 2.

	Q	R	S	T	U	V	W
1		Population Statistics					
2		μ (gen 1) =		μ (gen 2) =			
3		V_G (gen 1) =		V_G (gen 2) =		Average Gen	50.000
4		V_E (gen 1) =		V_E (gen 2) =		h^2 =	
5		V_P (gen 1) =		V_P (gen 2) =		h^2 =	
6		Environmental Heterogeneity =			0.300		

Figure 2.4 The range R1-W6 in Sheet 1 of the 'Random Mating' computer model.

The average phenotypic value and the environmental heterogeneity must be entered by the user into the table manually and can therefore be altered for different populations.

In order to assign a value to the B_1 and B_2 alleles, a table was set up in cells E13-I14 as shown in Figure 2.5. Since the A locus in this investigation is considered to be the marker locus, the breeding value at this locus should be 0. Furthermore, the headings 'BV', 'G', 'E' and 'P' were entered into cells E16-H16 (Figure 2.5) to facilitate the calculation of breeding values, genotypic values, environmental values and phenotypic values, respectively, for each individual in the population.

	D	E	F	G	H	I
12						
13		Genotype	A1	A2	B1	B2
14		Value	0	0		
15						
16		BV	G	E	P	
17						

Figure 2.5 The range E13-I16 on Sheet 1 of the 'Random Mating' model. Any value for the B_1 or B_2 alleles could be respectively entered into cells H14 or I14.

The value of the alleles at the B locus was used to assign breeding values to each individual in the parental population based on the genotypic values given in the range C17-C1016, shown in Figure 2.5, as follows:

- In cell E16, the name 'BV' was entered and highlighted in yellow (Figure 2.5).
- The following formula was entered into cell E17 and copied down to E1016.
  =LOOKUP(MID(C17,1,2),$F$13:$I$13,$F$14:$I$14)+LOOKUP(MID(C17,3,2),$F$13:$I$13,$F$14:$I$14)+LOOKUP(MID(C17,5,2),$F$13:$I$13,$F$14:$I$14)+LOOKUP(MID(C17,7,2),$F$13:$I$13,$F$14:$I$14)

The genotypic value was then calculated for each individual in the population.

- In cell F16, the name 'G' was entered and highlighted in yellow (Figure 2.5).
- In cell F17, the formula **=E17+$W$3** was entered and copied down to F1016. This formula adds the breeding value to the average genotype entered into cell W3.

An environmental deviation was then randomly assigned to each of the 1000 individuals in the parent population.

- In cell G16, the name 'E' was entered and highlighted in yellow (Figure 2.5).
- In cell G17 the formula **=NORMINV(RAND(),0,$U$6)** was entered and copied down to G1016.

Note that cell U6, referred to by this formula, contains the environmental heterogeneity value which has to be manually entered into the computer model. (See Figure 2.4)

Once each parent had a breeding value and an environmental effect value, it was possible to calculate their phenotypic values according to the formula P= G+E (Hartl and Clark, 1997).

- In cell H16, the name 'P' was entered and highlighted in yellow (Figure 2.5).
- In cell H17, the formula **=F17+G17** was entered and copied down to H1016.

Breeding values, genotypic values, environmental values and phenotypic values then had to be assigned to the offspring generation (or generation 2) in an identical fashion. The same population statistics that were calculated for generation 1, also needed to be calculated for generation 2.

The breeding values for each individual in generation 2 were calculated as follows:

- In cell R16, the name 'BV' was entered and highlighted in purple.
- The following formula was entered into cell R17 and copied down to R1016.
  =LOOKUP(MID(Q17,1,2),$F$13:$I$13,$F$14:$I$14)+LOOKUP(MID(Q17,3,2),$F$13:$I$13,$F$14:$I$14)+LOOKUP(MID(Q17,5,2),$F$13:$I$13,$F$14:$I$14)+LOOKUP(MID(Q17,7,2),$F$13:$I$13,$F$14:$I$14)

Genotypic values were then calculated for each of the individuals in generation 2.

- In cell S16, the name 'G' was entered.
- In cell S17, the formula **=R17+$W$3** was entered and copied down to S1016.

Environmental deviations were then randomly assigned to individuals in generation 2.

- In cell T16, the name 'E' was entered and highlighted in purple.
- In cell T17 the formula **=NORMINV(RAND(),0,$U$6)** was entered and copied down to T1016.

The phenotypic values for generation 2 were calculated as follows:

- In cell U16, the name 'P' was entered and highlighted in purple.
- In cell U17, the formula **=Q17+R17+$W$3** was entered and copied down to U1016.

The formulae used to calculate the population statistics for generation 1 and 2 (displayed in the table shown in Figure 2.4) are shown in Table 2.5.

Table 2.5 The formulae entered into each cell to calculate the different population parameters

Statistic	Cell	Formula Entered
$\mu_{(gen1)}$	S2	=AVERAGE(H17:H1016)
$V_{A\,(gen1)}$	S3	=VAR(F17:F1016)
$V_{E\,(gen1)}$	S4	=VAR(G17:G1016)
$V_{P\,(gen1)}$	S5	=VAR(H17:H1016)
$\mu_{(gen2)}$	U2	=AVERAGE(U17:U1016)
$V_{G\,(gen2)}$	U3	=VAR(S17:S1016)
$V_{E\,(gen2)}$	U4	=VAR(T17:T1016)
$V_{P\,(gen2)}$	U5	=VAR(U17:U1016)
$h^2_{(gen1)}$	W4	=S3/S5
$h^2_{(gen2)}$	W5	=U3/U5

The results of Sheet 1 of the 'Random Mating' computer model were presented in the form of tables and graphs on Sheet 2 and 3 in the same model.

2.2.4 Presentation of results

The change in genotypic frequencies, ***D***, V_P and allele frequencies over five generations of random mating were presented in tables. To do this, a second Sheet was added to the 'Random Mating' computer model and tables were set up as shown in Figure 2.6.

	A	B	C	D	E	F
1		Generation 1	Generation 2	Generation 3	Generation 4	Generation 5
2	A1A1B1B1					
3	A1A1B1B2					
4	A1A1B2B2					
5	A1A2B1B1					
6	A1A2B1B2					
7	A1A2B2B2					
8	A2A2B1B1					
9	A2A2B1B2					
10	A2A2B2B2					
11	Table 1. The change in genotypic frequencies over five generations					
12						
13						
14		Generation 1	Generation 2	Generation 3	Generation 4	Generation 5
15	D					
16	V_P					
17	Table 2. Change in in the linkage disequilibrium coefficient (D) and phenotypic variance (V_P) over five					
18	generations					
19						
20		Generation 1	Generation 2	Generation 3	Generation 4	Generation 5
21	p_1 = A1 =					
22	q_1 = A2 =					
23	p_2 = B1 =					
24	q_2 = B2 =					
25	Table 3. Change in allele frequencies over five generations					
26						
27						

Figure 2.6 The range A1-I22 in Sheet 2 of the 'Random Mating' computer model.

A macro was then designed to:

- Automatically generate the genotypic frequencies of the parental population on Sheet 1 for five generations (shown in the top table).
- To generated D and V_P for each generation (shown by the second table).
- To generate the allele frequencies over five generations for both the A and B loci (shown in the bottom table).

This was done by selecting |Tools| |Macro| |Record new Macro|. The macro was named 'Random Mating' and the shortcut key was entered as <Ctrl r>. The values were then manually copied from Sheet 1 and pasted into the tables on Sheet two. When all the values were manually copied and pasted, the 'stop recording' button was selected and the macro was complete. An activation button was created by right clicking on the toolbar, selecting the 'forms' option and then choosing the rectangular button. 'Click Here to Activate Macro' was written on the button and then it was assigned to the random mating macro by right clicking the button and selecting the 'assign macro' option.

To facilitate the graphical depiction of the changes in genotypic frequencies, breakdown of ***D***, changes in allele frequencies and changes in V_P over five generations calculated on Sheet 2 (Figure 2.6), a third Sheet was added to the 'Random Mating' computer model.

In order to illustrate the change in genotypic frequencies of five generations of random mating, a graph entitled 'Change in Genotypic Frequencies over Five Generations of Random Mating' was created on Sheet 3:

- The chart wizard was selected and the column graph option was chosen.
- The range A1-F10 on Sheet 2 was selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'Frequency'.

The breakdown of the linkage disequilibrium coefficient over five generations of random mating, when r was 0.5, was illustrated in the graph headed 'Breakdown of ***D***' on Sheet 3. This graph was created as follows:

- The chart wizard option was selected and the XY (scatter) option was chosen.
- The range A14-F15 on Sheet 2 was selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled '***D***'.

The change in allele frequencies at both the A and B loci were graphically depicted by the graph entitled 'Change in Allele Frequencies over Five Generations of Random Mating' which was created on Sheet 3:

- The chart wizard was selected and the column graph option was chosen.
- The range A20-F24 on Sheet 2 was selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'Frequency'.

In order to illustrate the change in phenotypic variance over time, the graph entitled 'Change in V_P over Five Generations of Random Mating' was generated on Sheet 3:

- The chart wizard was selected and the XY (scatter) option was chosen.
- The ranges A14-F14 and A16-F16 on Sheet 2 were selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'Phenotypic Variance'.

2.3 MODELS ILLUSTRATING THE EFFECTS OF SELECTION WITH THE ASSUMPTION OF NO DOMINANCE

2.3.1 Creation of model illustrating effects of phenotypic selection

A file was opened in Microsoft Excel and saved as 'Phenotypic Selection'. The format of Sheet 1 of the 'Phenotypic Selection' file was kept the same as that laid out in Sheet 1 in the 'Random Mating' computer model. However, certain changes were made in order to simulate a population undergoing phenotypic selection.

(a) Alteration of 'Random Mating' model to account for phenotypic selection

To ensure that the sum of the genotypic frequencies in the population was always equal to 1, the formula **=1-(B5+B6+B7+B8+B9+B10+B11+B12)** was entered into cell B13. This formula calculates the frequency of the double homozygote ($A_2A_2B_2B_2$) by subtracting the sum of all the other genotypic frequencies from 1.

To simulate selection of individuals based on their phenotypic values, the headings 'Select above,' 'Select' and 'Select Value' were added to the lay out of the original 'Random Mating' model (Figure 2.5, Page 39), as shown in Figure 2.7.

	D	E	F	G	H	I	J
12							
13		Genotype	A1	A2	B1	B2	
14		Value	0	0			
15					Select above		
16		B locus BV	G	E	P	Select	Select Value
17							

Figure 2.7 The range E13-J16 in Sheet 1 of the 'Phenotypic Selection' computer model.

Phenotypic Selection was simulated through the following steps:

- In cell H15, the name 'Select above' was added and highlighted in pink.
- Cell I15 was highlighted in pink and left blank.

In this cell, a value could be manually entered such that only those individuals in the population with a phenotypic value higher than the value entered in this cell, would be selected for breeding. This was done as follows:

- In cell I17, the formula **=IF(OR(H17>$I$15),1,'')** was entered and copied down to I 1016.

- In cell J17, the formula **=IF(I17=1,H17,'')** was entered and copied down to J1016.

Therefore if an individual had a higher phenotypic value than the value in I15, it would be selected and its phenotypic value would be re-entered into column J17-1016. If the individual's phenotypic value was lower than the value in I15, it would not be selected and all the adjacent cells would be left blank.

The mating of the selected individuals was simulated in the same way as shown in the 'Random Mating' model, although different formulae were used to account for the blank cells in the rows of individuals that were not selected.

- In cell K17, the formula **=IF(L17='','',A17)** was entered and copied down to K1016.
- In cell L17, the formula **=IF(OR(I17=1),D17,'')** was entered and copied down to L1016.
- In cell M17, the following formula was entered and copied down to M1016:
  =IF(L17 = '','',(ROUNDUP(RAND()*$I$1018,0)))
- In cell N17, the following formula was entered and copied down to N1016:
  =VLOOKUP(M17,$K$17:$L$1016,2)

Once phenotypic selection of individuals in the population had been simulated, it was necessary to calculate population parameters necessary to estimate the response to phenotypic selection. These population parameters could then be compared and contrasted with those calculated using different selection techniques. In this way the response to each selection strategy, as well as the effect of each strategy on the genetic variation in a population, could be compared with the aim of establishing the most beneficial selection strategy overall.

(b) Calculation of population parameters

To present the afore-mentioned population parameters in an easily accessible way, a 'selection statistics' table was set up as shown in Figure 2.8. The population statistics that were displayed were the μ, the V_A, V_E, V_P and h^2 for generation 1 and 2. The intensity of selection, the mean of the selected group ($\mu_{selected}$), the selection differential (S) and the response to selection (R), were also displayed in the selection statistics table.

	R	S	T	U	V	W
1	Selection Statistics					
2	$\mu_{(gen\ 1)}$ =		Environmental Heterogenity =			
3	$V_{G\ (gen\ 1)}$ =		S =		Average Gen =	
4	$V_{E\ (gen\ 1)}$ =		$\mu_{(gen\ 2)}$ =		$h^2_{(gen1)}$=	
5	$V_{P\ (gen\ 1)}$ =		$V_{G\ (gen\ 2)}$ =		$h^2_{(gen2)}$=	
6	i =		$V_{E\ (gen\ 2)}$ =		R =	
7	$\mu_{(selected)}$ =		$V_{P\ (gen\ 2)}$ =			
8						

Figure 2.8 The range R1-W7 in Sheet 1 of the 'Phenotypic Selection' computer model.

In cells W2 and W3 (Figure 2.8), the environmental heterogeneity and the average genotypic value, respectively, could be manually entered by the user. The population statistics shown in Figure 2.8 were calculated using the formulae shown in Table 2.6.

Table 2.6 The formulae entered into each cell to calculate the different population statistics.

Statistic	Cell	Formula Entered
$\mu_{(gen1)}$	S2	=AVERAGE(H17:H1016)
$V_{A\ (gen1)}$	S3	=VAR(F17:F1016)
$V_{E\ (gen1)}$	S4	=VAR(G17:G1016)
$V_{P\ (gen1)}$	S5	=VAR(H17:H1016)
i	S6	=U3/SQRT(S5)
$\mu_{(selected)}$	S7	=AVERAGE(J17:J1016)
S	U3	=(S7-S2)
$\mu_{(gen2)}$	U4	=AVERAGE(W17:W1016)
$V_{G\ (gen2)}$	U5	=VAR(U17:U1016)
$V_{E\ (gen2)}$	U6	=VAR(V17:V1016)
$V_{P\ (gen2)}$	U7	=VAR(W17:W1016)
$h^2_{(gen1)}$	W4	=S3/S5
$h^2_{(gen2)}$	W5	=U5/U7
R	W6	=(W4*U3)

The number of selected individuals was then calculated and displayed as shown in Figure 2.9.

	F	G	H	I	J	W
1015	53	-2.302	50.698	1	50.698	57.134
1016	56	0.515	56.515	1	56.515	51.750
1017					297	297
1018	Total number of selected individuals			703	703	703

Figure 2.9 The range F1015-J1018 and W1015-W1018 in Sheet 1 of the 'Phenotypic Selection' computer model.

To ensure that the number of selected individuals stayed consistent throughout the mating process and to obtain the new population size, the following formulae were used:

- In cell I1018, the formula **=COUNTIF(I17:I1016,'1')** was entered to count the number of selected individuals.
- In cell J1017, the formula **=COUNTIF(J17:J1016,'')** was entered to count the number of blanks (which represent the non-selected individuals) in the column.
- In cell J1018, the formula **=1000-J1017** was entered to calculate the number of selected individuals in the population.
- In cell W1017, the formula **=COUNTIF(W17:W1016,'')** was entered to count the number of blanks in the column.
- In cell W1018, the formula **=1000-W1017** was entered to calculate the number of individuals in generation 2

Due to the selection process, the population size decreased over time. Consequently, the formulae used to calculate the observed genotypic frequencies in the 'Random Mating' model (shown in Figure 2.3, Page 35) needed to be altered. To determine the observed genotypic frequencies for the offspring generation, the formulae, shown in Table 2.7, were used. These formulae counted the number of times each genotype appeared in the offspring population and then divided the total number of individuals with each genotype by the number of selected individuals.

Table 2.7 The cell numbers and formulae used to calculate the observed genotypic frequencies for the offspring generation.

Cell	Formula Entered
M8	=COUNTIF($M$17:$M$1016,'A1A1B1B1')/W1018
M9	=COUNTIF($M$17:$M$1016,'A1A1B1B2')/W1018
M10	=COUNTIF($M$17:$M$1016,'A1A1B2B2')/W1018
N8	=COUNTIF($M$17:$M$1016,'A1A2B2B2')/W1018
N9	=COUNTIF($M$17:$M$1016,'A1A2B1B2')/W1018
N10	=COUNTIF($M$17:$M$1016,'A1A2B1B1')/W1018
O8	=COUNTIF($M$17:$M$1016,'A2A2B1B1')/W1018
O9	=COUNTIF($M$17:$M$1016,'A2A2B1B2')/W1018
O10	=COUNTIF($M$17:$M$1016,'A2A2B2B2')/W1018

Once the effects phenotypic selection on the population statistics, shown in Figure 2.8, had been calculated for two generations and the response to one generation of phenotypic selection had been determined, it was necessary to study the effects of phenotypic selection over five generations.

(c) Response to selection

The effects of phenotypic selection on population parameters such as μ, V_P, h^2, D, i and σ_P and the response to phenotypic selection over five generations were displayed in Figure 2.10. The change in genotypic frequencies over five generations, as well as the relationship between environmental heterogeneity values, heritability values and response to selection, were also shown in Figure 2.10. In order to display these results, tables were created in Sheet 2 of the 'Phenotypic Selection' computer model, as shown in Figure 2.10.

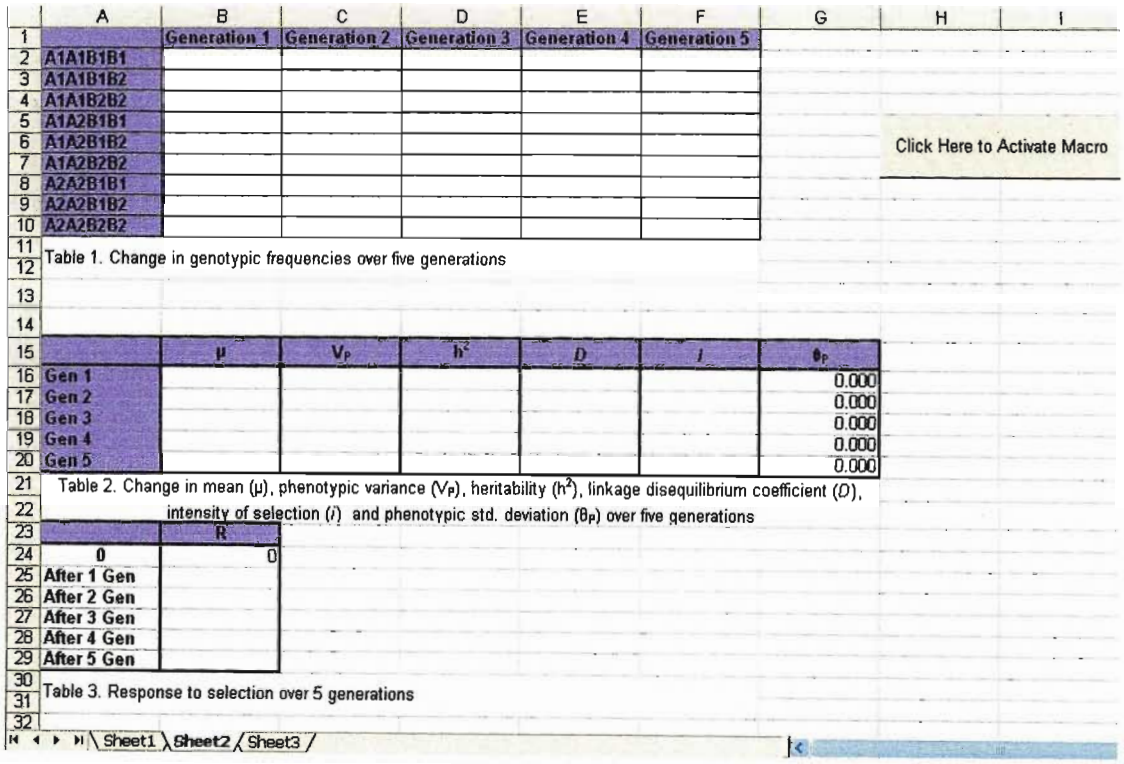


Figure 2.10 The range A1-I31 in Sheet 2 of the 'Phenotypic Selection' computer model.

A macro was designed to:

- Automatically generate the genotypic frequencies of the parental population on Sheet 1 for five generations (shown in Table 1 of Figure 2.10).
- To generate μ, V_P, h^2, D, i and σ_P over five generations (shown in Table 2 of Figure 2.10).
- To generate the R over five generations of phenotypic selection.

This was done by selecting |Tools| |Macro| |Record new Macro|. The macro was named 'PhenSelection' and the shortcut key was entered as <Ctrl p>. The values were then manually copied from Sheet 1 and pasted into the tables on Sheet two. When all the values were manually copied and pasted, the 'stop recording' button was selected and the macro was complete.

In order to make the interpretation of these results easier to understand, graphical representations of the significant relationships between certain population parameters were made.

(d) Graphical depictions of response to phenotypic selection

A third Sheet was added to the 'Phenotypic Selection' computer model in order to graphically depict the changes in genotypic frequencies, breakdown of ***D***, changes in allele frequencies and changes in V_P over five generations, shown in the tables on Sheet 2.

The response to phenotypic selection over five generations was graphically depicted on Sheet 3 by creating the graph entitled 'Response to Selection' as follows:

- The chart wizard was selected and the XY (Scatter) option was chosen.
- The range B24-B29 on Sheet 2 was selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'Response'.

To illustrate the relationship between σ_P and R, the following graph was named 'Relationship between Phenotypic Standard Deviation and Response to Selection' and generated on Sheet 3:

- The chart wizard was selected and the column graph option was chosen.
- The ranges G16-G20 and B25-B29 on Sheet 2 were selected as the data range.
- The X axis was labeled 'Generations'.

In order to demonstrate the effect of heritability on R, the following graph was created on sheet 3:

- The chart wizard was selected and the line graph option was chosen.
- The range D16-D20 and B25-B29 were selected as the data ranges.
- The X axis was labeled 'Generations.'

- The graph was entitled 'Relationship between Heritability and Response to Phenotypic Selection.'

2.3.2 Development of model illustrating effects of marker-assisted selection

A file was opened in Microsoft Excel and saved as 'Marker-Assisted Selection.' In order to simulate the effects of marker-assisted selection on a population, changes were made to Sheet 1 of the 'Phenotypic Selection' model and included in Sheet 1 of the 'Marker-Assisted Selection' model.

(a) Alterations to 'Phenotypic Selection' model to account for marker-assisted selection

The primary idea behind marker-assisted selection is that individuals in a population are selected on the basis of their marker value as opposed to their actual genotypic value (Williams, 2005). In this model, the A locus was used to represent the marker locus and therefore selection had to take place according to the value of alleles at the A locus.

In order to simulate the selection of individuals based on their marker genotype, headings were set up as shown in Figure 2.12. The table (E13-I14) in Figure 2.12 was used to assign a value to each marker variant. In cell I16, the name 'Marker Value' was entered and in this column, each individual was assigned a value according to their genotype at the A locus. Individuals selected based on their marker value were displayed in the column headed 'Select' and then their actual genotypic values (based on the B locus) were displayed in the column headed 'Select Value' (Shown in Figure 2.11).

	E	F	G	H	I	J	K
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13	Genotype	A1	A2	B1	B2		
14	Value	5	0	0	0		
15				Select above	0		
16	BV	G	E	P	Marker Value	Select	Select Value

Figure 2.11 The range E13-K16 in Sheet 1 of the 'Marker-Assisted Selection' computer model.

It is important to note the changes to the table (E13-I14) in Figure 2.11, when compared to the table (E13-I14) in Figure 2.7 (Page 44) showing Sheet 1 of the 'Phenotypic Selection' model. In cell F14 (Figure 2.11), the A_1 allele was given a value of 5 and in cell G14, the A_2 allele was given a value of 0. Both the alleles at the B locus were given values of 0. As mentioned earlier, the A locus was the marker locus and the B locus contained the gene(s) of interest. Selection in the 'Marker-Assisted Selection' model was therefore carried out on the basis of the marker genotype only, regardless of the individuals' breeding value at the B locus.

This investigation aimed to simulate the effects of marker-assisted selection utilizing linkage disequilibrium and therefore attention must be brought to the parental configuration shown in Figure 1.2 (Chapter 1, Page 7) whereby the A_1 allele is situated on the same chromosome as the B_1 allele and the A_2 allele is on the same chromosome as the B_2 allele. Thus, when linkage disequilibrium was present between the A and B loci, selection for the A_1 marker allele would more than likely result in the selection of the B_1 allele, which, as mentioned earlier, was the favourable allele at the locus containing the gene of interest.

The degree of association between the A_1 marker allele and the favourable B_1 allele would be dependent on the recombination fraction between the two loci in the population. This recombination fraction was manually entered into cell E1 of all the computer models and could be changed. A low recombination fraction would suggest

tightly linked loci and thus by selecting for the A_1 marker allele, the B_1 allele would have a very high probability of also being selected. This would be the ideal situation if marker-assisted selection is to effectively increase the response of the population to selection.

Any value could be given to the marker alleles by the user as long as the marker allele being selected had a higher value than the marker allele not being selected and the value placed in cell I15 (Figure 2.11, Page 51) corresponded to these values. For example, if the A_1 marker allele was given a value of 5 and the A_2 allele a value of 0 (as seen in Figure 2.12) and any individual with an A_1 marker allele was to be selected, then the value placed in cell I 15 should be between 0 and 4. This is because individuals with one or two A_1 alleles would have marker values of 5 and 10 respectively.

Selection based on the marker value was simulated using the following steps:

- In cell I16, the name 'Marker Value' was entered.
- In cell I17, the following formula was entered and copied down to cell I1016. **=LOOKUP(MID(C17,1,2),$F$13:$I$13,$F$14:$I$14)+LOOKUP(MID(C17,3,2),$F$13:$I$13,$F$14:$I$14)+LOOKUP(MID(C17,5,2),$F$13:$I$13,$F$14:$I$14)+LOOKUP(MID(C17,7,2),$F$13:$I$13,$F$14:$I$14**
- In cell J16, the name 'Select' was entered.
- In cell J17, the formula **=IF(OR(I17>$I$15),1,'')** was entered and copied down to J1016.

Once the individuals had been selected on the basis of their marker value, their phenotypic values needed to be re-entered into the next column as follows:

- In cell K16, the name 'Select Value' was entered.
- In cell K17, the formula **=IF(J17=1,H17,'')** was entered and copied down to K1016.

The selected individuals were then mated to each other to produce the offspring generation as shown in Figure 2.3 on Page 35.

(b) Estimation of individual values and population statistics

To calculate the population parameters, for example μ, V_A, V_E, V_P, h^2, i, S and R, a selection statistics table was set up as shown in Figure 2.8 (Page 46). The same formulae used to calculate the population statistics in the 'Phenotypic Selection'

model, as seen in Table 2.6 on Page 46, were used to calculate the populations statistics for the 'Marker-Assisted Selection' model.

The next step was to assign breeding values and genotypic values to the offspring generation. These values were based on two factors. Firstly, as seen in the 'Phenotypic Selection' model, it was necessary to assign values to the B_1 and B_2 alleles. Secondly, it was assumed that a Quantitative Trait Locus (QTL) rather than a single gene was located at the B locus. This assumption was made because most economically important traits are under the control of several different loci which each contribute to the variation of the trait. Therefore B_1 represents one of a number of different economically important alleles located adjacent to one another along the B locus. However, the amount of trait variation that is explained by a QTL such as B_1 varies depending on the trait and species. Therefore the "Marker-Assisted Selection' model was designed to incorporate this trait variation explained by the QTL into the genetic values of the offspring generation, in order to predict a more accurate response to selection. This was done by first setting up the headings as shown in Figure 2.12.

	U	V	W	X	Y
8					
9	Genotype	A1	A2	B1	B2
10	Value	0	0		
11					
12	%Trait VA explained by QTL				
13					
14					
15					
16	BV	G	E	P	

Figure 2.12 The range U9-Y16 in Sheet 1 of the 'Marker-Assisted Selection' computer model.

As with the 'Phenotypic Selection' model, the values of the B_1 and B_2 alleles had to be entered into cells X10 and Y10 respectively. The percentage trait variation explained by the QTL was entered into cell W12.

The following formulae were entered to calculate breeding values, genotypic values, environmental values and phenotypic values for the offspring generation:

- In cell U16, the name 'BV' was entered.

- In cell U17, the following formula was entered and copied down to U1016.
  =IF(T17="","",(LOOKUP(MID(T17,1,2),$V$9:$Y$9,$V$10:$Y$10)+LOOKUP(MID(T17,3,2),$V$9:$Y$9,$V$10:$Y$10)+LOOKUP(MID(T17,5,2),$V$9:$Y$9,$V$10:$Y$10)+LOOKUP(MID(T17,7,2),$V$9:$Y$9,$V$10:$Y$10)))
- In cell V16, the name 'G' was entered.
- In cell V17, the formula **=IF(U17="","",U17*($W$12)+$W$3)** was entered and copied down to V1016.
- In cell W16, the name 'E' was entered.
- In cell W17, the formula **=IF(U17="","",NORMINV(RAND(),0,$W$2))** was entered and copied down to W1016.
- In cell X16, the name 'P' was entered.
- In cell X17, the formula **=IF(W17="","",V17+W17)** was entered and copied down to X1016.

Once Sheet 1 of the 'Marker-Assisted Selection' model was completed, tables were set up in Sheet 2 to show the effect of marker-assisted selection on the different population parameters over five generations.

(c) Response to marker-assisted selection

In order to determine the response of the simulated population to marker-assisted selection, Sheet 2 of the file 'Marker-Assisted Selection' was opened and tables were set up as shown by Figure 2.13.

	A	B	C	D	E	F	G
1		Generation 1	Generation 2	Generation 3	Generation 4	Generation 5	
2	A1A1B1B1						
3	A1A1B1B2						
4	A1A1B2B2						
5	A1A2B1B1						
6	A1A2B1B2						
7	A1A2B2B2						
8	A2A2B1B1						
9	A2A2B1B2						
10	A2A2B2B2						
11	Table 1. The change in genotypic frequencies over five generations						
12							
13							
14		Generation 1	Generation 2	Generation 3	Generation 4	Generation 5	
15	p_2 = B1						
16	q_2 = B2						
17	Table 2. Change in allele frequencies at the B locus over five generations						
18							
19							
20		μ	V_G	θ_G	h^2	*D*	*i*
21	Gen 1			0			
22	Gen 2			0			
23	Gen 3			0			
24	Gen 4			0			
25	Gen 5			0			
26	Table 3. Change in population mean (μ), genotypic variance (V_G), genotypic std. deviation (θ_G) heritability (h2), linkage						
27	disequilibrium coefficient (D) and intensity of selection (i) over five generations						
28							
29		R					
30	0	0					
31	After 1 Gen						
32	After 2 Gen						
33	After 3 Gen						
34	After 4 Gen						
35	After 5 Gen						
36	Table 4. Response to Selection over fiver generations of marker-assisted selection						
37							

Figure 2.13 The range A1-G37 in Sheet 2 of the 'Marker-Assisted Selection' computer model.

A macro was designed to:

- Automatically generate the genotypic frequencies of the parental population on Sheet 1 for five generations (Table 1 of Figure 2.13).
- To generate the allele frequencies at the B locus over five generations (Table 2 of Figure 2.13)
- To generate μ, V_A, h^2, ***D*** and *i* over five generations (Table 3 of Figure 2.13).
- To generate R over five generations of marker-assisted selection.

The macro was named 'MAS' and the shortcut key was entered as <Ctrl m>.

The additive genetic standard deviations were calculated for the five generations:

- In cell D21, the formula **=SQRT(C21)** was entered and copied down to cell D 25.

When the macro is activated and all the values are automatically generated, so the additive genetic standard deviations are also automatically calculated.

Once all the different population statistics had been generated in Sheet 2 of the 'Marker-Assisted Selection' computer model, a third Sheet was added in order to graphically depict the effect of marker-assisted selection on a population over five generations.

(d) Graphical depictions of response to marker-assisted selection

The graphs created on Sheet 3 were graphical representations of the values displayed in Sheet 2, namely the change in allele frequencies, h^2, σ_A, V_A and R over five generations of marker-assisted selection.

To demonstrate the response to selection over five generations of marker-assisted selection, the following graph was created:

- The chart wizard was selected and the XY (Scatter) option was chosen.
- The range B30-B35 on Sheet 2 was selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'Response.'
- The graph was named 'Response to Marker-Assisted Selection.'

The change in allele frequency due to five generations of marker-assisted selection was illustrated by creating the following graph on Sheet 3:

- The chart wizard option was selected and the XY (scatter) option was chosen.
- The range A14-F16 on Sheet 2 were selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'Frequency.'
- The graph was entitled 'Change in Allele Frequency due to Marker-Assisted Selection.'

With the aim of depicting the dependency of R on h^2 and σ_A in the population the graph entitled 'Relationship between R, h^2 and σ_A' was generated:

- The chart wizard was selected and the XY (Scatter) option was chosen.
- The ranges B31-B35, E21-E25 and D21-D25 on Sheet 2 were selected as the data ranges.
- The X axis was labeled 'Generations'.

The effect of additive variance on the level of response to marker-assisted selection was shown by the graph created on Sheet 3 as follows:

- The chart wizard was selected and the XY (Scatter) option was chosen.
- The ranges B31-B35 and D21-D25 on Sheet 2 were selected as the data ranges.
- The X axis was labeled 'Genotypic Variance' and the Y axis was labeled 'Response to Selection.'
- The graph was entitled 'Relationship between Additive Variance and Response to Selection.'

Finally, in order to portray the direct relationship between R, i and $\boldsymbol{D}$, the following graph was created:

- The chart wizard was selected and the column graph option was chosen.
- The ranges G21-25, B31-B35 and F21-F25 on Sheet 2 were selected as the data ranges.
- The X axis was labeled 'Generations.'
- The graph was named 'Relationship between i, R and $\boldsymbol{D}$.'

2.3.3 Comparison between phenotypic selection and marker-assisted selection

To show the increase in the population mean under a number of different selection strategies, a spreadsheet consisting of tables and the corresponding graphs was created and named 'Comparison of Selection Strategies.' Since the increase in the population mean was measured and compared for phenotypic selection and marker-assisted selection, the values were taken from the 'Phenotypic Selection' and 'Marker-Assisted Selection' models.

Firstly, a recombination fraction of 0 was considered. The population mean was measured over five generations of marker-assisted selection when the QTL accounted for 60%, 80% and 100% of the trait variation. Thereafter, the population mean was measured over five generations of phenotypic selection. Subsequently the process was repeated for a recombination fraction set at 0.2 and 0.5 and the values were entered into tables and corresponding graphs were created. In Sheet 1 of the 'Comparison of Selection Strategies' spreadsheet, tables were set up as shown in Figure 2.14.

	A	B	C	D	E
1	Population mean at r = 0				
2		P	60%	80%	100%
3	Gen 1				
4	Gen 2				
5	Gen 3				
6	Gen 4				
7	Gen 5				
8					
9	Population mean at r = 0.2				
10		P	60%	80%	100%
11	Gen 1				
12	Gen 2				
13	Gen 3				
14	Gen 4				
15	Gen 5				
16					
17	Population mean at r = 0.5				
18		P	60%	80%	100%
19	Gen 1				
20	Gen 2				
21	Gen 3				
22	Gen 4				
23	Gen 5				

Figure 2.14 The range A1-E23 in Sheet 1 of the 'Comparison of Selection Strategies' spreadsheet.

The values for this table were calculated using the following steps:

1. Sheet 1 of the 'Marker-Assisted Selection' model was opened and the recombination fraction was set at 0 by entering '0' into cell E1.
2. The percentage of additive variance explained by the QTL for the trait of interest was set at 60% by entering '60' into cell W12 in Sheet 1 of the 'Marker-Assisted Selection' model.
3. The macro was activated by pressing <Ctrl m>
4. The resulting population means (**μ**) over five generations were copied from B21-B25 in Sheet 2 of the 'Marker-Assisted Selection' model and pasted into cells C3-C7 of Sheet 1 of the 'Comparison of Selection Strategies' spreadsheet (shown in Figure 2.14).
5. Similarly, μ over five generations was recorded with the percentage of V_A explained by the QTL set at 80% and 100%. The resulting population means were copied from B21-B25 in Sheet 2 of the 'Marker-Assisted Selection' model and pasted into cells D3-D7 and E3-E7, respectively, of Sheet 1 of the 'Comparison of Selection Strategies' spreadsheet.

6. Steps 2 - 5 were repeated for recombination fraction values of 0.2 and 0.5 and the appropriate population mean values were placed in cells C11-E15 and cells C19-C23 respectively.

Similarly, for the change in mean due to phenotypic selection:

1. Sheet 1 of the 'Phenotypic Selection' model was opened and the recombination fraction was set at 0, by entering 0 into cell E1.
2. The macro was run by pressing <Ctrl p> and the resulting population means over five generations were copied from cells B16-B20 in Sheet 2 of this spreadsheet and pasted into cells B3-B7 in Sheet 1 of the 'Comparison of Change in Mean' spreadsheet.
3. Similarly, the population mean over five generations was recorded with the recombination fraction was set at 0.2 and 0.5 and the values were placed in cells B11-B15 and cells B19-B23 respectively.

The graphs created from the tables on Sheet 1 (Figure 2.14) were presented in Sheet 2. The following figure heading was written at the top of Sheet 2. 'Figure 1-3. Comparison between the change in population mean when applying phenotypic selection (P) and marker-assisted selection when the QTL accounts for 60%, 80% and 100% of the trait variation. The recombination fraction (r) was varied as shown on the graph. Environmental heterogeneity = 0.3

The following graph was created to demonstrate the response to the different selection strategies when r = 0:

- The chart wizard was selected and the XY (Scatter) option was chosen.
- The range A2-E7 on Sheet 1 was selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'Population Mean.'

To illustrate the response to the different selection strategies when r = 0.2, the following graph was generated:

- The chart wizard option was selected and the XY (scatter) option was chosen.
- The range A10-E15 on Sheet 1 was selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'Population Mean.'

In order to show the response to the different selection strategies when r = 0.5, the following graph was created:

- The chart wizard option was selected and the XY (scatter) option was chosen.
- The range A18-E23 on Sheet 1 was selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'Population Mean.'

2.4 MODEL ILLUSTRATING THE EFFECTS OF SELECTION WITH THE ASSUMPTION OF DOMINANCE

The purpose of this model was to demonstrate the effect of dominance on response to selection after one generation for populations undergoing phenotypic or marker-assisted selection, as dominance may affect the efficiency of these two strategies.

2.4.1 Creation of model illustrating effects of phenotypic selection with assumption of complete dominance

A file was opened in Microsoft Excel and saved as 'Selection with Dominance'. Sheet 1 of the 'Phenotypic Selection' model was copied into Sheet 1 of the "Selection with Dominance' model and labeled 'Phen Sel.' This model demonstrated the effects of phenotypic selection when non-additive gene action occurs at the locus under selection. The format was the same as that in Sheet 1 in the 'Phenotypic Selection' model with only a few changes. The following figures and formulae explain the changes made to the previously described model which assumed additive gene action.

	A	B	C	D	E	F	G	H	I
1	**Phenotypic Selection**			Recombination Fracion =	0.5				
2									
3	Genotype	Frequency	Tally count		Allele frequencies				
4			0		Locus 1		Locus 2		
5	A1A1B1B1	0.300	0.3		p_1 = A1 =	0.4795	p_2 = B1 =	0.4795	
6	A1A1B1B2	0.000	0.3		q_1 = A2 =	0.5205	q_2 = B2 =	0.5205	
7	A1A1B2B2	0.000	0.3						
8	A1A2B1B1	0.000	0.3		Gamete probabilities				
9	A1A2B1B2	0.400	0.7		A1B1	A1B2	A2B1	A2B2	
10	A1A2B2B2	0.000	0.7	Expected	0.230	0.250	0.250	0.271	
11	A2A2B1B1	0.000	0.7	Observed	0.389	0.085	0.107	0.419	
12	A2A2B1B2	0.000	0.7						
13	A2A2B2B2	0.300	1	<= This number MUST = 1	Genotype	A1	A2	B1	B2
14					Value	0	0	3	0
15								Select above	50
16	Individual	Random #	Genotype	Gamete	BV	G	E	P	Select

Figure 2.15 The range A1-I16 in Sheet 1 of the 'Selection with Dominance' computer model.

	Q	R	S	T	U	V	W
1		Selection Statistics					
2		$\mu_{(gen\ 1)}$ =	54.062	Environmental Heterogenity =			0.200
3		$V_{A(gen\ 1)}$ =	5.435	$\mu_{(gen\ 2)}$ =	53.544	S =	0.816
4		$V_{D(gen\ 1)}$ =	2.445	$V_{A(gen\ 2)}$ =	2.177	Average Gen	50.000
5		$V_{E(gen\ 1)}$ =	0.038	$V_{D(gen\ 2)}$ =	6.530	h^2 =	0.686
6		$V_{P(gen\ 1)}$ =	7.923	$V_{E(gen\ 2)}$ =	0.042	h^2 =	0.248
7		i =	0.290	$V_{P(gen\ 2)}$ =	8.771	R =	0.560
8		$\mu_{(selected)}$ =	54.878				
9							
10							
11							
12							
13							
14							
15	ffspring Genotype						
16	Locus B		Genotype	BV	G	E	P

Figure 2.16 The range Q1-W16 in Sheet 1 of the 'Selection with Dominance' computer model.

While the breeding value of the individuals remained the same as those in the 'Phenotypic Selection' model, the genotypic values did not, since there was complete dominance at the B locus. The B_1 allele was completely dominant over the B_2 allele and therefore a B_1B_1 and a B_1B_2 individual had the same genotypic value. As in the other models, the genotypic values were determined by the values of the B_1 and B_2 alleles which could be manually entered into the table (cells E13-I14, Figure 2.15).

To calculate the genotypic values for generation 1, the following formula was entered into cell F17 (Figure 2.15) and copied down to F1016:

- **=IF(C17='A1A2B1B2',($H$14*2)+$W$4,IF(C17='A1A1B1B2',($H$14*2)+$W$4,IF(C17='A2A2B1B2',($H$14*2)+$W$4,$W$4+E17)))**

Similarly, to calculate the genotypic values for the offspring generation the following formula was entered into cell U17 (Figure 2.16), and copied down to U1016:

- **=IF(S17='','',IF(S17='A1A2B1B2',($H$14*2+$W$4),IF(S17='A1A1B1B2',($H$14*2)+$W$4,IF(S17='A2A2B1B2',($H$14*2)+$W$4,$W$4+T17))))**

Changes were also made to the selection statistics table used in both the 'Phenotypic Selection' and 'Marker-Assisted Selection' models because it was necessary to divide the genotypic variance up into an additive variance component (V_A) and a non-additive variance component (V_D) caused by the dominance at the B locus. The selection statistics table is shown in the range R1-W8 of Figure 2.16. The equations

used to calculate the non-additive variance due to dominance were based on the equation $V_G = V_A+V_D$ (Falconer, 1989). The additive variance was calculated from the breeding values displayed in the computer model and was subtracted from the genotypic variance to give the dominance variance.

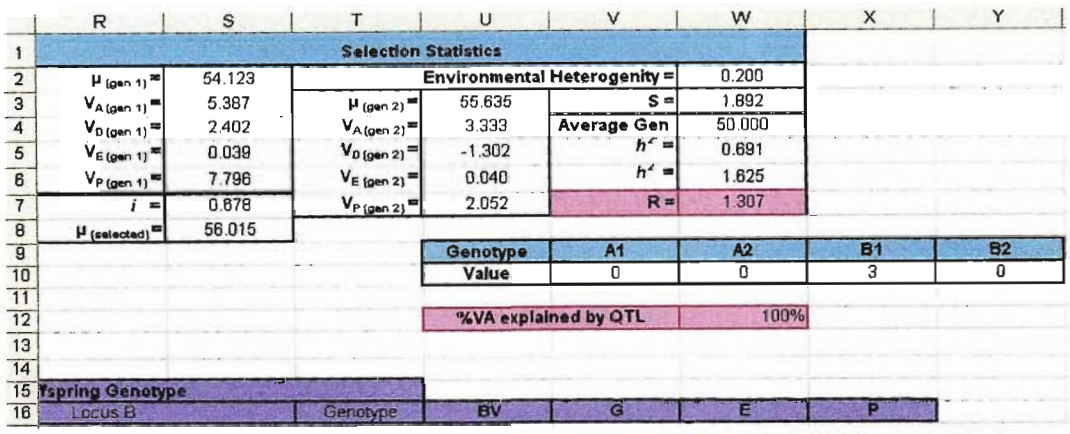
- In cell S3, the formula **=VAR(E17:E1016)** was entered to calculate the additive variance for generation 1.
- In cell S4, the formula **=(VARF17:F1016)-S3** was entered to calculate the non-additive variance for generation 1.
- In cell U4, the formula **=VAR(U17:U1016)** was entered to calculate the additive variance for generation 2.
- In cell U5, the formula **=(VARV17:V1016)-U4** was entered to calculate the non-additive variance for generation 2.

2.4.2 Creation of model illustrating effects of marker-assisted selection with assumption of complete dominance

In order to determine the effects of marker-assisted selection on a locus with non-additive gene action, Sheet 2 was opened in the 'Selection with Dominance' model and labeled 'MAS'. Sheet 1 of the 'Marker-Assisted Selection' model was then copied into Sheet 2 of the 'Selection with Dominance' model. Sheet 2 in this model therefore had the same format as Sheet 1 of the 'Marker-Assisted Selection' model and only the changes made for the purpose of this model will be discussed in this section. The following figures aid in the explanation of the changes made to the 'Marker-Assisted Selection' model.

	A	B	C	D	E	F	G	H	I
1	**Marker-Assisted Selection**			Recombination Fraction =	0.5				
2									
3	Genotype	Frequency	Tally count		Allele frequencies				
4			0		Locus 1		Locus 2		
5	A1A1B1B1	0.300	0.3		p_1 = A1 =	0.534	p_2 = B1 =	0.534	
6	A1A1B1B2	0.000	0.3		q_1 = A2 =	0.466	q_2 = B2 =	0.466	
7	A1A1B2B2	0.000	0.3						
8	A1A2B1B1	0.000	0.3		Gamete probabilities				
9	A1A2B1B2	0.400	0.7		A1B1	A1B2	A2B1	A2B2	
10	A1A2B2B2	0.000	0.7	Expected	0.285	0.249	0.249	0.217	
11	A2A2B1B1	0.000	0.7	Observed	0.432	0.095	0.101	0.372	
12	A2A2B1B2	0.000	0.7						
13	A2A2B2B2	0.300	1	<= This number MUST = 1	Genotype	A1	A2	B1	B2
14					Value	5	0	0	0
15								Select above	1
16	Individual	Random #	Genotype	Gamete	BV	G	E	P	Marker Value
17	1	0.026208007	A2A2B2B2	A2B2	0	50.000	0.057	50.057	0

Figure 2.17 The range A1-I17 in Sheet 2 of the 'Selection with Dominance' computer model.



	R	S	T	U	V	W	X	Y
1	Selection Statistics							
2	$\mu_{(gen\ 1)}$ =	54.123			Environmental Heterogenity =	0.200		
3	$V_{A(gen\ 1)}$ =	5.387	$\mu_{(gen\ 2)}$ =	55.635	S =	1.892		
4	$V_{D(gen\ 1)}$ =	2.402	$V_{A(gen\ 2)}$ =	3.333	Average Gen	50.000		
5	$V_{E(gen\ 1)}$ =	0.039	$V_{D(gen\ 2)}$ =	-1.302	h^2 =	0.691		
6	$V_{P(gen\ 1)}$ =	7.796	$V_{E(gen\ 2)}$ =	0.040	h^4 =	1.625		
7	i =	0.878	$V_{P(gen\ 2)}$ =	2.052	R =	1.307		
8	$\mu_{(selected)}$ =	56.015						
9				Genotype	A1	A2	B1	B2
10				Value	0	0	3	0
11								
12				%VA explained by QTL		100%		
13								
14								
15	Offspring Genotype							
16	Locus B		Genotype	BV	G	E	P	

Figure 2.18 The range R1-Y16 in Sheet 2 of the 'Selection with Dominance' computer model.

As in the 'Marker-Assisted Selection' model, the genotypic values were determined by the values entered into the table seen in Figure 2.18 (cells U9-Y10). However, to calculate the genotypic values for generation 1 with complete dominance, the following formula was entered into cell F17 (Figure 2.17) and copied down to F1016:

- **=IF(C17='A1A2B1B2',($X$10*2*$W$12)+$W$4,IF(C17='A1A1B1B2',(X&10*2*$W$12)+$W$4,IF(C17='A2A2B1B2',($X$10*2*$W$12)+$W$4,$W$4+E17)))**

Similarly, the genotypic values for the offspring generation were calculated. In cell V17 (Figure 2.18), the following formula was entered and copied down to V1016:

- **=IF(T17='A1A2B1B2',($X$10*2*$W$12)+$W$4,IF(T17='A1A1B1B2',(X&10*2*$W$12)+$W$4,IF(T17='A2A2B1B2',($X$10*2*$W$12)+$W$4,$W$4+E17)))**

The selection statistics were calculated in the same manner as on Sheet 1 of the 'Phenotypic Selection' model (Table 2.5, Page 41).

Sheets 1 and 2 of the 'Selection with Dominance' model demonstrated the effects of phenotypic selection and marker-assisted selection, respectively, in the presence of complete dominance. Sheet 3 of the same model was used to compare the response to selection of these different selection strategies.

2.4.3 Comparison between phenotypic selection and marker-assisted selection

To compare the effects of phenotypic selection and marker-assisted selection in the presence of complete dominance, Sheet 3 was created and labeled 'Comparison.' This

Sheet consisted of a table of values and the corresponding graph showing the comparison between response to marker-assisted selection and phenotypic selection when complete dominance is present at the locus under selection.

For marker-assisted selection, the percentage of trait variation explained by the locus under selection was set at 60%, 80% and 100%. The recombination fraction was set at 0, 0.2 and 0.5. The following figure shows the set up of the table on Sheet 3 of the 'Selection with Dominance' model.

	A	B	C	D
1	Response to selection at r = 0			
2	MAS			P
3	60%	80%	100%	
4				
5	Response to selection at r = 0.2			
6	MAS			P
7	60%	80%	100%	
8				
9	Response to selection at r = 0.5			
10	MAS			P
11	60%	80%	100%	
12				

Figure 2.19 The range A1-D12 in Sheet 3 of the 'Selection with Dominance' computer model.

The values for this table were calculated using the following steps:

To calculate the responses to phenotypic selection, the following steps were taken.

1. Sheet 1 of the 'Selection with Dominance' model was opened and the recombination fraction was set at 0, by entering 0 into cell E1 (Figure 2.15, Page 61).
2. The resulting response to selection was copied from cell W7 in Sheet 1 (Figure 2.16, Page 61) and pasted in cell D4 of Sheet 3 of the same model, in the table shown in Figure 2.19.
3. Similarly, R was recorded when the recombination fraction was set at 0.2 and 0.5, and the response values were placed in cells D8 and D12 respectively.

To calculate the response to marker-assisted selection, the following steps were taken:

1. Sheet 2 of the 'Selection with Dominance' model was opened and the recombination fraction was set at 0 by entering '0' into cell E1 (Figure 2.17, Page 63).
2. The percentage of additive variance explained by the QTL for the trait of interest was set at 60% by entering '60' into cell W12 (Figure 2.18, Page 63).
3. The resulting response to selection was copied from cell W7 in Sheet 2 (Figure 2.18, Page 63) and pasted in cell A4 of Sheet 3 in the table shown in Figure 2.19.
4. Steps 2-3 were repeated for when the percentage of additive variance explained by the QTL for the trait of interest was set at 80% and then at 100%. The resulting responses were copied from cell W7 and pasted in cell B4 and C4 (for 80% and 100% respectively) of Sheet 3 in the table shown in Figure 2.19.
5. Steps 2-4 were repeated when the recombination fraction was set at 0.2 and then 0.5. The response values, when r=0.2, were placed in cells A8-C8 and the response values, when r=0.5, were placed in cells A12-C12.

In order to graphically depict the tabulated results shown in Figure 2.19, the graph entitled 'Comparison of Response to Marker-Assisted Selection and Phenotypic Selection' was created:

1. The chart wizard was opened and the column graph option was selected.
2. For the '60%' category, the cells A4, A8 and A12 on Sheet 3 were selected as the data range.
3. For the '80%' category, the cells B4, B8 and B12 on Sheet 3 were selected as the data range.
4. For the '100%' category, the cells C4, C8 and C12 on Sheet 3 were selected as the data range.
5. For the 'P' category, the cells D4, D8 and D12 on Sheet 3 were selected as the data range.
6. The x axis was labeled 'Response to Selection.'

2.5 APPLICATION OF THE 'MARKER-ASSISTED SELECTION' MODEL TO HOLSTEIN CATTLE

The purpose of the following two models was to incorporate real data into the 'Marker-Assisted Selection' model and show that these models could be applied in the breeding industry as predictive and/or comparative tools.

2.5.1 Application of 'Marker-Assisted Selection' model using a direct marker

(a) Introduction to data used

With the aim of simulating the effect of marker-assisted selection for an increase in milk yield, utilizing a direct marker, data was taken from real Holstein cattle populations and incorporated into the 'Marker-Assisted Selection' model.

Buchanan *et al.* (2003) looked at data records from 11 Saskatchewan herds with a mean herd size of 71 and a mean milk yield of 67 pounds per day (lbs/day). The molecular marker in this study was a one base pair fragment of the DNA sequence of a gene that encoded a protein hormone known as Leptin. Different alleles of this gene have been reported to have an effect on milk yield in dairy cattle. Animals can have one of three marker genotypes - TT, TC or CC. (This would be the equivalent to A_1A_1, A_1A_2 and A_2A_2 in the 'Marker-Assisted Selection' model). Animals with the TT genotype were reported to produce more milk (3.3 lb/day) than animals with the CC genotype. The TC genotype was intermediate (2.0 lb/day) since there was additive gene action at the locus of interest. In this study the frequency of the T allele in Holstein cattle was 0.46.

The chromosomal position of the maker locus and the QTL in a heterozygous individual in the population is shown in Figure 2.20. The T marker allele is located within the B_1 allele and the C marker allele is located within the B_2 allele.

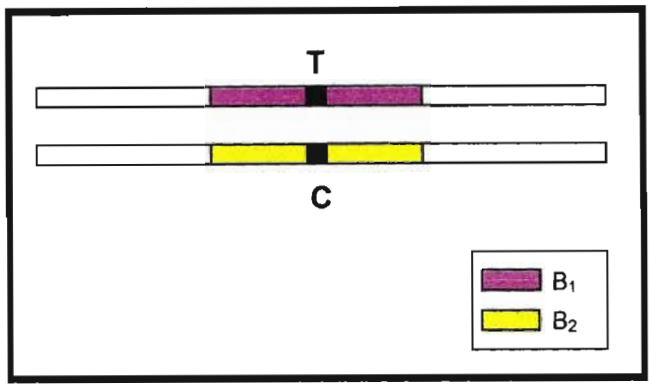


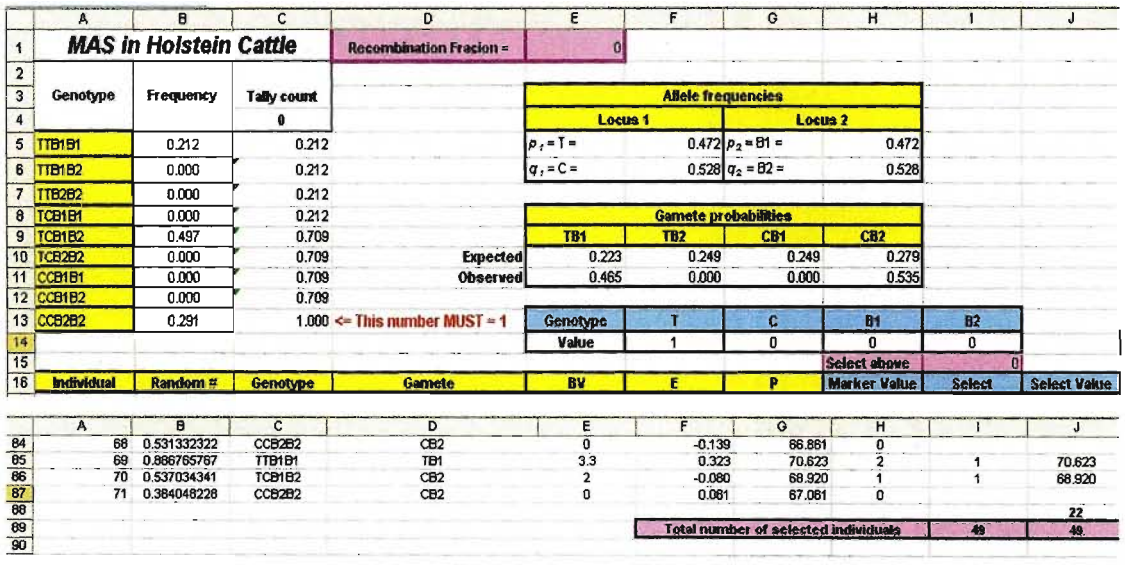
Figure 2.20 Chromosomal configuration of a heterozygous individual showing the position of the marker locus within the QTL of interest.

The molecular marker used in this study may be considered a direct marker, as the marker locus is situated inside the QTL, as shown in Figure 2.20. Therefore the recombination fraction was set as 0. Secondly, this model investigated the effects of marker-assisted selection on the milk production genes associated with the direct marker only and thus the percentage of the total trait variation explained by these genes was set at 100%. Thirdly, it was safe to assume that the T marker allele was directly linked to the B_1 allele for increased milk production and the C marker allele was directly linked to the B_2 allele (as illustrated in Figure 2.20). This model described a direct marker and the three genotypes in this population were symbolised as TTB_1B_1, TCB_1B_2 and CCB_2B_2.

(b) Incorporation of data into 'Marker-Assisted Selection' model

The 'Marker-Assisted Selection' model was copied into a new Microsoft Excel spreadsheet and saved as 'Direct Selection'. Since the frequency of the T allele in this Holstein cattle population was 0.46, the frequency of the C allele was taken as 0.54. The marker genotypic frequencies were then calculated to be TT = 0.212, TC = 0.497, CC = 0.291 using the Hardy-Weinberg principle.

The aforementioned data was used to create the 'Direct Selection' model in order to determine the efficiency of marker-assisted selection as a breeding strategy in this particular cattle population.



	A	B	C	D	E	F	G	H	I	J
1	**MAS in Holstein Cattle**			Recombination Fraction =	0					
2										
3	Genotype	Frequency	Tally count		Allele frequencies					
4			0		Locus 1		Locus 2			
5	TTB1B1	0.212	0.212		p_1 = T =	0.472	p_2 = B1 =	0.472		
6	TTB1B2	0.000	0.212		q_1 = C =	0.528	q_2 = B2 =	0.528		
7	TTB2B2	0.000	0.212							
8	TCB1B1	0.000	0.212		Gamete probabilities					
9	TCB1B2	0.497	0.709		TB1	TB2	CB1	CB2		
10	TCB2B2	0.000	0.709	Expected	0.223	0.249	0.249	0.279		
11	CCB1B1	0.000	0.709	Observed	0.465	0.000	0.000	0.535		
12	CCB1B2	0.000	0.709							
13	CCB2B2	0.291	1.000	<= This number MUST = 1	Genotype	T	C	B1	B2	
14					Value	1	0	0	0	
15								Select above	0	
16	Individual	Random #	Genotype	Gamete	BV	E	P	Marker Value	Select	Select Value

	A	B	C	D	E	F	G	H	I	J
84	68	0.531332322	CCB2B2	CB2	0	-0.139	66.861	0		
85	69	0.886765767	TTB1B1	TB1	3.3	0.323	70.623	2	1	70.623
86	70	0.537034341	TCB1B2	CB2	2	-0.080	68.920	1	1	68.920
87	71	0.384048228	CCB2B2	CB2	0	0.081	67.081	0		
88										22
89							Total number of selected individuals		49	49
90										

Figure 2.21 The ranges A1-J16 and A84-J90 in Sheet 1 of the 'Direct Selection' computer model.

Using the data from Buchanan *et al.* (2003), the following values were entered:

- The population size was set at 71 individuals.
- Genotypic frequencies were entered into the cells B5-B13.
- In cell E1, the value '0' was entered to set the recombination fraction at 0.

A table was set up (cells E13-I14 in Figure 2.21) in order to assign values to the marker alleles so that marker-assisted selection could be simulated. To ensure that only animals with at least one T marker allele were selected, the following entries were made:

- In cell F14, a value of 1 was entered.
- In cell G14, a value of 0 was entered.
- In cell I15, a value of 0 was entered so that only animals with a marker value of 1 or more were selected.

Headings were set up in the same way as the headings in the 'Marker-Assisted Selection' model, as shown in Figure 2.22.

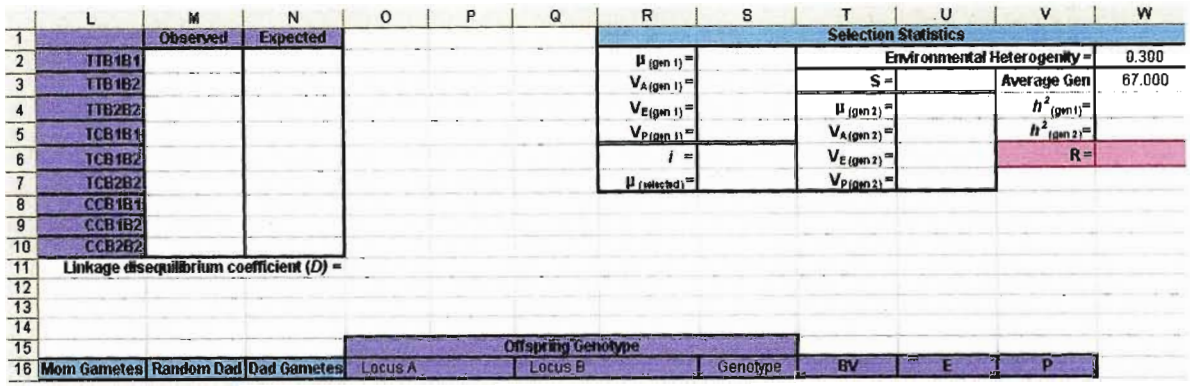


Figure 2.22 The range L1-16 in Sheet 1 of the 'Direct Selection' computer model.

Mating amongst the selected individuals was simulated in the same way as was described for the 'Phenotypic Selection' model and the calculations for the observed and expected genotypic frequencies were the same as those used in the 'Random Mating' model (Table 2.3, Page 34).

The following steps were taken to calculate breeding values of individuals:

- In cell W3, the value '67.000' was entered as the average genotypic value because the mean milk yield in the population is 67lbs/day.
- The formula, **=IF(S17="",",IF(S17='TTB1B1',3.3,(IF(S17='TCB1B2',2,0))))** was entered into cell T17 and copied down to cell T87.

This formula assigned a breeding value of 3.3lbs/day and 2.0lbs/day to all TTB_1B_1 and TCB_1B_2 individuals respectively. The CCB_2B_2 individuals were assigned a breeding value of 0, since the literature stated that the other two genotypes produce 3.3lbs/day and 2.0lbs/day more milk than the CCB_2B_2 individuals.

The following formula calculated the phenotypic values by adding the breeding value, the average genetic value (Cell W3, Figure 2.22) and the environmental value.

- In cell V17, the formula **=IF(U17="",",T17+U17+$W$3)** was entered and copied down to cell V87.

The selection statistics table was set up (R1-W7, Figure 2.22) and the same formulae from the 'Phenotypic Selection' model were used. (Table 2.5, Page 41)

(c) Response to marker-assisted selection

In order to calculate the response to marker-assisted selection, Sheet 2 was opened and tables were set up as shown in Figure 2.23.

	A	B	C	D	E	F
1		Generation 1	Generation 2	Generation 3	Generation 4	Generation 5
2	A1A1B1B1					
3	A1A1B1B2					
4	A1A1B2B2					
5	A1A2B1B1					
6	A1A2B1B2					
7	A1A2B2B2					
8	A2A2B1B1					
9	A2A2B1B2					
10	A2A2B2B2					
11	Table 1. The change in genotypic frequencies over five generations					
12						
13						
14		Generation 1	Generation 2	Generation 3	Generation 4	Generation 5
15	p_1 = T					
16	q_1 = C					
17	Table 2. Change in allele frequencies at the marker locus over five generations					
18						
19						
20		μ	V_A			
21	Gen 1					
22	Gen 2					
23	Gen 3					
24	Gen 4					
25	Gen 5					
26	Table 3. Change in population mean (μ) and additive variance (V_A) over five generations					
27						
28						
29		R				
30	0	0				
31	After 1 Gen					
32	After 2 Gen					
33	After 3 Gen					
34	After 4 Gen					
35	After 5 Gen					
36	Table 4. Response to Marker-Assisted Selection over five generations					
37						

Figure 2.23 The range A1-F37 in Sheet 2 of the 'Direct Selection' computer model.

A macro was created and named 'milkprod' and given a shortcut key <Ctrl p>. This macro was designed to automatically generate the following statistics over five generations of marker-assisted selection.

The macro was designed to:

- Calculate the genotypic frequencies of the parental population from Sheet 1 for five generations (Table 1, Figure 2.23).

- Generate the allele frequencies at the marker locus over five generations (Table 2, Figure 2.23).
- The population mean (μ), and additive variance (V_A) over five generations (Table 3, Figure 2.23).
- Generate R over five generations of marker-assisted selection (Table 4, Figure 2.23).

(d) Graphical depiction of response to marker-assisted selection

In order to graphically depict the effect of marker-assisted selection on the population over five generations, a third Sheet was added to the 'Direct Selection' model. The graphs created on Sheet 3 were the graphic representations of the values displayed in Figure 2.23.

To illustrate the response to marker-assisted selection over five generations, using a direct marker, the following graph was created on Sheet 3:

- The chart wizard was selected and the XY (Scatter) option was chosen.
- The range B30-B35 on Sheet 2 was selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'Response.'
- The graph was entitled 'Response to Marker-Assisted Selection.'

In order to demonstrate the expected change in allele frequencies in the Holstein cattle population as a result of marker-assisted selection, the graph entitled 'Change in Marker Allele Frequency' was generated on Sheet 3:

- The chart wizard option was selected and the XY (scatter) option was chosen.
- The range A14-F16 on Sheet 2 were selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'Frequency.'

The effect of marker-assisted selection on the frequency of the T marker allele was illustrated in the following graph on Sheet 3.

- The chart wizard was selected and the XY (Scatter) option was chosen.
- The ranges B31-B35 and B15-F15 on Sheet 2 were selected as the data ranges.
- The X axis was labeled 'Freq T' and the Y axis was labeled 'Response to Selection.'
- The graph was named 'Relationship between the Frequency of the T Marker Allele and Response to Selection.'

To demonstrate the effect of marker-assisted selection on the increase in milk yield, the graph entitled 'Increase in Mean Milk Production (lbs/day) in a Holstein Cattle Population as a result of Marker-Assisted Selection' was created on Sheet 3 as follows:

- The chart wizard was selected and the column graph option was chosen.
- The range B21-B25 on Sheet 2 was selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'lbs milk/day'

2.5.2 Application of 'Marker-Assisted Selection' model using an indirect marker

(a) Introduction to marker data

To simulate the effect of marker-assisted selection on an increase in milk fat percentage, utilizing an indirect marker, data was taken from an Israeli Holstein cattle population and incorporated into the 'Marker-Assisted Selection' model.

Weller *et al.* (2003) studied the population-wide linkage disequilibrium on bovine chromosome 14 between microsatellite ILSTS039 and DGAT1 in the Israeli Holstein population. Weller *et al.* (2003) found that the ILSTS039 allele, termed '225', and the DGAT1 K allele (which involved the substitution of a lysine residue with an alanine residue), were associated with increased fat percent. Furthermore, from the effects associated with cows homozygous for the 225 allele, it was found that the effect of the quantitative trait locus was co-dominant and that one copy of the DGAT1K allele increased the breeding value of individuals by 0.16% fat in the population.

The following information was taken from the study conducted by Weller *et al.* (2003):

- Total number of individuals in the population was 394
- The genotypic frequencies were as follows:

  $A_1A_1B_1B_1 = 0.008$

  $A_1A_1B_1B_2 = 0.018$

  $A_1A_1B_2B_2 = 0.005$

  $A_1A_2B_1B_1 = 0.010$

$A_1A_2B_1B_2 = 0.157$

$A_1A_2B_2B_2 = 0.114$

$A_2A_2B_1B_1 = 0.003$

$A_2A_2B_1B_2 = 0.056$

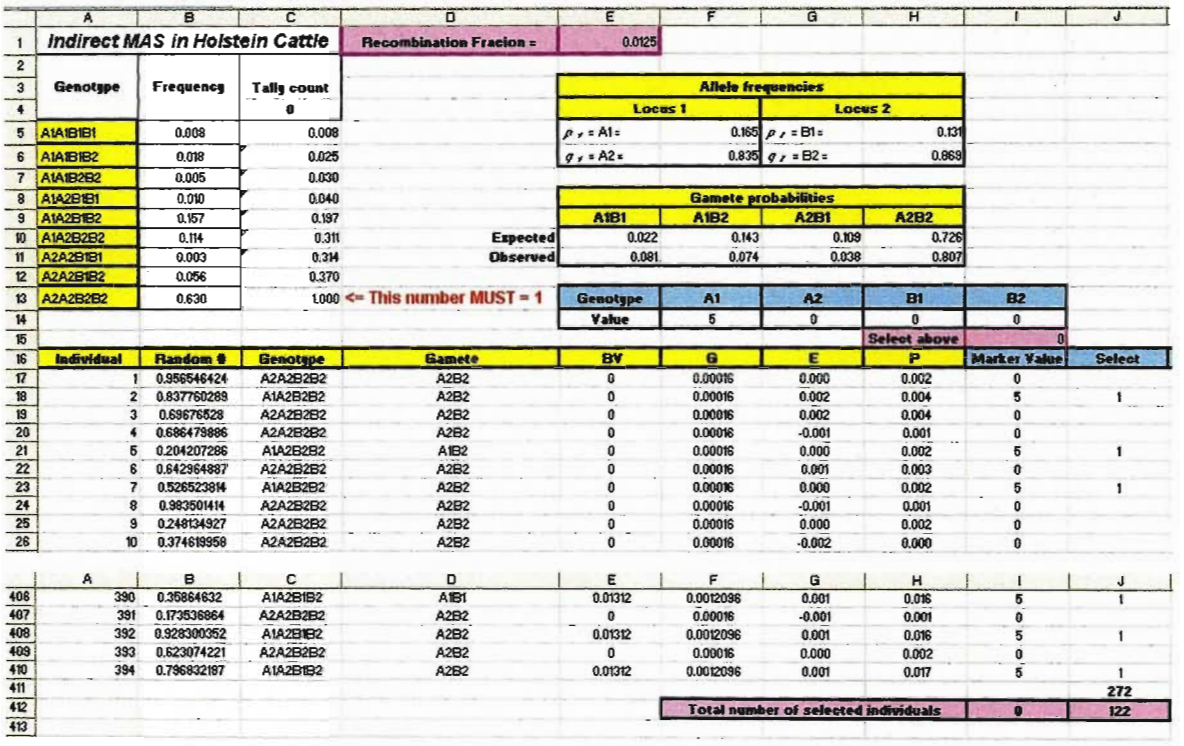
$A_2A_2B_2B_2 = 0.630$

- The mean percent fat was 0.002%.
- The recombination frequency between ILSTS039 and DGAT1 was found to be 2.5% and therefore the recombination fraction was equal to 0.0125.
- Environmental Heterogeneity was 0.001500.
- The QTL accounted for 8% of the total genetic variation.

Using the data obtained from the study conducted by Weller *et al.* (2003), the 'Indirect Marker' model was used to investigate the effects of marker-assisted selection on the 225 allele for increased mean percent fat (%fat) in the Israeli Holstein cattle population.

(b) Incorporation of data into 'Marker-Assisted Selection' model

The 'Marker-Assisted Selection' model was copied into a new Microsoft Excel spreadsheet and saved as 'Indirect Selection'. In the 'Indirect Selection' model, DGAT1, the putative QTL affecting milk production traits, was synonymous with the B locus and ILSTS039, the microsatellite marker, was synonymous with the A locus. The A_1 allele therefore represented the 225 allele of the ILSTS039 marker and A_2 represented the other unspecified allele of the ILST039 marker. Similarly, B_1 represented the DGAT1 K allele and B_2 represented the DGAT1 A allele. This data was used to create the 'Indirect Selection' model in order to determine the effects of marker-assisted selection utilizing linkage disequilibrium in this particular cattle population (Figure 2.24).



	A	B	C	D	E	F	G	H	I	J
1	*Indirect MAS in Holstein Cattle*			Recombination Fracion =	0.0125					
2										
3	Genotype	Frequency	Tally count		Allele frequencies					
4			0		Locus 1		Locus 2			
5	A1A1B1B1	0.008	0.008		p_r = A1 =	0.165	p_r = B1 =	0.131		
6	A1A1B1B2	0.018	0.025		q_r = A2 =	0.835	q_r = B2 =	0.869		
7	A1A1B2B2	0.005	0.030							
8	A1A2B1B1	0.010	0.040		Gamete probabilities					
9	A1A2B1B2	0.157	0.197		A1B1	A1B2	A2B1	A2B2		
10	A1A2B2B2	0.114	0.311	Expected	0.022	0.143	0.109	0.726		
11	A2A2B1B1	0.003	0.314	Observed	0.081	0.074	0.038	0.807		
12	A2A2B1B2	0.056	0.370							
13	A2A2B2B2	0.630	1.000	<= This number MUST = 1	Genotype	A1	A2	B1	B2	
14					Value	5	0	0	0	
15								Select above	0	
16	Individual	Random #	Genotype	Gamete	BV	G	E	P	Marker Value	Select
17	1	0.956546424	A2A2B2B2	A2B2	0	0.00016	0.000	0.002	0	
18	2	0.837760288	A1A2B2B2	A2B2	0	0.00016	0.002	0.004	5	1
19	3	0.69676528	A2A2B2B2	A2B2	0	0.00016	0.002	0.004	0	
20	4	0.686479886	A2A2B2B2	A2B2	0	0.00016	-0.001	0.001	0	
21	5	0.204207286	A1A2B2B2	A1B2	0	0.00016	0.000	0.002	5	1
22	6	0.642964887	A2A2B2B2	A2B2	0	0.00016	0.001	0.003	0	
23	7	0.526523814	A1A2B2B2	A2B2	0	0.00016	0.000	0.002	5	1
24	8	0.983501414	A2A2B2B2	A2B2	0	0.00016	-0.001	0.001	0	
25	9	0.248134927	A2A2B2B2	A2B2	0	0.00016	0.000	0.002	0	
26	10	0.374619958	A2A2B2B2	A2B2	0	0.00016	-0.002	0.000	0	

	A	B	C	D	E	F	G	H	I	J
406	390	0.35864632	A1A2B1B2	A1B1	0.01312	0.0012096	0.001	0.016	5	1
407	391	0.173536864	A2A2B2B2	A2B2	0	0.00016	-0.001	0.001	0	
408	392	0.928300352	A1A2B1B2	A2B2	0.01312	0.0012096	0.001	0.016	5	1
409	393	0.623074221	A2A2B2B2	A2B2	0	0.00016	0.000	0.002	0	
410	394	0.796832187	A1A2B1B2	A2B2	0.01312	0.0012096	0.001	0.017	5	1
411										272
412						Total number of selected individuals			0	122
413										

Figure 2.24 The ranges A1-J26 and A406-J412 in Sheet 1 of the 'Indirect Selection' computer model.

In order to assign values to the marker alleles so that marker-assisted selection could be simulated, a table was set up as shown in Figure 2.24 (cells E13-I14). To ensure that animals with at least one A_1 marker allele were selected, the following entries were made:

- In cell F14, a value of 5 was entered.
- In cell G14, a value of 0 was entered.
- In cell I15, a value of 0 was entered to ensure that only animals with a marker value of 5 or above were selected.

Headings were set up as shown in Figure 2.25. The table in cells W9-AA10 was constructed to assign genotypic values for %fat to the individuals in the population. Furthermore, in order to display the important population parameters, a selection statistics table was set up.

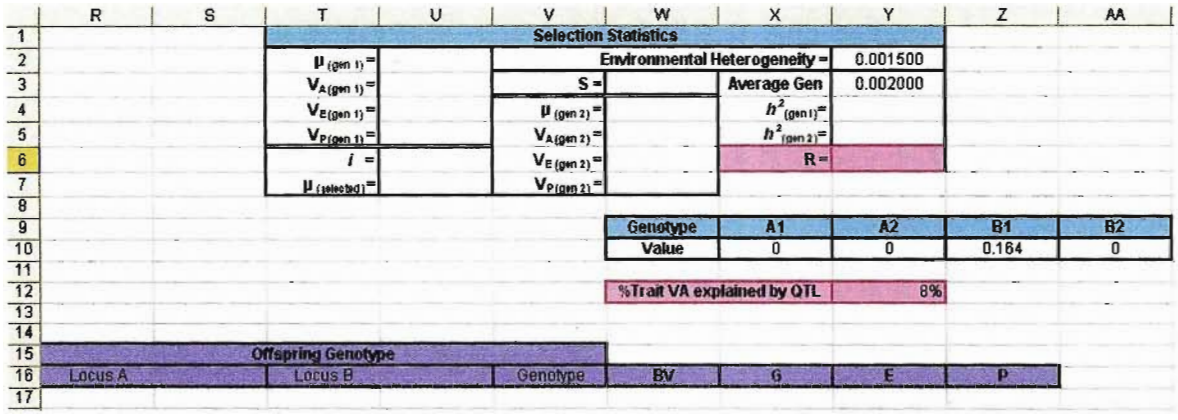


Figure 2.25 The range L1-16 in Sheet 1 of the 'Indirect Selection' computer model.

Mating amongst the selected individuals was simulated in the same way as was described for the 'Phenotypic Selection' model and the calculations for the observed and expected genotypic frequencies also remained the same as in the 'Random Mating' model (Table 2.3, Page 34).

The following data, taken from the study done by Weller *et al.* (2003), was incorporated into the 'Indirect Selection' model:

- The population size was set at 394 individuals..
- The genotypic frequencies observed in the population were entered into the cells B5-B13 (Figure 2.24).
- In cell E1 (Figure 2.24), the value '0.0125' was entered to set the recombination fraction at 0.0125.
- Environmental Heterogeneity was set at 0.0015 by entering 0.0015 into Y3 (Figure 2.25).
- The % trait additive variance explained by the QTL was set at 8 % by entering '8' into cell Y12 (Figure2.25).
- The value 0.164 was entered in cell Z10 (Figure 2.25) as this was the value of the DGAT1K allele, synonymous with the B_1 allele.
- In cell W3 (Figure 2.25), the value '0.002' was entered as the average genotypic value because that was the mean %fat in the population.

To calculate the breeding values of the individuals in the parental population, the following formula was entered into cell E17 and copied down to E410 (Figure 2.24):

- **=(LOOKUP(MID(C17,1,2),$V$9:$Y$9,$V$10:$Y$10)*$W$12)+(LOOKUP(MID(C17,3,2),$V$9:$Y$9,$V$10:$Y$10)*$W$12)+(LOOKUP(MID(C17,5,2),$V$9:$Y$9,$V$10:$Y$10)*$W$12)+(LOOKUP(MID(C17,7,2),$V$9:$Y$9,$V$10:$Y$10)*$W$12)**

To calculate the breeding values of the individuals in the offspring generation, the following formula was entered into cell U17 and copied down to cell U410 (Figure 2.25):

- **=IF(T17="","",((LOOKUP(MID(T17,1,2),$V$9:$Y$9,$V$10:$Y$10)*$W$12)+(LOOKUP(MID(T17,3,2),$V$9:$Y$9,$V$10:$Y$10)*$W$12)+(LOOKUP(MID(T17,5,2),$V$9:$Y$9,$V$10:$Y$10)*$W$12)+(LOOKUP(MID(T17,7,2),$V$9:$Y$9,$V$10:$Y$10)*$W$12)))**

To calculate the genotypic values of the parental generation, the following formula was entered into cell F17 and copied down to cell F410:

- **=($W$3+E17)*$W$12**

To calculate the genotypic values of the offspring generation, the following formula was entered into cell V17 and copied down to cell V410:

- **=IF(U17="","",U17*($W$12)+$W$3)**

The phenotypic values for both the parental and the offspring generation were calculated in the same way as the phenotypic values in the 'Marker-Assisted Selection' model were calculated.

The selection statistics table in Figure 2.25 was created as described in the 'Marker-Assisted Selection' model and the population parameters were calculated using the formulae shown in Table 2.5 (Page 41).

(c) Response to marker-assisted selection

In order to calculate the response to marker-assisted selection in the Israeli Holstein cattle population described by Weller *et al.* (2003), Sheet 2 was opened and the tables were set up as shown in Figure 2.26.

	A	B	C	D	E	F	G
1		Generation 1	Generation 2	Generation 3	Generation 4	Generation 5	
2	A1A1B1B1						
3	A1A1B1B2						
4	A1A1B2B2						
5	A1A2B1B1						
6	A1A2B1B2						
7	A1A2B2B2						
8	A2A2B1B1						
9	A2A2B1B2						
10	A2A2B2B2						
11							
12	Table 1. The change in genotypic frequencies over five generations						
13							
14		Generation 1	Generation 2	Generation 3	Generation 4	Generation 5	
15	p_2 = B1						
16	q_2 = B2						
17							
18	Table 2. Change in allele frequencies at the B locus over five generations						
19							
20		μ	V_A	θ_G	h^2	D	i
21	Gen 1			0.000000			
22	Gen 2			0.000000			
23	Gen 3			0.000000			
24	Gen 4			0.000000			
25	Gen 5			0.000000			
26	Table 3. Change in population mean (μ), genotypic variance (V_G), genotypic std. deviation (θ_G) heritability (h^2), linkage						
27	disequilibrium coefficient (D) and intensity of selection (i) over five generations						
28							
29		R					
30	θ	0					
31	After 1 Gen						
32	After 2 Gen						
33	After 3 Gen						
34	After 4 Gen						
35	After 5 Gen						
36							
37	Table 4. Response to Selection over five generations of marker-assisted selection						

Figure 2.26 The range A1-G37 in Sheet 2 of the 'Indirect Selection' computer model.

A macro was created and named 'MAS'. This macro was designed to automatically generate the following statistics over five generations of marker-assisted selection.

The macro was designed to generate:

- The genotypic frequencies of the parental population on Sheet 1 for five generations (Table 1, Figure 2.26).
- The allele frequencies at the marker locus over five generations (Table 2, Figure 2.26).
- The population mean, V_A, h^2, ***D*** and *i* over five generations (Table 3, Figure 2.26).
- To generate the R over five generations of marker-assisted selection (Table 4, Figure 2.26).

(d) Graphical depiction of response to marker-assisted selection

With the purpose of graphically depicting the effects of marker-assisted selection on the population over five generations, a third Sheet was then added to the 'Indirect Selection' model. The graphs shown on Sheet 3 are the graphic representations of the values displayed in Figure 2.26.

In order to demonstrate the response to marker-assisted selection utilizing an indirect marker in an Israeli Holstein cattle population, the graph entitled 'Response to Marker-Assisted Selection' was created on Sheet 3:

- The chart wizard was selected and the XY (Scatter) option was chosen.
- The range B30-B35 on Sheet 2 was selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'Response.'

The change in allele frequency at the locus containing the QTL of interest was illustrated in the graph named 'Change in Allele Frequency due to Marker-Assisted Selection' on Sheet 3:

- The chart wizard option was selected and the XY (scatter) option was chosen.
- The range A14-F16 on Sheet 2 was selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'Frequency.'

The relationship between the frequency of the B_1 allele and the increase in the mean milk fat percentage in the population was graphically depicted as follows:

- The chart wizard was selected and the XY (Scatter) option was chosen.
- The ranges A15-F15 and B21-B25 on Sheet 2 were selected as the data ranges.
- The X axis was labeled 'Frequency B_1' and the Y axis was labeled 'Population Mean (%)'.
- The graph was named 'Relationship between the Frequency of the B_1 Allele and Population mean.'

To illustrate the increase in milk fat percentage over five generations of marker-assisted selection in the population, the following graph was created:

- The chart wizard was selected and the column graph option was chosen.
- The range B21-B25 on Sheet 2 was selected as the data range.
- The X axis was labeled 'Generations' and the Y axis was labeled 'Milk Fat %'.
- The graph was entitled 'Increase in Mean %Fat in a Holstein cattle Population as a result of Marker-Assisted Selection.'

CHAPTER 3
RESULTS AND DISCUSSION

3.1 INTRODUCTION

The computer models that were developed in this investigation simulated populations undergoing either random mating, phenotypic selection or marker-assisted selection. The results generated from the different models focused primarily on comparing the response to phenotypic selection and marker-assisted selection over five generations and the consequent effect of these two selection strategies on genetic variation in a population. Important population parameters such as h^2 and i were also calculated and compared. Furthermore, the dependency of the success of marker-assisted selection on both the amount of linkage disequilibrium between two loci in a population and the percentage of trait variation accounted for by a QTL was simulated.

It was established that the results simulated in this investigation were consistent with the results obtained from numerous other studies. The predictive power of the computer models created for the purpose of this investigation may therefore be useful in the assessment of different breeding strategies prior to their implementation.

3.2 DEMONSTRATION OF THE EFFECTS OF RANDOM MATING AND LINKAGE DISEQUILIBRIUM

Aim 1: To simulate the effects of linkage disequilibrium between two loci in a randomly mating population

The first model simulated the effects of a randomly mating population in linkage disequilibrium. The association between the genotypes at two loci, designated A and B, were compared for a population in linkage equilibrium and linkage disequilibrium by studying the association of the genotypes at the two loci in an offspring generation.

Furthermore, the change in the genotypic frequencies was analyzed over time by studying the allele frequencies, ***D*** and V_P over five generations.

3.2.1 Association between genotypes at two loci

The computer model indicated that the association of the two genotypes, designated A and B, in a population was directly related to the amount of recombination that existed between the loci. When there was no recombination between the two loci, Figure 3.1 shows that the A_1A_1 genotype only associated with the B_1B_1 genotype, as did the A_1A_2 genotype with the B_1B_2 genotype and the A_2A_2 genotype with the B_2B_2 genotype.

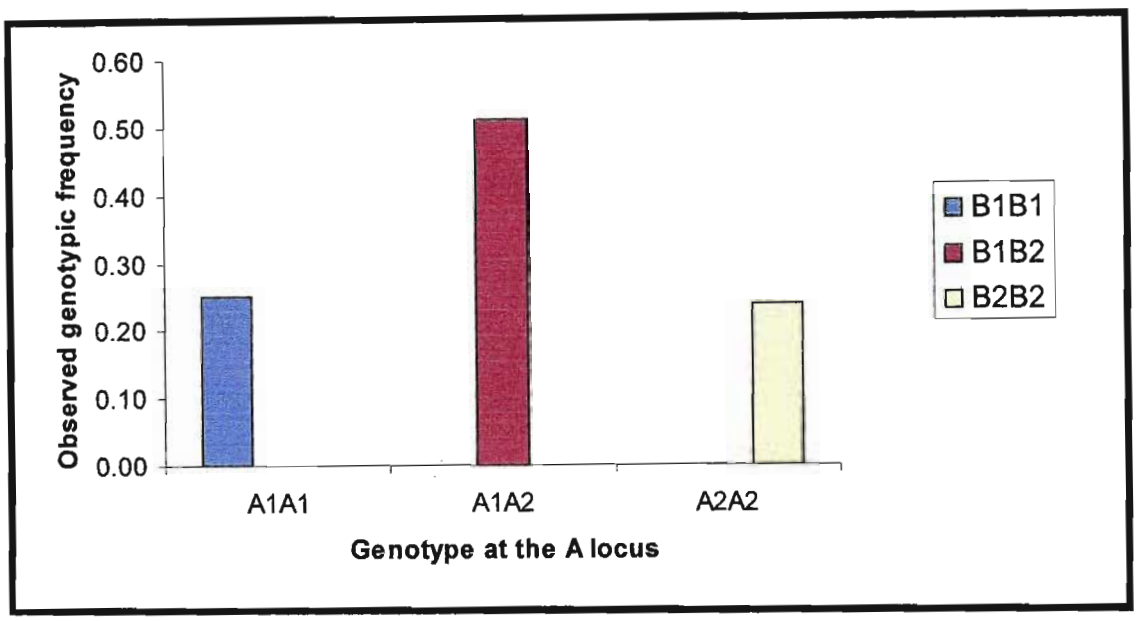


Figure 3.1 The association of the genotypes at the B locus with the genotypes at the A locus, when r = 0.

These results show that when the A and B loci were completely linked (r = 0), there was a non-random association of alleles at the A locus with the alleles at the B locus. The consequent genotypic frequencies were indicative of a population in linkage disequilibrium (Figure 3.1). Population-wide linkage disequilibrium can be found when two loci are positioned next to one another along a chromosome, such that crossing over cannot occur between the loci and the recombination fraction between them will be 0 (Babu *et al.*, 2004). This situation was thus accurately simulated in a randomly mating population since the results showed the exclusive association

between the A_1 and B_1 alleles and the A_2 and B_2 alleles when the A and B loci were in linkage disequilibrium.

In contrast, Figure 3.2 shows that when r was 0.5 between the two loci, the genotypes at the B locus were randomly associated with the genotypes at the A locus, such that there was no preferred association between alleles at the two loci.

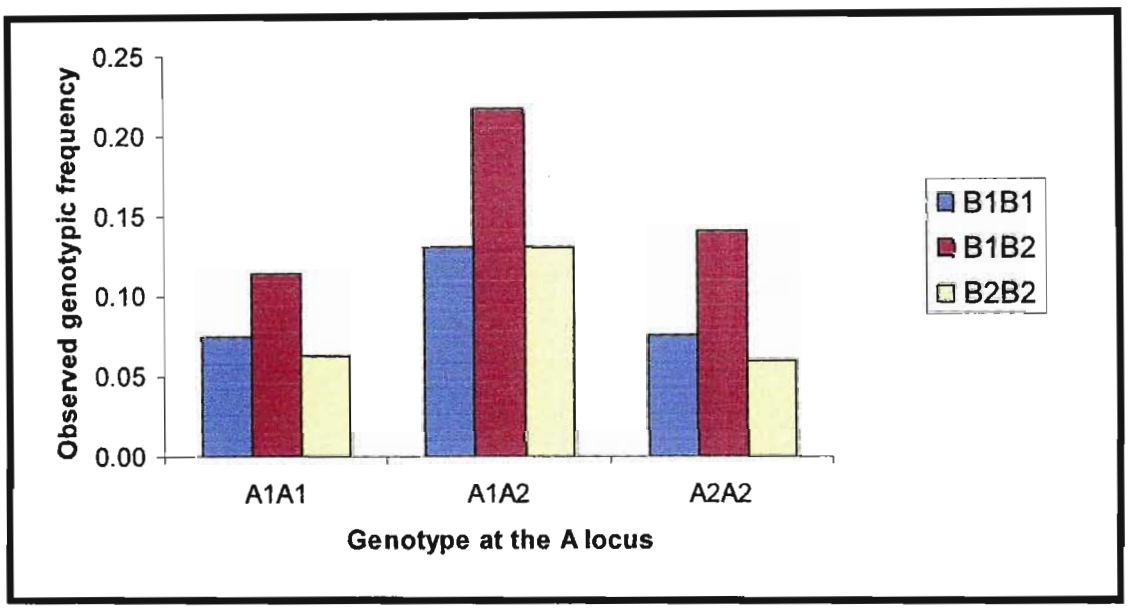


Figure 3.2 The association of the genotypes at the B locus with the genotypes at the A locus, when r = 0.5.

According to Snustad and Simmons (2000), when the recombination fraction between two loci is 0.5, crossing over between the loci is inevitable and consequently the loci will be in linkage equilibrium in the population. The results demonstrated in Figure 3.2 thus represented a population in linkage equilibrium where the A and B loci were either positioned far apart on one chromosome or on two different chromosomes altogether.

3.2.2 Change in allele frequency

While the afore-mentioned association of the genotypes at the B locus with the genotypes at the A locus were measured in one offspring generation, the following changes in allele frequency, genotypic frequency, ***D*** and V_P were measured over five generations of random mating. The recombination fraction was set at 0.5, the

environmental heterogeneity was set at 0.3, the population mean was set at 50, the frequencies of $A_1A_1B_1B_1$ and $A_2A_2B_2B_2$, were both set at 0.5 and the values of the B_1 and B_2 alleles were arbitrarily set at 3 and 0 respectively.

In order to study the change in allele frequency over time for a population in linkage disequilibrium, the genotypic frequencies of the two double homozygotes, $A_1A_1B_1B_1$ and $A_2A_2B_2B_2$, were both set at 0.5 in generation 1 (Figure 3.3). As a result the allele frequencies of the A_1, A_2, B_1 and B_2 alleles were all equal to 0.5 in generation 1. The change in allele frequencies at both loci over five generations of random mating is shown in Table 3.1. Although the allele frequencies fluctuated slightly, they remained very similar over the five generations, regardless of the amount of linkage disequilibrium between the two loci.

Table 3.1 Changes in allele frequencies over five generations of random mating.

Allele Frequency	Gen 1	Gen 2	Gen 3	Gen 4	Gen 5
$p_1 = A_1 =$	0.500	0.480	0.532	0.562	0.553
$q_1 = A_2 =$	0.500	0.520	0.468	0.438	0.447
$p_2 = B_1 =$	0.500	0.481	0.537	0.571	0.552
$q_2 = B_2 =$	0.500	0.519	0.464	0.430	0.448

The allele frequencies shown in Table 3.1 were therefore consistent with the principles of the Hardy-Weinberg principle which states that in a large, randomly mating population, in the absence of evolutionary forces, the allele frequencies remain constant from generation to generation (Raven and Johnson, 1999).

The particular associations of the alleles at the A and B loci into genotypes, however, did not remain constant and were responsible for changes in genotypic frequencies over time.

3.2.3 Change in genotypic frequencies

Changes in genotypic frequencies in a population are indicative of the amount of linkage disequilibrium that exists between two loci and are dependent on the frequency of a particular genotype; the double heterozygote $A_1A_2B_1B_2$ (Falconer and Mackay, 1996). In the first generation, only the two double homozygotes were present in the population which meant that recombination could not breakdown the amount of

linkage disequilibrium present in generation 1. When maximum linkage disequilibrium was initially present between the A and B loci, the genotypic frequencies fluctuated in the first three generations before approaching their equilibrium values in generation 4 (Figure 3.3).

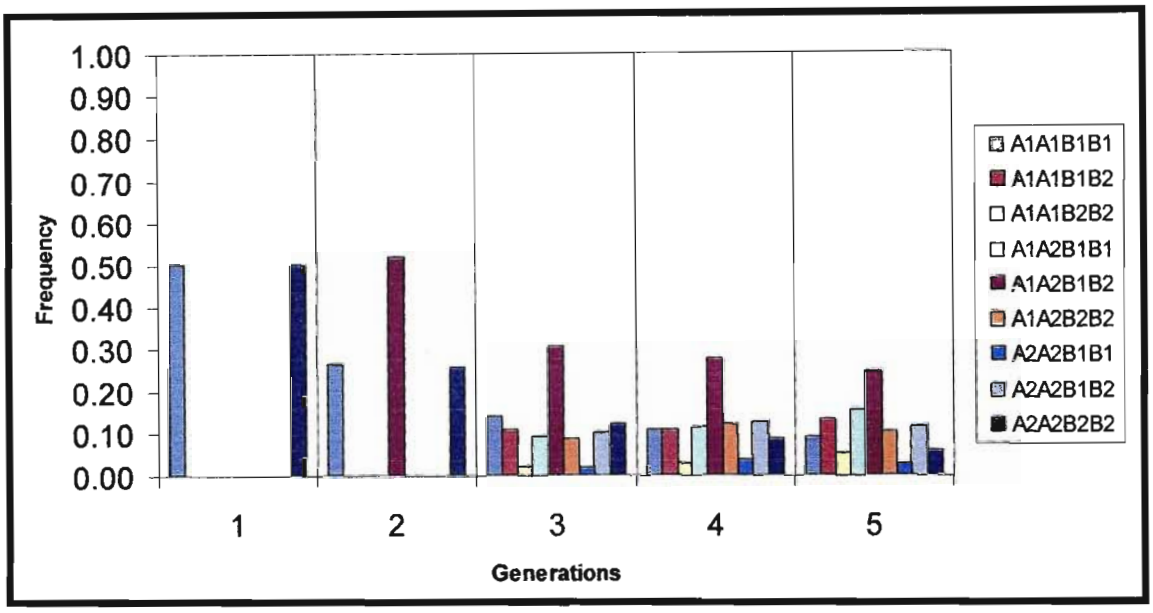


Figure 3.3 Change in genotypic frequencies over five generations of random mating.

In the first generation, only the double homozygous genotypes were present in the population each occurring at a frequency of 0.5 (Figure 3.3). In generation 2, the frequencies of the double homozygotes were reduced and the double heterozygotes appeared with a frequency of approximately 0.5. The frequencies of both the double homozygotes and the double heterozygotes decreased by a significant proportion of their initial value in generation 3 and the frequencies of all the other possible genotypes increased from an initial frequency of 0. The genotypic frequencies seen in generations 4 and 5 were similar, suggesting that the genotypic proportions had reached equilibrium values.

It could thus be concluded that the exclusive association of the A_1 allele with the B_1 allele and the A_2 allele with the B_2 allele in generation 1 (Figure 3.3) was broken down over time by recombination and the presence of the double heterozygote in generation 2 such that no preferred allelic association was evident in generation 5 (the loci were in linkage equilibrium). This result is in accordance with the Hardy-Weinberg law that states that the genotypic proportions will only remain constant over

time once linkage equilibrium is reached between the A and B loci (Donovan and Welden, 2002).

Another way to view the breakdown of the association between the alleles at the A and B loci, is to study the change in ***D*** over time.

3.2.4 Change in linkage disequilibrium coefficient

The linkage disequilibrium coefficient measures the amount of linkage disequilibrium that exists between two loci (Hartl and Clark, 1997). Figure 3.4 demonstrates how the linkage disequilibrium coefficient is broken down over five generations due to the recombination occurring between the A and B loci.

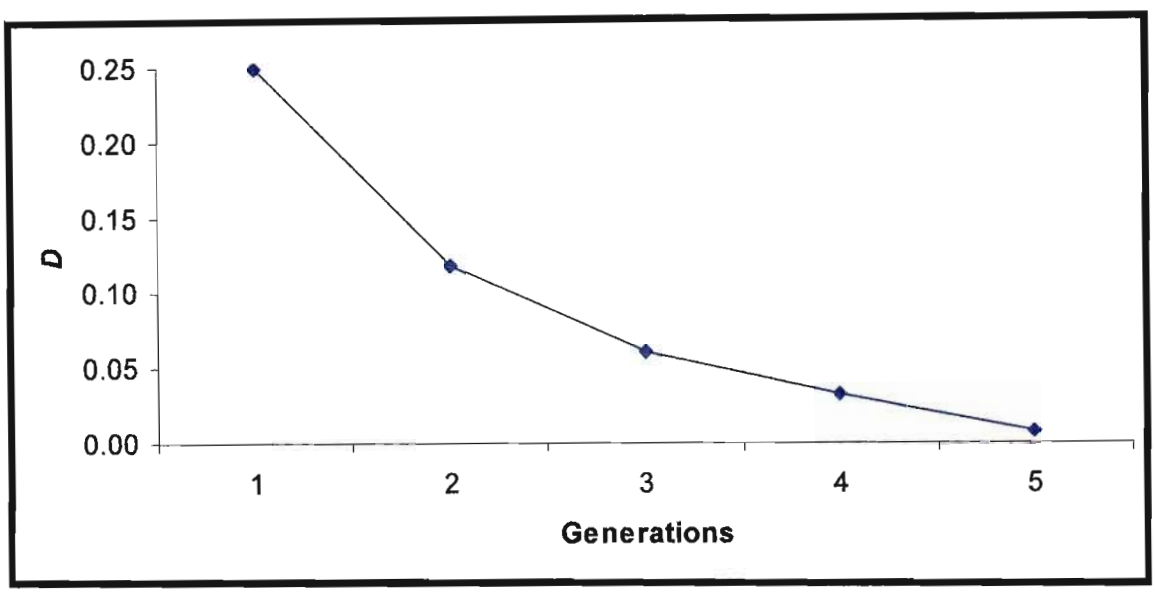


Figure 3.4 Breakdown of the linkage disequilibrium coefficient (***D***), measured over five generation of random mating.

In generation 1, when only the two double homozygotes were present in the population (Figure 3.3), ***D*** was at its maximum value of 0.25 (Figure 3.4), despite the fact that a recombination fraction of 0.5 existed between the two loci. This result was expected because recombination does not affect the gametic output of genotypes with one or more pairs of genes in the homozygous condition because the recombinant gametes will always be identical to the parental gametes (Gillespie, 2004). In generation 2, however, the double heterozygote was formed through the combination of the A_1B_1 and A_2B_2 gametes and as a result, the value of ***D*** decreased to

approximately 0.125, almost half of its initial value (Figure 3.4). This occurred because recombination was able to reshuffle the parental associations of the double heterozygotes and produce both parental (A_1B_1/A_2B_2) and recombinant (A_1B_2/A_2B_1) gametes (Falconer and Mackay, 1996).

The parental and recombinant gametes produced by the double heterozygote, were responsible for the formation of the nine different genotypes evident from generation 3 onwards (Figure 3.3). As a result, ***D*** steadily declined from 0.125 in generation 3 to just over 0.5 in generation 5 (Figure 3.4). The linkage disequilibrium present in generation 1 and 2 was thus largely depleted by generation 5, as linkage equilibrium was approached.

This result was an accurate reflection of the results published by Abecasis *et al.* (2005) which stated that over time, recombination breaks down the amount of linkage disequilibrium between two loci in a population until the loci are no longer linked and are said to be in linkage equilibrium.

3.2.5 Change in phenotypic variance

Phenotypic variance measures the amount of phenotypic variation that is present in a population. It is important to calculate the amount of variation in each generation in a population because without variation, populations are unable to adapt to changing environmental conditions and evolve (Raven and Johnson, 1999). The amount of V_P in generation 1 decreased by approximately 45% in generation 2, increased by about 2% in generation 3 and then remained constant over time (Figure 3.5).

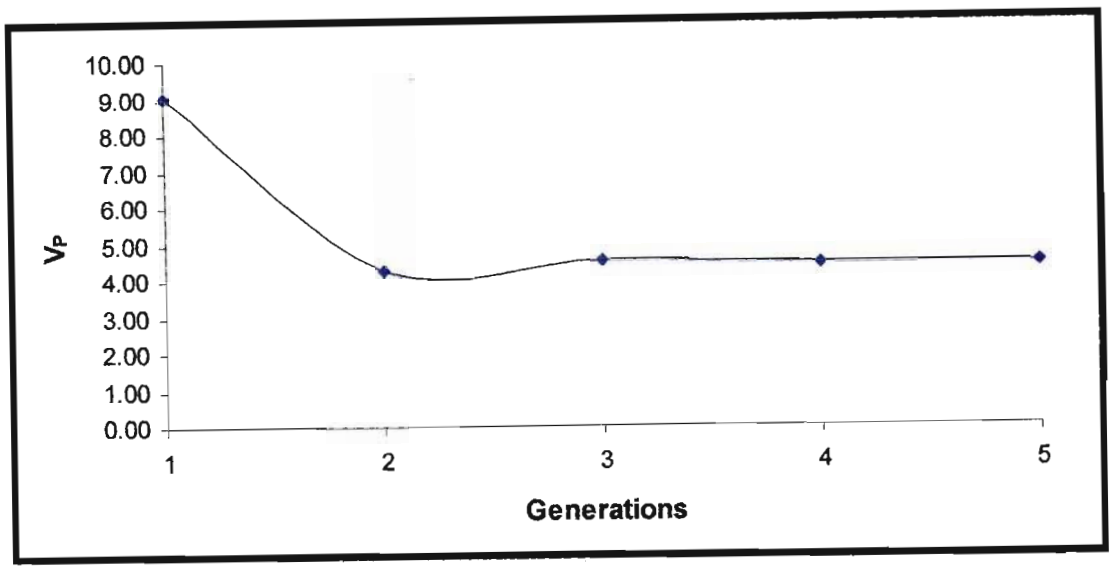


Figure 3.5 Change in phenotypic variance (V_p) over five generations of random mating

The phenotypic variance was just over 9 in generation 1 and it decreased to approximately 5 in generation 2 (Figure 3.5). The V_P then increased in generation 3 and remained constant from generation 4 onwards. This is because in a randomly mating population with no selection, the V_P may fluctuate slightly from one generation to the next, but a consistent level of variation will be maintained over time as all the alleles remain present in the population (Hartl and Clarke, 1997).

The change in V_P over time (Figure 3.5) can also be explained by the breakdown of linkage disequilibrium between the two loci illustrated in Figure 3.4. In the first three generations, the value of ***D*** indicated that the A and B loci were linked and therefore the phenotypic frequencies fluctuated in these two generations. After three generations, however, ***D*** was eroded as the loci approached linkage equilibrium and consequently the V_P after three generations remained constant. This result can be explained by the fact that once linkage equilibrium is reached, the genotypic proportions remain the same over time (Donovan and Welden, 2002), and in a constant environment, the V_P will also remain the same.

The effects of linkage disequilibrium between two loci in a randomly mating population were therefore accurately simulated by the 'Random Mating' model developed in this investigation, since the results of the simulation were consistent with the literature. In particular, the combined effects of recombination and the presence of

the double heterozygotes on the breakdown of linkage disequilibrium between two loci in a population were successfully demonstrated.

3.3 DEMONSTRATION OF THE EFFECT OF SELECTION WITH NO DOMINANCE

Aim 2: To simulate the effects of both phenotypic selection and marker-assisted selection with the assumption of no dominance at the loci of interest

3.3.1 Phenotypic selection

The success of phenotypic selection was assessed by studying the response to phenotypic selection and its effect on various population parameters over five generations.

The following results were obtained from the 'Phenotypic Selection' model when r=0.2, the environmental heterogeneity = 0.3, the population mean = 50, the frequency of the $A_1A_1B_1B_1$ genotype = 0.3, the frequency of the $A_1A_2B_1B_2$ genotype = 0.4 and the frequency of the $A_2A_2B_2B_2$ genotype = 0.3. These values were used in order to ensure that the effect of ***D*** and r on the genotypic frequencies was clearly demonstrated. The values of the B_1 and B_2 alleles were arbitrarily set at 3 and 0 respectively.

(a) Response to selection

The response to selection is a direct measure of the success of a particular selection strategy in increasing the overall population mean (van der Werf, 2000b). The decline in response to phenotypic selection over five generations, using the 'Phenotypic Selection' model, is shown in Figure 3.6. Individuals who had a phenotypic value higher than 53, were selected.

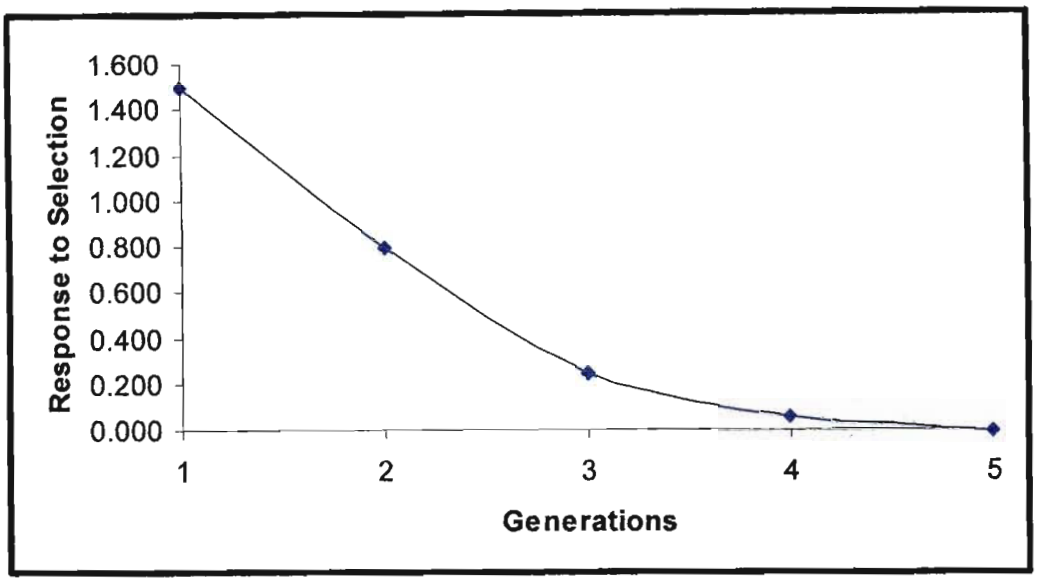


Figure 3.6 Response to selection over five generations of phenotypic selection.

After one generation of phenotypic selection, R was approximately 1.5. After the second generation, R dropped to approximately 0.8 and then again to approximately 0.2 after the third generation. After five generations, R was virtually 0. The response to selection was therefore significant in the first three generations, but after four generations of phenotypic selection, there was no longer any significant response.

(b) Effect of phenotypic selection on population parameters

While the aim of phenotypic selection is to increase the population mean, population parameters such as V_P, h^2, ***D*** and *i* are also affected. Since phenotypic selection makes use of the phenotypic values observed in a population, V_P was the parameter used to measure variation in the population. The effect of phenotypic selection on these different population parameters was measured over five generations (Table 3.2).

Table 3.2 Change in population mean (μ), phenotypic variance (V_P), heritability (h^2), linkage disequilibrium coefficient (***D***) and intensity of selection (*i*) over five generations of phenotypic selection.

Generation	μ	V_P	h^2	*D*	*i*
1	53.157	5.506	0.984	0.210	0.816
2	54.691	3.139	0.967	0.070	0.388
3	55.488	1.762	0.913	0.026	0.226
4	55.789	1.101	0.846	0.022	0.165
5	55.834	0.350	0.757	0.010	0.078

As is evident in Table 3.2, the population mean increased after the first two generations of phenotypic selection by 1.534 and 0.797 respectively and then increased by 0.346 from generation 3 through to generation 5. Phenotypic variance was at its highest value, 5.506, in the first generation and then dropped to 3.139 in generation 2 and 1.762 in generation 3. Phenotypic variance continued to decrease, although at a slower rate, after generations 3, until the amount of V_P in generation 5 was only 0.350. The heritability also showed a steady decline from 0.984 to 0.757 over five generations of phenotypic selection. In a similar way, ***D*** decreased from a value of 0.210 in generation 1, to 0.010 in generation 5. Finally *i* showed its maximum value, 0.816, in generation 1, dropped sharply to 0.388 in generation 2 and continued to steadily decline to 0.078 in generation 5.

The population mean was therefore the only parameter that increased in value over the five generations of phenotypic selection, while important parameters such as V_P, ***D*** and h^2 showed a steady decline.

(c) Factors affecting response to selection

For any type of selection strategy to be successful, genetic variation is needed in a population (Snustad and Simmons, 2000). In order to illustrate the effect of variation on R, σ_P was calculated for each generation, as the square root of V_P (Table 3.2) and compared to R (Figure 3.7).

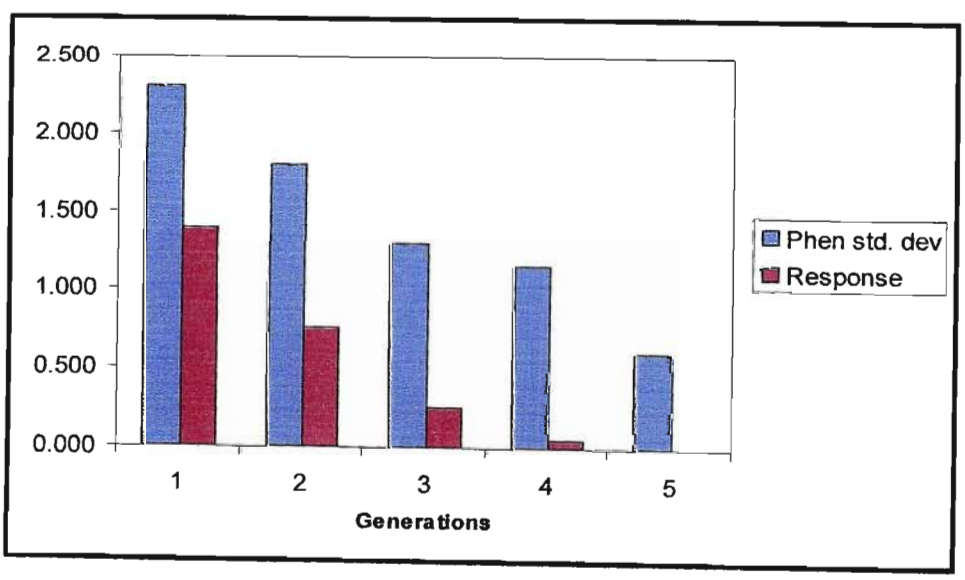


Figure 3.7 The change in phenotypic standard deviation and response to selection over five generations of phenotypic selection.

In generation 1, σ_P was at its highest value and R, after generation 1, was also at its highest value. Both parameters showed a steady decline over five generations such that σ_P displayed its lowest value in generation 5 and the corresponding response was 0. A direct relationship therefore existed between R and the amount of phenotypic variation present in the population.

(d) Discussion of results from the 'Phenotypic Selection' model

The process of phenotypic selection involves choosing individuals in a population, with phenotypic values above certain selection criteria, to breed (van der Werf, 2000b). This supposedly results in animals with similar, superior genotypes being chosen to pass on their genes to the next generation and can therefore lead to a narrowing of the genetic diversity in a species and reduce the genetic variation available for future selection (Williams, 2005). This reduction in additive variance over time causes a reduction in V_P over time, which was evident in Table 3.2. The reduction in genetic variation over five generations of phenotypic selection was responsible for the recorded decline in h^2 and i over time (Table 3.2). The results presented in Table 3.2 show that h^2 was affected the least by the decline in V_P. Since h^2 measures the proportion of additive variance over phenotypic variance, if both the additive and phenotypic variance was reduced in the population, the heritability estimate would not be expected to change significantly. In contrast, as phenotypic selection increased the population mean, more individuals in the population met the selection criterion and were consequently selected. According to Falconer (1989), as the proportion of selected individuals increases, so the intensity of selection decreases. This concept was demonstrated by the decline in the value of i in Table 3.2. Although practically, breeders control the proportion of selected individuals (keep i constant), one of the assumptions of this investigation was that the overall population size remained constant from one generation to the next. Instead of remaining constant, the value of i therefore changed over time and was recorded by the computer model.

According to the equation $\mathbf{R = (h^2)(\sigma_P)(i)}$, the response to phenotypic selection is directly related to h^2, σ_P and i in a population (van der Werf, 2000b). The response to phenotypic selection, as shown in Figure 3.6, was at its highest after the first generation of selection because V_P, h^2 and i were at their highest values in generation

1 (Table 3.2). The response progressively decreased after each generation, as did V_P, H^2 and *i*, although, since h^2 showed the smallest decline over five generations, it would have had the smallest effect on the decrease in response. After five generations, the genetic variation was almost entirely depleted in the population and consequently V_P, *i* and R, were virtually zero as expected.

The phenotypic standard deviation was plotted against R for five generations in Figure 3.7 and the results reiterated the fact that the lower the genetic variation in the population, the lower σ_P and therefore R.

The 'Phenotypic Selection' model developed in this study therefore accomplished its aim of accurately simulating the response to phenotypic selection as well as the effects of phenotypic selection on population parameters such as *i*, V_P and h^2.

3.3.2 Marker-assisted selection

The response to marker-assisted selection and the effect of marker-assisted selection on different population parameters was studied over five generations. The impact of selection for both individuals heterozygous at the marker locus (A_1A_2) and individuals homozygous for the A_1 marker allele (A_1A_1), was compared to the impact of selection for only A_1A_1 individuals, with regards to both μ and V_A. Additive variance was used to measure the amount of variation present in the population since marker-assisted selection provides direct information about individuals' genotypic values.

The following results were obtained from the 'Marker-Assisted Selection' model when r = 0.2, the environmental heterogeneity = 0.3, the population mean = 50, the frequency of the $A_1A_2B_1B_2$ genotype = 1 and the values of the B_1 and B_2 alleles were 3 and 0 respectively. The percentage of trait variance accounted for by the QTL was set at 100%. In the initial population, the frequency of the double heterozygote ($A_1A_2B_1B_2$) was set at 1, in order to illustrate the dependency of R on genetic variation in a population. The arbitrary values of the alleles were kept the same as those in the previous section.

(a) Response to selection for both A_1A_2 and A_1A_1 individuals

When marker-assisted selection for individuals heterozygous at the marker locus or homozygous for the A_1 marker allele was simulated, a significant response was evident in generation 2, 3 and 4, while in generation 5, R was equal to 0 (Figure 3.8).

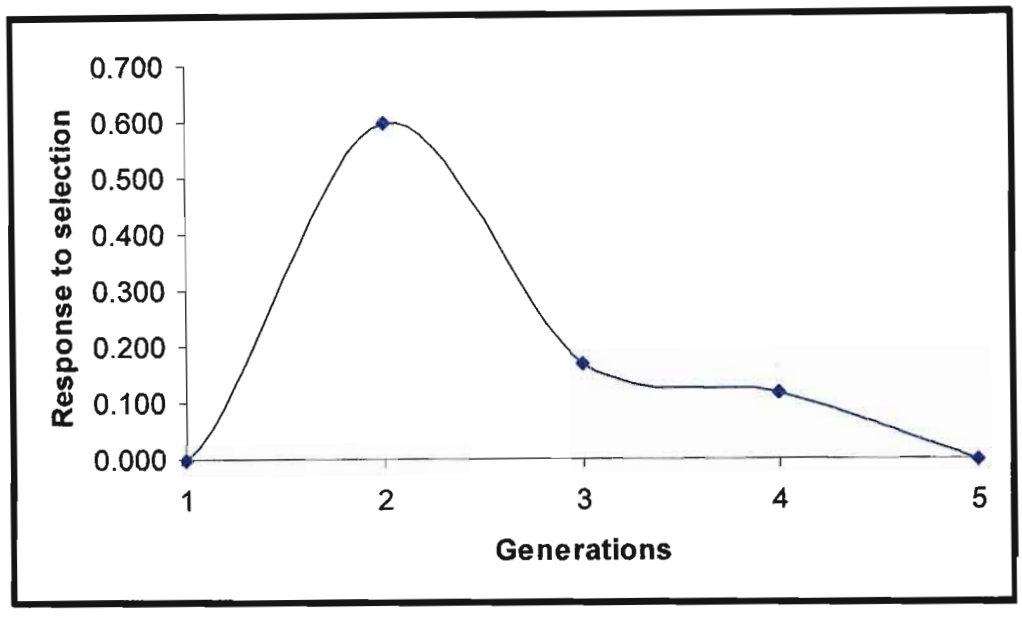


Figure 3.8 Response to selection over five generations of marker-assisted selection.

The response to selection after one generation of marker-assisted selection was 0 (Figure 3.8). After the second generation of marker-assisted selection, however, the response increased to just under 0.6. The response dropped to approximately 0.17 after generation 3 and to just over 1 after the fourth generation. After five generations, however, there was no further response to marker-assisted selection.

The response to selection is dependent on the effectiveness of a particular selection strategy in increasing the frequency of the favourable allele in the population (Williams, 2005). The allele frequencies at the locus under selection therefore needed to be examined.

(b) Change in allele frequency at the B locus

Theoretically, by directly selecting for the A_1 marker allele, marker-assisted selection indirectly increases the frequency of the favourable B_1 allele at the locus under

selection, if the A_1 and B_1 alleles are linked (Xie and Xu, 1998).

The capacity of Marker-Assisted Selection to increase the frequency of the B_1 allele over time is demonstrated in Figure 3.9.

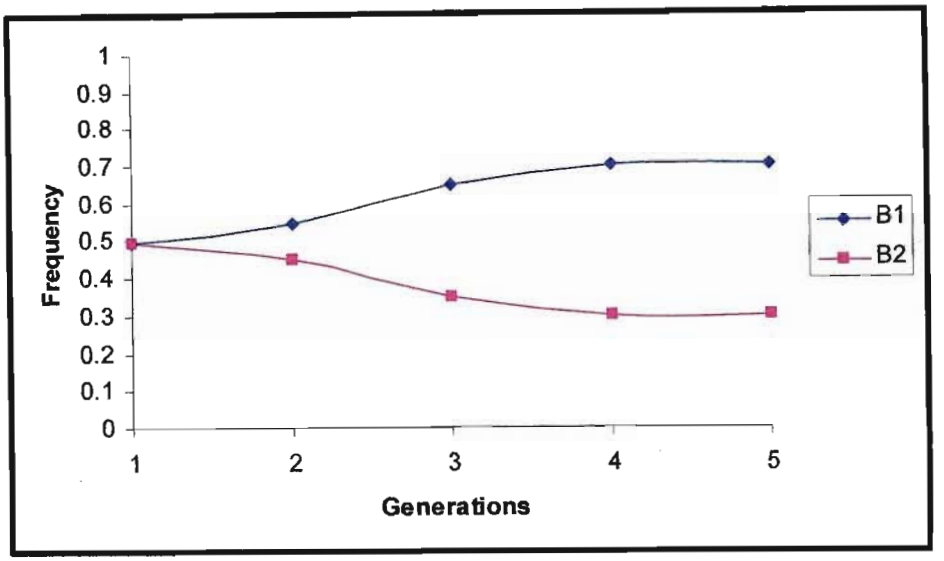


Figure 3.9 Change in allele frequencies over five generations of marker-assisted selection.

The B_1 and B_2 alleles both had frequencies of 0.5 in generation 1 (Figure 3.9). The frequency of the B_1 allele increased slightly to 0.55 after one generation and then to 0.65 by generation 3. In generation 4, the frequency of the B_1 allele had increased to 0.7 but remained at approximately 0.7 in generation 5. As the frequency of the B_1 allele continued to increase over the five generations, the frequency of the B_2 allele progressively decreased. Thus marker-assisted selection effectively increased the frequency of the B_1 allele, through selection for the A_1 allele, over five generations.

(c) Effect of marker-assisted selection on various population parameters

The change in various population parameters over five generations of marker-assisted selection was measured and it was found that a decrease in V_A and $\boldsymbol{D}$ were direct consequences of marker-assisted selection (Table 3.3).

Table 3.3 Change in population mean (μ), additive variance (V_A), heritability (h^2), linkage disequilibrium coefficient (***D***) and intensity of selection (*i*) over five generations of marker-assisted selection.

Generation	μ	V_A	h^2	*D*	*i*
1	52.986	0.000	0.000	0.159	0.000
2	52.994	4.318	0.972	0.117	0.274
3	53.537	4.504	0.973	0.111	0.091
4	53.754	3.916	0.978	0.078	0.059
5	53.850	3.888	0.984	0.067	0.001

The information provided by Table 3.3 indicates that the population mean increased from 52.986 to 53.850 over five generations of marker-assisted selection, although the majority of this increase was seen in the first three generations. The additive variance had a value of 0 in generation 1 and this value increased to 4.318 in generation 2. From generation 2 through to generation 5, the V_A gradually declined to a value of 3.888. The breakdown of linkage disequilibrium between the A and B loci was shown by the values of ***D***, which was initially 0.159 in generation 1 and decreased every generation until it reached 0.067 by the fifth generation. The intensity of selection had a value of 0 in generation 1 but increased to 0.274 in generation 2. The intensity of selection was only 0.91 in generation 3 and decreased further to 0.059 in generation 4 and virtually 0 in generation 5.

According to the results presented in Table 3.3, in generation 1, the frequency of the double heterozygotes ($A_1A_2B_1B_2$) in the population was 1 and therefore there was no additive variance. Since all the individuals in the population were genetically identical, *i* was also 0, because all the individuals were heterozygous for the marker locus (A_1A_2) and thus 100% of the population was selected to breed. As a result, the response to marker-assisted selection, as illustrated in Figure 3.8, after the first generation was 0.

Due to the decrease in V_A from generation 2 through to generation 5, h^2 was also expected to decrease, however, h^2 remained close to a value of 1 throughout the five generations of marker-assisted selection (Table 3.3). One explanation for this could be that because marker-assisted selection selects individuals on the basis of their DNA value, h^2 was not confounded by environmental effects and therefore remained high (Lande and Thompson, 1990).

The efficiency of marker-assisted selection is dependent on the amount of linkage disequilibrium that exists between the marker locus and the locus of interest and therefore R is directly related to ***D*** (Dekkers, 2004). The recombination fraction of 0.2, together with the presence of the double heterozygous individuals, ensured that ***D*** was broken down over time (Table 3.3) until the loci were in linkage equilibrium in generation 5. This is another reason why Figure 3.8 indicates that there was virtually no response to marker-assisted selection after five generations.

While Table 3.3 shows the actual change in values of the different statistics over five generations, Figure 3.10 demonstrates the direct relationship between the trends of certain statistics, namely R, *i* and σ_A.

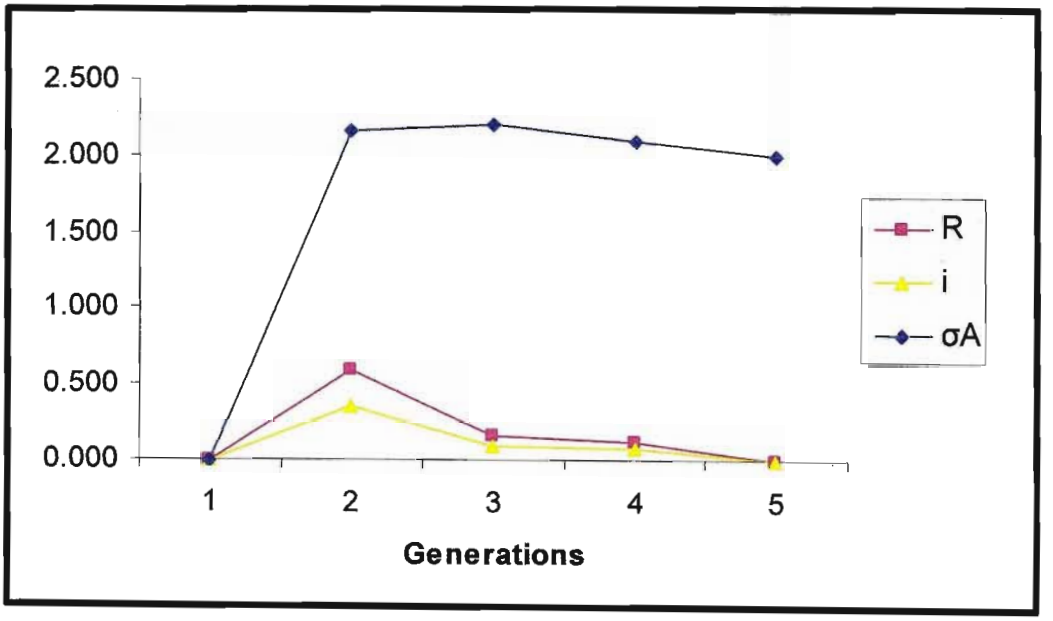


Figure 3.10 Relationship between response to selection (R), intensity of selection (*i*) and genetic standard deviation (σ_A).

In generation 1, σ_A and *i* both had values of 0 and after one generation, the response to marker-assisted selection, was also 0 (Figure 3.10). In generation 2, σ_A and *i* increased, as did R, after the second generation. From generation 3 to generation 5, both σ_A and *i* decreased slightly and R also decreased by the same proportion as a result. In generation 5, *i* reached a value of 0 and consequently, after the fifth generation of marker-assisted selection, R was also 0.

The dependency of R on V_A and *i*, (Figure 3.10) showed that the highest response to selection occurred after generation 2, which displayed the maximum value of V_A and *i*, and as V_A and *i* decreased over time, so did the response. However, while the decrease in V_A over time was gradual, *i* dropped rapidly to a value of 0 in generation 5 and was therefore largely responsible for the sharp decrease in response, seen in Figure 3.9, after the second generation.

Table 3.3 and Figure 3.10 demonstrate the effects of marker-assisted selection on various population parameters. Furthermore, these results illustrate the direct relationship between V_A, *i* and R over five generations of marker-assisted selection.

The effects of any marker-assisted selection strategy, however, are strongly dependent on the selection criteria and thus marker-assisted selection for individuals homozygous for the A_1 marker allele and for individuals heterozygous at the marker locus, should yield different results.

(d) Marker-assisted selection for individuals homozygous for the A_1 marker allele

The consequences of marker-assisted selection for individuals homozygous for the A_1 marker allele are examined in terms of the increase in population mean and the resulting effects on the additive variance. The frequency of the $A_1A_1B_1B_1$ and $A_2A_2B_2B_2$ individuals in the initial population were both 0.5. The frequency of the $A_1A_2B_1B_2$ individuals could not be set at 1, as done in section 3.3.2 (a), because only individuals homozygous for the A_1 marker allele could be selected. The effects of marker-assisted selection on the population mean and additive variance, when only the individuals homozygous for A_1 were selected, is shown in Table 3.4.

Table 3.4 Change in population mean (μ) and additive variance (V_A) over five generations of marker-assisted selection for individuals homozygous for the A_1 marker allele.

Generation	μ	V_A
1	52.937	5.452
2	56.007	0.000
3	55.996	0.000
4	56.009	0.000
5	56.004	0.000

In generation 1, V_A had a value of 5.452 and the population mean increased from 52.937 to 56.007 after the first generation of marker-assisted selection. In generation 2, however, the additive variance was entirely depleted in the population and remained at a value of 0 till generation 5. As a result, the population mean showed no significant change after the first generation of marker-assisted selection. It is interesting to note that μ does show small changes in value in each generation. This, however, is due to the computer simulation automatically generating new genotypes every generation and the changes are not significant in terms of studying the effects of marker-assisted selection.

These results (Table 3.4) suggest that selection for individuals homozygous for the A_1 marker allele in a population, where the frequency of each of the double homozygotes was 0.5, causes a substantial response of approximately 6% after the first generation. This was due to the fact that the A_1 allele became fixed at the A locus, and since recombination has no effect on homozygous individuals, the fixation of A_1 concomitantly fixed the B_1 allele at the B locus.

Although the population mean showed a significant increase from generation 1 to generation 2, the only response to selection for two copies of the A_1 marker allele was seen after the first generation of marker-assisted selection.

(e) Discussion of results from the 'Marker-Assisted Selection' model

As with phenotypic selection, marker-assisted selection for a particular trait results in a decrease in the genetic variation of a population (Dekkers and Hospital, 2002). This reduction in genetic variation was responsible for the decrease in V_A over time when both individuals heterozygous at the marker locus and individuals homozygous for the A_1 marker allele were selected as well as when only individuals homozygous for the A_1 marker allele were selected. Stella *et al.* (2002) simulated the implementation of several strategies for repeated application of marker-assisted selection and compared the short term and continual genetic response. It was found that economically important genes move toward fixation via the forces of selection and this fixation results in the reduction of gains in later generations. The advantages of marker-assisted selection therefore decreased with each generation.

Although selection for only individuals homozygous for the A_1 marker allele caused substantial increase in population mean after the first generation, it was also responsible for the depletion of genetic variation in the population. Since R is dependent on genetic variation in a population, the population mean showed no increase after the first generation when V_A was 0. In contrast, when both individuals homozygous for the A_1 marker allele and individuals heterozygous at the marker locus were selected, R was not as large initially, but the population mean did continue to increase over five generations.

The results simulated by the 'Marker-Assisted Selection' model were therefore successful in showing the effects of marker-assisted selection on R and various population parameters, namely V_A, h^2 and *i*. Furthermore, the 'Marker-Assisted Selection' model was able to provide a comparison between the response to selection and the effects of marker-assisted selection on additive variance in a population, when the selection criterion was varied.

Aim 3: Compare the efficiency of phenotypic selection and marker-assisted selection, with the assumption of no dominance at the loci of interest, in improving population mean values

3.3.3 Comparison of change in population mean between phenotypic selection and marker-assisted selection

While sections 3.3.1 and 3.3.2 demonstrated the effects of phenotypic selection and marker-assisted selection, respectively, on populations with differing initial genotypic frequencies, in order to compare the efficiency of these two strategies, it was necessary to simulate the effects of both phenotypic and marker-assisted selection on the increase in μ on the same population. The initial genotypic frequencies in the population were therefore set at $A_1A_1B_1B_1 = 0.3$, $A_1A_2B_1B_2 = 0.4$ and $A_2A_2B_2B_2 = 0.3$.

The response to phenotypic selection was compared to the response to marker-assisted selection by comparing the increase in μ over five generations as a result of each selection strategy. For marker-assisted selection, the percentage of trait variance

accounted for by the QTL under selection, was set at 60%, 80% and 100%, abbreviated to MAS (60%), MAS (80%) and MAS (100%) respectively.

For phenotypic selection, individuals with a phenotypic value higher than 53 were selected, while for marker-assisted selection, individuals both heterozygous for the marker locus (A_1A_2) and individuals homozygous for the A_1 marker allele were selected. The population means were measured when the r was set at 0, 0.2 and 0.5 and the environmental heterogeneity was set at 0.3.

(a) Comparison of selection strategies

The efficiency of each selection strategies was compared by examining their individual effects on μ over time, when there was no recombination between the marker locus and the QTL (Table 3.5). The initial populations do not have identical mean values because each new generation is randomly created by the computer model. These differences, however, are not significant and would not affect the results in any way.

Table 3.5 Change in population mean over five generations as a result of phenotypic selection (P) and marker-assisted selection (MAS) when the QTL accounted for 60%, 80% and 100% of the trait variation and the recombination fraction (r) = 0.

Generation	P	MAS		
		60%	80%	100%
1	52.852	52.884	52.858	53.023
2	53.513	53.517	53.688	54.080
3	53.951	53.759	54.121	54.511
4	54.361	53.995	54.466	54.750
5	54.367	54.002	54.498	55.045

The largest increase in μ over five generations was achieved with MAS (100%), which increased μ by 2.022, from 5.023 to 55.045 (Table 3.5). When the QTL accounted for 80% of the trait variation, marker-assisted selection increased μ from 52.858 to 54.498 and was therefore more successful than phenotypic selection, which increased μ from 52.852 to 54.367. The smallest effect on μ was made by MAS (60%), increasing it by just over 1.

Figure3.12 demonstrates that MAS (100%) and MAS (80%) are both more effective than phenotypic selection, when there is no recombination occurring between the marker locus and the QTL under selection. At each generation, MAS (100%) and MAS (80%) both resulted in higher mean values than phenotypic selection did.

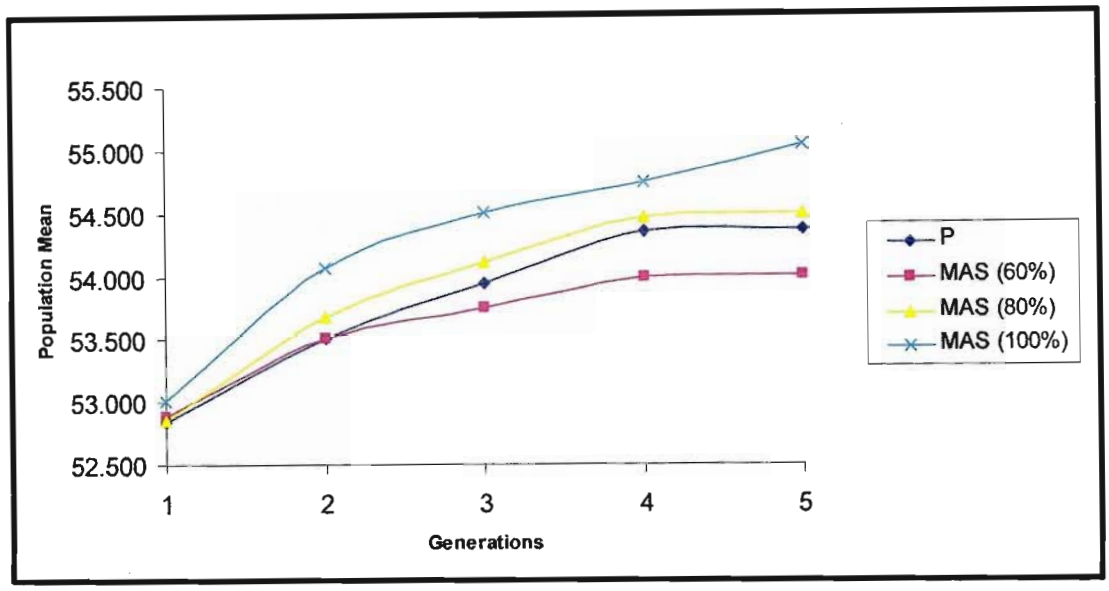


Figure 3.11 Change in population mean over five generations as a result of phenotypic selection (P) and marker-assisted selection (MAS) when the QTL accounted for 60%, 80% and 100% of the trait variation and the recombination fraction (r) = 0.

The most successful strategy to increase μ was demonstrably MAS (100%) (Figure 3.11). This strategy had a slightly higher μ in generation 1 than the other three strategies, but in generation 2, MAS (100%) had increased the original mean by a significantly higher amount and continued to show superior increases over all five generations. The second most successful strategy overall was MAS (80%) and phenotypic selection was the third most successful selection strategy overall, although the difference between the two was minimal. The least successful strategy in increasing μ over all five generations was MAS (60%).

When recombination was present between the marker locus and the QTL of interest, MAS (100%) was still the most effective selection strategy, but MAS (80%) was only more efficient than phenotypic selection after three generations (Table 3.6).

Table 3.6 Change in population mean over five generations as a result of phenotypic selection (P) and marker-assisted selection (MAS) when the QTL accounted for 60%, 80% and 100% of the trait variation and the recombination fraction (r) = 0.2.

Generation	P	MAS		
		60%	80%	100%
1	52.923	52.745	52.922	52.936
2	53.482	53.531	53.589	53.965
3	53.866	53.644	53.985	54.531
4	54.155	53.512	54.008	54.633
5	54.490	53.578	54.065	54.735

The phenotypic selection, MAS (100%) and MAS (80%) strategies were conducted on populations which had similar μ values of 52.936, 52.992 and 52.923 respectively (Table 3.6). When the QTL accounted for 100% of the trait variation, marker-assisted selection was again the most successful strategy because it increased μ to 54.735 after five generations. Phenotypic selection was the second most successful strategy, increasing the overall population mean to 54.490, while after five generations, MAS (80%) only increased μ to 54.065. MAS (60%) was the least successful strategy because μ increased by only 0.833 after five generations.

The superiority of MAS (100%) compared to the other selection strategies, as well as the efficiency of phenotypic selection, relative to MAS (80%), increasing after generation 3 is graphically depicted in Figure 3.12.

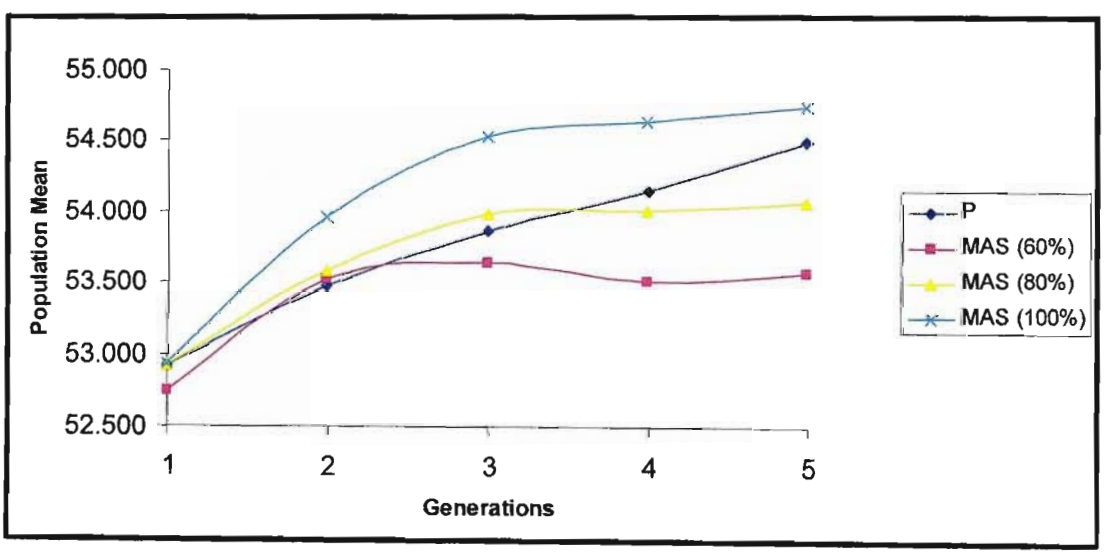


Figure 3.12 Change in population mean as a result of phenotypic selection (P) and marker assisted selection (MAS) when the QTL accounts for 60%, 80% and 100% of the trait variation and r = 0.2.

The MAS (100%) strategy was clearly the most effective and produced consistently higher values of μ every generation (Figure 3.12). Although phenotypic selection was the second most successful strategy overall, in generation 2 and 3, μ was higher as a result of MAS (80%). It was only from generation 4 that phenotypic selection began producing higher μ values than MAS (80%) because while phenotypic selection continued to increase μ, MAS (80%) had very little impact on μ after the third generation. After one generation, MAS (60%) had significantly increased μ from 52.745 to 53.531, but showed no further success in increasing μ. This is demonstrated by the increase from generation 1 to 2; thereafter μ remains fairly constant until generation 3, after which it declines (Figure 3.12).

When the marker locus and the QTL under selection were in linkage equilibrium (r=0.5), phenotypic selection was the most successful selection strategy in terms of increasing μ and marker-assisted selection was only able to produce a significant R when the QTL explained 100% of the trait variation (Table 3.7).

Table 3.7 Change in population mean as a result of phenotypic selection (P) and marker-assisted selection (MAS) when the QTL accounts for 60%, 80% and 100% of the trait variation and the recombination fraction was 0.5.

Generation	P	MAS		
		60%	80%	100%
1	52.983	52.761	52.908	52.981
2	53.625	52.906	53.466	53.511
3	53.829	52.776	53.646	53.723
4	54.017	52.457	53.545	53.871
5	54.221	52.384	53.388	53.910

The phenotypic selection and MAS (100%) selection strategies were conducted on populations with initial mean values of 52.983 and 52.981, respectively (Table 3.7). Although these initial values of μ were similar, phenotypic selection increased μ to 54.221 after five generations while MAS (100%) increased μ to only 53.910 over the same amount of time. Furthermore, MAS (80%) increased μ from 52.908 in generation 1, to 53.466 and 53.646 in generation 2 and 3 respectively (Table 3.7). However, over the next two generations, MAS (80%) decreased μ to 53.388. MAS (60%) had no significant effect on μ but did result in a slight decrease in mean value from 52.761 to 52.384. The above results are graphically depicted in Figure 3.13.

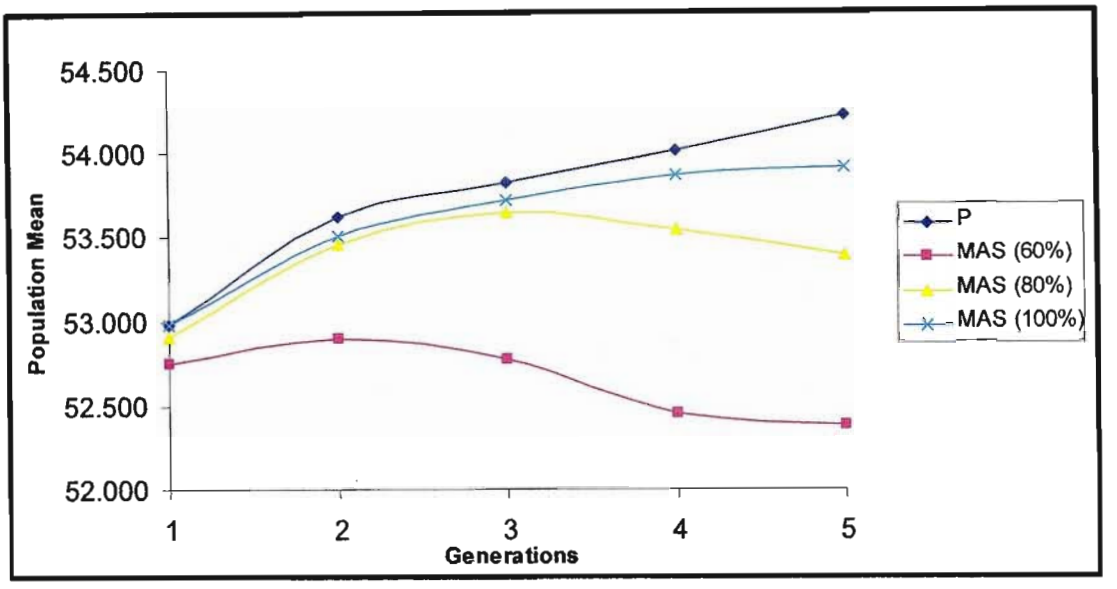


Figure 3.13 Change in population mean as a result of phenotypic selection (P) and marker-assisted selection (MAS) when the QTL accounts for 60%, 80% and 100% of the trait variation and the recombination fraction was 0.5.

When the recombination fraction between the marker locus and the locus of interest was 0.5, phenotypic selection was more successful in increasing μ than any of the marker-assisted selection strategies (Figure 3.13). The efficacy of MAS (80%) in increasing μ in generation 2 and 3 was as good as that achieved through phenotypic selection and MAS (100%), but after generation 3, the only effect MAS (80%) had was to slightly decrease μ. MAS (60%) increased μ in generation 2, after which it showed a negative effect on μ until generation 5.

(b) Concluding remarks

The results shown in Figures 3.11, 3.12, 3.13 and Tables 3.5, 3.6, 3.7, directly compare the efficiency of phenotypic and marker-assisted selection in increasing μ. Furthermore, these results focus on the effect of recombination and the percentage additive variance account for by the QTL, on the efficacy of marker-assisted selection.

Marker-assisted selection was more successful than phenotypic selection when the marker locus and the QTL were in linkage disequilibrium. When the recombination fraction between the marker locus and the QTL under selection was 0.2, MAS (80%) was only more efficient than phenotypic selection before recombination began to

erode the linkage disequilibrium present between the two loci, after which phenotypic selection produced high mean values.

The literature published by van der Werf (2000a) states that marker-assisted selection will only be beneficial if recombination does not break up the associations between marker loci and QTL. Moreover, Edwards and Page (1994) reported, using computer simulation, that marker-assisted selection produced rapid gains early in the selection process compared with phenotypic selection, however the rate of gain diminished greatly within three to five generations. They proposed that the distance between markers and QTL, and consequently the amount of recombination, was the factor that most limited gains from MAS. The results produced from the 'Phenotypic Selection' and 'Marker-Assisted Selection' models successfully showed that when there was no recombination between the marker locus and the QTL, marker-assisted selection was far more effective than when r was equal to 0.5 and the two loci were in linkage equilibrium (Figure 3.10 and 3.13).

The relative efficiency of marker-assisted selection over phenotypic selection was shown by the computer models developed in this investigation to be dependent on the amount of additive variance accounted for by the QTL. Through comparing the results obtained from the 'Phenotypic Selection' and 'Marker-Assisted Selection' models, it was found that the higher the percentage of additive variance explained by the QTL, the more efficient the marker-assisted selection strategy. Similarly, Meuwissen and Goddard (1996) simulated the additional genetic gains of marker-assisted selection over phenotypic selection and found that the larger the effect of the QTL on the trait under investigation, the higher the additional gain from marker-assisted selection. Thus, unless genetic markers account for most of the genetic variation of an economically important trait, selection should be based on a combination of marker and conventional phenotypic data (Williams, 2005).

The effectiveness of marker-assisted selection and phenotypic selection in enhancing economically important quantitative traits in sweet corn was also compared by Yousef and Juvik. (2001). In this case, selection efficiencies were evaluated on the basis of gains over one cycle and estimated evaluation costs. The average gains from marker-assisted selection and phenotypic selection across all the populations, calculated as

percent increase or decrease from the randomly selected controls, were 10.9% and 6.1%, respectively. Although, the 'Phenotypic Selection' and 'Marker-Assisted Selection' models developed in this investigation did not take the cost of the different breeding strategies into consideration, they demonstrated that marker-assisted selection was more efficient than phenotypic selection, provided the QTL under investigation accounted for a large proportion of the total variation.

The results simulated by both the 'Phenotypic Selection' and 'Marker-Assisted Selection' models created in this study were thus in agreement with the results of various other studies. Marker-assisted selection was determined to be more successful that phenotypic selection when linkage disequilibrium was present in the population and the QTL under selection captured a large amount of the trait variation. It may therefore be concluded that the deterministic power of these models provides the potential to enhance breeding programmes through enabling the most effective strategy to be established prior to its implementation.

3.4 DEMONSTRATION OF THE EFFECTS OF SELECTION WITH COMPLETE DOMINANCE

Aim 4: Simulate and compare the effects of phenotypic selection and marker-assisted selection, under the assumption of complete dominance at the loci under investigation.

The following results were acquired when complete dominance was simulated in the population ($B_1>B_2$). The consequences of complete dominance, is that the heterozygote has the same genotypic value as a homozygous dominant individual. Thus, heterozygous individuals (B_1B_2) had the same genotypic value (and therefore a similar phenotypic value) as individuals homozygous for the B_1 allele (B_1B_1).

The results displayed in Table 3.8 were obtained from the 'Selection with Dominance' model when the initial genotypic frequencies in the population were $A_1A_1B_1B_1 = 0.3$, $A_1A_2B_1B_2 = 0.4$ and $A_2A_2B_2B_2 = 0.3$. For phenotypic selection, individuals with a phenotypic value higher than 53 were selected and for marker-assisted selection,

individuals with at least one copy of the A_1 marker allele (A_1A_2 or A_1A_1 individuals) were selected. The amount of the trait variation accounted for by the QTL was set at 20%, 60%, 80% and 100%. The environmental heterogeneity was set at 0.3 and the recombination fraction was set at values of 0, 0.2 and 0.5.

Table 3.8 The response to phenotypic selection (P) and marker-assisted selection (MAS) when the recombination fraction (r) was equal to 0, 0.2 and 0.5. For marker-assisted selection, the percentage of trait variation accounted for by the QTL was set at 20%, 60%, 80% and 100%.

Response to selection at r = 0				
	MAS			P
20%	60%	80%	100%	
0.133	0.671	0.940	1.186	0.551
Response to selection at r = 0.2				
	MAS			P
20%	60%	80%	100%	
0.115	0.629	0.887	1.150	0.538
Response to selection at r = 0.5				
	MAS			P
20%	60%	80%	100%	
0.105	0.598	0.806	1.101	0.513

The highest R achieved by all the selection strategies, was when there was no recombination present between the marker locus and the locus of interest (r = 0) (Table 3.8). In contrast, when r was 0.5 and there was no linkage between the two loci, the selection strategies all produced the lowest R. The amount of recombination between the loci, however, had the smallest effect on phenotypic selection, since the response to phenotypic selection was 0.551, 0.538 and 0.513 when r was 0, 0.2 and 0.5 respectively (Table 3.8).

When r = 0, MAS (100%), MAS (80%) and MAS (60%) all had a higher R than phenotypic selection (Table 3.8). MAS (100%) had a response of 1.186, which was just over double the response recorded with phenotypic selection (0.551). The response to selection achieved with MAS (20%) was only 1.33, which was far lower than the response to phenotypic selection.

When r = 0.2, the response to the different selection strategies was consistently less than the results obtained when r = 0. However, the order of effectiveness remained the same. It is evident that MAS (100%) had the highest R (1.150), followed by MAS

(80%) with a R value of 0.887 and MAS (60%) with a R of 0.629 (Table 3.8). The response to phenotypic selection was 0.538. MAS (20%) showed the lowest R of 0.115.

When r was 0.5, the order of success of the different selection strategies remained the same and therefore MAS (100%), MAS (80%) and MAS (60%) all achieved a greater R than phenotypic selection (Table 3.8). MAS (20%) again showed the lowest R.

Marker-assisted selection, when compared to phenotypic selection, is most effective when there is no recombination between the marker locus and the QTL of interest and when the QTL accounts for a large percentage (more than 60%) of the trait variation. When the QTL under selection exhibits complete dominance, marker-assisted selection is particularly useful relative to phenotypic selection, even when r is 0.5 (Figure 3.14).

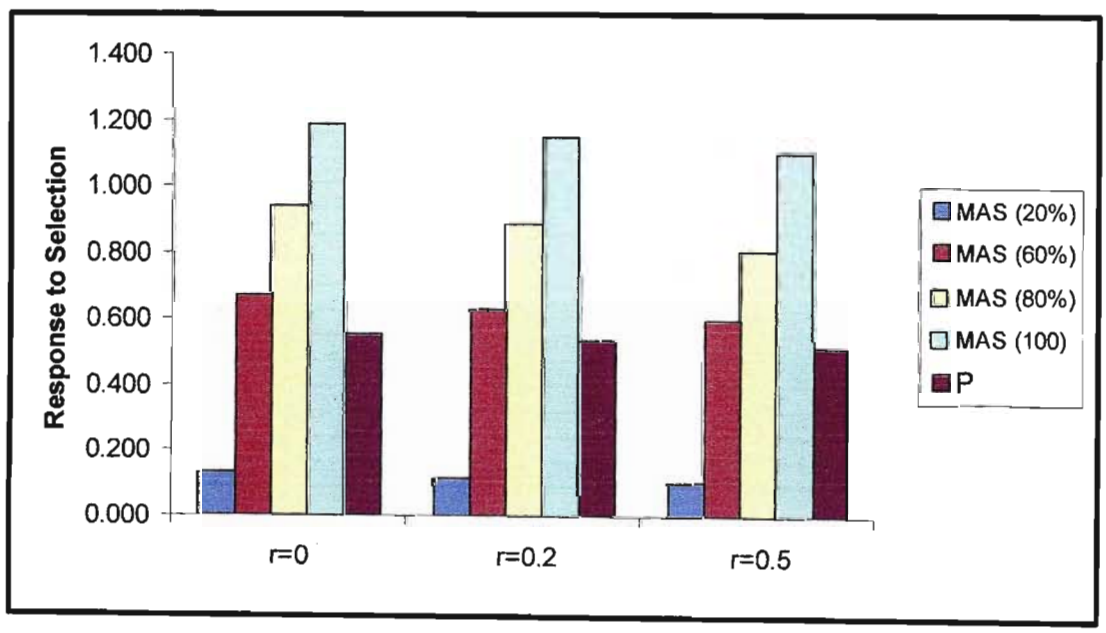


Figure 3.14 The response to phenotypic selection (P) and marker-assisted selection (MAS) when the recombination fraction (r) was equal to 0, 0.2 and 0.5. For marker-assisted selection, the percentage of trait variation accounted for by the QTL was set at 20%, 60%, 80% and 100%.

The response to selection for MAS (100%), MAS (80%) and MAS (60%) was higher than the response to phenotypic selection no matter what the recombination fraction

(Figure 3.14). MAS (20%) consistently produced the lowest R. Furthermore, R for all the different strategies was highest when r = 0 and at its lowest when r = 0.5.

In cases where the QTL under selection exhibited complete dominance, the response to MAS (60%) was consistently superior to the response to phenotypic selection. Furthermore, the less recombination that existed between the marker locus and the locus under selection, the more successful these marker-assisted selection strategies were. Although the response to phenotypic selection did decrease as the recombination fraction increased, the effect was minimal in comparison to the effect of recombination on the response to marker-assisted selection.

Williams (2005) suggests that marker-assisted selection is particularly beneficial, when compared to conventional selection methods, when the heritability of a trait is low. Dominance at a locus causes an increase in the non-additive genetic variance component in a population, which is responsible for decreasing the heritability of the trait controlled by that locus (Falconer, 1989). The response values presented in Table 3.8 and illustrated in Figure 3.14, for traits which exhibit complete dominance and therefore have low heritabilities, indicate that the response to marker-assisted selection was higher than the response to phenotypic selection, even when recombination existed between the marker locus and QTL and the QTL accounted for as little as 60% of the additive trait variance. Thus, when complete dominance existed at the QTL under selection, marker-assisted selection was especially advantageous when compared to phenotypic selection.

3.5 APPLICATION OF MODELS TO HOLSTEIN CATTLE POPULATIONS

Aim 5: Mimic the effects of linkage disequilibrium marker-assisted selection, utilizing both a direct and an indirect marker, in two different Holstein cattle populations.

3.5.1 Application of model using a direct marker

The effects of marker-assisted selection for a marker directly linked to an increase in milk yield in a Holstein cattle population was studied by Buchanan *et al.* (2003). The molecular marker identified in this study is considered a direct marker, whereby the marker allele, designated T, was directly linked to an allele for increased milk production, designated B_1, and the other marker allele, C, was directly linked to the another allele, B_2. Furthermore, Buchanan *et al.* (2003) established the allele frequencies at the marker locus, T=0.46 and C=0.54, the initial population mean (67lbs/day; 30kg/day) and the value of the alleles at the B locus for increased milk yield.

The above-mentioned data was incorporated into the 'Direct Selection' model developed in this investigation. The recombination fraction was set at 0 (direct marker) and individuals heterozygous for the marker locus, TC, as well as individuals homozygous for the T marker allele, TT, were selected in an attempt to simulate direct selection for the favourable B_1 allele.

(a) Response to selection

The response to marker-assisted selection was calculated over five generations in order to assess whether or not marker-assisted selection would be an effective breeding strategy with regards to increasing the mean milk yield in the population described by Buchanan *et al.* (2003). The response to selection over five generations as a result of selection based for the T marker allele was significant in the first three generations, after which it remained low (Figure 3.15).

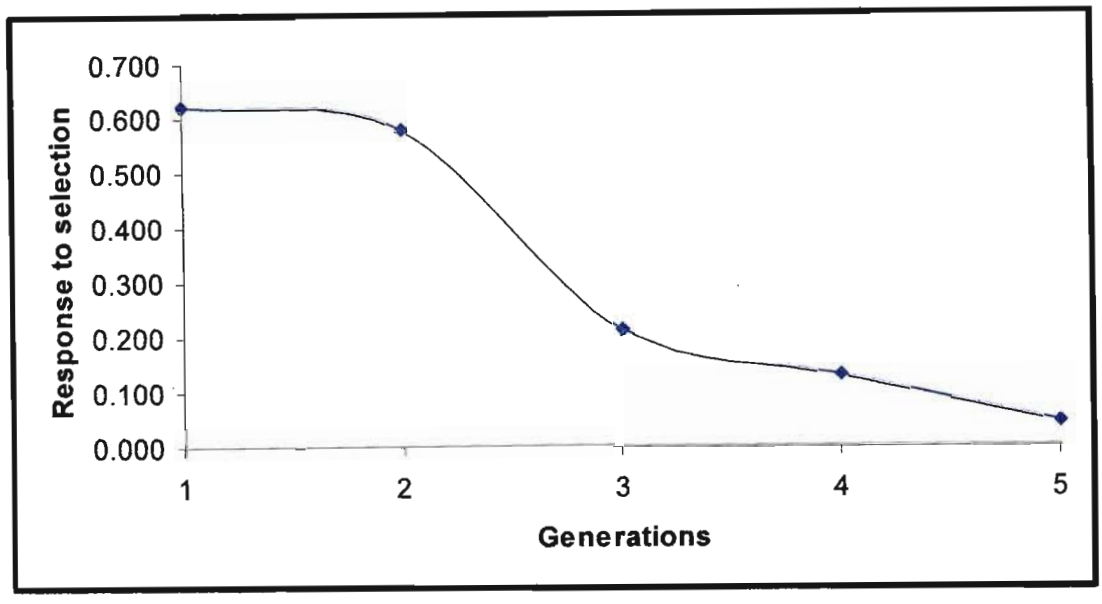


Figure 3.15 Response to marker-assisted selection over five generations using a direct marker.

After one generation of marker-assisted selection, R was just over 0.6 and the response decreased only marginally after the second generation. After generation 3, the response decreased to approximately 0.2, then again to 0.12 after generation 4. After the fifth generation of marker-assisted selection, the response was close to 0.

(b) Change in frequency of T and effect on population mean

As the response to selection illustrated in Figure 3.15 was a consequence of selection for the T marker allele, the change in frequency of T was therefore also examined. Theoretically, the increase in the frequency of the T marker allele, due to marker-assisted selection, should result in an increase in the frequency of the directly linked B_1 allele. Since the B_1 allele has been associated with an increase in milk yield, an increase in the frequency of the T allele should consequently result in a higher mean milk yield in the population. Five generations of marker-assisted selection for the T marker allele resulted in an increase in the frequency of T from 0.4 to near fixation (Figure 3.16).

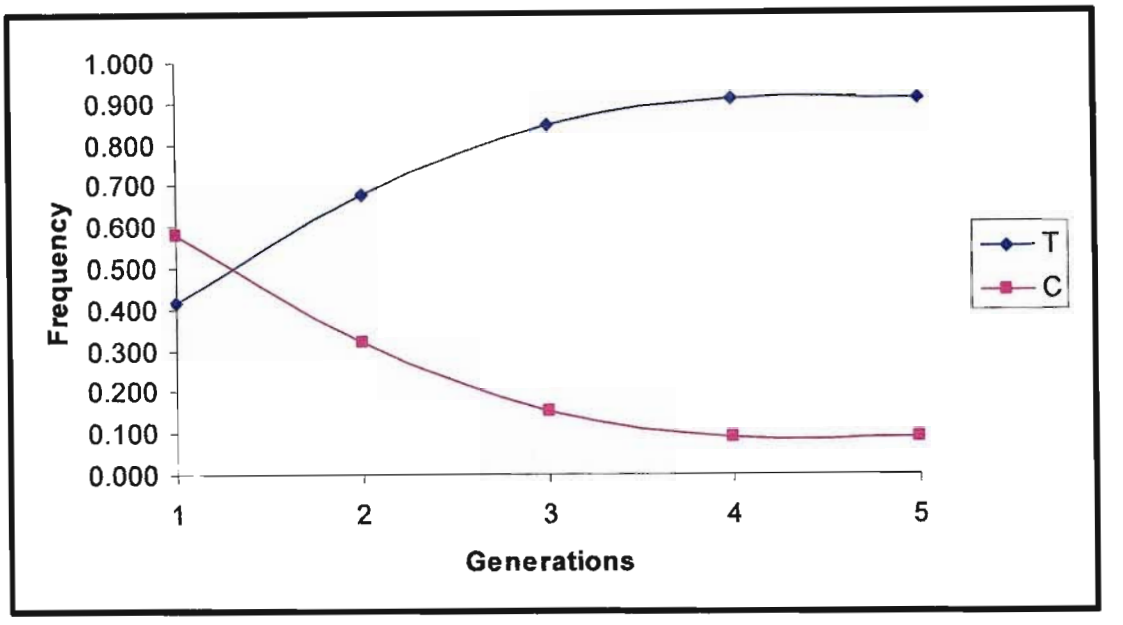


Figure 3.16 Change in allele frequencies at the A marker locus over five generations of marker-assisted selection.

In the initial population, the frequency of the T marker allele was lower than the frequency of the C marker allele. However, after just one generation of marker-assisted selection, the frequency of the T and C alleles was approximately 0.7 and 0.3 respectively. In generation 5, the frequency of the T allele had increased to just over 0.9.

The direct correlation between the frequency of the T allele and the population mean was demonstrated by the computer model (Figure 3.17) such that as the frequency of T increased, so did μ.

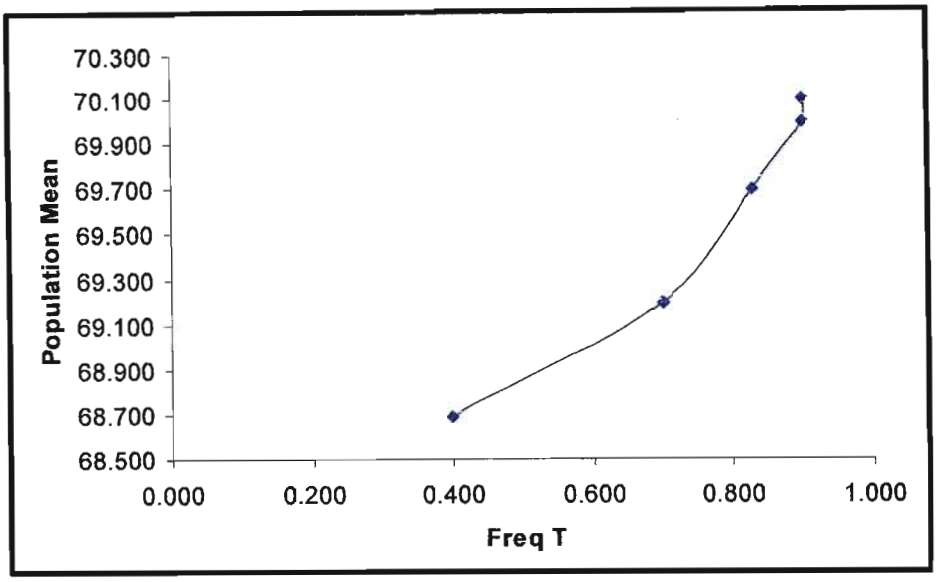


Figure 3.17 Relationship between the change in the frequency of the T marker allele and the change in population mean.

Figure 3.17 shows the direct relationship between the frequency of the T allele in the population and the population mean. As the frequency of the T allele increased in the population, so the population mean increased. When the T allele had a frequency of about 0.4, the population mean had its lowest value of about 68.7. When the frequency of the T allele was 0.9, the population mean was just higher 70.

The significant increase in milk yield (lbs/day) in the population over five generations of direct selection for the T marker allele is illustrated in Figure 3.18.

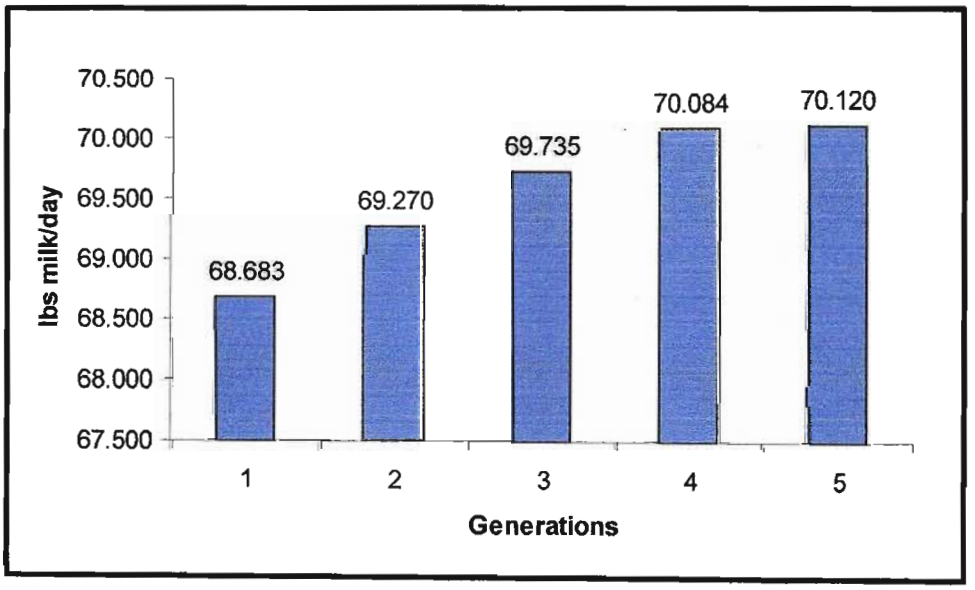


Figure 3.18 Change in mean milk production (lbs/day) over five generations of Marker-assisted selection.

Figure 3.18 shows the steady increase in mean milk yield in the population from 68.683 lbs/day (31.220kg/day) to 70.120lbs/day (31.873kg/day) over five generations of marker-assisted selection. Although μ increased every generation, the greatest increase was seen after the first two generations of marker-assisted selection. The increase in milk yield/day in generation 5 was minimal.

These results demonstrate the direct association between an increase in milk yield and the increase in the frequency of the T marker allele. From generation 1 to generation 3, when the frequency T allele increased the most (Figure 3.16), the population mean value also showed its largest increase (Figure 3.18). This direct association was responsible for the high response to marker assisted selection (Figure 3.15). In contrast when the T allele increased gradually in generation 4 and 5, the population mean also increased by a minimal amount. The same result was illustrated in Figure 3.17.

3.5.2 Application of model using an indirect marker

The effects of marker-assisted selection based on linkage disequilibrium between the marker locus and the QTL of interest was studied by Weller *et al.* (2003), using data obtained from a Holstein cattle population. Weller *et al.* (2003) found that the marker allele, designated A_1, was linked to the QTL responsible for an increase in milk fat percent, designated B_1; and the other marker allele, designated A_2, was linked to the QTL with no added effect on milk fat percent, designated B_2. Furthermore, they found that the recombination fraction between the marker locus and the QTL was 0.0125 and the QTL accounted for 8% of the total additive variance of milk fat percent in cattle.

The above-mentioned data was incorporated into the 'Indirect Selection' model developed in this study for the purpose of determining the effects of marker-assisted selection on the Holstein cattle population described by Weller *et al.* (2003). Individuals heterozygous for the marker locus (A_1A_2) and individuals homozygous for the A_1 allele were selected.

(a) Response to selection

The response to selection was calculated over five generations in order to analyze the success of marker-assisted selection based on linkage disequilibrium in increasing the mean milk fat percentage in the population (Figure 3.19).

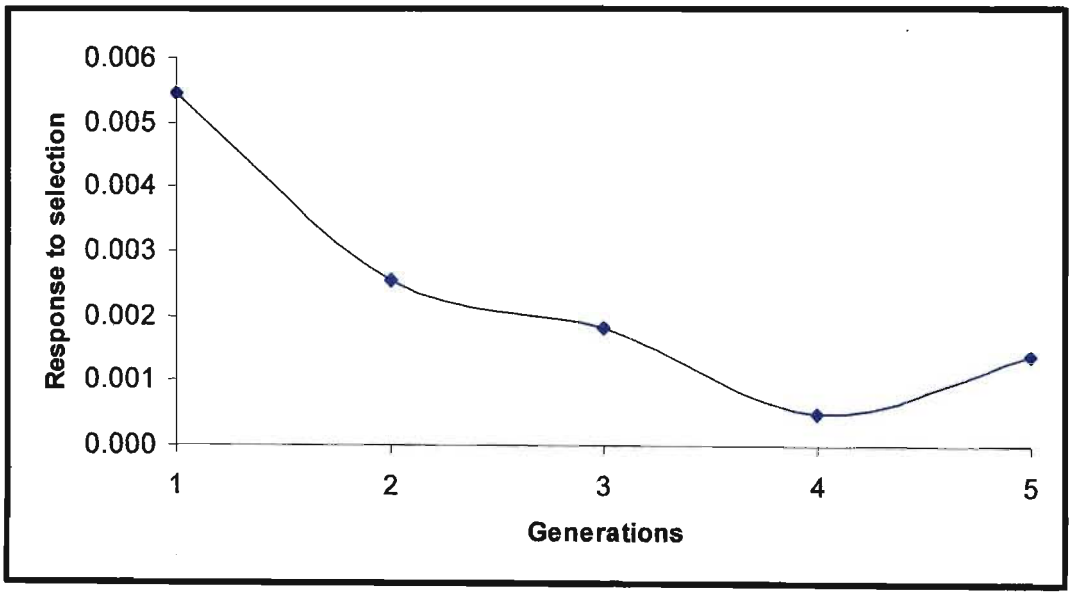


Figure 3.19 Response to marker-assisted selection over five generations using an indirect marker

The greatest response to selection was after the first generation of marker-assisted selection (approximately 0.0055). The response dropped to approximately 0.003 after the second generation and to below 0.002 after generations three and four. After the fifth generation R increased slightly to around 0.0015.

(b) Change in frequency of B_1 and the effect on population mean

Since the B_1 allele was synonymous with the QTL responsible for an increase in milk fat percent, an increase in the frequency of B_1 over time, resulted in an increase in the overall mean milk fat percent in the population (Figure 3.20).

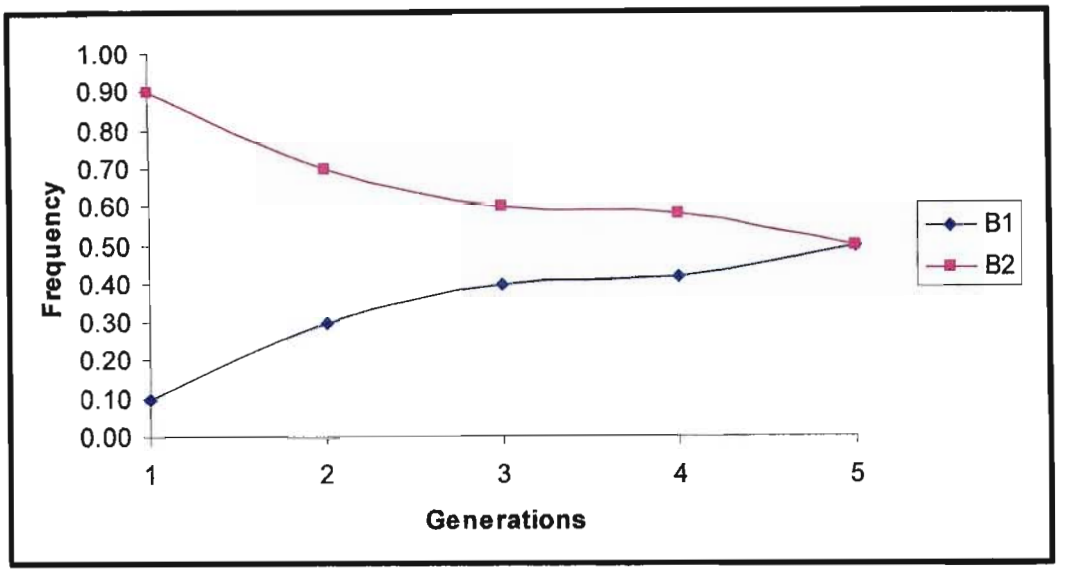


Figure 3.20 Change in allele frequencies at the B locus controlling milk fat percent over five generations of marker-assisted selection.

The frequency of the B_1 allele had a low initial frequency of about 0.1 (Figure 3.20). Indirect selection for this allele, increased the frequency of B_1 to approximately 0.3 and then to 0.4 after the first two generations. After five generations of marker-assisted selection, the frequency of the B_1 allele had increased to almost 0.5.

The increase in the frequency of B_1 (Figure 3.20) directly affected the values of μ. The increase in mean milk fat percent over five generations is shown in Figure 3.21.

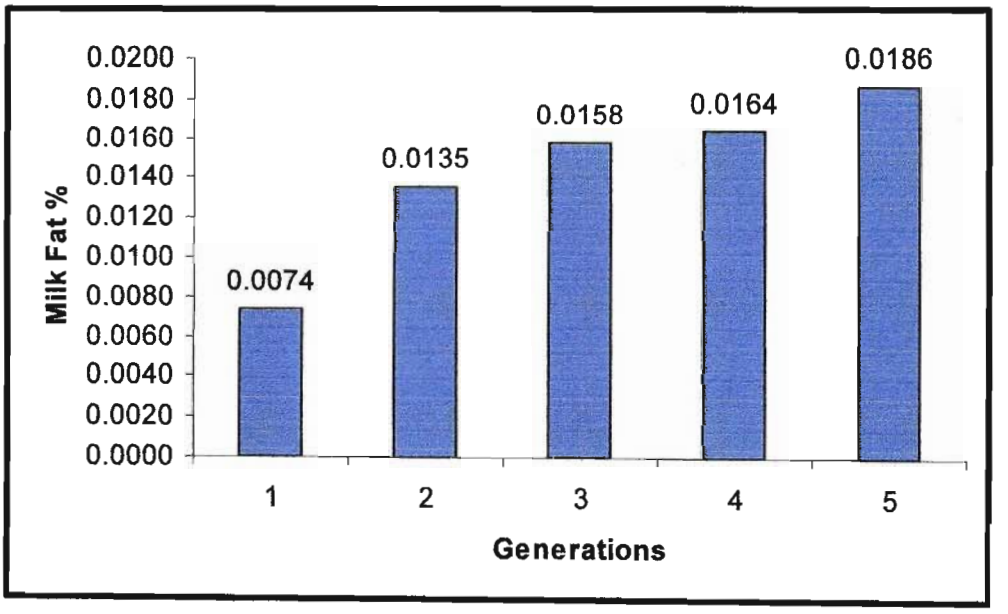


Figure 3.21 Change in Mean Milk Fat % over five generations of Marker-Assisted Selection

The largest increase in mean milk fat percent in the population, from 0.0074 to 0.0135, occurred after the first generation of marker-assisted selection (Figure 3.21). In the third generation, the mean increased to 0.0158 and then increased only slightly to 0.0164 in the fourth generation of marker-assisted selection. There was a more significant increase in population mean to 0.0186 in generation 5.

Marker-assisted selection for the A_1 marker allele effectively increased the frequency of the B_1 allele and consequently improved the overall population mean. In five generations, marker-assisted selection successfully increased the mean milk fat percentage by 0.0112%. According to the results simulated by the 'Indirect Selection' model, marker-assisted selection utilizing linkage disequilibrium would be a viable breeding option in this particular Holstein cattle population even though the QTL under selection only accounted for 8% of the total trait variation.

3.5.3 Concluding Remarks

The results of the 'Direct Selection' and 'Indirect Selection' models, which mimicked real Holstein cattle populations, showed that marker-assisted selection was effective in increasing the mean milk yield and the mean milk fat percent using direct and indirect marker data respectively. The responses to marker-assisted selection using a direct and indirect marker were illustrated Figure 3.15 and Figure 3.19 respectively, and both showed that although the response was greatest after the first generation, there was a significant response to marker-assisted selection throughout the first five generations.

Since these simulation models are merely real examples of the principles that were demonstrated by the 'Marker-Assisted Selection' model, the same principles involving the decrease in R as a result of decreased $V_{A,}$ i and $\boldsymbol{D}$, also apply.

The response to marker-assisted selection in both the populations was illustrated in the increase in mean milk production over five generations (Figure 3.18) and the increase in mean milk fat% over five generations (Figure 3.21). In the short term, marker-assisted selection would therefore be a viable selection strategy with regards

to increasing mean milk yield with the use of a direct marker and increasing mean milk fat percent with the using of an indirect marker in the short term.

3.5.4 Benefits of Marker-Assisted Selection

In the Holstein cattle population, cows with the lysine marker allele have a considerably higher proportion of milk fat compared to those with the alanine marker allele (Weller *et al.*, 2003). Subsequent studies, however, have shown that while the lysine allele is associated with an increased milk fat content, it is also related to decreased protein percentage and thus breeders could not only change the milk fat content but also alter, to some extent, the protein level in the milk (Williams, 2005). While this is a negative genetic correlation, and breeders aim to increase both fat percent and protein content in milk, knowledge of this negative correlation could allow breeders to identify and select for individuals with the most economically viable ratio of fat and protein content in milk.

Traits that are typically suitable for marker-assisted selection are traits that are expensive to measure and difficult to record (van der Werf, 2000a). Marker-assisted selection is especially useful for traits involved with milk yield and composition in cattle because it allows for the selection of male individuals as well as female individuals carrying beneficial genes with regards to milk production. In contrast, traditional selection would have relied on large progeny tests and pedigree data in order to estimate the breeding value of males (since no information on the individual can be provided) and this is both time consuming and expensive (Dekkers and Hospital, 2002). The benefits of marker-assisted selection over phenotypic selection in cattle populations are therefore clear, not only in terms of the significant response to selection shown by the simulated populations, but also in terms of overcoming some of the difficulties and costs involved with traditional selection methods.

The 'Marker-Assisted Selection' model developed in this investigation was thus able to successfully mimic the application of a marker-assisted selection strategy in Holstein cattle populations. The simulated results showed that marker-assisted selection could potentially improve milk production traits in cattle over five generations, which would have been difficult to do via conventional selection.

CHAPTER 4
SUMMARY OF RESEARCH FINDINGS AND IMPLICATIONS OF THIS INVESTIGATION

4.1 INTRODUCTION

The ultimate goal of this investigation was to study the efficiency of marker-assisted selection, utilizing linkage disequilibrium, as a breeding strategy to improve genetic gains in a population. The effects of linkage disequilibrium thus needed to be examined so that its applications to marker-assisted selection could be fully understood and exploited to ensure maximum response to selection. Furthermore, the effects of phenotypic selection were analyzed and used as a basis from which to compare the success of marker-assisted selection as a breeding strategy. This comparison was made with the assumption of both additive gene action and complete dominance at the QTL under selection so as to ascertain the benefits and pitfalls of each selection strategy under these two different circumstances. Finally, the applications of marker-assisted selection to real cattle populations were investigated in order to gauge the additional genetic gains that could be established in existing breeding programmes through the implementation of marker-assisted selection. This process is graphically depicted in Figure 4.1.

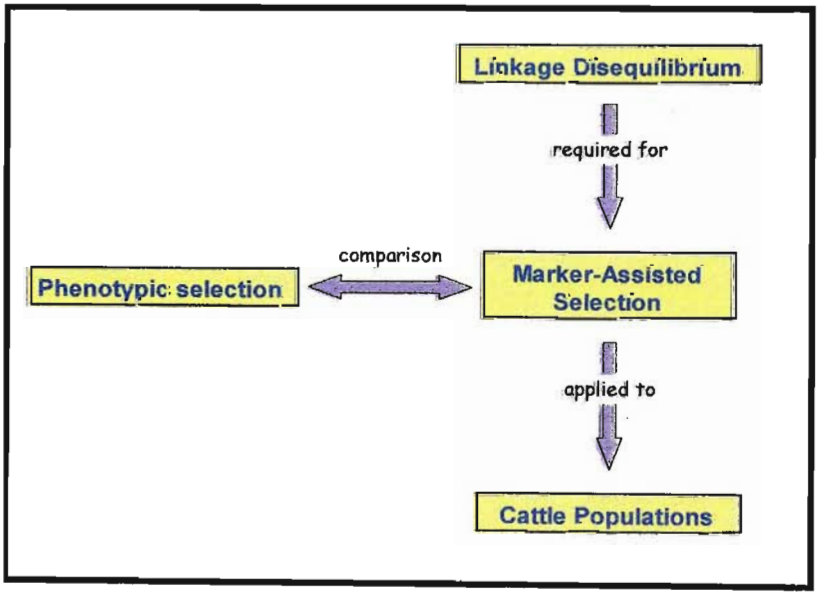


Figure 4.1 The steps taken to investigate the efficiency of marker-assisted selection utilizing linkage disequilibrium as a breeding strategy.

Six computer models were developed in this study with the purpose of simulating the effects of linkage disequilibrium, the effects of phenotypic selection and marker-assisted selection with the assumption of both additive gene action and complete dominance at the QTL of interest and finally the application of a marker-assisted selection programme to cattle populations using both a direct and an indirect marker.

It was determined that the simulated results were consistent with the findings of various other studies involving linkage disequilibrium, phenotypic selection and marker-assisted selection and therefore the validity of the computer models created in this investigation was confirmed.

4.2 GENERAL DISCUSSION OF RESEARCH FINDINGS

The first aim of this investigation was to simulate the effects of linkage disequilibrium between two loci in a randomly mating population using the 'Random Mating' model. The results showed conclusively that linkage disequilibrium was responsible for the preferential association of alleles at two loci in a population as opposed to the random association of alleles at loci in linkage equilibrium. Furthermore the presence of the double heterozygote in a population, as well as recombination between two loci, was found to be responsible for the breakdown of linkage disequilibrium between two loci. The latter result, in particular, has significant implications for the utilization of linkage disequilibrium in marker-assisted selection programmes.

Linkage disequilibrium marker-assisted selection often relies on the crossing of two inbred lines for the formation of linkage disequilibrium between marker loci and loci containing economically important genes in the F_1 generation (Montaldo and Meza-Herrera, 1998). This concept is illustrated in Figure 4.2.

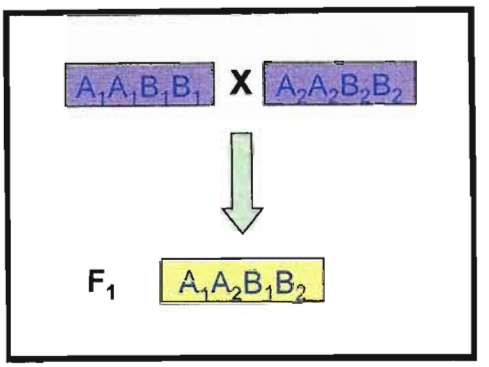


Figure 4.2 Crossing of two inbred lines in order to establish linkage disequilibrium in the F_1 generation that can be used in marker-assisted selection.

While recombination will have no effect on the amount of linkage disequilibrium present in the parental generation, because the inbred lines are homozygous, it will erode the linkage disequilibrium present in the heterozygous F_1 generation (Figure 4.2). Although this strategy is successful in creating linkage disequilibrium in the F_1 generation, the success will be short lived as recombination in subsequent generation starts to breakdown the linkage in the F_1. Care must therefore be taken to re-evaluate the amount of linkage disequilibrium present in every generation before marker-assisted selection is implemented.

Secondly, this investigation aimed to simulate the effects of both phenotypic selection and marker-assisted selection with the assumption of additive gene action at the loci of interest. This was achieved through the development of the 'Phenotypic Selection' and 'Marker-Assisted Selection' models, which allowed the effect of each selection strategy on R and on various population parameters such as V_A, i and h^2, to be examined.

In order to compare the efficiency of marker-assisted selection and phenotypic selection, each selection strategy was implemented upon the same simulated population and their effects were examined and compared. It was found that the advantages of marker-assisted selection over phenotypic selection were dependent on the amount of linkage disequilibrium present between the marker locus and the QTL and the amount of trait variance accounted for by the QTL under selection.

Marker-assisted selection was consistently more effective than phenotypic selection in improving the population mean, when recombination had not eroded the linkage disequilibrium between the marker locus and the QTL. Over time, as this association was eroded by recombination, phenotypic selection was found to be more beneficial. Furthermore, it was found that the more trait variance explained by the QTL, the higher the response to marker-assisted selection. Marker-assisted selection, however, rapidly decreased the variability of alleles at the QTL over time and thus, according to Stella *et al.* (2002), the advantages of marker-assisted selection programmes can only be sustained in the long term if new QTL are continually discovered and selected. The disadvantage of this is that the effect of the newly discovered QTL may be smaller than the original QTL, because the QTL with large effects are the most likely to be found early or moved toward fixation via the forces of conventional selection (Stella *et al.*, 2002). As the percentage of trait variation explained by the QTL decreases, so will the response to marker-assisted selection and other breeding strategies may thus become more effective in increasing genetic gains in the long term.

The fourth aim of this investigation was to simulate and compare the effects of both phenotypic selection and marker-assisted selection with the assumption of complete dominance at the QTL under selection. The 'Selection with Dominance' model was created with the purpose of comparing the response to phenotypic selection and marker-assisted selection, after one generation, when the QTL exhibited complete dominance. Marker-assisted selection proved to be consistently more effective than phenotypic selection in increasing the overall population mean when complete dominance was assumed at the QTL under selection. In the absence of dominance, marker-assisted selection was only more effective than phenotypic selection when the QTL captured a large portion of the additive variance. In contrast, in the presence of complete dominance, even marker-assisted selection for QTL explaining only a small portion of the additive trait variance proved to be more beneficial than phenotypic selection.

The results simulated by the 'Selection with Dominance' model indicated that marker-assisted selection was most appropriate when traits with low heritabilities that were difficult to measure, were under selection. This is because the accuracy of evaluation under marker-assisted selection schemes is higher than under phenotypic selection

schemes (Abdel-Azim and Freeman, 2002). Despite the fact that currently, obtaining information on molecular markers is more expensive than obtaining measurements on phenotypic characteristics, the expected return from marker-assisted selection programmes for traits with low heritabilities, is substantial (Xie and Xu, 1998). Marker-assisted selection may therefore prove to be more economically viable than phenotypic selection for traits that exhibit dominance, provided that the QTL identified by the marker, accounts for a reasonable percentage of the trait variance.

The benefits of marker-assisted selection can also be extended to traits which can only be measured in one sex, such as milk yield, because while phenotypic selection has to rely on progeny data and pedigree information to assess the breeding value of individuals, marker-assisted selection is able to directly evaluate the individual on the basis of their DNA (Meuwissen and Goddard, 1996). An example of this is the improvement of milk production traits in cattle.

The final aim of this study was to apply the 'Marker-Assisted Selection' model to two different Holstein cattle populations in order to mimic the effects of marker-assisted selection utilizing both a direct and an indirect marker.

While linkage disequilibrium has been suggested for the high-resolution mapping of economically important genes, serious reservations have been expressed about the power of this method with respect to the required marker density (Meuwissen *et al.*, 2002). Livestock, however, represent a unique opportunity to follow genomic changes through the various levels of selection because of their large progeny groups (Gomez-Raya *et al.*, 2002). Cattle, in particular, have low effective population sizes because often sperm from one sire is used to artificially inseminate many cows, and this produces extensive genome-wide linkage disequilibrium and facilitates fine mapping with relatively low density marker maps (Meuwissen *et al.*, 2002). Data from Holstein cattle populations was therefore used to study the effects of marker-assisted selection simulated by both the 'Direct Selection' and 'Indirect Selection' models.

The results simulated by the 'Direct Selection' model showed that an increase in milk yield could be achieved through the implementation of a marker-assisted selection programme in a Holstein cattle population where the molecular marker was directly

linked to the QTL controlling milk yield. Furthermore, the 'Indirect Selection' model simulated the response to marker-assisted selection utilizing linkage disequilibrium for an increase in milk fat percent. The results simulated by the 'Indirect Selection' model established that marker-assisted selection for the QTL controlling milk fat percent would increase the mean milk fat percent in the population over five generations, even though the QTL only captured 8% of the trait variation.

The six computer models developed in this investigation were therefore successful in terms of achieving their aims. As a result, these models could be used in breeding programmes to assess the advantages of different selection strategies prior to their implementation and in so doing maximise the potential genetic gains.

4.3 APPLICATION OF THE COMPUTER MODELS DEVELOPED IN THIS INVESTIGATION

Since marker-assisted selection is still in its infancy, it is important that breeders weigh up the benefits of increased genetic gain with the financial implications of such a breeding scheme (Brascamp *et al.*, 1993). Computer simulations have provided powerful tools for analysing the design and efficiency of marker-assisted selection programmes and have enabled breeders to realize genetic gains with greater speed and precision (Babu *et al.*, 2004). The computer models developed in this investigation allow for the prediction of response to both phenotypic selection and marker-assisted selection utilizing linkage disequilibrium. The application of these models as a research tool was demonstrated by the incorporation of real data from Holstein cattle populations into the 'Marker-Assisted Selection' computer model. The results indicated that marker-assisted selection would be successful in increasing mean milk yield in the one population based on a direct marker and in increasing mean milk fat percent in another population. The costs of implementing marker-assisted selection in these Holstein cattle populations could therefore be weighed up against the genetic gains predicted by the simulation models to assess whether marker-assisted selection would be worthwhile financially overall.

The computer models created in this study will not only have applications in the agricultural sector as mentioned above, but may also be used as an educational tool.

Computer assisted learning is becoming increasingly prominent in education because of its potential to change the face of teaching and learning (Newby *et al.*, 2000). While there are numerous driving forces behind such innovations, the desire to increase learning outcomes and teaching efficiency has generally been regarded as the major contributing factor (Duit and Confrey, 1996). Berger *et al.* (1994) found that computer based education commonly provided some kind of benefit over other instructional methods and there was a reduction in the amount of learning time required by the students. Furthermore they found that computer based learning improved attitudes towards learning and increased motivation and was particularly effective amongst learners with lower ability.

The subject of genetics, in particular, involves concepts which are complex and abstract, making it difficult for students to understand and interpret them (Hancock, 2006). However, even though the cognitive load required for understanding these concepts may be high, it has been documented that computers potentially allow learners to function at levels that their cognitive system would otherwise not have allowed them to reach, and can thus redefine and enhance students' performance (Berger *et al.*, 1994). For abstract genetics concepts such as linkage disequilibrium and marker-assisted selection, it is therefore clear that the computer models in this investigation may help to achieve learning outcomes that typical approaches to learning may not.

Computers have been found to be particularly useful in relating graphical representations to the real-world situation that they characterize and this can enable learners to rectify their inaccurate or incomplete mental models through interacting with the formal principles pertaining to the system (Duit and Confrey, 1996). Important genetics principles such as linkage disequilibrium, recombination, phenotypic and marker-assisted selection, dominance and environmental effects and heritability estimates, to name a few, are graphically represented in these computer models, and may therefore enable students to grasp these concepts more effectively than they would have in a traditional learning environment.

In order for computer-based simulations to be a successful mode of instruction, they should be designed in such a way that they are able to fulfil the following

requirements laid out by delMas (1997). Firstly, the curriculum should integrate physical activities into computer simulations or tutorials, so as to allow students to relate their newly acquired knowledge to real world phenomena. The computer models in this investigation fulfilled this requirement by showing real life applications of marker-assisted selection to Holstein cattle populations and allowing students to manipulate certain population parameters and observe the effects.

Secondly, computer-based instruction should encourage pictorial and conceptual representations. The graphical representation provided by the computer models in this study, included the comparison in response to the different selection strategies under changing environmental conditions. Furthermore conceptually difficult concepts such as dominance effects and recombination were graphically depicted using different types of graphs and tables.

In accordance with the above-mentioned guidelines, the computer models developed in this investigation could provide an adaptive learning environment that would encourage students to construct and consequently grasp important concepts, within the particular domain of linkage disequilibrium and marker-assisted selection. Clearly these models will not only be applicable to research, but also to the improvement of learning within the realm of Genetics.

4.4 PROPOSALS FOR FUTURE RESEARCH

While the computer models developed in this investigation successfully simulated results that were accurate in terms of other published studies and literature, a number of limiting assumptions were made in the development of these models. Although assumptions form an integral part of model development, they do serve to decrease the accuracy of the predictions made by these models and therefore the fewer assumptions made in the development of a model, the more accurate the model in mimicking real life situations. In order to expand the applications of the computer models created in this study, it would therefore be necessary to address the following assumptions and in so doing increase the accuracy of the simulated results.

Firstly, the models developed in this investigation assumed that mating between males and females occurred at a ratio of 1:1. This is not always an accurate assumption, for livestock populations in particular, because one male is often used to artificially inseminate several females (Meuwissen *et al.*, 2002). As a result, the effective population size will be lowered and the genetic variability in a population will be decreased (Falconer, 1989). Although the response to selection and the effect of selection on genetic variation in a population, simulated in this study, was in keeping with the literature, it is clear that by taking the effective population size into account, the accuracy of these results would be increased.

Secondly, this investigation focussed on utilizing the linkage disequilibrium present between a marker locus and a single QTL. As more economically important QTL are detected throughout the genome, however, genetic gains in a population could be significantly increased through the use of marker-assisted selection for several of these QTL simultaneously (van der Werf, 2000a). Deterministic computer models, such as these, will therefore have to be expanded in order to incorporate more than one locus affecting the trait of interest.

The comparison of phenotypic selection and marker-assisted selection was made with the assumption of both additive gene action and complete dominance at the locus of interest. In reality, however, varying degrees of dominance exist at loci controlling traits of economic importance and thus the 'Selection with Dominance' model could be adapted and expanded in order to take these varying degrees of dominance into account. In so doing, the applications of this model could be extended to any population, regardless of the gene action at the loci of interest.

Finally, the models created in this study assumed that there was no genotype-environment interaction at the loci under selection. Genotype-environment interactions are becoming increasingly evident in traits of economic importance and therefore emphasis is being placed on the development of models which can take this interaction into account (Chapman *et al.*, 2003). If the computer models created for the purpose of this study could be modified in order to take this interaction into account, the results would be far more accurate and its validity as a deterministic model would be increased.

Advances in molecular technology and genome programmes are expected to create a wealth of information that can be exploited for the genetic improvement of plants and animals (Soller and Beckman, 1988). The development of deterministic simulation models that accurately predict the response to selection, will therefore be necessary to optimize selection schemes (Montaldo and Meza-Herrera, 1998). By addressing the afore-mentioned assumptions, the computer models developed in this investigation could be expanded in order to meet the future demands for a deterministic model that is able to accurately predict the outcome of different selection strategies.

REFERENCES

1. Abdel-Azim, G. and A.E. Freeman. 2002. Superiority of QTL-assisted selection in dairy cattle breeding schemes. *Journal of Dairy Science*. 85: 1869–1880.

2. Abecasis, G.R., D. Ghosh and T.E. Nichols. 2005. Linkage disequilibrium: ancient history drives the new genetics. *Human Hereditary*. 59: 118–124.

3. Ardlie, K., S. Liu-Cordero, M.A. Eberle, M. Daly and J. Barrett. 2001. Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. *American Society of Human Genetics*. 69: 582–589.

4. Asins, M.J. 2002. Present and future of quantitative trait locus analysis in plant breeding. *Plant Breeding*. 121: 281–291.

5. Babu, R., K.S. Nair, B.M. Prasanna and H.S. Gupta. 2004. Integrating marker-assisted selection in crop breeding – Prospects and Challenges. *Current Science*. 87: 5-10.

6. Barnet, N. 2000. DNA fingerprinting for routine pedigree recording. In: *Animal breeding: uses of new technologies*. (Eds. Kinghorn, B., J. van der Werf and M. Ryan). 131-142. Desktop Publishing, Sydney.

7. Berger, C.F., C.R. Lu, S.J. Belzer and B.E. Voss. 1994. Research on the uses of technology in science education. In: *Handbook of research on science teaching and learning*. (Ed. Gable, D.L). 466-492. Macmillan Publishing Company, New York.

8. Brascamp, E.W., J.A.M. van Arendonk and A.F. Groen. 1993. Economic appraisal of the utilization of genetic markers in dairy cattle breeding. *Journal of Dairy Science*. 76: 1204–1213.

9. Buchanan, F.C., A.G. Van Kessel, C. Waldner, D.A. Christensen, B. Laarveld and S.M. Schmutz. 2003. Hot topic: An association between a leptin single nucleotide polymorphism and milk and protein yield. *Journal of Dairy Science*. 86: 3164-3166.

10. Butterfield, M.K., A. D'Hont and N. Berding. 2001. The sugarcane genome: A synthesis of current understanding and lessons for breeding and biotechnology. *Proceedings of the South African Sugarcane Technology*. 75: 1–5.

11. Chapman, S., M. Cooper, D. Podlich and G. Hammer. 2003. Evaluating plant breeding strategies by simulating gene action and dryland environment effects. *Agronomy Journal*. 95: 99-113.

12. Dekkers, J.C.M. 2004. Commercial application of marker- and gene-assisted selection in livestock: Strategies and lessons. *Journal of Animal Science*. 82: 313– 28.

13. Dekkers, J.C.M. and F. Hospital. 2002. The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics*. 3: 22–32.

14. delMas, R. 1997. A framework for the development of software for teaching statistical concepts. In: *Research on the role of technology in teaching and learning Statistics* (Eds. Garfield J.B. and G. Burril). 85-99. International Statistics Institute, Netherlands.

15. Doerge, R.W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. *Nature Reviews Genetics*. 3: 43–51.

16. Donovan, T.M. and C.W. Welden. 2002. S*preadsheet exercises in ecology and evolution*. Sinauer Associates Inc., U.S.A.

17. Duit, R. and J. Confrey. 1996. Reorganising the curriculum and teaching to improve learning in science and mathematics. In: *Improving teaching and learning in science and mathematics*. (Ed. Treagust, D.F., R. Duit and B.J. Fraser). 79-93. Teachers College Press, New York.

18. Edwards, M.D. and N.J. Page. 1994. Evaluation of marker-assisted selection through computer simulation. *Theoretical and Applied Genetics*. 88: 376–382.

19. Falconer, D.S. 1989. *Introduction to quantitative genetics, 3rd ed.* John Wiley & Sons Incorporated, New York.

20. Falconer, D.S. and T.F.C. Mackay. 1996. *Introduction to quantitative genetics, 4th* ed. Longman Group, New York.

21. Farnir, F., W. Coppieters, J. Arranz, P. Berzi, N. Cambisano, B. Grisart, L. Karim, F. Marcq, L. Moreau, M. Mni, C. Nezer, P. Simon, P. Vanmanshoven, D. Wagenaar and M. Georges. 2000. Extensive genome-wide linkage disequilibrium in cattle. *Genome Research*. 10: 220–227.

22. Frisch, M., M. Bohn and A.E. Melchinger. 2000. PLABSIM: Software for simulation of marker-assisted backcrossing. *Heredity*. 91: 86–87.

23. Gaut, B.S. and A.D. Long. 2003. The low-down on linkage disequilibrium. *The Plant Cell*. 15: 1502-1506.

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

25. Gilean, A. and T. McVeana. 2002. A genealogical interpretation of linkage disequilibrium. *Genetics*. 162: 987–991.

26. Gillespie, J.H. 2004. *Population genetics: A concise guide, 2nd ed.* Johns Hopkins University Press, U.S.A.

27. Gomez-Raya, L., H.G. Olsenb, F. Lingaasc, H. Klunglandb, D.I. Vågeb, I. Olsakerc, S. Talleb, M. Aaslandb and S. Lienb. 2002. The use of genetic markers to measure genomic response to selection in livestock. *Genetics*. 162: 1381-1388.

28. Good, R. and C. Berger. 1998. The computer as a powerful tool for understanding science. In: *Teaching science for understanding: A human constructivist view*. (Eds. Mintzes, J.J., J.H. Wandersee and J.D. Novak). Academic Press, U.S.A.

29. Hancock, C.E. 2006. Identification and remediation of student difficulties with quantitative genetics. Unpublished PhD thesis, University of KwaZulu-Natal.

30. Hartl, D.L. and A.G. Clark. 1997. *Principles of population genetics, 3rd ed.* Sinauer Associates, Sutherland.

31. Hendrick, P.W. 2000. *Genetics of populations, 2nd ed.* Bartlett Publishers Inc., London

32. Hospital, F., L. Moreau, F. Lacoudre, A. Charcosset and A. Gallais. 1997. More on the efficiency of maker-assisted selection. *Theoretical and Applied Genetics*. 95: 1181–1189.

33. Jansen, R.C. 2001. Quantitative trait loci in inbred lines. In: *Handbook of statistical genetics*. (Ed. Balding, D.J., M. Bishop and C. Cannings). 567-593. John Wiley & Sons Limited, U.S.A.

34. Kinghorn, B. 2000. The building blocks of quantitative genetics. In: *Animal breeding: uses of new technologies*. (Eds. Kinghorn, B., J. van der Werf and M. Ryan). 11-18. Desktop Publishing, Sydney.

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

36. Li, C.C. 1976. *First course in population genetics*. Boxwood Press, U.S.A.

37. Lunce, L.M. 2004. Computer simulations in distance education. *International Journal of Instructional Technology & Distance Learning*. 1: 1550–6908.

38. Lynch, M. and B. Walsh. 1998. *Genetic analysis of quantitative traits*. Sinauer Associates, Inc., USA.

39. Mauricio, M. 2001. Mapping quantitative trait loci in plants: Uses and caveats for evolutionary biology. *Nature Reviews*. 2: 370–381.

40. Meuwissen, T.H.E. and M.E. Goddard. 1996. The use of marker haplotypes in animal breeding schemes. *Genetics of Selective Evolution*. 28: 161-176.

41. Meuwissen, T.H.E. and J.A.M. van Arendonk. 1992. Potential improvements in rate of genetic gain from marker-assisted selection in dairy cattle breeding schemes. *Journal of Dairy Science*. 75: 1651–1659.

42. Meuwissen, T.H.E., A. Karlsen, S. Lien, I. Olsaker and M.E. Goddard. 2002. Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics*. 161: 373–379.

43. Montaldo, H.H. and C.A Meza-Herrera. 1998. Use of molecular markers and major genes in the genetic improvement of livestock. *Electronic Journal of Biotechnology*.

44. Newby, T.J., D.A. Stepich, K.D. Lehman and J.D. Russell. 2000. *Instructional technology for teaching and learning: Designing instruction, integrating computers and using media, 2nd ed.* Prentice-Hall, New Jersey.

45. Ohta, T. 1981. Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proceedings of the National Academy of Science*. 79: 1940–1944.

46. Oren, T.I. and B.P. Zeigler. 1979. Concepts for advanced simulation methodologies. *Simulation*. 32: 69-82.

47. Partner, P.L.R., M.L. Smith, W. Spoor and M.I. Clarkson. 1993. Computer simulation of selection in a hypothetical crop species. *Cabios*. 9: 597-605.

48. Podlich, D.W. and M. Cooper. 1998. QU-GENE: A simulation platform for quantitative analysis of genetic models. *Bioinformatics*. 14: 632–653.

49. Raven, P.H. and G.B. Johnson.1999. *Biology, 5th ed.* McGraw-Hill Inc., U.S.A.

50. Rothwell, N.V. 1993. *Understanding genetics: A molecular approach*. John Wiley & Sons Inc., U.S.A.

51. Sargent, R.G. 1992. Verification and validation of simulation models. In: *Progress in modelling and simulation*. (Ed. Cellier, F.E.). 37-47. Academic Press, London.

52. Snustad, D.P. and M.J. Simmons. 2000. *Principles of genetics, 3rd ed.* John Wiley & Sons, Inc., U.S.A.

53. Soller, M. and Beckmann, J.S. 1988. Genomic genetics and the utilization for breeding purposes of genetic variation between populations. In: *Proceedings of the second international conference on quantitative genetics*. (Eds. Weir,

B.S., E.J. Eisen,, M.M. Goodman and G. Namkoong). 161-189. Sinauer Associates Inc., U.S.A.

54. Stella, A., M.M. Lohuis, G. Pagnacco and G.B. Jansen. 2002. Strategies for continual application of marker-assisted selection in an open nucleus population. *Journal of Dairy Science*. 85: 2358-2367.

55. Sterman, J.D. 1991. A skeptic's guide to computer models. In: *Managing a nation: The microcomputer software catalog*. (Ed. Barney, G.O.). 209-229. Westview Press Inc., U.S.A.

56. Terwilliger, J.D., S. Zöllnerc, M. Laanc and S. Pääboc. 1998. Mapping genes through the use of linkage disequilibrium generated by genetic drift: 'Drift mapping' in small populations with no demographic expansion. *Human Heredity*. 48: 138–154.

57. Tinker, N.A. and D.E. Mather. 1993. GREGOR: Software for genetic simulation. *Heredity*. 84: 237-239.

58. Tuncer, O. 1981. Concepts and criteria to assess acceptability of simulation studies: A frame of reference. *Simulation Modeling and Statistical Computing*. 24: 180–189.

59. van der Werf, J. 2000a. Basics of marker assisted selection. In: *Animal breeding: Uses of new technologies*. (Eds. Kinghorn, B., J. van der Werf and M. Ryan). 119-127. Desktop Publishing, Sydney.

60. van der Werf J. 2000b. Selection theory and components of genetic change. In: *Animal breeding: Uses of new technologies*. (Eds. Kinghorn, B., J. van der Werf and M. Ryan). 20-34. Desktop Publishing, Sydney.

61. Weller, J.I., Y. Kashi and M. Soller. 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *Journal of Dairy Science*. 73: 2525–2537.

62. Weller, J.I., M. Golik, E. Seroussi, E. Ezrat and M. Ron. 2003. Population-wide analysis of a QTL affecting milk-fat production in the Israeli Holstein population. *Dairy Science*. 86: 2219–2227.

63. Williams, J.L. 2005. The use of marker-assisted selection in animal breeding and biotechnology. *Revolutionary Science and Technology*. 24: 379–391.

64. Yin, X., P. Stam, M.J. Kropff and A. Schapendonk. 2003. Crop modeling, QTL mapping, and their complementary role in plant breeding. *Agronomy Journal*. 95: 90–97.

65. Xie, C. and S. Xu. 1998. Efficiency of multistage marker-assisted selection in the improvement of multiple quantitative traits. *Heredity*. 80: 489-498.

66. Yousef, G.G. and A.J. John. 2001. Comparison of phenotypic and marker-assisted selection for quantitative traits in sweet corn. *Crop Science*. 41: 645-655.

67. Zhang, W. and C. Smith. 1992. Simulation of marker-assisted selection utilizing linkage disequilibrium. *Theoretical and Applied Genetics*. 86: 492-496.