INTROGRESSION OF FAVOURABLE ALLELES INTO LIVESTOCK POPULATIONS

Eileen Elizabeth Wall

M. Agr. Sc. (National University of Ireland)

A thesis submitted in fulfilment of

the requirements for the degree of

Doctor of Philosophy

University of Edinburgh

2002

DECLARATION

I declare that this thesis is my own composition and that the research described in it is my own work, except where otherwise stated.

Eileen E Wall, 2002

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to the many individuals who made this study possible and provided valuable assistance over the past three years.

I am deeply indebted to Peter Visscher and John Woolliams for supervising this work and providing advice and guidance throughout the course of my PhD.

This work would not have been possible without the funding of the Teagasc Walsh Fellowship Scheme, Irish Cattle Breeders Federation and Ministry of Agriculture, Food and Fisheries Postgraduate Studentship Scheme. I thank Brian Wickham (ICBF) and Seamus Hanrahan (Teagasc) for their encouragement to pursue this PhD and their continued support.

Special thanks to the staff of all the institutes that housed me during my three years (IERM and ICABP, University of Edinburgh and the Roslin Institute). The diversity of locations and colleagues has been a challenge at times but has provided me with diversity of knowledge, skills and new insight into this work. Special thanks to Liz Archibald for looking out for me from day one at Roslin.

Many thanks to all my collaborators, especially Jennie Pryce and Grant Walling for being great friends as well as great co-workers.

I have had made many close friends during my studies. My fellow students and office mates have always provided inspiration for this work and suitable amusing distractions when needed. I thank all of you but I would especially like to thank Tim Roughsedge (Steps wannabe!), Dr. Dr. Stuart Thomas and Huw Jones (you're the greatest dancer!). A word of thanks to all the MSc students for being great classmates in 1998/99 and great mates over the three years, especially Snakey, Daniel and Angela. Thanks to all my friends outside the work loop for providing excellent diversions when I needed them.

Last, my sincerest and deepest thanks to my family, especially my parents for their assistance, love and support throughout all my life, particularly whilst pursuing this PhD. I dedicate this thesis to my parents, Chris and Myra Wall, with all my love.

PUBLICATIONS

Wall, E.E., Woolliams, J.A., Visscher, P.M., 2002. Genetic lag in a Meishan x Large White pig backcross population: A simulation study. *Proceedings of the Annual Meeting of the British Society of Animal Science, 2002.*

Wall, E.E, Visscher, P.M. and Woolliams, J.A., 2002. Genomic contributions in livestock gene introgression programmes. *Accepted for Genetics, Selection, Evolution.*

Wall, E.E, Visscher, P.M. and Woolliams, J.A., 2001. Identity by descent around the target locus in a gene introgression programme. *Proceedings of the 52nd Annual Meeting of the European Association for Animal Production.* pp 61.

Wall, E., Woolliams, J., Hanrahan, J.P. and Visscher, P., 2001. Genetic lag in livestock introgression programmes *Proceedings of the Irish Agricultural Research Forum, 2001.* pp 49.

Pryce, J.E., **Wall, E.E.,** Lawrence, A.B., Simm, G., 2001 Breeding strategies for organic dairy cows. *Proceedings of the 4th NAHWOA Workshop, Wageningen, The Netherlands, 24th-27th March, 2001.*

Pryce, J., **Wall, E.,** Coffey, M., Simm, G., 2001. Is organic breeding black and white? Breeding stratergies for organic dairy cattle. *Proceedings of the British Cattle Breeders Club, 18th-19th of January, 2001.* p. 36-41.

Wall, E.E., Woolliams, J. A., Hanrahan, S. P., Visscher, P. M., 2000. Genomic contributions in livestock introgression programmes. *Young Research of the Year, Teagasc Seminar Day, Dublin, Ireland, 15th November, 2000.*

Wall, E.E., Woolliams, J. A. & Visscher, P. M., 2000. Genomic contributions in a gene introgression program *6th QTL and MAS International Workshop, Tune, Denmark, 7th-8th September, 2000.*

Wall, E.E, Woolliams, J.A., Visscher, P.M., 2000. Genomic contributions in a cattle introgression program. *Proceedings of the Annual Meeting of the British Society of Animal Science, 2000.* pp 42.

ABSTRACT

Not all commercial lines or breeds will contain the best alleles for all traits considered of economic importance. An inferior line may contain many alleles affecting a trait of interest that are not present in any commercial line. Introgression strategies are concerned with the transfer of a favourable allele from a "wild" type line to a "domestic" or "commercial" type line. The aim of introgression is to incorporate the favourable allele into a population, containing alleles affecting commercial traits, without bringing in the remainder of the less advantageous alleles from the donor population, thus limiting the loss of beneficial alleles in the commercial population. After a number of generations of backcrossing heterozygous individuals are selected and *inter se* crossed to produce individuals that are homozygous for the allele of interest. In this thesis, the genome composition in the resultant population in terms of contributions from ancestral populations, allelic diversity and homozygosity and identical by descent (IBD) are investigated using analytical methods and computer simulation.

Theoretical predictions of proportional genomic contributions from ancestors used at each generation of crossing in an introgression programme agreed closely with simulated results. It was shown that the number of backcross generations and the length of the chromosome affected proportional genomic contributions to the carrier chromosomes while population structure had no significant effect. The equations derived describing genomic contributions provided a framework to predict the genomic composition of a population after the introgression of a favourable donor allele. These ancestral contributions were assigned a value and used to predict genetic lag in different types of livestock introgression programmes.

The loss of alleles at neutral and selected loci in backcross breeding programmes and introgression studies was examined by quantifying and deriving a simple theoretical approximation of the number of ancestral alleles contributing to neutral loci, at and around a locus of selection. There was close agreement between the simulations and predictions for all population structures studied for carrier and non-carrier chromosomes in various backcross schemes. The loss of allelic diversity was reduced by decreased population sizes, increased offspring group sizes, increased backcross generations and distance from the target locus. Although donor allelic diversity dropped quickly at a neutral locus the overall allelic diversity did not decline as new recipient breed alleles entered the system to maintain a more substantial allelic diversity.

The level of homozygosity (due to both alleles at a locus tracing back to the same ancestral population) on the carrier chromosome after introgression was studied using simulations. Homozygosity was higher on smaller chromosomes (0.5 Morgan VS 1 and 2 Morgan). For the extreme lengths (8 and 16 Morgans) the proportion of homozygosity began to rise due to the high genomic contributions from individuals used in the last recipient generation. Increasing the number of backcross generations decreased homozygosity and genomic contribution from donor individuals, in turn increasing the genomic contribution from recipient individuals. IBD (2 alleles at a locus are identical by descent leading to homozygosity by descent in the population homozygous for the target allele) in the region of the target locus decreased as distance from the target locus increased over all parameters examined. As allelic diversity at a locus decreased, donor IBD increased but only in close proximity to the target locus. IBD did not continue to decrease with distance from the target locus due to the reduced selection pressure of the target locus. The loss of allelic diversity was reduced by decreased population sizes, increased offspring group sizes, increased backcross generations and increased distance from the target locus. In turn, IBD around the target locus increased as offspring group size increases and population size decreases.

Understanding the genomic contributions in these breeding programmes will allow breeders to introduce optimal recipient animals at appropriate backcross generations. Predicting genetic lag and the influence of the linkage drag segment will allow focussed selection decisions if selection is applied additional to the selection at the target locus. Allelic diversity information can be used as a guide to maintaining an optimum level of variation around the target gene/s and elsewhere in the genome, thus minimising the risk of producing tracts of IBD after the *inter se* cross. These factors will help to design introgression schemes that achieve the goal of introducing new alleles into a recipient population and result in an optimal breeding population post introgression.

CHAPTER ONE

INTRODUCTION

1.1. Introduction

For many years, breeders have strived to improve plant and livestock populations by selecting on the best phenotype for traits of economic importance. Genetic variation in both plants and animals is one cause of phenotypic differences in performance. The efforts to improve economic traits have been upon the use of phenotypic and relationship information coupled with complex statistical methodology. This methodology has been a powerful tool in animal improvement but has been used with little knowledge of the underlying genetic variation affecting the observed trait.

Genotypic information allows a breeder to identify, at a genetic level, the individuals carrying the best alleles for the traits of interest. Utilising genotypic information may improve the accuracy of selection and rate of genetic progress. Additional benefits are also gained because an animals genetic "worth" for a trait can be observed at any age or in either sex, which allows speedier selection decisions to be made. There has been success in mapping regions of various livestock genomes that explain a proportion of the variance in many qualitative and quantitative traits. These traits range from production (Georges *et al.*, 1995), reproductive (Rothschild *et al.*, 1996; Wilson *et al.*, 2001), product quality (Le Roy *et al.*, 2000), health (Zhang *et al.*, 1998; Coltman *et al.*, 2001; Yonash *et al.*, 1999; Sharif *et al.*, 1998) and welfare traits (Brenneman *et al.*, 1996; Schmutz *et al.*, 1995).

Applying results from mapping studies to breeding programmes may increase rates of genetic improvement. This will especially be the case where phenotypic information is unable to be collected, *e.g.*, sex-limited traits, collection of

information is costly or when the trait has a low heritability (Smith and Simpson, 1986). This genotypic information assists within breed selection or allow breeders to exploit phenotypes from other breeds or lines (maybe non-commercial breeds) using introgression (discussed below).

1.2. Gene Mapping and Genetic Markers

Recent developments in the mapping of livestock genomes have created the opportunity of utilising genomic information in breeding programmes to benefit genetic progress. A genetic map reflects the order of loci (markers and/or functional genes) on a chromosome within the genome of an organism. Genetic markers are informative points in the genome with multiple variants at a known location. They are ordered by the distance between them to produce a linkage map. The distance relates to the probability of recombination between any two marker loci. Recent advances in the use of marker technology have developed a multitude of markers providing comprehensive coverage of the genome. Figure 1.1 shows examples of a linkage map for sheep chromosome 6 and cattle chromosome 2.

The ultimate aim of gene/QTL mapping projects is to identify regions of the genome affecting targeted traits, the causative gene mutations and their mode of action. There are examples of single genes which control variation between animals which are large enough to be individually recognised, such as the polled gene in cattle located on cattle chromosome 1 (Brenneman *et al.,* 1996). Microsatellite markers have been identified in complete linkage to the polled locus (Georges *et al.*, 1993a; Schmutz *et* al., 1995). Performance traits are rarely controlled by single genes and

are commonly the combined effect of many genes and environmental factors controlling the quantitative difference between individuals. Loci affecting quantitative traits are referred to as quantitative trait loci (QTL). Until underlying genes affecting quantitative traits have been mapped, closely linked markers are used to identify chromosomal regions.

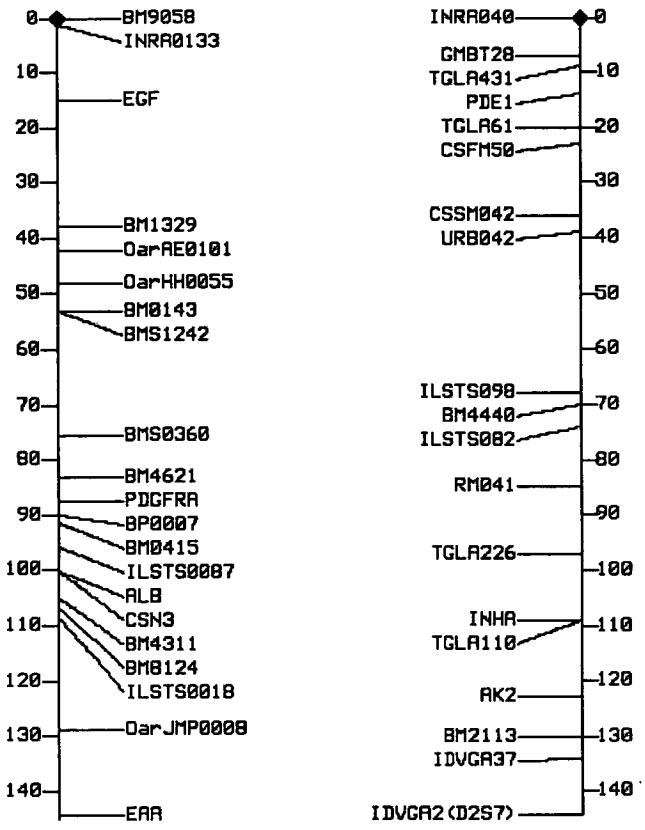

FIGURE 1.1. Linkage maps for sheep chromosome 6 (right map) and cattle chromosome 2 (left map) taken from ArkDB at the Roslin Institute (ArkDB webpage, Hu *et al,* 2001)

Genetic markers are used because of linkage to the trait loci. Genetic markers can be used to identify whether an animal is carrying a locus of interest and can help trace the inheritance of the chromosomal section to the offspring. The necessity for large

numbers of markers throughout the genome has led to the rapid development of many livestock species' maps (Hu *et al*, 2001).

It is useful if genetic markers are highly polymorphic because animals and lines are likely exhibit different genotypes and a co-dominant marker allows all possible genotypes, both heterozygous and homozygous, at the locus to be identified. Protein polymorphisms (*eg* blood groups) tend to be co-dominant but are not highly polymorphic and sufficient in number of map an entire genome. In contrast to protein polymorphisms, DNA-based genetic markers generally comply with the above criteria (Falconer and Mackay, 1996). Restriction fragment length polymorphisms (RFLPs), variable number of tandem repeats (VNTRs), minisatellite or microsatellite loci (simple sequence repeats, SSRs, of differing lengths) are generally used in mapping studies (Ott, 1991). RFLPs are less polymorphic and are difficult and expensive to identify (Falconer and Mackay, 1996). Minisatellites produce DNA-fingerprints, which have been used previously in introgression studies, *e.g.*, Hillel *et al.* (1990) and Yancovich *et al.* (1996). These produce a banding pattern shown post-electrophoresis, which is almost unique to an individual. However, DNA-fingerprints can be difficult to allocate to an allelic pair.

Other types of DNA markers that are increasingly used are amplified fragment length polymorphism (AFLP) and random amplified polymorphic DNA (RAPD). RAPDs are dominant (carriers and one type of homozygote are inseparable), indicating the presence or absence of an allele and therefore do not give full information about the marker site (Ragot and Hoisington, 1993). These dominant markers are useful in a

backcross analysis as only heterozygotes for the marker allele should exist and it is unnecessary to distinguish homozygotes. The majority of livestock QTL and marker assisted breeding studies have used microsatellites because they are abundant in many livestock species (Lynch and Walsh, 1998). They tend to be highly polymorphic and are codominant.

More recently in human genetics, there has been a move to the discovery of Single Nucleotide Polymorphisms (SNPs) and subsequent use in linkage and association studies. SNPs are single base pair variants and can occur every 100 to 600 bp in the human genome (Heller and Eisenberg, 1998). They are also relatively genetically stable, which makes them ideal for constructing a high-resolution, global genomic map. SNPs are currently used extensively in human genetics studies and maybe utilised in future animal studies. The low mutation rate, abundance and ease of scoring of SNPs aids the identification of genes and eases the tracking of its descent in a population or pedigree. Due to these properties, the association between SNPs and a phenotypic trait, such as disease susceptibility in humans, can then be studied. Research in the discovery of SNPs and their association to various phenotypic traits is rapidly increasing and the number of human reference SNPs that are mapped is approaching 2 million (SNP Home Page, 2001, http://www. ncbi. nlm. nih. gov /SNP /index.html). SNP maps in livestock populations are currently being developed and this information could assist the discovery of genes affecting *e.g.,* disease resistance/susceptibility.

1.3. Marker assisted introgression

1.3.1. What is introgression and what is its role in livestock populations?

Not all commercial lines or breeds will contain all the best alleles for all traits of economic importance (Tanksley and McCouch, 1997). Given the geographical and the genetic diversity of breeds of livestock it is reasonable to believe that some alleles have become fixed in populations before selection programmes were introduced to the breed. In fact, an "inferior" line can contain "superior" alleles affecting a trait of interest that are not present in commercial lines.

Rothschild *et al.* (1994) found evidence for a QTL affecting litter size in a cross between Large White and Chinese Meishan pigs, which increased litter size by one piglet per litter. This QTL for litter size was found to have originated from the non-commercial Meishan breed and would benefit commercial populations if introduced into commercial lines which are superior for meat traits. Other studies involving non-commercial Meishan pigs have found a QTL on chromosome 7 decreasing backfat depth (*e.g.*, Walling, 1998). This finding was interesting because the Meishan is significantly fatter than commercial Western breeds hence the allele was "cryptic" (*i.e.*, the allele improved a trait despite being at a high frequency in a breed with poorer commercial performance for the trait).

Introgression strategies generally are concerned with the transfer of a favourable allele from a "wild" type line to a "domestic" or "commercial" type line. The aim of introgression is to bring in the favourable allele for a major gene or QTL into a

recipient population without bringing in the remainder of the less advantageous alleles from the donor population. This prevents the loss of beneficial alleles from the commercial population. After generations of backcrossing, heterozygous individuals are selected and *inter se* crossed to produce individuals homozygous for the allele of interest. Assuming no linkage between the target locus and all other loci in the genome, the contribution of the donor parent is reduced by half with each successive generation of backcrossing (Hill, 1993; Toojinda *et al.*, 1998).

1.3.2. Why use introgression in cattle populations?

Introgression may be used as a tool of genetic improvement for a breed or as a tool for addressing new market demands. The dairy cattle population has been undergoing selection for commercial milk production traits for many generations. This has generally entailed selection index and BLUP methodology (Henderson, 1988). More recently there has been efforts to include longevity and fertility information as selection criterion to increase the productive lifetime of dairy cows, reducing the need for involuntary culling and ultimately reducing costs. There has therefore been a move away from sole production based indices to indices including type, fertility, health and longevity traits (*e.g.*, Pryce *et al.*, 1998; Ducrocq and Sölkner, 1998; Jairath *et al.*, 1998). These methods have centred on phenotypic measurements of traits on cows as the only way of "measuring" genetic differences between animals, but MAS/MAI may help to address new breeding objectives faster.

For phenotypic information to be of use for genetic improvement, it is important that the data provides maximum information on the underlying genetic variation (Lynch

and Walsh, 1998). Unfortunately, this is not always achieved and examining the effects of individual genes or QTL could aid selection decisions. Information generated by markers can be included in future selection programmes. Different traits are also becoming important to consumer and producers.

FIGURE 1.2. The use of backcrossing (BC) to introgress a polled gene to a horned breed*

Cross	Mating programme	Average proportion of genes from horned breed in sel.
F1	Polled × Horned	
1 (BC1)	Selected polled offspring × Horned	$1 - ½^1 = ½$
2 (BC2)	Selected polled offspring × Horned	$1 - ½^2 = ¾$
3 (BC3)	Selected polled offspring × Horned	$1 - ½^3$
T (BC*T*)	Polled offspring	$1 - ½^T$

*Nicholas (1996)

Figure 1.2 illustrates a breeding programme whereby a polled gene is introduced into a horned cattle population. First the donor line containing the gene to be introgressed (polled) is crossed with the recipient line (horned) followed by *T* generation of backcrossing. This demonstrates how the donor parental contribution is reduced by 50% with each successive generation of backcrossing.

1.3.3. The benefit of markers in introgression and breeding schemes.

In any genetic improvement programme information must be used to differentiate between the "good" and "bad" animals. Figure 1.2 (introgression polled allele) demonstrates the introgression of a dominant allele that can be measured phenotypically. As described earlier, the location of the polled allele is known and linked markers have been identified. Even for such an easily measurable trait markers can be used in the introgression programme at the *inter se* cross to help identify homozygous individuals. If the allele is recessive then linked markers can be used to track carriers during backcrossing.

Not all traits are controlled by a single gene or can be measured phenotypically in both sexes and all ages. Traits like milk production cannot be measured at an early age or in both sexes so performance information from relatives is required to ascertain genetic merit. Markers could be used to gain information on traits that are difficult to measure without harming or pressuring the animal, such as carcass traits or disease resistance traits. Marker information becomes a useful tool when trying to introgress an allele affecting such traits. These traits, which are more difficult to measure, are becoming more important in livestock breeding schemes worldwide.

1.3.4. Examples of plant and animal marker assisted introgression schemes.

Introgression programmes in plant populations have been used successfully to incorporate new traits, such as disease resistance, into commercial plant varieties. Using AFLPs as markers for disease resistance, Toojinda *et al.* (1998) produced commercially viable lines of barley with resistance to stripe rust after one generation

of marker assisted backcrossing. Loss of performance, with respect to selection responses in a commercial line not undergoing introgression, was not examined. With studies in tomato lines, Chetelat *et al.* (1995) found that using marker assisted introgression (MAI) compared to phenotypic introgression allowed the speedy creation of commercially viable plant lines.

Practical introgression schemes have been used less in commercial livestock populations. One example is the introgression an allele for increased fecundity (Booroola gene) from a Merino strain to the Romney breed of sheep (Piper *et al.*, 1985). This study did not use markers or genotypic information to aid the introgression of the gene and only used phenotypic information. This caused a number of problems because females had to produce progeny before they could be genotyped (Booroola genotype was assigned by measurement of ovulation rate) and males could be only allocated a Booroola genotype by progeny testing. Montgomery *et al.* (1994) later mapped the Booroola gene to chromosome 6. An example in pig populations is the study of the halothane gene, affecting meat quality by causing porcine stress syndrome (PSS) leading to Pale Soft and Exudative meat (Hanset *et al.*, 1995). The gene itself was identified in 1991 and culling took place of all carriers from nucleus populations. This can be described as an "extragression" of the gene from the population. To date, the majority of studies into the viability of introgression schemes in commercial livestock populations have been based on simulation studies. Introgression in livestock populations is difficult at present due to the lack of identified alleles, costs, time and the genetic lag.

1.3.5. Congenic lines

Mouse models have been used to support the genetic study of complex human diseases. The differences between two inbred lines are studied by creating congenic lines. Congenic lines are created where genomic sections (containing interesting loci) are introgressed into different genetic backgrounds (usually different inbred lines) to aid the study of complex human diseases. Congenic breeding strategies follow the same procedure of backcrossing followed by an *inter se* cross that is used for introgression. Traditional protocols for creation of congenic strains call for 12 backcross generations with an average donor derived genomic segment of 20 cM (Wakeland *et al.*, 1997). This procedure usually takes between 2.5 and 3 years to complete (Markel, 1997).

Markel *et al.* (1997), Wakeland *et al.* (1997) and Visscher (1999) studied various issues of using markers in the creation of congenic strains, or speed congenics. These studies showed that the process of congenic strain creation could be accelerated with the use of markers. The strategies and predictions used for marker-assisted introgression (discussed below) would also apply for speed congenics.

1.4. Factors affecting the efficiency of marker assisted introgression

1.4.1. Quality of marker information

The type, quantity and density of genetic markers chosen in MAI programmes can play an important role in deciding the success or failure of the programme. Studies by Hillel *et al.* (1990; 1993) examined the use of DNA fingerprints as markers to

identify high "genomic similarity" to the recipient line in backcross individuals. This is akin to selection against the unfavourable genomic background of the donor breed. This approach was put into practice by Yancovich *et al.* (1996), introducing the avian naked neck gene from a layer line into a commercial broiler line of chickens and compared the method with using conventional selection on body weight to assist genome recovery. The backcross animals from the two selection regimes had the same phenotypic performance (both 3.1-3.9% heavier at the second backcross generation than randomly selected animals). It was found that the additional genome recovery of the recipient breed using the DNA fingerprints was higher than that from phenotypic selection. The number of markers used in this study was low and coverage of the genome was poor, hence, the proportion of donor genome was not well estimated. The theory of Hillel *et al.* (1990) did not account for recombination around the marker loci. Hillel *et al.* (1990) assumed that the proportion of donor genome from a line used for backcrossing is the same for linked and unlinked loci

Many studies on introgression (*e.g.*, Hospital *et al.*, 1992; Hillel *et al.*, 1993) assume that the introgressed gene could be located with complete precision. Keeping track of QTL can be difficult due to the loss of target loci through recombination and with less informative markers. Spelman and van Arendonk (1997) have shown that MAS with a falsely identified putative QTL had a comparable genetic loss to that of a breeding scheme not using QTL information. Visscher *et al.* (1996) derived an optimum marker spacing for introgression based on the estimate of the standard error of the effect and the position of the QTL. A distance of 10-20 cM between the QTL and selection marker was concluded as appropriate for introgression.

1.4.2. The use of reproductive and genetic technologies to improve the efficiency of introgression.

As advances in various reproductive and genetic technologies become available their application and usefulness in breeding programmes has to be addressed. These techniques range from easier and cheaper genome mapping and genotyping (Westman & Kresovich, 1997) to physiological, artificial breeding techniques, such as MOET, transgenics and even cloning of animals. These types of techniques could play a vital role in the success and speed of gene introgression programmes in some livestock species when litter size is low, generation interval is long and/or only a small population can be maintained during introgression.

Gama *et al.* (1992) studied the benefit of transgenics in conjunction with an introgression programme. Transgenic technology can be used to insert the QTL or gene of interest from an inferior line and incorporate it into the genome of a superior line or breed. Gama *et al.* (1992) considered using the transgenic organism as the donor animal of an introgression programme. The creation of transgenic animals is not yet a commercially viable option in even elite livestock populations due to the expense and the technical difficulties involved in the procedures. Transgenics could be used on a small scale in introgression programmes by "creating" the donor population, instead of looking for naturally occurring populations carrying a favourable donor allele. Introgression could then be used as a method of introducing the transgene into a wider population. The creation of transgenic lines may be of benefit in relation to overcoming problems associated with health and welfare. For

example, a transgenic line could be created carrying a natural disease resistance allele from a completely non-compatible mating species and the transgene could then be introgressed into a commercial population.

Visscher and Haley (1999) showed that the *inter se* crosses are a more important component of genetic lag (see 1.5.4.). Large numbers of animals need to be selected on their genotype and leaves little opportunity to select for other traits. Selection intensity is reduced at the *inter se* cross stage, especially in the females. A MOET scheme may be useful in increasing the numbers of progeny per female and therefore reducing genetic lag (Visscher and Haley, 1999). A large population size is needed during the introgression period to maximise expected responses to selection for commercial traits and allow continuous evaluation of the value of the introgressed allele (Gama *et al.*, 1992). The larger family size that a MOET scheme would produce also allow markers to trace and select for the introgressed gene throughout backcrossing.

Bishop and Woolliams (1991) examined, by simulation, that the introgression of the sex-determining region Y (SRY) transgene into a cattle population using a transgenic animal as a donor. It was shown that after 15 years, if all matings in the scheme were made to transgenic animals, the transgenic males were 9 years behind progress in the non-transgenic population. This was achieved with an increase in the ratio of males: females from 1:1 to 3:1. This study assumed one calf per mating, but the use of a MOET scheme would be more beneficial and made the scheme more efficient.

1.5. Difficulties with gene introgression

1.5.1. Cost of an introgression scheme

The cost of an introgression programme has generally been overlooked in many studies. This is an important issue and it is generally dependent on the cost of genotyping. The loss of profit due to the genetic lag could also be estimated. Gama *et al.* (1992) concluded that an allele would need to have an economic (or beneficial) effect equivalent to between one and two generations of selection in the commercial population to compensate for genetic lag. Visscher and Haley (1999) examined the cost of an introgression scheme in a pig nucleus of 400 animals (40 litters per generation). Present day costs were assumed and the total cost of such a programme (after 4 backcross generations and *inter se* cross) was estimated to be $180,200 over the entire scheme. This calculation ignored the cost of genetic lag incurred because of the introgression programme and potential extra earnings to be gained from the introgression population carrying the favourable allele.

Koudandé *et al* (2000) examined the cost of an introgression scheme in terms of the cost of genotyping and rearing of animals. This study was set up to examine the efficiency of the introgression of 3 QTL in mouse lines. This study found it was necessary to minimise the number of animals in the scheme to reduce costs. Using alternative backcross and *inter se* cross strategies (*e.g.*, creating multiple lines during backcrossing, each carrying a different target QTL and then crossing these to create carriers of all 3 QTL) helped reduce the number of animals and genotypings. Creating two lines during backcrossing (each carrying 2 of the QTL therefore 1 QTL

in common) and using all animals that are at least heterozygous for all 3 QTL for a second *inter se* cross had the greatest effect on minimising the rearing and genotyping costs out of all the alternatives studied. The costs of genotyping and the number of animals that need to be maintained throughout introgression could be further reduced if QTL locations are precisely known but in many studies there is a large confidence interval associated with the QTL position.

1.5.2. Outbred populations

Many introgression studies to date have assumed that donor and recipient lines are divergent or inbred and that the alleles for each marker used are fixed in both breeds. This is less likely for outbred lines. Pedigree information may be useful to trace back the descent of alleles within lines and this information can be included to estimate the presence of the desired allele. The transmissed marker alleles are not always traceable. This is determined by the level of polymorphism for a marker, with markers of low polymorphism being more difficult to trace (van der Beek and van Arendonk, 1996). Inferior traceability has a negative effect on the MAI programme but the use of closely linked markers can alleviate this problem (Kashi *et al.*, 1990).

Many studies have shown that response to genotypic selection in MAS produced less than optimal response to selection in the medium and long-term compared to the short-term gain (*e.g.*, van der Beek and van Arendonk, 1996; Gibson, 1994; Meuwissen and van Arendonk, 1992). The apparent explanation for this is the genetic variance in early generations was reduced with MAS relative to traditional selection indices, which resulted in less genetic gain in latter generations (Ruane,

1994). Dekkers and van Arendonk (1998) showed the change in response to selection over time is mainly caused by selection causing fixation for the segregating QTL and therefore no additional genetic variance is generated during the course of selection.

The initial difference between the recipient and donor breeds or lines is important because if it is large, phenotypic selection and marker selection have similar results in reducing genetic lag the selection of the homozygous individuals at the *inter se* cross. When working with a introgression programme where the initial phenotypic difference is small, markers add little benefit and perform no better than random selection after individuals have been pre-selected for the gene of interest (Visscher and Haley, 1999).

1.5.3. Linkage drag

The quantity of the donor genome in backcross lines is variable but decreases, on average, by a half with each additional backcross generation. This figure does not take account of the effect of selection, especially selection on the target allele. In the course of gene introgression donor alleles linked to the desired allele are incorporated into the genome of the recipient line by a phenomenon called linkage drag (Brinkman and Frey, 1977). The linkage drag segment is important as it may incorporate other less favourable alleles and drag them into the commercial population. The risk of these unfavourable alleles being dragged into the recipient population is related to its length. Stam and Zeven (1981) found that the chromosomal segment containing the desired introgressed allele (linkage drag) could still be large, *e.g.*, average segment length of 32 cM for a 100 cM chromosome after

six backcross generations. A number of studies have examined the theoretical prediction of the expected length of the linkage drag in backcross breeding programmes (*e.g.*, Hanson, 1959; Stam and Zeven, 1981; Naveira and Barbadilla, 1992).

The use of markers to retrieve the recipient genome can speed the recovery of the genome by up to 2 generations when compared to phenotypic selection (Hospital *et al.*, 1992). During the early stages of introgression the interest is mainly based around the presence of the introgressed allele. Additional selection should be place of markers distant to the target locus. A set of genetic markers close to the target locus may become useful in later backcross generations with the aim of reducing the linkage drag (Hospital *et al.*, 1992). With studies in tomato lines, Chetelat *et al.* (1995) found that the recovery of the recipient genome was increased when using genetic markers in comparison to traditional introgression programmes not using markers. The introgressed segment was reduced from 43.2 cM to 0.8-7.1 cM.

1.5.4. Genetic lag

Introgression programmes may result in a reduction in performance of production traits in the crossbred population because the donor line may be inferior for traits that have been selected for many generations in the recipient line. This results in a "genetic lag" for the introgression programme. The backcross stage of introgression attempts to minimise the donor contamination while keeping the allele of interest in the recipient population. This donor contamination arises from the linkage drag segment and from donor genome contribution elsewhere in the genome. Only when

the introgressed allele has a large effect on the quantitative trait of interest does the total response from selection for genomic similarity (selection of animals with similar genome of recipient or selection against donor genome contamination) surpass phenotypic selection. Gama *et al.* (1992) concluded that a gene would need to have an economic effect equivalent to one to two generations of selection in the commercial population to make it commercially viable to offset for the loss due to genetic lag. A major part of genetic lag seen in an introgression programme can be attributed to the two *inter se* cross generations creating animals homozygous for the allele of interest (Haley and Visscher, 1998).

1.6. Improving the efficiency of gene introgression

1.6.1. Foreground Selection

Hospital and Charcosset (1997) examined the use of foreground selection (selection on the target allele) for the introgression of QTL alleles using markers to control the genotype at the QTL. These results showed that the introgression of the QTL can be controlled easily with two markers optimally placed to the QTL for one backcross generation. Increasing the number of backcross generations required more markers at different positions to account for the recombination between the marker and the QTL. For the introgression of multiple QTL alleles their results showed it was necessary to use at least 3 optimally positioned markers to identify each QTL allele.

1.6.2. Background selection

Background selection uses markers to select against donor contamination in the genome of the recipient population, after selection of individuals based on the target locus genotype. Few studies have been carried out on this procedure in outbred animal populations. The method requires markers placed throughout the genome, ideally fixed for alternative alleles in the donor and recipient lines. Wide coverage of the genome and tracking of the markers could prove costly. Background selection could be also carried out on QTL for commercial traits in the recipient breed.

Hospital *et al.* (1992) used markers to speed up the recovery of the recipient line genome during the backcross stage and considered selection intensity, population size, number and position of markers for chromosomes carrying or not carrying the introgressed allele. Two markers were considered on the chromosome that carries the allele to be introgressed. The study concluded that the speed of retrieving the recipient genome was two backcross generations faster using marker information but this was dependent on the informativeness and frequency of markers used. This study concentrated on the introgression of a single allele.

Groen and Timmermans (1992) compared marker assisted introgression to introgression based on phenotypic information. It was concluded that introgression using marker information did not produce sufficient additional benefit to justify the additional work required to collect genotype information. However, their method

was dependent on parameters that were used in the study. This included the effectiveness of the phenotypic selection in the early generations of backcrossing.

Groen and Smith (1995) carried out a simulation study on the efficiency of selection for genomic similarity to the recipient line compared to selection using phenotypic performance during introgression. The study found that selection for genomic similarity was less efficient in obtaining a high average genotypic value for the backcross line. Selection for genomic similarity however was more efficient in increasing the frequency of the desired allele during the *inter se* cross and displayed a higher recovery of the recipient genome. The linkage disequilibrium, caused by the initial cross between the donor and recipient lines, was decreased within the backcross phase of the introgression, as was the correlated selection response on the quantitative trait.

1.6.3. Reduction of linkage drag in marker-assisted backcrossing

More recently, some studies have examined methods of reducing linkage drag in marker-assisted backcrossing by selecting favourable flanking recombinants, after ensuring the presence of the target locus. This is a form of background selection but specifically on the carrier chromosome. Most of the donor alleles still present in the genome will be on the carrier chromosome (Young and Tanksley, 1989) and therefore it is wise to concentrate efforts on the carrier chromosome.

Frisch and Melchinger (2001) developed predictions of the length of linkage drag when selection was for the presence of the target allele followed by selection of

recipient markers at flanking loci. This study showed the placement of markers should be symmetrical and get closer to the target locus as the number of backcross generations increases. The study examined the cumulative and probability density distribution of the expected linkage drag length when selection is on only one flanking marker. The predictions are difficult to apply to livestock species where multiple animals need to be selected and therefore favourable recombinants will occur across all generations in different family lines.

The study of Hospital (2001) examined the efficiency of marker-assisted backcrossing in reducing linkage drag and produced similar results to that of Frisch and Melchinger (2001) but produced a simpler set of predictions that accounted for multiple generation information. This study found the most important factors in the reduction of linkage drag were the distance of flanking markers to target locus, the population structure and the number of backcross generations. The design of the programme should depend on the desired result. If the introgression scheme needs to be completed in a relatively short time it is necessary to have a large population size to allow for at least one favourable recombination between closely placed flanking markers and the target locus. Hospital (2001) found that performing several backcross generations was a better strategy because costs are drastically decreased due to the decreased population size, number of flanking markers and genotypings. This also could allow for background selection elsewhere in the genome.

1.6.4. Introgression of multiple QTL

The introgression of multiple QTL alleles complicates the introgression scheme because of the difficultly ensuring a sufficient number of carriers of all QTL alleles. Hospital and Charcosset (1997) showed it possible to manage the introgression of up to four unlinked QTL with a few hundred individuals. This study attempted to minimise the number of individuals to be genotyped in order to obtain at least one individual at the end of the backcross phase. This would not be appropriate in livestock introgression as there would be a need for more animals at the end of introgression. The study did not examine genetic variation, diversity and inbreeding which are important considerations in livestock populations, but less so in plants.

Koudandé *et al.* (1999) examined the number of animals to be genotyped during backcrossing when introgressing three QTL alleles. The population size becomes large when selecting for all three QTL alleles throughout the backcross stage in a single line. The study suggested to select and genotype only males during backcrossing until the second last generation of backcrossing to reduce the number of genotypings. Alternative backcross strategies were compared in the subsequent study of Koudandé *et al.* (2000), with each strategy aiming to produce 1000 individuals heterozygous for the target loci at all there QTL at the end of backcrossing. Traditional backcrossing methods (one line carrying all three QTL) require approximately twice as many genotypings compared to alternative methods studied. The number of animals needed to be maintained differed depending on the number of lines being maintained during backcrossing. These alternative methods

required more generations in the introgression scheme, to accommodate for a rotational crossing stage, which would be a problem in agricultural species (Koudandé *et al*, 2000). These studies examined the introgression of QTL in mice and were not limited by large generation intervals and small offspring group size.

In livestock introgression programmes, there is a need for a minimum population size homozygous for the target locus/loci to subsequently produce a sufficient number of individuals for mating to the commercial population. Livestock introgression programmes also requires the population homozygous for the target locus to reproduce whilst maintaining a reasonable inbreeding coefficient. Reproductive techniques can be used to create a more feasible population structure. Artificial insemination (AI) can be used in cattle or sheep to minimise the number of males that are used at each generation whilst monitoring inbreeding. MOET schemes can be used to increase the number of offspring per mating.

1.7. Genetic diversity and inbreeding in introgression programmes and preservation of genetic resources

1.7.1. Loss of variation in backcrossing

Few studies have examined the loss of genetic variation in introgression programmes. It is necessary to maintain genetic variation in livestock introgression programmes to minimise the probability of identity by descent (IBD) at and around target loci. Increasing IBD of donor alleles around the target locus will result in an increased probability of homozygosity after the *inter se* cross, therefore increasing inbreeding. If the population homozygous for the target locus is expected to replace

itself through mating within the population, it would be necessary to have an effective population size of around 50 breeding individuals to keep inbreeding at an acceptable level (FAO, 1998).

The expected proportion of alleles at loci that are homozygous by descent in an inbred individual is the inbreeding coefficient of that animal (*e.g.*, Franklin, 1977). In randomly mating populations, heterozygosity is lost because gamete sampling causes allele frequencies to drift towards zero or one. Such random genetic drift of gene frequencies is of negligible importance in very large populations but the level of homozygosity can be inflated by consanguineous matings (Falconer and Mackay, 1996). The loss of diversity in ancestral alleles during backcrossing increases the potential for homozygosity in individuals homozygous for the target allele post introgression. The reduction of variation can affect performance but more importantly, reduces the evolutionary potential of the final population. This minimises the possibility of the individuals homozygous for the target locus to be bred to a wider population. To avoid this requires careful management and planning and some prediction of the potential loss of variation (Allendorf, 1986).

Many studies have focused on the loss of variation in closed breeding systems by quantifying the rate of loss of genic heterozygosity at a locus with a defined population size (*e.g.*, Wright, 1969; Nei, 1973). The loss occurs at a rate of $1/(2N_e)$ per generation (N_e = effective population size). This does not give a measure of the actual number of different ancestral genotypes in the population. The rate of loss of alleles is less known but could be used as another measure of genetic variation.

Under selection, alleles present in one generation may become more or less frequent in subsequent generations. This will occur at and around the target locus. In addition to this, loss will also occur at neutral loci in the genome due to random genetic drift. The loss of allelic variants can be used as a measure of the loss in variation because of the breeding programme (Lacy, 1989). The loss in the allelic variants of ancestral alleles during backcrossing will increase the potential for homozygosity, especially directly around the gene/s undergoing selection.

1.7.2. Issues in the management of genetic resources

It is necessary to maintain the genetic diversity within a species to allow future adaptation to answer new breeding goals (Notter, 1999). Potential future breeding objectives, which may have to be addressed by the livestock industry, may include welfare, health and preference for a more extensive based production system. Given the uniform nature of many production systems in developed countries, the focus for management of genetic resources may be in developing countries where most of the worlds livestock breeds are found (Mason and Crawford, 1993)

Backcrossing can be used in the reconstitution of a breed (from semen, ova and/or embryos) in a grading-up process with DNA markers being used to trace founder population genome and traits. This method could be used to maintain inbred experimental lines, indigenous species or near extinct breeds in a gene bank. These lines can be recreated by using the germ line supplies in a backcross programme with a compatible mating breed.

The loss of alleles is an important consideration in conservation schemes or captive breeding schemes. Reconstituting a breed from cryoconserved semen can follow a similar backcrossing scheme to introgression to allow the "grading-up" to a new breed. Genome banks can be created from which breeders can draw in the future, allow for the re-establishment of an extinct breed or new breed development to fit new production circumstances. It is important to maintain genetic variation by maximising the allelic diversity in the gene bank.

1.8. Theoretical considerations on gene introgression

1.8.1. Prediction of linkage drag length in backcrossing generations

The following section deals with the theoretical expectations of the average quantity of donor genome remaining after a certain number of generations of backcrossing. The estimation of expect length of the linkage drag segment and its variance are given using Hanson's method (1959) and cited by Stam and Zeven (1981).

Consider a marker position l from the distal end of a chromosome. The probability density function (p.d.f.) of the length of chromosome to one side of the marker site still intact after t generations of backcrossing can be expressed as:

$$f(x,t) = te^{-tx}, 0 < x < l \qquad 1.1.$$

with a probability condensation of e^{-tl} at $x = l$ corresponding to no break in the segment of length l. The mean length of the segment of donor genome to the side of the marker is therefore:

$$\int_0^l xf(x,l)\mathrm{dx} + le^{-tl} = \frac{1}{t}(1-e^{-tl}) \qquad 1.2.$$

For a marker position in the centre of the chromosome of length L Morgans (that is l = ½L), there are two sides to the marker locus and attached length of chromosome to either side of the central marker (linkage drag). This linkage drag can be expressed as a function of tL by:

$$\mathrm{E}\left(\frac{\mathrm{x}}{2},t,L\right) = \frac{2}{tL}\left(1-e^{-\frac{tL}{2}}\right) \qquad \text{(Hanson, 1959)} \qquad 1.3.$$

Table 1.2 shows the estimations of the linkage drag length as calculated by Stam and Zeven (1981). It can be seen that when tL is small (*i.e.*, short chromosome during the first few generations of backcrossing) the linkage drag will be large and cover a large proportion of the carrier chromosome. However, the linkage drag will be much smaller when the chromosome is large and/or after many backcross generations.

TABLE 1.2. Mean length (μ) and standard deviation (σ), expressed as a proportion of the chromosome length and mean length in cM (μL) of donor segments surrounding a central marker locus after t generations of backcrossing.*

t	Expectation* L=50cM			Expectation* L=100cM			Expectation* L=200cM		
	μ	μL	σ	μ	μ L	σ	μ	μ L	σ
1	0.885	44	0.181	0.787	79	0.226	0.632	126	0.255
2	0.787	39	0.226	0.632	63	0.255	0.432	86	0.230
3	0.704	35	0.247	0.518	52	0.250	0.317	63	0.196
4	0.632	32	0.255	0.432	43	0.230	0.245	49	0.163
5	0.571	29	0.255	0.367	37	0.215	0.199	40	0.136
6	0.518	26	0.250	0.317	32	0.196	0.166	33	0.116
8	0.432	22	0.230	0.245	25	0.163	0.125	25	0.088
10	0.367	18	0.215	0.199	20	0.136	0.100	20	0.071

*Stam and Zeven, 1981.

1.8.2. Genetic composition in backcrossing programmes

Assuming no linkage between alleles the proportion of donor genome remaining in a backcross population will reduce by a half with every generation of backcrossing (see Figure 1.2). Hill (1993) gave a formula for the estimation of the variance of this contribution of donor genome after n generations of backcrossing in the absence of selection. This was given as:

$$\mathrm{var}(\overline{Z}_{(n)}) = (1/2L^2)(1/4^n)\sum_{i=1}^{n}\binom{n}{i}(1/i^2)\left(2iL - v + \sum_{j=1}^{v} e^{-2il_j}\right)$$

where: $\mathrm{var}(\overline{Z}_{(n)}) =$ variance of contribution of the nonrecurrent population derived from the crossbred parent

$v =$ number of chromosomes

$l =$ length of chromosome v, where $\sum l = L$ (in Morgans)

This formula can be used to estimate the mean and variance of contributions of ancestors to the non-carrier chromosome (*i.e.*, ignore the carrier chromosome in L). The formula above is the same as Stam and Zeven (1981) derived to estimate the variance of the estimated mean proportion of linkage drag after a number of backcross generations for a randomly placed marker. The example given represents 10 chromosomes with chromosome length (l) of 1.5, 1 and 0.5 Morgans and the mean contribution and the standard deviation of contribution from the donor individuals is given in Table 1.3. The genomic contributions from recipient population will decline in the same manner as the donor contribution.

TABLE 1.3. Mean and Standard Deviation (SD) of contribution of nonrecurrent parent (donor) to a backcross population*

Cross	F_1	BC_1	BC_2	BC_3	BC_4	BC_5
Expected Mean	0.5	0.25	0.125	0.0625	0.0312	0.0156
Expected SD	0	0.0351	0.0286	0.0204	0.0140	0.0094

* Hill, 1993

1.9. Conclusions

This chapter described a number of studies and issues important to introgression of alleles. The main aim of this thesis is to examine the effectiveness and efficiency of the introgression of alleles particularly into livestock populations.

A number of studies have examined the efficiency of marker-assisted introgression to maximise the contribution of recipient alleles to the genome of the introgression population. These studies never considered the ancestral source of the recipient genome. If the majority of the recipient alleles selected trace back to earlier backcross recipient ancestors, the introgression population will lag behind the current population for commercial traits. This will be of particular importance in livestock populations where long generation intervals are common and where the rate of genetic improvement is high. Chapter Two aims to develop predictions of the genomic contributions from each of the ancestral populations used in the introgression scheme. These predictions and predictions developed by other authors for linkage drag (Hanson, 1959; Stam and Zeven, 1981) will be verified by simulation studies. Such predictions will provide a useful framework to determine the genetic worth of the final population in Chapter Six.

The loss of genetic variation has been shown to be an important consideration in terms of the viability of the breeding population, particularly in livestock introgression programmes. Many of the simplistic predictions of the loss of alleles in a population do not account for selection solely on a target locus during backcrossing which will inevitably lead to a reduction in the variation at and around the target locus. Chapter Three will attempt to describe the loss of allelic variation at neutral and target loci. The predictions developed will be validated by simulation studies.

Inbreeding is also an important consideration in gene introgression. Ranking of individuals based on foreground or background selection, leads to co-selecting sib groups (van der Beek and van Arendonk, 1996). The levels of inbreeding in the recipient population that the gene is being introduced into will also affect the efficiency of an introgression scheme. The type of selection (foreground vs. background, phenotypic vs. genotypic vs. combination) that is used is important, but in all schemes, the marker allele of interest has to be closely monitored as well as the levels of inbreeding within the scheme. Chapters Four and Five will examine the increase in homozygosity in an introgression programme, by describing homozygosity due to alleles at a locus derived from the same ancestral population and alleles at a locus deriving from the same individual.

The scale and design of livestock introgression programme will be restricted by the size of the population that can be maintained during the backcross phase and the reproductive capacity of the species. These factors will be a major limitation to the scope for selection, even with the application of reproductive techniques. The long

generation intervals of many livestock species will also limit the design of the scheme because the population homozygous for the target locus may be required within a certain timeframe. This will limit the potential for reducing the donor genome through backcrossing. The effect of population structure and the number of backcross generations applied will be studied throughout this thesis, for genomic contributions, allelic variation and homozygosity. The results can be used in the design of introgression schemes where the goal may be to minimise the loss of variation or maximise recipient contributions.

Livestock introgression may be more difficult to carry out successfully when compared to plant population because genetic contributions, variation and inbreeding may have to be maintained within certain limits. The expense and timescale is likely to be greater and the final product will require different properties in the final population homozygous for the target allele (*i.e.*, high levels of inbreeding would be a problem). Therefore, the study of the effect of introgression scheme design on the loss of variation, inbreeding and genomic contributions should be studied before introgression is considered a viable option.

CHAPTER TWO

THEORETICAL CONSIDERATIONS ON GENOMIC CONTRIBUTIONS IN GENE INTROGRESSION PROGRAMMES

2.1. Introduction

There is a wealth of genetic diversity among breeds and lines of livestock and it is reasonable to assume that some alleles have become fixed in populations before artificial selection was introduced. Commercial lines or breeds are unlikely to contain all the best alleles for traits considered of economic importance (e.g., Tanksley and McCouch, 1997). Developments in molecular genetics have led to the uncovering of individual genes or regions of the genome that have an effect on traits of interest that may wish to be utilised in commercial livestock lines; e.g., the halothane sensitivity locus in pigs (disease resistance, Hanset et al, 1995); the double muscling gene in cattle (meat yield, Charlier et al, 1995) and callipyge gene in sheep (meat yield, Cockett et al, 1994); the RN gene in pigs (meat quality, Milan et al, 1995); polled gene in cattle (welfare, Brenneman et al, 1996) and the estrogen receptor locus in pigs (reproduction, Rothschild et al, 1996). A relevant point of interest in the latter example is that the beneficial allele increasing fecundity is found at much higher frequency in a non-commercial line of pigs (Meishan). Many of these examples refer to recent mutations in livestock species, some of which are harmful mutations and care must be taken when they are harnessed for commecial use.

Gene introgression can be used as a tool for genetic improvement by the introduction of new alleles into a population to address challenges facing current breeding goals (Hospital *et al*, 1997). Having detected an allele of interest from a non-commercial (donor) line, the aim of introgression is to fix that allele into a commercial (recipient) population whilst minimising the contribution of the donor genome, thereby

minimising the loss of beneficial alleles from the commercial population. Introgression involves

(i) a number of generations of backcrossing of individuals carrying the desired allele to the recipient breed to obtain further heterozygotes that have an increasingly higher proportion of the recipient breed genome; followed by

(ii) an *inter se* cross among those heterozygotes to breed individuals that are homozygous for the desired allele (*e.g.*, Groen and Timmermans, 1992; Hillel *et al*, 1990; Hospital *et al*, 1992; Visscher *et al*, 1996). This technique has become a more realistic option with the advent of DNA markers to track the alleles that derive from donor and recipient breeds.

In the course of gene introgression, many donor alleles linked to the desired allele are incorporated into the genome of the recipient line by a phenomenon called linkage drag (Brinkman and Frey, 1977). The linkage drag segment is important as it may incorporate other less favourable alleles and drag them into the commercial population and the risk of this is related to its length. Several authors (*e.g.*, Hanson, 1959; Naveira and Barbadilla, 1992; Stam and Zeven, 1981) have examined the prediction of the expected length of the linkage drag in backcross breeding programmes. Stam and Zeven (1981), developing work of Hanson (1959), showed that linkage drag can be large, *e.g.*, a segment length of 32 cM, for a 100 cM chromosome after six generations of backcrossing. This theoretical work was verified by several practical examples (*e.g.*, Zeven *et al*, 1983; Brown *et al*, 1989).

Marker information associated with the desired allele (*e.g.* flanking markers) and markers specific to the recipient line can be used with Marker Assisted Selection (MAS) protocols to minimise the donor contamination in the genome during backcross phase (Franklin, 1999; Groen and Smith, 1995; Hospital *et al*, 1992). This possibility is most powerful in plant populations where the offspring group size is large, thereby offering considerable selection opportunity. Successive backcross generations can be carried out over relatively short periods. The study of Hospital *et al* (1992) showed that 98.5 % of the recipient genome could be recovered in four generations of backcrossing when using MAS to speed the recovery of recipient genome on non-carrier chromosomes during introgression (compared to six generations without using MAS). The study of Frisch and Melchinger (2001) also examined the use of MAS to reduce the length of the linkage drag.

There are important differences between plant and livestock introgression programmes revolving around the number of progeny and generation interval.

(i) The selection intensities during a livestock introgression programme are lower than those achieved in plant breeding. This severely limits the selection of favourable recombinants at flanking markers, particularly if alleles at multiple loci are to be introgressed. These limits arise from the practical constraints on the size of the introgression population and/or the biological constraints on offspring group size.

(ii) Plant breeders would tend to select a single individual with the most favourable set of recombinants to continue using for backcrossing. Livestock introgression programmes may wish to continue using all carriers (especially in species with

small offspring group size) and end up with a viable breeding population at the end of introgression. Threats to viability arise from genetic bottlenecks in both the donor contributions and in the *inter se* cross (which will form a new inter breed population). Considerations of the parental and offspring numbers change the statistical properties of some of the parameters (*e.g.*, obligate drag which can be described as the minimum of the distribution of linkage drag).

(iii) The recovery of the recipient breed genome and/or the reduction in the linkage drag length does not remove all ancestral genome that will contribute to genetic lag. This is of particular importance to livestock introgression programmes because the generation interval is much larger than that of plants. The commercial populations will make significant rates of genetic change in the time it takes for introgression. Having predictions of the recipient genomic contributions from different generations of the introgression scheme is therefore important to allow a breeder to optimise the design of the programme.

The purpose of this present study was to investigate the genomic state of the carrier and non-carrier chromosomes after introgression of a desired allele through to the *inter se* cross through analytical methods and simulation studies. Aspects of genomic status considered are genetic lag and its components, including the predictions of the recipient genomic contributions.

The parameters considered are

(i) the number of backcrosses,

(ii) the length of the carrier chromosome and total genome, and

(iii) the structure of the populations, *i.e.*, numbers of parents and progeny.

The results derived are validated by simulation studies.

2.2. Materials and Methods

2.2.1. Populations, Structures and Notation

The introgression of a marker for a desired allele at a target locus, s Morgans from the chromosome end, is performed by crossing donor and recipient individuals to create F_1 individuals born at time 0, followed by T generations of backcrossing (creating individuals born at times 1 to T), and an *inter se* cross born at time T+1. D refers to individuals of the donor breed used to initiate the introgression programme and R refers to individuals of the recipient breed used as parents at each generation of backcrossing. BC refers to the backcross heterozygous offspring, subsequently used as parents, and IC refers to the offspring of the *inter se* cross that are homozygous for the desired allele. This is summarised in Table 2.1.

TABLE 2.1. Design of an introgression of a gene from a donor breed (D) into a recipient population (R) with an F_1, T generations of backcrossing (BC) followed by an *inter se* cross (IC), with a description of the ancestral origins of the alleles.

Gen^n	Cross	Progeny	Description	Contributions
0	$R_0 \times D$	F_1	Recipients crossed with donors to create F_1	D and R_0
1	$R_1 \times F_1$	BC_1	F_1 individuals backcrossed to recipients	D, R_0 and R_1
2	$R_2 \times BC_1$	BC_2	BC_1 individuals backcrossed to recipients	D, R_0 $R_1$and R_2
.....	.....	.....	.....	.....
T	$R_T \times BC_{T-1}$	BC_T	BC_{T-1} individuals backcrossed to recipients	D, R_0, R_1,...R_{T-1}, R_T
$T+1$	$BC_T \times BC_T$	IC	BC_T individuals crossed for the *inter se* cross	D, R_0, R_1,...R_{T-1}, R_T

Subscript t is used to denote time, so R_t refers to the group of recipient breed parents used to produce BC_t offspring at time t (t = 1 to T). The special case, R_0, is used to denote parents of the recipient breed used to produce the F_1 cross. The lengths of the carrier chromosome and total genome are l and L Morgans, respectively. The total length of all non-carrier chromosomes in the genome is therefore given by (L - l). In the introgression programme the number of mating pairs at each backcross generation is N with n offspring per mating (*i.e.*, N D × N R_0 produce Nn F_1 offspring). All carrier individuals at the end of backcrossing are used for the *inter se* cross.

π(X) is the proportion of total alleles in IC from population group X, where X can be the donor (D), recipient groups R_0, R_T or offspring group F_1, BC_1 to BC_T. The subscripts C or NC denote carrier or non-carrier chromosomes respectively. The subscripts p and d refer to a description of genomic contributions to the proximal or distal side of the target locus s respectively. To develop predictions the total genomic contributions of population groups will be described by separately considering proximal contributions (with a chromosome segment length $s_p = s$) and distal contributions (with a chromosome segment length $s_d = l - s$). These two sections are then summed to describe the carrier chromosome in terms of contributions from each ancestral population. A list of notation is given in Table 2.2.

TABLE 2.2. Summary of main symbols and notation used in the text.

General terms	
T	Number of backcross generations in the introgression
l	Length of the carrier chromosome
L	Length of total genome
N	Number of mating pairs at each backcross generation
n	Number of offspring per mating
s	Location of the target locus on the carrier chromosome
t	Time variable
$\pi(X)$	Contribution of population X to the genome of IC
Population group terms	
F_1	Offspring of the initial cross between the donor and recipient breed.
BC	Backcross heterozygous offspring
IC	*Inter se* cross offspring homozygous for the target allele
D	Donor individuals
R	Recipient individuals used in F_1 cross (R_0) and each backcross generation (R_1, R_T)
Subscript terms	
NC	Denotes a non-carrier chromosome
C	Denotes a carrier chromosome
E	Entire genome (aggregate of NC and C)
d	Distal to the target locus s
p	Proximal the target locus s
Genetic lag terms	
G_D	Difference in background genotype between recipient and donor populations
ΔG	Genetic gain per generation in the commercial population
ΔI	Genetic lag in introgression population for commercial traits.

2.2.2. Theoretical considerations on genomic contributions

Haldane's mapping function (1919) is used which assumes no interference in crossing-over events. It is also assumed that loci are uniformly distributed over the chromosome map, (i.e., segments of equal length will contain an equal number of loci). To extrapolate the genomic contributions to IC individuals it is necessary to note that the *inter se* cross may be treated as one additional generation of backcrossing since recombinations in the parents are with the recipient genome (this is not so for any further generations).

2.2.2.1. Carrier chromosome

Consider the proportion of the carrier chromosome in IC that derives from R_1, *i.e.*, $\pi_C(R_1)$, and F_1, *i.e.*, $\pi_C(F_1)$ for a scheme with a single backcross (Figure 2.1.). These contributions will depend on s_p and s_d but for the present we will consider only the segment of the chromosome proximal to the target locus, *i.e.*, s_p, and the subscript will be dropped to derive the following theory.

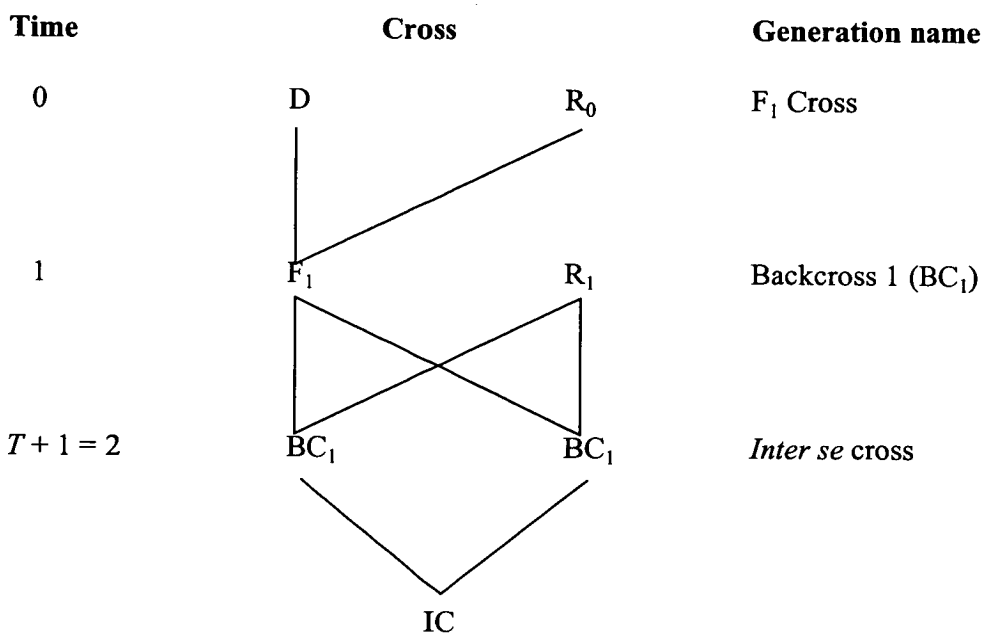

FIGURE 2.1. Diagram of an introgression scheme with one backcross generation showing population groups, generation names and time period.

Stage 1 in Figure 2.2. illustrates the derivation of $\pi_C(R_1)$ and $\pi_C(F_1)$ while Stage 2 illustrates the separation of $\pi_C(F_1)$ into $\pi_C(R_0)$ and $\pi_C(D)$. In Stage 1 $\gamma[s]$ and $\delta(s,1)$ are the total proximal length of F_1 genome and the proximal linkage drag, with $\varepsilon(s,1) = \gamma[s] - \delta(s,1)$ the residual F_1 genome outside the linkage drag. The expectation of the length of the linkage drag segment after an F_1 and t generations of backcrossing for a locus in position s is termed $\delta(s,t)$. Following Hanson (1959), the linkage drag proximal to s is given by:

$$\int_0^s e^{-tu}\,du = t^{-1}\left(1 - e^{-ts}\right) \qquad \text{(Stam and Zeven, 1981)} \qquad (2.1.)$$

Analogously the linkage drag distal to s is given by $t^{-1}(1 - e^{-t(l-s)})$.

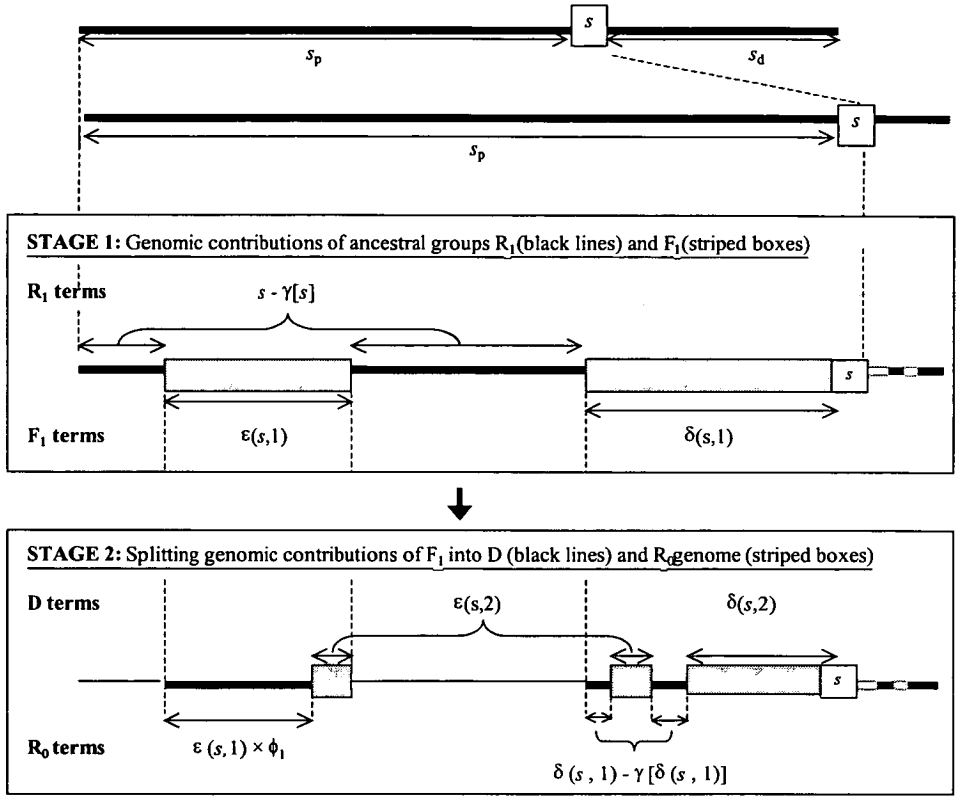

FIGURE 2.2. Diagram illustrating the derivation of ancestral contributions proximal to a target locus at positions *s*.

The total length of genome co-inherited with a locus situated at the end of a genome fragment of length s is denoted $\gamma[s]$. Consider a target locus at position s and another locus at position u, $0<u<s$. The probability that the allele transmitted at u is from the same parental gamete as the target locus is equal to the probability of an even number of recombinations, *i.e.*, $\frac{1}{2}(1 + e^{-2(s-u)})$ (Hanson, 1959). Therefore the total length co-inherited proximal to s is given by:

$$\int_0^s \tfrac{1}{2}(1 + e^{-2(s-u)})du = \tfrac{1}{2}s + \tfrac{1}{4}(1 - e^{-2s}) = \gamma[s] \qquad (2.2)$$

Analogously the total length co-inherited distal to s is given by $\frac{1}{2}(l-s) + \frac{1}{4}(1 - e^{-2(l-s)})$.

The principle behind the apportionment of the genomic contributions in Stage 2 follows the schema:

For the linkage drag segment $\delta(s,1)$, the total length of contribution of D is given by $\gamma[\delta(s,1)]$, with a linkage drag $\delta(s,2)$.

In the segment of length $\varepsilon(s,1)$ the contribution of R_0 to F_1 is $\gamma[s] - \gamma[\delta(s,1)]$ with the remainder originating from D. Therefore the approximate proportion of $\varepsilon(s,1)$ contributed by R_0 is

$$\phi_1 \approx (\gamma[s] - \gamma[\delta(s,1)]) / (s - \delta(s,1)) \qquad (2.3.)$$

Therefore bringing these contributions together to predict the genomic contributions of the populations D, R_0 and R_1 (as in Figure 2.2.) we have:

$$\pi_C(R_1) = l - \sum_{s_p,s_d} \gamma[s]$$

$$\pi_C(R_0) = \sum_{s_p,s_d} \{\varepsilon(s,1)\,\phi_1 + \delta(s,1) - \gamma[\delta(s,1)]\}$$

$$\pi_C(D) = \sum_{s_p,s_d} \{\gamma[\delta(s,1)] + \varepsilon(s,1)(1 - \phi_1)\}$$

This can be extended to further generations of backcrossing, by defining the following recursive relationships:

$$\varepsilon(s,t) = \gamma[\delta(s,t\text{-}1)] - \delta(s,t) + (1 - \phi_{t-1})\varepsilon(s,t\text{-}1) \qquad (2.4.)$$

$$\phi_t = (\gamma[s] - \gamma[\delta(s,t\text{-}1)]) / (s - \delta(s,t\text{-}1)) \qquad (2.5.)$$

Therefore, for T generations of backcrossing we have:

$$\left.\begin{aligned} \pi_C(R_T) &= l - \sum_{s_p,s_d} \gamma[s] \\ \pi_C(R_t) &= \sum_{s_p,s_d} \{\varepsilon(s,T\text{-}t)\,\phi_{T\text{-}t} + \delta(s,\,T\text{-}t) - \gamma\,[\delta(s,\,T\text{-}t)]\} \qquad (0 \le t < T) \\ \pi_C(D) &= \sum_{s_p,s_d} \{\gamma\,[\delta(s,\,T)] + \varepsilon(s,\,T)(1-\phi_T)\} \end{aligned}\right\} \quad (2.6.)$$

For parental group X, the proportional contributions to the carrier chromosome is then given by $l^{-1}\pi_C(X)$.

2.2.2.2. Non-carrier chromosomes

The length of donor genome on the non-carrier chromosomes, *i.e.*, $\pi_{NC}(D)$ is $\frac{1}{2}^{T+1}$. The contribution of R_0 ... R_T to the non-carrier chromosomes follows the same pattern and are:

$$\left.\begin{aligned} \pi_{NC}(R_T) &= \tfrac{1}{2}\,(L-l) \\ \pi_{NC}(R_{T-1}) &= \tfrac{1}{2}^2\,(L-l) \\ \pi_{NC}(R_{T-t}) &= \tfrac{1}{2}^{t+1}\,(L-l) \\ \pi_{NC}(R_0) = \pi_{NC}(D) &= \tfrac{1}{2}^{T+1}\,(L-l) \end{aligned}\right\} \quad (2.7.)$$

2.2.3. Variance of linkage drag and obligate drag

Stam and Zeven (1981) derived the variance of the linkage drag with the same assumptions and same probability density as Hanson (1959) used for the derivation of $\delta(s,\,t)$ (see Equation 2.1.). The accuracy of these predictions are checked via simulation studies.

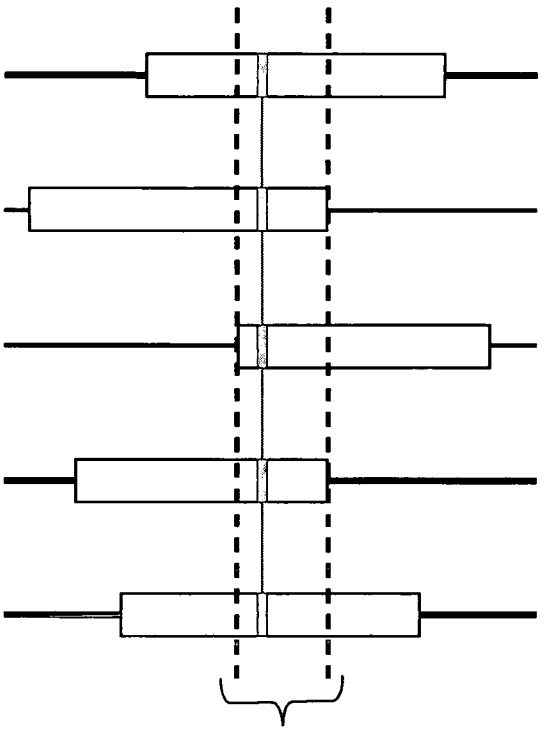

FIGURE 2.3. Diagram illustrating the obligate drag on a set of carrier chromosomes. Clear box denotes the linkage drag segment, the target locus is denoted in grey and dotted back lines denote the obligated drag.

In addition, the obligate drag, which is defined as the segment length corresponding to the closest cross-over to the target locus (Figure 2.3.), has been examined. Obligate drag can be defined as the minimum individual donor segment proximal and distal to *s*. A simple prediction of obligate drag would involve examining the probability of a recombination a certain distance from the target locus. Let $r(d)$ = the probability of the a recombination d Morgan from the target locus and m the number of informative meioses. At one side of the target locus the obligate drag (OD) can be expressed as OD = d for which $r(d) \times m = 1$. For small distances $r(d) \sim d$ (in Morgans). Therefore OD = $1/m$. For this study $m = 0.5 \times N$. This will give the

length of the obligate drag one side of the target locus in Morgans. For example, when $N = 20$ the obligate drag after one backcross generation would be $2(1/20) = 0.1$ Morgans. The validity of the prediction for obligate drag with the simulation results was checked.

2.2.4. Simulations

The initial cross for the introgression scheme was assumed to be between two divergent lines that are fixed for alternative alleles at each loci. The carrier chromosome was simulated using crossing-over events assuming Haldane's mapping function (1919). The N parents for the next generation were selected at random among the offspring heterozygous for the target locus. All heterozygous individuals were used as parents of the *inter se* cross. Only the offspring homozygous at the target locus for the donor allele were considered in summaries of the IC population.

The results examined the validity of the predictions of linkage drag and recipient genomic contributions. The following were derived from the genome of each IC individual:

(i) the proportional genomic contribution from each the ancestral population groups, $\pi_C(D)$, $\pi_C(R_t)$, $\pi_{NC}(D)$, $\pi_{NC}(R_t)$ and

(ii) the length of the linkage drag, $\delta(s,T+1)$ each side of s. From these the expected linkage drag, its variance and the obligate drag were calculated.

The population parameters used were: parental population size (N = 10, 20, 50); offspring group size, (n = 2, 3, 4, 5, 10); the length of the carrier chromosome (l =

0.5, 1, 2, 4, 8) and the total number of backcross generations (T = 3, 6, 10, 20). The location of the target locus s was varied and the major results are derived for $s = l/2$ and s = 0.1. The simulations were run for 500 replicates for each set of parameters studied.

2.3. Results

2.3.1. Validity of the theoretical predictions

2.3.1.1. Linkage drag and proportional genomic contributions

Theoretical predictions of the proportional contributions of donor and recipient genomics to the carrier chromosomes of IC, $\pi_C(D)$ and $\pi_C(R_t)$, the expectation of linkage drag and its standard deviation within replicate are compared to simulation results in Table 2.3. (for N = 20, n = 2, l = 1 Morgan, T = 6 and s = 0.1 Morgan *i.e.* the target locus is towards one end). The predictions of proportional contributions in other schemes are shown in Figure 2.4. and compared to simulation results. The predictions for all schemes studied are very accurate.

The predictions of linkage drag from this paper were accurate for a target locus in all positions whereas when the locus was not centrally placed the prediction from Hanson (1959) (assuming central position) and Stam and Zeven (1981) (assuming random position) could be subject to substantial error. For example when for the simulation with the parameters N = 20, n = 2, l = 1 Morgan, T = 6 and s = 0.1 Morgan the expected linkage drag expressed as a proportion of l was 0.21 whereas the prediction of Hanson (1959) and Stam and Zeven (1981) were 0.28 and 0.27 respectively.

TABLE 2.3. Comparisons of the simulated and predicted genomic contributions (Morgans) from ancestral population groups to the carrier chromosome in IC.*

		Simulation	Prediction
Ancestral group	$\pi(D_c)$	0.24 ± 0.003	0.23
	$\pi(R_{0\,c})$	0.033 ± 0.002	0.033
	$\pi(R_{1\,c})$	0.045 ± 0.002	0.045
	$\pi(R_{2\,c})$	0.058 ± 0.002	0.062
	$\pi(R_{3\,c})$	0.083 ± 0.002	0.087
	$\pi(R_{4\,c})$	0.118 ± 0.002	0.122
	$\pi(R_{5\,c})$	0.170 ± 0.002	0.169
	$\pi(R_{6\,c})$	0.250 ± 0.002	0.246
Linkage drag		0.220 ± 0.004	0.215
s.d. of linkage drag		0.129 ± 0.002	0.114
Obligate drag		0.014 ± 0.001	0.008

*$N = 20$, $n = 2$, $l = 1$ M, $T = 6$ and $s = 0.1$ M

Table 2.4 shows close agreement between simulation and prediction results of linkage drag in two different population structures (A: $N = 20$, $n = 2$; B: $N = 50$, $n = 10$) over different chromosome lengths (l = 0.5, 1, 2 Morgans) and backcross generations (T = 1-6, 8, 10). The standard deviations are close to expectation. There were no significant differences between the simulations A and B showing that population structure had no effect on linkage drag length. This is in agreement with Stam and Zeven (1981) that the length of the linkage drag is dependent on the initial length of the carrier chromosome and the number of backcross generations.

TABLE 2.4. Mean length of donor segments surrounding a central marker locus after *t* generations of backcrossing as a proportion of chromosome length *l* Morgans.*

t	Expectation*		*N* = 20, *n* = 2		*N* = 50, *n* = 10		L (M)
	μ	σ	μ	σ	μ	σ	
1	0.89	0.18	0.90 ± 0.002	0.18 ± 0.002	0.90 ±0.002	0.18 ±0.002	0.5
2	0.79	0.23	0.80 ± 0.003	0.22 ± 0.002	0.80 ± 0.002	0.23 ± 0.002	
3	0.70	0.25	0.72 ± 0.003	0.24 ± 0.002	0.71 ± 0.003	0.25 ± 0.002	
4	0.63	0.26	0.64 ± 0.004	0.24 ± 0.002	0.64 ± 0.004	0.26 ± 0.002	
5	0.57	0.26	0.58 ± 0.004	0.24 ± 0.002	0.58 ± 0.004	0.25 ± 0.002	
6	0.52	0.25			0.52 ± 0.004	0.25 ± 0.002	
8	0.43	0.23			0.44 ± 0.004	0.23 ± 0.002	
10	0.37	0.22			0.37 ± 0.004	0.21 ± 0.002	
1	0.79	0.23	0.79 ± 0.002	0.22 ± 0.002	0.80 ± 0.002	0.22 ± 0.001	1
2	0.63	0.26	0.64 ± 0.003	0.25 ± 0.002	0.64 ± 0.002	0.25 ± 0.001	
3	0.52	0.25	0.52 ± 0.003	0.24 ± 0.002	0.53 ± 0.003	0.25 ± 0.002	
4	0.43	0.23	0.43 ± 0.003	0.22 ± 0.002	0.44 ± 0.003	0.23 ± 0.002	
5	0.37	0.22	0.37 ± 0.003	0.20 ± 0.002	0.37 ± 0.003	0.21 ± 0.002	
6	0.32	0.20			0.32 ± 0.003	0.20 ± 0.002	
8	0.25	0.16			0.25 ± 0.002	0.16 ± 0.002	
10	0.20	0.14			0.20 ± 0.002	0.13 ± 0.001	
1	0.63	0.26	0.63 ± 0.003	0.25 ± 0.002	0.64 ± 0.002	0.25 ± 0.001	2
2	0.43	0.23	0.43 ± 0.003	0.22 ± 0.002	0.44 ± 0.002	0.23 ± 0.001	
3	0.32	0.20	0.32 ± 0.003	0.19 ± 0.002	0.32 ± 0.002	0.20 ± 0.001	
4	0.25	0.16	0.25 ± 0.002	0.16 ± 0.002	0.25 ± 0.002	0.16 ± 0.001	
5	0.20	0.14	0.20 ± 0.002	0.13 ± 0.001	0.20 ± 0.002	0.13 ± 0.001	
6	0.17	0.12			0.17 ± 0.002	0.11 ± 0.001	
8	0.13	0.09			0.13 ± 0.001	0.08 ± 0.001	
10	0.10	0.07			0.10 ± 0.001	0.07 ± 0.001	

*Stam and Zeven, 1981.

Figure 2.4 also shows that the proportional contributions from the recipient ancestral groups also change significantly when the target locus is non-central compared to

when it is centrally placed. These changes were correctly predicted using Equations 2.6.

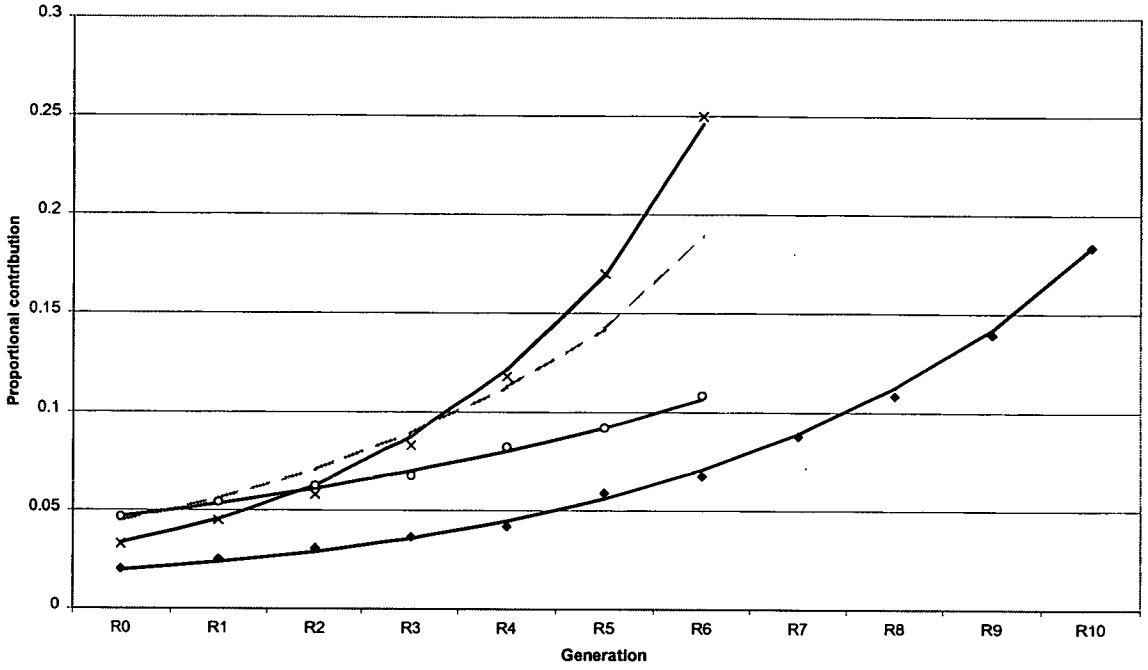

FIGURE 2.4. Comparison of simulated and predicted results $\pi_C(X)$ expressed as a proportion of l for a variety of schemes. Symbols represent simulations and lines represent predictions. (♦) $T = 10$, $l = 1$, $s = 0.5$ M; (○) $T = 6$, $l = 0.5$, $s = 0.25$ M and (×) $T = 6$, $l = 1$, $s = 0.1$ M. The broken line indicates predictions for (×, $T = 6$, $l = 1$) when $s = 0.5$ M.

2.3.1.2. Obligate drag

It can be seen in Table 2.3 that the prediction of obligate drag is close to the simulation result when $n = 2$ and 4 but only in early backcross generations. As a general result varying N to 10 or 50 made no difference to this result. In summary, the prediction for obligate drag is approximately correct but only when $t > 3$.

TABLE 2.3. Comparison of simulated and predicted results for the obligate drag length (expressed in cM) expressed as a proportion of the carrier chromosome over a number of backcross generations and population structures when $l = 1$ and $s = 0.5$ Morgans.

Population structure	Population group	Simulation	Prediction
$N = 20, n = 2$	BC_1	10.5	10
	BC_2	5.5	5
	BC_3	5.4	3.3
	BC_4	4.4	2.5
	BC_5	3.9	2
	BC_6	2.3	1.7
$N = 20, n = 4$	BC_1	9.1	10
	BC_2	5.9	5
	BC_3	4.3	3.3
	BC_4	3.8	2.5
	BC_5	3.5	2
	BC_6	2.4	1.7

2.3.2. Effects of carrier chromosome length and position of target locus on proportional genomic contributions

The predictions were used to explore the impact of carrier chromosome length (l = 0.5, 1, 2, 4, 8, $T = 6$) and position of target locus genomic contributions. Figure 2.5 shows that as l increases the proportional genomic contributions of the donor to the carrier chromosome decreases. This decline is mirrored by the early backcross generations and compensated for by increases in total contributions from later R_t generations. For example, when s is centrally placed and l = 0.5 Morgan $\pi_C(D) \sim$ 50% and $\pi_C(R_6) \sim 10\%$ but when $l = 8$ Morgans $\pi_C(D) \sim 5\%$ and $\pi_C(R_6) \sim 44\%$.

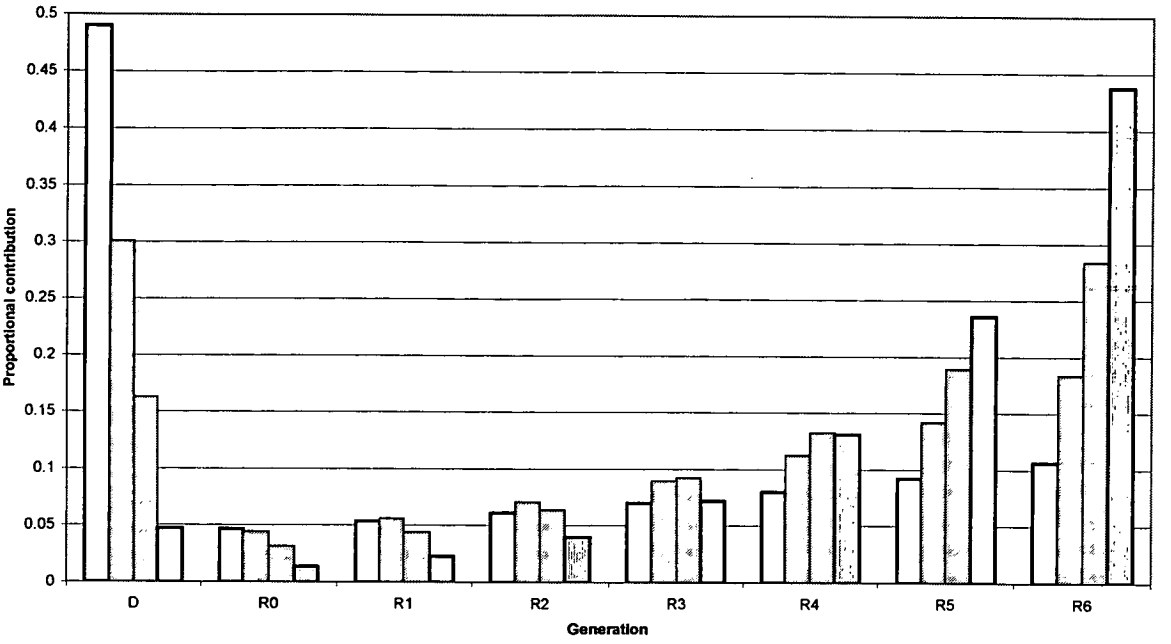

FIGURE 2.5. Proportional genomic contributions from ancestral population groups to the carrier chromosome of IC as a function of *l* M. □=0.5, □=1, □=2, □=4

Whilst the proportional linkage drag length changes dramatically for the different chromosome lengths studied the actual length is relatively constant, with this example (T = 6, N = 20, n = 2) the linkage drag is in fact over a very narrow range and approximately 28 cM except when l = 0.5 when linkage drag is 24 cM.

As the location of the target locus on the carrier chromosome approaches the chromosome end the donor contribution decreases primarily due to a decrease in linkage drag (Figure 2.6) compensated for by an increase in $\pi_C(R_t)$. There is little change in the linkage drag away from a centrally placed target on larger chromosomes until the target locus approached the very edge of the chromosome in which case the amount of donor genome decreased dramatically. For example, Figure 2.6 shows that when t = 6, only when s = 8 cM or less does the linkage drag

differ markedly for a centrally placed target (approximately 20 cM compared to 28 cM) and this is relatively insensitive to *l*.

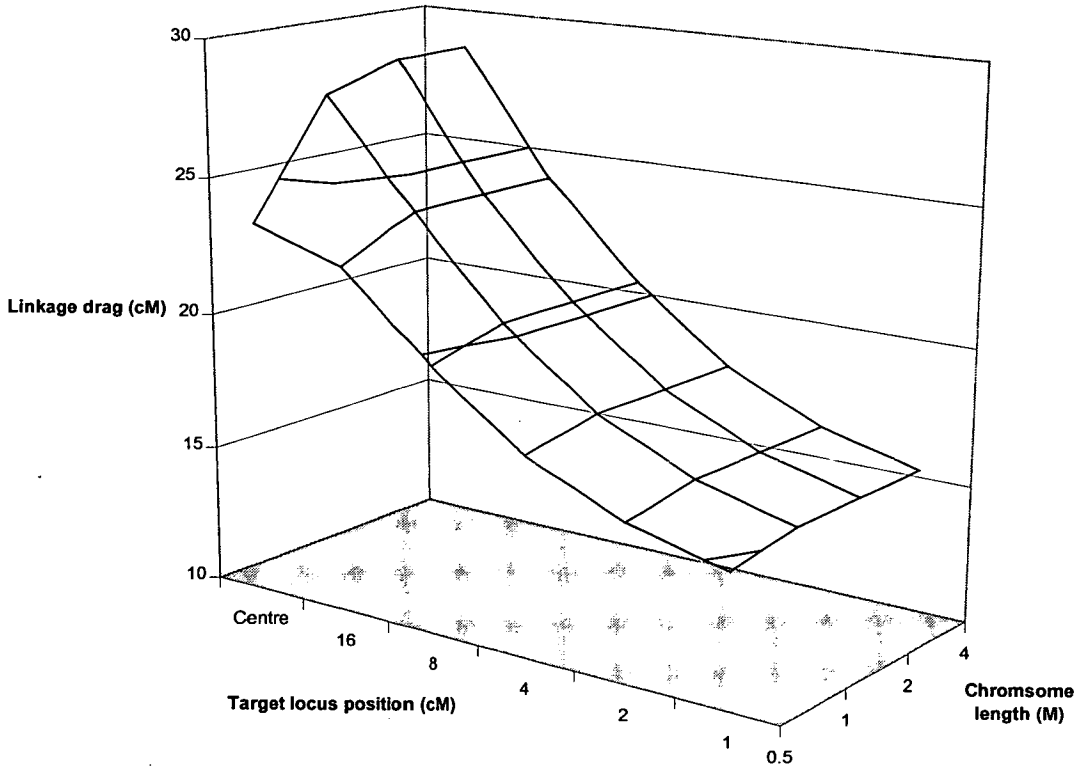

FIGURE 2.6. Length of linkage drag in IC as a function of carrier chromosome length (*l* M) and target locus position (*s* M) after six backcross generations.

2.3.3. Effect of number of backcross generations on proportional genomic contributions

The proportional donor contributions are highest when the number of backcross generations (*T*) is small (Figure 2.7 with $l = 1$ Morgan and $s = 0.5$ Morgan). From Equations 2.4 the proportional contribution from the last backcross generation is independent of *T*. The recursive nature of Equations 2.4 predicting ancestral recipient groups show that their contributions is a function of the number of backcross generations that have been performed since the introduction of the said

ancestral recipient group (for constant T, s and l). This result was confirmed with simulation results that showed the same trend. The behaviour of the linkage drag as a proportion of l can be derived from Equation 2.1 and will decline with T for all s, which is in agreement with the predictions given by Hanson (1951) for a centrally placed target locus.

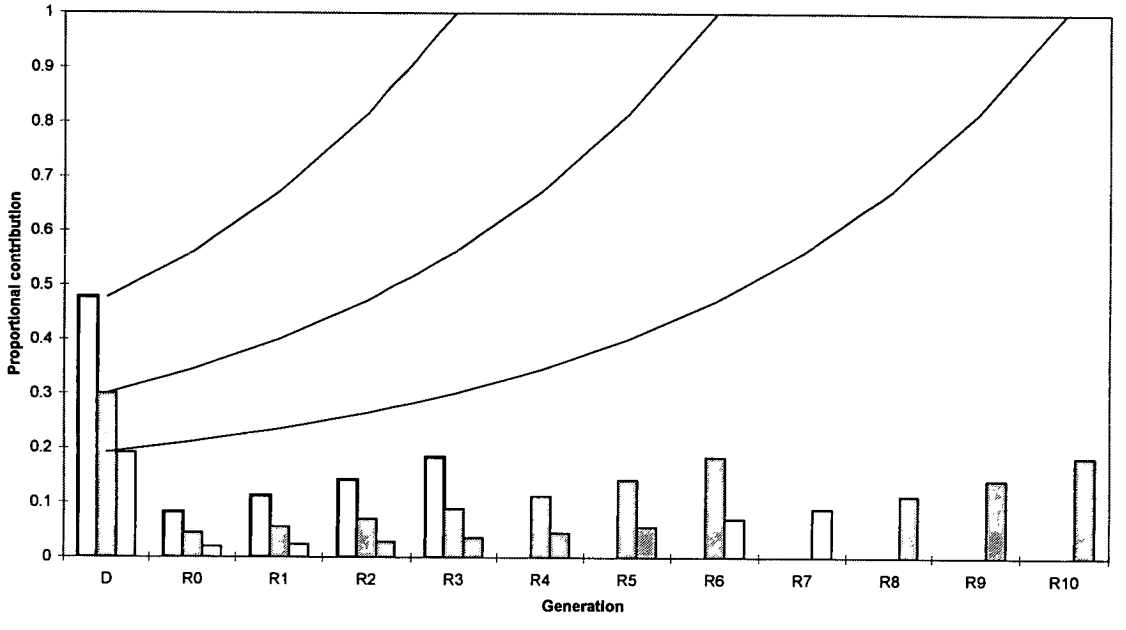

FIGURE 2.7 Proportional genomic contributions from ancestral population groups to the carrier chromosome of IC as a function of T. Lines indicate the cumulative values of genomic contributions over the generations. $T = 3$, $T = 6$, $T = 10$

2.4. Discussion

Using analytical methods and simulation studies the genomic contributions of individuals used at each generation to the genome of individuals after a gene introgression programme was quantified. The prediction of donor and recipient individuals genomic contributions agreed closely with the simulated results for all

population structures (N and n), number of backcross generations (T) and length of carrier chromosome (l) studied thereby validating the predictions derived by this study. It was shown that carrier chromosome length influenced the proportional linkage drag. These influences were quantified with a high degree of precision for all ancestral group contributions and a prediction of genetic lag was developed from them.

The linkage drag and predictions of the linkage drag have been well studied by many authors (*e.g.*, Frisch & Melchinger, 2001, Hanson, 1959, Naveira & Barbadilla, 1992, Stam & Zeven, 1981). The predictions developed are based on the distribution used by Hanson (1959) and are shown to agree with Hanson's method when considering his condition of a centrally placed target locus. The prediction developed in this study, using Equation 2.1, is also accurate for non-centrally placed target loci. The importance of the linkage drag is that undesirable alleles, associated with the donor breed located in this region, may be "dragged" into the recipient breed. The drag is a function of T and l and so since the length of the carrier chromosome cannot be changed the one method of control over this risk available is the number of backcross generations required. The introgression programme can therefore be designed so that the expected length of the linkage drag segment does not contain any known deleterious or undesirable donor alleles. If suitable flanking markers are available, another method to control this risk (not examined here) is the selection of favourable recombinants minimising the expected length and variation of the linkage drag segment (Hospital & Charcosset, 1997, Hospital *et al*, 1992).

The obligate drag is the portion of the donor genome on the carrier chromosome in IC that would never be removed by selection for recombinants. Simple predictions for the obligate drag agreed when t was low. The prediction failing after a couple of backcross generations makes sense considering the pattern of decline of the linkage drag segment. The decline of the length of the linkage drag segment is exponential and slows down over backcross generations. However, the prediction used in this paper assumed a linear decline in obligate drag. It is safe to assume that obligate drag would follow a similar pattern. The prediction of this would require the derivation of the minimum function of the distribution of the linkage drag segment, which would be very complicated. Information on obligate drag can be used to aid the creation of congenics lines. At present a large number of backcross generations are carried out to reduce the donor segment of interest to a small enough region. Using the "supersonic congenics", methodology described by Behringer (1998) would increase offspring group size and allow for a large number of informative meioses in earlier backcross generations. This would reduce the linkage drag segment and the obligate drag dramatically therefore reducing the timescale of the operation.

As with the linkage drag the number of backcross generations and the length of the carrier chromosome also affect the recipient contributions to the carrier chromosome. Figures 2.5 and 2.7 shows how the genomic contribution from individuals used in the F_1 cross to the carrier chromosome decreases as T increases and l decreases. These trends are vital in the design of livestock introgression schemes as the proportion of recipient genome from earlier backcross generations also adds to the genetic lag.

These predicted contributions from the different ancestral population groups can be used for improving the predictions of genetic lag and other parameters such as identity by descent. Using background selection can increase the recovery of the recipient genome either by minimising the linkage drag using flanking markers or selecting for recipient alleles on non-carrier chromosomes (Frisch & Melchinger, 2001; Hospital & Chacosset, 1997; Visscher *et al*, 1996). These equations give an accurate prediction of the recipient contributions to the carrier chromosome. This could help to optimise the type of background selection carried out during introgression.

Population parameters may not be as easy to vary in some species as described in this paper and this is particularly so in livestock. The obligate drag was shown to decrease slightly for larger offspring group sizes ($n = 4$), which is good for livestock species with large offspring group size such as pigs or chickens. Selection of reduced linkage drag segment lengths from carriers in these populations would reduce the obligate drag further as there would be a higher number of informative meioses (*i.e.*, $m = 0.5 \times t \times N \times n$).

This study can be used to help prediction the contributions in livestock introgression programmes, in the creation of congenic strains or in backcross mapping populations. These predictions can than be used as a method of design of the scheme or to help in subsequent genetic analysis.

CHAPTER THREE

THE LOSS OF ALLELES IN BACKCROSS BREEDING PROGRAMMES

3.1. Introduction

Livestock genetic resource management is important in the maintenance of genetic diversity within a species and allows future adaptation of population to answer new breeding goals. Breeding goals of the future would probably include welfare, health and a switch to more extensive based production systems. The genetic diversity of livestock species needs to be maintained to deal with the many requirements of the range production systems in the world (Notter, 1999). The uniform nature of many production systems in developed countries and intense selection for production traits has resulted in low genetic diversity (Hammond, 1998; Patterson, 2000). An example of this is the low effective population size in the world's dairy population due to the domination in a small number of elite sires of the AI market (Goddard, 1992). Therefore, the focus for management of genetic resources may be based in developing countries where most of the world's livestock breeds are found (Mason and Crawford, 1993)

The success and appeal of various programmes to answer to breeding objectives and maintain diversity has increased with the advent of DNA markers. Markers can be used to help keep track of various parts of the genome thus allowing selection for optimum contributions from all or certain ancestors. An example of this is gene introgression, where repeated backcrossing to a line or individual is used to transfer a particular gene from one line (donor) on to the genetic background of another (recipient) in plant or animal species. Backcrossing can also be used in the reconstitution of a breed from germplasm (semen, ova and/or embryos) with DNA markers being used to trace particular founder population genomes and traits.

Storage of inbred experimental lines, indigenous species or near extinct breeds in gene banks can allow breeders to utilise the germplasm at a later date either by improving current commercial stock (using grading-up backcross procedures) or the creation of a new synthetic line. Utilising germplasm of wild cultivars has been used widely in plant breeding and has resulted in the creation of plant gene banks (Frey, 1998).

Livestock conservation programmes tend to be closed with a limited population size and therefore can result in a loss of genetic variation. The reduction in variation can affect performance, but more importantly reduces the evolutionary potential of the final population (Lacy, 2000). This would reduce the scope to use these populations for gene banks or to service a wider population. To avoid this reduction in variation requires careful management and planning, requiring prediction of the potential loss of variation (Allendorf, 1986). Many studies have focused on the loss of variation in closed breeding systems by studying and quantifying the loss of genic heterozygosity at a locus with a defined population size (*e.g.*, Wright, 1969; Nei, 1973). These studies show that overall loss of genic heterozygosity occurs at a rate of $1/(2N_e)$ per generation. A measure of the overall loss of heterozygosity does not indicate the actual number of alleles at a locus. The rate of loss of alleles is less well known but could be used as another measure of genetic variation in conservation programmes, and as a measure of diversity of stock contributing to gene banks or breeding programmes.

Recent developments in molecular genetics have led to the uncovering of individual genes or regions of the genome that have an effect on many qualitative and quantitative traits, both in commercial and non-commercial breeds of livestock. Selection on markers related to these traits in conservation breeding programmes (*e.g.*, *in-situ* conservation programmes, gene bank donor selection procedures, grading-up) that might utilise the association of markers with traits would result in a loss of genetic diversity. With selection, alleles present in one generation may become more or less frequent, or even extinct, in subsequent generations. This will occur at and around the selection locus and at neutral loci in the genome due to random genetic drift. The loss of allelic variants can be used as a measure of the loss of variation as a result of the breeding programme (Lacy, 1989). This loss of alleles will increase the potential for homozygosity, especially directly around the gene/s undergoing selection.

The purpose of this study was to investigate the loss of variation at neutral and selected loci in backcross breeding programmes. The study aims to quantify and derive a simple theoretical approximation of the number of ancestral alleles still contributing to neutral loci, at and around a locus of selection after a backcrossing breeding programme. An approximation is carried forward to the *inter se* cross of an introgression programme. The parameters considered are the; (*i*) number of backcrosses, (*ii*) the length of the carrier chromosome, (*iii*) the structure of the population, *i.e.*, numbers of parents and offspring per generation and (*iv*) the number of alleles present at a locus in the initial population. The results derived are validated by simulations studies.

3.2. Methods

3.2.1. General definitions and notation

A backcross programme is performed by crossing donor (D) and recipient (R) individuals to create F_1 individuals born at time 0, followed by T generations of backcrossing to available recipient individuals (creating heterozygous backcross, or BC, individuals born at time 1 to T). Selection during the backcross breeding programme is on target locus s. IC refers to the offspring of an *inter se* cross, if following the introgression design described in Chapter Two.

Subscript t is used to denote time and so R_t refers to the group of recipient breed parents used to produce BC_t offspring at time t (t = 1 to T). The special case R_0 is used to denote parents of the recipient breed used to produce the F_1 cross. The length of the carrier chromosome is l Morgan. In the backcross programme the number of mating pairs at each backcross generation is N with n offspring per mating (*i.e.*, N D $\times$ N R_0 produce Nn F_1 offspring). At the end of the backcross phase all carriers are used in an *inter se* cross. Subscript C refers to a locus on the carrier chromosome, which can be the target locus s, or x Morgan from s. Subscript NC refers to a neutral locus, assumed to be on a non-carrier chromosome and therefore unlinked to s on the carrier chromosome.

TABLE 3.1. Additional summary of symbols and notation used in the text (see Table 2.2 for design terms).

Subscript terms	
x	Distance between s and a locus on the carrier chromosome (Morgans)
Population terms	
X_t	Ancestral population introduced at the time of BC_t
Equation terms	
a_t(X)	Number of alleles remaining from group X at a locus in BC_t.
ai(X)	Number of alleles remaining from group X at a locus in IC.
a'(X)	Number of equally frequent alleles from group X at locus in X.
$\Phi(u)$	Probability generating function (p.g.f.) with the coefficient of u equal to the probability that j (j = 1, n) copies of an allele are transmitted from one backcross generation to the next, where u is a dummy variable.
$b(u)$	p.g.f. with the coefficient of u equal to the probability of the transmission of j carriers to n offspring
$m(u)$	p.g.f. with the coefficient of u equal to the probability of the selection of j familial carriers in N
A	Allelic diversity at a locus

a(X) is the number of alleles remaining at a locus (neutral locus on NC or, s or x cM from s on C) from an ancestral population group X in BC_t. X can be donor (D) or recipient groups R_0 to R_T. The initial number of alleles in population X, at their introduction during backcrossing, is described as a'(x). For example, $a_{C,x,t}$(D) is the number of donor alleles remaining at locus x Morgan away from the target locus on the carrier chromosome in BC_t. ai(X) is the number of alleles remaining at a locus from a given ancestral population (X) in IC. The main parameters used in the study are detailed in Table 2.2 and Table 3.1.

3.2.2. Loss of alleles at a neutral locus

The probability that any given ancestral allele is not transmitted from one generation can be described by the following probabilities:

(i.) Probability that an ancestral allele is **not** transmitted from parent to offspring,

(ii.) Probability that an ancestral allele is transmitted to 1, 2,... or n offspring but not selected as one of the N parents of the next generation.

The sum of these probabilities gives the probability of exclusion of any given ancestral allele from one generation to the next.

With multiple backcross generations the prediction of the loss alleles is complicated as many chains of transmission of alleles throughout the pedigree are possible which would result in the loss of an ancestral allele. This study focuses on the loss of donor alleles from a locus. Recipient alleles will be introduced at each backcross generation, however, there will only be one influx of donor alleles in an introgression programme. Donor alleles maybe be lost at each generation, resulting in a reduced diversity of donor representation, particularly around the target locus.

Each event that can lead to the loss of donor alleles can be seen as an independent and exclusive event, once the population size is large enough (if the population size is small or $N = n$ each selection of one of the N parents may have a bearing on the selection of another).

First, consider the first p.g.f. of the transmission or non-transmission of an allele being traced from backcross parents to their offspring. This is described by:

p.g.f. for P(allele transmitted to a carrier offspring) = $b(u) = (\frac{1}{2} + \frac{1}{2}u)$ 3.1.

where u is the value of the p.g.f.

Alleles may also be lost if a carrier (from the group of n offspring) are not selected as a parent of the next backcross generation. The probability of selecting k carriers from n offspring when selecting a total of N carrier parents from the Nn individuals is a random process across families since all families have equal chances of producing carriers. Therefore, this probability is equivalent to that of random selection of carrier parents and the p.g.f. of P(selecting k from n when selecting N from Nn) is:

$$m(u) = \sum_{k=0}^{n} \frac{\binom{n}{k}\binom{Nn-k}{N-k}}{\binom{Nn}{N}} u^k \text{ OR } \sum_{k=0}^{n} \left(\frac{N!n!(Nn-N)!(Nn-n)!}{Nn!k!(N-k)!(n-k)!(Nn-N-n+k)!} \right) u^k \quad 3.2.$$

The p.g.f. of the number of donor alleles at a neutral locus for one generation is then

$$\Phi_1(u) = m(b(u)) \quad 3.3.$$

This can be repeated over T backcross generations (*i.e.*, the p.g.f. over 2 generations is $\Phi_2(u) = \Phi_1(\Phi_1(u))$). The probability of losing the allele is $\Phi_2(0)$ *i.e.*, constant in the polynomial. Thus values are easily calculated from knowledge of $m(u)$ and $b(u)$ alone.

3.2.3. Loss of alleles at a particular locus on the carrier chromosome

The probability that a donor allele is lost at a locus on the carrier chromosome, due to a recombination event between it and the target locus s, is a function of the probability that there has been an odd number of recombination events between the locus and s. The p.g.f. function in Equation 3.3 can be easily adapted to include this

information. The probability that there are an odd number of crossover events, assuming no interference, is described by Haldane's mapping function (1919), $c = \frac{1}{2}(1 - e^{-2x})/2$ where x is the distance, in Morgans, between the s and the locus of interest. The probability of an even number of crossover events is simply $(1 - c)$. The p.g.f. for the probability of transmission of an allele at a locus x Morgans away from the target locus s from a parent to a selected parent of the next generation in a backcross breeding programme is:

$$b(j_x) = (c + (1 - c)j_x) \qquad 3.4.$$

where j_x is the value of the p.g.f. for a locus x Morgan away from the target locus, s. Equation 3.4 is then used in the p.g.f. in Equation 3.3 in the place of the p.g.f. function described by Equation 3.1. The prediction of the number of donor alleles at the target locus s is also described by Equation 3.4 assuming that $x = 0$.

3.2.4. Loss of alleles from BC_T to IC

The aim of many livestock introgression programmes would be to create a large enough pool of homozygous animals that could be bred to a larger commercial population or that could breed effectively in a closed population with an acceptable rate of inbreeding. The Food and Agriculture Organization of the United Nations (FAO) guidelines suggest an effective population size of 50 animals per generation, which leads to a rate of inbreeding of 1% per generation (FAO, 1998). To achieve 50 homozygous individuals would require an average of 200 heterozygous mating pairs at the *inter se* cross. This would mean that there would be little room for selection amongst the heterozygotes at the end of backcrossing the majority of

livestock populations and all would go forward for mating in the *inter se* cross. Therefore the probability that alleles are lost at this stage can be simplified $½^{n+1}$.

3.2.5. Measure of allelic diversity

The simulation and prediction results will be compared in terms of allelic diversity at a locus adapted from a method described by Allendorf (1986) as:

$A = (a_t(X) - 1)/ (a'(X) - 1)$

where $a_t(X)$ = the number of alleles at a locus from population X after t generations of backcrossing since their introduction and,

$a'(X)$ = number of alleles at a locus in population X (at equal frequencies) at their introduction to the backcross breeding programme; $a'(X) = N$ unless stated.

Therefore if $a'(D) = 5$ and $a_t(X) = 3$, $A = 2/4$, meaning that allelic diversity, A, is ½.

A ranges from 1, where all alleles are retained in the population, or 0, when only one allele is remains, to –1, where all donor alleles have been lost. The loss of allelic diversity is then related to the IBD due to donor genome around the target locus. Although overall allelic diversity approximately approaches $½^t$ this does not give information on the diversity of alleles from different sets of ancestors. The diversity seen in alleles from particular ancestral populations (*e.g.*, donor population in introgression) is important when tracts of genome around a selection locus will trace back to a certain set of ancestors.

3.2.6. Simulations

The initial cross for the introgression scheme was assumed to be between two lines fixed for alternative alleles at each locus. A carrier chromosome of length l was simulated using Haldane's mapping function (1919) to generate recombinations. Selection took places at a target locus s and N parents are selected at random at each backcross generation from the heterozygous offspring of the pervious generation producing n offspring at each backcross generation. All heterozygous offspring were used for parents of the *inter se* cross to produce the homozygous offspring population (IC).

The population parameters used were: parental population size (N = 10, 20, 50); offspring group size (n = 2, 4); the length of carrier chromosome (l = 0.5, 1 Morgan) and total number of backcross generations (T = 4, 6). The number of ancestral alleles at a locus, deriving from donor and recipient lines, were recorded for BC_t for a neutral locus and loci 0 to $l/2$ Morgan away from the target locus s. The simulations were run for 500 replicates. The results examine the validity of the predictions for the loss of alleles in a backcross breeding programme.

3.3. Results

3.3.1. Validity of the theoretical predictions

3.3.1.1. Prediction of the loss of alleles from a neutral locus

The loss of donor alleles at a neutral locus from the population BC_2 when $N = 20$ and $n = 2$ (total of $2N$ alleles in donor breed, resulting in 20 donor alleles at a locus in F_1) is given by Equation 3.3 ($k = 0$) as:

$$\Phi_2(0) = {}^{19}/_{78} + {}^{20}/_{39}\left({}^{1}/_{2} + {}^{1}/_{2}0\right)^1 + {}^{19}/_{78}\left({}^{1}/_{2} + {}^{1}/_{2}0\right)^2 = {}^{175}/_{312}$$

The loss of alleles by BC_3 are predicted by applying the p.g.f. described in Equation 3.4, $b(j) = (½ + ½({}^{175}/_{312})) = 0.78$ which is then used in the p.g.f. in Equation 3.3, $m(j)$, for all values of $k = 0$ to n, as above,

$$\Phi_3(0.78) = {}^{19}/_{78} + {}^{20}/_{39}(0.78)^1 + {}^{19}/_{78}(0.78)^2 = 0.792$$

This results in $a_{NC,3}(D) = 1 - 0.792 = 0.208$ of donors alleles (4.16 alleles, $A = 0.166$) surviving at a neutral locus chromosome in BC_3. Following on, $a_{NC,4}(D) = 0.104$ surviving at a neutral locus until BC_4 (2.03 alleles, $A = 0.054$). The comparison of these results and other results ($N = 40$, $n = 2$ and $n = 4$, $N = 10$ or 20) to simulation studies (Table 3.1) shows close agreement, strongly suggesting that the method to predict the loss of alleles from a neutral locus is accurate.

TABLE 3.2. Simulation* (± standard errors) and prediction results of the allelic diversity (A) of donor (D) and recipient (R0) individuals remaining in genome of the parents of each backcross generation when n = 2 and N = 20 or 40, and when n = 4 and N = 10 or 20.

Parents of..	*A* for..	N = 20, n = 2		N = 40, n = 2		N = 10, n = 4		N = 20, n = 4	
		Sim^n	Pred^n	Sim^n	Pred^n	Sim^n	Pred^n	Sim^n	Pred^n
BC_2	$a_{NC,2}(D)$	0.414	0.409	0.428	0.424	0.353	0.354	0.393	0.386
	$a_{NC,2}(R_0)$	0.389	0.409	0.427	0.424				
BC_3	$a_{NC,3}(D)$	0.168	0.166	0.189	0.213	0.100	0.105	0.154	0.151
	$a_{NC,3}(R_0)$	0.172	0.166	0.190	0.213				
BC_4	$a_{NC,4}(D)$	0.055	0.054	0.081	0.078	-0.007	-0.007	0.049	0.045
	$a_{NC,4}(R_0)$	0.060	0.054	0.082	4.04				

* Standard errors of simulation results ranged from ± 0.001 to ± 0.004

3.3.1.2. Predicting the number of copies of an allele in a backcross population

Equation 3.3 and 3.4 can be extended to predict the probability that of a certain number of alleles surviving to the parents of a particular backcross generation. Accounting for the number of ancestral alleles transmitted to populations can be used to predict the frequency of a particular allele in a given generation. Following on from the previous example ($N = 20$ and $n = 2$),

$\Phi_2(1) = {}^{20}\!/_{39}\,{}^{1}\!/_{2} + {}^{19}\!/_{78}\,{}^{1}\!/_{2} = {}^{59}\!/_{156}$ and

$\Phi_2(2) = {}^{19}\!/_{78}\,{}^{1}\!/_{2}^{2} = {}^{19}\!/_{312}$

These predictions, and predictions when N = 10 and 20 and n = 4, are in close agreement with the simulation results (Table 3.3). This is not the objective of this study and is presented to demonstrate the robustness of the prediction described in Equation 3.3. The number of alleles remaining from ancestral populations will not be examined any further.

TABLE 3.3. Comparison of simulation (± standard errors) and prediction results for the number of donor (D) and recipient (R0) alleles remaining in genome of the parents of BC2 when N = 20 and n =2, 4 and N = 10 and n = 4

	$N = 10, n = 4$		$N = 20, n = 4$		$N = 20$ and $n = 2$.	
	Pred^n	**Sim^n**	**Pred^n**	**Sim^n**	**Pred^n**	**Sim^n**
$\Phi_2(0)$	5.80	5.82 ± 0.039	11.67	11.54 ± 0.055	11.22	11.13 ± 0.054
$\Phi_2(1)$	3.90	3.44 ± 0.045	6.92	6.99 ± 0.062	7.56	7.53 ± 0.055
$\Phi_2(2)$	0.64	0.71 ± 0.023	1.29	1.40 ± 0.033	1.22	1.22 ± 0.014
$\Phi_2(3)$	0.05	0.06 ± 0.007	0.12	0.14 ± 0.011		
$\Phi_2(4)$	0.002	0.002 ± 0.001	0.004	0.004 ± 0.002		

3.3.1.3. Predicting the number of alleles after an *inter se* cross

The loss of alleles after an *inter se* cross can be given by $¼^n$ (*i.e.*, the probability of none of the *n* offspring being homozygous at the target locus). Table 3.4 shows the number of donor (D) and recipient (R_0) alleles individuals that still contributed to the genome of the offspring of the last backcross generation (the parents of IC, $½^n$ lost from the predictions given in Table 3.2) and IC (another $½^n$). The approximation of $¼^n$ resulted in an over-prediction of the loss of alleles through an *inter se* cross when compared to the simulation results in Table 3.4 (N = 10, 20 and 40, n = 2 and 4). The prediction seemed to over-predict the loss when offspring group size is larger (compare n = 2 and 4). This over-prediction did not continue to increase as offspring group size increased (results not shown). It is worth considering that the prediction results for allelic diversity in BC_4 was slightly larger than the simulation results

(Table 3.2). The reasons for these over-predictions will be discussed later. Also, the allelic diversity had reached a low level by the last backcross generation, as shown in Table 3.2, which may affect the further loss of ancestral alleles from the genome of the offspring of the *inter se* cross.

TABLE 3.4. Comparison of simulation* and prediction results for the loss of donors (D) from the genome of the offspring of the last generation (BC4) and the offspring of the *inter se* cross for three different population structures.

Pop^n	Parents of IC		$ai_{NC,4}$(D)	
	Sim^n	Pred^n	Sim^n	Pred^n
$N = 40, n = 2$	0.060	0.052	0.046	0.032
$N = 10, n = 4$	-0.011	-0.013	-0.013	-0.018
$N = 20, n = 4$	0.045	0.039	0.043	0.033

* Standard errors of simulation results range from ± 0.001 to ± 0.003.

3.3.1.4. Loss of alleles on a carrier chromosome

The prediction of the loss of allelic diversity at a locus on the carrier chromosome using Equation 3.4 was shown to agree closely with simulation results in Figure 3.1 (N = 10, n = 4, l = 1 Morgan), showing that the loss in allelic diversity increased as distance from the target locus increased. More alleles were retained at the target locus than at a neutral locus (see comparison in Table 3.5). Simulation and prediction results agreed over all sets of parameters studied.

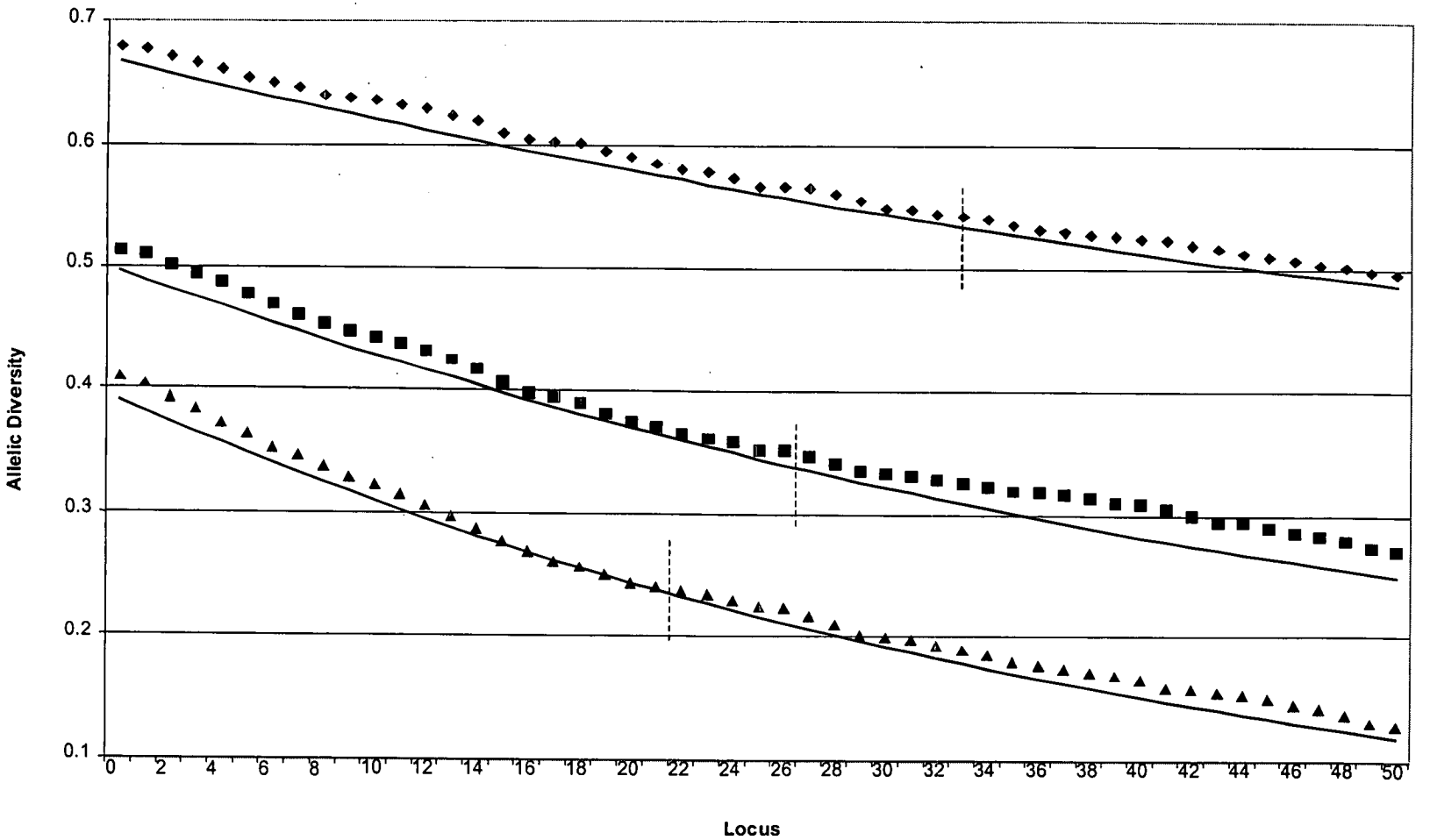

FIGURE 3.1. Simulation (points) and prediction (solid line) results of the loss of donor allelic diversity from the loci on the carrier chromosome expressed as a function of the distance from the target locus in BC_2 (♦), BC_3 (■) and BC_4 (▲) when $N = 10$, $n = 4$, $l = 1$ M (using equation 3.4). The vertical dotted lines indicate the approximate end of the linkage drag segment on carrier chromosomes in each of the populations.

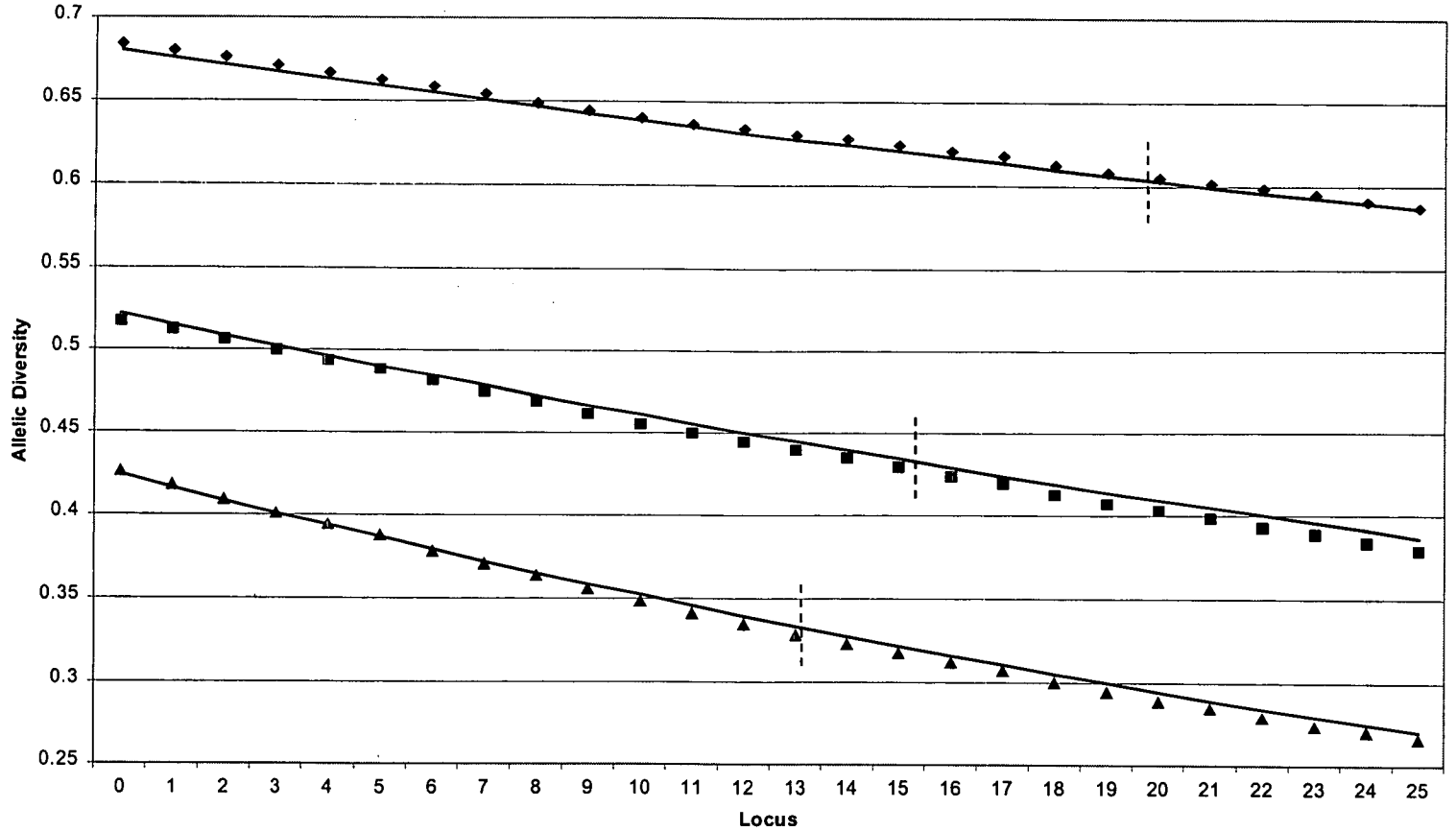

FIGURE 3.2. Simulation (points) and prediction (solid line) results of the loss of donor allelic diversity from the loci on the carrier chromosome expressed as a function of the distance from the target locus in BC2 (♦), BC3 (■) and BC4 (▲) when N = 50, n = 4 and l = 0.5 M (using equation 3.4). The vertical dotted lines indicate the approximate end of the linkage drag segment on carrier chromosomes in each of the populations.

TABLE 3.5. Simulation* and prediction results for the loss of donor allelic diversity at particular loci along the carrier chromosome in the population of BC parents when T = 2 to 4 when N = 10, n = 4 and l = 100 cM (using equation (8)).

	Sim BC_2	Pred BC_2	Sim BC_3	Pred BC_3	Sim BC_4	Pred BC_4
Target	0.679	0.667	0.513	0.497	0.410	0.391
1 cM	0.677	0.662	0.510	0.489	0.403	0.382
10 cM	0.636	0.620	0.443	0.428	0.323	0.310
20 cM	0.590	0.579	0.374	0.259	0.244	0.244
30 cM	0.549	0.543	0.333	0.322	0.199	0.192
40 cM	0.524	0.512	0.309	0.282	0.166	0.151
50 cM	0.495	0.487	0.271	0.248	0.128	0.118
$a_{NC,2}$**(D)**	0.353	0.354	0.107	0.105	-0.007	-0.006

* Standard errors of simulation results range from ± 0.003 to ± 0.004 for $a_{NC,2}$(D) and from ± 0.007 to ± 0.01 for all other allelic diversity on the carrier chromosome.

3.3.2. Effect of the number of backcross generations on loss of allelic diversity

As the number of backcross generations increased, the loss in donor allelic diversity at a neutral and selection locus increased (Table 3.2 and Figure 3.1). The loss in allelic diversity of recipient ancestral groups at neutral loci of the populations used in backcrossing also decreased at the same rate as donor alleles (results not shown). The loss in diversity of recipient alleles in the region of the target locus was higher due to the selection pressure on donor alleles. Although diversity was lost in alleles from one ancestral population (in this case donor) the overall locus allelic diversity was not be as low. Even though donor alleles were lost new alleles (*e.g.* recipient) enter the system resulting in a higher allelic diversity than described by Equation 3.7.

The prediction of the loss of alleles appeared to lose accuracy as the number of backcross generation increased, showing a slight upward bias resulting in an over-prediction of the loss of donor allelic diversity. This deviation in the prediction of allelic diversity from the simulation study results could be due to the population parameters used. Due to the small population sizes the number of ancestral alleles remaining was very low. Therefore the selection of an individual from a family for the next generation would influence the probability of another individual (and the ancestral allele it carries) being selected. However, the equations presented here assume independence of the sampling of gametes from parents to offspring and the random selection of N parents for the next generation.

3.3.3. Effect of carrier chromosome length on loss of allelic diversity

Figures 3.1 and 3.2 show that as the distance from the target locus increased there was a gradually decrease in ancestral allelic diversity. However, there was still higher donor allelic diversity on the carrier chromosome compared to a neutral locus (Table 3.4). Figure 3.3 ($T = 10$, $l = 4$ M, $N = 50$, $n = 4$) shows that the as distance from the target locus increases, in map distance or time in the form of extra backcross generations, that loci become more like a non-carrier chromosome, as expected. This is due to recombination events breaking the linkage between the locus and the target locus.

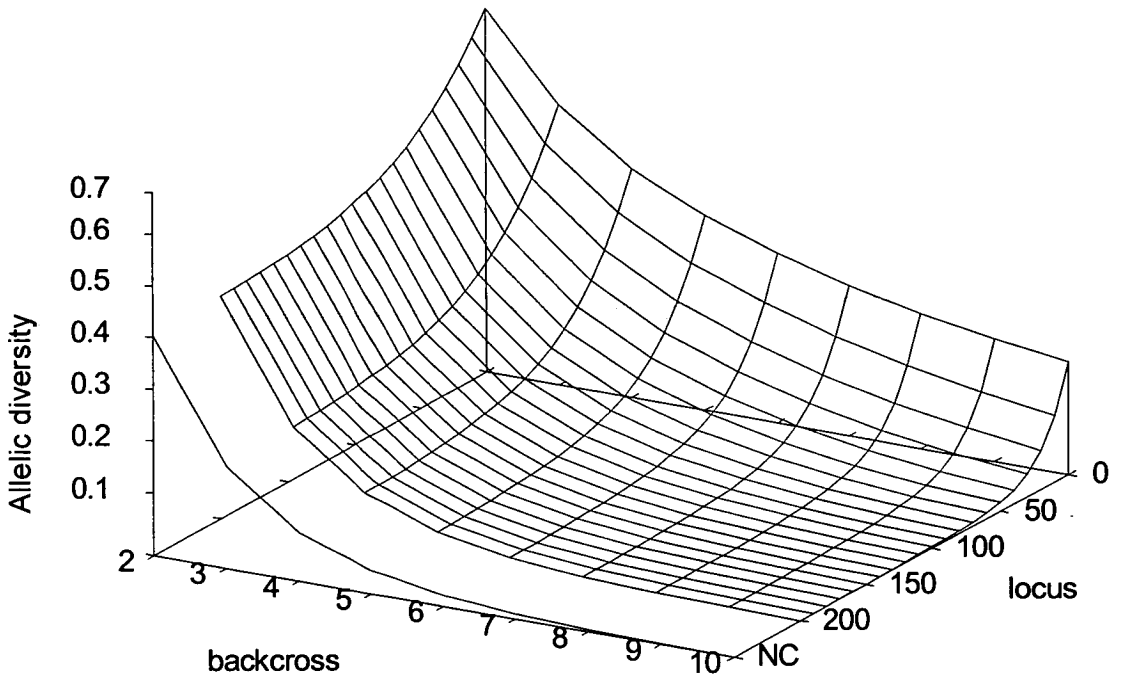

FIGURE 3.3. The loss of donor allelic diversity from loci on a carrier chromosome (l = 4 M) over 10 backcross generations when N = 50 and n = 4. The single line on the graph refers to of the loss of alleles from a non-carrier chromosome over the 10 backcross generations.

With introgression a quantity of donor genome is dragged into the recipient line along with the target allele. This length of intact donor genome around the target locus is called linkage drag (Brinkman and Frey, 1977) and has been described theoretically by many authors (Hanson, 1959, Stam and Zeven, 1981 and Chapter Two). The dashed line on Figures 3.1 and 3.2 indicate the position of the approximate end of the linkage drag segment (as derived from Stam and Zeven, 1981) and shows that loci in this area do not equate to a neutral locus, indicating that the effect of selection on allelic diversity continues outside the linkage drag. In other words, when $l \rightarrow 0$ M and/or $T \rightarrow 0$ the loss in allelic diversity from loci on the carrier chromosome will tend to that of the target locus. However when $T \rightarrow \infty$ and/or $l \rightarrow \infty$ the loss of donor alleles away from the target locus will tend to that of a non-carrier chromosome locus, *i.e.* with zero donor alleles remaining at a locus.

3.3.4. Effect of the population structure and number of alleles at a locus on loss of allelic diversity

The loss of allelic diversity at a neutral locus decreased as population size, *N*, increases but the loss increased as the number of offspring, *n*, increases (Table 3.2 and Table 3.6). Table 3.6 shows the expected change in allelic diversity at a neutral locus in BC_2 with differing population structures. The number of parents (*N*) and offspring group size (*n*) had a substantial effect on allelic diversity, especially when these values were small. Changes in population structure seemed to have a constant effect in the loss of allelic diversity across all *N* and *n* studied, with little difference between the decrease in the loss of allelic diversity as population size increases or as offspring group size decreases. This trend continued for all backcross populations studied and at loci on the carrier chromosome (results not shown).

TABLE 3.6. Prediction of donor allelic diversity in BC_2 at a neutral locus for different population structures ($N = 10, 20, 40, 80, 160, 320$; $n = 2, 4, 6, 8$).

		n =			
		2	**4**	**6**	**8**
N =	**10**	0.379	0.355	0.348	0.344
	20	0.410	0.386	0.378	0.375
	40	0.424	0.400	0.393	0390
	80	0.431	0.407	0.400	0.397
	160	0.434	0.410	0.403	0.400
	320	0.436	0.412	0.405	0.402

The results also show that as population size became large the change in loss of allelic diversity became quite small (less the 1% increase in allelic diversity when *N*

≥ 80 and less than 1% decrease when $n = 8$). This shows that the loss of ancestral allelic diversity is a problem that should be considered when population size is small and/or with large offspring group size.

Figure 3.4 shows that allelic diversity get higher as population size increases when the initial number of donor alleles, $a'(D)$, is $2N$. The increase in allelic diversity is not constant as population size increases and slows down as population size increases. There is a slight decrease in allelic diversity as population size increases if a fixed number of alleles is assumed (in this example $a'(D) = 5$). The sampling of the fewer alleles, in this second example, in a larger number of individuals results in a greater loss of allelic diversity. If the number of alleles is small increasing the population size in the breeding programme will not help to maintain donor allelic diversity.

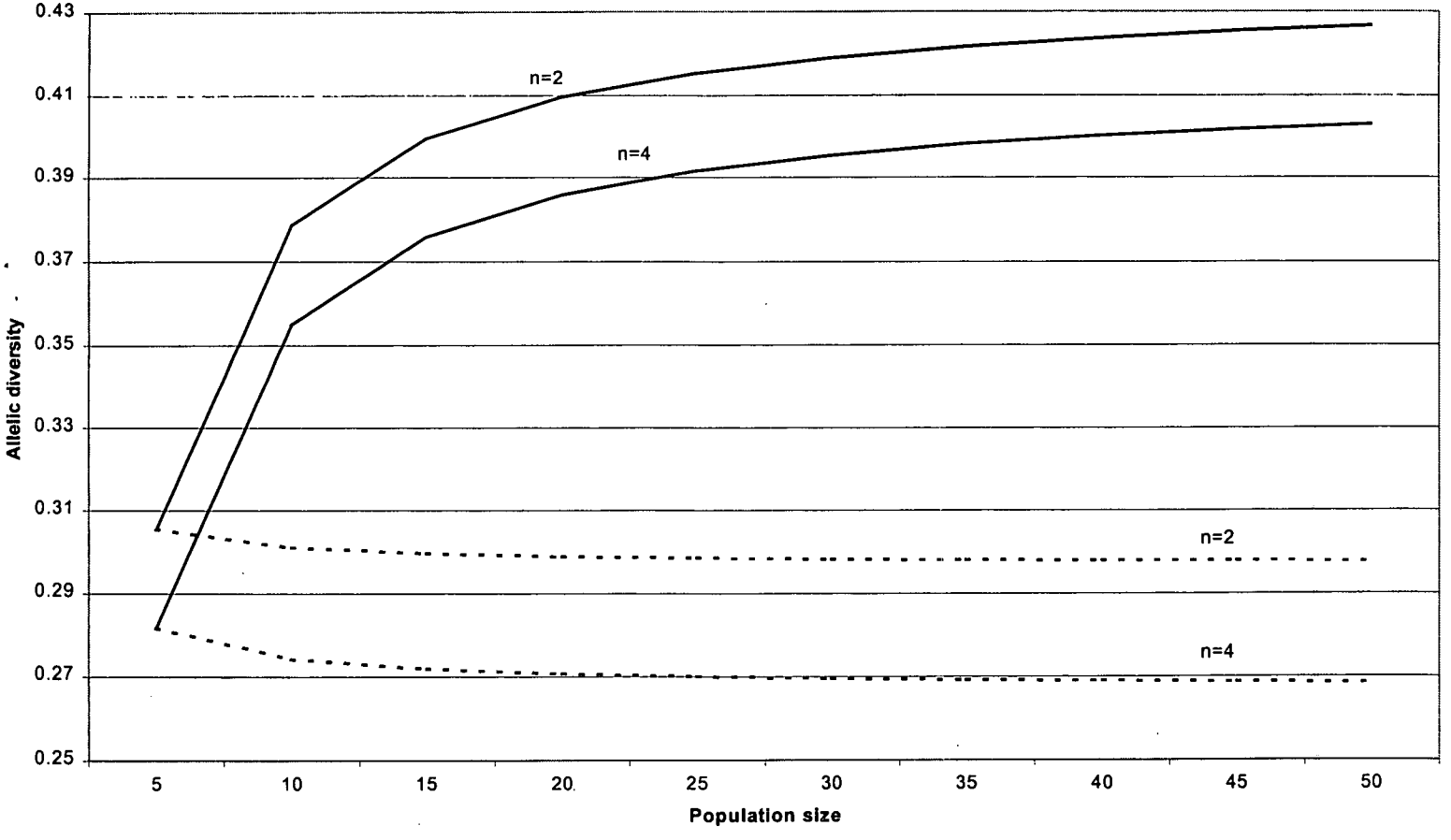

FIGURE 3.4. Changes in donor allelic diversity at a neutral locus in BC_2 when the initial number of alleles is $2N$ (solid lines) or fixed number of 5 (dashed lines) over different population structures of $N = 5, 10, 15, \ldots, 50$ and $n = 2$ and 4.

3.4. Discussion

This study develops a methodology to predict the loss of ancestral alleles from the genome of individuals involved in the backcross scheme for a neutral locus and around loci of selection over many population structures. There is close agreement between the simulations and the predictions that have been developed for all population structures studied for carrier and non-carrier chromosomes in a number of different types of backcross schemes. Allele number is reduced by decreased population sizes, increased offspring group sizes, increased backcross generations and distance from a locus of selection.

3.4.1. Neutral locus

The prediction of the loss of donor alleles from a neutral locus during backcross generations shows great reduction in the number of donor alleles throughout backcrossing. In all situations just over half of the donor alleles are lost by BC_2. This rate of loss appears to continue as the numbers of backcross generations increases. In simulation studies and prediction results (results not shown) for small population sizes, it was seen that the dramatic rate of loss did not continue when the allelic diversity dropped below 0 (less than one donor allele remaining at a locus). Therefore it is quite difficult for all alleles from a particular ancestral population to be lost, although frequencies may become extremely low.

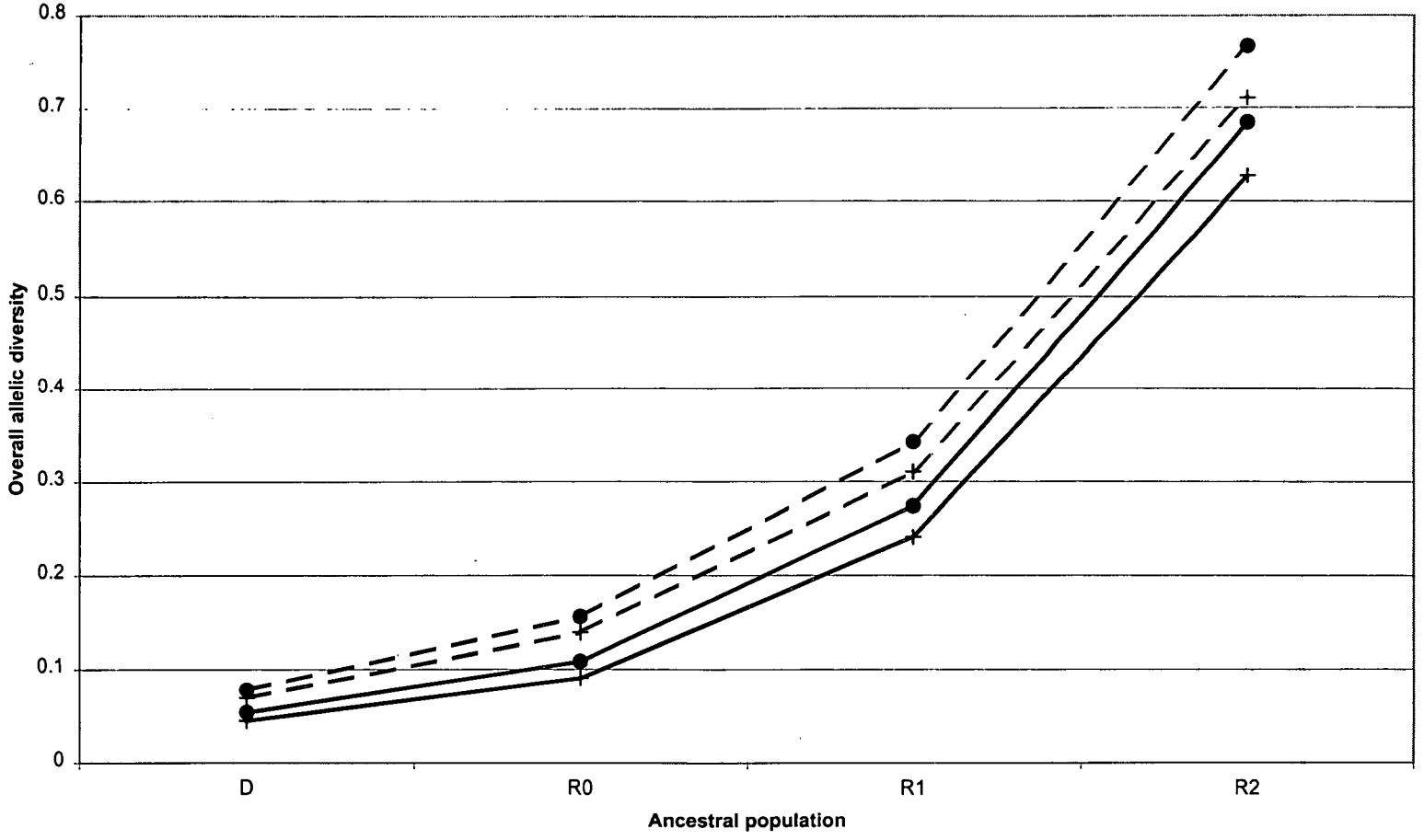

FIGURE 3.5. Cumulative allelic diversity (over allelic diversity for each ancestral population) at a neutral locus in the parents of the *inter se* cross for a backcross scheme of 3 backcross generations where solid line indicates $N = 20$ and dashed line indicates $N = 40$ and (•) indicates $n = 2$ and (+) indicates $n = 4$.

Ancestral alleles were shown to be lost from populations with large offspring group sizes. This is the case because as the number of offspring per family group increases the chance of an ancestral allele being represented multiple times in the next generation increases. For this to be the case another allele would be excluded and has a lower probability of going forward to the next generation. In other words, if there are two matings with two offspring per mating and one mating has both offspring selected as the two parents of the next backcross generation, the allele carried by the other mating is now permanently lost from the introgression population.

Although the number donor alleles drops quite quickly, the overall allelic diversity is not as low as new recipient breed alleles enter the system to maintain a more substantial allelic diversity. Equation 3.7 can be used to describe the allelic diversity at a neutral locus for all ancestral populations introduced during backcrossing. Figure 3.5 shows there is a decrease overall number of alleles (donor and recipient) but not as dramatic as the loss of donor alleles alone. Although a high proportion of donor alleles are lost during backcrossing the introduction of new alleles from recipient breeds keeps overall allelic diversity quite high at a neutral locus.

3.4.2. *Inter se* cross

The prediction of the loss of donor alleles from the end of the backcross generation to the offspring of the *inter se* cross does not hold as well as those presented for the loss during the backcross generations. The predictions given by Equation 3.7 for the

loss by the end of the backcross generations demnstrated a downward bias in the majority of the cases. The under-prediction of the loss of alleles shown in the *inter se* cross populations might be an inflation of these effects.

The methodology used in Equation 3.7 to predict the loss of alleles in the backcross and *inter se* cross considers each selection stage and gametic sampling event as independent processes. However, independent sampling may not always best describe selection events, especially when population size or the number of offspring is small or when the number of ancestral alleles remaining in a population becomes low. Once one parent carrying a particular ancestral allele has been selected it reduces the chance of the other parents carrying other donor alleles being selected. This results in a negative correlation between the successful transmissions of ancestral alleles from one generation to the next whereas this described methodology relies on their independence.

Also, the methodology used to predict the loss of alleles from the donor ancestors between the backcrossing and the *inter se* cross assumes that no matings will be between individuals that carry the same heterozygous genotype at the non-carrier locus (*i.e.*, carry the same donor allele). In general, the probability of two individuals being paired up carrying the same ancestral allele at a heterozygous locus is low, especially when the population size (N) is large. If a mating between individuals carrying the same donor allele at a locus occurs then the chance that this allele will be lost from the population become ¼. However, it is assumed that this does not occur in the prediction used and therefore the chance that a donor allele is lost is

assumed to be ½ on average. This can account for the downward prediction of the loss of alleles when compared to the simulation results. This is especially the case when the number of donor alleles becomes quite small and potentially multiple copies exist in a small population and therefore there is more of a chance that parents of the *inter se* cross will carry the same donor allele at a locus.

3.4.3. Carrier chromosome loci

There is close agreement between simulation and prediction results for the loss of ancestral alleles from the carrier chromosome showing that the loss is lower than that of a neutral locus and decreases moving away from the locus of selection. The linkage drag has an important role in keeping the number of donor alleles on the carrier chromosome higher than that of a non-carrier chromosome locus. This can be seen clearly by the difference between the neutral loci and loci on the carrier chromosome results in Table 3.5, even in BC_4 when the linkage drag segment is approximately 0.432 of the carrier chromosome. Even though the carrier chromosome is only partly made up of a linkage drag segment, there is still linkage to the extreme points of the chromosome keeping the number of donor alleles on a carrier chromosome higher than at a neutral locus. However on an infinitely long carrier chromosome and/or for an infinite number of backcross generations, the variation at loci on the carrier chromosome outside the linkage drag segment will tend towards that of a neutral locus, as shown in Figure 4.

3.4.4. IBD in livestock introgression programmes

This work shows a rapid decline in the number of original donor alleles still contributing to the genome on a carrier chromosome with selection on a target locus during backcrossing. If large tracts of the genome of the backcross population trace back to a low number of ancestors it increases the chances of large tracts of the genome being IBD and result in high levels of inbreeding after the *inter se* cross. Other studies have examined various tools to minimise the linkage drag length on the carrier chromosome (*e.g.,* Frisch & Melchinger, 2001) or speed up the level of recipient genome recovery to optimise introgression programmes (*e.g.,* Hospital & Charcosset, 1997; Visscher *et al*, 1996). These types of studies have added greatly to the material to aid the design of optimum plant and animal introgression schemes by minimising the amount of donor genome retained on carrier and non-carrier chromosomes.

However, in large animal introgression programmes there may be multiple end goals including the minimisation of donor genome contamination that will reduce the genetic worth of the population. Firstly, the number of animals that will practically be available for mating and a small number of offspring per mating will limit the opportunities for the selection of favourable recombinants to minimise donor genome in large animal introgression schemes. Secondly, these studies do not consider the implications of the potential increase in inbreeding through this type of selection, which could be quite large in smaller introgression schemes. These are important factors because, in the majority of livestock cases, there would be a desire to have an end product that one would be able to market as an animal to be bred across the

population at large. The increase in IBD will reduce the genetic variance of this trait as alleles become fixed at target, and maybe other loci thus reducing the effective number of breeding animals to further mate to the population. So in this case one of the goals of the breeding scheme would be to introgress the favourable effects of the target allele whilst maximising the recipient breed traits and the genomic variation.

Allele number is reduced by decreased population sizes, increased offspring group sizes, increased backcross generations and increased distance from the target locus. In turn, homozygosity due to allele being identical-by-descent (IBD) around the target locus is increased as offspring group size increases and population size decreases. However, the homozygosity due to donor genome over backcross generations becomes constant moving further away from the locus of selection. This homozygosity can be directly related to the inbreeding in the final population. The high levels of homozygosity around the target locus could result in unfavourable recessive donor alleles dragged into the recipient population. There may be a negative effect in fitness of the final homozygous population if unfavourable recessive alleles up or down stream from the target allele also become homozygous-by-descent.

This study can be used in conjunction with other studies that base their selection on foreground or background selection to try and minimise the fixation of ancestral alleles at particular loci around the genome. By maximising the variation at loci around the genome throughout the backcross generation, in turn, you are minimising the risk of producing tracts of IBD after the *inter se* cross. This will help design an

introgression scheme that achieves the goal of introducing a new gene into and recipient population whilst maximising the variation at a genomic level resulting in a optimised effective breeding pool post introgression.

CHAPTER FOUR

HOMOZYGOSITY IN GENE INTROGRESSION PROGRAMMES USING INBRED LINES

4.1. Introduction

Many studies examining introgression tend to look closely at the backcross generations and methods of marker-assisted introgression to minimise genetic lag due to donor contamination (*e.g.*, Hospital *et al.*, 1992; Hillel *et al*, 1993; Visscher *et al.*, 1996; Frisch & Melchinger, 2001). Very few studies have examined the properties of genomic contributions and loss of alleles (and therefore variation) or the properties of the *inter se* cross as done in Chapters Two and Three. Very few studies have examined homozygosity by descent in the *inter se* cross. Tracts of the genome can become homozygous due to loci being identical-by-descent (*i.e.*, 2 alleles at a locus tracing back to a common ancestor). If an allele frequency is very high, tracts of the genome may become homozygous due to loci being identical-by-state (IBS) meaning they have the same allele at a locus without being identical-by-descent (IBD).

Homozygosity, in part, will be due to overlapping linkage drag segments on the carrier chromosome. This means that there is a possibility of a large region of homozygosity around the locus of interest. Homozygosity can also occur on non-carrier chromosomes or at neutral loci. There will be large genomic contributions from the recipient parents used in the latter generations of backcrossing on non-carrier chromosomes (Chapter Two). These recipient alleles may be IBD in the backcross population and therefore cause large regions of homozygosity in the offspring of the *inter se* cross.

This Chapter examined the level of homozygosity on the carrier and non-carrier chromosomes after introgression of a target allele. Homozygosity was described in terms of alleles at a locus tracing back to the same ancestral population (donor or a recipient population used in one of the backcross generations). This would be a reasonable assumption when dealing with crosses using inbred lines (plants, rats, mice) or when creating congenic strains. In these situations, the homozygosity described will relate to the homozygosity due to alleles being IBS. The effect of (*i*) the number of backcrosses, (*ii*) the length of the carrier chromosome and total genome, and (*iii*) the structure of the populations, *i.e.*, numbers of parents and offspring per generation on homozygosity was described.

4.2. Materials and Methods

4.2.1. Simulations

A stochastic process was used to simulate the carrier chromosome of individual parents from an introgression programme after each generation of backcrossing. The initial cross of N matings (with n offspring per mating) was assumed to be between two divergent inbred lines that are fixed for alternative alleles at each loci. The carrier chromosome was simulated using crossing-over events generated assuming Haldane's mapping function without interference (1919). Parents were selected for T backcross generations from among offspring heterozygous for the centrally placed desired allele, and any further selection was at random. Genotypes were simulated to be independent to any phenotypic or population data.

Effect of population structure: Homozygosity was studied by simulating a parental population size (N) of 20, 50 or 100 with 3 or 5 offspring per mating (n). One offspring per mating was avoided because the introgression programme could fail, since the number of heterozygous candidates with the desired genotype would decrease by a half each backcross generation. The results presented, for each set of population parameters, are for a carrier chromosome of length (l) 1 Morgan, with T = 6 backcross generations followed by an *inter se* cross.

Effect of length of carrier chromosome (l): Lengths of 0.5, 1 or 2 Morgans were considered representative of livestock species. Simulation of extreme chromosome lengths were also run (l =8,16 with N = 20, n = 5, T=6) to extrapolate trends in genomic contributions seen in the standard lengths.

Effect of the number of backcross generations (T): The number of backcross generations was set to 1, 3, 6 or 10 generations. These simulations were run with a population structure of N = 20, n = 5 and l = 1 Morgan.

For each set of parameters, 500 replicates were simulated and the following values were recorded from the genome of the population homozygous for the target locus after the *inter se* cross (IC): (*i*) the proportion of homozygosity within the linkage drag; and (*ii*) and proportion of homozygosity arising from each ancestor generation. The proportion of homozygosity calculated by looking at the number of loci on a chromosome that are homozygous due to both alleles originating from the same generation. This is expressed as a proportion of the length of the chromosome. A

simulation of a non-carrier chromosome (NC) was carried out to examine homozygosity elsewhere in the genome

4.3. Results

4.3.1. Effect of population structure

The proportion of homozygosity on the carrier chromosome pair after introgression differed slightly with population structure. Table 4.1 showed that total homozygosity decreased slightly as population size (N) and offspring number (n) increased. However, the difference between overall homozygosity was only slightly significant with a range of 0.287 for $N = 100$ and $n = 5$ and 0.331 for $N = 20$ and $n = 30$. The effect of increasing the number of offspring (n) from 3 to 5 was had significant effect on the decrease of overall homozygosity ($P<0.005$), donor homozygosity ($P<0.05$) and approaching significance for the homozygosity due to many of the recipient ancestral populations. The effect of increasing the number of mating pairs at each generation (N) from 20 to 50 to 100 was had significant effect on the decrease of overall homozygosity on the carrier chromosome only ($P<0.05$), but approached significance for donor homozygosity.

Non-carrier chromosomes (NC) had the highest proportion of homozygosity, which was attributed to high genomic contributions from individuals used in the last generation of backcrossing (R_6 in this case), as seen in Chapter Two. The donor and early ancestral recipient populations contributed little to the non-carrier chromosomes of the homozygous population for the target allele and therefore had a very low level of homozygosity.

The final number of individuals homozygous at the target locus, after the *inter se* cross, agreed with the prediction of Mendelian inheritance, with an average of ¼ of all offspring from the *inter se* cross being homozygous for the target allele. The range of the final population size was wide and showed a low minimum number of homozygous individuals for some of the population structures (*e.g.*, the minimum value for N=20 and n=3 is 2).

TABLE 4.1: Effect of population structure in the introgression programme (with l = 1 Morgan, T = 6) on proportion of homozygosity-by-descent due to a pair of alleles at a given locus originating from the individuals used in the same generation of backcrossing on the carrier chromosomes of IC (individuals homozygous for the target allele. Also shown is the total homozygosity on the carrier chromosome (HOM) and the final number of IC (FN) (range in brackets).

N	n	D	R_0	R_1	R_2	R_3	R_4	R_5	R_6	HOM	FN
20	3	0.180	0.011	0.011	0.015	0.018	0.028	0.025	0.043	0.331	7.4±0.1
		±0.003	±0.001	±.001	±0.001	±0.001	±0.002	±0.001	±0.002	±0.003	2-15
	5	0.174	0.008	0.010	0.012	0.016	0.018	0.023	0.047	0.307	12.6±0.1
		±0.003	±0.001	±0.001	±0.001	±0.001	±0.001	±0.001	±0.001	±0.004	5-22
50	3	0.177	0.009	0.012	0.015	0.0190	0.025	0.023	0.041	0.319	18.9±0.2
		±0.002	±0.001	±0.001	±0.001	±0.001	±0.001	±0.001	±0.001	±0.003	9-31
	5	0.170	0.006	0.008	0.009	0.014	0.019	0.025	0.043	0.292	30.9±0.2
		±0.002	±0.000	±0.000	±0.001	±0.001	±0.001	±0.001	±0.001	±0.002	18-45
100	3	0.175	0.008	0.011	0.014	0.0180	0.025	0.023	0.043	0.317	37.9±0.2
		±0.001	±0.000	±0.000	±0.001	±0.001	±0.001	±0.001	±0.001	±0.002	21-59
	5	0.166	0.005	0.007	0.010	0.013	0.004	0.024	0.043	0.287	61.8±0.3
		±0.001	±0.000	±0.000	±0.000	±0.000	±0.000	±0.001	±0.001	±0.001	42-80
NC		0.0002	0.0003	0.0006	0.002	0.005	0.018	0.063	0.247	0.335	50.0±0.0
		±0.000	±0.000	±0.000	±0.000	±0.000	±0.001	±0.001	±0.002	±0.002	

4.3.2. Effect of carrier chromosome length

The length of the chromosome had an effect on homozygosity due an ancestral population and total homozygosity. Donor genome homozygosity (shown on the

lower graph on Figure 4.1) was high when l = 0.5 Morgans and made up a large proportion of the total homozygosity. This was mainly attributable to the high overall contribution of donor genome to smaller chromosomes (Chapter Two), which resulted in high donor genome homozygosity and therefore high overall homozygosity. The total homozygosity decreased when l = 1 and 2 Morgans but then increased slightly for the more extreme lengths of 8 and 16 Morgans due to the higher genomic contribution from R_7.

4.3.3. Effect of the number of backcross generations

The lower graph of Figure 4.2 shows homozygosity on the carrier chromosomes of the population homozygous for the target allele and overall homozygosity over introgression programmes with different number of backcross generations. As expected, homozygosity was highest with few generations of backcrossing, with the majority of the homozygosity due to donor genome homozygosity. This high donor genome homozygosity was mainly attributed to the overlap of the linkage drag segments on a carrier chromosome pair. Increasing the number of backcross generations shortened the linkage drag segment therefore reduced donor genome contribution to the overall homozygosity. The overall homozygosity, in these cases, was due more to the recipient individual genome (lower graph of Figure 4.2).

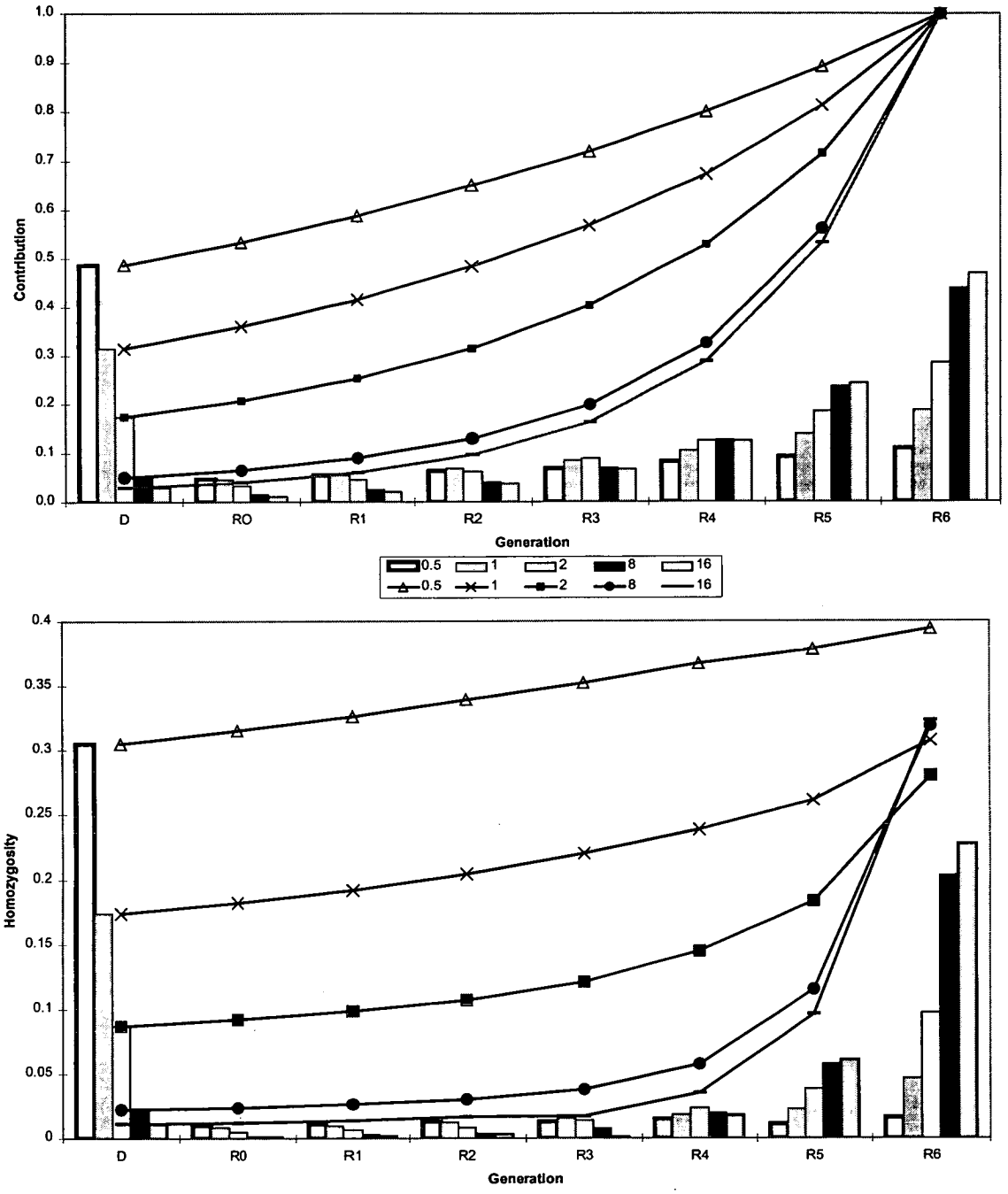

FIGURE 4.1. Genomic contributions (upper) and the proportion homozygosity (lower) from individuals used at each generation in the introgression programme to the carrier chromosome pair of homozygous individuals for the target allele post introgression over five carrier chromosome lengths (0.5 – 16 Morgans). Lines indicate the cumulative values of genomic contributions and total homozygosity over the generations.

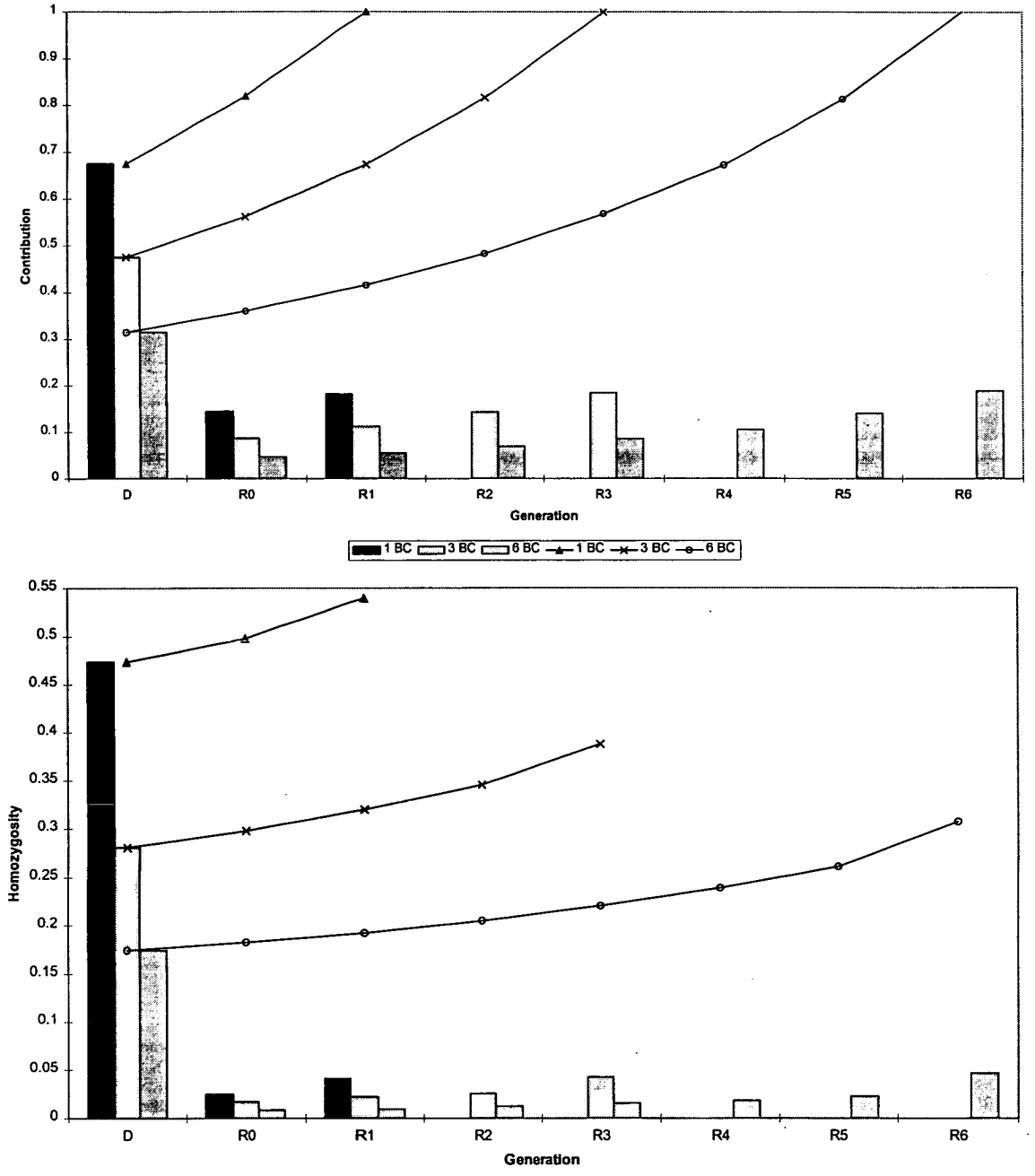

FIGURE 4.2. Genomic contributions (upper) and the proportion homozygosity (lower) from individuals used at each generation in the introgression programme to the carrier chromosome pair of homozygous individuals for the target allele post introgression after three types of introgression programmes with 1, 3 and 6 backcross generations. Lines indicate the cumulative values of genomic contributions over the generations.

4.4. Discussion

The proportion of homozygosity along the carrier chromosome pair was higher on the 0.5 Morgan chromosome compared to the 1 and 2 Morgan chromosome. However, for the extreme lengths of 8 and 16 Morgans the proportion of homozygosity began to rise again due to the high genomic contributions from individuals used in the last recipient generation. Increasing the number of backcross generations decreased homozygosity and genomic contribution from donor individuals, in turn increasing the genomic contribution from recipient individuals. Increasing the population size (N) and the number of offspring per mating (n) had a significant effect on decreasing overall homozygosity. However, increasing n had a significant effect on reducing homozygosity due to donor and some recipient ancestral populations.

The expected proportion of loci that are homozygous by descent in an individual is the inbreeding coefficient of that individual (Franklin, 1977). Homozygosity was highest on non-carrier chromosomes due to the increased genomic contribution (and therefore homozygosity) from individuals used in later backcross generations. However, homozygosity from each generation presented here was not equivalent to loci that are identical-by-descent (IBD) and therefore inbreeding, because there are a number of individuals used at each backcross generation. As shown in Chapter Three, the loss of alleles from donor and early recipient ancestral populations is dramatic by the time of the *inter se* cross. Therefore, the homozygosity shown in these results could relate to the true homozygosity post introgression due to alleles being IBD. However, the high level of homozygosity due to latter ancestral recipient

populations is not a reflection of homozygosity due to alleles being IBD. These simulations assume that the donor and recipient populations are inbred lines. These results do not relate to the homozygosity that will be seen in an introgression scheme using outbred populations. In this case, there are up to $2N$ different alleles at each locus in the donor and recipient populations and the probability of homozygosity post introgression will be much lower.

The results showed that as the number of mating pairs (N) increased, the proportion of homozygosity decreased, both on the carrier and non-carrier chromosome. When N increased mating pairs for the *inter se* cross were selected from a larger pool of individuals and, therefore, the increased sampling reduced the probability of homozygosity. As the number of offspring per mating (n) increased, there was a significant reduction in overall homozygosity as well homozygosity due to many of the ancestral populations. The increase in offspring group size resulted in a larger pool of carriers of the target locus in the backcross generations, and at the *inter se* cross therefore increasing the sampling variance.

Donor homozygosity on the carrier chromosome was shown to be high. This homozygosity was mainly due to the overlapping linkage drag segment, which traced back to the donor population and therefore caused a tract of homozygosity around the target locus. If tracing each donor ancestor (*i.e.*, following alleles that are IBD) this homozygosity might not be as high as there would still be some donor allelic diversity (Chapter Three) however reduced. However, the results do suggest that there was a high probability of creating an inbred tract of genome around the target

locus in the individuals homozygous for the target locus. This could have a negative impact on the introgression programme if this inbred tract contains deleterious recessive alleles (or alleles with a negative effect on production traits) associated with the donor breed.

To minimise the proportion of loci that are homozygous-by-descent on the non-carrier chromosome, these results indicate that increasing the number of the recipient individuals used in later backcross generations is more important than the number of recipient parents used in earlier generations. Increasing population size and the number of offspring per mating reduces the risk of homozygosity on the carrier chromosome. It is particularly necessary to have a high number of donor individuals or maintain allelic variation around the target locus to minimise the probability of tracts of homozygosity on the carrier chromosome.

These results only reflect homozygosity due to alleles at a locus tracing back to the same ancestral population. The extent of loci that are IBD needs to be estimated by tracing the contributions of specific ancestors to the final introgression population. This will be an estimate of true homozygosity in an introgression programme using outbred populations. This will be examined in Chapter Five.

CHAPTER FIVE

HOMOZYGOSITY DUE TO IDENTITY-BY-DESCENT IN GENE INTROGRESSION PROGRAMMES

5.1. Introduction

Chapter Four studied the homozygosity post introgression due to two alleles at a locus tracing back to a donor or recipient ancestral population only (from here on this will be referred to as generation homozygosity). The results of the study in Chapter Four can easily be applied to an introgression programme where recipient and donor populations are inbred (*e.g.*, mice, plants) but not to an outbred livestock introgression scheme. In this case, it is necessary to examine the homozygosity due to alleles being identical-by-descent (IBD) (from here on this will be referred to as IBD homozygosity).

Alleles that have descended from a single ancestral allele are said to be identical-by-descent or IBD. Relatedness in a population is because of sharing identical alleles. The coefficient of inbreeding (F) is the probability that two alleles at a locus within an individual are IBD (Falconer & Mackay, 1996). Therefore, studying IBD homozygosity will provide a measure of inbreeding in the population homozygous for the target locus.

Since blocks of linked alleles rather that individual genes are transmitted from one generation to the next, blocks of linked alleles rather that single genes become identical as inbreeding proceeds (Stam, 1980). At any stage of inbreeding, the genome of an individual will consist of alternative regions of heterozygous and homozygous segments. There will be a block of donor genome around the target allele which may, or may not, trace back to a common ancestor. The decrease in

donor allelic diversity in many introgression schemes will increase the possibility of linkage drag segments tracing back to the same ancestor.

Chapter Four showed that non-carrier chromosomes had the highest level of homozygosity of all cases studied. This was due to the high level of recipient genomic contributions (particularly from latter generations of backcrossing). However, these figures yield little information on the probability of IBD homozygosity. The allelic diversity of the latter recipient ancestral populations was shown to be high in Chapter Three. This will reduce the risk of alleles at a locus, within an individual, being homozygotes for identical alleles (IBD).

The study in Chapter Four showed that proportion of generation homozygosity on the carrier chromosome was affected by:

(i) Chromosome length: Initially donor generation homozygosity decreased as chromosome length increased. The decrease was attributed to the decreased proportional donor genomic contributions on smaller chromosomes (as seen in Chapter Two). This resulted in lower total generation homozygosity for longer chromosomes. However, for extreme chromosome lengths (8 or 18 Morgans) total generation homozygosity started to rise again, this time due to the increased genomic contributions from the later recipient generations.

(ii) Number of backcross generations: Total generation homozygosity was decreased as the number of backcross generations increased. This was due to decreased level of donor genome contributions, especially within the linkage drag segment.

(iii) Population structure had a minor effect on generation homozygosity. As population size increased, generation homozygosity on the carrier chromosome decreased. However, increasing offspring group size had a larger effect on the decrease in generation homozygosity.

It is the aim of this study to quantify the proportion of a carrier chromosome that is homozygous due to identity-by-descent (IBD homozygosity) due to donor alleles by examining the effect of (*i*) the number of backcross generations, (*ii*) the population size, (*iii*), offspring group size (*iv*) carrier chromosome length and (*v*) position of the target locus. The effect of distance from the target locus on the probability of donor IBD homozygosity at a locus is examined. The IBD homozygosity is related to the results of loss of allelic diversity described in Chapter Three.

5.2. Materials and Methods

The simulation programme followed the design of that used in previous chapters and described, in detail, in Chapter Two. The design of an introgression programme was given in Table 5.1, with notation in Table 2.2. The introgression of a marker for a desired allele was performed by crossing donor and recipient individuals to create F_1 individuals followed by T generations of backcrossing to new recipient individuals. Each of the N F_1 parents were simulated to have 2 different alleles at each locus. Recipient populations were backcrossed to this population for T generations. The backcross population was simulated as homozygous at each locus for ease of simulation and as the interest was in the IBD homozygosity due to donor alleles around the target locus.

The effect of population structure on IBD homozygosity was studied on the carrier chromosome. Parental population size (N mating pairs) were simulated to have values ranging from N =10 to 100 (with T = 6, n = 4, l = 1 Morgan, s = l/2). The number of offspring per mating (n) had values of 2 to 10 (with T = 6, N = 20, l = 1 Morgan, s = l/2). The number of offspring per mating was not set to 1 because the number with the desired genotype would decrease, on average, by a half each generation.

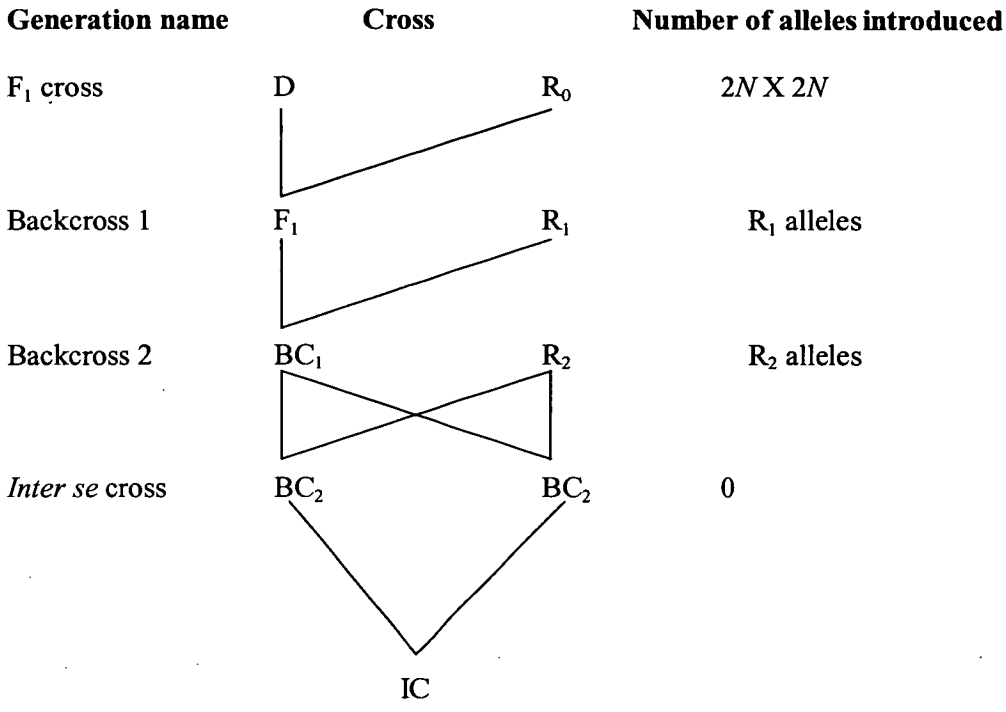

FIGURE 5.1. Diagram of the simulation design of an introgression scheme with two backcross generations (T = 2) showing population groups, generation names and the alleles introduced at each generation.

The number of backcross generations (T) was set at values ranging from 1 to 10 (with N = 20, n = 4, l = 1 Morgan, s = l/2).

The effect of carrier chromosome length (l) IBD homozygosity was studied bysimulating values of l ranging from 0.5 to 4 Morgans. These values were considered representative of chromosome lengths found in livestock species from chickens to cattle. An extreme chromosome length of 10 Morgans was also simulated. The location of the target locus on the carrier chromosome (s) ranged from the chromosome end to the centre of the chromosome, *i.e.*, 0 to $l/2$ Morgans (with $T = 6$, $N = 20$, $n = 4$, $l = 1$ Morgan).

For each set of parameters, 500 replicates were simulated and the following values were recorded from the genome of individuals post introgression (IC);

(i) Overall IBD homozygosity (proportion of loci on the carrier chromosome that are homozygous due to the fact that both donor alleles are identical by descent) on the carrier chromosomes due to donor alleles,

(ii) IBD homozygosity at a locus x cM away from the target locus on the carrier chromosome. IBD homozygosity due to overlapping linkage drag segments (that is IBD homozygosity due to donor alleles around the target locus),

(iii) Generational homozygosity (Proportion of loci that are homozygous due to the fact that both alleles trace back to the same generation) due to donor and recipient alleles at a given locus on the carrier chromosome was recorded,

(iv) Heterozygosity (proportion of loci that carrier alleles that do not trace back to the same ancestral generation) at loci on the carrier chromosome.

The estimates of homozygosity were calculated for individuals homozygous for the target gene.

5.3. Results

Table 5.1 shows the proportion of IBD homozygosity was much lower than the level of generation homozygosity shown in Chapter Four. IBD homozygosity due to donor alleles over the carrier chromosome (*i*) when described over the entire chromosome was low. However, the IBD homozygosity at the target locus (*ii*) was much higher (11-25%). Chapter Four showed that the number of donor alleles at the target locus decreased as the number of backcross generations increased. This would result in an increased probability of IBD homozygosity at the target locus. These results show that the main source of IBD homozygosity on the carrier chromosome is due to the high levels of donor IBD homozygosity at and around the target locus, or the linkage drag (*iii*).

TABLE 5.1. Donor IBD homozygosity across the carrier chromosome, IBD homozygosity at the target locus and donor generation homozygosity (N = 10 or 20, n = 4, T = 3 or 4, l = 1 Morgan, s = $l/2$) (standard deviation).

	***N*=20, *n*=4**		***N*=40, *n*=4**
	***T*=4**	***T*=6**	***T*=6**
(i) Donor IBD homozygosity	0.045	0.046	0.021
	(0.020)	(0.025)	(0.015)
(ii) IBD @ target	0.156	0.256	0.112
	(0.064)	(0.100)	(0.074)
(iii) IBD homozygosity due to linkage drag	0.036	0.038	0.017
	(0.020)	(0.022)	(0.012)
(iv) Donor generation homozygosity	0.229	0.167	0.164
	(0.150)	(0.107)	(0.106)

5.3.1. IBD homozygosity and heterozygosity at a locus on the carrier chromosome

The level of donor IBD homozygosity and donor generation homozygosity at individual loci dropped as distance from the target locus increased (Figure 5.2). For example, donor IBD homozygosity dropped from over 25% at the target locus to 8% 10 cM away to less then 0.3% 50 cM away. In fact, the donor IBD has reached half the value of the donor IBD at the target locus (half-life) only 6 cM away from the target locus. The dashed line on Figure 5.2 shows how the homozygosity due to recipient alleles tracing back to the same generation (generational homozygosity) is low at loci close to the target locus. In fact, the generational homozygosity remains low along the entire carrier chromosome, never rising above 4%. This low percentage is in line with the homozygosity observed on the carrier chromosome in Chapter Four. The natural log of the donor IBD declines in a linear fashion (not shown on the graph).

Figure 5.2 and Table 5.1 show that the majority of IBD homozygosity was due to high levels of donor IBD homozygosity close to the target locus. This was mainly due to the linkage drag segment (approximately 28 cM in the example in Figure 5.2). From this point forward results will refer to the effect of the parameters on donor IBD and generation homozygosity.

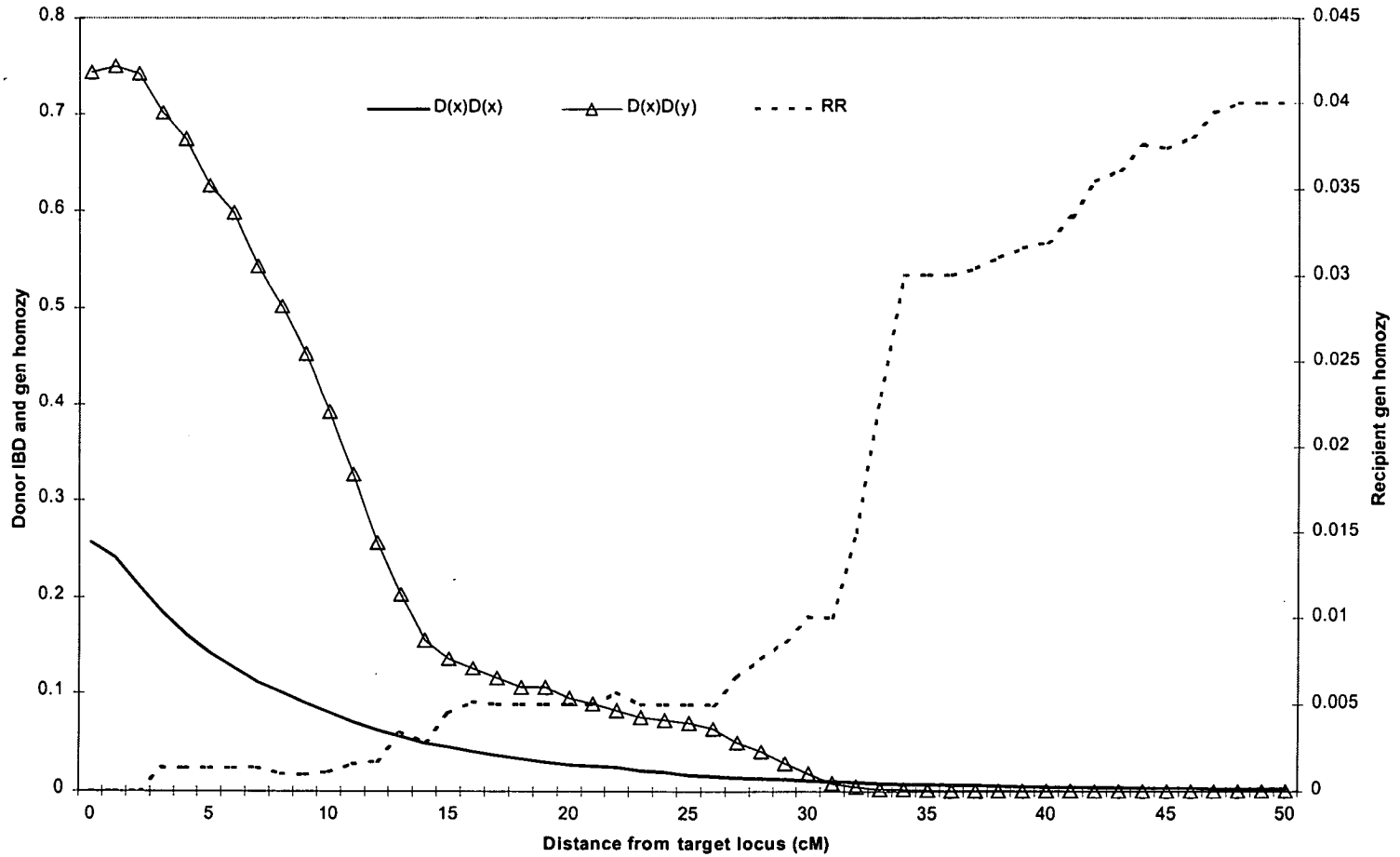

FIGURE 5.2. Distribution of donor IBD homozygosity (D(x)D(x)), recipient generational homozygosity (RR) and donor genome homozygosity (D(x)D(y)) at a locus moving away from the target locus (N=20, n = 4, T = 6, l = 1 Morgan, s = l/2).

5.3.2. Effect of number of backcross generations on IBD homozygosity

Donor IBD homozygosity on the carrier chromosome initially increased as the number of backcross generations (T) increased (Figure 5.3). However, there was a drop in donor IBD when $T = 10$. This decline continued as the number of backcross generations increased further (*e.g.*, donor IBD homozygosity = 0.036 when $T = 15$). IBD homozygosity due to the linkage drag segment contributed the majority of donor IBD homozygosity over all backcross generations. Figure 5.3 shows how the donor generation homozygosity decreased as the number of backcrosses increased. This was the opposite of the trend seen for donor IBD homozygosity. In reality the donor IBD at the target will approach 1 as the number of backcrosses increase. The donor homozygosity decline reaches its half life after 3 backcross generations. The natural log of the decline of donor IBD as backcross generations is shown as an insert in Figure 5.3. This also illustrates that the increase of IBD is initially fast but quickly asymptotes. However, increasing backcross generations *ad finitum* will eventually result in donor IBD proportionally decreasing again as the linkage drag decreases.

IBD close to the target locus also increased as the number of backcross generations increased (upper graph, Figure 5.4). However, the increase slowed down when the number of backcross generations was 6 or higher. Loci outside the limit of the linkage drag segment quickly approached the IBD homozygosity seen at a neutral locus. The increase in IBD homozygosity over backcross generations was attributed to the increased loss of alleles as the number of backcross generations increased (lower graph, Figure 5.4). The loss of donor alleles slowed down when $T > 6$.

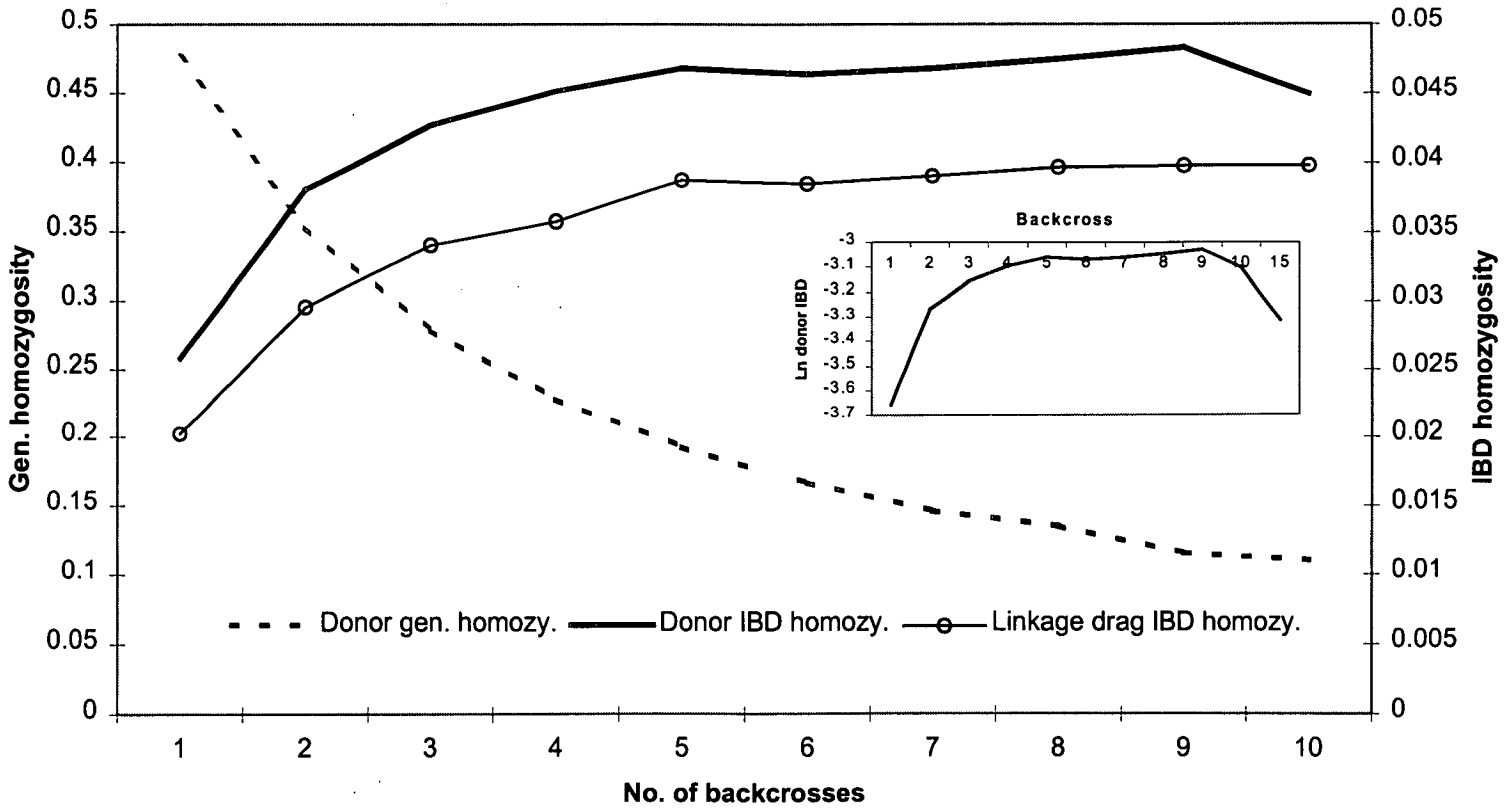

FIGURE 5.3. Effect of the number of backcross generations on donor generation homozygosity, donor IBD homozygosity and linkage drag IBD homozygosity when N=20, n = 4, l = 1 Morgans and $s = l/2$.

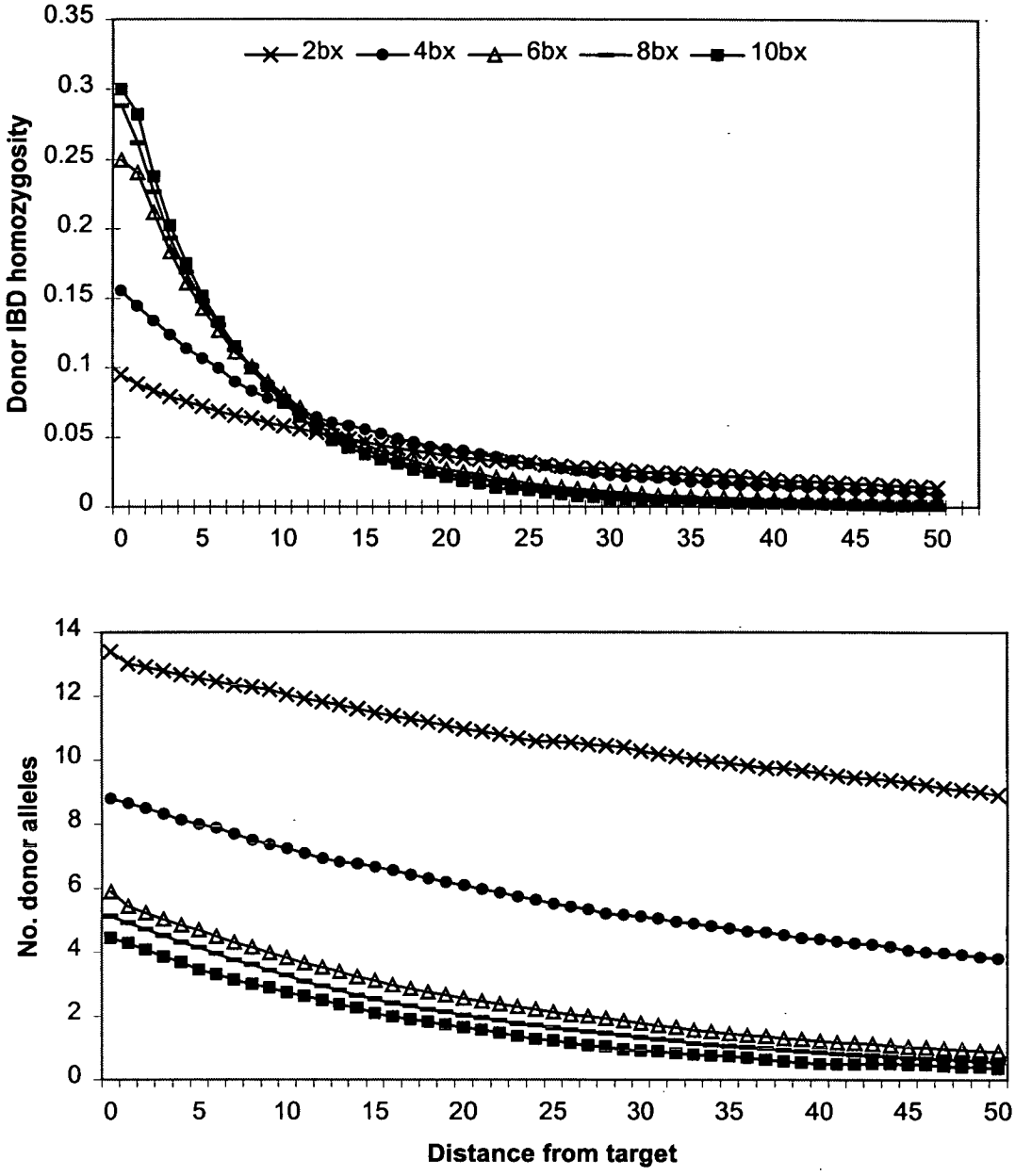

FIGURE 5.4. Effect of backcross generation (ranging from 2 – 10 bx) on the distribution of donor IBD homozygosity (top graph) across the carrier chromosome and number of donor alleles present in the final population homozygous for the target allele (bottom graph).

The decline in IBD homozygosity was much slower when the number of backcross generations was low (upper graph, Figure 5.4). Chapter Two showed how the linkage drag decreased in length as the number of backcross generations increased.

This linkage drag made up a larger part of the carrier chromosome when the number of backcross generations was low. This resulted in an increased area of potential overlap causing donor IBD homozygosity and therefore, decreased at the same rate of decline of donor IBD homozygosity.

5.3.3. Effect of chromosome length on IBD homozygosity

Donor IBD homozygosity and donor generation homozygosity decreased as the carrier chromosome length (l) increased (Figure 5.5). There was a linear decline in the natural log of the donor IBD (insert in Figure 5.5). IBD homozygosity due to the linkage drag segment contributed nearly all donor IBD homozygosity over all chromosome lengths studied. The trend of decline in donor generation homozygosity as the number of backcrosses increased exactly matched the trend seen for donor IBD homozygosity. The declines seen in IBD and homozygosity are completely mirrored by the decline in the length of the linkage drag (not shown). This suggests that the decrease in IBD is because of the decline in donor genome around the target locus. Figure 5.5 shows that there is rapid elimination of donor IBD outside that around the target locus as chromosome length increases and that linkage drag IBD represents almost all the donor IBD.

There was little difference in the IBD homozygosity at a locus x cM away from the target over all chromosome lengths studied. The IBD homozygosity graph approached an asymptote to a level of homozygosity seen at a neutral locus.

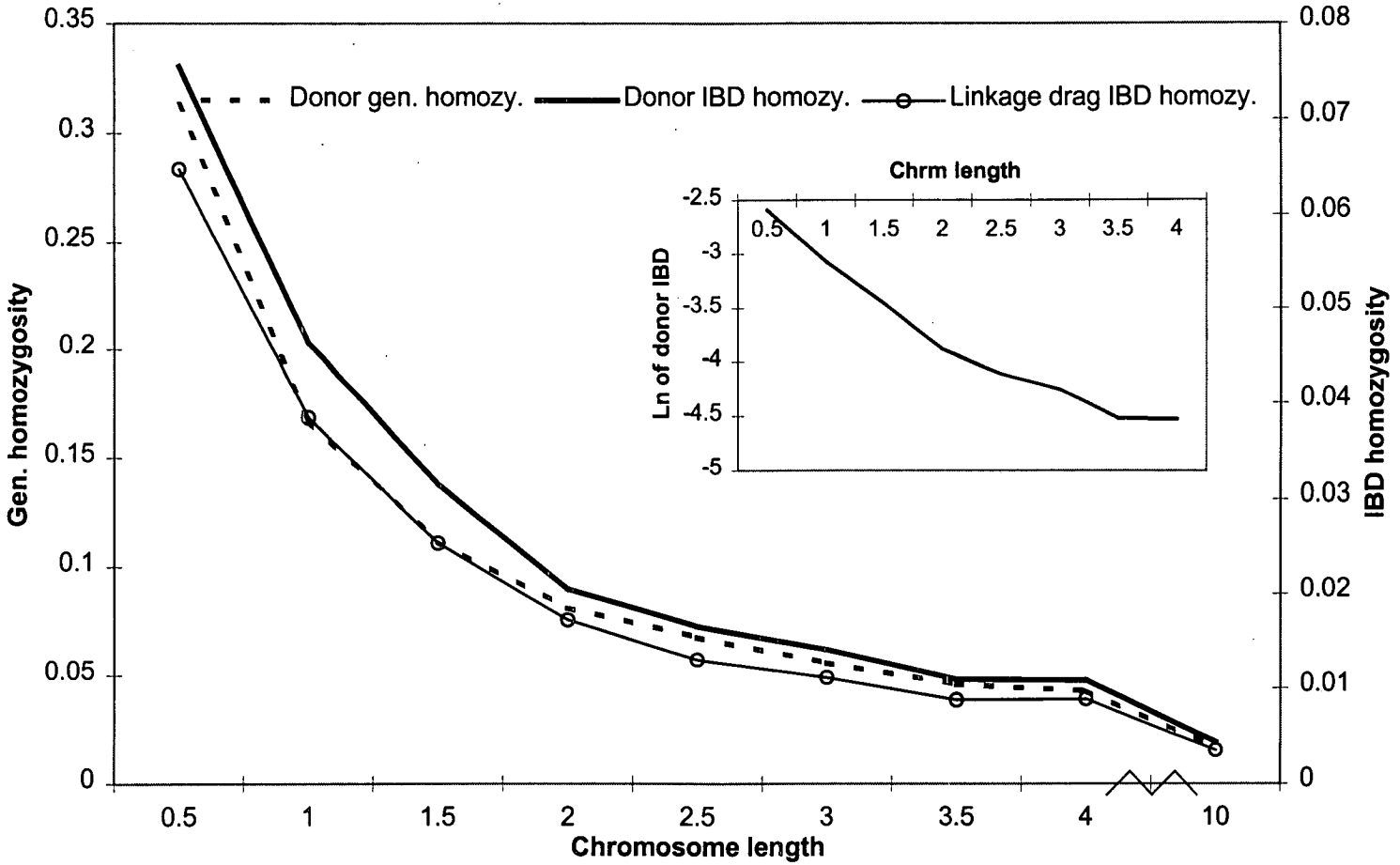

FIGURE 5.5. Effect of chromosome length on donor generation homozygosity, donor IBD homozygosity and linkage drag IBD homozygosity when N =20, n = 4, T = 6 and $s = l/2$.

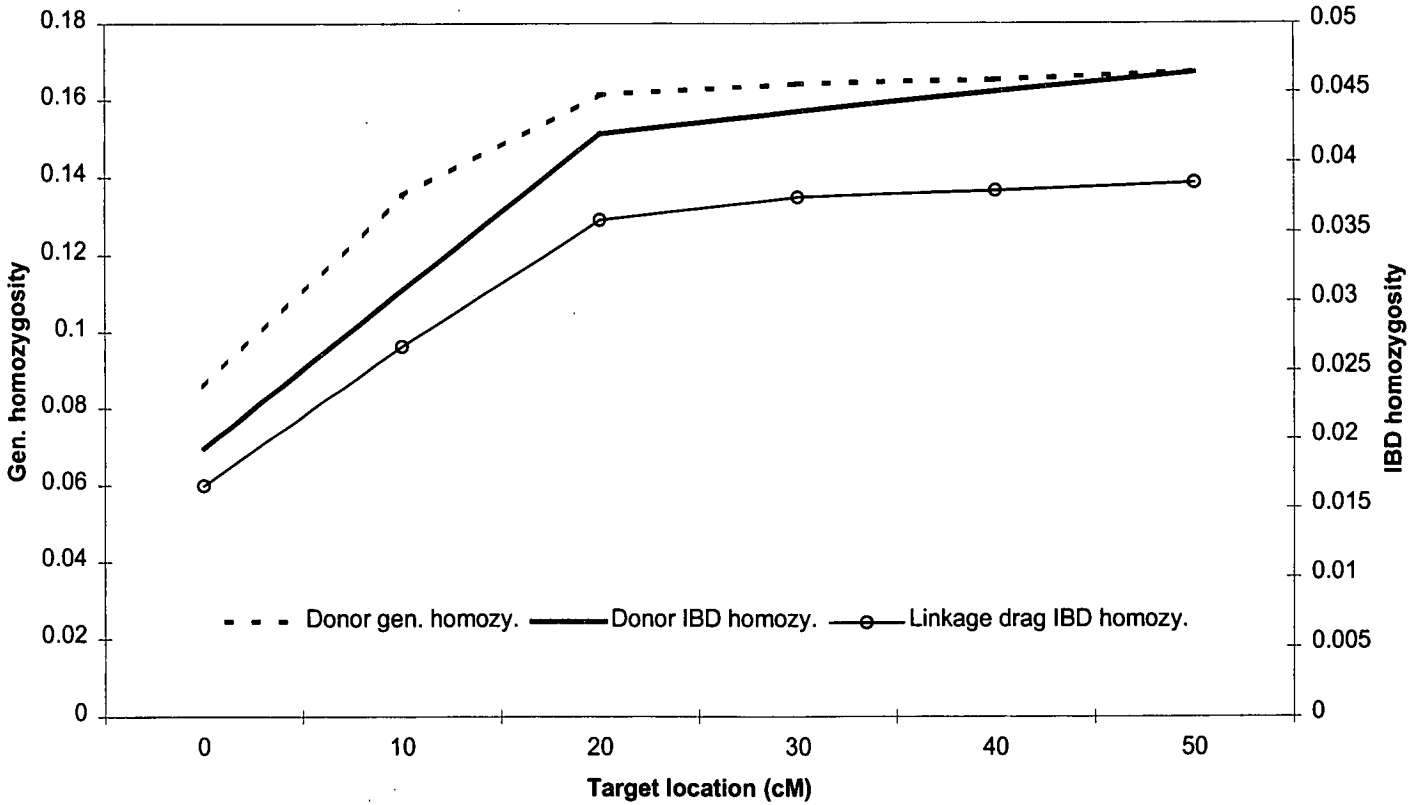

FIGURE 5.6. Effect of target locus position (s = 0 to 50 cM) on donor generation homozygosity, donor IBD homozygosity and linkage drag IBD homozygosity when N=20, n = 4, T = 6 and l = 1 Morgan.

5.3.4. Effect of target locus position on the carrier chromosome on IBD homozygosity

Donor IBD homozygosity and donor generation homozygosity decreased as the position of the target locus (s) approached the chromosome edge (Figure 5.6). IBD homozygosity due to the linkage drag segment made up a larger percentage of total IBD homozygosity as the target locus moved to the chromosome end. In theory the donor IBD due to linkage drag for an edge target locus should be twice that of a centrally placed (*i.e.*, the linkage drag is half that of the centrally placed target). In this example the proportion of IBD on the carrier chromosome is just over twice. However, the standard error across replicates is quite large for this statistic and also gets larger as the target locus approaches the centre. This is due to the large variation associated with the linkage drag within a population at any given time. The loss of alleles on the carrier chromosome was slightly lower when $s = 0$ M but this was not significant (results not shown).

5.3.5. Effect of population size on IBD homozygosity

Donor IBD homozygosity and IBD homozygosity due to the linkage drag decreased at the same rate as population size (N) increased (Figure 5.7). The decline in IBD homozygosity slowed down as when the population size higher than 70. The half life of the decline in IBD homozygosity as population size increased was quite low and reached this point when the population size was between 20 and 30 mating pairs. A slight decrease in donor generation homozygosity was seen but not at the same rate of decline of donor IBD homozygosity.

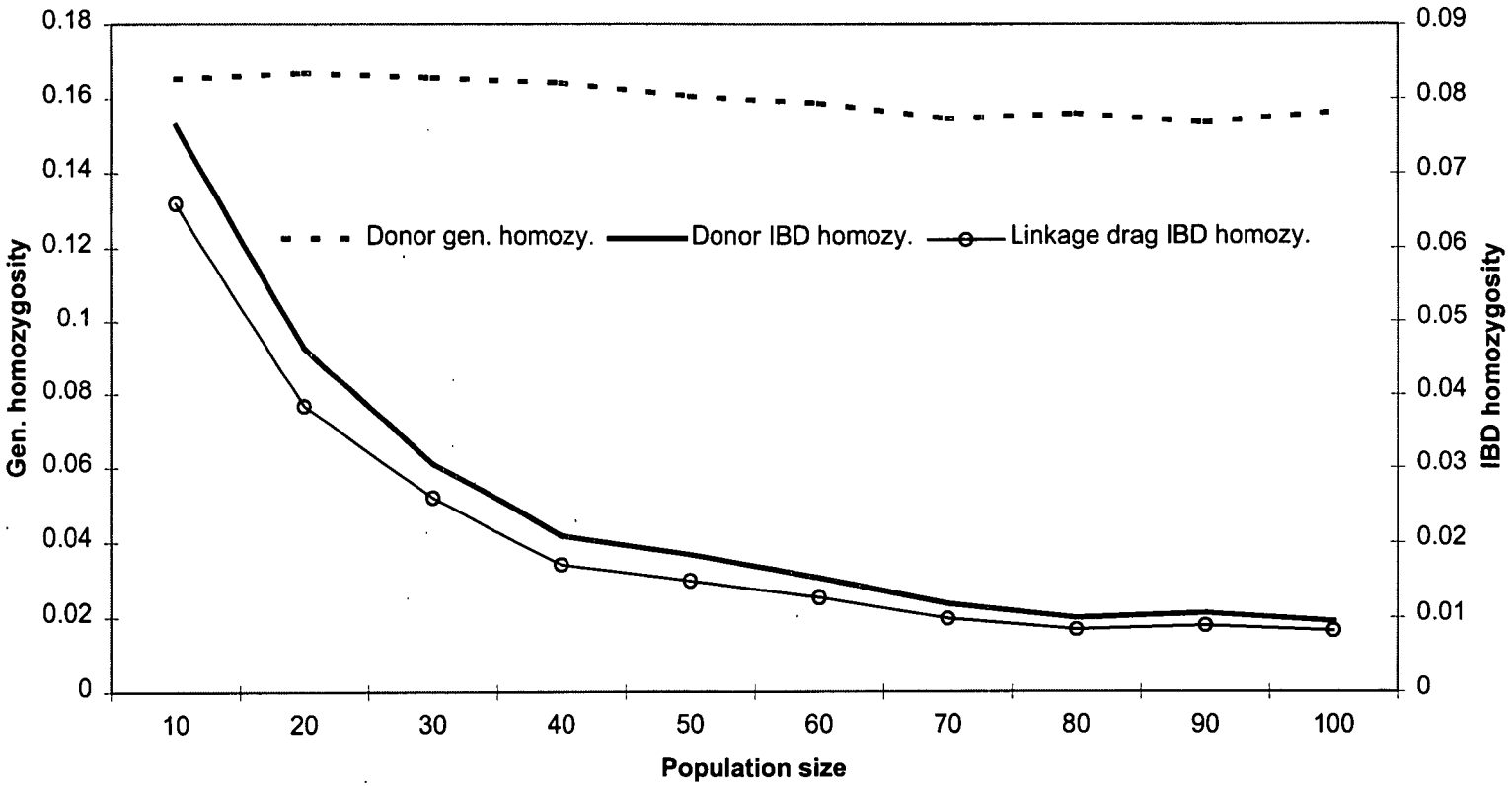

FIGURE 5.7. Effect of population size on donor generation homozygosity, donor IBD homozygosity and linkage drag IBD homozygosity when $n = 4$, $T = 6$ and $l = 1$ Morgan and $s = l/2$.

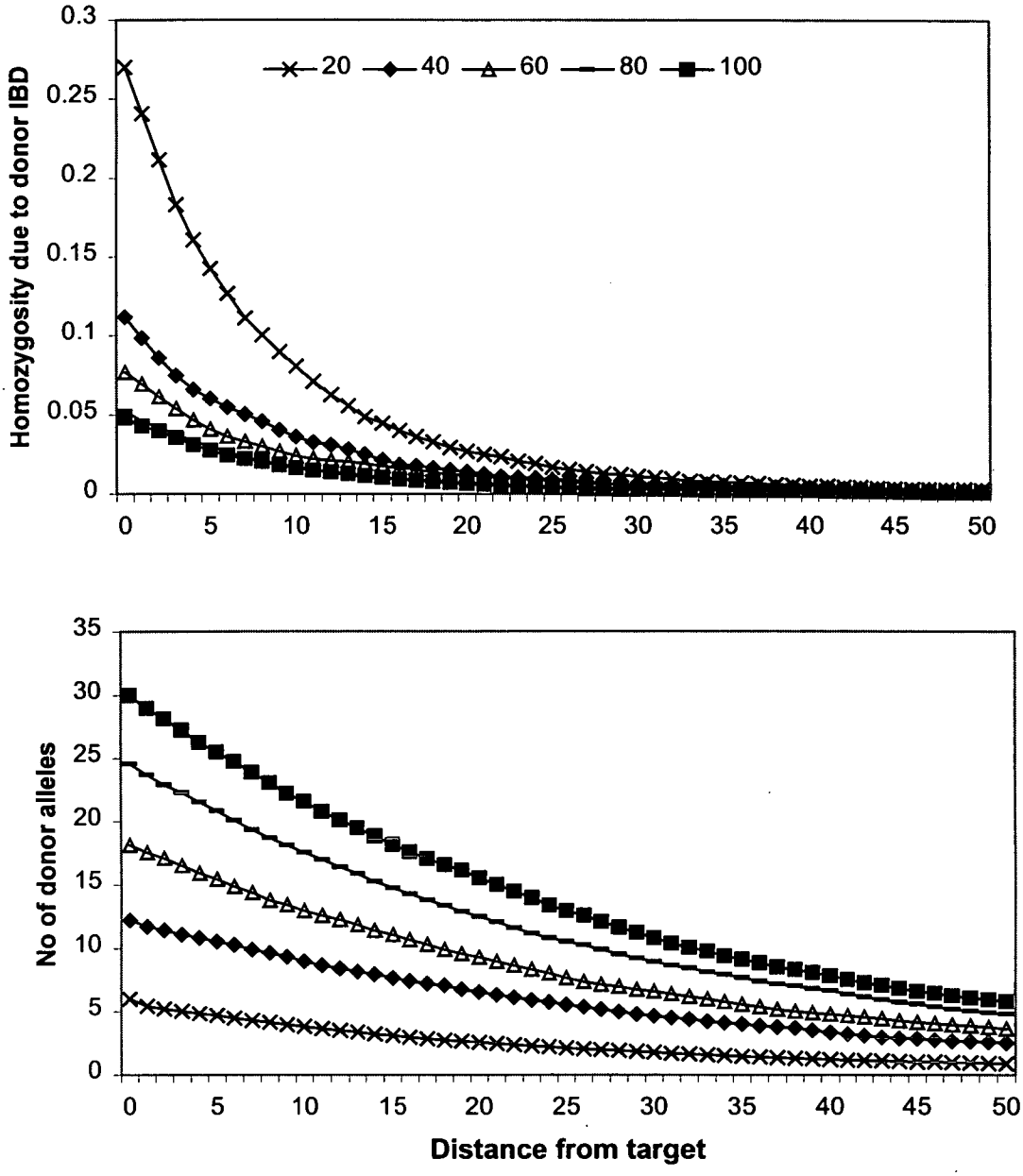

FIGURE 5.8. Effect of population size (*N*) on the distribution of donor IBD homozygosity (top graph) across the carrier chromosome and number of donor alleles at a locus in the final population homozygous for the target allele (bottom graph).

There was dramatic increase in donor IBD homozygosity at loci close to the target locus when the population size was low (upper graph, Figure 5.8). This increased IBD homozygosity declined quickly as the distance from the target locus increased to approach the level seen at a locus in other population sizes. The increased IBD

homozygosity at a locus close to the target was attributed to fewer donor alleles remaining at a locus when N was low (lower graph, Figure 5.8).

5.3.6. Effect of number of offspring per mating on IBD homozygosity

It is important to point out that when $n = 2$ in 9% of the replicates all donor alleles at the target locus were lost during introgression and these were ignored. Donor IBD homozygosity was relatively constant as the number of offspring per mating (n) increased (Figure 5.9). Donor generation homozygosity declined at first and then asymptoted. There was an increase in IBD homozygosity due to linkage drag when the number of offspring increased from 2 to 4. This was attributed to an increase in the number of carrier produced per family when $n > 2$. This resulted in an increase in the variance of the sampling of individuals from amongst carrier as one of N parents of the next backcross generation. When $n = 2$, on average, only one carrier of the target allele is produced per family. This resulted in a decreased sampling variance within families because, on average, there will be one carrier per family and therefore one offspring per family selected for the next generation. In reality, increasing population size reduces IBD, independent of n.

There was decreased IBD homozygosity at loci close to the target locus when $n = 2$ but little difference for all other values of n (upper graph, Figure 5.10). The decline in IBD homozygosity at a locus was slower for $n = 2$ then for all other n. The increased IBD homozygosity at a locus close to the target when $n = 2$ was attributed to the fact that fewer donor alleles remained loci on the carrier chromosome when the number of offspring per mating was 2 (lower graph, Figure 5.10).

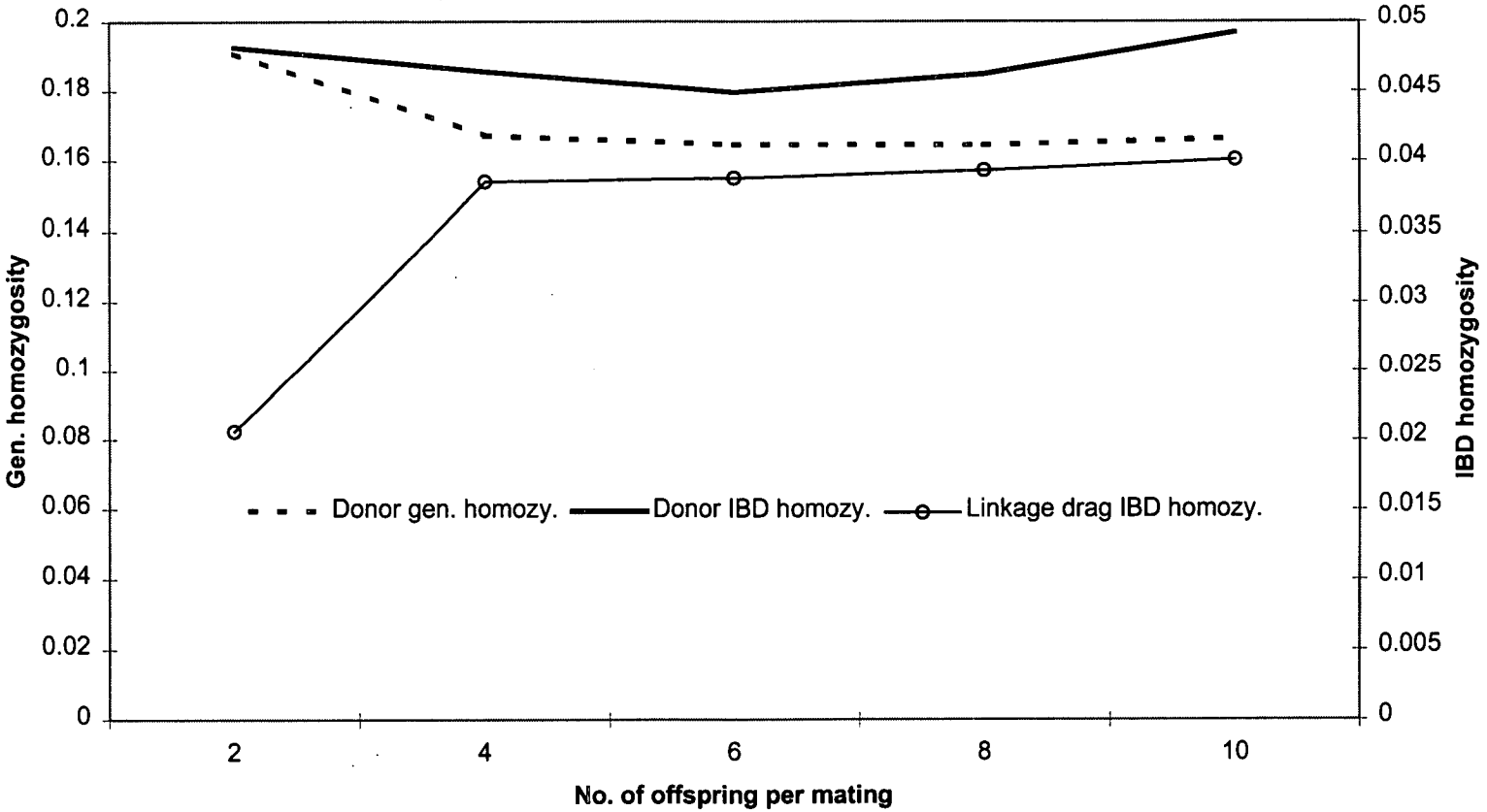

FIGURE 5.9. Effect of number of offspring per mating on donor generation homozygosity, donor IBD homozygosity and linkage drag IBD homozygosity when $N = 20$, $T = 6$ and $l = 1$ Morgan and $s = l/2$.

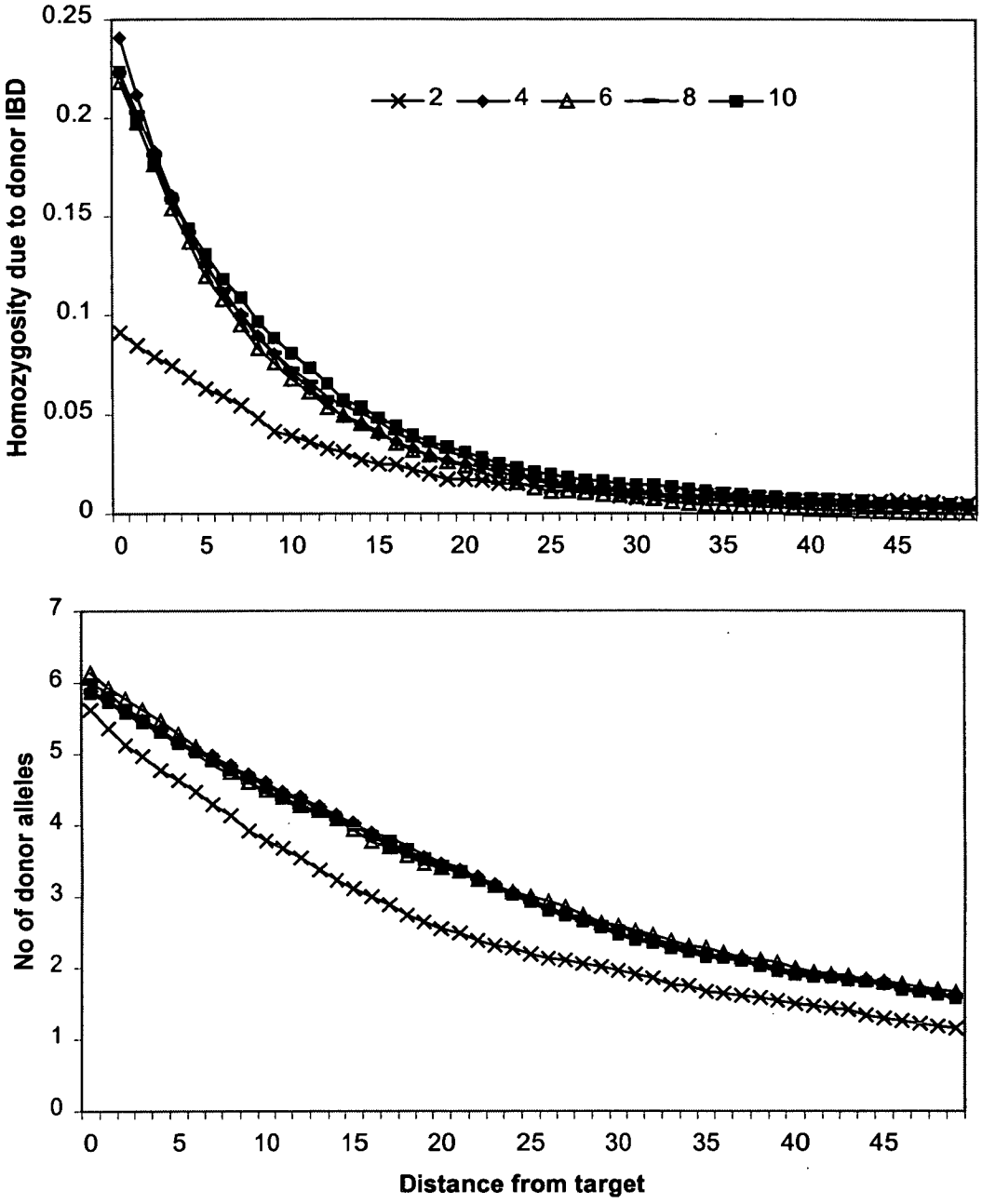

FIGURE 5.10. Effect of chromosome length on the distribution of donor IBD homozygosity (top graph) across the carrier chromosome and number of donor alleles (bottom graph).

5.3.7. Prediction of IBD homozygosity on the carrier chromosome

The prediction of IBD homozygosity on the carrier chromosome is more complicated than that of a neutral locus due to the selection pressure on the target locus. It is

necessary to account for the joint distribution of Mendelian sampling terms between all loci and the target locus within an individual. The prediction of IBD homozygosity at the target locus is the probability of a donor allele surviving until the last backcross generation (Chapter Three) and the probability of picking two individuals carrying the same ancestral allele at the target locus. This is complicated by the many chains of the transmission of alleles throughout the pedigree of the introgression programme.

5.3.8. Non-carrier chromosomes

There was no significant difference in total generation homozygosity across all population sizes, offspring group sizes and number of backcross generations studied on the non-carrier chromosomes. The total generation homozygosity was approximately 0.33, with over ¾ of that coming from generation homozygosity due to R_6 in all cases. This was examined in more detail in Chapter Four. Total IBD homozygosity decreased as population size increased (Table 5.2) dropping from 2% when $N = 20$ to 0.4% when $N = 100$. The main cause for this decrease was due to the drop in IBD homozygosity from the recipient populations used in the last few backcross generations (*e.g.*, $R_3 - R_6$ when $T = 6$, as shown in Table 5.2). The ancestral generations used in the F_1 cross and recipient populations used in first few backcross generations contributed very little to the total IBD homozygosity.

TABLE 5.2. The effect of population size (N) on IBD homozygosity on non-carrier chromosomes from the last four recipient population ($R_3 - R_6$) and total IBD homozygosity when $n = 4$ and $T = 6$.

	R_3	R_4	R_5	R_6	TOTAL
$N = 20$	0.001	0.002	0.005	0.010	0.021
$N = 40$	0.001	0.001	0.002	0.005	0.009
$N = 60$	0.001	0.001	0.002	0.003	0.006
$N = 80$	0.000	0.001	0.001	0.003	0.005
$N = 100$	0.000	0.001	0.001	0.002	0.004

Total IBD homozygosity increased only as offspring group size increased from 2 to 4 offspring per mating, increasing from 1.4% to 2.1% (Table 5.3). The main cause for this increase was due to the rise in IBD homozygosity from the recipient population in the last backcross generation (R_6 in Table 5.3). The IBD homozygosity due to other recipient populations was relatively constant over offspring group size. The ancestral generations used in the F_1 cross and recipient populations used in first few backcross generations contributed very little to the total IBD homozygosity.

TABLE 5.3. The effect of the number of offspring per mating (n) on IBD homozygosity on non-carrier chromosomes from the last four recipient population $R_3 - R_6$ when $N = 20$ and $T = 6$.

	R_3	R_4	R_5	R_6	TOTAL
$n = 2$	0.001	0.001	0.005	0.005	0.014
$n = 4$	0.001	0.002	0.005	0.010	0.021
$n = 6$	0.002	0.004	0.004	0.010	0.020
$n = 8$	0.002	0.002	0.006	0.011	0.021
$n = 10$	0.001	0.003	0.005	0.012	0.023

Total IBD homozygosity ranged from 1.9% - 2.1 % over all the simulations varying the number of backcross generations (Figure 5.11). Nearly ½ of the total IBD homozygosity was attributable to the recipient population used in the last backcross generation (R_T). The IBD homozygosity from R_{T+1} was approximately 1% across all

values of T studied. This trend continued could be seen for earlier backcross generations (*e.g.*, R_{T-1} when $T = 6$ has the same value as R_{T-1} when $T = 10$).

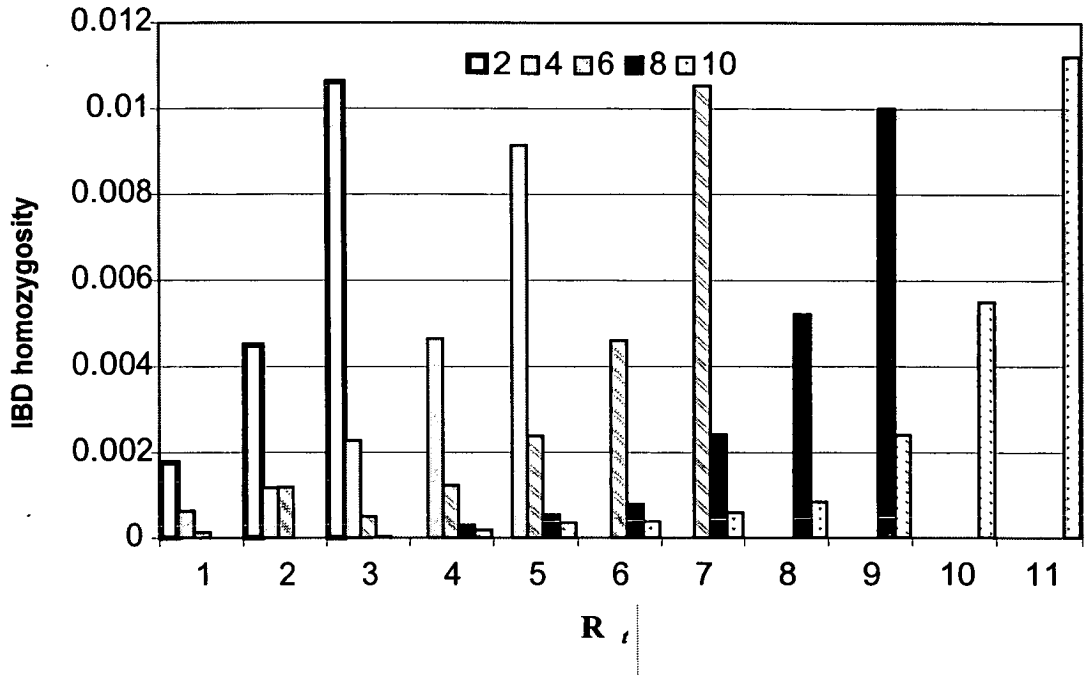

FIGURE 5.11. The effect of the number of backcross generations (T) on the IBD homozygosity attributable to each ancestral recipient population.

5.4. Discussion

IBD homozygosity after introgression was affected by many of the factors examined, namely, population size, number of offspring per mating, number of backcross generations, carrier chromosome length and position of the target locus. Non-carrier chromosomes had lower donor IBD homozygosity compared to carrier chromosomes due to the influence of the selection on the target locus on the carrier chromosome. IBD homozygosity decreased as the carrier chromosome length increased. In reality the average number of loci that were homozygous due to IBD alleles remained the same over chromosome lengths, but when expressed as a proportion of chromosome

length the value decreased. The effect of position of target locus only began to have a significant effect when the position was close to the chromosome end. This was attributed to the reduction in the linkage drag length when a target locus was at or near to a chromosome end.

5.4.1. Non-carrier chromosomes

The IBD homozygosity on non-carrier chromosomes was decreased as population size increased and as the number of offspring initially increased. This decrease can be attributed to the increased numbers of alleles when population size increased. Each recipient population contributed N new alleles to a locus. Therefore, as N increased the probability of alleles at a locus tracing back to the same ancestor decreased. The total IBD homozygosity was shown to be very low and ranged from 0.4% to 2.3% over all simulations. At least 50% of the total IBD homozygosity was due to R_T IBD homozygosity.

Increasing population size (or the number of alleles at a locus) reduces the probability of IBD homozygosity at a neutral locus. The reduction in the probability was dramatic when population size increased from 20 to 40 reducing the IBD homozygosity by over a half. For higher population sizes, the decrease was not as large. It is reasonable to assume that a recipient population will not introduce as many as N new alleles to the introgression programme, especially for larger values of N. These results suggest that increasing population size above 60 does not have a big effect on reducing IBD homozygosity.

Loci at non-carrier chromosomes can be considered as neutral loci in the genome and are the inbreeding coefficient at a neutral locus. This study has shown that the level of IBD homozygosity at a neutral locus is low. This means the inbreeding coefficient of the population homozygous population would be low if examining homozygosity at a neutral locus. This study simulated $2N$ donor and $2N$ recipient alleles at locus in the F_1 cross. If the cross was a diverse line cross it would be reasonable to assume that, these lines would be fixed for alternative divergent alleles. This assumption is not valid if the cross is between lines that have diverged in the recent past. In this latter scenario, a limited number of alleles may exist at a locus and some alleles may be at a high frequency. In addition, the two lines may contain the some of the same alleles, especially at a neutral locus. The study also assumed that each recipient population used for backcrossing introduced N new alleles to a locus. However, these recipient populations may have alleles in common. Therefore, the estimates of IBD homozygosity (and therefore inbreeding) may be lower than the potential IBD homozygosity in a real livestock introgression programme.

5.4.2. Effect of the number of backcross generations on IBD homozygosity

The results showed an increase in IBD homozygosity over the carrier chromosome when the number of backcross generations increased from 1 to 4 generations. With only one backcross generation the IBD homozygosity on the carrier chromosome approached that of a non-carrier chromosome (~ 2%). By backcross generation 4, it had doubled to 4%. Increasing the backcross generations after this had little effect on the IBD homozygosity until $T = 10$ when it began to drop off. A similar trend was seen for IBD homozygosity due to linkage drag. The decrease in linkage drag

length was shown to slow down after the first few generations of backcrossing in Chapter Two and Figure 5.4. The loss of diversity of alleles also slows down after the first few backcross generations (Chapter Four and Figure 5.4). This could cause the static state of IBD homozygosity seen between 4 and 9 backcross generations.

Increasing the number of backcross generation above 10 caused the donor IBD homozygosity to drop again. This was due to the loss of diversity of donor alleles near to the target locus and the reduced linkage drag. Increasing the number of backcross generations increased the recombination events on the carrier chromosome. Therefore, the linkage drag continued to be broken down whilst the number of donor alleles remaining approached 1 across the entire population. Donor IBD homozygosity on the carrier chromosome is unlikely to ever reach zero because the population homozygosity for the target allele will always be homozygous at the target locus. The recipient IBD homozygosity continues to increase at backcross generations increase (described later). Ultimately the IBD on the carrier chromosome will be 1 at the target locus and 0 outside that.

The results suggest that fewer backcross generations minimise the risk of IBD homozygosity and inbreeding and maximises the allelic variation across the genome. Even though the donor IBD homozygosity on the carrier chromosome was much higher than that seen on a non-carrier chromosome, this will be diluted when averaged out across the genome. Increasing the number of backcross generations reduces donor contamination in the genome and therefore genetic lag. The increased

inbreeding experienced on the carrier chromosome with increased backcross generations could be traded off against the reduction of genetic lag.

5.4.3. Effect of population structure on IBD homozygosity

There was dramatic decline in donor IBD homozygosity on both the carrier and non-carrier chromosome as the population size increased. The decline in IBD homozygosity was asymptotic as the number of mating pairs increased. These results indicated that there is little benefit, in terms of reduction in inbreeding, of increasing the population size above 50. This is consistent with the FAO guidelines for an effective population size in straight bred conservation programs where they recommend an effective population size of 50 in conservation breeding programmes (FAO, 1998). There was a small increase in IBD homozygosity on the carrier chromosome when population increased from 2 to 4 as was seen on the non-carrier chromosome.

The most important population parameter on decreasing IBD homozygosity appears to be population size as the number of alleles at a locus is likely to be increased in larger population sizes. Take the example of a cattle introgression programme to introduce the polled gene from a genetically polled population to a horned population (discussed in more detail later). The cattle genome can be approximated to 30 Morgans with the polled gene at the end of chromosome 1 (1.5 Morgans). The IBD homozygosity decreases on carrier and non-carrier chromosome (non-carrier chromosome results taken from simulation studies) when population size increases thus resulting in a low probability of genome wide IBD homozygosity. Increasing

the number of backcross generations increases IBD homozygosity on the carrier chromosome but has little effect on non-carrier chromosomes. Therefore, there is little change in the genome wide IBD homozygosity over backcross generations.

TABLE 5.5. IBD homozygosity on carrier, non-carrier and cross of the genome of various forms of introgression scheme to introduce the polled allele in cattle into a horned population.

When T=6 and n = 4	N = 20	N = 40	N = 60	N = 80	N = 100
Non-carrier	0.021	0.009	0.006	0.005	0.004
Carrier	0.074	0.032	0.023	0.017	0.014
Genome	0.024	0.010	0.007	0.006	0.005

When N = 20 and n =4	T = 2	T = 4	T = 6	T = 8	T = 10
Non-carrier	0.019	0.021	0.021	0.020	0.020
Carrier	0.051	0.066	0.075	0.084	0.086
Genome	0.021	0.023	0.024	0.023	0.023

IBD homozygosity is an important consideration in gene introgression programmes due to its relationship with inbreeding. The results of this Chapter and Chapter Three showed how the variation, at a genome level, was reduced greatly in smaller populations. The reduced genomic variation could result in non-viable breeding population with a high inbreeding coefficient. This may not be a consideration in plant introgression but should be a major consideration when designing livestock introgression programme. The number of animals that can be maintained during a livestock introgression programme may be low due to the cost of maintenance and genotyping. However, this cost must be balanced against the potential inbreeding that may occur by reducing the population. This study suggests that a population of more then 40 mating pairs will keep the inbreeding coefficient below 1% in the population homozygous for the target allele.

CHAPTER SIX

APPLICATIONS OF THEORETICAL PREDICTIONS AND SIMULATION RESULTS: LINKAGE DRAG AROUND THE BOOROOLA LOCUS AND GENETIC LAG AND INBREEDING IN LIVESTOCK INTROGRESSION PROGRAMMES

6.1. Introduction

Chapter Two described and validated the description of genomic contributions in a gene introgression programme. Predictions of genomic contributions can be used in a number of ways to aid the development of practical and feasible livestock introgression programmes. Introgression of genes into a commercial line from an inferior line can result in a genetic lag for commercial traits. This Chapter describes how the predictions developed in Chapters Two and Three and the estimates in Chapter Five can be used to examine the linkage drag, genetic lag, loss of variation and homozygosity in different types of introgression programmes.

The aim of this Chapter is to illustrate the application of the predictions and estimates of the previous Chapters to practical livestock introgression schemes. This will show how the earlier work of this thesis can be easily used to help design the optimum introgression scheme for different circumstances. The examples are:

1. Linkage drag around the Booroola locus: This example predicts the linkage drag after the introgression of the Booroola allele in a sheep population. It is shown that an unfavourable allele was dragged into the population within the Booroola linkage drag segment.
2. Genetic lag in cattle introgression schemes: This example applies the predictions of genomic contributions in Chapter Two to the estimation of the genetic lag after the introgression of a polled allele from the British Friesian to a commercial Holstein population.
3. Comparison of methods to estimate genetic lag: This example compares the method of estimating genetic lag in the above example to a previous

methodology using the example of the introgression of the double muscling gene into a Holstein Friesian population.

4. Identity-by-descent in a Meishan X Large White pig backcross population: This example uses methods and estimates from Chapters Four and Five to examine the allelic diversity and identity-by-decent in a pig backcross population.

6.2. Example 1: Linkage drag around the Booroola locus

6.2.1. Introduction

The linkage drag segment is the block of DNA around the target locus that traces back to a donor ancestor. The linkage drag segment may incorporate other less favourable alleles and drag them into the commercial population, the risk of which is related to its length. The prediction of the expected length of linkage drag in backcross breeding programmes, developed by Stam and Zeven (1981), was verified by practical examples in wheat (Zeven *et al.*, 1983) and barley (Brown *et al.*, 1989) and a rare livestock example in sheep (Walling *et al.*, 2000). Experimental observations give a general validation of the accuracy of the theoretical prediction. However, their results are open to errors depending on the extent of DNA information and the size of the studies as there may not be enough markers or individuals to estimate the linkage drag in the population accurately.

Chapter Two developed aspects of the work of Hanson (1959) and Stam and Zeven (1981) to predict the recipient contributions and validate the prediction for linkage drag via simulation studies. The simulation studies were based on introgression programme designs that would apply to livestock populations. This example

examines the impact of the linkage drag around the Booroola locus in a recipient Merino population. The range of linkage drag segments lengths for the population was examined, and if any unfavourable donor alleles were dragged into the recipient population during introgression.

6.2.2. Materials and methods

The expectation of the length of the linkage drag segment after t generations of backcrossing for a locus in position s can be given as $\delta(s, t)$, the length proximal to s is $t^{-1}(1 - e^{-ts})$ and distal to s is $t^{-1}(1 - e^{-t(l-s)})$ (Equation 2.1, for definition of terms see Table 2.2). This prediction was validated by simulation in Chapter Two. Equation 2.1 will be used to predict the linkage drag around the Booroola allele.

The Booroola gene is a single autosomal mutation (*FecB*) and is recognised for increasing sheep fecundity (ovulation rate and litter size) and found in specific selected lines of Australian Merino sheep (Davis *et al.*, 1982; Piper *et al.*, 1985). The *FecB* mutation has been mapped to sheep chromosome 6 (Montgomery *et al.*, 1994) and cloned (Wilson *et al.*, 2001) and, on average, increases litter size by 1-2 extra lambs with each *FecB* allele. Female Booroola carriers may be unable to support the larger litters. If this is the case, then lamb survival becomes an important issue for the producer. Producers have reported that Booroola carriers produce lighter animals than contemporary non-carriers.

Walling *et al.* (2000) described a study to investigate evidence of quantitative trait loci (QTL) affecting weight traits (weaning weight and mating weight) in sheep.

Animals for this study came from a Booroola backcross flock where homozygous Booroola rams were mated with non-carrier ewes to create a heterozygous F1 population. Heterozygous female offspring were backcrossed to non-carrier rams for up to 4 generations. Animals from half-sib families, created by mating heterozygous Booroola rams to non-carrier females, were also used in this study. Phenotypic measurements of birth rank, weaning weight (WWT) and mating weight (MWT) were taken. Animals were genotyped for 21 markers on chromosome 6, which was 2.3 M in length. A linkage analysis of the markers was carried out using Cri-map (Green *et al.*, 1990). The QTL analysis used an extension of the least squares regression model (Haley *et al.*, 1994) to account for more complex pedigree structures (Dodds, 1999). A pleiotropic effect of the Booroola QTL on weight measurements is discounted in this model.

Table 6.1 shows a significant effect for both WWT and MWT detected on chromosome 6 (single QTL model) from the study of Walling *et al.*, 2000. However, after pre-adjusting the MWT for WWT no significant effect was detected. The Booroola gene was at 77 cM and is in linkage disequilibrium with the allele, which decreases weaning weight (WWT) by an average of 1.4 kg (Walling *et al.*, 2000).

TABLE 6.1. Summary of results for one QTL affecting live weight on chromosome 6. All effects given as the allele substitution effect (kg). The location of the Booroola gene was 77 cM.*

Trait	Position (cM)*	95% CI (cM)	Effect (se) (kg)	F-ratio	Probability
WWT	98.0	86-127	-1.37 (0.43)	10.18	0.002
MWT	65.0	3-182	-1.23 (0.48)	6.50	0.011

*From Walling *et al.*, 2000.

6.2.3. Results

The predicted linkage drag around the Booroola allele was examined using the details of the Walling *et al.* (2000) study. Table 6.2 shows the predicted linkage drag around the Booroola locus for up to 4 backcross generations, encompassing the situation of all animals in the pedigree. The linkage drag is described in terms of total length of the linkage drag around the target locus and the length of intact donor (Booroola) genome proximal and distal of the target locus. For example, the linkage drag proximal to the Booroola gene after 1 backcross generation would be $1^{-1}(1 - e^{-1(0.77)}) = 0.537$ Morgans and distally = 0.788 Morgans resulting in a total linkage drag length of ~ 1.3 Morgans.

TABLE 6.2. Total expected length (in centiMorgans) of linkage drag around the Booroola gene (77 cM on chromosome 6 which is 232 cM) and the length proximal and distal to the Booroola locus.

	Linkage drag	Length distal	Length proximal	Location
BC1	132.5	53.7	78.8	23.3 – 148.8
BC2	87	39.3	47.7	37.7 – 124.7
BC3	63	30	33	47 – 110
BC4	48.8	23.9	24.9	53.1 – 106

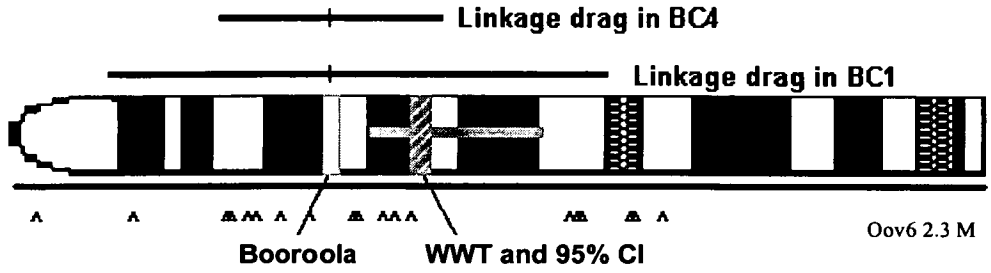

FIGURE 6.1. Diagrammatic representation of chromosome 6 and the respective positions of the Booroola gene (solid grey) and the WWT locus (hatched grey). The coverage of the linkage drag segment around the Booroola gene after 1 (BC1) and 4 (BC4) backcross generations are also shown.

Figure 6.1 shows that the weaning weight locus falls within the linkage drag of the Booroola gene with WWT being 22 cM distal to the Booroola locus. After 4 backcross generations the predicted linkage drag incorporates the estimated position of the undesirable weaning weight locus and only part of the 95% confidence interval of the position of the weaning weight QTL. Only a small proportion of the animals have gone through 4 generations of backcrossing to non-carrier parents. Therefore, only a small proportion of the animals would have a linkage drag that had reduced to this level. This prediction is likely to be too low. After 1 backcross generation the linkage drag segment proximal to the Booroola locus incorporates the genome up to position ~ 1.5 Morgans.

The predictions show that the linkage drag is not symmetrical around the Booroola locus in the first few generations. The length of donor genome distal to the locus is longer then proximal donor segment in the early generations. The linkage drag becomes more symmetrical with each additional backcross generations as the higher number of recombination events allows the proximal linkage drag segment to be

broken down. After 4 backcross generation the segment proximally has not been broken down sufficiently to lose the undesirable weaning weight locus.

6.2.4. Discussion

The weaning weight locus was shown to be within the distal linkage drag segment of the Booroola locus. One method of removing the weaning weight allele is to carry out more backcross generations and identify favourable recombinants to break up the genome distal to the Booroola locus and so is likely to have originated from the donor Booroola animal. The introgression programme can therefore be designed so that the expected length of the linkage drag segment does not contain any deleterious or undesirable donor alleles. If suitable flanking markers are available, another method (not examined here) to control the risk is the selection of favourable recombinants minimising expected length and variation of the linkage drag (Hospital, 2001). This can only occur if there is a sufficient number of animals to allow for the selection pressure against the weaning weight locus as well as the selection of Booroola carriers.

6.3. Example 2: Genetic lag in dairy cattle introgression schemes Introgression of the polled gene to horned dairy population

6.6.1. Introduction

An inferior line can contain many alleles affecting a trait of interest that are not present in any commercial lines (*e.g.*, QTL affecting litter size in Chinese Meishan pigs, Rothschild *et al.*, 1994). The donor line may be inferior for traits that have been selected upon for many generations in the recipient line. The introgression of

unfavourable donor alleles, on carrier and non-carrier chromosomes, can result in a reduction in performance for production traits in the recipient breed through "donor contamination" (Wakeland *et al.,* 1997). The recipient population undergoing introgression will lag behind the commercial population undergoing selection solely for production traits. Having detected a gene of interest from an inferior line, the aim of introgression is to fix the gene of interest in a commercial population whilst minimising donor contamination.

The overall impact of an introgression programme, including donor contribution and loss of selection opportunity in the recipient breed, can be measured as genetic lag, *i.e.,* the difference between the non-introgressed commercial population and the offspring of the *inter se* cross. Gama *et al.* (1992) and Visscher and Haley (1999) developed equations to describe genetic lag between an introgression population and a commercial population considering only the number of backcross generations. However, they did not consider the impact of linkage drag on genetic lag.

The predictions of genomic contributions developed in Chapter Two can be used to estimate the genetic lag of an introgression programme, including the impact of the linkage drag. The equations derived, describing genomic contributions from ancestral generations, provide a framework to predict the genomic composition of a population after the introgression of a favourable donor allele. These ancestral contributions can be assigned a genetic worth allowing the prediction of genetic lag and provide a validated deterministic prediction for genetic lag.

6.3.2. Materials and methods

The genetic lag of an introgression scheme can be estimated by predicting the genetic worth of the introgression population relative to a commercial population. Predicted genomic contributions (see Chapter Two) are weighted and used as an estimate of genetic lag. Summing the Equations 2.6 and 2.7 gives the entire genomic contributions of ancestral groups post-introgression:

$$\pi_E(X) = \pi_{NC}(X) + \pi_C(X) \qquad (6.1.)$$

where $\pi(X)$ is the contribution of population X to the offspring of the *inter se* cross (IC), E, NC and C denoting the entire genome, non-carrier and carrier chromosomes respectively.

Each $\pi_E(X)$ can be weighted by the genetic worth of individuals at any given point in time, assuming, for example, an infinitesimal genetic model. Assume that G_D gives the difference in background genotype between recipient and donor genome and that the commercial population has a genetic gain of ΔG per generation. Then the genetic lag, ΔI, is given by:

$$\Delta I = (T+1)\Delta G - [\pi_E(D)G_D + \sum_{t=0}^{T+1} \pi_E(R_t)(T)\ \Delta G\ ] \qquad (6.2.)$$

The first term of Equation 6.2 represents genetic gain in the commercial population and the second is the total gain for commercial traits during introgression. This assumes that the scheme is unable to make any selection other than for the donor target allele, which is likely to be the situation for ruminants but less likely for pigs and poultry where a level of concurrent selection may be feasible given sufficient

resources. The substitution effect of the target allele/s is not included in this prediction.

The commercial Holstein Friesian cattle population has been undergoing selection for commercial milk production traits for generations. However, different traits are now becoming important to consumers and producers alike. Disease resistance, for example, is important to the consumer as it means fewer treatments on farm (welfare concerns) and it reduces maintenance costs for the consumer. Welfare issues are important to modern day breeders and producers. One example of a single gene that may minimize a serious welfare concern in cattle is the polled gene (Brenneman *et al.*, 1996). The Holstein Friesian population has a very low frequency of the polled allele and therefore has a high frequency of the horned allele resulting in a predominantly horned population. However, there are cases of polled herds within the British and North American Friesian population (John King, *pers comm.*). These animals may not necessarily be all homozygous for the polled alleles but could act as suitable donors of the polled allele. Producing Holstein Friesian bulls that are homozygous for the polled allele would mean that calves would not have to be chemically dehorned, thereby eliminating the necessity of a very stressful and costly farm procedure that is part of modern day calf rearing. Also, legislation might rule against dehorning of animals in the future, especially in certain production systems such as organic or free-range. Therefore, creating a pool of genetically dehorned animals will help the breeder to meet these future demands.

The polled allele is a dominant gene to horned gene and therefore carriers of the gene will be polled. The donor polled Friesian herds are said to be (PP) or (Pp) for the polled locus and the Holstein Friesian is assumed to be (pp). The polled gene in cattle located on chromosome one (Brenneman *et al.,* 1996). Georges *et al.* (1993a) mapped microsatellite markers have been mapped for this gene. Table 1.1 illustrates a breeding programme whereby a polled gene is introduced from into a horned cattle population.

The aim of this study is quantify genetic lag in a breeding programme attempting to introgress the polled gene from the polled Friesian (PP/Pp) into Holstein Friesian animals heterozygous for the polled allele (Pp) or homozygous horned (pp). It will be assumed that the aim is to create bulls that are homozygous for the polled gene so that all offspring of matings with these bulls are phenotypically polled. This will be achieved by if a period of backcrossing is practised, followed by an *inter se* cross. The production parameters used are in line with current (2001) UK breed averages.

6.3.3. Results

Equation 6.2 can be used to estimate the genetic lag of an introgression population when compared to a commercial population undergoing selection solely for production traits. This introgression scheme aims to introduce the polled gene from the British Friesian population into Holstein Friesian population to create genetically polled bulls with a superior production trait profile than in the subset of the Friesian population.

The polled gene is located at the end of cattle chromosome 1. The length of the chromosome is assumed to be 1.5 Morgan for the purpose of this study. The entire cattle genome is assumed to be 30 Morgan. An initial breed difference, at the time of the F_1 cross, for kg of milk of 1000 kg (6500 VS 5500) is assumed with the British Friesian being inferior to the Holstein Friesian. A Multiple Ovulation Embryo Transfer (MOET) scheme is applied for the duration of the introgression scheme to reduce the female generation interval to 4 years (the application of a maiden heifer MOET scheme will be discussed later). The MOET scheme produces 4 live offspring per mating with a sex ratio of 50%. Therefore there are two female offspring per mating, a half of which will be carriers of the polled allele (*i.e.,* each female replaces itself in the next backcross generation). All heterozygous females are backcrossed to top Holstein Friesian bulls to minimise genetic lag. A genetic gain of 1.5% of the mean per annum in top Holstein Friesian bulls is assumed. The number of backcross generations in the scheme is varied.

Table 6.3 shows the average genomic contributions on carrier and non-carrier chromosomes of the bulls homozygous for the polled allele following the introgression scheme for 3 backcross generations. The assumed genetic worth of the genomic contributions are also shown.

TABLE 6.3. Genomic contributions on carrier (C) and non-carrier (NC) chromosomes expressed as a proportion of their length and total genome length (E). Row totals indicate the genomic contribution across the genome of donor (D) and recipient individuals used in the F1 cross (R_0) and the backcross generations (R_1 to R_3). The genetic worth is expressed in kg of milk.

X	$\pi_C(X)$	$\pi_{NC}(X)$	$\pi_E(X)$	Genetic worth (kg milk)
D	0.016	0.059	0.075	415
R_0	0.004	0.059	0.063	416
R_1	0.006	0.119	0.125	862
R_2	0.009	0.238	0.247	1809
R_3	0.015	0.475	0.500	3886
Total	0.050	0.950		**7385 kg of milk**

Table 6.3 shows that homozygous polled bulls will have a predicted genetic merit of ~ 7385 kg of milk. The introgression scheme would take 16 years to complete (4 years per generation, 3 backcross generation plus *inter se* cross). The predicted genetic merit of commercial Holstein Friesian bulls after 16 years of a MOET breeding scheme for conventional selection for production traits would be ~ 8250 kg (6500 + 1.5% ΔG per annum) resulting in a genetic lag of nearly 12% (for kg of milk) for the introgression homozygous bulls. The genetic lag for an introgression scheme with 2 backcross generations is slightly higher at 13.5% (7770 kg VS 6840 kg).

Assuming a rate of genetic gain/annum of 1.5% in the pure-bred British Friesian bulls, it would take over 20 years of selection on kg of milk to reach 7385 kg (equal to that of homozygous bulls post introgression) with the BF having a genetic worth of 7000 kg of milk after 16 years. Although still lagging behind the commercial population for milk production traits the substitution effect of the polled allele may outweigh this difference in terms of reduced costs and management.

One potential way of reducing the genetic lag and the timescale of the introgression programme is to use the more advanced technology of a juvenile MOET scheme using maiden heifers. An introgression programme with 3 backcross generations would now take 8 years (2 year generation interval) and the genetic lag would be 6% (6917 kg VS 7322 kg).

6.3.4. Discussion

The initial cross is between the Holstein Friesian and the inferior British Friesian, which results in a loss in commercial characteristics in the introgression population. The genetic lag is quantified to be about 11% after the introgression scheme. The study shows that reducing the number of backcross generations in the scheme increases the genetic lag, whereas the reduction of the generation interval (by the introduction of a juvenile MOET scheme) decreases the genetic lag and timescale.

If additional selection criteria are to be applied throughout, it is necessary to have a large enough population of carriers at each backcross generation to place further selection upon whilst maintaining enough animals to breed from at the end of the scheme. The number of individuals that are needed to maintain the simple scheme set out above for the introgression of the polled gene (3 backcross generations, 4 offspring per mating) is quite large. Assuming 50 homozygous bulls are required at the end of introgression, 400 cows need to be maintained throughout the backcross phase. 200 donor cows produce 400 F1 female carrier offspring, which are mated in BC1 to produce 400 female carrier offspring. These 400 carrier cows are mated in

BC2 and then in BC3 to produce 400 female carrier offspring and 400 male carrier offspring. These go forward to the *inter se* cross to produce 50 homozygous (0.25) male (0.5) offspring.

Using sexed semen throughout the scheme (sexing for females in the offspring of the backcross, no sexing in last backcross generation and sexing males for the offspring of the *inter se* cross) would reduce the numbers to be maintained at each backcross generation. Six (6.25) donor animals will produce all female carrier offspring in the F1 (25) which are mated in BC1 to produce 50 female carriers, mated in BC2 to produce 100 female carriers, mated in BC3, with no sexing, 200 male carrier and 200 female carrier offspring. These 400 carriers are intercrossed with sexing to produce all males 0.25 of which will be homozygous. If the number of maintenance animals were increased there would be some scope for selection on other characteristics such as reduced donor genome contribution and/or selection for recipient breed traits (most likely by genotype information as carrier animals could not go through phenotypic testing without increasing the generation interval).

6.4. Example 3: Comparison of methods to predict genetic lag Introgression of the Belgian Blue double-muscling gene into the Holstein Friesian population

6.4.1. Introduction

Previous methods of estimating genetic lag in gene introgression programmes have been developed (*e.g.*, Gama *et al.*, 1992; Visscher and Haley, 1998) but have failed to take account of the linkage drag on the carrier chromosome. The linkage drag will

reduce the recipient contributions to the carrier chromosome compared to a non-carrier chromosome. The theory developed in Chapter Two describes fully the genomic contributions from all ancestral populations in a gene introgression scheme and these are combined to estimate the genetic lag in an introgression population (Equation 6.2). The aim of this study is to apply Equation 6.2 to estimate the genetic lag for a gene introgression programme and compare it to the method used by Gama *et al.* (1992) and Visscher and Haley (1999).

A hypothetical example of a livestock introgression programme would be the introgression of the Belgian Blue double muscling allele into the commercial dairy Holstein-Friesian breed. The double muscling gene (Brennerman *et al*, 1996) is a single autosomal recessive gene causing increased muscles of the back and hindquarters and absence of fat. Heterozygotes may exhibit some increased muscling. Due to the increased meat in the carcass there may be some future role for the double muscling in a dairy breed as if this gene is present in populations male calves could be sold at a higher profit and there would be an increase of output (milk and meat) from a farm. This is only used as a hypothetical example of a single gene that is introgressed from a donor population to and recipient population and could be considered for any allele of interest found in a non-commercial line and the same technique could be considered for genes controlling disease resistance *etc.*. The example of introgressing the double muscling gene describes a scheme where the donor and recipient breeds differ greatly for production traits; therefore there will be an increased genetic lag than in the previous example.

The double muscling gene has been mapped to the end of chromosome 2 (Charlier *et al.*, 1995). An initial breed difference of 5500 litres of milk/lactation (7000 vs. 1500 litres) and a genetic improvement of 1.5% in the mean production per annum is assumed. The cows in the introgression programme are mated to the top available bulls at each backcross generation, with a generation interval of 4 years on the cow side. The cattle genome is approximated to 30 Morgan and chromosome 2 to 1.5 Morgan, with the double muscling gene at 10 cM. The example only examines the genetic lag for the commercial dairy traits.

The Materials and methods for estimating genetic lag are described in the previous example. It is important to point out that the example of the introgression of the double muscling gene is a hypothetical example that may not necessarily be desired practically. However, it does illustrate the effect of the introgression of a gene from a breed of diverse commercial properties.

6.4.2. Results

For an introgression programme incorporating 4 backcross generations (20 years) the genetic lag is ~ 1516 litres (9428 vs. 7912) or just under 12 years of selection. This can be compared to 1397 litres of lag using the prediction of Visscher and Haley (1999). This underestimated the lag by 119 litres, or 8%. The method of Visscher and Haley (1999) assumed that the genomic contributions of the carrier chromosome followed that of the non-carrier chromosomes, reducing by a half with each subsequent backcross generation. As the position of the target locus becomes more

central the linkage drag become larger and the fractional error between the two predictions increases slightly (increasing to 10.5% when s = 75 cM).

The difference between Equation 6.2 and the method of Visscher and Haley for predicting genetic lag becomes larger as the number of backcrosses increases initially but stays asymptotes, at approximately 8.5%, above 6 backcross generations (Figure 6.2). The donor genome makes up a large proportion of the carrier and non-carrier chromosomes in an introgression programme with only a few backcross generations, therefore the effect of the carrier chromosome is not as important. If the number of backcrosses increases the on non-carrier chromosomes decreases rapidly but the linkage drag segment and therefore the donor genome on the carrier chromosome does not decrease as quickly. This means that the linkage drag becomes a more important source donor genome in the prediction of genetic lag.

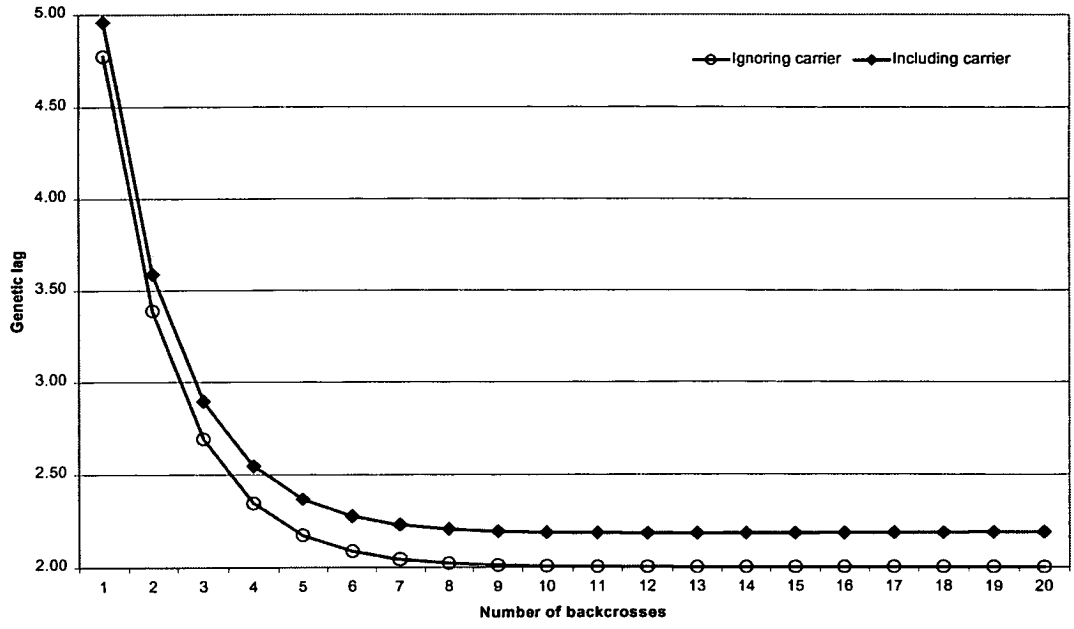

FIGURE 6.2. The effect of the number of backcross generations on methods (Equation 6.2. vs. Visscher and Haley, 1999, or including vs ignoring differences in genomic contributions of carrier and non-carrier chromosomes) of predicting genetic lag (generations of genetic gain in commercial pop^n).

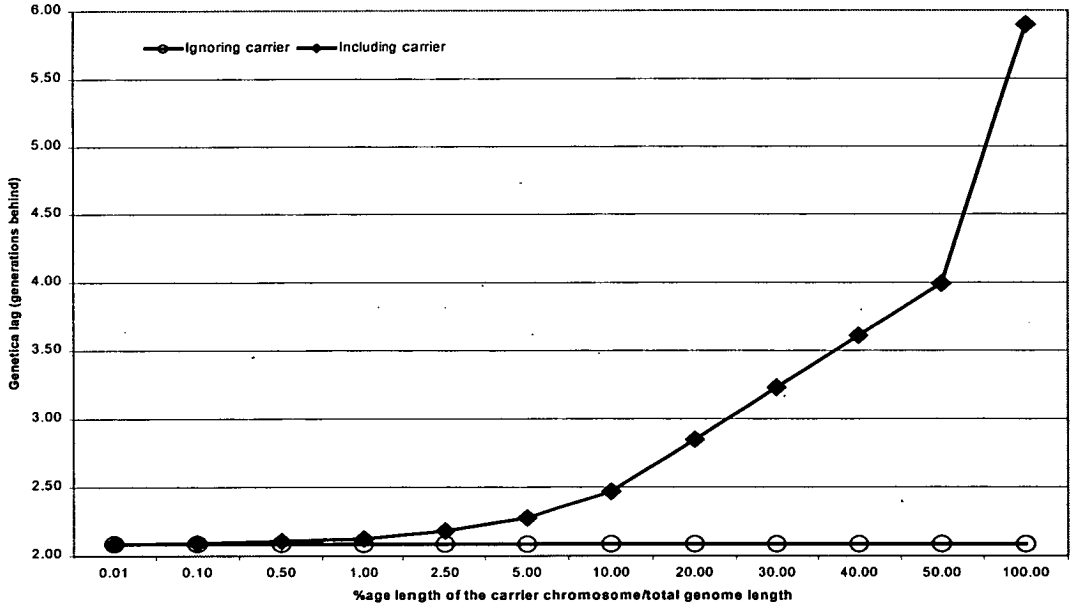

FIGURE 6.3. The effect of the carrier chromosome length as a proportion of genome length (p) on two methods (Equation 6.2. vs. Visscher and Haley, 1999, or including vs ignoring the difference in genomic contributions of carrier and non-carrier chromosomes) of predicting genetic lag.

Figure 6.3 shows that when the carrier chromosome (l) is relatively small in comparison the entire genome length (L) (or p is small) the difference between the predictions in Equation 6.2 and that of Visscher and Haley (1999) of genetic lag is small. As p increases the differences between the two predictions of genetic lag is dramatically increased (*e.g.*, when p = 20% the percentage difference between the two methods is almost 27%).

Using the method of Visscher and Haley (1999) can lead to a highly inaccurate prediction of potential lag. For example, the chicken genome is made up of 6 pairs of macrochromosomes and 30 pairs of microchromosomes (Smith *et al.*, 2000). Chromosome 1 is 3.8 Morgan and approximately 17% of the genome. Underestimating the genetic lag may effect a breeder's decision on the type of programme to use (*e.g.*, the minimum number of backcross generations needed to achieve a certain acceptable genetic lag given the proportional length of the carrier chromosome) and may make the difference between success and failure of the programme and commercial viability.

6.4.3. Discussion

The prediction of genetic lag does not include selection in the introgression programme for commercial traits or the economic improvement due to the extra value earned from new donor trait included. However, whilst incomplete, Equation 6.2. serves as a basis for estimating the monetary cost of an introgression programme. Modifications would need to include the premium attached to a new

commercial product as well as the cost of the programme in a cost-benefit analysis in different populations to pinpoint the best and most cost effective type of introgression programme for different populations.

Using background selection can increase the recovery of the recipient genome either by minimising the linkage drag using flanking markers or selecting for recipient alleles on non-carrier chromosomes (Frisch and Melchinger, 2001; Hospital and Charcosset, 1997; Visscher *et al.*, 1996). Using the equation of genetic lag and the equations for carrier chromosome contributions in Chapter Two could help to optimise the type of background selection carried out during introgression.

6.5. Example 4: Genetic lag and identity-by-descent in a Meishan X Large White pig backcross population.

6.5.1. Introduction

During introgression many alleles linked to the target donor allele are incorporated into the recipient line due to linkage drag. Alleles at loci in the region of the target locus may trace back to a common ancestor and become identical by descent (IBD) during backcrossing. This leads to a loss of diversity around the target locus. Chapter Three showed that alleles are lost during backcrossing. The loss in the allelic variants during backcrossing will increase IBD, especially directly around the target loci. This loss of variation will increase the inbreeding coefficient of the population. This increase in inbreeding would not be an issue in plant introgression schemes as many of these schemes are based on using inbred recipient lines. However, an inbred

population (homozygous for the target allele post introgression) would constitute a high risk strategy in livestock introgression programmes.

This example aims to investigate the effect of population size and number of backcross generations on genetic lag, linkage drag, allelic diversity and homozygosity due to alleles being IBD on the carrier chromosome after introgression. The programme introgresses an allele reducing back fat found in the Chinese Meishan breed into a commercial Large White population.

6.5.2. Materials and methods

The expectation of the allelic diversity at loci on the carrier chromosome is described in detail in Chapter Three. Equation 3.8 will be used to predict the number of alleles at a locus on the carrier chromosome in a backcross breeding programme.

Homozygosity due to alleles being identical by descent (IBD) was described using simulation results in Chapter Five. These results will be used to estimate the homozygosity around a target locus after an introgression programme.

The initial cross assumed N donor Meishan boars were mated to N Large White sows with an average of eight piglets per litter throughout backcrossing. At each backcross generation N carriers of a QTL allele on pig chromosome 7 at 58 cM (additive effect of decreasing fat depth at shoulder, FS, by 2.7 mm, Walling *et al*, 1998) were selected. Meishan and Large White differed by approximately 280 g (679 vs. 395) for average daily gain (ADG) (Haley *et al*, 1992) with a rate of genetic progress in

commercial lines of 15 g in ADG per annum (A.D. Hall *pers comm*). Chromosome 7 is 150 cM and the size of the pig genome is 25M.

Linkage drag was predicted using Equation 2.1. Genetic lag in ADG was predicted using Equation 6.2. Allelic diversity on the carrier chromosome at the end of backcrossing was predicted using Equation 3.8. The simulation studies described in Chapter Five were used to examine homozygosity due to alleles being IBD around the target locus. The number of parental pairs (N=10-100) and the number of backcross generations in the introgression scheme (T=1-10) were varied in the simulation studies and run for 500 replicates.

6.5.3. Results

Genetic lag decreased as the number of backcross generations rose (Figure 6.4). Changes in population size had no significant effect on genetic lag (results not shown). This is in agreement with previous results. Chapter Two showed how population structure had little effect on the genomic contributions. Genetic lag was predicted using a weighting of the genomic contributions (Equation 6.2), therefore changes in population structure would have little effect on this parameter.

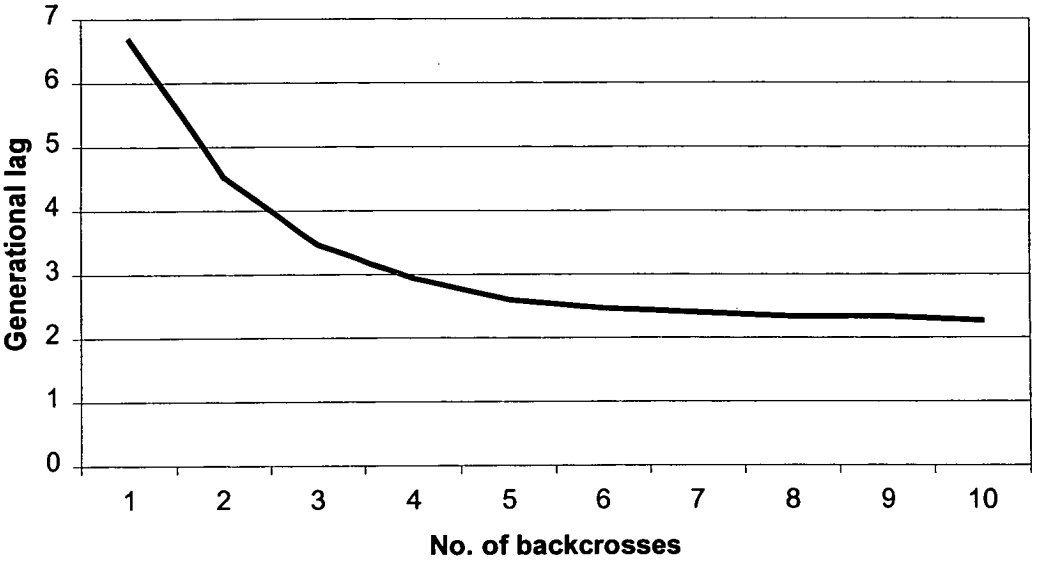

FIGURE 6.4. Effect of the number of backcross generations (T x-axis) on genetic lag, expressed in the number of generations of selection that the introgression population is behind the commercial population.

Linkage drag decreased and proportion of loci homozygous due to donor alleles being IBD increased as backcross generations increased (Figure 6.5). Allelic diversity on the carrier chromosome also decreased in a similar fashion to linkage drag. The proportion of loci homozygous due to IBD on the carrier chromosome was relatively constant over latter backcross generations (~0.06 when N=20 and $T > 3$).

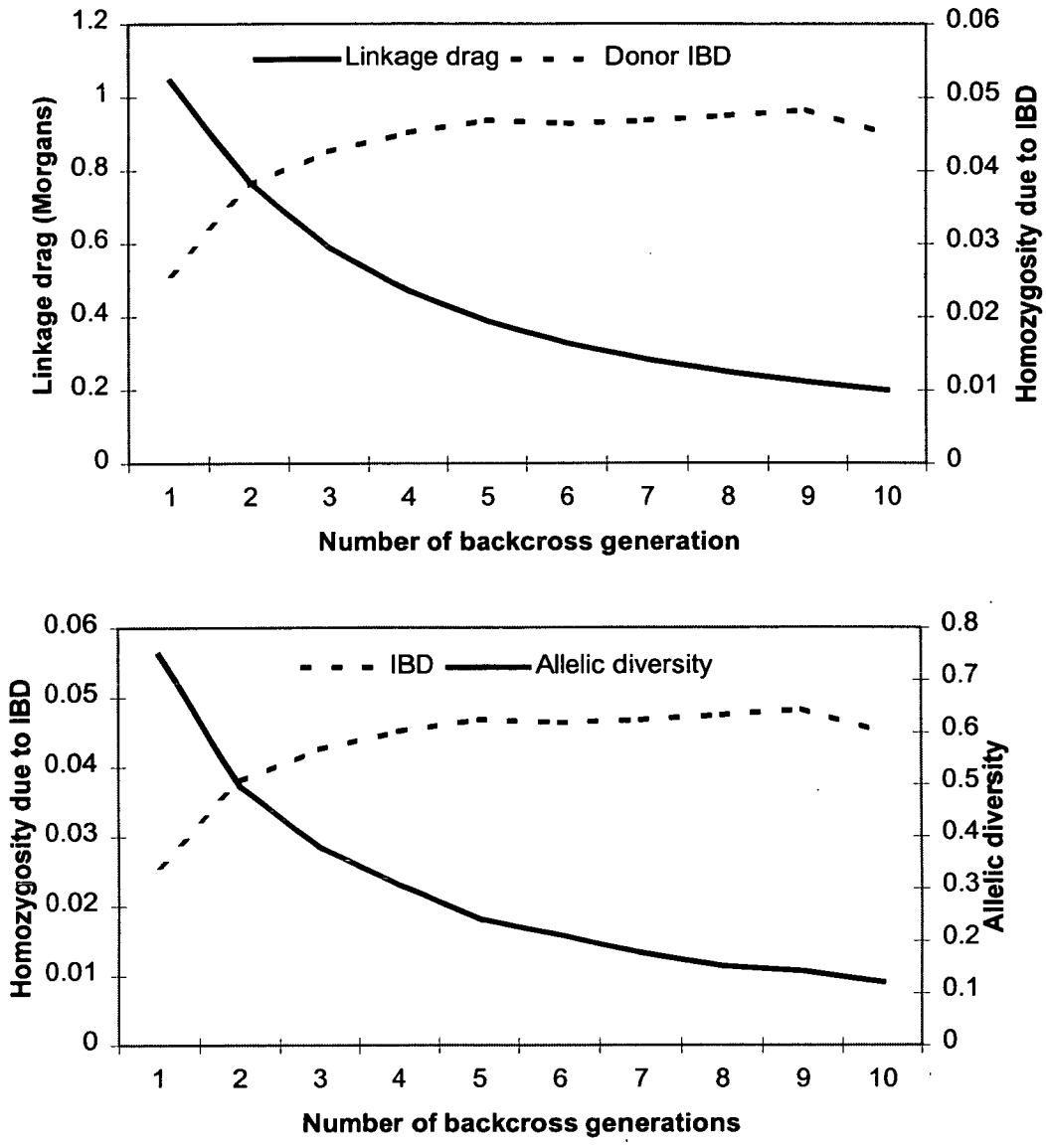

FIGURE 6.5. Effect of T on linkage drag length in Morgans (solid line, upper graph) and homozygosity due to donor alleles being IBD on the carrier chromosome (dashed line) and allelic diversity (solid line, lower graph).

Proportion of loci IBD decreased as population size increased (Figure 6.6). Figure 6.6 shows how the number of donor alleles on the carrier chromosome increased as population size increased. However the allelic diversity did not change as dramatically as the number of alleles, although it did increase slightly (from (0.35 to 0.41, full results not shown). IBD on the carrier chromosome was mainly due to the

high levels around the target locus and this occurred due to the limited number of contributions from donor ancestors.

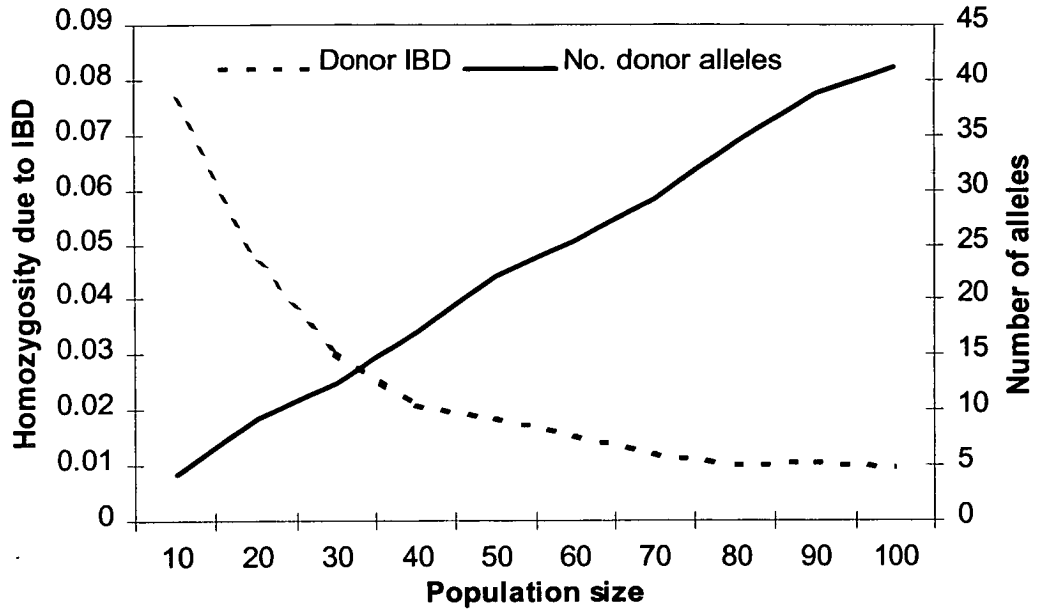

FIGURE 6.6. Effect of *N* on homozygosity due to donor alleles being IBD on the carrier chromosome (dashed line) and allelic diversity (solid line, lower graph).

The results show genetic lag reduces in early backcross generations but after 5 generations the lag approached an asymptote. Although there is genetic lag for ADG, the beneficial effect of the back fat allele may compensate for this lag. For this scheme it appeared that it was necessary to have no more than 5 backcross generations and 80 mating pairs at each generation to maintain an acceptable level of genetic lag and IBD around the target locus but this has to be balanced by an acceptable level of genetic lag.

6.5.4. Discussion

Genetic lag decreased whilst the homozygosity increased as the number of backcross generations increased. This is an unfavourable relationship as, ideally, genetic lag

and homozygosity should be kept to a minimum. The number of backcross generations seemed to have the largest effect genetic lag, allelic diversity and homozygosity. A livestock introgression scheme will aim to reduce the genetic lag by increasing the number of backcross generations but it may do so at the cost of an increased inbreeding coefficient at the end of the introgression programme. In this case the programme should either reduce the number of backcross generations to maintain a reasonable level of inbreeding or increasing the population size to decrease the risk of the homozygosity post introgression.

6.6. Discussion

This Chapter has examined and validated some practical applications of the predictions and estimates of various indicators of genomic status post introgression given earlier in the thesis. The aim of this was to illustrate the utilisation of methods given in Chapters Two and Three and applications of the estimates of homozygosity in Chapter Four and Five.

1. The introgression of the Booroola allele into a commercial meat sheep population was shown to have potentially dragged an unfavourable weaning weight QTL into the commercial population resulting in reduced weights. This method of predicting genetic lag could be used to infer the origins of segments of the genome in introgression programmes.
2. Genetic lag was shown to be an important consideration in livestock introgression schemes (introgression of the polled allele and the double muscling allele into a commercial Holstein Friesian dairy population). The prediction of

genetic lag, using prediction of genomic contributions on the carrier chromosome, was shown to be more accurate than previous methods. This study has shown that genetic lag in livestock introgression programmes originates from the linkage drag as well as the reduction of the genetic merit due to donor contamination throughout the genome.

3. The loss of alleles and subsequently identity-by-descent was shown to be high in the example of the Meishan X Large White pig backcross population. This loss of variation must also be considered in the design of a livestock introgression scheme. If the variation becomes too low the final product may be unviable.

These practical examples show that a livestock introgression has many parameters to consider in the design stage. It is necessary to decide the main aims of the scheme in terms of acceptable levels of genetic lag, genetic variation and the inbreeding and time scale and cost. It is necessary to plan the scheme to achieve these guidelines. If the introgression scheme is limited by the population structure or the length of the generation interval it may be necessary to balance genetic variation against increased costs, time scale or accepting increased genetic lag.

CHAPTER SEVEN

GENERAL DISCUSSION

A number of studies have examined methods of speeding up gene introgression programmes using marker-assisted selection protocols (*e.g.,* Hillel *et al*, 1990; Hospital *et al.*, 1992). Many of these studies do not relate the results to practical livestock introgression programmes. Many factors have to be considered before embarking on a livestock introgression programme, such as, reproductive capacity, number of individuals homozygous for the target allele required at the end of backcrossing, genetic lag, time-scale and the increased risk of inbreeding, especially around the target allele. This thesis has attempted to address these issues drawing general conclusions and specific conclusions that apply to livestock introgression programmes.

Prediction equations were developed, and validated by simulation results, to describe the genomic contributions of the donor and recipient populations to the genome of the introgression population (Chapter Two). The genomic contributions from donor and recipient populations were affected by the number of backcross generations, the length of the carrier chromosome, and the position of the target locus. The predictions of genomic contribution can be used to infer origins of segments of the genome after an introgression or backcross scheme (*e.g.*, the dragging of an unfavourable weaning weight QTL with the Booroola allele into a commercial sheep population). Understanding the genomic status of the genome of such breeding scheme designs can be utilised in a number a ways, including QTL mapping.

The prediction equations of genomic contributions were further developed to estimate the genetic lag of practical livestock introgression schemes. The method of

predicting genetic lag was shown to be more accurate the previous methods used by other authors (*e.g.*, Visscher and Haley, 1999). The genetic lag decreased as the number of backcross generations increased. However, the gain in genetic lag slowed down after approximately 4 or 5 backcross generations. This suggests that the increasing the number of backcross generations has little additional benefit after a certain time frame.

The prediction of genetic lag ignored the impact of additional selection for commercial traits during the introgression procedure. Selection based on phenotypic or genotypic information on commercial traits during the backcross period to reduce the genetic lag. The scope for this in practical livestock introgression may be limited and therefore may only add little benefit in terms of the reduction of genetic lag. The selection pressure for commercial traits after the selecting individuals carrying the target allele will be low in many livestock population. This will be due to the limited reproductive capacity and the cost of maintenance of a large population, both of which will limit the scope and feasibility of livestock introgression programmes.

The reduction in selection pressure for commercial traits is inevitable when practicing introgression. The gain by including the donor gene in the commercial population often is hard to quantify and should be expected to outweigh the potential loss. An introgression scheme should be planned carefully from the outset, deciding on technological and molecular methods to use, the number and which type of animals needed and whether or not the target donor gene will bring a benefit that will outweigh the loss. With careful planning at the start, and at each backcross

generation, and by being dynamic in the decision making process (breeding from the best animals available at each stage and not too rigid in the selection criteria) the introgression programme would achieve its end objective.

Koudandé (2000) examined the use of different reproductive techniques (artificial insemination, AI; multiple ovulation and embryo transfer, MOET; *in vitro* maturation, IVM; *in vitro* fertilisation, IVF and semen/embryo sexing) to accelerate cattle introgression. These results showed that targeting multiple QTL in cattle introgression using natural service is not feasible due to the low reproductive capacity on the female side. MOET could be used in cooperation with AI to combat the low reproductive capacity on the female side and juvenile MOET can be used to reduce the generation interval. However, increasing the number of offspring per female (up to 25 per cow with MOET but may take a very long time, Cunningham, 1999) will increase the number of related individuals and therefore increase the inbreeding. Embryo/semen sexing, IVM and IVF may also provide some additional benefit to an introgression scheme in terms of reducing costs (reducing animal maintenance costs) and decrease the time scale (decreasing the generation interval). However, the use of such technologies must be cost effective and the costs considered at the design stage of the introgression scheme. These technologies cannot be used without some management of the risk of inbreeding in the final population as has already been examined with straight breeding populations.

Current and future developments in reproductive and molecular techniques surpass the technologies described here. There is a potential for an introgression scheme to

be carried out completely in the lab. Animals can be selected at the embryonic stage for a number of traits (including the target allele) across the genome using high density mapping (*e.g.*, DNA chips). Animals could also be selected for the genomic diversity at this stage. These embryos can then be implanted and therefore reduce the costs of maintenance of animals and the generation interval. Development of reproductive techniques, such as harvesting oocytes from young animals, will also help reduce generation interval and costs. Developments in technology associated with cloning could completely negate the need of raising animals at all during the introgression period. Nuclear transfer may develop to the stage were cell differentiation may be controlled *in vitro* and therefore meiosis, followed by fertilisation, can occur in a "test-tube" and allow for very quick generational turn around (Visscher and Haley, 1998). However, the genetic variation will be dramatically reduced if care is not taken with such a procedure. Finally, the need for introgression programmes may become of thing of the past if gene transfer is developed sufficiently.

The monetary costs of a gene introgression programme have been mentioned briefly in this study. Introgression requires additional resources above that of breeding schemes in terms of maintaining a larger animal population, genotyping, reproductive technology and additional time. The cost benefit of the introduction of the favourable gene into a new population is hard to quantify. This benefit is priceless when dealing with the introgression of a gene due to legislative restrictions (*e.g.*, disease resistance genes if the use of medications is limited/prohibited, polled gene in cattle if dehorning is banned, behavioural QTL such as cannibalism, although

such QTL are yet to be found). A cost-benefit analysis should be carried before embarking on an introgression programme. The cost of introducing some types of genes (such as those described previously) into a population may appear to be unsustainable after a cost-benefit analysis, however the benefits of consumer acceptance and welfare standards are difficult to quantify. The use of new technologies will increase the operating costs of the introgression programme. If considered a viable expense, the increase in operating costs will speed up the introgression programme and increase the opportunity for selection on additional commercial traits, therefore reducing the waiting period for return of investment. This cost benefit analysis must also consider the genetic aspects, such as genetic lag, genetic variation and inbreeding.

Prediction equations were developed and validated to describe the loss of alleles at a neutral locus and around a target locus during backcrossing. These results showed that reduced population size, number of backcross generations and initial alleles frequency had a dramatic effect on the allelic diversity and the end of introgression and therefore the genomic variation. This loss of variation, particularly around the target locus will increase the probability of the inbreeding (due to homozygosity) in the final population homozygous for the target allele. This is of particular importance if an unfavourable allele associated with the donor breed (or even a specific donor individual) lies close to the target locus. The reduction of allelic variation will increase the probability that this allele is dragged into the recipient line. The prediction of the loss of alleles can be used to extrapolate the loss of genomic variation in introgression or similar breeding schemes. Understanding this potential

loss of variation before commencement of such a scheme will help develop an ideal strategy to maximise the variation both during and after the scheme is completed.

The rate of loss at a neutral locus in an introgression scheme can be applied to many situations where the loss of genetic variation in closed breeding schemes may be of interest. The most obvious association is in breed conservation programmes. In this situation, donor populations of limited size may need to go through a backcrossing scheme to a similar breed to try to increase effective population size. In this case, it would be necessary to maximise the contribution of different donor alleles in the genome and maintain a reasonable level of allelic diversity. Future studies may wish to adapt the prediction for the loss of alleles to use for conservation genetics (*e.g.*, reconstitution of a population from cryoconserved sperm/embryos by grading-up to a recipient population). Populations with a larger effective population size can be used to reduce the inbreeding of a smaller conservation scheme. However, these equations would have to be developed to account for different mating structures (*i.e.*, other then backcrossing). Therefore, different schemes can be compared to choose the mating scheme that reduces inbreeding or maximises variation.

The utilisation of marker information to maximise variation could help in the design of conservation schemes or introgression scheme with a low population size. The predictions in Chapter Three do not account for this. However, markers would need to be placed throughout the genome to maximise variation in conservation programmes. The feasibility this type of scheme is limited because of this, as it would require a large number of genotypings to be successful.

It is natural to move from allele loss, indicative of inbreeding, to estimation the inbreeding of livestock introgression schemes. Chapter Five predicted the homozygosity at neutral loci due each of the recipient populations and showed that IBD at neutral loci was relatively low and mainly due to the recipient population used in the last backcross generation.

Inbreeding at the target locus was shown to be quite high. However, this study assumed that there were $2N$ alleles were at the target locus in the donor population. Realistically donor populations will not carry such large number of alleles for traits that are fixed in the population. For example, it would not be reasonable to assume that there are $2N$ different ancestral mutations of the polled allele in polled cattle populations. It is reasonable to assume that the area around this locus will have also become fixed for linked donor alleles. Therefore, the allelic variation will not be as high in the region of the target locus as described in Chapters Three, Four and Five. This would result in a much higher probability of homozygosity around the target locus after gene introgression then described by this work. The predictions of allelic diversity can be used for any number of alleles at a locus in any backcross breeding. This can then be related to the risk of homozygosity after the *inter se* cross.

The prediction of homozygosity around the target locus was more complex due to the complication of the scheme design coupled with the additional effect of distance on selection pressure of the target locus. Future studies wishing to study homozygosity

out with the target locus would need to incorporate these additional complexities into the mathematical model.

This work could be developed further to help in the design of optimum marker-assisted introgression schemes. The first step in the advance of this work would be to include the effect of selection outside that of selection on the target allele. This selection can take many forms (as discussed in Chapter One). Selection can be for commercial traits (phenotypic or genotypic), the reduction of donor contamination or reduction of the linkage drag. However, this type of selection should also be conducted whilst maintaining the genetic variation to reduce the risk of inbreeding in the population. Marker information may also be used to maximise the genetic variation (*e.g.* Wang, 2001, Wang & Hill, 2000). Accounting for these multiple forms of selection using mathematical models may be complicated and these types of schemes may have to be simulated. The predictions presented in this thesis will provide information on the general design of an introgression scheme that reduces the genetic lag whilst maximising the genetic variation. These predictions can be used to set broader guidelines and simulation studies used to fine-tune the guidelines accounting for selection for commercial traits and/or maximisation of variation.

This study has examined aspects of gene introgression that is of particular importance to livestock introgression. Factors such as inbreeding and reduced variation may not be of such importance in plant introgression or the creation of congenics strains. The creation of inbred lines is acceptable in plant introgression schemes. The population structure in many plant introgression schemes allows scope

for additional selection and more backcross generations (due to the short generation interval). These predictions can be used to predict the genetic lag in the design of plant introgression schemes. The information on allele diversity is could be used to monitor the ideal diversity of seeds entering seed banks for future utilisation.

The specific influence of the thesis was livestock introgression programmes. A number of important results have emerged that produce useful guidelines for such schemes. The following lists some of the important results of these studies that relate to the design of gene introgression programmes in livestock species.

- Linkage drag will tend to be large in livestock introgression programmes. The scope to reduce linkage drag by increasing backcrosses may be limited due to the generation interval of many livestock species. However, these studies show that at least 4 backcross generations would be necessary in many livestock introgression schemes. There may also be additional scope for reducing linkage drag segments by selecting on flanking markers, however this must be done whilst maintaining the allelic diversity.
- Genetic variation also needs to be maintained during backcrossing both around the target locus and elsewhere in the genome. A simple way to do this would be to select individuals on the target locus genotype and then try to select across families amongst these individuals. Increasing the number of donors above 50 had little influence on the final allelic diversity at and around the target locus
- The risk of inbreeding (due to homozygosity-by-descent) in gene introgression programmes can be reduced by keeping the number of backcross generations low

(four) and increasing the population size (at least 50 mating pairs at each generation).

- Livestock introgression programmes (especially in sheep and cows) can be improved with the use of reproductive techniques to reduce generation intervals, increase population size, increase opportunities for additional selection and maintain genetic variation.
- Livestock introgression programmes tend to always lag behind commercial populations. However, the genetic lag is reduced by increasing the number of backcross generations. In the practical livestock introgression schemes the difference in genetic lag between commercial and introgression populations remained relatively constant after five or six backcross generations.

Gene introgression provides an opportunity to utilise favourable traits from non-commercial breeds (disease resistance, fertility, welfare traits). If properly designed and managed the programme can successfully introduce this allele into a commercial breed. The commercial population will be able to accrue the benefit of the target trait whilst maintaining a reasonable degree of performance in the commercial traits. However, there will be a genetic loss for commercial traits and the scheme will take time and careful management of the target allele, commercial traits and genetic variation. If well managed gene introgression provides a natural breeding method to utilise genes from all populations to allow breeders to adapt to the ever changing and increasing diversity of world farming systems.

Allendorf, F.W., 1986. Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology*, **5:** 181-190.

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

Archibald, A.L., Haley, C.S., Brown, J.F., Couperwhite, S., McQueen, H.A., Nicholson, D., Coppieters, W., Van de Weghe, A., Stratil, A., Winterø, A.K., Fredholm, M., Larsen, N.J., Nielsen, V.H., Milan, D., Woloszyn, N., Robic, A., Dalens, M., Riquet, J., Gellin, J., Caritez, J-C., Burgaud, G., Ollivier, L., Bidanel, J.-P., Vaiman, M., Renard, C., Geldermann, H., Davoli, R., Ruyter, D., Verstege, E.J.M., Groenen, M.A.M., Davies, W., Høyheim, B., Keiserud, D., Andersson, L., Ellegren, H., Johansson, M., Marklund, L., Miller, J.R., Anderson Dear, D.V., Signer, E., Jeffreys, A.J., Moran, C., Letissier, P., Mulando, Rothschild, M.F., Tuggle, C.K., Vaske, D., Helm, J, Liu, H.-C., Rahman, A., Yu, T.-P., Larson, R.G. & Schmitz, C.B., 1995. The PiGMaP consortium linkage map of the pig (*sus scrofa*). *Mammalian Genome,* **6:** 157-175.

ArkDB, 2000. URL: http://thearkDB.org [24th of October, 2001].

Barendse, W., Armitage, S.M., Kossarek, L.M., Shalom, A., Kirkpatrick, B.W., Yan, A.M., Clayton, D., Li, L., Neibergs, H.L., Zhang, N., Grosse, W.M., Weiss, J., Creighton, P., Mccarthy, F., Ron, M., Teale, A.J., Fries, R., Mcgraw, R.A., Moore, S.S., Georges, M., Soller, M., Womack, J.E., Hetzel, D.J.S., 1994. A genetic-linkage map of the bovine genome. *Nature Genetics*, **6**: 227-235.

Barendse, W., Vaiman, D., Kemp, S.J., Sugimoto, Y., Armitage, S.M., Williams, J.L., Sun, H.S., Eggen, A., Agaba, M., Aleyasin, S.A., Band, M., Bishop, M.D., Buitkamp, J., Byrne, K., Collins, F., Cooper, L., Coppettiers, W., Denys, B.,

Drinkwater, R.D., Easterday, K., Elduque, C., Ennis, S., Erhardt, G., Ferretti, L., Flavin, N., Gao, Q., Georges, M., Gurung, R., Harlizius, B., Hawkins, G., Hetzel, J., Hirano, T., Hulme, D., Jorgensen, C., Kessler, M., Kirkpatrick, B.W., Konfortov, B., Kostia, S., Kuhn, C., Lenstra, J.A., Leveziel, H., Lewin, H.A., Leyhe, B., Lil, L., Burriel, I.M., McGraw, R.A., Miller, J.R., Moody, D.E., Moore, S.S., Nakane, S., Nijman, I.J., Olsaker, I., Pomp, D., Rando, A., Ron, M., Shalom, A., Teale, A.J., Thieven, U., Urquhart, B.G.D., Vage, D.I., VandeWeghe, A., Varvio, S., Velmala, R., Vikki, J., Weikard, R., Woodside, C., Womack, J.E., Zanotti, M., Zaragoza, P., 1997. A medium-density genetic linkage map of the bovine genome. *Mammalian Genome*, **8**: 21-28

Behringer, R., 1998. Supersonic congenics? *Nature Genetics,* **18:** 108.

Bishop, S.C. and Woolliams, J.A., 1991. Utilization of the sex-determining region Y gene in beef cattle breeding schemes. *Animal Production,* **53:** 157-164.

Block, D.S Curran; D.J., Eisenberg, R.S., Heller, M., 1998. Patenting Genomic Technologies. *Science*, **282:** 1419.

Brenneman, R.A., Davis, S.K., Sanders, J.O., Burns, B.M., Wheeler, T.C., Turner, J.W., Taylor, J.F., 1996. The polled locus maps to BTA1 in a *Bos indicus* x *Bos taurus* cross. *Journal of Heredity,* **87**:156-61.

Brinkman, M.A. and Frey, K.J., 1977. Yield component analysis of oat isolines that produce different grain yields. *Crop Science* **17:** 165-168.

Brown, A.H.D., Lawrence, G.J., Jenkin, M., Douglass, J., Gregory, E., 1989. Linkage drag in backcross breeding in barley. *Journal of Heredity* **80:** 234-239.

Charlier, C., Coppieters, W., Farnir, F., Grobet, L., Leroy, P.L., Michaux, C., Mni, M., Schwers, A., Vanmanshoven, P., Hanset, R., Georges, M., 1995. The mh gene

causing double-muscling in cattle maps to bovine chromosome-2. *Mammalian Genome* **6:** 788-792.

Chetelat, R.T., Deverna, J.W., Bennett, A.B., 1995. Introgression into tomato (*lycopersicon-esculentum*) of the l-chmielewskii sucrose accumulator gene (sucr) controlling fruit sugar composition. *Theoretical and Applied Genetics,* **91:** 327-333.

Cockett, N.E., Jackson, S.P., Shay, T.L., Nielsen, D., Moore, S.S., Steele, M.R., Barendse, W., Green, R.D., Georges, M., 1994. Chromosomal localization of the Callipyge gene in sheep (*Ovis aries*) using bovine DNA markers. *Proceedings of the National Academy of Science U. S. A.* **91:** 3019-3023.

Cunningham, E.P., 1999. The applications of biotechnologies to enhance animal production in different farming systems. *Livestock Production Science,* **58:** 1-24.

Cushwa, W.T., Dodds, K.G., Crawford, A.M., Medrano, J.F., 1996. Identification and genetic mapping of random amplified polymorphic DNA (RAPD) markers to the sheep genome. *Mammalian Genome,* **7**: 580-585.

Davis, G.H., Montgomery, G.W., Allison, J.A., Kelly, R.W., Bray, A.R., 1982. Segregation of a major gene influencing fecundity in progeny of Booroola sheep. *New Zealand Journal of Agricultural Research* **25:** 525 – 529.

Dekkers, J.C.M. & van Arendonk, J.A.M., 1998. Optimizing selection for quantitative traits with information on an identified locus in an outbred population. *Genetical Research,* **71:** 257-275.

Dodds, K.G. 1999. Haley-Knott regression mapping of quantitative trait loci in extended pedigrees. *Proceedings of the Association for the Advancement of Animal Breeding and Genetics* **13:** 313-316.

Ducrocq, V. and Sölkner, J., 1998. Implementation of a routine breeding value evaluation for longevity of dairy cows using survival analysis techniques. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production,* **23:** 447-448.

deGortari, M.J., Freking, B.A., Cuthbertson, R.P., Kappes, S.M., Keele, J.W., Stone, R.T., Leymaster, K.A., Dodds, K.G., Crawford, A.M., Beattie, C.W., 1998. A second-generation linkage map of the sheep genome. *Mammalian Genome*, **9**: 204-209

FAO, 1998. Secondary Guidelines for Development of National Farm Animal Genetic Resources Management Plans - Management of small population at risk.

Falconer, D.S. and Mackay, T.F.C., 1996. In *Introduction to Quantitative Genetics, Fourth Edition.* Longman Group, Ltd, London/Edinburgh.

Frey, M.J., 1998. National plan for genepool enrichment of US crops. *National Plant Breeding Study-III.* Special Report, 191, Iowa State University, Ames.

Franklin, I. R., 1999. Improving the efficiency of backcrossing programs using DNA markers. *Proceedings of the Association of the Advancement of Animal Breeding and Genetics* **13:** 357-360.

Frisch, M. and Melchinger, A.E., 2001. The length of intact donor chromosome segment around a target gene in marker-assisted backcrossing. *Genetics* **157:** 1343-1356.

Gama, L.T., Smith, C., Gibson, J.P., 1992. Transgene effects, introgression strategies and testing schemes in pigs. *Animal Production,* **54:** 427-440.

Georges, M., Drinkwater, R., King, T., Mishra, A., Moore, S.S., Nielsen, D., Sargeant, L.S., Sorensen, A., Steele, M.R., Zhao, X., *et al.*, 1993a. Microsatellite

mapping of a gene affecting horn development in *Bos taurus*. *Nature Genetics*, **4:** 206-210.

Georges, M., Dietz, A.B., Mishra, A., Nielsen, D., Sargeant, L.S., Sorensen, A., Steele, M.R, Zhao, X., Leipold, H., Womack, J.E., Lathrop, M., 1993b. Microsatellite mapping of the gene causing weaver disease in cattle will allow the study of an associated quantitative trait locus. *Proceedings of the National Academy of Sciences of the United States of America,* **90:** 1058-1062.

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

Gibson, J.P., 1994. Short-term gain at the expense of long-term response with selection of identified loci. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production,* **21:** 201-204.

Goddard, M.E., 1992. Optimal effective population size for the global population of black and white cattle. *Journal of Dairy Science,* **75:** 2902-2911.

Green, P., Falls, K., Crooks, S., 1990. Cri-map Version 2.4. Washington University School of Medicine, St. Louis.

Groen, A.F., and Timmermans, M.M.J., 1992. The use of genetic markers to increase the efficiency of introgression – a simulation study. *Genetics,* **8:** 523-527.

Groen, A.F., and Smith, C., 1995. A stochastic simulation study of the efficiency of marker-assisted introgression in livestock. *Journal of Animal Breeding and Genetics,* **112:** 161-170.

Haldane, J.B.S., 1919. The combination of linkage values and the calculation of distances between the loci of linked factors. *Journal of Genetics,* **8:** 299-309.

Haley, C.S. and Knott, S.A., 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity,* **69:** 315-324.

Haley, C.S., d'Agaro, E., Ellis, M., 1992. Genetic components of growth and ultrasonic fat depth traits in Meishan and Large White pigs and their reciprocal crosses. *Animal Production*, **54:** 105-115.

Haley, C.S., Knott, S.A., Elsen, J.M., 1994. Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics,* **136:** 1195-1207.

Haley, C.S., Williams, J., Woolliams, J, Visscher, P.M., 1996. Gene mapping and its place in livestock improvement. *British Cattle Breeders Club, 51st Annual Conference,* pp, 5-15.

Haley, C.S. and Visscher, P.M., 1998. Strategies to utilize marker-quantitative trait loci associations. *Journal of Dairy Science,* **81:** 85-97.

Hammond, K., 1998. Animal genetic resources for the twenty-first century. *Acta Agriculturae Scandinavica Section A: Animal Sciences,* **28:** 11-18.

Hanset, R., Dasnois, C., Scalais, S., Michaux, C., Grobet, L., 1995. Introgression into the Pietrain genome of the normal allele at the locus for halothane sensitivity. *Genetics Selection Evolution*, **27**: 77-88.

Hanson, W.D., 1959. Early generation analysis of lengths of heterozygous chromosome segments around a locus held heterozygous with backcrossing or self. *Genetics,* **44:** 833-837.

Heller, M. A. and Eisenberg, R.S., 1998. Can Patents Deter Innovation? The Anticommons in Biomedical Research. *Science,* **280**: 698 – 701.

Henderson, C.R., 1988. Theoretical basis and computational methods for a number of different animal models. *Journal of Dairy Science,* **71,** Suppl. 2, pp: 1.

Hill, W.G., 1993. Variation in genetic composition in backcrossing programs. *Journal of Heredity,* **84:** 212-213.

Hill, W.G., 1998. Selection with recurrent backcrossing to develop congenic lines for quantitative trait loci analysis. *Genetics,* **148:** 1341-1352.

Hillel, J., Schaap, T., Haberfield, A., Jeffreys, A.J., Plotzky, Y., Cahaner, A., Lavi, U., 1990. DNA fingerprints applied to gene introgression in breeding programs. *Genetics,* **124:** 783-789.

Hillel, J., Verrinder Gibbons, M., Etches, R.J., Shaver, D.McQ., 1993. Strategies for the rapid introgression of a specific gene into a commercial poultry flock from a single carrier. *Poultry Science,* **72:** 1197-1211.

Hospital, F., Chevalet, C., Mulsant, P., 1992. Using markers in gene introgression breeding programs. *Genetics*, **132**:1199-1210.

Hospital, F. and Charcosset, A., 1997. Marker-assisted introgression of quantitative trait loci. *Genetics*, **147**:1469-1485.

Hospital, F., 2001. Size of donor chromosome segments around introgression loci and reduction of linkage drag in marker-assisted backcross programs. *Genetics*, **158:** 1363-1379.

Hu, J., Mungall, C., Law, A., Papworth, R., Nelson, J.P., Brown, A., Simpson, I., Leckie, S., Burt, D.W., Hillyard, A.L., Archibald, A.L., 2001. The ARKdb: genome databases for farmed and other animals. *Nucleic Acids Research*, **29**: 106-10.

Jairath, L., Dekkers, J.M.C., Schaeffer, L.R., Lui, Z., Burside, E.B., Kolstad, B., 1998. Genetic evaluation for herd life in Canada. *Journal of Dairy Science,* **81:** 550-562.

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

Koudandé, O.D., Thomson, P.C., van Arendonk, J.A.M., 1999. A model for population growth of laboratory animals subjected ti marker-assisted introgression: how many animals do we need? *Heredity*, **82:** 16-24.

Koudandé, O.D., Iraqi, F, Thomson, P.C., Teale, A.J., van Arendonk, J.A.M., 2000. Strategies to optimise marker-assitsed introgression of multiple unlinked QTL. *Mammalian Genome*, **11:** 145-150.

Koudandé, O.D., 2000. Introgression of trypanotolerance genes. Doctoral dissertation, Wageningen University.

Lacy, R.C., 1989. Analysis of founder representation in pedigrees: founder equivalents and founder genome equivalents. *Zoo Biology*, **8:** 111-123.

Lacy, R.C., 2000. Should we select genetic alleles in our conservation breeding programs? *Zoo Biology*, **19:** 279-282.

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

Laurie, C.C., True, J.R., Liu, J.J., Mercer, J.M., 1997. An introgression analysis of quantitative trait loci that contribute to a morphological difference between *Drosophila simulans* and *D. mauritana. Genetics*, **145**: 339-348.

Looft, C., Milan, D., Jeon, J.T., Paul, S., Reinsch, N., Rogel-Gaillard, C., Rey, V., Amarger, V., Robic, A., Kalm, E., Chardon, P., Andersson, L., 2000. A high-density linkage map of the RN region in pigs. *Genetics Selection and Evolution* **32:** 321-329.

Lord, E.A., Davis, G.H., Dodds, K.G., Henry, H.M., Lumsden, J.M., Montgomery, G.W., 1998. Identification of Booroola carriers using microsatellite markers (Reprinted). *Wool Technology and Sheep Breeding*, **46**: 245-249.

Lynch, M. and Walsh, B., 1998. In *Genetics and Analysis of Quantitative Traits.* Sinauer Associates Inc., Massachusetts.

Markel, P., Shu, P., Ebeling, C., Carlson, G.A., Nagle, D.L., Smuto, J.S., Moore, K.J., 1997. Theoretical and empirical issues for marker-assisted breeding and congenics mouse strains. *Nature Genetics,* **17:** 280-284.

Martinez, O. and Curnow, R.N., 1992. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theoretical and Applied Genetics,* **85:** 480-488.

Mason, I.L. and Crawford, R.D., 1993. Global status of livestock and poultry species. In: *Managing global genetic resources: Livestock.* National Academy Press, Washington DC, pp 141-169.

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

Milan, D., LeRoy, R., Woloszyn, N., Caritez, J.C., Elsen, J.M., Sellier, P., Genin, J., 1995. The RN locus for meat quality maps to pig chromosome-15. *Genetics Selection Evolution*, **27**:195-199.

Montgomery, G.W., Lord, E.A., Penty, J.M., Dodds, K.G., Broad, T.E., Cambridge, L., Sunden, S.L.F., Stone, R.T., Crawford, A.M., 1994. The Booroola fecundity (*FecB*) Gene Maps To Sheep Chromosome-6. *Genomics*, **22**: 148-153.

Naveira, H., Barbadilla A., 1992. The theoretical distribution of lengths of intact chromosome segments around a locus held heterozygous with backcrossing in a diploid species, *Genetics* **130:** 205-209.

Nei, M., 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the Nation Academy of Sciences USA*, **70:** 3321-3323.

Neimann-Sørensen, A., Robertson, A., 1961. The association between blood groups and several production characteristics in three Danish cattle breeds. *Acta Agriculturae Scandinavica,* **11:** 163-196.

Nicholas, F.W., 1996. In *Introduction to Veterinary Genetics*. Oxford University Press Inc., New York.

Notter, D.R., 1999. The importance of genetic diversity in livestock populations of the future. *Journal of Animal Science*, **77:** 61-69.

Patterson, D.L., 2000. Is there a lifeguard at the gene pool? *Canadian Journal of Animal Science*, **80:** 245-255.

Piper, L.R., Bindon, B.M., Davis, G.H., 1985. The single gene inheritance of high litter size of the Booroola Merino. In *Genetics and Reproduction of Sheep,* Land, R.B. & Robinson, D.W. (eds). Butterworths, London, pp. 115-125.

Pryce, J.E., Simm, G., Coffey, M.P., Kadarmideen, H., 1998. Selection strategies to suit your system. *DRC subject day, Harper Adams University College, Shropshire.* 10th of December, 1998.

Ott, J., 1991. *An Analysis of Human Genetic Linkage.* John Hopkins University Press, Baltimore/London.

Ragot, M., Hoisington, D.A., 1993. Molecular markers for plant breeding: comparisons of RFLP and RAPD genotyping costs. *Theoretical and Applied Genetics,* **86:** 975-984.

Rothschild, M.F., Jacobson, C., Vaske, C.K., Tuggle, C.K., Short, T.H., 1994. A major gene influencing litter size in pigs. *Proceedings of the 5th World Congress on Genetics Applied to Livestock Production,* **21:** 225-228.

Rothschild, M.F., Liu, H.-C., Tuggle, C.K., Yu, T.-P. & Wang, L., 1995. Analysis of pig chromosome 7 genetic markers for growth and carcass performance traits. *Journal of Animal Breeding and Genetics,* **112:** 341-348.

Rothschild, M., Jacobson, C., Vaske, D., Tuggle, C., Wang, L.Z., Short, T., Eckardt, G., Sasaki, S., Vincent, A., McLaren, D., Southwood, O., van der Steen, H., Mileham, A., Plastow, G., 1996. The estrogen receptor locus is associated with a major gene influencing litter size in pig. *Proceedings of the National Academy of Sciences of the United States of America,* **93:** 201-205.

Sawhney, R.N. and Sharma, J.B., 1996. Introgression of diverse genes for resistance to rusts into an improved wheat variety, *Kalyansona. Genetica,* **97**: 255-261.

Sax, K., 1923. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris. Genetics,* **8:** 522-560.

Smith, J., Bruley, C.K., Paton, I.R., Dunn, I., Jones, C.T., Windsor, D., Morrice, D.R., Law, A.S., Masabanda, J., Sazanov, A., Waddington, D., Fries, R., Burt, D.W., 2000. Differences in gene density on chicken macrochromosomes and microchromosomes. *Animal Genetics* **31:** 96-103.

SNP Home Page, 2001. URL: http://www.ncbi.nlm.nih.gov/SNP/index.html, [18th of September, 2001].

Soller, M., 1991. Mapping quantitative trait loci affecting traits of economic importance in animal populations using molecular markers. In *Gene Mapping Techniques and Applications*, Shook L.B., Lewin H.A., McLaren D.G. (Eds.), Dekker, New York, pp. 21-49.

Spelman, R.J. & van Arendonk, J.A.M., 1997. Effect of inaccurate parameter estimates on genetic response to marker assisted selection in an outbred population. *Journal of Dairy Science,* **80:** 3399-3410.

Stam, P. and Zeven, A.C., 1981. The theoretical proportion of the donor genome in near isogenic lines of self-fertilizers bred by backcrossing. *Euphtica,* **30:** 227-238.

Stam, P., 1980 The distribution of the fraction of the genome identical by descent in finite random mating populations. *Genetical Research,* **35:** 131-155.

Tanksley, S.D., 1993. Mapping polygenes. *Annual Review of Genetics,* **27:** 205-233.

Tanksley, S.D., & Nelson, J.C., 1996. Advanced backcross QTL analysis: A method for the simultaneous discovery and transfer of valuable QTLs from unadapted germplasm into elite breeding lines. *Theoretical and Applied Genetics,* **92**: 191-203.

Tanksley, S.D. and McCouch, S.R., 1997. Seed banks and molecular maps: Unlocking the genetic potential from the wild. *Science,* **277:** 1063-1066.

Toojinda, T., Baird, E., Booth, A., Broers, L., Hayes, P., Powell, W., Thomas, W., Vivar, H., Young, G., 1998. Introgression of quantitative trait loci (QTL) determining stripe rust resistance in barley: an example of marker-assisted line development. *Theoretical and Applied Genetics*, **96**: 123-131.

van der Beek, S. & van Arendonk, J.A.M., 1996. Marker-assisted selection in an outbred poultry breeding scheme. *Animal Science,* **62:** 171-180.

van Heelsum, A.M., Visscher, P.M., Haley, C.S., 1997a. Marker-assisted introgression using non-unique marker alleles. I: Selection on the presence of linked marker alleles. *Animal Genetics*, **28**: 181-187.

van Heelsum, A.M., Haley, C.S., Visscher, P.M., 1997b. Marker-assisted introgression using non-unique marker alleles. II: Selection on probability of presence of the introgressed allele. *Animal Genetics*, **28**: 188-194.

Visscher, P.M., Haley, C.S., Thompson, R., 1996. Marker-assisted introgression in backcross breeding programs. *Genetics,* **144:** 1923-1932.

Visscher, P.M., Haley, C.S., 1999. On the efficiency of marker-assisted introgression. *Animal Science,* **68:** 59-68.

Visscher, P.M., 1999. Speed congenics: accelerated genome recovery using genetic markers. *Genetical Research*, **74:** 81-85.

Walling, G.A., 1998. Detection and mapping of quantitative trait loci in domestic farm livestock species. *Doctoral thesis,* University of Edinburgh, Scotland.

Walling, G.A., Archibald, A.L., Cattermole, J.A., Downing, A.C., Finlayson, H.A., Nicholson, D., Visscher, P.M., Walker, C.A., Haley, C.S., 1998. Mapping of quantitative trait loci on porcine chromosome 4. *Animal Genetics*, **29**: 415-424.

Walling, G.A., Archibald, A.L., Visscher, P.M., Haley, C.S., 1998. Mapping genes for growth rate and fatness in a Large White x Meishan F_2 pig population. *Proceeding of the British Society of Animal Science, 1998.* pp 7.

Walling, G.A., Dodds, K.G., Galloway, S.M., Beattie, A.E., Lord, E.A., Lumsden, J.M., Montgomery, G.W., McEwan, J.C., 2000. The consequences of introducing the Booroola Fecundity *(FecB)* gene on liveweight. *The proceedings of the British Society of Animal Science, 2000.* pp 43.

Wakeland, E., Morel, L., Achey, K., Yui, M., Longmate, J., 1997. Speed congenics: a classic technique in the fast lane (relatively speaking). *Immunology Today,* **18:** 472-477.

Wang, J.L., 2001. Optimal marker-assisted selection to increase the effective size of small populations. *Genetics*, **157:** 867-874.

Wang, J.L. and Hill, W.G. 2000. Optimal marker-assisted selection to increase effective population size by reducing Mendelian segregation variance. *Genetics*, **154:** 1475-1489.

Weil, M.M., Brown, B.W., Serachitopol, D.M., 1997. Genotype selection to rapidly breed congenics strains. *Genetics,* **146:** 1061-1069.

Weller, J.I., 1990. Experimental designs for mapping quantitative trait loci in segregating populations. *Proceedings of the 4th World Congress on Genetics Applied to Livestock Production*, **XIII:** 113-116.

Westman, A.L., Kresovich, S., 1997. Use of genetic marker techniques for the description of plant genetic variation. In *Biotechnology and Plant Genetic Resources: Conservation and Use,* Callow, Ford-Lloyd and Newbury (eds.), CAB International, Oxford, UK.

Wilson, T., Wu, X-Y., Juengel, J.L., Ross, I.K., Lumsden, J.M., Lord, E.A., Dodds, K.G., Walling, G.A., McEwan, J.C., O'Connel, A.R., McNatty, K.P., Montgomery, G.M., 2001. Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogentic protein 1B receptor (ALK-6) that is expressed in both oocytes and granulose cells. *Biology of Reproduction*, **64:** 1225-1235.

Woolliams, J.A., Bijma, P., 2000. Predicting rates of inbreeding in populations undergoing selection. *Genetics*, **154:** 1851-1864.

Woolliams, J.A., Bijma, P., Villanueva, B., 1999. Expected genetic contributions and theit impact on gene flow and genetic gain. *Genetics*, **153:** 1009-1020.

Wright, S., 1969. *In* Evolution and the genetics of populations. Vol .II: The theory of gene frequencies. Chicago, University of Chicago Press.

Yancovich, A., Levin, I., Cahaner, A., Hillel, J., 1996. Introgression of the avian naked neck gene assisted by DNA fingerprints. *Animal Genetics*, **27**: 149-155.

Young, N.D., Tanksley, S.D., 1989. RFLP analysis of the size of chromosomal segments retained around the *tm*-2 locus of tomato during backcross breeding. *Theoretical Applied Genetics*, **77:** 353-359.

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

Zeven, A.C., Knott, D.R.,Johnson R., 1983. Investigation of linkage drag in near isogenic lines of wheat be testing for seedling reaction to races of stem rust, leaf rust and yellow rust. *Euphytica* **32:** 319-327.

Zhang, Q., Boichard, D., Hoeschele, I., Ernst, C., Eggen, A., Murkve, B., Pfister Genskow, M., Witte, L.A., Grignola, F.E., Uimari, P., Thaller, G., Bishop, M.D.,

1998. Mapping quantitative trait loci for milk production and health of dairy cattle in a large outbred pedigree. *Genetics*, **149:** 1959-1973.