The Characterisation of a Longevity QTL in the New Zealand Holstein Friesian Population

A thesis presented in partial fulfilment of the requirements for the degree of Masters of Science in Genetics at Massey University, Palmerston North, New Zealand.

Jennifer Marie Anderson
2002

Abstract

A large scale genome scan analysis to identify QTL relevant to commercial dairy cattle in New Zealand was undertaken using 7 large grandsire families. From this experiment a QTL site situated on bovine chromosome 23 (BTA23) for longevity was identified. The present study investigated this QTL in the subsequent generations.

The addition of more markers to the region under question and the addition of grandsons to the analysis helped support the evidence that a QTL for longevity was indeed present within the region. However a more precise location was not identified. Although this is not helpful for marker assisted selection it will not affect the candidate gene search until the bovine genome sequence is available. This is due to the large inversion event within the BTA23 which means the target region in a comparative map search with HSA6 must still include the entire chromosome.

Analysis of the granddaughters within these families confirmed a link between the speculative QTL site and variation in herd life but could not identify a cause for this variation at a phenotypic level despite a data set of over 800,000 animals. This result indicates that the variation in longevity is most likely the product of a variation in disease resistance at a sub clinical level. As ill health would impact on all production traits, animals affected would be removed for a variety of reasons. Because the only check of health in dairy cattle is their ability to be productive and remain in the herd it is impossible to identify these problems unless animals die from them.

The MHC gene cluster lies within QTL identified and was the prime candidate for linked genes. Analysis of the DRB3 region in the two grandsire families showed a similar genotype in both grandsires. Genotype 1201/090_ was common to both grandsires. Further analysis of the DRB3 by restriction endonuclease digest in the sons and grandsons showed that allele 1201 or alleles similar to 1201 were common in the population whilst alleles 090_ were not seen as often. Variation in phenotype for the 090_ allele suggested a more complex model than a simple inferior/superior allele relationship.

Acknowledgements

Firstly I would like to thank Vialactia and Livestock Improvement for providing me with the opportunity and funding for this project, especially my supervisor Dr Richard Spelman for all your support and guidance.

Thank you to my Massey supervisor Dr Kathryn Stowell for your day to day encouragement and words of wisdom. Thank you also for completing the mammoth task of proofing this thesis. I know how you feel about cows and I appreciate every minute you had to spend thinking about them! A big thank you also to Prof Hugh Blair and Prof Dorian Garrick for your input into the set up and analysis of the project.

To all the guys in the Twilight Zone! What would my days of been like without you? Thanks for making it a pleasure to come to the lab and keeping me smiling! Thank you to Carol Flyger for running the lab so efficiently and keeping us all in line. To Robyn, Kirsty, Tash, Brenda and Chrisse a special thanks for your friendship and support and for keeping me well stocked with caffeine at the right moments! Thank you also Vikki for your friendship and infectious optimism. To all the newbees in the lab: GOOD LUCK!

Thank you to all the IMBS members and Massey University for creating a wonderful environment and providing me with this opportunity.

A big thanks to my family for their support during my time at Massey. It would have been impossible to study without the knowledge that you were 100% behind me.

A special thanks to all my flatmates especially Russell and Rossy for cooking and cleaning up after me in my moments of stress. A special thanks to Russell for all your support and IT advice.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AI	Artificial Insemination
B cells	Bone lymphocyte cells
BoLA	Bovine Leukocyte Antigen
bp	Base pairs
BTA	Bos Taurus
BV	Breeding Value
cM	Centi Morgan
COMPASS	Comparative Map Analysis of similar sequences
DFM	Days from First Mating
DNA	Deoxyribonucleic acid
ER	Endoplasmic Reticulum
EST	Expressed sequence Tag
FISH	Fluorescent *in situ* Hybridization
HAS	Homo sapiens
HLA	Human Leukocyte Antigen
LB	Luria Broth
LIC	Livestock Improvement Corporation
MARC	Meat Animal Research Centre
MAS	Marker Assisted Selection
Mb	Mega bases
MHC	Major histocompatibility complex
ML	Maximum Likelihood
OMIA	Online Mendalian Inheritance in Animals
OMIM	Online Mendelian Inheritance in Man
QTL	Quantitative Trait Locus
RFLP	Restriction Fragment Length Polymorphism
SNP	Single Nucleotide Polymorphism
SSCP	Single Stranded Conformational Polymorphisim

T cells	Thymus cells
TAP	Transporter for Antigen processing Protein
TCR	T cell receptor

Chapter 1 Introduction

1.1.1 Longevity in Dairy cattle

Longevity, or herd life, is defined as the length of productive life an animal leads in a herd, from first calving to the last herd test. As herd life is controlled primarily by economic factors it is unlikely that there will be an opportunity for an animal to reach old age, considered to be around 20 years. Rather, animals will be culled or removed for other reasons at approximately 4 years of age (Esselmount and Kossaibati, 1997). As farms are forced to become increasingly competitive and minimise input costs the benefits of a long production life are becoming increasingly important. It is for this reason that longevity has been added to the breeding goals for most sire proving systems. However, longevity is not a simple phenotype to define and the strong influence of farm practises and environment make it difficult to separate out a genetic component without analysing large numbers of offspring. There is also the problem that most of the other traits measured in sire proving schemes are likely to also influence the longevity of an animal.

The costs of screening sires through a sire proving scheme are large; bulls, after sexual maturity, need to sire a number of daughters, which in turn need to reach sexual maturity themselves before calving and subsequent milking can be undertaken. This process means at least five years must elapse before a reliable sire breeding value (BV) can be calculated for most production traits and even then there is no guarantee that the sire will be suitable for further use. Longevity BVs can take even longer to calculate due to the need to identify the time of herd removal for the majority of the offspring. To address these issues, research into marker assisted selection (MAS) has been undertaken in the hope that systems will be able to be implemented to genetically test for relevant trait variation at an early age, possibly before embryo implantation (Spelman and Garrick,

1998; Haley and Visscher, 1998). However, to use these markers confidently they need to be extremely close in proximity to the relevant genomic regions and for total reliability the test needs to be centred around the actual gene itself, making its identification vital.

1.1.2 Reasons for disposal

A dairy cow's productive life can be terminated for a number of reasons including, low fertility, disease, low production, type traits, ease of milking and accidents. In most studies infertility is the leading cause of disposal, especially in younger animals. The second most common culling factor is either mastitis or low production, varying in individual studies (Stevenson and Lean, 1998; Bascom and Young; 1998, Seegers *et al.*, 1998 and Whitaker *et al.*, 2000).

Difficulty in identifying the true reason for herd removal can be compounded by the fact that an animal is usually culled for multiple reasons, especially in the case of voluntary culling. Low production may be the primary factor but type may also influence the decision. Studies have attempted to address this problem by allowing farmers to give primary, secondary and tertiary reasons for disposal and weighting accordingly (Bascom and Young, 1998). However, this approach is still open to a large amount of interpretation by the farmer and is subject to individual culling choices. Within herd effects are also extremely important in influencing the cause of disposal; animals that are less favoured by the farmer may not be given as many services before being culled for fertility reasons and different type traits may have higher priority with different farmers. Expanding or closed herds will also contain variables in the rate of culling. Environmental effects may influence the occurrence of disease including mastitis and the quality of herd mates will also affect an individual's chances of voluntary culling (Lehenbauer and Oltjen, 1998).

Reasons given for culling are usually divided into two groups according to whether the culling was involuntary or voluntary. Involuntary culling, often caused by infertility, disease or accident is usually easy to identify. However, voluntary culling practises for traits such as production, temperament and conformation are usually more difficult to classify as these are determined by the individual farmer. Voluntary culling is always likely to be a factor in herd management and genetic improvement. The emphasis on different traits will change with farm and year.

Although reasons for culling may vary, the average annual culling rate in a commercial herd is typically between 22-24%. (Esselmount and Kossaibati, 1997, Whitaker *et al.*, 2000 and Karuppanan *et al.*, 1997). As most animals are culled for involuntary reasons further research into the genetic component of the longevity trait may be economically beneficial.

1.1.3 Calculating Breeding Values

Breeding values are calculated for most economically important traits in commercial livestock. They are calculated as the genetic advantage a sire or dam is expected to pass on to its offspring relative to the normal population. For example, if a bull sires calves 1 kg heavier, on average, than the normal population its offspring are likely to produce calves 0.5 kg heavier than the average (only half the estimated breeding value is inherited from the sire). This gives the estimated BV for the offspring without the need to test the offspring first. If a bull's estimated breeding values indicate an advantage it is likely that it will be bred to more dams. The information gathered from these matings will then lead to identifying true breeding values. Obviously the more offspring a bull sires the more accurate its BV figures will be.

There has been much debate on the most appropriate method for measurement of longevity for breeding purposes. The obvious correlation between productivity and herd

life makes it difficult to discern whether an individual is genetically superior for longevity or whether the extended herd life is a product of other economically relevant traits. Functional longevity, defined as the ability to delay involuntary culling, is the ideal breeding goal (Ducrocq, 1987). True longevity, the ability to delay all culling, is the real combined effect of both longevity traits as well as production factors.

Unfortunately, the measurement of functional longevity is difficult to estimate as there are so many contributing factors. Many argue that attempting to calculate this accurately is impractical. Therefore, true longevity values are preferred and in most studies longevity BVs are calculated as the difference in the productive life of the daughters compared to a baseline average. The use of productive, as opposed to total lifespan is usually preferred with the reasoning that this will not disadvantage early calving animals. Although some suggest that using productive life makes the comparison across herds difficult with different management practises affecting the ages at first calving (Karuppanan *et al.*, 1997). This difference can be easily taken into account with the calculation of herd effect.

When dealing with longevity BVs the issue of living daughters needs to be addressed. Obviously it is not practical to wait for all daughters to be removed from the herd before calculating the BV for longevity. Therefore methods have been developed to allow living daughters to be assessed for longevity. These living daughters, or censored records, are added to the analysis according to a number of factors, such as the length of life the animal has already lived, the production of that animal and the estimated breeding values of its sire and dam. Collectively these data help formulate an expected lifespan and will change as the animal lives through each season.

Calculation of breeding values for longevity can be determined using several different methods, however, at present the use of survival analysis is considered the most appropriate due to its ability to use censored records and time dependent covariates can also be added to the analysis (Vollema and Groen, 1997). The survival analysis approach is based on the concept of a hazard rate, or the probability of being culled at a certain

time given that the animal has lived up to that time. This is usually modelled as a product of a baseline hazard function based on natural aging processes, with the effects of influential factors, such as herd, year and production included as added functions according to the individual (Ducrocq and Solkner, 1994)

1.1.4 Economic Importance of Longevity

Longevity is of economic importance in terms of replacement costs, the ability to have a higher percentage of older higher producing animals in the herd and in terms of genetic gain allowed by more voluntary culling practises (VanRaden and Wiggans, 1995). With the increasing pressures of today's economic environment it is becoming more and more important in farm management to reduce the costs and maximise profits. Replacement costs are a major input cost to the farmer and their reduction will help to increase net profit. When calculating the advantage of replacement costs several factors need to be taken into account: the expected cost of the replacement (Tozer and Heinreichs, 2001), the price collected on the culled animal from meat sales, and the overall production expected from the replacement. Other costs in terms of medical treatment, time and age of the culled animal may also need to be assessed depending on the circumstances. Genetic gain in the herd will probably be a minimal factor due to the fact that most genetic improvement within a commercial dairy herd is achieved by the sires and not the dams. For an economic advantage to be achieved the replacement animal must be of a sufficiently higher standard than the culled animal such that the increased production will cover the costs of premature replacement. Obviously, replacement would usually only be worthwhile in very low producing animals.

Despite the variables involved in culling decisions it is clear that as the costs of replacements decrease, the economic advantage of voluntary culling increases (Rajala-Schultz *et al.*, 2000). However, this is not a linear function as decreases in replacement costs usually correspond to a slump in the industry and therefore an increase in

production at this time may not be as advantageous as expected. In the formulation of a culling-for-profit strategy the culling should not be based on whether an animal is profitable but whether a replacement cow is expected to be more profitable. However this assumption only works well in a fixed herd when the addition of a new animal means the removal of an original herd member. In expanding herds the decision to cull may be postponed.

Involuntary culling is the main cause for removal from a commercial herd, usually in the form of infertility or mastitis. Costs in these cases are the loss of potential production and replacement of the culled animal. These fates do not necessarily affect the weaker producing animals and therefore the costs are often greater than a voluntary cull.

Diseases have also been shown to affect the culling process. Animals affected by conditions such as mastitis, milk fever and retained placenta are more likely to be removed than their healthy herd mates. Often this is seen as voluntary culling in terms of reduced production or fertility, however involuntary culling through disease also occurs. (Grohn *et al.*, 1998).

1.2 : Identifying Chromosome regions influencing longevity

1.2.1 Genetic tools

A quantitative trait locus (QTL) is defined as a region within the genome that contributes to the variation of a continuously distributed trait. If the trait affected is economically important, such as longevity, these regions are targets for identification and characterisation for further use in animal or plant improvement systems. The discovery of polymorphic genetic markers, such as microsatellites (randomly repeated di-or

trinucleotides), SNPs (single nucleotide polymorphisms) and RFLPs (restriction fragment length polymorphisms) within the genome have made it possible to attempt to determine linkage of these QTLs with particular markers. This, thereby, enables the position of the QTL in a genome to be predicted. Obviously, to do this the variation must be of a significant magnitude and the markers close enough to the locus.

Markers are chosen for position and heterozygosity. Although laboratory methods vary according to the marker used, essentially all markers must be amplified via the polymerase chain reaction (PCR) using specific primers. Markers are then typed according to fragment size which can be measured on a simple agarose or polyacrylamide gel or using an automated DNA sequencer. SNP analysis can be achieved using a number of different methods, however most involve the use of a sequence-specific probe.

Haplotyping a specific region of a chromosome in an individual can be achieved by following the linkage of specific marker alleles with neighbouring marker alleles along the chromosome in the next generation. This method makes use of recombination which occurs naturally almost at random throughout most of the genome. Although most markers used in haplotyping have known genetic positions, novel markers can be placed on the genetic map according to the inheritance of a specific allele with alleles of other known markers. For example, if an allele appears to be inherited 99 out of 100 times with another allele it can be said that the markers are approximately 1 cM (centimorgan) apart. Measurements in cM can be used to generate a genetic map. More accurate physical maps can be achieved through sequencing or cytogenetic maps, which are generated by FISH (fluorescent *in situ* hybridisation) analysis. Genetics maps are not as accurate due to the fact that they have a tendency to vary according to chromosome position, sex and species. This is due to the presence of recombination hotspots or areas in the genome that appear to have reduced recombination rates. The reason for these hot spots is not fully understood but could be due to chromosome conformational changes. Generally however, in a human male it appears that 1 cM represents approximately 1,100 kbps (kilo base pairs). Despite the problems with the accuracy of genetic maps, they are still suitable for QTL studies because these are based simply on the probability of inheriting a given

chromosome region at the same time as inheriting a given marker allele. In this case, the likelihood of recombination is more important than the physical distance between the marker and the QTL, although the likelihood of recombination is governed principally by physical distance.

Family size must be sufficiently large to detect QTL. Initial studies in QTL linkage were carried out in plants such as maize (Stuber, 1995) and barley (Hayes *et al.*, 1993) which were capable of producing large second generations. In recent years the study of QTL has moved to livestock such as cattle (Hayley, 1995) where artificial insemination (AI) has allowed single sires to produce large numbers of offspring.

After successful QTL detection and verification, candidate genes can then be identified within the locus. With the first draft of the human genome completed and published (International Human Genome Sequencing Consortium, 2001) comparative genetic maps are now available between bovine and human (Band *et al.*, 2000). Comparative maps allow the study of potentially important bovine genes that have not yet been characterised. Comparative mapping techniques work on the basis that if known genes on a given bovine chromosome appear to be in a similar order and position to homologues on the human chromosome there is a good chance that the genes flanking this region in the human will also be present at the same relative location in the bovine sequence. Using this rationale, candidate genes can be chosen not only from the known bovine genes but also from the genes identified in the human genome. Public bovine EST (expressed sequence tag) databases are also growing and these aid in confirming the existence of bovine homologues of human genes, as long as they are expressed within the bovine tissues examined (Smith *et al.*, 2001; Takasuga, *et al.*, 2001; Rebiez and Lewin, 2000).

1.2.2 QTL Studies

A successful QTL search relies on several factors; the locus under observation and the linked marker(s) must be heterozygous in at least one parent, the variation at the locus must be sufficient to allow significant phenotype differences between the two haplotypes, and the marker map must be dense enough to allow markers to segregate with the locus (Fig.1.1).

Procedures commonly used in QTL analysis to identify genome regions contributing to phenotype variation are the simple t test, linear regression analysis, nonlinear regression and likelihood approaches. Although t tests can be carried out using only the marker data it is possible to use the regression (Knapp *et al.*, 1990; Knott and Haley, 1992; Martinez and Curnow, 1992; Jansen, 1992) or maximum likelihood (ML) (Lander and Botstein, 1989; Jansen, 1993) methods in conjunction with a linkage map for interval mapping techniques. Interval mapping works by defining one putative QTL allele as 1 and the other as 0. The probability of inheriting allele 1 is then calculated for every cM that is not represented by a genetic marker based on the flanking marker haplotype and the chances of recombination between the marker and the QTL. By doing this the QTL probability at each cM can be predicted and used to identify a QTL position that is not represented by actual markers.

Figure 1.1. Marker and QTL relationship. Markers and QTLs must be heterozygous for QTL identification

Although there is still dispute between groups over the use of multiple regression versus ML techniques both methods continue to be regularly used in cattle. ML was the initial choice for QTL studies (Botstein and Lander, 1989) but lost favour over regression analysis techniques (Hayley and Knott, 1992) as these appeared computationally simpler

and had the ability to cope with more complex models. Simulated models testing the two methods have illustrated shortfalls with the regression method in terms of its ability to cope with closely linked and interacting QTL (Kao, 2000). In practical studies where both methods have been trialled on real data the results have been similar for both types of analysis (Elo *et al.*, 1999).

An important issue in QTL studies is the calculation of significant threshold levels for the test statistic, as the assumption of normal distribution and independent testing does not necessarily apply to QTLs. The simplest way to achieve these levels is using the permutation method (Churchill and Dorge, 1994) which works essentially by randomly shuffling the phenotype data against the genotype data and calculating the test statistic. When this is carried out a large number of times, at least 1,000 – 10,000 times, the likelihood of a false positive can be estimated.

1.2.3 Experimental Design

Original QTL studies tended to focus on the inheritance of genome regions using either a backcross experiment or through an F2 mating (Knapp, 1991) (A and B, Fig.1.2). In these

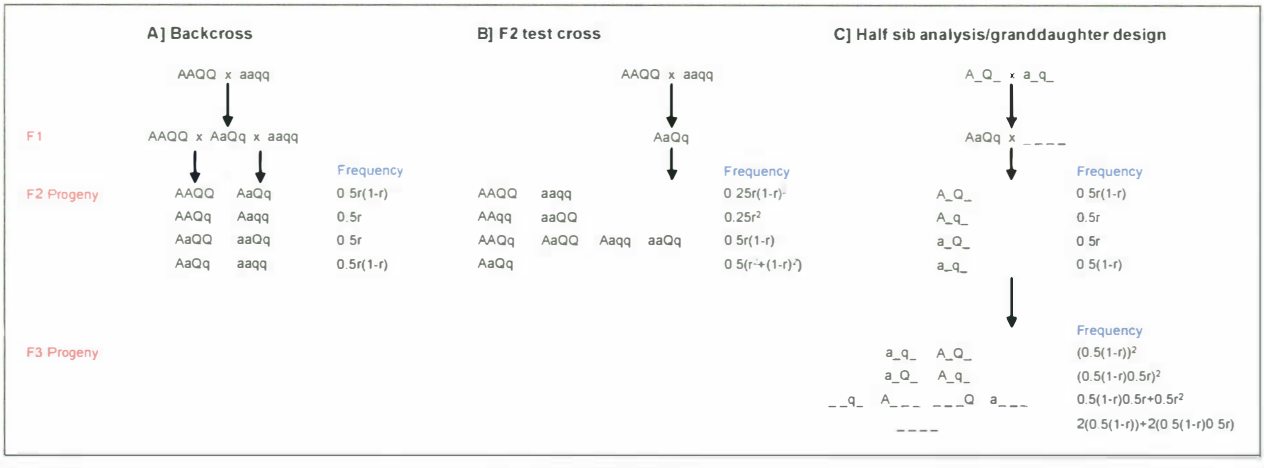


Figure 1.2. Experimental designs for QTL identification A) Backcross, design involving inbreeding to the parental line. **B)** F2 cross where known homozygous parents are used to generate an F1 which is then inbred. **C)** Half-sib analysis where parents carrying different alleles generate a heterozygous sire which in turn mates with a sample of the population. Granddaughter design uses the phenotype variation in the F3 to identify genetic variation in the F2.

studies both parents are chosen for having different phenotypes and therefore assumed to be carrying different alleles at the appropriate locus. Although these experiments are easy to set up using plants they are not practical for animal studies. Firstly, it is impossible to use a single dam to breed the large family necessary and secondly setting up a family line purely for QTL study is extremely expensive in livestock. Therefore it is desirable to design an experiment around existing family lines. Experiments can be set up around controlled mating using different breeds to maximise differences and therefore allow more QTL to be identified. However, this can lead to problems when practical breeding decisions need to be addressed. In addition, most of the QTL identified are highly likely to be fixed in the given breeds.

Half sib analysis is still the only practical option for detecting QTL responsible for variation in economic traits within existing commercial populations of livestock. Fortunately, in the dairy industry large half sib families are available for study. Widely used bulls might be responsible for siring as many as 100,000 daughters. This allows daughter design experiments to be set up (de Koning *et al.*, 2001, Geogres *et al.*, 1995). In this situation, a sire is common to a large number of daughters but the dams used are different. Dividing the daughters according to the marker haplotype inherited from the sire at each genome position allows a QTL search to be carried out. The dam affect is assumed to be equal because of the large number of daughters contributing to each haplotype group. However, problems arise in these experiments due to the expense, difficulty in collecting DNA samples and genotyping the daughters. Attempts are being made to minimise these costs by using milk samples collected at herd testing. DNA can be extracted from cells in milk samples, but the costs are still large for this type of design due to the large number of daughters that are needed. In addition collecting uncontaminated DNA samples from commercial farms can be difficult and time consuming.

In cattle some of the formentioned problems have been addressed using the granddaughter design experiment (C, Fig.1.2.) (Weller *et al.*, 1990). High performance sires may have a large number of sons that can also be used extensively. The

granddaughter design assumes that because of the large number of daughters contributing to the breeding value of these sons that they are an accurate measure of the sons' genetic influence on the trait under study. Under this assumption, only the grandsire and the sons need be genotyped saving considerable time and expense and allowing a larger number of offspring to contribute to the experiment. This method allows the same power with approximately ¼ of the assays for genotyping compared to the daughter design.

Another method that is becoming more prevalent in QTL studies is the use of selective DNA pooling (Lipkin *et al.*, 1998). Using this method, DNA samples need only be taken from the extreme high and low phenotypic daughter groups of a large half-sib family. These samples are then pooled into the two groups and the allele frequencies in each group measured according to densitometric intensity. If a given allele appears to be over represented in a given outlier group there is then a possibility the marker is linked to a QTL. Use of this method to identify QTL for milk protein percentage has shown potentially useful results when compared to other traditional studies (Mosig *et al.*, 2001)

1.2.4 QTL studies on BTA 23

A significant QTL for longevity has been identified on bovine chromosome 23 (BTA23) in a granddaughter design experiment involving 7 sire families (unpublished, Spelman, 2000). Within family analysis identified 2 of the 7 sires to have significant segregation around marker BM1258. A number of large scale QTL searches have been attempted in dairy cattle throughout the world with BTA23 being the focus due to the presence of the bovine MHC (major histocompatibility complex) on the q arm of chromosome 23. The MHC region, commonly known as the bovine leukocyte antigens (BoLA) in cattle and HLA (human leukocyte antigens) in humans, has been intensively studied in the past for its possible role in disease resistance, fertility and production-related trait variation in a large variety of domestic and wild animals (Lewin, 1999). This is due to a wide array of polymorphisms within the genes in the MHC region making it one of the most diverse

coding regions in the genome and therefore a prime target for association studies between the genes themselves and molecular markers along the chromosome.

Single marker analysis studies in cattle have shown suggestive linkage between BTA23 markers and several economically important traits. For example, possible QTL affecting somatic cell count, udder depth and herd life have been identified in American Holstein populations (Ashwell *et al.*, 1996 and Ashwell *et al.*, 1998).

Other studies that have not looked at herd life directly have shown associations with traits and BTA23. A QTL regulating live weight was reported in Finnish Ayrshire with a likely position between markers BM1258 and BoLADRBP1. This study also reported QTL on BTA23 at lower than the 5% threshold level in the across family analysis for milk protein %, calf mortality, milking speed and veterinary treatment (Elo *et al.*, 1999). A QTL for fat yield has also been suggested on BTA23, although in that particular study the 5% threshold level was not exceeded (Zhang *et al.*, 1998). No significant QTL for productive life were reported although it was one of the seven traits analysed.

Studies looking at milk traits have identified likely QTL along BTA23 for protein yield and percentage, milk yield and fat yield and percentage (Cowan *et al.*, 1990; Sharif *et al.*, 1998; Starkenburg *et al.*, 1997). Although discrepancies exist in these studies over the exact QTL locations, all studies place QTL within the middle region of chromosome 23. Of course this is most likely due to the information content at the ends of the chromosome being significantly lower, especially in the earlier studies but it does show evidence that MHC related genes may be associated with significant variation in production traits.

1.3: The Bovine MHC

1.3.1 Overview of the Role of the MHC

The immune system is vertebrates protection against the wide variety of antigenic organisms and substances it encounters in the environment. It can be split into two pathways depending on the type of response the antigen stimulates. The cell mediated response, or the cellular response is carried out by the Killer T cells that work to engulf and destroy foreign bodies or infected cells. The antibody-mediated response, or the humoral response, works to stimulate B cells which in turn convert to plasma cells and begin the expression of specific antibodies which bind to the antigen and neutralise it. Not all invading bodies are able to elicit an immune response. To do this the invader must first be recognised by the immune system. This is the role of the MHC.

The MHC is the name given to a large group of genes whose protein products are involved in the presentation of peptides at the cell surface for identification by the immune system. The MHC class I (MHC I) and II (MHC II) molecules bind to short peptides produced by degradation in either the proteasome or within endosomes respectively. Once bound, the complexes are then transported to the cell surface for T cell recognition and initiation of the appropriate immune response (Figure 1.3.). Both self and nonself peptides are presented by the MHC molecules, however usually only the nonself peptides will stimulate an immune response.

Although both classes of the MHC work in a similar manner they vary in the type of antigens they predominantly present. Whilst MHC II proteins tend to be responsible for exogenous antigen presentation, such as from bacteria, MHC I proteins usually display peptides derived from endogenous viruses and self proteins. Their different targets mean that when investigating resistance to particular antigens in practical studies different

MHC genes will be of more interest. For example in a bacterial study the MHC II genes would be the target. However, there are always exceptions to the rule and sometimes bacterial peptides will be displayed by MHC I molecules. This usually occurs if the cells have engulfed parts of the bacteria at an earlier stage.

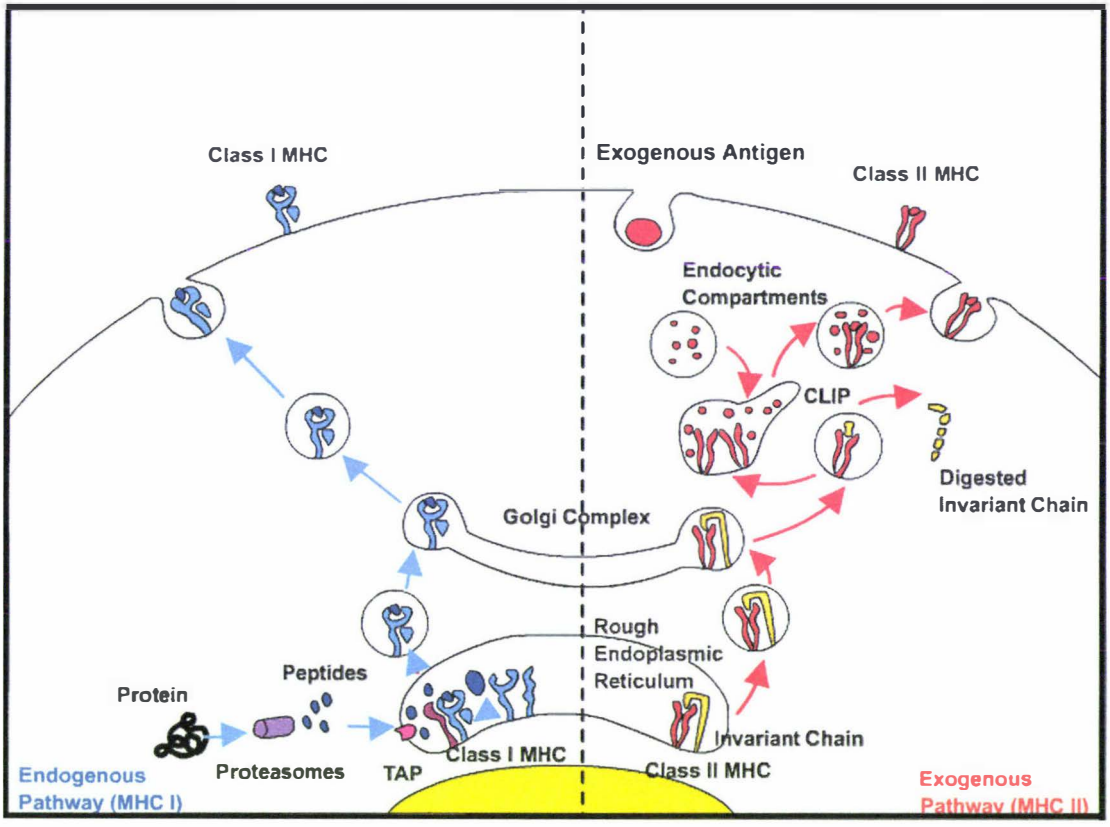


Figure 1.3 Role of the MHC presentation molecules (adapted from Goldsby *et al*. 2000)

MHC I molecules are expressed at some level in nearly all cells in the body. Peptides that are synthesized in the proteasome are exported to the endoplasmic reticulum (ER) using the transporter for antigen processing protein (TAP). The MHC I associate with the TAP within the ER until binding of the peptide.

MHC II molecules are generally only expressed in B-cells, macrophages and dendritic cells. Rather than binding with peptides in the ER, MHC II molecules become bound to an invariant chain protein (CLIP) and are transported via the endosomal route. It is in

these secondary compartments that the class II molecules are released from the invariant chain via interactions with other MHC related molecules and other proteases, and allowed to associate with peptides.

Proteins within an endosome are sourced from the intracellular environment via endocytosis whilst the peptides produced via the proteasome are from proteins within the cell. This is the major biological difference between the two pathways.

1.3.2 MHC genes

In humans the MHC genes are found on chromosome HSA6 spanning approximately 4 million bases in the p arm. The region is divided into three according to the genes present. The class I region is located closest to the centromere and spans almost 1.8 Mb (Shian *et al.*, 1999) followed by the class III region (the largest region) spanning 2 Mb, and then finally the class II. The class III region contains a number of genes involved in the immune system but none that are directly involved in antigen presentation.

The class I region is made up of 118 genes in the human including 73 of a known function. Three of these (HLA-A,-B, and -C) are genes for the classical class I peptide binding proteins. These are similar in sequence to the non classical -E, -F and -G genes but only the classical genes are regularly expressed in most cells. Other genes found in the class I region are the MICA and MICB genes; although similar in structure they appear to be expressed only in times of stress (Ota *et al.*, 2000). MICA is considered the most polymorphic gene of the class I region in human.

The class II region is approximately 1 Mb in size and contains the genes for another set of peptide binding proteins. The four classical MHC II genes include three DRB and one DRA. The DRB genes are the most polymorphic in the entire human genome and most of

this variation is seen in the region coding for the peptide binding sites of these molecules. This suggests that there is a benefit to maintain this variation within a population.

The BoLA is sufficiently similar to the human MHC (HLA) to allow primer design from the human DNA sequence for successful amplification of the conserved MHC genes. The BoLA can be divided into three gene groups. However, early linkage studies suggested a large genetic gap between the majority of the BoLA genes and a number of class II genes. Radiation hybrid maps (Band *et al.*, 1998) and FISH (fluorescence *in situ* hybridisation) studies (Hess *et al.*, 1999) showed that this genetic distance was also physical and the BoLA class II genes have been separated from class I and III by a large

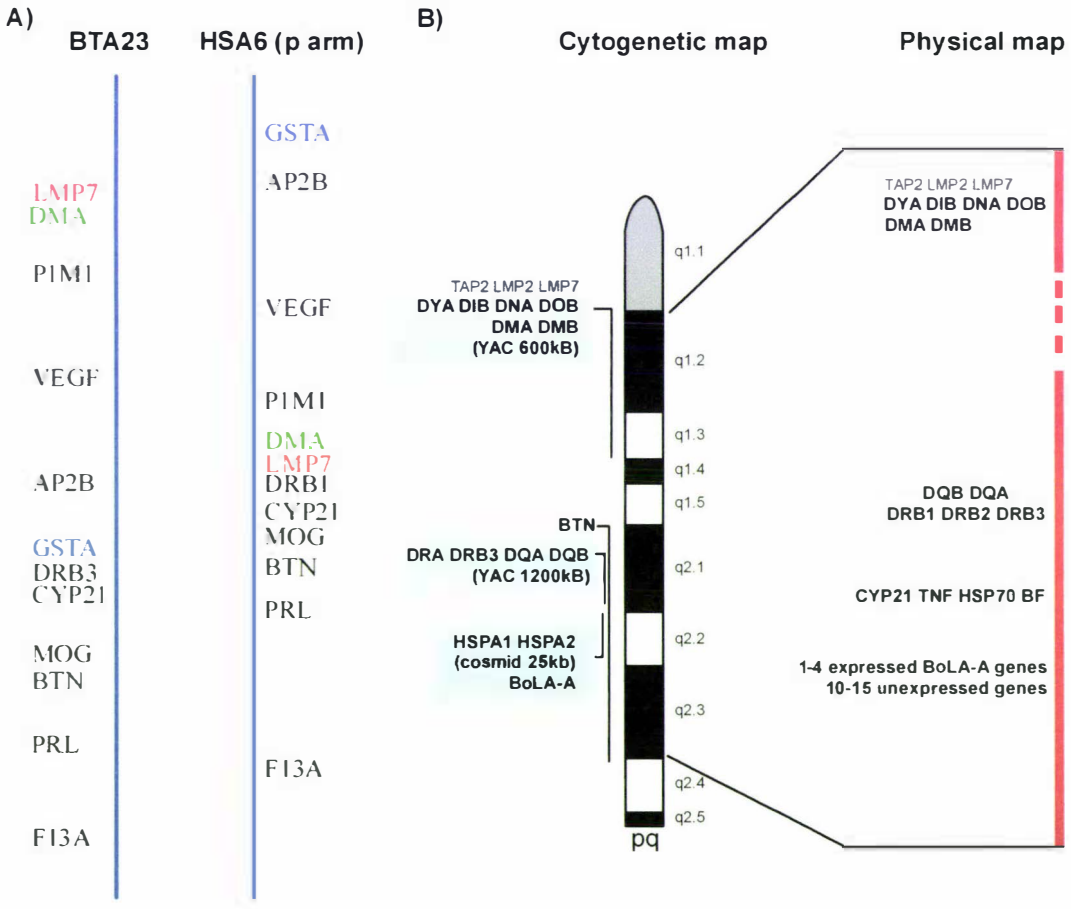


Figure 1.4 A) Comparative map of BTA23 and HSA6 adapted from Band *et al*, 2000). **B)** The bovine MHC on BTA23

inversion event. The majority of the genes are found on band 23q21 whilst the region now thought of as class IIb lies within bands 23q12-13 (Figure 1.4 B) Studies in sheep and goat and deer have also shown a similar arrangement of the MHC as observed in bovine.

Another unique feature of the BoLA is that different haplotypes vary in both number and expression of the MHC I genes (Lewin *et al.*, 1996) . The DQ locus of the class II region also shares this feature (Marello *et al.*, 1995). The consequences of this are not yet understood, although it has been suggested that this feature could lead to variation in individual disease resistance. BoLA also contains novel class II genes DYA and DIB, but their role in disease resistance is not fully understood.

1.3.3 The structure of the MHC molecules

Human HLA-A2 (class I) structure was the first MHC protein to be identified using X-ray crystallography (Bjorkmann *et al.*, 1987). The actual MHC I gene encodes for a heavy chain peptide of approximately 360 residues in size. When correctly folded the molecule can be divided into 3 extracellular domains, a transmembrane domain and a smaller cytoplasmic domain. The extracellular domains are noncovalently attached to a lighter chain of 99 residues known as the beta2m (Fig.1.5). This lighter chain is encoded at another location in the genome and appears to interact with all of the MHC I antigen-binding proteins. The actual binding site of the MHC I molecule is found in the extracellular domains 1 and 2 and is comprised of two parallel alpha helical structures with 8 anti-parallel beta sheets between, giving rise to a small cleft. The MHC I proteins bind peptides of around 8-10 residues in size. However longer peptides are sometimes bound by bulging the middle region and extending it out of the groove (Collins *et al.*, 1994, Joyce *et al.*, 1994)

The first structure of an MHC I bound to a peptide was determined using a murine class I molecule and a peptide synthesized from the H-2Kb virus (Fremont *et al.*, 1992). It was shown from this structure that an interaction between the backbone of the peptide and the MHC binding site (via hydrogen bonding) is responsible for holding the peptide in place. The side chains of the peptide appear to contribute to only a small number of contacts, however they are highly significant for an interaction. It has been hypothesized that binding via the peptide backbone enables an MHC protein to recognise more sequences rather than being specific for the more varied side chains. As an individual has only a maximum of 6 alleles for MHC I genes it would be expected that the binding sites of each would need to be able to bind more than one form of peptide (Young *et al.*, 1995).

The class II molecules are similar in structure to the class I. Both integrate into the cell membrane with extracellular domains that bind peptide. However, the class II molecules have only two extracellular domains and do not associate with the beta2m chain. Instead the MHC II molecules form heterodimers between an alpha and a beta subunit. The binding site situated between the two subunits, though constructed differently, appears similar to the MHC I binding site. A slight change in the helical structures of the MHC II molecules has left the binding site open-ended allowing for binding of larger peptides as long as 15 residues (Stern *et al.*, 1994).

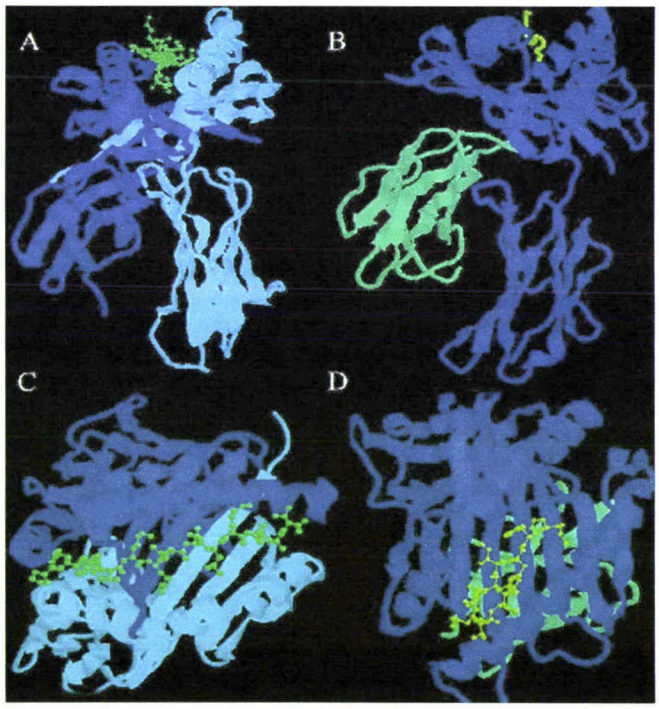


Figure 1.5 Structure of the MHC I and II binding molecules A. MHC II with peptide (green) **B.** Binding cleft MHC II **C.** MHC I and beta2m with peptide **D.** Binding cleft MHC I. http://webmed.unipv.it/immunology/mhcstr.html

1.3.4 T cell receptor binding

Once the peptide and the MHC are bound and presented on the cell surface the T-cell receptor (TCR) is able to bind. The different classes of MHC bind to different T cells determined by a co-receptor on the T cell, $CD8^+$T cells or $CD4^+$T cells. $CD8^+$T cells will recognise MHC I molecules and $CD4^+$T cells will recognise MHC II. Although there is still a possible interaction with the TCR and the peptide/MHC molecule complex this is lowered in the absence of the CD co-receptor (Lustgarten *et al.*, 1991).

T cells are produced in the thymus where they undergo strict selection. Only those that will not respond to self peptides presented by the MHC will be allowed to leave the thymus. This is important and is the reason why most organ transplants were unsuccessful. As every cell produces a unique array of peptides, the TCRs will recognise a cell that is not self and stimulate an immune response against it. Early experiments using transplants in mice identified the MHC and its properties for the first time (Little, 1916)

The actual binding between the TCR and the MHC takes place at the peptide binding site (Garcia and Teyton, 1998) with the TCR making contacts with conserved residues on the alpha helical structures of the MHC. Contacts between the TCR and the peptide also exist however these are out numbered approximately 3:1 in favour of the TCR to MHC contacts. This has lead to the conclusion that the MHC, rather than the peptide, is the dominant element in TCR recognition. Because the majority of bonds are with the conserved residues on the MHC molecule there is no need for the TCR to be polymorphic in order to bind a wide range of peptide/MHC complexes. Only the MHC antigen binding proteins require this feature.

1.3.5 MHC association Studies

Many earlier experiments involving the MHC region were carried out using the MHC alleles themselves as genetic markers. This entailed using serological methods, rather than DNA tests, to genotype individuals for given MHC haplotypes. Although these studies were usually carried out on only a small number of individuals they appeared to show a significant link between the MHC and a number of relevant traits such as fertility (Batra *et al.*, 1989, Stear *et al.*, 1989) and disease occurrence (Kelm *et al.*, 1997, Lunden *et al.*, 1990). However, with the advent of DNA technology many of these results were unable to be reproduced in large samples. Comments on these studies point to the absence of strict controls and inappropriate sampling (Arriens *et al.*, 1996). Despite these reservations the MHC region has been clearly linked to disease occurrence in respect to the incidence of clinical symptoms of the bovine leukemia virus (Xu *et al.*, 1993) and mastitis (Sharif *et al.*, 1998). Resistance to bovine leukemia virus (BLV), an economically important disease in the dairy industry, has been linked to DRB3 alleles (11* 23* and 28*) after an association with a class I (BoLA A14) allele was discovered (Bernco and Lewin *et al.*, 1989). Mastitis resistance has also been linked to allele 23* (Sharif *et al.*, 1998) although the mechanisms of action have not yet been defined.

In other species the MHC genes are also linked to disease and fer ility. In humans the link to fertility is well studied, however results vary in different studies. Early work in the field showed a significant link between certain haplotypes and spontaneous abortion rates in couples (Christiansen, 1996). However when similar studies were undertaken using larger populations the results were not corroborated (Ober *et al.*, 1999). There is speculation that parents of similar haplotype may be more likely to be infertile due to the inability of the mother to recognise the fertilised egg as non-self. HLA-G, a class I molecule, is also a target for further research. Its role is only just beginning to be understood in humans but its expression in the embryo suggests that a role in communication between foetus and maternal cells may exist (Choudhury and Knapp, 2001).

Other genetic diseases that have been linked to the human MHC are listed at OMIM (http://www.ncbi.nlm.nih.gov/omim/). However, these data should be treated with caution. An association between the MHC genes and a given phenotype does not necessarily mean a direct link and more often than not it has been shown that disease linked to the MHC genes are in fact controlled by other genetic factors within the region. Haemochromatosis, an excessive uptake of iron into the body for storage, is an example of disease thought to be linked to a specific MHC class I haplotype. Further research proved that this phenotype was in fact related to another closely linked gene (HFE) which has been shown to encode a protein that in its wild type form is thought to inhibit a specific receptor which is involved in initiating cellular iron uptake (Lyon and Frank, 2001). Beryllium susceptibility has however been linked to an MHC haplotype, which appears to instigate an excessive immune response in individuals that carry the haplotype.

In sheep, a link between footrot and MHC haplotypes has been postulated (Escayg *et al.*, 1997). Studies using daughter design experiments have been set up and have identified haplotypes that may show resistance to clinical footrot. However these data are only at the preliminary stage and more experiments are needed to define a clear link.

Aim of this study

A likely QTL site for longevity has been identified on BTA23 in several studies involving Holstein Friesian populations. As this chromosome is the site of the bovine MHC it is highly likely that this variation could be caused by the alleles of this extremely polymorphic region. Although there is good evidence for the QTL, the exact position is not yet clear. To identify this region further mapping within the region was needed. This was achieved by analysis of additional markers and genotyping more individuals.

Longevity has the potential to be affected by nearly every gene in the bovine genome. Therefore to identify candidate genes, the true cause of the variation in the longevity phenotype needed to be defined. To do this the fate data of the granddaughters contributing to the longevity BVs were analysed.

The polymorphic nature of the bovine MHC made it a starting point for candidate gene studies particularly genes coding for the MHC II DRB3 and the MHC I antigen binding genes. Their polymorphic characteristics and their role in the presentation of peptides to the immune system meant that these genes could potentially cause variation in the longevity trait and warranted further investigation.

Chapter 2: Materials and Methods

2.1 Materials

IPTG, X-gal, dNTPs,, ampicillin, lysozyme, BSA, mineral oil, unlabeled primers and 6-FAM labeled primers were purchased from Sigma Chemical Company, St Louis, MO, USA.

Oligonucleotides labelled with NED and VIC were purchased from Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

Bacteriological agar, 1kb plus ladder and Luria Bertani (LB) broth base were purchased from GIBCOBRL, Invitrogen Corporation, Invitrogen NZ limited, Penrose, Auckland, New Zealand.

Restriction endonucleases and buffers, *Taq* Polymerase, DNA quantification standards and agarose LE powder for electrophoresis were from a number of different sources: New England Biolabs, MA, USA; Boehringer Mannheim, Germany and Roche, Mt. Wellington, Auckland.

PCR purification kit QIAquick was purchased from QIAGEN, New Zealand distributors: Biolab Scientific Ltd, Albany, Auckland, New Zealand

NuSieve GTG was obtained from FMC, Bio Products, 191 Thoaston St, Rockland, ME, USA

pGEM®-T vector system was purchased from Promega Corporation, WI, USA.

The plasmid Mini Prep Kit was purchased from Bio-Rad Laboratories, 2000 Alfred Drive, Hercules, California, USA.

The *Escherichia coli* XL-1 blue strain was purchased from Stratagene, La Jolla, CA, USA.

Disposable pipette tips PCR tubes and plates were purchased from, SSI, 1310 Thurman St, Lodi, CA, USA.

2.2 **Methods**

2.2.1 DNA Collection

DNA samples were obtained from the grandsires and all possible sons and grandsons of two large Holstein Friesian families. Animals that had left the breeding programme or had too few daughters to warrant further study were not included in the analysis.

All DNA samples were isolated from sire semen collected by Livestock Improvement Corporation (LIC) at Vialactia in Auckland. DNA samples were then transported on dry ice to Massey University, Palmerston North. Phenotype data concerning fate was collected through LIC databases for 800,000 granddaughters in commercial herd members. Breeding values used were all calculated by LIC.

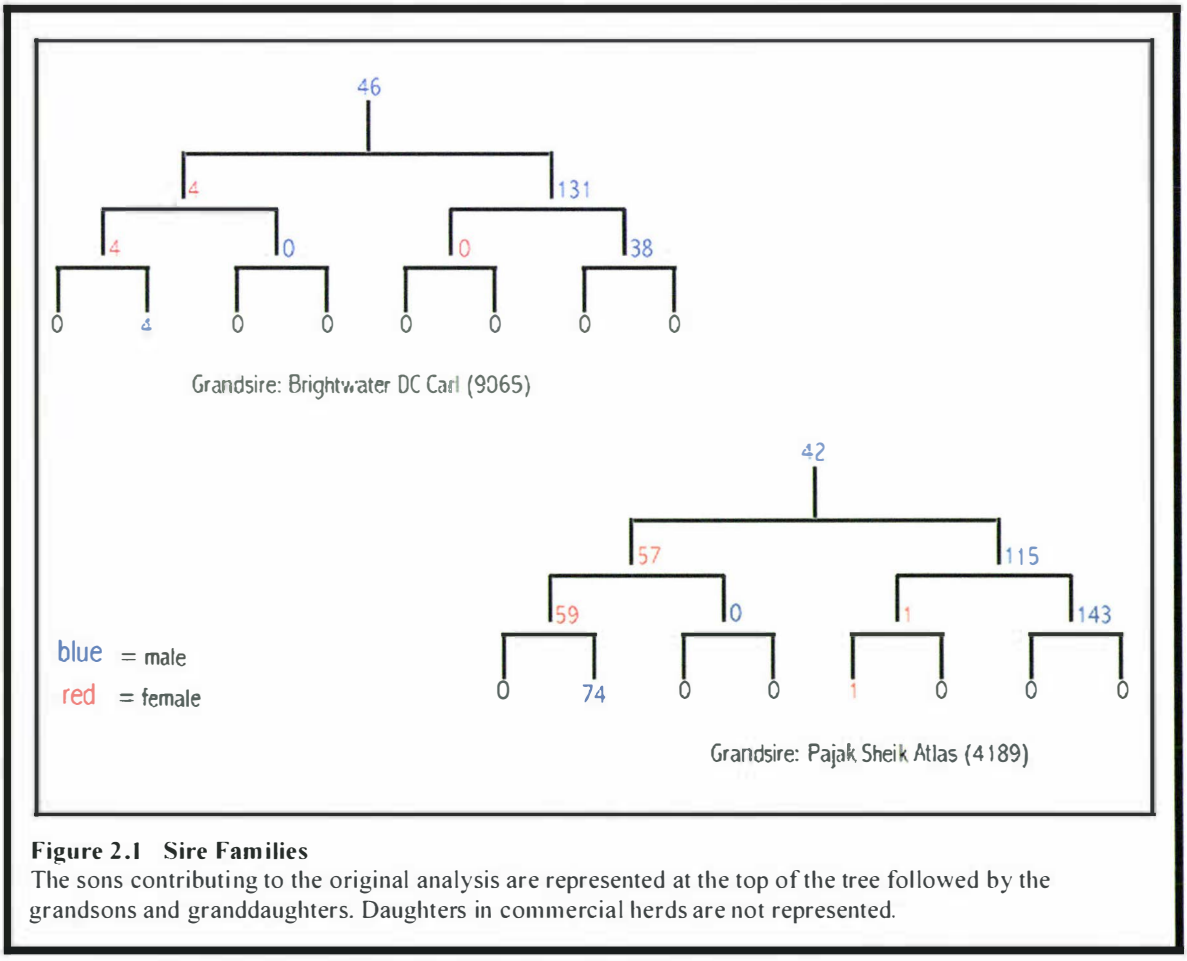


Figure 2.1 Sire Families
The sons contributing to the original analysis are represented at the top of the tree followed by the grandsons and granddaughters. Daughters in commercial herds are not represented.

2.2.2 Phenotype Data

Data were collected from daughters in commercial herds and breeding values calculated by LIC. Breeding values for AI (artificial insemination), DFM (days from first mating), survival from first to second lactation and survival from first to fifth lactation were used for phenotype data for the older sons. For the younger grandsons BVs for longevity and fertility were used as phenotype data.

Survival from first to second lactation is measured as the ability of the animal to simply survive in the herd until the next herd test. Survival to fifth lactation is the ability of the animal to last to the fifth lactation. Longevity is a newer measurement which represents the length of productive life, i.e. the time from first calving until removal. This BV is calculated from actual herd life data from animals already removed and expected herd life data or censored records. Censored records are calculated from the actual number of lactations the animal has already lived through, the production of the animal and several TOP traits (traits other than production) such as conformation and temperament.

Fertility BVs are calculated by looking at two daughter records: the ability of the animals to conceive within the first 3 weeks of mating and the ability of the animals to conceive from AI. These records are combined to give the likelihood of the animals calving in the early calving period. These data can be split into two BVs based on just the AI data or the DFM.

2.2.3 Polymerase Chain Reaction (PCR)

Amplification from the bovine genomic DNA samples by PCR was carried out in an Applied Biosystems 2700 thermal cycler. For microsatellite amplification all reactions were made to a total volume of 25 μL as follows: 9 μL milliQ water, 2.5 μL buffer (200 mM Tris-HCl (pH 8.4), 500 mM KCl, 50 mM $MgCl_2$) , 2.5 μL dNTP (3 mM), 2.5 μL forward primer (50 ng/μL) , 2.5 μL reverse primer (50 ng/μL), 0.20 μL *Taq* polymerase (5 units/μL). Template was diluted 1:50 and 6 μL used for each reaction. PCR of products intended for cloning or sequencing where amplified in 50 μL reaction volumes. Cocktails were made for all reactions to limit the effect of possible pipetting error. With reactions involving more than 40 samples 96 well plates were used. With fewer than 40 samples 200 μL tubes were used. All reactions involving plates were set up in a laminar flow hood to prevent contamination. No amplified PCR products were entered into the hood at any stage. Once the PCR reactions were set up they were placed immediately into the thermal cycler with the following cycle:

Denaturation	94°C 5mins	x1 cycle	
Denaturation	94°C 30sec	x1 cycle	x40 cycles
Annealing	x*°C 30sec	x1 cycle	
Extension	72 °C 30sec	x1 cycle	

Holding temperature 4 °C
* Temperature varied according to primer set

2.2.4 Troubleshooting for PCR

All primer sets were tested with the above conditions with recommended annealing temperatures (Table 2.1). When initial reactions were unsuccessful annealing temperature was altered to 45°C. Often this allowed for non-specific binding and the temperature was subsequently raised to allow specific binding. If PCR was unsuccessful at this temperature $MgCl_2$ concentrations were altered ranging from 0.5 mM to 2.0 mM. Due to the low concentration of template that was being used primer dimer formation was observed in most reactions. With unsuccessful PCR reactions resulting only in primer dimer formation, the PCR was repeated with increased template.

2.2.5 Electrophoresis of PCR products

A 5 µL aliquot of each reaction was analysed by electrophoresis through an agarose or NuSieve gel. Products between 200 and 130 base pairs in size where analysed using 1.5% agarose whilst products between 130 and 50 base pairs where analyzed using a 3.5% NuSieve gel. 1 kb plus™ ladder was used for estimating product size. Electrophoresis was carried out at 100 volts for approximately 45 mins in a 1xTAE buffer diluted from 50X stock(242 g Tris base, 57.1 mL glacial acetic acid, 18.6 g EDTA, 1 L distilled H_2O) . Staining of the DNA was achieved by the addition of 3 µL ethidium bromide to the gel at 55°C before the gel was allowed to set. Gels were then analysed using a UV transiluminator.

2.2.6 Quantification Gels

To quantify DNA for sequencing, quantification standards where electrophoresed alongside the PCR products in a 1% agarose gel and stained with ethidium bromide as above. Standards used were 10 ng/5µL, 20 ng/5µL, 50 ng/µL and 100 ng/µL.

2.2.7 PCR Purification

All PCR products for sequencing were purified using Qiagen purification columns according to the manufactures instructions. Briefly, a silica gel membrane is used to bind DNA to the column in the presence of high salt buffer under centrifugation. Removal of polymerases, salts and DNA fragments smaller than 100 bps is achieved in a series of wash steps. DNA is then released from the column in the presence of a low salt buffer or water.

2.2.8 Microsatellite Analysis

Microsatellites were separated using an ABI Prism 377-36/64 and analysed using GeneScan^{TM} software. To allow detection of the products, primers were synthesised with fluorescent tags at the 5' ends of the forward primers for each primer set. Because these tags have different signal strengths it was necessary to dilute the PCR products before analysis. This allowed the signal strength to be within the optimum detection bracket.

2.2.9 Primer Sets used for Microsatellite Analysis

Primers used for microsatellite amplification are detailed in table 2.2.1. All primer sequences were identified using the MARC web site (Knipling, 2002) and synthesised by either Sigma or ABI.

Locus	Primer Sequences	Product size (bp)	Annealing Temp	Label	Reference
BOLA-DRB1 (MB025(510))	5'-ATGGTGCAGCAGCAAGGTGAGCA-3' 5'-GGGACTCAGTCTCTCTATCTCTTTG-3'	121-137	56°C	6-FAM	Creighton et al, 1992 Acc. M30010
TAMLS113.3 (L29386(909))	5'-TTACTGCTGAGCCACCGG-3' 5'-GATGGGGGTCACAAACTGAC-3'	148-172	56°C	6-FAM	Skow et al, 1994 Acc. L29386
RM002 (CA002(600))	5'- AGCAATGTCAAACATTGCTCC-3' 5'- TCCATCTTTACAACTTTTGAATC-3'	112-102	58°C	6-FAM	Kossarek et al, 1993 Acc. U32908
BM7233 (BM7233(302))	5'- GGAATGAAAGAGCCTAGCAGC-3' 5'- AGGACTACTGTATGGATGTGCG-3'	104-126	56°C	NED	Stone et al, 1995 Acc. G18795
BM1815 (BM1815(10))	5'- AGAGGATGATGGCCTCCTG-3' 5'- CAAGGAGACAAGTCAAGTTCCC-3'	140-170	60°C	NED	Kappes et al, 1997 Acc. G18389
BM1258 (BM1258(6))	5'- GTATGTATTTTTCCCACCCTGC-3' 5'- GAGTCAGACATGACTGAGCCTG-3'	100-128	58°C	VIC	Bishop et al, 1994 Acc. G18385

Table 2. 1 Microsatellite Primers.

2.2.10 MHC primers

Primers for amplification of the MHC DRB3 and MHCI antigen presenting genes are listed in Table 2.2

Loci	Primer	Primer Sequences	Product size (bp)	Annealing Temp	Reference
DRB3(exon 2)	DRB3F	5'-ATCCTCTCTCTGCAGCACATTTCC-3'	284	65°C	Lewin, 1996
	DRB3R	5'-TCGCCGCTGCACAGTGAAACTCTC-3'			
BoLA I	MHC1F	5'-GCTCCCACTCBMTGAGGTATT-3'	700	58°C	Sawhney et al., 2001
	MHC1R	5'-TCCAGGTATCTGCGGAGC-3'			

Table 2. 2 Unlabeled Primers used for MHC Gene Amplification

2.2.11 Cloning

PCR products were cloned into the vector pGEM-T (Promega) (Appendix A) using the recommended protocol. Half of the ligation reaction was used (5 µL) to transfect 100 µL XL-1 Blue cells. The transformation was left on ice for 5 mins before 50 µL was used to inoculate LB/agar plates at 37°C containing 10 µg/mL Amp, 0.5 mM IPTG and 80 µg/mL X-gal. Plates were left at 37°C overnight to allow single colony growth. The presence of the beta galactosidase gene within the cloning site of the pGEM-T vector allowed for blue/white slection of successfully transformed clones. These coionies were then subsequently amplified in 5 mL LB broth overnight in the presence of ampicillin (10 µg/mL). Rapid boil plasmid isolation and restriction endonuclease digests were used to confirm the identity of clones before plasmid purification using the Biorad mini prep kit.

PCR of desired clone using plasmid specific primers was carried out to increase concentration of template for sequencing. These PCR products were then purified again through a Qiagen PCR clean up column.

2.2.12 Plasmid Preparation

The method presented in Holmes and Quigley, 1981 was used for all rapid boil analysis. This method is based on lysis of the bacterial cell wall with lysozyme and denaturation of the genomic DNA and cellular proteins by boiling. Plasmid DNA remains intact and soluble by virtue of its small size and can be precipitated from solution with the use of isopropanol.

2.2.13 Restriction Endonuclease Digest

All PCR product digests were performed using 2.5 µL of PCR product in the presence of 0.6 µL buffer (specified for each enzyme), recommended units of restriction enzyme and made to a total volume of 6 µL with distilled water. Digests where left at the specified temperature for each enzyme for 1 hour before analysis by NuSieve gel electrophoresis.

pGEM T plasmid digests were set up using 10 µL of template produced by rapid boil plasmid preparation, 1.5 µ L of buffer H (Roche Diagnostics), 0.1 µL of *Not I* enzyme (10 units/ µL), 0.1 µL of *Nco I* enzyme (10 units/ µL) and 3.2 µL of distilled water. After 1 hr at 37 ° C 10 µL of digest was analysed and quantified on a 1% agarose gel.

Chapter 3: Microsatellites on BTA23

3.1 Introduction

To identify a QTL, a link between the inheritance of known polymorphic marker alleles and phenotype variation can be investigated. This works on the basis that if the marker is linked to a polymorphic QTL site there will be a significant difference in the phenotypes of the offspring, dependent on the allele inherited.

In a previous study a QTL for survival was found to be segregating in 2 of 7 large grandsire families on BTA23 (Spelman, unpublished). 8 microsatellite markers were originally used covering a 65 cM region representing most of the chromosome. However, to follow the inheritance of the chromosome through the next generation, the markers must be heterozygous in the sire. In the first family (grandsire 4189) only 3 of the 8 markers and in family 2 (grandsire 9065) only 7 markers were heterozygous. The distance between the markers spanned up to 18 cM. Although both families showed strong evidence for the QTL site near marker BM1258 (Figure 3.1), because of the low density of the map it was impossible to identify a specific region.

For a candidate gene search the genome region must be reduced and more individuals tested to strengthen the evidence of the QTL. To do this the grandsons with reliable BVs were added to the analysis using new markers within the suggested region. These markers could also be used on the sons used in the original analysis to increase the information density in this map.

In the original study, BVs for survival from first to second lactation, survival from first to fifth lactation, AIC (artificial insemination success) and DFM (days from first mating), were used as phenotype data. These values were again used when examining the new markers in the original sons. Longevity and fertility BVs were used for the analysis of the grandsons (see section 2.2.3).

3.2 Identifying microsatellites

The most common markers used in QTL studies are microsatellites, regions within the DNA that contain small di or tri-nucleotide repeats. As there is no known role for these repeats no selection against variations of the number of repeats is seen. This has meant that it is possible for a microsatellite region to have many different alleles, unlike other markers which only have a couple of different alleles. The variety of allele sizes that occurs in microsatellites means they are more likely to be informative when used in a QTL study.

A list of most published microsatellites in the bovine genome can be found at the MARC site (http://www.marc.usda.gov/genome/genome.html). This site lists a total of 25 microsatellites for BTA23, including the original 8 used in the initial study. Information on the estimated heterozygosity, the alleles and recommended primers are also available at this site. Using this site 5 new markers were chosen that could be analysed in the grandsires, sons and grandsons. Markers were chosen based on the chance of heterozygosity and their position along the genetic map with additional markers selected between marker BM1258 and the BoLA region (Figure 3.1).

This area was the focus for two main reasons; the original study suggested the QTL to be most likely near marker BM1258, and the presence of the highly polymorphic BoLA region which would be likely to contain candidate genes. Two markers were chosen that lay slightly outside the region to reduce the chance of narrowing the region down so much that a QTL lying slightly outside this region would go undetected.

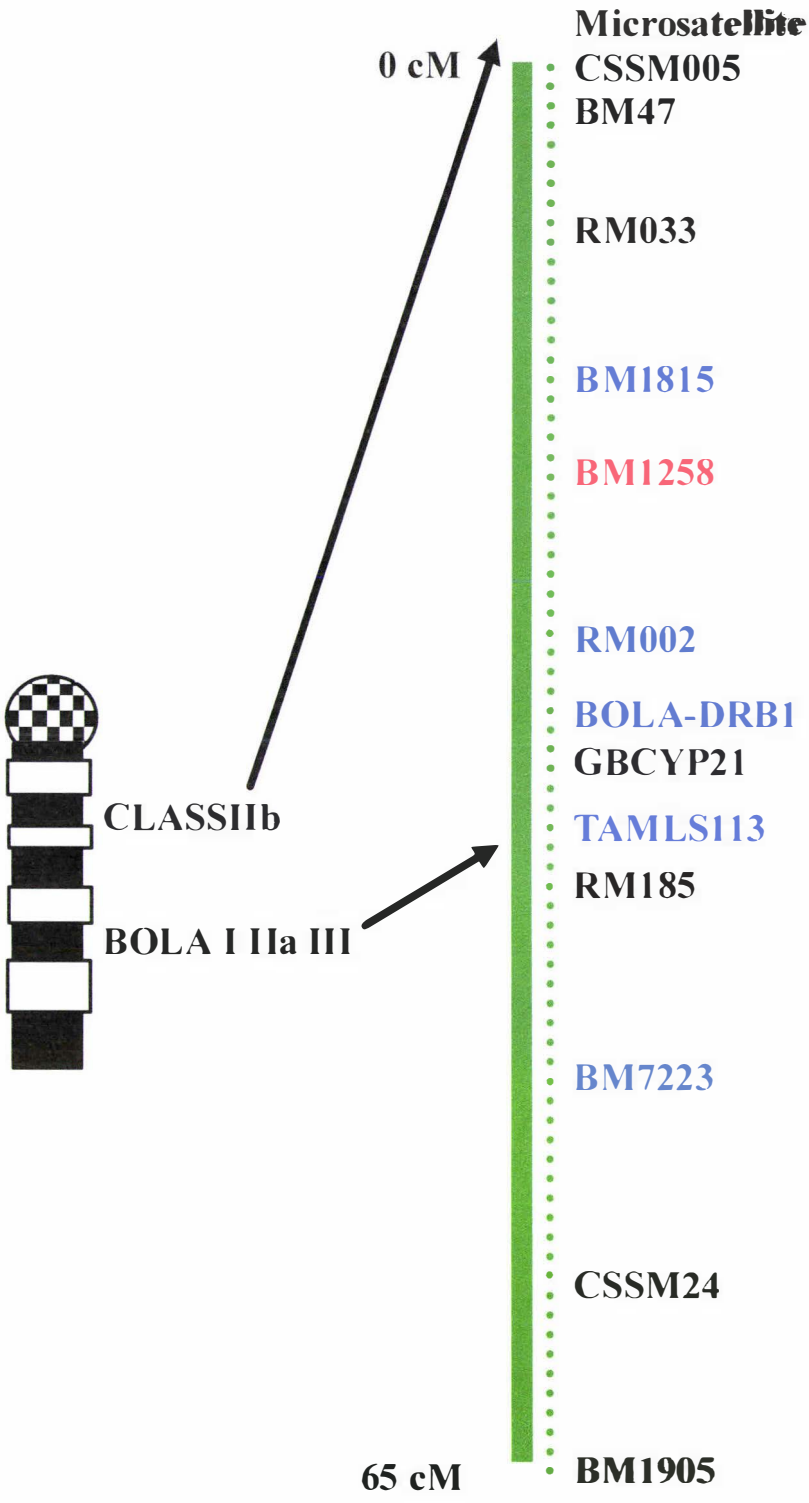


Figure 3.1 BTA23 microsatellites used in Grandsire families. Black denotes microsatellites used in original study whilst blue represents microsatellites used in this study. BM1258 was used in both studies. FISH probes of BoLA genes show approximate position of linkage map in comparison to cytogenetic map. (Adapted from www.marc.usda.gov, 2002)

3.3 Screening the Grandsires

All 5 markers (Figure 3.1) were initially tested in the grandsires to confirm that they were heterozygous. This was carried out to ensure that they would be useful to track the inheritance of BTA23 through the subsequent generations.

All microsatellites were amplified with labelled primers and sent for analysis by the Massey University DNA analysis service. The fluorescent tags were used such that all markers could be multiplexed for each individual.

Marker CA002, with an estimated heterozygosity of 37%, was unfortunately homozygous in both grandsires (Figure 3.2) and marker BM7233 proved difficult to amplify making it impractical for use in large scale PCR. The other markers were all heterozygous in sire 9065, however, only marker MB025 was heterozygous in sire 4189 (Figure 3.2).

Due to the cost of the Gene Scan analysis it was decided that analysis of one marker in the 4189 sons was not economical and this family was placed to one side. All markers in sons and grandsons were analysed using the same primer sets and conditions.

In order to multiplex the markers 1/10 1/20 and 1/50 dilutions were trialled for each microsatellite PCR product. To get an accurate allele size the fluorescent signal must be between 150 and 4000 relative fluorescent units. Because each microsatellite needed a different dilution to achieve a signal within this range it was not possible to multiplex the PCR. Instead each microsatellite was amplified separately and the products pooled together before analysis on the sequencer (Fig 3.2). Variations to final dilutions were made if the PCR product appeared light on the agarose gel after amplification.

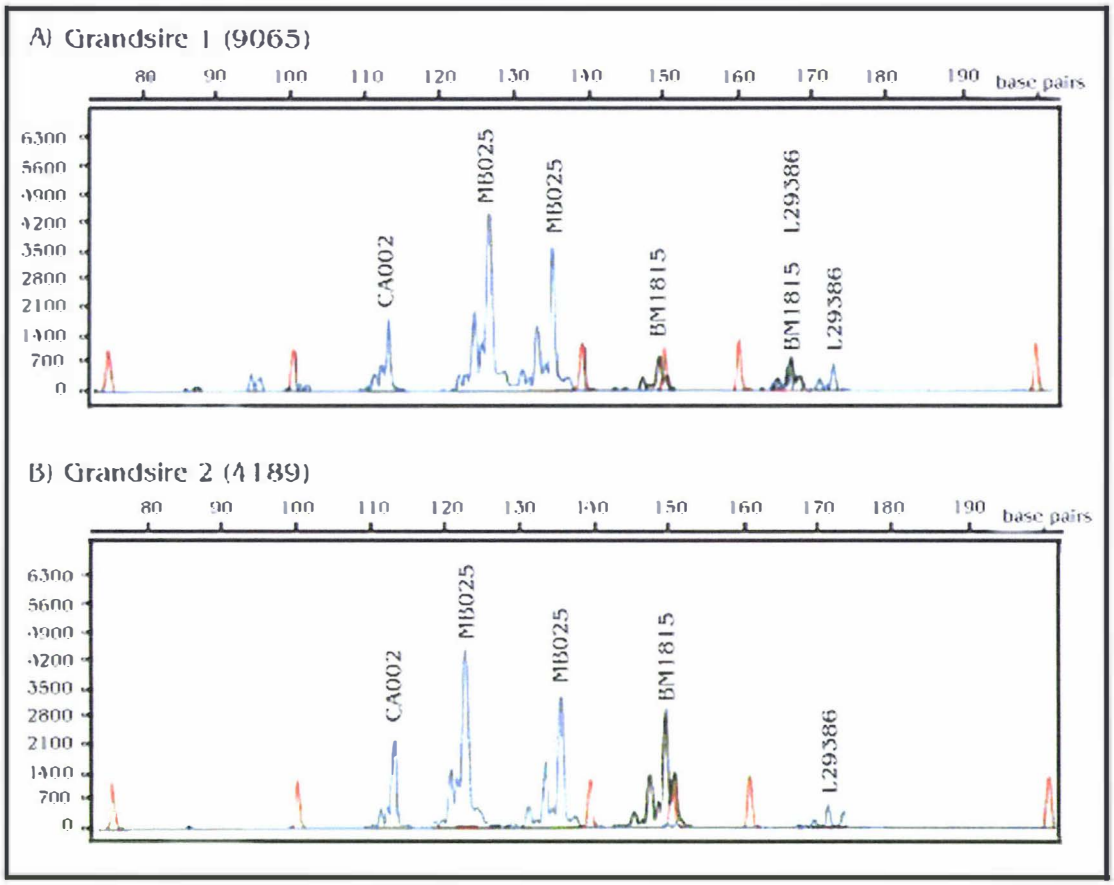


Figure 3.2 GeneScan run of new markers in grandsire 9065 and 4189

PCR products were diluted as follows with MilliQ: MB025 1/50, L29386 1/20, CA002 1/20, BM1815 1/5 All microsatellite analysis was carried out using the GeneScan programme. Although stutter bands were seen in all experiments the highest peak was taken as the actual microsatellite. All microsatellites were individually amplified and multiplexed afterwards to the desired concentration. All products amplified at different levels and showed variable emissions from each fluorescent tag. This meant some products were diluted as much as 50 times whilst others only needed a dilution of 10 times.

3.4 Microsatellite Analysis

The additional markers were genotyped in the sons of grandsire 9065. To identify the linkage phase of these new markers the inheritance of the new alleles was compared to the original markers. If a new marker allele appeared to be inherited most often with a particular allele of one of the nearby original marker alleles they were said to be in the same linkage phase thus they belonged to the same haplotype. This allowed the new marker alleles to be divided into the two haplotype groups. If the genotype of the son is the same as the sire for a particular marker it is impossible to identify which allele the son inherited from the sire and which from the dam and the marker becomes uninformative.

Haplotype 1 was then arbitrarily assigned as the carrier of QTL_1 (Q_1) and all marker data was converted to a probability of inheriting Q_1 (Appendix B.1). This meant all genotype 1 alleles took on the value of 1 and all genotype 2 alleles were valued at 0. The probability of inheriting Q_1 for uninformative markers was calculated according to the genotype of surrounding markers and the distance between them and the unknown marker. (Figure 3.3).

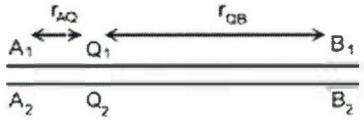


Figure 3.3 Calculating the probability of recombination: A and B represent known markers. Q represents possible QTL position. r represents the genetic distance.

To calculate the probability of these unknown markers being haplotype 1 the distances between the unknown and the known markers were first converted from cM to recombination units using the following formula:

$$r = 0.5(1\text{-}e^{-2m})$$

where m = distance in cM

This converts the map distances into a probability of a recombination event occurring between the two loci (Haldane, 1919). There are a total of eight different possible haplotypes for three markers (Table 3.1). If the loci were unlinked the probability of

inheriting Q_1 at an unknown marker position with A_1 would be 0.5, however, when linked this probability increases. The chances of Q_1 being inherited with A_1 is dependent on no recombination occurring between the two ($1-r_{AQ}$) whilst the chances of Q_2 being linked to A_1 is dependent on recombination (r_{AQ}). So the probability of inheriting A_1 with Q_1 is $1-r_{AQ}$ multiplied by the chances of getting Q_1 over Q_2 (0.5).

If the genotype of a marker on the other side of the unknown marker is known it can also be used. If the genotype of the flanking markers is A_1B_1 then the probability of obtaining $A_1Q_1B_1$ can be calculated as the probability of no recombination between A|Q and Q|B multiplied by the chances of obtaining Q_1 over Q_2 (0.5). This can be written as $0.5(1-r_{AQ}) \times (1-r_{QB})$. This formula alone gives the probability of getting $A_1Q_1B_1$ over the other 7 possible haplotypes (Table 3.1). However, if the flanking markers are known there are only two possible haplotypes ($A_1Q_1B_1$ or $A_1Q_2B_1$). By dividing by the total probability of getting A_1B_1 in the first place the probability of getting Q_1 over Q_2 can be calculated.

Example:

Let $r_{AQ} = 0.01$ and $r_{QB} = 0.09$
If the haplotype of the flanking markers = A_1B_1

The probability of Q_1 $= 0.5(1-r_{AQ}) \times (1-r_{QB})/\text{prob } A_1B_1$
$= 0.5\ (1-0.01) \times (1-0.09)/\text{prob } A_1B_1$
$= 0.45045/\text{prob } A_1B_1$

The probability of getting A_1B_1 is calculated by summing the probability of $A_1Q_1B_1$ and $A_1Q_2B_1$. This is not quite equal to $0.5(1-r_{AB})$ as the possibility of two recombination events occurring is taken into account rather the just one. It is possible to have three recombination events if the distance between the two markers is large, however, for the purpose of this study it was assumed that this did not occur.

Prob $A_1B_1 = 0.5 \times (1-r_{AQ}) \times (1-r_{QB}) + 0.5 \times r_{AQ} \times r_{QB}$
$= 0.5 \times (1-0.01) \times (1-0.09) + 0.5 \times 0.01 \times 0.09$
$= 0.4509$

Therefore Probability of $Q_1 = 0.45045/0.04509$
$= 0.999002$

Flanking marker Haplotype	Probability of genome region containing genotype 1 or 2 probQ$_1$	probQ$_2$
A_1B_1	$(0.5x(1-r_{aq})x(1-r_{qb}))/$ probA_1B_1	$(0.5x\ r_{aq} \times r_{qb})/$ prob A_1B_1
A_1B_2	$(0.5x(1-r_{aq})x\ r_{qb})/$ prob A_1B_2	$(0.5x\ r_{aq}\ x(1-r_{qb}))/$ prob A_1B_2
A_2B_1	$(0.5x\ r_{aq}\ x(1-r_{qb}))/$ prob A_2B_1	$(0.5x(1-r_{aq})x\ r_{qb})/$ prob A_2B_1
A_2B_2	$(0.5x\ r_{aq} \times r_{qb})/$ prob A_2B_2	$(0.5x(1-r_{aq})x(1-r_{qb}))/$ prob A_2B_2

Table 3.1. Calculation of probabilities. Q = region of unknown haplotype. r = recombination

By using these formulae the probability of each unknown marker being Q_1 was calculated (Appendix B1 and B2).

3.5 Single Marker Analysis

Within family analysis was carried out on the sons of the grandsire 9065 to study the relationship between genotype and phenotype using the regression principles of Knott *et al.*, 1994. The probability of getting genotype 1 can be plotted against the phenotype data for each marker (Figure 3.3A). A straight line can be fitted to this plot using the following model:

$$\mathbf{Y_{jk} = \mu + b_k X_{ik} + e_{jk}}$$

Where Y_{jk}= the BV of the j^{th} sire at the k^{th} chromosomal position, μ= the overall mean, b_k= the regression coefficient for the grandsire at k^{th} chromosomal position, X_{ik}= the probability of the i^{th} son receiving the chromosomal segment from the grandsire at the k^{th} position and e_{jk} = random residual.

This analysis was achieved using the TREND function in Excel which fits the line according to the least squares principle. Due to the fact that only one grandsire contributed to the analysis, the grandsire effect from other loci was assumed to be equal

for all sons. The dam effect is also absent as this is assumed equal between the two allele groups. Random residual is also assumed equal.

The better the real data fits this model line the more likely a link between the marker and a QTL. This was tested by calculating an F statistic for each marker position. An F statistic is calculated as a ratio of the sum of the squares from the regression / mean of the squares of the residual (SS_{reg}/MS_{res}). The higher this value is, the more likely a QTL is to be linked to the marker under investigation.

From this single marker regression analysis (Figure 3.4A) it was shown that the most likely site for the longevity QTL was between markers E and B spanning a region of 19 cM. This was due to the fact that these markers appeared to encompass the markers showing the highest F values. To test whether these values were indeed statistically significant threshold levels were identified (Section 3.7).

3.6 Interval mapping

To examine the region more closely interval mapping was undertaken in the 19 cM region. The probability of inheriting Q_1 was calculated for each cM in every son exactly the same as with the single marker analysis. These values were then regressed with the phenotype data (Figure 3.4B).

This analysis showed that according to these data the most likely QTL site was approximately 23 cM along the chromosome. However, the QTL site is likely to shift as more individuals are added to the investigation.

3.7 Threshold Levels

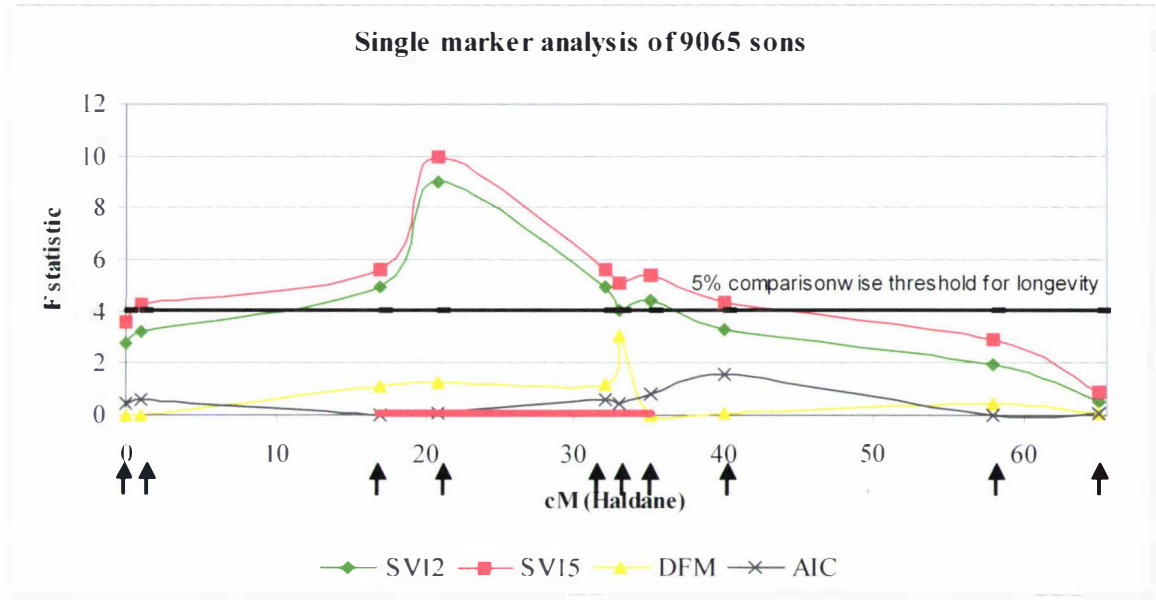
Probability tables calculated for F statistics are based on independent and regular distribution of the data. These conditions are not met in QTL analysis so it is necessary to calculate new critical values using the permutation method (Churchill and Doerge, 1994).

This is simply achieved by the shuffling of the phenotype data against the probabilities of the genotype data and calculating the F statistic a large number of times (1000 times for 5% threshold value).

To calculate the comparisonwise critical values, the F statistic is collected for each shuffle for all analysis points. These values are grouped and the F statistic value at the 95^{th} percentile is given as the threshold level at 5%. Normally this method can be used to calculate the threshold level at the 1% and 0.01% level, however in this experiment a maximum of 1200 shuffles was carried out for each point. This number of shuffles is not enough to confidently give values at these levels. Although a slightly different threshold value was calculated for each point, the differences were so small the average threshold level was used to plot the threshold line on the graph.

For more accurate values an experimentwise threshold can be calculated. This involves collecting the highest F statistic for each shuffle and taking the 95^{th} percentile in this new set. These values exceeded all real calculated F statistics. However the danger of a false negative in this experiment is a very real concern, and so at this stage the cut off level was left at the comparisonwise 5% threshold level.

A)



B)

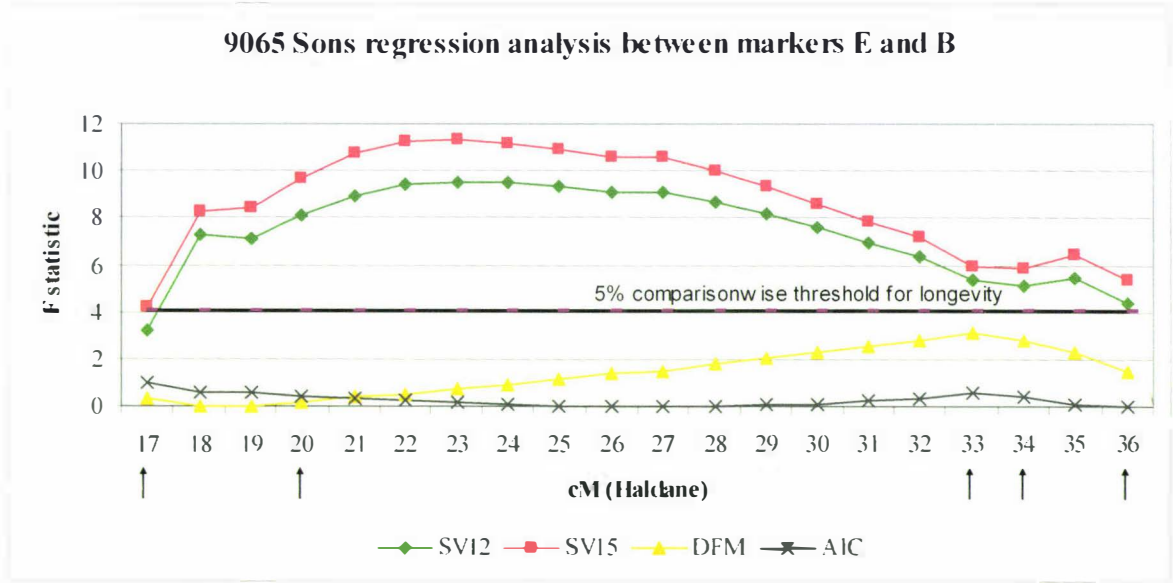


Figure 3.4. Regression Analysis in Sons of 9065.

A) Single marker analysis using original and new markers in 9065 sons. This method uses the haplotype information surrounding the actual markers only.

B) Interval mapping in region of interest. Shows most likely QTL site is still nearest marker BM1258.

Arrows at bottom of graph represent the position of marker used in analysis. Large red line shows region used .in interval mapping.

3.8 Grandson Analysis

Three sons, 7658249 (249), 7816140 (140) and 7932279 (279) born to sire 9065 were shown to have a number of sons of their own participating in the breeding programme. All grandsons sired through these sons were also genotyped for the new markers (Appendix B.2). Sire 279 was heterozygous for all 3 markers whilst sire 249 and 140 were homozygous for all but marker BoLA-DRB1. It was decided at this point that marker BM1258 would also be added to these new individuals. Unfortunately this was not useful in the sire 249 family as this sire was homozygous for this microsatellite as well and so only single marker analysis could be undertaken.

Data analysis was carried out in these sons using the same methods as with the grandsire sons, however, only the middle region of the chromosome was analysed (Appendix B.2).

Both fertility and longevity BVs were analysed in this study. However, as seen in Figure 3.4 no significant QTL could be seen for either family in regards to fertility.
A suggestive QTL for longevity does appear in the 279 family whilst 140 shows no evidence (Fig.3.5).

The phenotype data for family 249 was regressed on the genotype data for marker BoLA-DRB3 and showed a slight trend with a calculated F value of 3.98 for longevity. However without other markers for comparison of the significance of this value, the evidence is not strong.

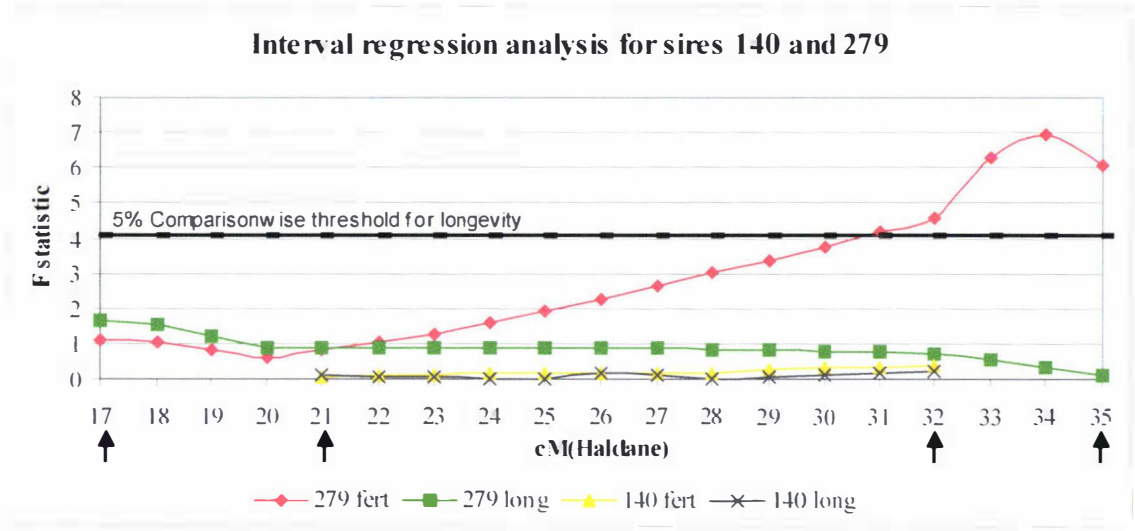


Figure 3.5 Analysis of suspected QTL site in families of sire 140 and 279.

3.9 Information content

To study the information content achieved by the genotype data, the variance of the probability coefficients was calculated at each cM. If the haplotype of every point on the chromosome is known, the variance would be 0.25, as the maximum value is 1 and the minimum 0. By calculating the variance for each point and calculating it as a fraction of 0.25 (maximum information) the true information content at any given point can be estimated.

This was carried out in the sons of the grandsire 9065 and the grandsons born of sons 279 and 140 (Figure 3.6). The information content towards the ends of the examined regions is quite low in family 279. This may account for the large jump in the F values towards the end of the region.

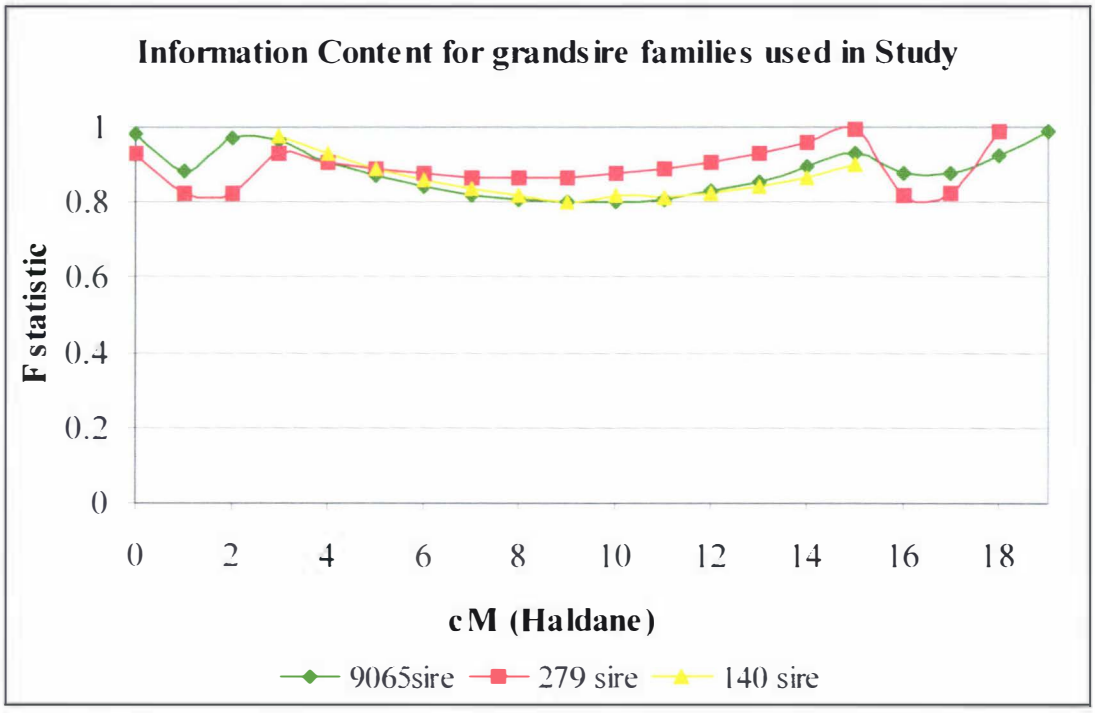


Figure 3.6 Information Content for 3 sire families used in study

3.10 Conclusions

From this study it appears likely that a QTL for longevity does exist in the middle of the BTA23 chromosome. The fact that only one grandson also showed evidence for this is unfortunate but not surprising due to the possibility that the QTL could have been homozygous in the other sires.

In the original study 2 of 7 sires appeared to be heterozygous. Assuming that it is the same gene in both families it is likely that the alleles are in relatively high frequencies within the population. If the allele that caused the variation was rare and exists predominantly in heterozygous individuals it is a large coincidence that the allele appeared twice in the 7 sires. It is quite possible that the allele causing the variation is more common and the other grandsires were homozygous for one or the other allele or there simply was not enough genomic or phenotypic data to show a significant QTL effect. This means there could be real benefits in the identification of the gene involved.

Although the threshold level in this experiment was quite low, at 5%, it is interesting that this was reached in two non-related families. The fact that the son 279 also shows segregation around this region is evidence that the longevity QTL may indeed be real and warrants further study.

Chapter 4: Fate Data Analysis

4.1. Introduction

Virtually every gene and regulatory element in the bovine genome has the potential to affect natural fitness which directly affects the longevity of an individual. Therefore if a useful candidate gene list is to be created, the exact cause of the longevity variation needs to be identified.

As the commercial environment rarely allows dairy cows to live to old age, longevity in a commercial herd is influenced primarily by an animal's ability to stay productive rather than its ability to defy aging. Therefore, the variation seen in these two families is most likely a product of variation in a production related trait. However, the fact that previous QTL studies within BTA23 in these two grandsire families for other production traits have shown no links between this region and trait variation suggests that there is a possibility the animals may be affected by other phenotypes, such as disease, not measured in the sire proving schemes.

In order to study the phenotype more closely, fate data, belonging to the granddaughters collected by LIC at herd testing, was analysed. The assumption was that the data would be accurate enough to allow a significant difference in the proportions affected by the relevant fate between each haplotype in the two families.

Because of the large number of animals and the fact that the data was given as discrete values it was decided that for the purpose of initial investigation only the genotype of the sires at marker BM1258 would be considered. As already shown this marker does

segregate significantly with longevity and therefore the effects of the fate should also be detectable.

4.2 Data set

Fate data for granddaughters of grandsires 4189 and 9065 were obtained from Livestock Improvement databases collected through herd tests from 1988 until the present. A total of 319,165 granddaughters representing grandsire 4189 and 416,853 representing grandsire 9065 were identified. The fate for each granddaughter was originally given as a 2 digit code, the first representing the fate of the animal (no fate, died, culled, sold) and the second the reason (production, mastitis, etc). Animals could be divided into 40 different fate groups on this basis. However, many of these categories were very similar and animals were subsequently summed into only 12 different categories as listed in Table 4.2.

Closer investigation of this data set showed the majority of this set was the offspring of only 6 sires and that most of these offspring were born after 1998. Further analysis of these younger granddaughters showed that their sires were very similar in the markers that they inherited from the grandsire therefore making it difficult to study the variation in these younger animals since one haplotype was substantially under represented. For this reason only granddaughters used in the original sire proving schemes were used to allow all sons to contribute almost equally. This also minimised the variation caused by yearly environmental changes (both economic and farm management changes). Unfortunately no information on the herd the granddaughters belonged to was available and so the herd effects in the study are assumed equal. A total of 3093 granddaughters were used from the grandsire 9065 family and 1929 granddaughters for the grandsire 4189 family. 32 sires contributed to the 9065 family and 21 to the 4189 family with a range of 35-224 daughters each (Figure 4.1).

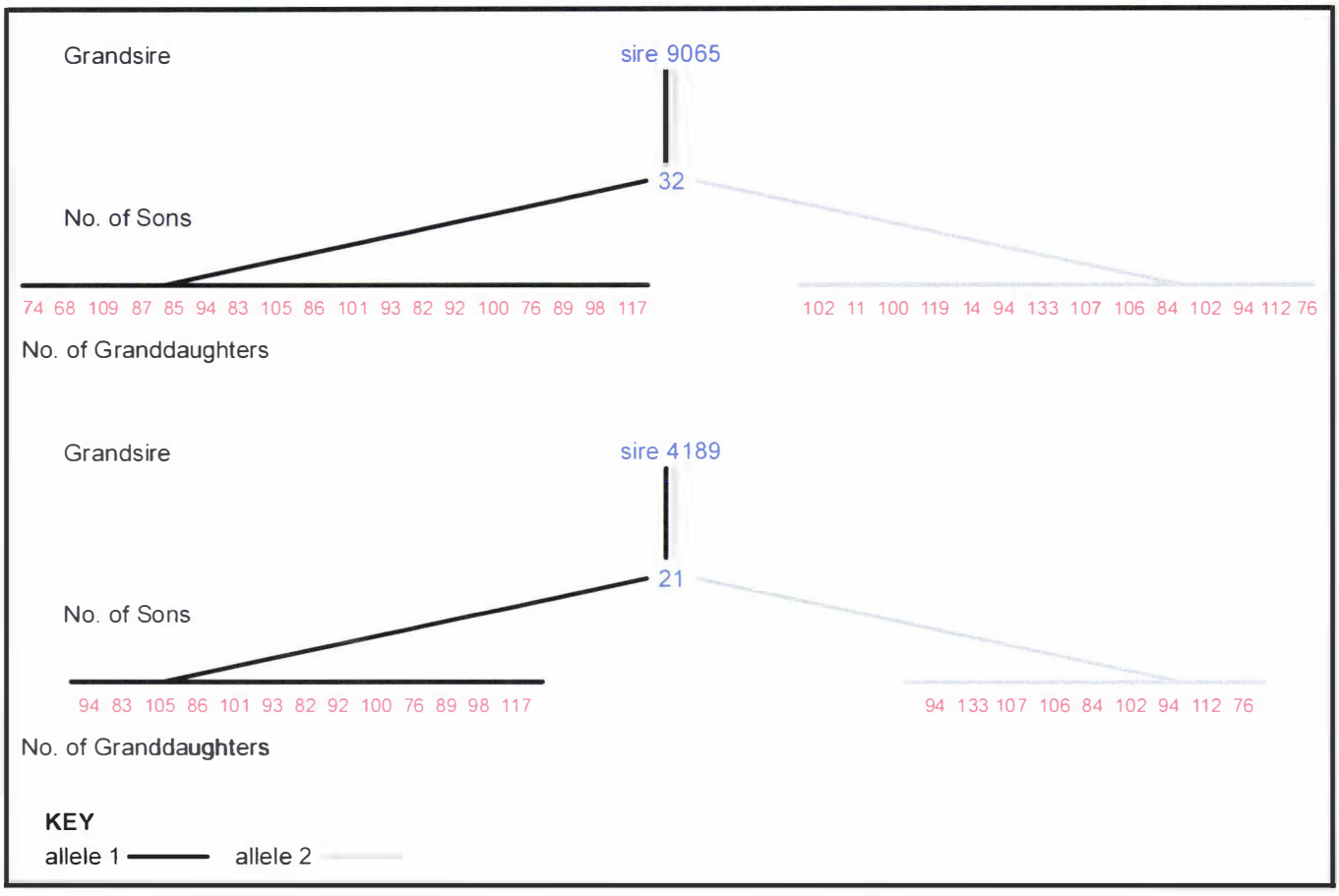


Figure 4.1 Origin of granddaughters used in analysis.
All granddaughters used were born in 1991 for family 9065 and 1988 for family 4189 to minimise year effect. Daughters belonging to each son summed in red. Only sons with known BM1258 allele were used in the analysis.

4.3 Age Affect on Fate

4.3.1 Age at death

To study the affect of the longevity QTL on BTA23, the granddaughter ages were separated into two groups according to the genotype of their sires at marker BM1258. Marker BM1258 was originally shown to be the closest marker to the QTL, therefore if a significant difference was to be detected it would most likely be identified at this marker.

To verify a significant effect at marker BM1258 the granddaughters were separated into two groups according to the BM1258 allele carried in their sires. The average ages of the two allele groups were calculated in both families (Table 4.1). In both cases the allele previously shown to be linked to variation in longevity also appeared to be linked to increased life span despite the absence of herd effect in the calculation.

The significance of the differences between the two alleles was investigated using a permutation analysis. The Resample Stats package (www.resample.com) allowed the ages to be shuffled against the allele of the sires and the new averages calculated 10,000 times. This programme allowed the calculation of the difference between the two groups and calculated the possibility of achieving the same difference from a random sample. Both families exceeded the 5% threshold with family 9065 exceeding 1%. Thus a link between age at herd removal and marker BM1258 was suggested.

4.3.2 Onset of variation

To see if there was a noticeable difference in the numbers of animals culled in the early years, the daughters were grouped according to the age at removal from the herd (Figure 4.2) and the percentage removed each year calculated. A large number of individuals (30-40%) in both families reached well over nine years old. This number appeared larger than was expected when compared to other studies, however, further analysis of the raw data showed that most animals contributing to this group had indeed participated in herd testing in 2001 and therefore must have been alive at that stage. If animals in this column had not recently contributed to a herd test the date of their last herd test was used to calculate a new age and they were given a fate code of unknown.

A large difference between sire allele and the percentages culled in the first year was observed in both families. Comparisons between the families cannot be made at this stage due to year and sire effects.

By randomly assigning the sire allele to the age data and recalculating the percentage removed in each age bracket 10,000 times the significance of the difference in ages was estimated. Both families showed that the difference was significant at the 0.01% threshold level. This was evidence that the trait variation was possibly affecting the removal of animals within the first years.

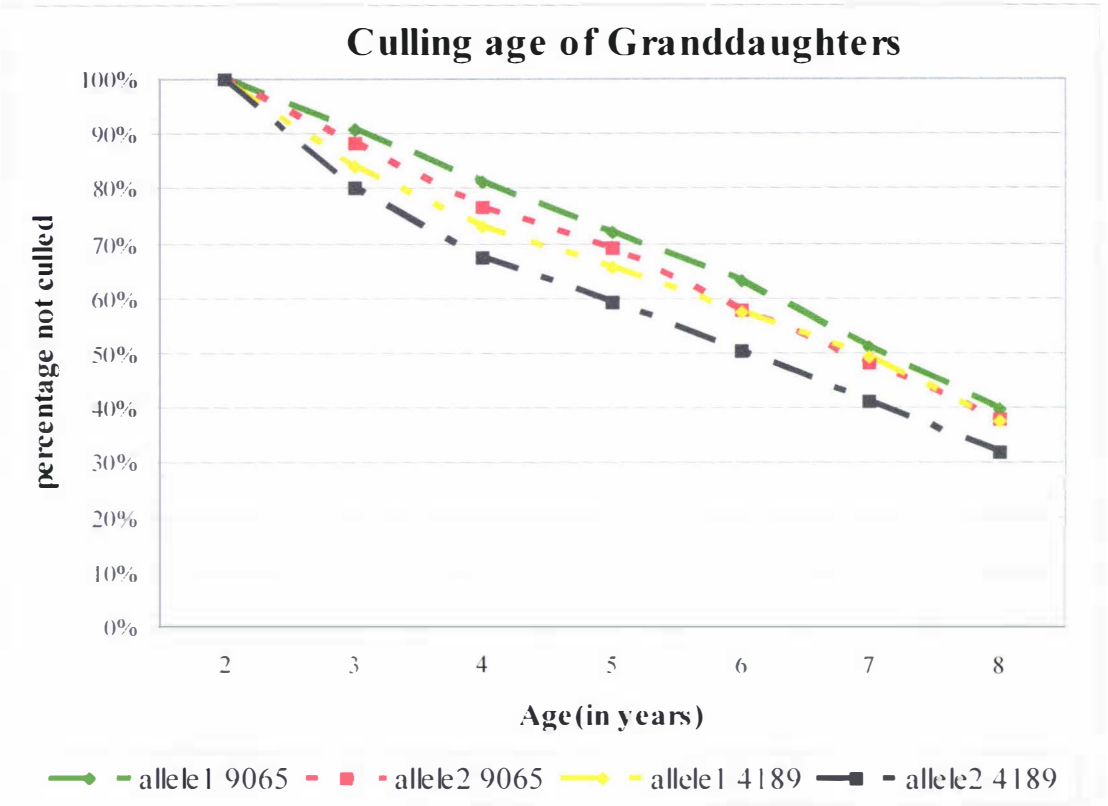


Figure 4.2.1 The age at removal for the granddaughters separated according to haplotype at marker BM1258 in the sire. 0-2 years not represented in analysis.

	Grandsire 4189		Grandsire 9065	
Allele at BM1258	1	2	1	2
Number of daughters	799	1130	1639	1454
Average age(yrs)	6.78	6.31	6.79	6.55
P Values	0.0489		0.0073	

Table 4.1 The number of granddaughter and the average age for the 4 haplotype groups.

4.3.1 Fate data analysis

To analyse the fate data, animals were divided into groups according to the haplotype of their sire at marker BM1258 and then divided again according to their fate code (Figure 4.3). Studies on the sires had already been undertaken to look for QTL for other production related traits with no success so it was unlikely that there would be any real difference in these trait categories. However, fate data is slightly different to the data used to calculate the BVs.

Only the daughters contributing to the same year were analysed together. This prevented the age of the animals varying the proportion of each fate. Many fates are age specific, for example an animal is more likely to be culled for fertility in the first lactation than for udder breakdown, whilst the latter may be more important to slightly older animals. Also, by only using daughters of the same year the living animals were also left in the analysis. If different ages were used the percentage of animals in this living group would be extremely sensitive to younger animals.

Once again the permutation method was used to identify significant differences in the two allele groups. Only removal for production related reasons exceeded the 5% threshold in this experiment. However these results were not surprising when the large number of individuals in the unknown category was considered.

Although the large numbers of animals in the unknown category were a concern there was not a significant difference in the number of animals contributing to this group between the two allele types. This suggested that if there was true variation in another group it may still be detectable assuming that an equal number from one genotype were contributing to the unknown group.

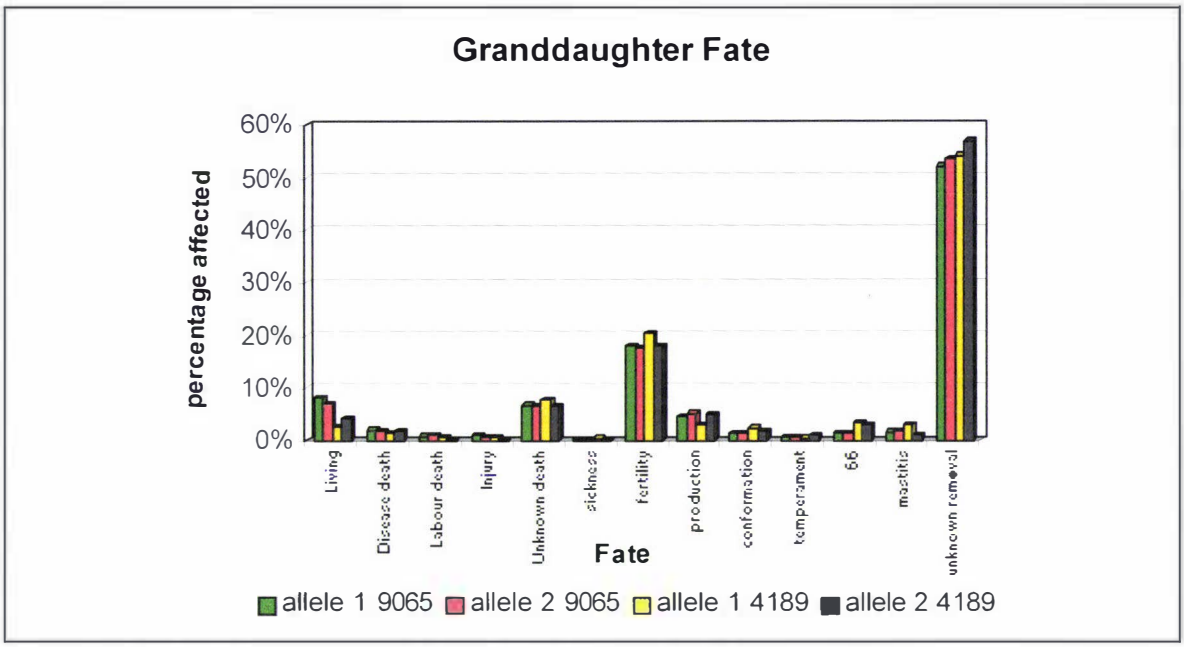


Figure 4.3 Percentage of granddaughters in each allele group removed for each fate. The granddaughters of each family were divided into two groups according to the BM1258 allele their sire inherited from the grandsire. These groups were then divided again according to the fate of the granddaughters. A comparison of the percentages affected by each fate could then be made between the two allele groups.

Fate group	Explanation of fate group
Living	Animal still present in herd
Death in birth	Death due to calving difficulties
Disease death	Death from disease
Injury	Fatal injury
Unknown death	Animal died on farm for unexplained reasons
Sickness*	Animal was removed for disease
Fertility*	Abortion or not mated successfully
Production*	Low production
Conformation*	Inferior structure including udder, legs, feet
Temperament*	Difficult handling
66	Sickness related removal
Mastitis*	Removed for mastitis related reasons
unknown removal	sold for diary purposes or culled for no known reason

*may be culled or sold

Table 4.2 Explanation of fate groups

4.4. Independent Fate analysis

To analyze the possibility that a particular fate may be affecting animals earlier the three leading causes for removal were assessed for onset times (Figure 4.4) All animals affected by fertility, production and death were summed and the percentage of animals affected each year calculated.

Most animals affected by fertility and production were affected when very young or very old. This was expected as these are the ages most affected by all culling. Death, however, affected the younger animals more than the old. 60% of all animals removed by death were younger than five years old. Unfortunately death as a fate code is not very informative as it can be the product of disease, or trauma. The tendency to affect younger animals may also be a data collection error. Unknown death may have been used instead of unknown removal in the early years of data collection. However there is no way of proving this possibility and the data must be assumed to be correct.

To study any variation surrounding BM1258 the average age of the animals affected by each fate was calculated for each allele. Unfortunately the data could not be divided to look at the allele effect each year due to the small size of the data set. Although there was a trend for all the averages for allele two to be lower than for allele 1 in both families only fertility in the family of grandsire 9065 showed a significant difference above the 5% threshold margin (Table 4.3).

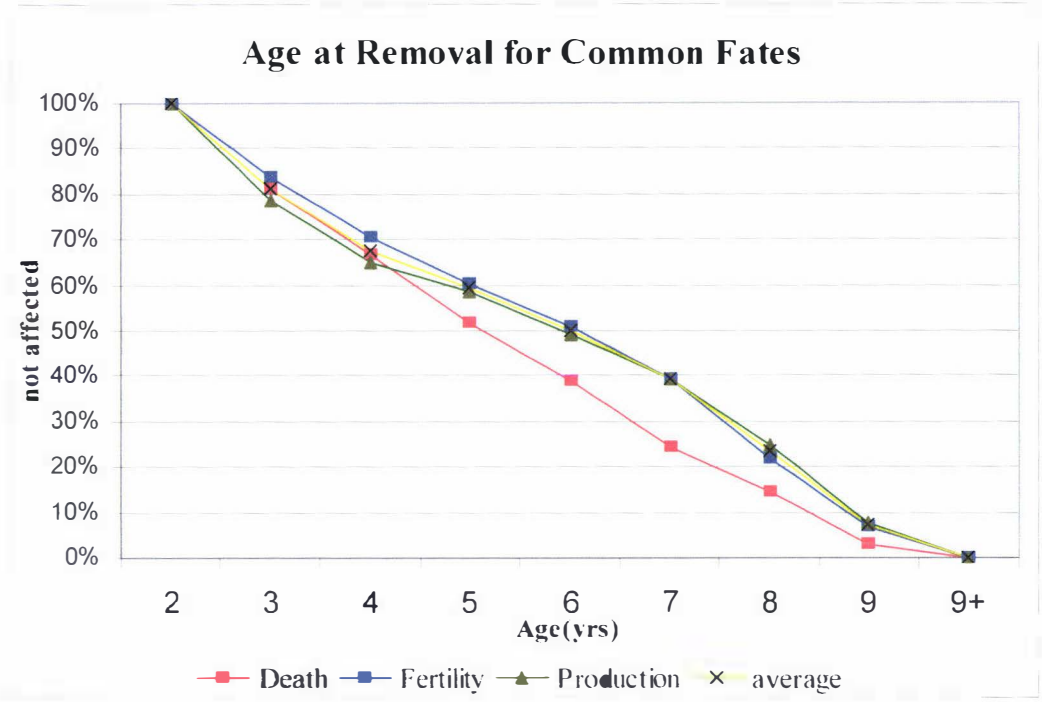


Figure 4.4 Onset of leading culling factors. Granddaughters affected by the three main fates were grouped according to age at removal. The percentage removed from the herd for each age was calculated starting at 0% for the 2 years olds.

	Average age days			
	9065 allele 1	4189 allele 1	9065 allele 2	4189 allele 2
Death	1975.4(176)	2146.7(103)	1967.7(148)	1979.8(83)
Fertility	2320.5(295)	2454.1(163)	2124.5(257)	2313.4(203)
Production	2304.9(74)	2298.4(25)	2207.6(75)	2228.1(55)

() number of animals

Table 4.3 Average age of animals (in days) affected by top three culling factors for each allele at BM1258

4.5 Discussion and Conclusion

Longevity values are normally calculated from the length of productive life, from calving to removal. In this study the calving dates were not known and so birth date to removal was used. Most studies prefer to use the length of productive life as the most valid measurement for longevity with the argument that a bias against early maturing animals is observed when age and not first calving date is used as a measurement. Conversely others argue that this measurement makes comparisons between herds difficult due to the different farm management of first mating. However herd effects can be easily added to the analysis whereas calculating the effect of time of first parturition as a time dependent covariate for each animal would be more difficult.

The fact that no herd effect was calculated into this study is a point of concern. The sample size was hopefully large enough to make this almost equal in both allele groups but this can not be guaranteed. Another concern is the large number of animals contributing to the unknown group which made drawing any conclusions from these data almost impossible. Although there is an equal percentage of unknowns for both alleles it is still the largest group.

When compared to other culling factor studies in dairy herds the proportions removed by each fate in this study appear to agree well. Fertility is consistently the number one cause for removal, usually responsible for approximately 20% of all culling. Mastitis, production and death are the next three main causes of removal in all other studies and this is also represented in this data set.

Death in this study appears to be responsible for at least 13% of animals removed. Similar numbers have been reported (Bascom and Young, 1998) however this is not often the case with death usually sitting below 5% in most American and European studies.

One group affected by death includes the unknown deaths. This fate code affects the young much more strongly than the older animals. This may be a real effect; however, the possibility that this category was misused and animals that were actually 'unknown removals' were placed under this category in the early years of data collection may be the reason why this group is so large.

The difference in the percentage culled in the first year between the two genotypes was significant and it suggests evidence for early onset of the phenotype. It is unfortunate that no information on heifer culling rates was available as if the variation was caused by a trait not directly related to production, culling in this first year may have been seen.
The first year also shows a large difference in percentage culled between the two grandsire families. This may be due to environmental differences as the 4189 granddaughters were born 4 years earlier, or to a sire effect. Only the first year shows such a large difference.

Production appeared to show suggestive segregation between the two alleles. It is difficult to believe that this phenotype alone is the cause of the longevity variation as the number culled primarily for production is small. However, production could be a secondary factor in many culling decisions and it is likely that its affects are underrepresented in this study. Single fate codes are not likely to give a full description of all the reasons why an animal was culled. More likely the animals will have been culled for multiple reasons, particularly in voluntary culling practises. Which fate is listed as the primary culling reason is at the discretion of the individual farmer. To combat this in the future multiple traits could be listed with relevant weighting for primary and secondary culling.

The trend, for the average ages of animals affected by death fertility and production, to be slightly lower in the inferior allele group does suggest that the phenotype affects all fates and could point to overall lower fitness that does not influence the fate but simply the onset of the fate. If this is the case the most likely cause is a health issue as a production

related issue not related to health would not have be expected to affect the age of animals removed for death and only minimal effects would be seen in fertility. This could also be the reason why a single fate did not clearly stand out. Another reason no single fate stood out could be due to the grouping of the fate codes. If the categories are too broad the influence of a minor health problem would be unnoticed if grouped with other fates.

In summary, few conclusions could be drawn from these data because of the large number of animals in the unknown fate group. Single fate data is also difficult to analyse as the effects of secondary culling factors are underestimated. If the animals are suffering from health problems only fatal cases would be represented, although it would be likely to affect the timing of fertility and production related culling.

Chapter 5: Candidate Genes for Longevity Phenotype

5.1 Introduction

Without a clear link to any major culling decisions and the variation in longevity, it is reasonable to assume the variation is a product of a more subtle nature, such as health. The general health of an animal has the ability to affect its production, fertility and its ability to protect itself from disease. With this in mind BTA23 was scanned for possible candidate genes. Several clear candidates, including the BoLA genes immediately stood out, however, the fact that the clear cause of this longevity variation has not been identified meant nearly every gene along BTA23 could still have the potentially affect longevity.

5.2 Comparative Map analysis

Comparative maps are generated by comparing EST library sequences generated from bovine tissue with the known sequences of other genomes, such as human. COMPASS (comparative map analysis of similar sequences) results of 47,747 ESTs generated from placenta, spleen and ovary bovine tissue (Lewin, 2001) are available in the public database. Using this database a list of over 300 genes known to be expressed in these tissues are predicted to be on BTA23.

The map has been generated by looking at the known genes on the bovine map, comparing their position in the human map and then placing the unknown genes according to the position of their human homologues on the bovine genome. This

assumes that if known gene homologues appear to be linked in a similar fashion in human then the genes of unknown position will also be similarly linked to the human homologues. Obviously not all ESTs can be placed in this way but fortunately the degree of homology between BTA23 and HSA6 is high and so it is likely that most genes that are really expressed on BTA23 will have homologues on HSA6.

Unfortunately the large inversion in BTA23 means that the homologous region is spread out on HSA6. This means that the entire length of HSA6 p arm should be considered when searching for candidates as there is the possibility that gene shuffling may have occurred before the inversion event. This area is approximately 60 Mbp and at the moment contains nearly 1400 predicted genes. The possibility that the QTL has no homologue on HSA6 cannot be forgotten, although it is reasonable to start a search amongst the genes contained on HSA6.

Using the HSA6 sequence a comparative map can be made with the mouse. The homology is not as strong as with the cattle/human. Chromosomes 17, 9, 13 and 1 in the mouse genome all appear to be linked to HSA6 with the mouse MHC found on chromosome 17.

At this stage the list of possible genes based on comparative mapping is still too large to simply choose candidates without further phenotype information.

5.3 OMIM study

Information on nearly all documented genetic diseases in humans can be found at the OMIM site (McKusick, 2002). Creating a list of all the diseases known to be associated with HSA6p to look for possible bovine comparisons was an option that needed to be investigated. Unfortunately the list created did not contain any immediate disease

candidates. Most were either too severe to be tolerated in commercial cattle or human specific, such as dyslexia. Even though there is a possibility that the genes have less dramatic mutations, this is not a realistic starting point. The OMIA (Online Mendalian Inheritance in Animals) site is not as informative as OMIM and contains no information on BTA23 and genetic disease in cattle (Nicholas, 2002)
In the phenotype analysis there was no link between disease frequency and the QTL for longevity. However, there is still the possibility of a link at the subclinical level.

5.4 Candidate genes from the MHC

Despite the fact that no significant difference could be seen in the fate codes for the granddaughters of the two families it was still possible to look at the BoLA locus for candidate genes.

The BoLA gene products contribute to the presentation of antigens to the immune system. An inferior allele may not be fatal but it may contribute to lower overall fitness in the animal and therefore a shorter lifespan. This may be seen in lower production, lower fertility or sub clinical health problems. This could be the reason why there was no significant difference in the percentage removed for different fate codes despite an overall decrease in lifespan.

There are several loci in the BoLA that are regularly heterozygous in an individual; the class I genes, the DRB3 genes and the DQA and B genes. All of these encode antigen binding proteins and it is commonly thought that variation in these genes is an advantage to a population in order to bind a large number of antigens. However, evidence shows that certain alleles are able to bind specific antigens better than others and so it is obvious that a phenotype change related to these genes is possible. Although it is important to keep a high level of polymorphism in the population at these loci it could be beneficial to

increase the percentage of alleles that appear to protect against common antigens such as mastitis causing bacteria.

Unfortunately the DQA and DQB loci are difficult to study as it is possible that the number of loci vary in cattle. Even if haplotyping by following the inheritance of the alleles through to the next generation were possible it becomes difficult to track phenotype changes as these are altered depending on which DQA molecule joins to which DQB molecule to form the functional heterodimer. It was for this reason that further study on these loci was terminated.

The MHC I haplotypes also vary in the number of loci per individual. However the MHC I gene products do not associate with another polymorphic molecule and therefore are easier to follow in subsequent generations.

The DRB3 alleles are the most studied genes in the BoLA. Their products associate with a non polymorphic DRA molecule and contribute to the binding of exogenous antigens such as bacteria. Extensive work looking at these genes as candidate genes for variation in disease resistance and fertility has been undertaken in a number of species including cattle with sometimes promising results. The high expression of this gene also suggests an important physiological role.

5.4.1 Amplification of candidate genes

In most studies of the BoLA region exon 2 of the DRB3 gene is the target for amplification. This exon, which codes for the antigen binding site, is highly polymorphic and only 320 base pairs in size. Amplification of this small exon alone allows common DRB3 alleles to be identified.

In the case of the MHC I. amplification of a 700 base pair fragment contains most of the common variations. Both of these products were amplified from gDNA. Using DRB3 exon two primers and BoLA I primers respectively (Section 2.11)

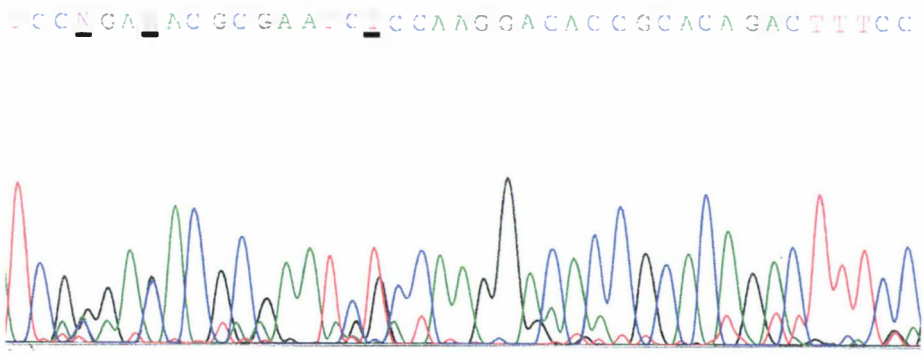


Figure 5.1 Polymorphic fragment of MHC gene. This sequence indicates the inability of the sequencer to identify all polymorphic sites. Note the underlined T which should have been read as N for thymine and guanine.

5.4.2 Grandsire Analysis

To ensure that the genes were heterozygous in the grandsires a genomic DNA sample from each sire was taken and the DRB3 and MHC I sequences amplified by PCR. Analysis of the sequences confirmed that both genes were indeed polymorphic in both sires at several base pair positions. As shown in figure 5.1 not all polymorphic sites were represented by an N this was not a problem in the case of these alleles as multiple polymorphic sites are present. However it highlights the need to scan sequences by eye and not to rely on base calculation software. Because of the large number of polymorphic sites it was impossible to define the exact alleles the grandsires carried from these sequences. This meant that cloning would be necessary.

5.4.3 Allele Identification

To identify the exact alleles in the sires the gene products amplified by PCR were cloned into p-GEMT vector and used to transform XL-1 bacterial cells. Using the pGEMT vector system allowed positive colonies to be identified using the blue/white selection method. Three plasmids from each sire were isolated from the positive clones and sequenced using the M13F primer which binds upstream of the insertion site. Analysis of these cloned sequences showed that the bovine DRB3 allele *1201 was present in both sires (Appendix C.1). Alleles 0901 and 0902 were present in grandsires 9065 and 4189 respectively. *0901 and *0902 are of the same major group meaning they are extremely similar in sequence (Davies *et al.*, 1997) (Appendix C.2).

The quality of sequencing from the clones for the MHC I appeared poor and exact haplotyping was not possible from the limited sequence data obtained.

5.4.4 Restriction Digests

Due to the large interest in the DRB3 gene in bovine an effort has been put into identifying useful restriction enzymes for haplotyping. These results are well documented at the BoLA nomenclature site (http://www.projects.roslin.ac.uk/bola/bolahome.html). Three enzymes, *Bst* Y1, *Rsa* 1 and *Hae* III have been shown to be the most informative.

With regard to identifying alleles *1201 and *090_ only *Bst*Y1 and *Rsa* 1 restriction enzymes were required. Unfortunately, because of the frequency of the *1201 allele bands it was difficult to identify individuals that were actually carrying this allele from ones that were carrying similar alleles, however the detection of the *090_ alleles was straight forward. This meant that although there may be an over estimate of the

individuals carrying *1201 a comparison could still be made between them and the individuals that inherited *090_ alleles.

Amplified cloned DRB3 exon 2 from both grandsires was digested with *Bst* Y1 and *Rsa* 1 to confirm the results of the clone sequences (Figure 5.2). Fragment lengths were tabulated using NIH Image1.62 software so the exact length of fragments could be compared to gels of the later generations.

5.4.5 Analysis of the DRB3 in the sons of grandsire 9065

Although the appearance of the *1201 allele in both sires was interesting the frequency of this allele in the New Zealand Holstein Friesian population is not known. To investigate this and to see if the inheritance of the *1201 of *090_ alleles are linked to variation in longevity values, the DNA of the sons of 9065 were amplified and digested with the restriction enzymes. Unfortunately the restriction map of the *1201 allele appeared in the majority of the samples making it difficult to identify individuals with allele *090_ only. Because of this the analysis of the sons of sire 9065 was carried out using allele 090_ present or absent as the categories. To ascertain whether a link between carrying 090_ alleles and variation in survival exists a simple T test was carried out between those that carried the allele and those that did not. This returned a P value of 0.069995 (n=28) suggesting a link between this allele and the survival trait.

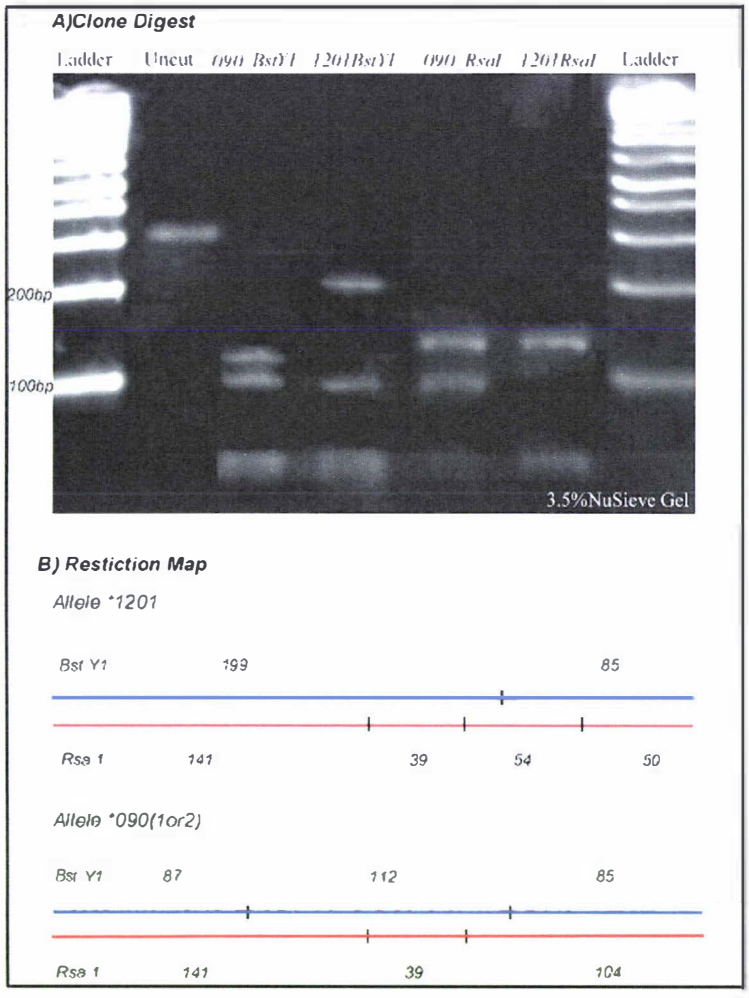


Figure 5.2 Restriction Map of alleles 09_ and 1201. A) Approximately 50 ng of PCR product amplified from the DRB3 clones from each grandsire was digested with *Rsa* I and *Bst*YI endonuclease restriction enzymes and the fragments separated on a 3.5% NuSieve gel in TAE and visualized under UV light after ethidium bromide staining. **B)** Map of restriction sites for alleles *1201 and *090_ showing the expected fragment sizes for the two different enzymes.

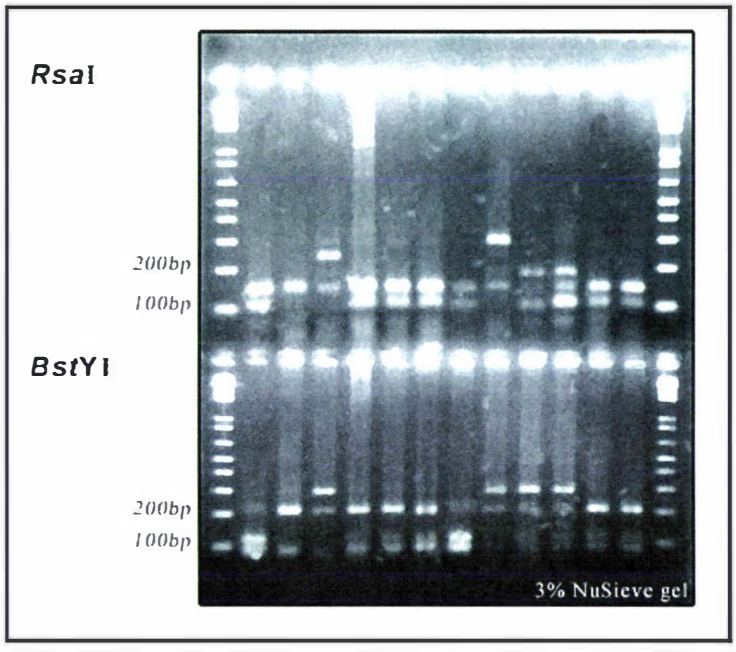


Figure 5.3 Digest of PCR for Grandsons. Individual DRB3 alleles were amplified by PCR and digested using *Rsa*I and *Bst*YI. Digests were then run on 3% NuSieve in TAE and stained with ethidium bromide and visualized using UV light. Lanes marked with M are loaded with 1kb plus ladder. All other lanes a grandson digests. The *BstYI* digest correspond to the same individual as the *Rsa* I digest run directly above them. There was some difficulty in haplotyping effectively due to the high frequency of the 1201 banding pattern. This could be due to a high frequency of the allele or because several alleles also had very similar digest patterns

5.4.6 Analysis in the Grandsons

The grandson's DRB3 exon 2 was also amplified and digested (Figure 5.3). As previously mentioned only three sons contribute to these grandsons. Sires 279 and 249, were shown to carry both *090_ and *1201 whilst sire 140 appeared to have inherited the *090_ allele only.

When each family group was analyzed only the family belonging to sire 249 showed a link between allele *090_ and longevity (P= 0.027033, n=31). This result is not surprising as a large percentage of the sons in the other two family groups were carrying both allele types and so a strong statistical correlation would not be expected. Unfortunately the *090_ alleles were linked to increased longevity in the grandsons rather than decreased longevity as observed in the sons.

5.8 Conclusions

The fact that the *090_ allele is linked to variation in two generations provides evidence that further study of this allele may be beneficial. Due to the fact that the *090_ allele is linked to both a negative and a positive affect in the two generations it is likely that the DRB3 locus is merely closely linked to the true QTL. However, the high heterogeneity of the MHC genes is thought to be a product of a population protecting itself from a wide variety of pathogens. With this in mind it is plausible that the allele may play different roles depending on the environment.

What did become evident in this study was the relatively low frequency of the *090_ alleles. This meant that the fact that both grandsires happened to possess this allele may

be significant. If the *090_ allele was truly linked to disease susceptibility it must somehow be acting through a dominant role, as the heterozygous nature of the MHC locus

would almost always inhibit the phenotype of a recessive allele. There are several examples of alleles playing a dominant negative role in the BoLA including BLV MHC related susceptibility (Xu et al, 1993). This is thought to be linked to the type of immune response that is elicited after T cell recognition rather than the antigen going undetected.

A recent study has reported a link between the *090_ alleles and susceptibility to bovine dermatophilosis. Although this disease is not seen in New Zealand herds the authors suggested that a variation in the structure of the antigen binding site coded for by the *090_ allele could be responsible. The authors postulated that the variation in the coded protein sequence would create a binding site which would associate quite weakly with the antigenic peptide therefore allowing the peptide to be exposed to proteases and thus cause weak recognition by the T cells. This theory might also be applied to other peptides that are bound by the *090_ coded molecule. However, as already mentioned, the MHC alleles are thought to be co-dominant in expression meaning that unless the individual is homozygous another binding site should act to present the antigenic peptide and thus no phenotypic variation should be seen. It could be possible that the results of Maillard *et al.*, 2002 are due to the same QTL as the one under investigation in this study.

Chapter 6: Discussion

Longevity is notoriously difficult to analyse in commercial cattle due to the wide variety of factors that can contribute to the length of time an animal spends in a herd. Large sample numbers normally aid in identifying subtle genetic factors which can influence important traits, however, the statistical noise that comes with adding individuals from different herds, environments and years mean that many QTL go undetected. Even with the QTL identified it is difficult to take research further especially when a clear understanding of the cause of the trait variation is not known. In this study, despite analysis of the granddaughter fate data, an obvious cause for the variation in longevity could not be discovered and without this information a candidate gene search is difficult. At this stage there would be little accomplished with the addition of more genetic markers. The information content achieved with the current markers should be sufficient for the initial study of other populations. The immediate issue to be addressed in this study is obtaining more informative fate data.

6.1 Fate Data

Analysis of the fate information in this study identified a problem in the collection and interpretation of these data. The unknown removal category, although large, is expected when attempting to collect data from commercial environments. Minimising this category could be difficult without any real incentives for farmers to collect and record detailed fate reports. For the purpose of this study it was assumed that the unknown category was represented by an equal percentage of all other fate groups. This is unlikely to be the case considering fate data is collected in milking periods only. This means animals lost in the dry months of the year would make up a disproportionate percentage of the unknown

group. These animals face quite different selection pressures than the milking herd and so a bias to these fates would affect the results.

Another major issue affecting the fate data records is the voluntarily culled animals. Because a large proportion of animals are voluntarily culled, the issue of farmer bias also needs to be addressed. It is not possible to obtain useful culling data from farmers using only a single fate code. As previously stated (see chapter 4) this could be dealt with by the collection of multiple culling reasons and scaled according to farmer priority. This information could be used in conjunction with the overall herd effect data to identify the objectives a farmer appears to be working towards in his culling decisions and their affect on the fate data. Obviously this would increase the work load for data collection but with a more in depth study not as many individuals would need to contribute. Without these data the fate codes can be used for little more than studying farmer perceptions of what is important in culling decisions and monitoring very general trends in disease and fertility.

For the purpose of this study other phenotype information would have been beneficial in the candidate gene search other than the end fate code. Information on the condition of the animal at the time of culling, where the animal was in its yearly cycle (ie milking or dry) and medical history would have been particularly useful for this study. The condition of the animal and its medical history would enable identification of any immune system deficiencies the animal may have. This is not addressed in current sire proving schemes directly, unless the animal dies because of it. However, to a degree the effect of variation in the immune system can be seen in fertility, production and longevity trait values. This is acceptable in a practical environment but means there is no information for the actual status of the immune system for studies such as this one. Obviously collecting consistent condition information and medical scores that can be compared to other animals is unrealistic in a working environment but may be necessary for the identification of this particular QTL.

In summary, to proceed further with this study the collection of fate data needs to be examined. The actual fate data collected from the entire offspring pool is of little use with so many factors biasing the end fate code. In future possibly only the age at removal should be collected and details on involuntarily culled animals. This data could then be used to monitor the general longevity of the herds and should identify any significant genetic problems. In order to obtain further information on culling decisions smaller studies could be set up to monitor fate data of selected farms, possibly only those already directly involved in sire proving schemes. This would mean that although fewer data would be collected the information would be more accurate and contain more relevant details enabling the detection of subtle changes and as well as making the results more reliable. Collecting enough specific data to analyse a single family may be difficult using this collection method, as commercial farms usually use a selection of bulls but this could be over come with planning in the preliminary stages of the experiment.

If research were to proceed based on the hypothesis that the families are affected by variation in disease resistance then direct challenge experiments could be an option. By directly infecting animals with known pathogens at controlled levels the ability of the superior haplotype compared to the inferior could be directly measured.

6.2 MHC allele variation

Obviously targeting the phenotype would be a costly experiment to set up especially when there is the possibility that no significant results may arise at the end of the study. For this reason further research into this QTL might be better spent with a large scale analysis of the BoLA alleles in the two immediate families.

To further study the MHC alleles a better method of haplotyping needs to be designed. Restriction endonuclease maps are useful for small sample numbers but problems arise

when individuals of unknown parentage are entered into the analysis and many alleles contain similar restriction endonuclease sites. Single Stranded conformation polymorphism (SSCP) gels could be used (Aldridge *et al.*, 1998) but these can be technically difficult to set up and can produce inconsistent results. They also have the disadvantage that not all alleles are different enough to distinguish between them on a gel. To date, the most practical methods to ensure the detection of all alleles are cloning and sequencing or mass spectrometry (Kristensen *et al.*, 2001). Mass spectrometry measures the size/charge ratio of a molecule and is sensitive enough to discriminate between DNA molecules that vary at a single nucleotide in a sequence greater than 700bp. This method can not only be used to identify genetic variants but is also capable of identifying a given allele if a standard sample is available. Although the sequencing method would be relatively simple and reliable to carry out it would also be quite time consuming and expensive. The mass spectrometry could be used as an accurate and quick method to screen for the different alleles at the beginning of a study thereby minimising the number of samples that would then need to be sequenced.

An increasing number of research groups are working to identify links between the BoLA genes and disease and fertility. Most of these studies use the candidate gene approach to select the BoLA and exploit the high polymorphism in order to carry out association studies. This project was different in that it was based around results of an entire genome scan, using microsatellites that identified the BoLA region as a likely site of a longevity QTL with no bias for the genes within it. The fact that this region was the only region to show significant variation in the 7 families tested indicates that the suggested role of the MHC in longevity may indeed be a reality. However, despite all the theories surrounding the MHC no study has yet identified the reason for the link between certain alleles and increased fitness. At the moment all evidence surrounding the MHC and phenotype variation is statistical only.

The allele 1201 is thought to be quite common in the bovine population and therefore its presence in the two sires was not as surprising as initially thought. The *090_ alleles

however are not as common and have been linked to disease susceptibility (Maillard *et al.*, 2002).

The reason for the maintenance of a large number of MHC haplotypes present in a given population is not fully understood. Obviously there must be some positive selection for the heterozygote; this is most likely due to its ability to defend itself from a wider variety of pathogens. Whether this theory is enough to explain all the variation seen within the MHC is still open to debate. The fact that allele *090_ behaved differently in two different populations may support the theory that the alleles play different roles in different environments and therefore supports the hypothesis that the heterozygous individual has an advantage.

The results of the candidate gene analysis suggested a link between the DRB3 locus and the phenotype variation. The association of the *090_ allele with increased longevity in the sons and decreased longevity in the grandsires can be explained by several models (Fig. 6.1). These models are explained as follows.

The haplotype analysis of the sons and the sire suggest that there was no crossover between the BM1258 marker and the DRB3 allele. This suggests that if the QTL was linked to the DRB3 allele because there was no change in the haplotype in the region of interest between the sire and the son then the marker/QTL/DRB3 haplotype inherited by the sons should have been the same. This may not have been the case; the possibility of a double crossover event between the known marker and DRB3 allele would allow a new haplotype to form that would not be detected by analysis of the flanking markers (Fig 6.1 A). The distance between the DRB3 locus and the marker BM1258 is approximately 10cM so although the chances of this occurring are slim it can not be ruled out.

Because the alleles are so varied at the DRB3 locus it is likely that the genotype *1201/*090_ in the grandsire was not the actual genotype in the sons. Although it

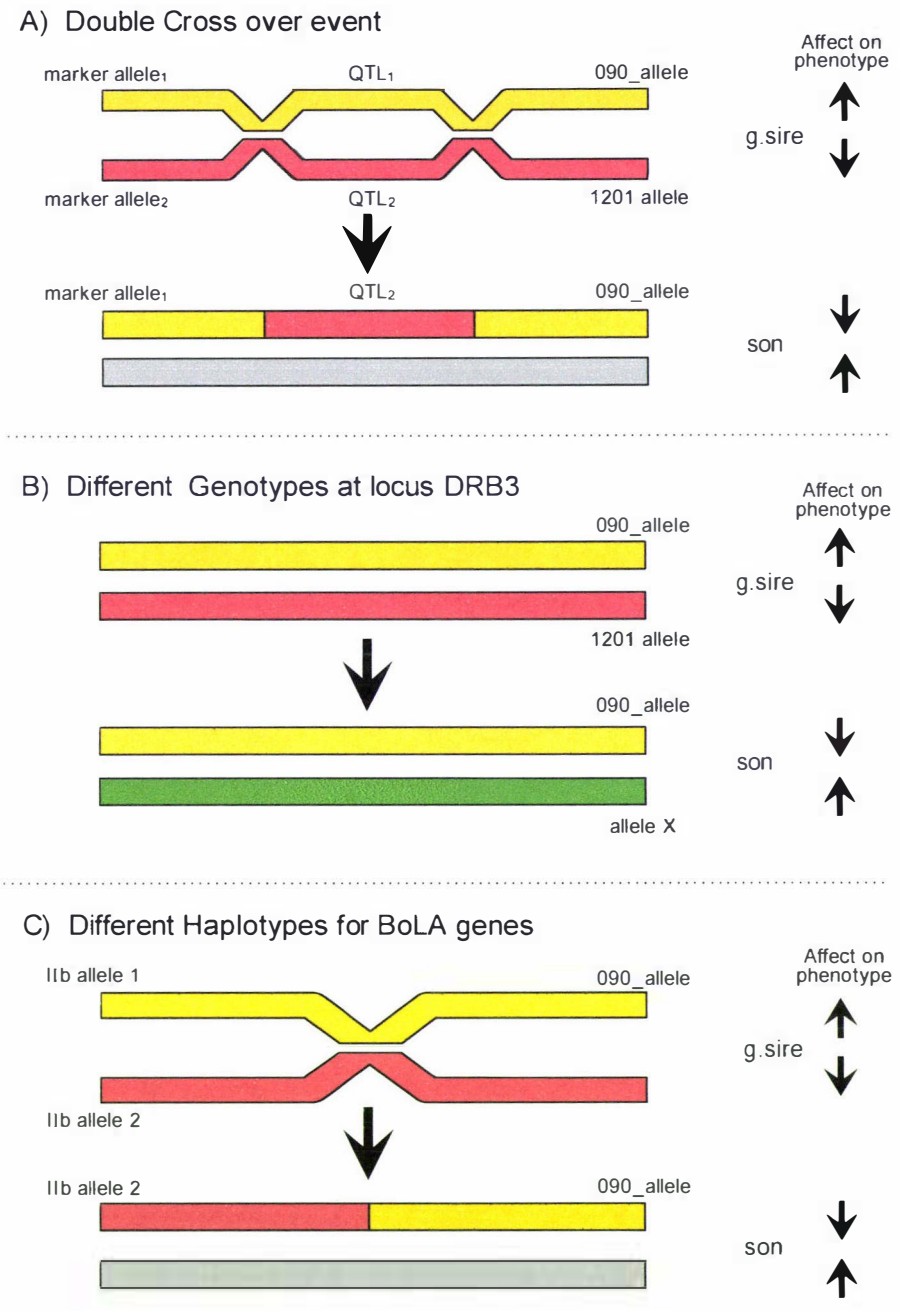


Figure 6.1 Possible models explaining phenotype results A) Double crossover event occurring between markers to remove the positive QTL from the haplotype 1. B) The maternal allele inherited at the DRB3 locus is a superior allele in the offspring and an inferior allele in the grandsire. C) The phenotype effect is also governed by the allele at the other class II loci, thus a positive affect is only seen in the correct haplotype.

appeared that the son did in fact inherit the *1201 allele and the *090_ allele, restriction endonuclease digests cannot distinguish all alleles and so it is entirely possible that the son actually inherited another genotype at this locus. The allele inherited from the mother could have been inferior to the *090_ allele and therefore the offspring carrying the *090_ allele would have had a phenotypic advantage (Fig 6.1 B).

A common theme that is repeated in studies of the MHC genes is that of the superior haplotype rather than the advantageous properties of a single locus. Because the class IIb genes are situated a significant genetic distance from the IIa genes recombination events between the two loci may have created new haplotypes that may have changed the ability of the individual to combat different pathogens (Fig 6.1 C).

The above models by no means cover all the possibilities that could explain the change in phenotype. The evidence is not strong enough at this stage to rule out simple random variation. There is also the fact that interactions between different alleles at different loci can lead to an infinite number of possibilities in regards to phenotype. These results simply provide evidence to the complexity of the relationship between different loci and alleles and explain the observations that QTL genome searches may not yield as much useful information for breeding systems as first thought.

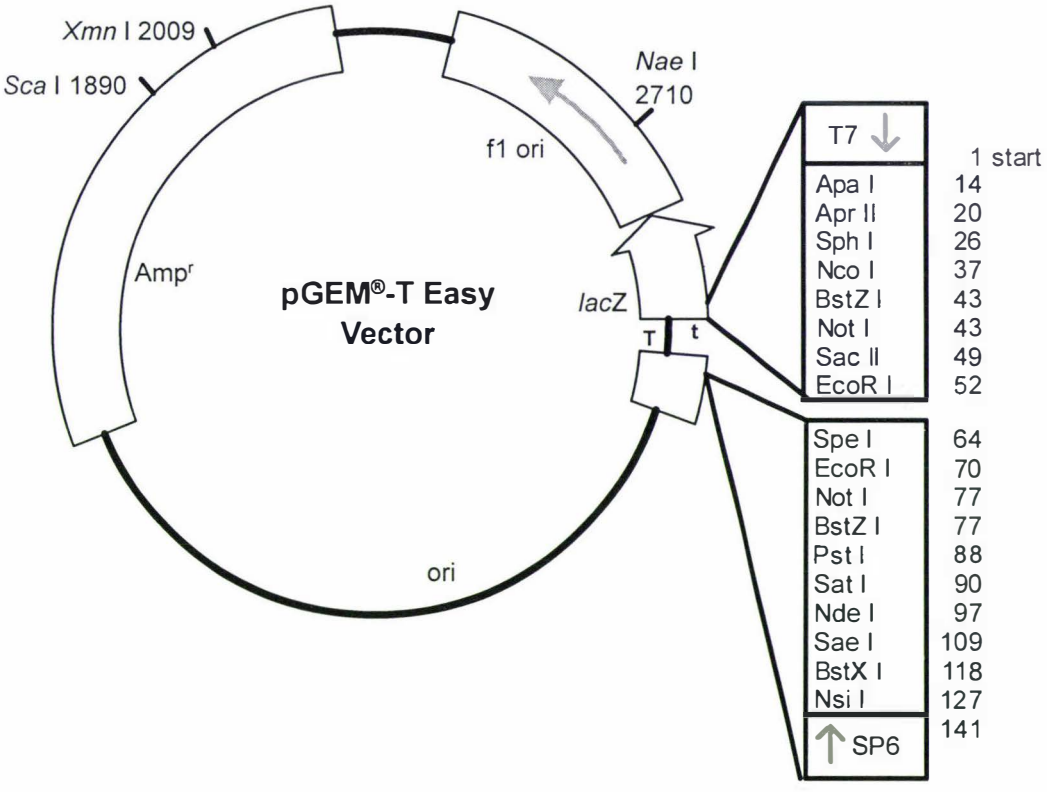
The study of the MHC genes would be the next logical step in this experiment. Allele frequencies for New Zealand herds would be important for a reliable study. This would involve a large number of individuals but the results would have the potential to aid in other MHC studies in bovine. Although the DRB3 locus is the most expressed class II gene it would be important to look at the other DQA, DQB and Class I genes to identify haplotypes that may be important.

In summary the results of this study suggest a link between the MHC region and longevity in cattle. Whether this is the result of polymorphisms in the MHC genes

themselves, other linked polymorphisms or a false positive was not determined but the region does warrant further investigation. The MHC contains a wide variety of polymorphic genes and it would be difficult to accept that at least some of the alleles do not have an advantage over the others. However, the large number of alleles in the MHC and the fact that they are co-dominantly expressed contribute to the difficulty in identifying conclusively the phenotypic advantage of each allele. The size of the bovine families now in commercial herds would make an excellent model for not only the bovine MHC but for further study of the MHC in general.

Appendices

Appendix A: pGEM-T Easy vector used for cloning BoLA DRB3 and BoLA I.



Appendix B - Genotype Data

Appendix B.1: Table 1-Haplotype Probabilities for Sons of Grandsire 9065

All probabilities calculated as per the methods in Section 3.4 for markers used on BTA23. Probabilities calculated from genotyping in this project are highlighted in bold. Haplotype one in yellow and haplotype 2 in green.

Appendix B.2: Table 2-Microsatellite Genotypes in the Grandsons

Microsatellite sizes for the markers used in grandsons belonging to sires 140 and 279. All microsatellites measured using the GeneScan software. Marker sizes given in base pairs.

Appendix B.3:Table 3-Haplotype Probabilities for Grandsons of Grandsire 9065

Probabilities calculated as per the methods of Section 3.4 for each cM along the possible QTL site on BTA23. Haplotype one in yellow and haplotype 2 in green.

Table A

Sire #	0cM	1cM	2cM	3cM	4cM	5cM	6cM	7cM	8cM	9cM	10cM	11cM	12cM	13cM	14cM	15cM	16cM	17cM	18cM	19cM
7422172	0	0.0003	0.00042	0.0003	0	0.0012	0.0022	0.0029	0.0035	0.0039	0.0039	0.0041	0.0041	0.0039	0.0035	0.0029	0.0021	0	0.0001	0
7473268	1	0.9997	0.99958	0.9997	1	0.9988	0.9978	0.9971	0.9965	0.9961	0.9961	0.9959	0.9959	0.9961	0.9965	0.9971	0.9979	1	0.9999	1
7473285	1	0.9997	0.99958	0.9997	1	0.9989	0.998	0.9973	0.9969	0.9966	0.9965	0.9966	0.9969	0.9973	0.998	0.9989	1	1	0.9999	1
7473290	0	0.00623	0.00504	0.00504	0	0.0011	0.002	0.0027	0.0031	0.0034	0.0035	0.0034	0.0031	0.0027	0.002	0.0011	0	0	0.0001	0
7473331	1	0.9997	0.99958	0.9997	1	0.9989	0.998	0.9973	0.9969	0.9966	0.9965	0.9966	0.9969	0.9973	0.998	0.9989	1	1	0.9999	1
7473343	0.9984	0.971	0.98	0.99	1	0.9215	0.8419	0.7659	0.6934	0.6184	0.6184	0.5396	0.4604	0.3816	0.3066	0.2341	0.0464	0	0.0001	0
7473358	0	0.06385	0.12703	0.2499	0.3122	0.3122	0.3778	0.4377	0.5	0.5623	0.6222	0.6878	0.7501	0.8116	0.873	0.9362	0.9654	1	0.9999	1
7474793	0	0.0003	0.00042	0.0003	0	0.0011	0.002	0.0027	0.0031	0.0034	0.0035	0.0034	0.0031	0.0027	0.002	0.0011	0	0.0002	0.0001	0
7483535	0	0.0003	0.00042	0.0003	0	0.0011	0.002	0.0027	0.0031	0.0034	0.0035	0.0034	0.0031	0.0027	0.002	0.0011	0	0	0.0001	0
7483540	0.9984	0.971	0.98	0.99	1	0.9989	0.998	0.9973	0.9969	0.9966	0.9965	0.9966	0.9969	0.9973	0.998	0.9989	1	1	0.9999	1
7487761	0.9507	0.029	0.02	0.01	0	0.0011	0.002	0.0027	0.0031	0.0034	0.0035	0.0034	0.0031	0.0027	0.002	0.0011	0	0	0.0001	0
7501790	0	0.0003	0.00042	0.0003	0	0.0011	0.002	0.0027	0.0031	0.0034	0.0035	0.0034	0.0031	0.0027	0.002	0.0011	0	0	0.0001	0
7510716	0	0.06385	0.12703	0.2499	0.3122	0.3122	0.3778	0.4377	0.5	0.5623	0.6222	0.6878	0.7501	0.8116	0.873	0.9362	1	1	0.9999	1
7517904	1	1	0.9997	0.9997	1	0.9989	0.998	0.9973	0.9969	0.9966	0.9965	0.9966	0.9969	0.9973	0.998	0.9989	1	1	0.99	0.9955
7543112	0	0.0003	0.00042	0.0003	0	0.0011	0.002	0.0027	0.0031	0.0034	0.0035	0.0034	0.0031	0.0027	0.002	0.0011	0	0	0.0001	0
7543119	0.9984	0.971	0.98	0.99	1	0.9989	0.998	0.9973	0.9969	0.9966	0.9965	0.9966	0.9969	0.9973	0.998	0.9989	1	1	0.9999	1
7543131	1	1	0.9997	0.9997	1	0.9989	0.998	0.9973	0.9969	0.9966	0.9965	0.9966	0.9969	0.9973	0.998	0.9989	1	1	0.9999	1
7543134	1	0.9997	0.99958	0.9997	1	0.9988	0.9978	0.9971	0.9965	0.9961	0.9961	0.9959	0.9959	0.9961	0.9965	0.9971	0.9979	1	0.5	0
7552770	0	0.0003	0.00042	0.0003	0	0.0011	0.002	0.0027	0.0031	0.0034	0.0035	0.0034	0.0031	0.0027	0.002	0.0011	0	0	0.5	1
7557296	0	0.0003	0.00042	0.0003	0	0.0012	0.0022	0.0029	0.0035	0.0039	0.0039	0.0041	0.0041	0.0039	0.0035	0.0029	0.0021	0	0.0001	0
7557300	0	0.00623	0.00504	0.00504	0	0.0011	0.002	0.0027	0.0031	0.0034	0.0035	0.0034	0.0031	0.0027	0.002	0.0011	0	0	0.0001	0
7562257	1	0.9997	0.99958	0.9997	1	0.9989	0.998	0.9973	0.9969	0.9966	0.9965	0.9966	0.9969	0.9973	0.998	0.9989	1	0.6689	0.33108	0
7575114	1	0.9997	0.99958	0.9997	1	0.9989	0.998	0.9973	0.9969	0.9966	0.9965	0.9966	0.9969	0.9973	0.998	0.9989	1	1	0.9999	1
7577661	1	0.74727	0.5	0.25273	0	0.0011	0.002	0.0027	0.0031	0.0034	0.0035	0.0034	0.0031	0.0027	0.002	0.0011	0	0	0.0001	0
7577665	0	0.25273	0.5	0.74727	1	0.9989	0.998	0.9973	0.9969	0.9966	0.9965	0.9966	0.9969	0.9973	0.998	0.9989	1	1	0.99	0.916
7587203	0	0	0.00141	0.00262	0.0035	0.0043	0.0048	0.0053	0.0055	0.0055	0.0053	0.0048	0.0043	0.0035	0.0026	0.0014	0	0	0.01	0.084
7596293	1	0.93615	0.87297	0.81161	0.7501	0.6878	0.6222	0.5623	0.5	0.4377	0.3778	0.3122	0.2499	0.1884	0.127	0.0638	0.0598	0	0.0001	0
7620064	1	0.9997	0.99958	0.9997	1	0.9989	0.998	0.9973	0.9969	0.9966	0.9965	0.9966	0.9969	0.9973	0.998	0.9989	1	1	0.9999	1
7620076	0	0.0003	0.00042	0.0003	0	0.0012	0.0022	0.0029	0.0035	0.0039	0.0039	0.0041	0.0041	0.0039	0.0035	0.0029	0.0021	0	0.5	1
7658249	1	0.9997	0.99958	0.9997	1	0.9989	0.998	0.9973	0.9969	0.9966	0.9965	0.9966	0.9969	0.9973	0.998	0.9989	1	1	0.9999	1
7668616	1	0	0.93282	0.8631	0.7988	0.7334	0.6645	0.6009	0.5348	0.4652	0.3991	0.3355	0.2666	0.2012	0.1369	0.0672	0	0	0.0001	0
7709558	0	0.00623	0.00504	0.00504	0	0.0785	0.1581	0.2341	0.3066	0.3816	0.3816	0.4604	0.5396	0.6184	0.6934	0.7659	0.9536	1	0.99	0.999
7741135	1	0.74727	0.5	0.25273	0	0.0011	0.002	0.0027	0.0031	0.0034	0.0035	0.0034	0.0031	0.0027	0.002	0.0011	0	0	0.0001	0
7776155	1	0.9997	0.99958	0.9997	1	0.9989	0.998	0.9973	0.9969	0.9966	0.9965	0.9966	0.9969	0.9973	0.998	0.9989	1	1	0.9999	1
7817903	1	0.9997	0.99958	0.9997	1	0.9989	0.998	0.9973	0.9969	0.9966	0.9965	0.9966	0.9969	0.9973	0.998	0.9989	1	1	0.9999	1
7835413	1	0.9997	0.99958	0.9997	1	0.9989	0.998	0.9973	0.9969	0.9966	0.9965	0.9966	0.9969	0.9973	0.998	0.9989	1	1	0.9999	1
7888878	0	0.0003	0.00042	0.0003	0	0.0851	0.1728	0.2506	0.3308	0.415	0.5	0.585	0.6692	0.7494	0.8272	0.9149	1	1	0.9999	1
7932279	1	0.9997	0.99958	0.9997	1	0.9507	0.9019	0.8572	0.8111	0.763	0.763	0.714	0.6686	0.6169	0.5703	0.5265	0.4735	0.4297	0.38313	0.237

Table B

	Marker Size							
Sire #	BM1815		BM1258		MB025		L29386	
12608396	148.28	154.28	99.89		119.63	134.15	167.48	171.54
12621081	148.28		98.01	99.89	120.44	130.97		
12636077	150.39	165.93	99.89		120.46	134.16	171.47	
12642226			97.9	99.78	125.67	134.25	169.46	171.54
12656963			99.89	103.65	125.52	140.81	163.41	161.33
12657491			98.89	102.51	133.64		166.27	169.31
12658516	104	166	102	104			170	172
12660785	148.28	146.27	99.89	101.76	121.48	134.15	164.33	170.43
12669955	148.38	151.87	99.89	103.63	118.54	134.04		
12678010	148.28	166.06	101.66	103.63	121.49	125.61	171.54.	
12678993	148.27	166.02	99.89	103.61	125.62	134.16		
12708239			102.59	104.24	119.32	125.29	167.34	172.41
12723050	148.28		100	103.64	125.96	133.15	165.4	171.62
12731047	144.8		99.89		134.15	130.03		
12737958	148.3		98.02	101.77	121.56	134.2		
12744719	148.28		98.04	101.75	121.45	129.96		
12760522	148.29		98.02	99.89	134.17		165.27	169.4
12763632	148	162	100	102	126	135	165	
12766642			99.56	100.63	119.33	133.63		
12801552	148.28		100		134.18	140.92		
12801937			99.67	100.63	129.8	133.94	163.23	165.22
12807606		148.28	97.9	99.89	134.15			
12808521	148.28		99.89		129.01	134.16	169.45	171.53
12815953	148.28		98.01		134.16			
12861205			100.63	102.51	121.26	123.13	163.11	164.06
12918558	147	149	99.89	100	133	135	166	170
12954298			104.35		119.23	133.85	166.35	170.41
13063692	148.28	166.06	99.89	103.55	125.88		167.48	171.54
13086895	148	166	98	104	126	134	166	172
13092975			101.47	102.52	121.3	123.27	166.36	172.41
13105975	148.27	166.02	101.65	103.51	118.46	121.46		
13110636	165.97			103.55	119.57	125.67	171.55	
13125673	148.38		99.78		119.51	134.15		
13143202			98.78	100.63	133.93		165.19	
13143216	161.9	165.97	101.3	103.55	125.67		167.48	171.54
13192605	148.28	165.56	101.67		119.61	134.15		
13198401	165.93	148.29	101.76	103.63	121.51	125.63		
13203347	148.28	165.97	99.89	103.56	119.56	134.15		
13311856	166.05	164.5	99.78	103.54	134.16		165.49	167.56
13516117	148	166	99	98	125			
13518184	148		99	103	125	136		
13519507	148	150			118	125		
13525454	148.38	165.93	99.78	103.52	119.52	125.56		
13531548			100.63	104.39	119.28	125.31	170.35	172.34
13535238	165.97		99.78	103.53	125.58	136.42	165.4	
13549615			100.63	104.3	123.19	136.19	160.19	166.27
13561268	148.28		99.89	101.78	119.67	125.67		
13571646			98.67	10.63	125.52	134.03	170.48	172.56
13634883	148	166	98	101	126	124	166	
13643989	148.28	165.56	99.78	101.66	118.41	123.36		

13661674	148.28	166.06	99.89	101.77	119.61	134.16		
13668324			100.63	102.51			165.36	166.31
13712409	148		99	103	126	136		
13748806	148		97	100	123	125		
13754609			98.78	100.63			166.18	172.34
13755267			98.78	100.63	123.2	129.65	166.18	
13829564	148.28	161.79	99.89	101.76	119.54	134.16	171.47	165.37
13836476			100.63	104.42	125.42		171.45	172.39
13847549	148.28	165.96	103.53		118.52	124.6	165.4	
13882407	156.14	165.93	99.78	103.52	123.39	125.56	161.32	163.3
13905751			97.67	98.67	123.19	125.38	166.37	172.34
13941843	148		99	101	118	122	165	
13965997	148.27		99.78	125.56			163.41	
13973058	148.28		99.89	101.66	125.68			
14128334	148.38		97.91	99.78	125.56	134.14		
14163409			100.63	104.3	119.23	123.19	166.18	
14227470					119.59	125.68	165.27	
14259326	148	165	99		124	136		
14259344	148.28		98.02	99.78	123.36	134.06		
14370990	148.28		99.89	103.54	119.6	123.42		
14424238	148.28	164.08	101.57	103.24	125.67			
14434397	148	166	99	101	118	122		
14528362	148	165	99		118	124		
14617866	148.27	132.91	97.8	99.78				
15065730	148	165	103	99	119	123		
15656114	148.28	165.96	99.78	103.58	123.46	125.63		

Table C

Sire #	1cM	2cM	3cM	4cM	5cM	6cM	7cM	8cM	9cM	10cM	11cM	12cM	13cM	14cM	15cM	16cM	17cM	18cM	19cM
12608396	1	0.99979	0.99979	1	0.9989	0.99801	0.99734	0.99685	0.99658	0.99649	0.99658	0.99685	0.99734	0.99801	0.9989	1	0.99979	0.99979	1
12636077	0	0.33108	0.66892	1	0.9989	0.99801	0.99734	0.99685	0.99658	0.99649	0.99658	0.99685	0.99734	0.99801	0.9989	1	0.99979	0.99979	1
12642226	0.971	0.98	0.99	1	0.99859	0.99738	0.99647	0.99573	0.9952	0.99473	0.99448	0.99448	0.99473	0.9952	0.99573	0.99647	0.99738	0.99859	1
12656963	0.129	0.123	0.114	0.106	0.098	0.089	0.082	0.074	0.065	0.056	0.047	0.038	0.029	0.02	0.01	0	0.00021	0.00021	0
12658516	0	0.00021	0.00021	0	0.06718	0.1369	0.20121	0.26663	0.33547	0.39908	0.46522	0.53478	0.60092	0.66453	0.73337	0.79879	0.8631	0.93282	1
12660785	1	0.99979	0.99979	1	0.9989	0.99801	0.99734	0.99685	0.99658	0.99649	0.99658	0.99685	0.99734	0.99801	0.9989	1	0.99979	0.99979	1
12678010	0.029	0.02	0.01	0	0.0011	0.00199	0.00266	0.00315	0.00342	0.00351	0.00342	0.00315	0.00266	0.00199	0.0011	0	0.33108	0.66892	1
12708239	0.029	0.02	0.01	0	0.0011	0.00199	0.00266	0.00315	0.00342	0.00351	0.00342	0.00315	0.00266	0.00199	0.0011	0	0.00021	0.00021	0
12723050	1	0.66892	0.33108	0	0.01	0.02	0.029	0.038	0.047	0.056	0.065	0.074	0.089	0.098	0.106	0.114	0.123	0.129	0.137
12760522	1	0.99979	0.99979	1	0.9989	0.99801	0.99734	0.99685	0.99658	0.99649	0.99658	0.99685	0.99734	0.99801	0.9989	1	0.66892	0.33108	0
12763632	1	0.66892	0.33108	0	0.00141	0.00262	0.00353	0.00427	0.0048	0.00527	0.00552	0.00552	0.00527	0.0048	0.00427	0.00353	0.00262	0.00141	0
12766642	0.971	0.98	0.99	1	0.9989	0.99801	0.99734	0.99685	0.99658	0.99649	0.99658	0.99685	0.99734	0.99801	0.9989	1	0.99	0.98	1
12808521	1	0.99979	0.99979	1	0.9989	0.99801	0.99734	0.99685	0.99658	0.99649	0.99658	0.99685	0.99734	0.99801	0.9989	1	0.99979	0.99979	1
12918558	1	0.99979	0.99979	1	0.9989	0.99801	0.99734	0.99685	0.99658	0.99649	0.99658	0.99685	0.99734	0.99801	0.9989	1	0.99	0.98	0.971
12954298	0.029	0.02	0.01	0	0.08506	0.1728	0.25057	0.33079	0.41501	0.5	0.58499	0.66921	0.74943	0.8272	0.91494	1	0.99	0.98	0.971
13063692	0.129	0.123	0.114	0.106	0.098	0.089	0.082	0.074	0.065	0.056	0.047	0.038	0.029	0.02	0.01	0	0.00021	0.00021	0
13110636	0	0.00021	0.00021	0	0.0011	0.00199	0.00266	0.00315	0.00342	0.00351	0.00342	0.00315	0.00266	0.00199	0.0011	0	0.00021	0.00021	0
13143202	0.971	0.98	0.99	1	0.9989	0.99801	0.99734	0.99685	0.99658	0.99649	0.99658	0.99685	0.99734	0.99801	0.9989	1	0.66892	0.33108	0
13143216	0	0.00021	0.00021	0	0.0011	0.00199	0.00266	0.00315	0.00342	0.00351	0.00342	0.00315	0.00266	0.00199	0.0011	0	0.33108	0.66892	1
13311856	0	0.06718	0.1369	0.20121	0.26663	0.33547	0.39908	0.46522	0.53478	0.60092	0.66453	0.73337	0.79879	0.8631	0.93282	1	0.66892	0.33108	0
13535238	0	0.00141	0.00262	0.00353	0.00427	0.0048	0.00527	0.00552	0.00552	0.00527	0.0048	0.00427	0.00353	0.00262	0.00141	0	0.00021	0.00021	0
13550330	0.971	0.98	0.99	1	0.9989	0.99801	0.99734	0.99685	0.99658	0.99649	0.99658	0.99685	0.99734	0.99801	0.9989	1	0.99	0.98	0.971
13624952	0.152	0.145	0.137	0.129	0.123	0.114	0.106	0.098	0.089	0.082	0.074	0.065	0.056	0.047	0.038	0.029	0.02	0.01	0
13829564	1	0.99979	0.99979	1	0.9989	0.99801	0.99734	0.99685	0.99658	0.99649	0.99658	0.99685	0.99734	0.99801	0.9989	1	0.99	0.98	0.971
13847549	0.029	0.02	0.01	0	0.0011	0.00199	0.00266	0.00315	0.00342	0.00351	0.00342	0.00315	0.00266	0.00199	0.0011	0	0.00021	0.00021	0
13882407	0	0.00141	0.00262	0.00353	0.00427	0.0048	0.00527	0.00552	0.00552	0.00527	0.0048	0.00427	0.00353	0.00262	0.00141	0	0.00021	0.00021	0
14227470	0.129	0.123	0.114	0.106	0.098	0.089	0.082	0.074	0.065	0.056	0.047	0.038	0.029	0.02	0.01	0	0.00021	0.00021	0
12621081				1	0.99	0.98	0.971	0.962	0.953	0.944	0.935	0.926	0.918	0.911	0.902	0.894			
12657491				0.894	0.902	0.911	0.918	0.926	0.935	0.944	0.953	0.962	0.971	0.98	0.99	1			
12669955				0.894	0.902	0.911	0.918	0.926	0.935	0.944	0.953	0.962	0.971	0.98	0.99	1			
12731047				1	0.9989	0.99801	0.99734	0.99685	0.99658	0.9298	0.94193	0.99685	0.99734	0.99801	0.9989	1			
12744719				0.106	0.098	0.089	0.082	0.074	0.065	0.056	0.047	0.038	0.029	0.02	0.01	0			
12801552				0	0.08506	0.1728	0.25057	0.33079	0.41501	0.04453	0.07269	0.66921	0.74943	0.8272	0.91494	1			
12801937				1	0.9989	0.99801	0.99734	0.99685	0.99658	0.9298	0.94193	0.99685	0.99734	0.99801	0.9989	1			
12807606				1	0.9989	0.99801	0.99734	0.99685	0.99658	0.9298	0.94193	0.99685	0.99734	0.99801	0.9989	1			

Table C continued

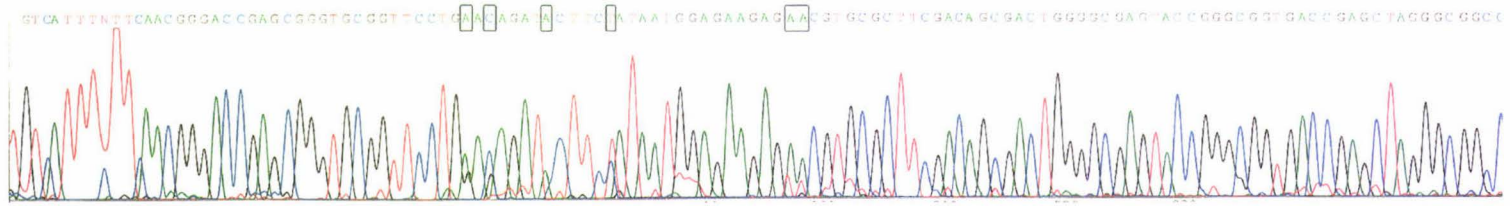
12815953				1	0.9989	0.99801	0.99734	0.99685	0.99658	0.9298	0.94193	0.99685	0.99734	0.99801	0.9989	1			
12861205				0	0.0011	0.00199	0.00266	0.00315	0.00342	0.0702	0.05807	0.00315	0.00266	0.00199	0.0011	0			
13092975				0	0.0011	0.00199	0.00266	0.00315	0.00342	0.0702	0.05807	0.00315	0.00266	0.00199	0.0011	0			
13105975				0	0.0011	0.00199	0.00266	0.00315	0.00342	0.0702	0.05807	0.00315	0.00266	0.00199	0.0011	0			
13125673				1	0.9989	0.99801	0.99734	0.99685	0.99658	0.9298	0.94193	0.99685	0.99734	0.99801	0.9989	1			
13192605				0	0.08506	0.1728	0.25057	0.33079	0.41501	0.04453	0.07269	0.66921	0.74943	0.8272	0.91494	1			
13198401				0	0.0011	0.00199	0.00266	0.00315	0.00342	0.0702	0.05807	0.00315	0.00266	0.00199	0.0011	0			
13203347				1	0.9989	0.99801	0.99734	0.99685	0.99658	0.9298	0.94193	0.99685	0.99734	0.99801	0.9989	1			
12581810																1			
13007670																1			
13042628																0			
13061178																0			
13088903																0			
13211727																1			
13516117																1			
13518184																1			
13519507																1			
13525454																1			
13531548																1			
13549615																0			
13561268																1			
13571646																1			
13643989																0			
13712409																1			
13748806																1			
13755267																0			
13836476																1			
13941843																0			
13973058																1			
14128334																1			
14163409																0			
14259326																0			
14259344																0			
14370990																0			
14424238																1			
14434397																0			
15065730																0			

Appendix C- Variations in the exon 2 of DRB3

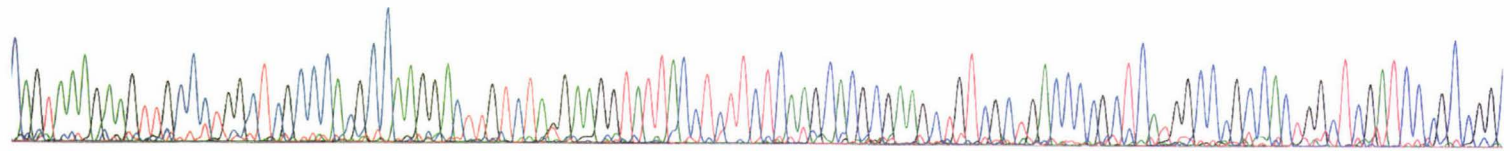
Appendix C.1 Sequence Alignment for exon 2

Alignment of sequences for DRB3 clones and genomic DNA. Heterozygous base positions highlighted with boxes.

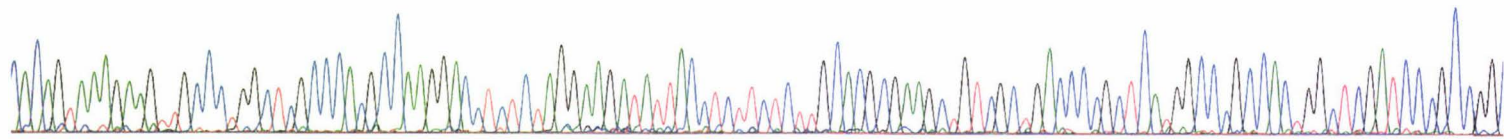
gDNA for DRB3 exon 2

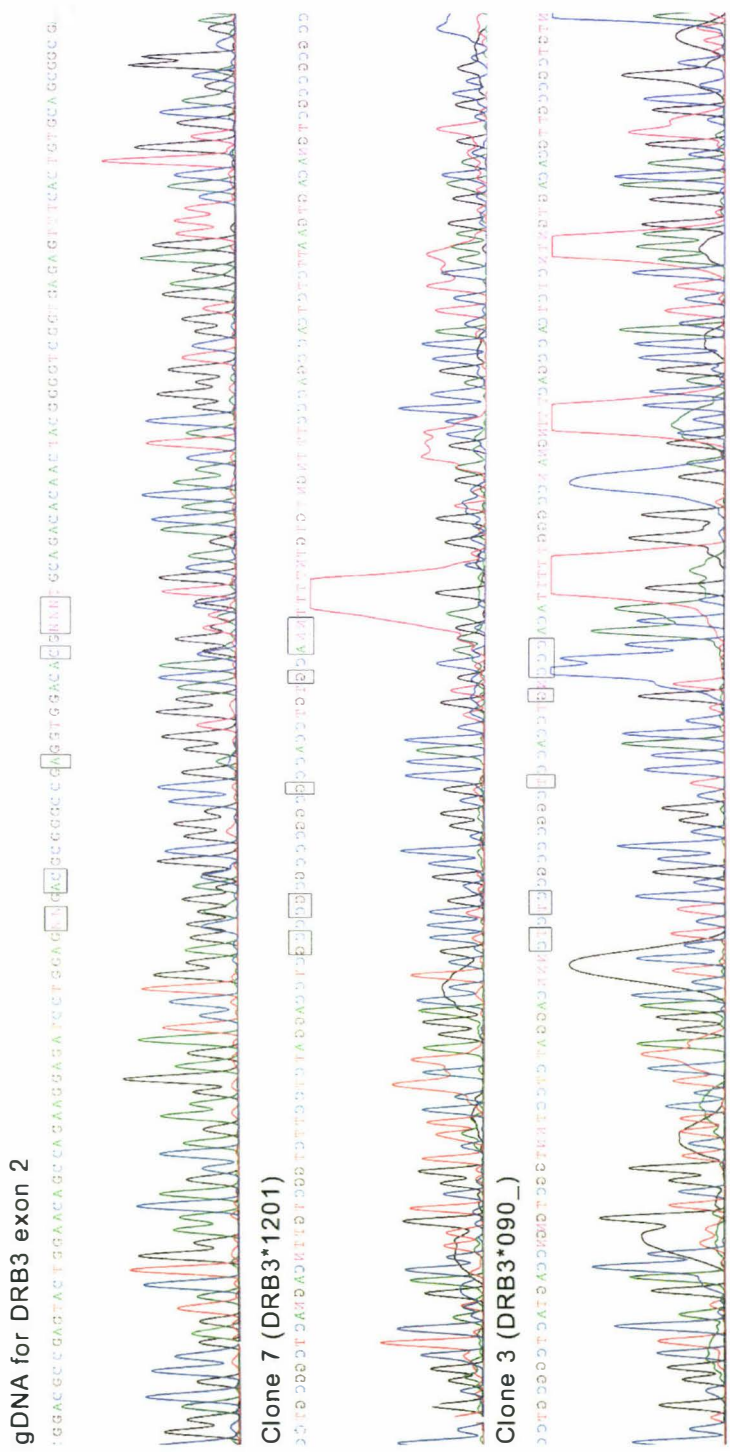


Clone 7 (DRB3*1201)



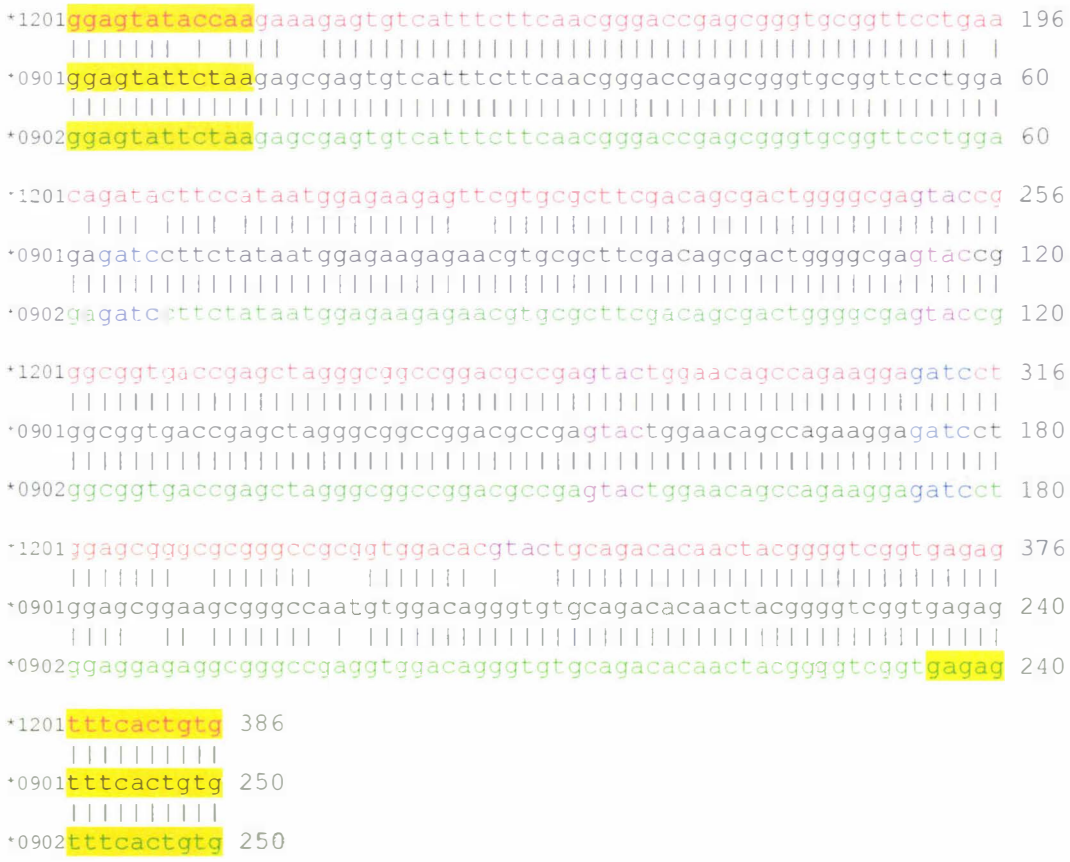
Clone 3 (DRB3*090_)




gDNA for DRB3 exon 2
Clone 7 (DRB3*1201)
Clone 3 (DRB3*090_)

Appendix C.2 Alignment of Grandsire DRB3 exon 2 alleles

Alignment of the 1201, 0901 and 0902 alleles found in the grandsires. Highlighted area indicates primer positions. Sequences in blue show restriction endonuclease sites.



References

Aarestrup, F. M., N. E. Jensen and H. Ostergard (1995). "Analysis of associations between major histocompatability complex (BoLA) class I haplotypes and subclinical mastitis of dairy cows." Journal of Dairy science 78: 1684-1692.

Aldridge, B. M., S. M. McGuirk, R. J. Clark, L. A. Knapp, D. I. Watkins and D. P. Lunn (1998). "Denaturing gradient gel electrophoresis: a rapid method for differentiating BoLA-DRB3 alleles." Anim Genet 29(5): 389-394.

Andersson-Eklund, L. and J. Rendel (1993). "Linkage between amylase-1 locus and a major gene for milk fat content in cattle." Anim Genet. 24(2): 101-3.

Arriens, M. A., A. Hofer, G. Obexer-Ruff and S. Lazary (1996). "Lack of association of bovine MHC class I alleles with carcass and reproductive traits." Animal Genetics 27(6): 429-31.

Ashwell, M. S., Y. Da, C. P. Van Tassell, P. M. Vanraden, R. H. Miller and C. E. J. Rexroad (1998). "Detection of putative loci affecting milk production and composition, health, and type traits in a United States Holstein population." J Dairy Sci. 81(12): 3309-14.

Ashwell, M. S. and C. P. Van Tassell (1999). "Detection of putative loci affecting milk, health, and type traits in a US Holstein population using 70 microsatellite markers in a genome scan." J Dairy Sci. 82(11): 2497-502.

Band, M., J. H. Larson, J. E. Womack and H. A. Lewin (1998). "A radiation hybrid map of BTA23: identification of a chromosomal rearrangement leading to separation of the cattle MHC class II subregions." Genomics 53(3): 269-75.

Band, M. R., J. H. Larson, M. Rebeiz, C. A. Green, D. W. Heyen, J. Donovan, R. Windish, C. Steining, P. Mahyuddin, J. E. Womack and H. A. Lewin (2000). "An ordered comparative map of the cattle and human genomes." Genome Research 10: 1359-1368.

Bascom, S. S. and A. J. Young (1998). "A summary of the reason why farmers cull cows." Journal of Dairy science 81(8): 2299-2305.

Batra, T. R., A. J. Lee, J. S. Gavora and M. J. Stear (1989). "Class I alleles of the bovine major histocompatibility system and their association with economic traits." J Dairy Sci. 72(8): 2115-24.

Bernoco, D. and H. A. Lewin (1989). "Genetic aspects of bovine leukaemia virus infection and disease progression." Anim Genet 20(3): 337-339.

Bishop, M. D., S. M. Kappes, J. W. Keele, R. T. Stone, S. L. F. Sunden, G. A. Hamkins, S. Solinas Toldo, R. Fries, M. D. Grosz, J. Yoo and C. W. Beattie (1994). "A genetic linkage map for cattle." Genetics 136: 619.

Bjorkman, P. J., M. A. Saper, B. Samraoui, W. S. Bennett, J. L. Strominger and D. C. Wiley (1987). "Structure of the human class I histocompatibility antigen, HLA-A2." Nature 8-14;329(6139)(506-12).

Choudhury, S. R. and L. A. Knapp (2001). "Human reproductive failure II: immunogenetic and interacting factors." Hum Reprod Update 7(2): 135-60.

Christiansen, O. B., B. Pedersen, O. Mathiesen, M. Husth and N. Grunnet (1996). "Maternal HLA class II alleles predispose to pregnancy losses in Danish women with recurrent spontaneous abortions and their female relatives." Am J Reprod Immunol 35(3): 239-44.

Churchill, G. A. and R. W. Doerge (1994). "Emphirical Threshold values for Quantitative Trait Mapping." Genetics 138: 963-971.

Collins, E. J., D. N. Garboczi and D. C. Wiley (1994). "Three-dimensional structure of a peptide extending from one end of a class I MHC binding site." Nature 13(371): 626-629.

Cowan, C. M., M. R. Dentine and L. A. Schuler (1990). "Structural variation around prolactin gene linked to quantitative traits in an elite holstein sire family." Theor Appl Genet 79: 577-582.

Creighton, P., A. Eggen, R. Fries, S. A. Jordan, J. Hetzel, E. P. Cunningham and P. Humphries (1992). "Mapping of bovine markers CYP21, PRL, and BOLA DRBP1 by genetic linkage analysis in reference pedigrees." Genomics 4(2): 526-528.

Davies, C. J., L. Andersson, S. A. Ellis, E. J. Hensen, H. A. Lewin, S. Mikko, N. E. Muggli-Cockett, J. J. van der Poel and G. C. Russell (1997). "Nomenclature for factors of the BoLA system, 1996: report of the ISAG BoLA Nomenclature Committee." Animal Genetics 28: 159-168.

de Koning, D. J., N. F. Schulmant, K. Elo, S. Moisio, R. Kinos, J. Vilkki and A. Maki-Tanila (2001). "Mapping of multiple quantitative trait loci by simple regression in half-sib designs." J Anim Sci. 79(3): 616-22.

Ducrocq, V. (1987). "An analysis of length of productive life in dairy cattle." Ph.D. Thesis, Cornell Univ., Ithaca, NY, U.S.

Ducrocq, V. and J. Sölkner (1994). "The Survival Kit- a Fortran package for the analysis of survivaldata." Proc 5th World Congress on Genetics Applied to Livestock Production 22: 51-52.

Elo, K. T., J. Vilkki, D. J. de Koning, R. J. Velmala and A. V. Maki-Tanila (1999). "A quantitative trait locus for live weight maps to bovine chromosome 23." Mamm Genome 10(8): 831-5.

Escayg, A. P., J. G. H. Hickford and D. W. Bullock (1997). "Association between alleles of the ovine major histocompatibility complex and resistance to footrot." Research in Veterinary Science 63: 283-287.

Esslemont, R. J. and M. A. Kossaibati (1997). "Culling in 50 dairy herds in England." Vet Rec. 140(2): 36-9.

Fremont, D. H., M. Matsumura, E. A. Stura, P. A. Peterson and I. A. Wilson (1992). "Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb." Science 14;257(5072): 919-27.

Garcia, K. C. and L. Teyton (1998). "T-cell receptor peptide-MHC interactions: biological lessons from structural studies." Curr Opin Biotechnol. 9(4): 338-43.

Georges, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto, A. T. Pasquino, L. S. Sargeant, A. Sorensen, M. R. Steele and X. Zhao (1995). "Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing." Genetics 139(2): 907-20.

MHC binding site structure (2001). [Online Image]. Uni of Pavia, Italy. Available from: <http://webmed.unipv.it/immunology/mhcstr.html>.

Goldsby, R. A., T. J. Kindt and R. B. Osborne (2000). Kuby Immunology, 4th edition. W.H. Freeman and Company.

Grohn, Y. T., S. W. Eicker, V. Ducrocq and J. A. Hertl (1998). "Effect of diseases on the culling of Holstein dairy cows in New York State." J Dairy Sci. 81(4): 966-78.

Haldane, J. B. S. (1919). "The combination fo linkage vaules and the calculation of distances between the loci of linked factors." J Genet 8: 299-309.

Haley, C. S. (1995). "Livestock QTLs--bringing home the bacon?" Trends Genet. 11(12): 488-92.

Haley, C. S. and S. A. Knott (1992). "A simple regression model for interval mapping in line crosses." Heredity 69: 315-324.

Haley, C. S., S. A. Knott and J. M. Elsen (1994). "Mapping quantitative trait loci in crosses between outbred lines using least squares." Genetics 136(3): 1195-11207.

Haley, C. S. and P. M. Visscher (1998). "Strategies to utilize marker-quantitative trait loci associations." J Dairy Sci 81 Suppl 2: 85-97.

Hayes, P. M., B. H. Liu, F. Knapp, B. Chen, T. Jones, J. Blake, J. Franckowiak, D. Rasmusson, M. Sorrells, S. E. Ullrich, D. Wesenberg and A. Kleinhofs (1993). "Quantitaive trait locus effects and environmental interaction in a sample of North American barley germplasm." Theor Appl Genet 87: 392-401.

Hess, M., T. Goldammer, A. Gelhaus, K. Ried, G. Rappold, A. Eggen, M. D. Bishop, M. Schwerin and R. D. Horstmann (1999). "Physical assignment of the bovine MHC class IIa and class IIb genes." Cytogenet Cell Genet. 85(3-4): 244-7.

Holmes, D. S. and M. Quigley (1981). "A rapid boiling method for the preparation of bacterial plasmids." Anal Biochem. 114(1): 193-197.

Jansen, R. C. (1993). "Maximum likelhood in a generalized linear finite misture model by using EM algorithm." Biometrics 49: 227-231.

Jansen, R. C. (1992). "A general mixture model for mapping quantitative trait loci by using molecular markers." Theor Appl Genet 85: 252-260.

Joyce, S., K. Kuzushima, G. Kepecs, R. H. Angeletti and S. G. Nathenson (1994). "Characterization of an incompletely assembled major histocompatibility class I molecule (H-2Kb) associated with unusually long peptides: implications for antigen processing and presentation." Proc Natl Acad Sci U S A 10(91(10)): 4145-4149.

Kao, C. H. (2000). "On the differences between maximum likelihood and regression interval mapping in the analysis of quantitative trait loci." Genetics 156(2): 855-65.

Kappes, S. M., J. W. Keele, R. T. Stone, R. A. McGraw, T. S. Sonstegard, T. P. L. Smith, N. L. Lopez-Corrales and C. W. Beattie (1997). "A second generation linkage map of the bovine genome." Genome Res 7: 235.

Karuppanan, P., M. C. Thurmond and I. A. Gardner (1997). "Survivorship approaches to measuring and comparing cull rates for dairies." Preventive Veterinary Medicine 30: 171-179.

Kelm, S. C., A. E. Freeman and D. H. Kelley (1997). "Realized versus expected gains in milk and fat production of Holstein cattle, considering the effects of days open." J Dairy Sci. 80(8): 1786-94.

Knapp, S. J. (1991). "Using molecular markers to map multiple quantitave trait loci: Models for backcross, recombinant inbred, and doubled haploid progeny." Theor Appl Genet 81: 333-338.

Knapp, S. J., W. C. Bridges and D. Birkes (1990). "Mapping quantitative trait loci using molecular marker linkage maps." Theor Appl Genet 79: 583-592.

Knipling, E.B. (2002). "U.S. Meat Animal Research Centre".[Internet].Agricultural Research Service, U.S. Avaliable from:<http://www.marc.usda.gov/genome/genome.html>.

Kossarek, L. M., W. M. Grosse, T. S. Finco and R. A. McGraw (1993). "Rapid Communication:Bovine dinucleotide polymorphism RM002." J. Anim. Sci 71: 3173.

Kristensen, V. N., D. Kelefiotis, T. Kristensen and A. Borresen-Dale (2001). "High-Throughput Methods for Detection of Genetic Variation." Biotechniques 30: 318-332.

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

Lehenbauer, T. and J. W. Oltjen (1998). "Dairy cow culling strategies: making economical culling decisions." J Dairy Sci 81(1): 264-71.

Lewin, H. A. (1996). Genetic Organization, polymorphism, and function of the bovine major histocompatability complex.In The Major Histocompatability Complex Region of Domestic Animals.Chapter 4, CRC Press.

Lewin, H. A., G. C. Russell and E. J. Glass (1999). "Comparative organization and function of the major histocompatibility complex of domesticated cattle." Immunol Rev. 167: 145-58.

Lewin, H.A. (2001). "Compass Predictions" [Internet]. Bioinformatics Unit, Uni. of Illinois, IL, U.S. Available from: <http://titan.biotec.uiuc.edu/>.

Lipkin, E., M. O. Mosig, A. Darvasi, E. Ezra, A. Shalom, A. Friedmann and M. Soller (1998). "Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: analysis of milk protein percentage." Genetics 149(3): 1557-67.

Little, C. C. and E. E. Tyzzer (1916). "Further experimental studies on the inheritance of susceptibility to a transplantable tumor, carcinoma (JWA) of the Japanese waltzing mouse." J. Med. Res. 33: 393-453.

Lunden, A., S. Sigurdardottir, I. Edfors-Lilja, B. Danell, J. Rendel and L. Andersson (1990). "The relationship between bovine major histocompatibility complex class II polymorphism and disease studied by use of bull breeding values." Animal Genetics 21(3): 221-32.

Lustgarten, J., T. Waks and Z. Eshhar (1991). "CD4 and CD8 accessory molecules function through interactions with major histocompatibility complex molecules which are not directly associated with the T cell receptor-antigen complex." Eur J Immunol 21(10): 2507-15.

Lyon, E. and E. L. Frank (2001). "Hereditary hemochromatosis since discovery of the HFE gene." Clin Chem 47(7): 1147-56.

Mackinnon, M. J. and M. A. Georges (1992). "The effects of selection on linkage analysis for quantitative traits." Genetics 132(4): 1177-85.

Maillard, J. C., I. Chantal, D. Berthier, S. Thevenon, I. Sidibe and H. Razafindraibe (2002). "Molecular immunogenetics in susceptibility to bovine dermatophilosis: a candidate gene approach and a concrete field application." Ann N Y Acad Sci. 969: 92-96.

Mallard, B. A., J. C. Dekkers, M. J. Ireland, K. E. Leslie, S. Sharif, C. L. Vankampen, L. Wagter and B. N. Wilkie (1998). "Alteration in immune responsiveness during the peripartum period and its ramification on dairy cow and calf health." J Dairy Sci 81(2): 585-95.

Marello, K. L., A. Gallagher, D. J. McKeever, R. L. Spooner and G. C. Russell (1995). "Expression of multiple DQB genes in Bos indicus cattle." Anim Genet 26(5): 345-9.

Martinez, O. and R. N. Curnow (1992). "Estimation the locations and the sizes of the effects of quantitave trait loci using flanking markers." Theor Appl Genet 85: 480-488.

McKusick, V.A (2002). "Online Mendelain Inheritance in Man"[Internet]. National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, U.S. Available from: <http://www.ncbi.nlm.nih.gov/Omim/>.

McPherson, J. D., M. Marra and W. R. Hillier L, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis ER, Wilson RK, Fulton R, Kucaba TA, Wagner-McPherson C, Barbazuk WB, Gregory SG, Humphray SJ, French L, Evans RS, Bethel G, Whittaker A, Holden JL, McCann OT, Dunham A, Soderlund C, Scott CE, Bentley DR, Schuler G, Chen HC, Jang W, Green ED, Idol JR, Maduro VV, Montgomery KT, Lee E, Miller A, Emerling S, Kucherlapati, Gibbs R, Scherer S, Gorrell JH, Sodergren E, Clerc-Blankenburg K, Tabor P, Naylor S, Garcia D, de Jong PJ, Catanese JJ, Nowak N, Osoegawa K, Qin S, Rowen L, Madan A, Dors M, Hood L, Trask B, Friedman C, Massa H, Cheung VG, Kirsch IR, Reid T, Yonescu R, Weissenbach J, Bruls T, Heilig R, Branscomb E, Olsen A, Doggett N, Cheng JF, Hawkins T, Myers RM, Shang J, Ramirez L, Schmutz J, Velasquez O, Dixon K, Stone NE, Cox DR, Haussler D, Kent WJ, Furey T, Rogic S, Kennedy S, Jones S, Rosenthal A, Wen G, Schilhabel M, Gloeckner G, Nyakatura G, Siebert R, Schlegelberger B, Korenberg J, Chen XN, Fujiyama A, Hattori M, Toyoda A, Yada T, Park HS, Sakaki Y, Shimizu N, Asakawa S, Kawasaki K, Sasaki T, Shintani A, Shimizu A, Shibuya K, Kudoh J, Minoshima S, Ramser J, Seranski P, Hoff C, Poustka A, Reinhardt R, Lehrach H. (2001). "A physical map of the human genome." Nature 409(6822): 934-41.

Mosig, M. O., E. Lipkin, G. Khutoreskaya, E. Tchourzyna, M. Soller and A. Friedmann (2001). "A whole genome scan for quantitative trait loci affecting milk protein percentage in Israeli-Holstein cattle, by means of selective milk DNA pooling in a daughter design, using an adjusted false discovery rate criterion." Genetics 157(4): 1683-98.

Nicholas, F.W. (2002). "Online Mendelain Inheritance in Animals"[Internet]. Faculty of Veterinary science, University of Sydney, NSW, Australia. Avaliable from: <http://www.angis.su.oz.au/Databases/BIRX/omia/>.

Nonnecke, B. J., H. A. Lewin and L. L. Timms (1989). "Preliminary report on BoLA polymorphism in Guernsey cattle." J Dairy Sci 72(12): 3255-8.

Ober, C., T. Karrison, R. R. Odem, R. B. Barnes, D. W. Branch, M. D. Stephenson, B. Baron, M. A. Walker, J. R. Scott and J. R. Schreiber (1999). "Mononuclear-cell immunisation in prevention of recurrent miscarriages: a randomised trial." Lancet 31;354(9176): 365-9.

Ota, M., S. Bahram, Y. Katsuyama, S. Saito, Y. Nose, M. Sada, H. Ando and H. Inoko (2000). "On the MICA deleted-MICB null, HLA-B*4801 haplotype." Tissue Antigens. Sep; 56(3): 268-71.

Palmer, B. R., J. G. H. Hickford and R. Bickerstaffe (1997). "A candidate gene approach to animal quality traits." Proceedings of the New Zealand Society of Animal Production 57: 294-296.

Rajala-Schultz, P. J., Y. T. Grohn and H. G. Allore (2000). "Optimizing breeding decisions for Finnish dairy herds." Acta Vet Scand. 41(2): 199-212.

Rebeiz, M. and H. A. Lewin (2000). "COMPASS of 47,747 cattle ESTs." Animal Biotechnology 11: 75-241.

Riquet, J., W. Coppieters, N. Cambisano, J. Arranz, P. Berzi, S. K. Davis, B. Grisart, F. Farnir, L. Karim, J. Taylor, P. Vanmanshoven, D. Wagenaar, J. E. Womack and M. Georges (1999). "Fine-mapping of quantitative trait loci by identity by descent in outbred populations: application to milk production in dairy cattle." Proc Natl Acad Sci U S A 96(16): 9252-7.

Sawhney, S. M., D. W. Taylor and G. C. Russell (2001). "Polymorphism of bovine major histocompatibility complex (MHC) class I genes revealed by polymerase chain reaction (PCR) and restriction enzyme analysis." Animal Genetics 32: 27-31.

Seegers, H., F. Beaudeau, C. Fourichon and N. Bareille (1998). "Reasons for culling in French Holstein cows." Prev Vet Med 36(4): 257-71.

Sharif, S., B. A. Mallard, B. N. Wilkie, J. M. Sargeant, H. M. Scott, J. C. Dekkers and K. E. Leslie (1998). "Associations of the bovine major histocompatibility complex DRB3 (BoLA-DRB3) with production traits in Canadian dairy cattle." Anim Genet 30(2): 157-60.

Skow, L. C., J. Goy and D. Honeycutt (1994). "Dinucleotide repeat polymorphism near a bovine MHC class I sequence." Animal Genetics 25: 290.

Smith, T. P., W. M. Grosse, B. A. Freking, A. J. Roberts, R. T. Stone, E. Casas, J. E. Wray, J. White, J. Cho, S. C. Fahrenkrug, G. L. Bennett, M. P. Heaton, W. W. Laegreid, G. A. Rohrer, C. G. Chitko-McKown, G. Pertea, I. Holt, S. Karamycheva, F. Liang, J. Quackenbush and J. W. Keele (2001). "Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle." Genome Res. 11(4): 626-630.

Spelman, R. J., W. Coppieters, L. Karim, J. A. van Arendonk and H. Bovenhuis (1996). "Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population." Genetics 144(4): 1799-808.

Spelman, R. J. and D. J. Garrick (1998). "Genetic and economic responses for within-family marker-assisted selection in dairy cattle breeding schemes." J Dairy Sci 81(11): 2942-50.

Starkenburg, R. J., L. B. Hansen, M. E. Kehrli and H. Chester-Jones (1997). "Frequencies and effects of alternative DRB3.2 alleles of bovine lymphocyte antigen for Holsteins in milk selection and control lines." J. Dairy Sci. 80: 3411-3419.

Stear, M. J., T. S. Pokorny, S. E. Echternkamp and D. D. Lunstra (1989). "The influence of the BoLA-A locus on reproductive traits in cattle." J Immunogenet. 16(1): 77-88.

Stern, L. J., J. H. Brown, T. S. Jardetzky, J. C. Gorga, R. G. Urban, J. L. Strominger and D. C. Wiley (1994). "Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide." Nature 368(6468): 215-221.

Stevenson, M. A. and I. J. Lean (1998). "Risk factors for culling and deaths in eight dairy herds." Aust Vet J 76(7): 489-94.

Stone, R. T., J. C. Pulido, G. M. Duyk, S. M. Kappes, J. W. Keele and C. W. Beattie (1995). "A small insert bovine genomic library highly enriched for microsatellite repeat sequences." Mamm Genome 6: 714.

Stuber, C. W. (1995). "Mapping and manipulating quantitative traits in maize." Trends Genet. 11(12): 477-81.

Takasuga, A., S. Hirotsune, R. Itoh, A. Jitohzono, H. Suzuki, H. Aso and Y. Sugimoto (2001). "Establishment of a high throughput EST sequencing system using poly(A) tail-removed cDNA libraries and determination of 36 000 bovine ESTs." Nucleic Acids Research, 29(22): e108-108.

Tozer, P. R. and A. J. Heinrichs (2001). "What affects the costs of raising replacement dairy heifers: a multiple-component analysis." J Dairy Sci 84(8): 1836-44.

VanRaden, P. M. and G. R. Wiggans (1995). "Productive life evaluations: calculation, accuracy, and economic value." J Dairy Sci 78(3): 631-8.

Vollema, A. R. and A. F. Groen (1997). "Genetic correlations between longevity and conformation traits in an upgrading dairy cattle population." J Dairy Sci. 80(11): 3006-14.

Vollema, A. R. and A. F. Groen (1998). "A comparision of breeding value predictos for longevity using a linear model and survival analysis." Journal of Dairy science 81: 3315-3320.

Watson, A. L. and E. J. Yunis (1983). "Immunogenetics and longevity." Prog Clin Biol Res. 133: 77-88.

Weller, J., Y. Kashi and M. Soller (1990). "Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle." J Dairy Sci. 73(9): 2525-37.

Whitaker, D. A., J. M. Kelly and S. Smith (2000). "Disposal and disease in 340 British dairy herds." Vet Record 146(13): 363-367.

Xu, A., M. J. van Eijk, C. Park and H. A. Lewin (1993). "Polymorphism in BoLA-DRB3 exon 2 correlates with resistance to persistent lymphocytosis caused by bovine leukemia virus." J Immunol 15;151(12): 6977-85.

Young, A. C., S. G. Nathenson and J. C. Sacchettini (1995). "Structural studies of class I major histocompatibility complex proteins: insights into antigen presentation." FASEB J 9(1): 26-36.

Zhang, Q., D. Boichard, I. Hoeschele, C. Ernst, A. Eggen, B. Murkve, M. Pfister-Genskow, L. A. Witte, F. E. Grignola, P. Uimari, G. Thaller and M. D. Bishop (1998). "Mapping quantitative trait

loci for milk production and health of dairy cattle in a large outbred pedigree." Genetics 149(4): 1959-73.