DISSERTATION

IMPROVING GENETIC PREDICTIONS BY ACCOUNTING FOR MENDELIAN SAMPLING OR INBREEDING DEPRESSION

Submitted by

Azwihangwisi Maiwashe

Department of Animal Sciences

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2004

COLORADO STATE UNIVERSITY

October 26, 2004

WE HEREBY RECOMMEND THAT THE **DISSERTATION** PREPARED UNDER OUR SUPERVISION BY AZWIHANGWISI MAIWASHE ENTITLED IMPROVING GENETIC PREDICTIONS BY ACCOUNTING FOR MENDELIAN SAMPLING OR INBREEDING DEPRESSION BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Adviser

Department Head

ABSTRACT OF **DISSERTATION**

IMPROVING GENETIC PREDICTIONS BY ACCOUNTING FOR MENDELIAN SAMPLING OR INBREEDING DEPRESSION

Current genetic evaluation procedures (Best Linear Unbiased Prediction (BLUP) animal model) employed in the livestock industries throughout the world provide accurate predictions of breeding values. However, accuracy of predictions can still be enhanced. For example, BLUP methodology uses genetic relationships computed from pedigree information to allow for optimal utilization of all sources of information (performance of the individual and its relatives). Each source of information is weighted differently depending on the proportion of genes "shared" in common between relatives. However, these relationships among collateral relatives assume an average covariance. This averaging can sometimes limit the accuracy of prediction. Joint use of pedigree and currently available genetic (DNA) marker information can provide relationships that account for the variation in the covariance among collateral relatives due to Mendelian sampling. Furthermore, most current evaluation procedures do not account for decline in performance associated with increasing levels of inbreeding (inbreeding depression). The effective population sizes in many livestock populations are small and often in decline, causing increased inbreeding. In view of these opportunities for enhancement of accuracy of prediction, two independent investigations were conducted in this study. For

convenience, the two investigations were categorized as the inbreeding study and the genetic relationship study.

The primary objectives of the inbreeding study were to 1) assess the level of inbreeding, and 2) quantify inbreeding depression on early female reproduction, long-term cow productivity and carcass traits in cattle registered with the Red Angus Association of America. Inbreeding coefficients were computed using 829,882 pedigree records of animals born between 1930 and 2001. Most animals (89%) had both parents known. The inbreeding trend was evaluated over the period 1960 to 2001. Effects on performance of inbreeding of the individual (F_d) and that of its dam (F_m) were estimated. Reproductive traits included: heifer calving success (HCS, n = 1,197); heifer calving ease (HCE, n = 636); and cow stayability (CS, n = 14,268). Carcass traits were: hot carcass weight (HCW, n = 951); rib-eye area (REA, n = 947); backfat thickness (BFT, n = 767); and marbling score (MRB, n = 947).

Individual inbreeding coefficients ranged from 0 to 51% with an average inbreeding level of 3.1% for the entire population. Considering the period 1992 to 2001, where pedigrees were more complete, the rate of inbreeding was 0.08%±0.002 per year. This rate of inbreeding indicates that inbreeding is accumulating at a minimal rate in the registered Red Angus population. However, the fact that individuals exist with inbreeding coefficients as high as 51% is suggestive of occasional deliberate or inadvertent intense inbreeding. The partial regressions of female performance on F_d expressed on a probability scale were 0.75, -1.44, and -0.56 %/% for HCR, HCE and CS, respectively. Corresponding partial regressions of female performance on F_m were -0.40, 0.20, and 0.12 %/%. All partial regression coefficients were nonsignificant ($P > 0.10$) except for

CS on F_d (P = 0.0007) indicating that a percentage increase in inbreeding coefficient of the cow was associated with a 0.56% decline in the probability of the cow to maintain production until or beyond six years of age. Partial regressions of carcass traits on F_d were -0.71 kg/%, -0.04 cm^2/%, 0.0005 cm/%, and -0.011 per % for HCW, REA, BFT, and MRB, respectively. Corresponding values for F_m were -0.45 kg/%, -0.12 cm^2/%, -0.0012 cm/%, and -0.003 per %. The only significant partial regression coefficients were those of HCW on F_d (P = 0.0238) and REA on F_m (P = 0.0111). Results from this study provided little evidence of unfavorable relationships between inbreeding and performance in the registered Red Angus population at least for the current levels of inbreeding.

The objectives of the relationship study were to 1) derive weighting factors for different sources of information available for predicting genetic merit of an individual, and 2) evaluate the change in accuracy of genetic prediction when the inverse of the marker-based numerator relationship matrix (A_M) was substituted for the inverse of the standard (pedigree) numerator relationship matrix (A_P) in the mixed model equations. The data used to achieve the first objective comprised of daughter yield deviation (DYD) records on one grandsire and half-sib sons. The grandsire and his sons had several hundred daughters with yield deviation records. The DYD is an average of the daughters' yields adjusted for fixed and non-genetic random effects of the daughters and genetic effects of the dam. Predicted transmitting abilities (PTA) were obtained using a sire model. The PTA for each son was expressed as a linear function of his DYD if available and PTA of his relatives. The weights for a half-sib bull using A_P are always zero except for his DYD and the grandsire's PTA indicating that only its daughters' records (adjusted

for the merit of the mate) and parent contribute directly to the evaluation of the individual. When A_P was replaced with A_M, the evaluation of a half-sib bull had non-zero weights for all half-sibs with the weight on each half-sib varying with the proportion of alleles shared in common indicating that half- and full-sibs can also contribute directly to the evaluation of the individual. This result demonstrate that the genetic merit of a bull with no daughters could be more reliably predicted using A_M rather than A_P if he had half-sibs with daughters because the inferiority or superiority of his Mendelian sampling could be assessed to some extent.

The data set used to achieve the second objective comprised of records on 1,811 progeny-tested Holstein bulls. Each record comprised DYD for milk, fat and protein yield and genotypic information on 52 microsatellite markers. The markers were located in interesting quantitative trait loci (QTL) regions on six chromosomes. Three sets of analyses were conducted to obtain breeding values. The first set of breeding values (EBV-ALL) was obtained using all sources of information (e.g. own DYD and those of all relatives) incorporating the inverse of A_P in the mixed model equations. The second set of breeding values (EBV-PED) was computed as in the first analysis except that the sire's own DYD (but not those of its relatives) was excluded when predicting its breeding value. The third set of breeding values (EBV-MRK) was computed as in the second analysis except that the inverse of A_M was substituted for the inverse of A_P. Correlations of EBVs and of their ranks were computed between EBV-ALL and EBV-PED or EBV-MRK to evaluate the change in accuracy and ranking of sires when A_P was replaced by A_M. Considering all sires without sons in the data set (n = 849), the accuracy of prediction increased by 4.3% for milk yield but did not change for fat and protein yields

when A_P was replaced by A_M (computed across chromosomes). Considering A_M computed within chromosomes, the use of marker information resulted in at least an improvement in accuracy of prediction for milk and protein yield and sometimes a decline in accuracy for fat yield. These results suggest that different A_M may be required for different traits. The rank order correlations were consistently higher across traits when A_M was used, suggesting that use of markers provide better ranking of sires compared to use of pedigree information only. Results from the current study demonstrated that marker information could be used successfully to enhance the accuracy of genetic prediction in routine genetic evaluation particularly for young animals without their own performance or progeny information.

Genetic evaluation has evolved over time as a result of research focusing on different aspects of the prediction process. Results from this study suggest that little will be gained in accuracy of genetic prediction by accounting for inbreeding depression in genetic evaluation at least for the ranges of inbreeding investigated except for CS, HCW, and REA. However, the relationship study demonstrated that using genetic marker information could enhance accuracy of genetic prediction.

Azwihangwisi Maiwashe
Department of Animal Sciences
Colorado State University
Fort Collins, CO 80523
Fall 2004

ACKNOWLEDGEMENTS

This work has involved a lot of dedication and hard work on my part. This, in no way, suggests that this adventure has been an individual effort. Many people contributed directly or indirectly in ensuring successful completion of this work. I hereby wish to extend my sincere appreciation and gratitude to these individuals:

This study would not have been possible without the financial support from the USDA, South African Department of Agriculture and Agricultural Research Council, and Colorado State University Department of Animal Sciences. To the scholarship administrators from the U.S. Dr Child, Mrs. Arlyne Meyers and Mrs. Tana Allshouse and their South African counterpart, Mrs. Loraine Molope, I say thank you very much.

I wish to thank Drs. Golden and Blackburn for initiating the study. Their encouragement is acknowledged. I also thank all members of my graduate work advisory committee for their able guidance and continuous interest throughout the study. To Dr. Garrick, my major adviser, I feel privileged to have worked with the most articulate, patient and intelligent professor. Your simple explanation of complex concepts and your philosophy of learning through practice always made learning an exciting experience. You were always an inspiration through the tough times of graduate school. For all these, I thank you.

To my fellow graduate students, Matt, Scott a.k.a. "Big Boy", Batchimeg, Ellie, Carrie, Brian, Devori, April, Katie, Karen, Jamie, Jessica and Valentin, your

contributions are highly appreciated. We shared the dull and best moments of graduate school. You deserve, at least, a thank you from me.

These acknowledgements will not be complete without mention of some amazing people I met during my stay in the United States. Mrs. Jody Magowitz your friendship changed my view of the world and I am grateful for that. Kevin and Jotham thanks for the friendship.

Lastly and not the least, I wish to thank my family for their love and support through this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION AND OBJECTIVES

1.1. Introduction

1.1.1. Inbreeding

The primary objective of a livestock enterprise is to achieve profit. From an economic standpoint, this objective can be achieved through minimizing the cost of production or increasing the efficiency of the production process. The cost can be reduced by identifying animals of superior genetic merit for traits that are associated with the profit objective of the enterprise. If the accuracy of evaluation and intensity of selection are high, the average performance of the progeny is higher than the parental generation. However, productivity may plateau or decline in the long-term due to factors such as reduced genetic variation and low performance associated with inbreeding depression.

Recent developments in statistical methodology have allowed more accurate genetic evaluation of selection candidates. For example the animal model best linear unbiased prediction (BLUP) is the method of choice for prediction of breeding values in

most of the beef cattle breeds. The BLUP methodology allows optimal utilization of information from an individual and that of its relatives when predicting genetic merit. One of the properties of BLUP is that when the heritability of a trait is low, more weight is placed on information from relatives than own performance when estimating the breeding value (EBV). This leads to high correlations among the genetic merit of sibs since ancestral information plays a predominant role in their evaluation. On the other hand, when the heritability of the trait is high and sibs do not have their own performance information each sibs' EBV is half that of its parents also resulting in high correlations among the EBV of the sibs. When selection is based on these EBV, families rather than individuals are selected. Belonsky and Kennedy (1988) showed using simulation that ten years of selection using BLUP-EBV lead to superior genetic gain and high levels of inbreeding compared to mass selection when the heritability of the trait was low. When the heritability of the trait was high genetic gain and average levels of inbreeding were comparable between the two selection schemes.

Another technology that is of importance in the livestock industry is artificial insemination (AI). It allows sires to leave more progeny than is biologically possible with natural mating. For instance, there are elite Holstein bulls that have sired as many as 250,000 milking daughters worldwide (Wiegel, 2001). It is not a surprise that the estimated effective population sizes for most of the dairy cattle breeds in the U.S. are low; 161, 61, 65, 39, and 30 for Ayrshire, Brown Swiss, Guernsey, Holstein, and Jersey, respectively (Wiegel, 2001).

However, the scenario is slightly different in the beef cattle industry. Whilst BLUP technology is used widely, the utilization of AI has been estimated to be about

13% in the U.S (National Animal Health Monitoring System, 1998), much less than the comparable statistic in dairy industry. Currently, there is paucity of information about the impact of these technologies in the beef industry. Nomura et al. (2001) evaluated changes in the level of inbreeding and effective population size associated with use of BLUP EBV in the population of Japanese Black cattle consisting of 0.53 million cows. They observed a decline in the effective population size from 30 to 17.2 over a period of 12 years. They attributed this decline to intensive use of a few prominent sires.

The lack of interest with regard to inbreeding in the beef industry is mainly due to the structure of the beef cattle industry where many commercial animals are crossbred; thus the effects of inbreeding on the purebreds are offset by crossbreeding in the form of heterosis. So, the impact of high levels of inbreeding in the form of inbreeding depression is limited to the purebreeding sector of the industry. However, reduction in genetic variation due to inbreeding may be cause for long-term concern in the whole industry. It is necessary that the level of inbreeding be assessed in beef cattle. In addition, inbreeding depression should be investigated for traits of economic importance.

1.1.2. Additive genetic relationships

Genetic relationships allow information from relatives to be incorporated in the genetic evaluation process. Traditionally relationships are computed using pedigree information. These genetic relationships are incorporated into the mixed model equations through the inverse of the numerator relationship matrix. One property of the pedigree-based numerator relationship matrix is that the contribution of each individual to the

genetic evaluation of its relative is through its contribution to the evaluation of the parent(s) in common. Thus, the estimated breeding value (EBV) of an animal without its own records is half the genetic merit of the parents. The EBV for young animals without their own performance does not allow for within family selection because sibs that have the same parents will have the same EBV.

Currently the availability of genetic marker information holds promise in improving the accuracy of prediction for young animals without own performance information. When causative mutations responsible for variation for traits of economic importance in livestock are identified, selection can be easily practiced using genotypic information. Given the quantitative nature of most of the traits of importance in livestock, it may take some time before genotypic selection is feasible. Alternatively, genetic marker information can be used to compute additive genetic relationships among animals to be evaluated (Nejati-Javaremi, 1995). Nejati-Javaremi (1995) demonstrated using simulation that genetic marker information could be used successfully to increase the accuracy of genetic prediction. Computation of these relationships does not require knowledge of the location of QTL on the genome. Segments flanked by consecutive markers are used to compute the relationship. However, no empirical study has been conducted to evaluate the efficacy of using genetic markers to compute relationships for routine genetic evaluation.

1.2. Objectives

The objectives of the current study were to evaluate the impact of inbreeding in U.S. registered Red Angus cattle and to investigate the utility of genetic markers in computing the additive genetic relationships used in routine genetic evaluation using data on U.S. registered Holstein sire families. The objectives were achieved as follows:

1. Assess the trend in the level of inbreeding in U. S. Red Angus cattle population. The inbreeding coefficient for each animal was computed using pedigree information supplied by the Red Angus Association of America. Individuals' inbreeding coefficients were averaged by birth year to establish the inbreeding trend.

2. Obtain estimates of inbreeding depression on female reproductive performance and carcass traits. The female reproductive traits evaluated were heifer calving success, heifer calving ease and cow stayability. The carcass traits were hot carcass weight, rib-eye area, backfat thickness, and marbling score.

3. Compare the accuracy of evaluation when the inverses of standard or marker-based numerator relationship matrices were incorporated in the mixed model equations. Marker-based relationships were computed conditional on pedigree and DNA marker information. The type of DNA markers used was the microsatellite marker.

1.3. Dissertation organization

The dissertation is organized in the form of several stand-alone chapters. The literature review was divided into two chapters; literature review on inbreeding and that on additive genetic relationships presented in chapters II and VI, respectively. For ease of presentation, the study on estimates of inbreeding depression was presented separately for female reproduction and carcass traits in chapters IV and V, respectively. The study on the use of genetic markers is presented in two separate chapters; chapter VII on the derivation of weighting factors when the inverse of pedigree or marker-based relationship matrices are used in the mixed model equations and chapter VIII which compares the accuracy of evaluation when pedigree or marker-based relationships are used in predicting breeding values. The *awk* programming language scripts and java computer programs created for preparation and analyses of the data are given in the appendix at the end of the dissertation.

1.4. Literature Cited

Belonsky, G. M., and B. W. Kennedy. 1988. Selection on individual phenotype and best linear unbiased predictor of breeding values in a closed swine herd. J. Anim. Sci. 66:1124-1131.

National Animal Health Monitoring System. 1998. Part IV. Changes in the U.S. beef cow-calf industry, 1993-1997.

Nejati-Javaremi, A. 1995. Alternative methods for defining relationship, assigning haplotypes and measuring linkage in animal breeding. Ph.D. Dissertation. University of Guelph.

Nomura, T., T. Honda, and F. Mukai. 2001. Inbreeding and effective population size of Japanese Black cattle. J. Anim. Sci. 79:366-370.

Weigel, K. A. 2001. Controlling inbreeding in modern breeding programs. J. Dairy Sci. 84 (E. Suppl.): E177-E184.

CHAPTER II

LITERATURE REVIEW

INBREEDING

2.1. Introduction

Wright (1922a) analyzed data from long-term experiments conducted to investigate the effects of inbreeding and crossbreeding in guinea pigs. Inbreeding resulted in decline in performance in all the traits considered (i.e. percentage born alive, percentage raised of all young born alive, birth weight, daily gains and weaning weight). Subsequent outcrossing of the inbred lines led to individuals with performance superior to that of the control stock. In view of these results, many long-term inbreeding experiments were initiated in beef cattle in the early 1930's in the midwestern states of the U.S. (Brinks and Knapp, 1975; Brinks, 1975). A total of 48 inbred lines were initiated. The idea was to create inbred lines that would be ultimately crossed to exploit genetic differences among the highly inbred lines. Development, breeding and management practiced within these inbred lines is described elsewhere (Urick et al., 1966; Brinks, 1975). Some of the lines were continued until completion of the

experiments while others were discontinued due to low performance, genetic abnormality or lack of facilities (Brinks, 1975).

The main objective of this chapter is to review results from these inbreeding experiments. Theoretical aspects about the process of inbreeding will be considered. Concepts such as the inbreeding coefficient, inbreeding depression, and effective population size are discussed. Changes in the mean and variance of a trait due to inbreeding are also addressed. General conclusions drawn from these studies are summarized at the end of the chapter.

2.2. Inbreeding

Inbreeding refers to the mating together of individuals that are more closely related than would be the case if mating was at random (Falconer and Mackay, 1996). It is of importance in population genetics for two reasons; increases in inbreeding are associated with 1) a decline in performance in attributes related to the fitness of an individual such as survival, fertility and vigor, and 2) reduction in additive genetic variation. The degree of inbreeding in a population is a function of the size of the population and the specified breeding structure. In populations of infinite size where breeding individuals mate at random and evolutionary forces such as mutation, migration and selection are unimportant, the genetic properties of the population remains the same from one generation to the next. Moreover, due to large population size, the likelihood of relatives mating at random is low and inbreeding is of no consequence. On the contrary, in populations of finite sizes the gene frequency fluctuates randomly from generation to

generation due to finite sampling of gametes (Caballero, 1994). The fluctuation in gene frequency from random sampling of gametes is termed random genetic drift (Falconer and Mackay, 1996). The genetic outcome of drift and inbreeding are the same as pointed out by Falconer and Mackay (1996) "inbreeding and drift are not different consequences but different ways in which consequences may be seen".

2.3. Inbreeding coefficient

The process of inbreeding can be measured by the inbreeding coefficient (Wright, 1922b). The inbreeding coefficient of a population can be computed based on the number of breeding individuals or using pedigree information. The two methods are described below.

2.3.1 Inbreeding coefficient based on population size

The inbreeding coefficient was deduced by hypothesizing an idealized population as shown in Figure 2.1 (Wright, 1931; Fisher, 1930).

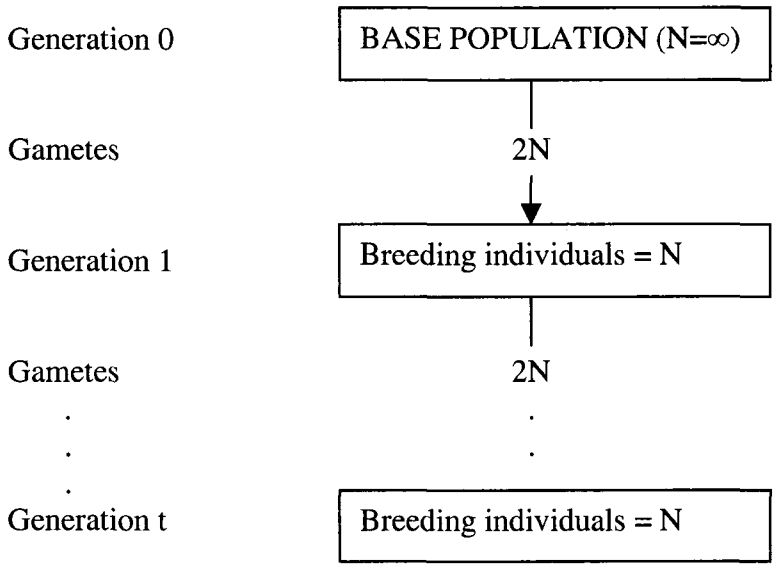


Figure 2.1. Schematic representation of the idealized population, N is the number of breeding individuals and t referes to the generation number. Adapted from Falconer and Mackay (1996).

An idealized population is a conception of a monoecious (individuals capable of self-fertilization) population with constant size over discrete generations where mating occurs at random including self-fertilization. Each individual has an equal probability of contributing gametes to the next generation and mutation and selection are ignored (Falconer and Mackay, 1996). Given the idealized population, the inbreeding coefficient between successive generations can be expressed as follows:

$$F_t = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{t-1} \qquad [2.1]$$

where F_t and (F_{t-1}) are the inbreeding coefficients at generations t and (t-1), respectively and N is the number of breeding individuals or population size. The first term in the right hand side of the above equation represents the new inbreeding or rate of inbreeding and the second term is a remainder due to previous inbreeding. The increment or rate of inbreeding can be represented as follows:

$$\Delta F = \frac{1}{2N} \qquad [2.2]$$

Since the population size (N) is constant over time, it follows that the ΔF will also be constant and thus inbreeding at any particular time can be computed as follows:

$$F_t = 1 - (1 - \Delta F)^t \qquad [2.3]$$

where F_t is the average inbreeding of the population in generation t.

In practice, not all individuals contribute progeny to the next generation due to reasons such as age, survivability or fertility. Therefore, the number of breeding individuals is not necessarily equal to N. In this case the census population size (N) will lead to an under-estimation of the level of inbreeding in the population. To overcome this problem Wright (1931) introduced the concept of effective population size, which is the size of an idealized population that would give rise to the same rate of inbreeding as the population under consideration (Falconer and Mackay, 1996). Thus, the N in [2.2] can be replaced by the effective population size (N_e), resulting in

$$\Delta F = \frac{1}{2N_e} \qquad [2.4]$$

Rearranging [2.4] leads to

$$N_e = \frac{1}{2\Delta F} \qquad [2.5]$$

The formulae presented above have limitations due to deviations from the conditions of the idealized population that are common in practice. Thus, the approximations of the effective population size will differ from one scenario to the next. The approximations developed by Wright (1938) are as follows:

1) Unequal number of breeding males and females

$$N_e = \frac{4N_m N_f}{N_m + N_f} \qquad [2.6]$$

where N_m and N_f is the number of males and females, respectively.

2) Different population sizes over generations

$$\frac{1}{N_e} = \frac{1}{t}\left[\frac{1}{N_{e,1}} + \frac{1}{N_{e,2}} + \frac{1}{N_{e,3}} + \ldots + \frac{1}{N_{e,t}}\right] \qquad [2.7]$$

3) Different family sizes (unequal progeny contribution)

$$N_e = \frac{8N}{V_{km} + V_{kf} + 4} \qquad [2.8]$$

where V_{km}, V_{kf} are the variances in family size of males and females, respectively.

2.3.2 Inbreeding coefficient based on pedigree information

Consideration of the pedigree information in the computation of the inbreeding coefficient started in the early 1900s as noted by Pearl (1913) that "there seems not to have been worked out any adequate general method of measuring quantitatively the degree of inbreeding which is exhibited in a particular pedigree". His series of papers on calculation of inbreeding using pedigree information laid the groundwork for development of the inbreeding coefficient (Pearl, 1913; Pearl, 1914a,b; Pearl, 1917a,b). He coined the term inbreeding coefficient and proposed the following formula for the computation of the inbreeding coefficient of the progeny in generation n due to the mating of its parents in generation $n-1$:

$$Z_n = \frac{100(p_{n-1} - q_{n-1})}{p_{n-1}} \qquad [2.9]$$

where p_{n-1} and q_{n-1} are the maximum possible and actual number of different individuals in generation $n-1$ of the progeny's pedigree and Z_n is the inbreeding coefficient of a particular individual. The total number of Z_n's computed is equal to the number of generations all the way to the "base generation". The principle involved in the derivation of Z_n is that p_{n-1} will take the values 2,4,16,..,2^n representing number of parents, grand-parents, great grand-parents,…,ancestors in generation *n*. Clearly, q_{n-1} is likely to be smaller that p_{n-1} for inbred compared to non-inbred individuals.

As pointed out by Wright (1922b), the inbreeding coefficient of Pearl was not unique to each pedigree structure. For example, Pearl's inbreeding coefficient for an individual produced by continuous mating of double first cousins was the same as that of an individual produced by crossing different lines, where each line was created by brother-sister mating. In this scenario, the inbreeding coefficient is not useful as a measure of the physiological impact of inbreeding on animal performance as the crossbred individual is likely to be more vigorous compared to the inbred individual.

Wright (1922b) made the most significant contribution to population genetics through his inbreeding coefficient. Though there are similarities in the logic behind Wright and Pearl's inbreeding coefficient, Wright's inbreeding coefficient was unique to the particular pedigree structure. He defined the inbreeding coefficient of an individual as the correlation between uniting gametes from the individual's parents. The inbreeding coefficient of individual X, denoted F_X, for any autosomal diploid locus relative to an arbitrary ancestral generation is

$$F_X = \sum_{a=1}^{j} \left(\frac{1}{2}\right)^{n_s+n_d+1} (1+F_a) \qquad [2.10]$$

where the summation is over all the j common ancestors of individual X; n_s and n_d are the number of generation from sire and dam to the a^{th} common ancestor, respectively; F_a is the inbreeding coefficient of the a^{th} ancestor. The F_X is independent of gene frequency and thus, represents the effects of inbreeding on all autosomal loci in the absence of selection and mutation.

Malécot (1948) derived a formula for F_X equivalent to [2.10] using probability theory and defined the inbreeding coefficient as the probability that homologous alleles at a locus in an individual are identical by descent. He expressed the inbreeding coefficient of an individual as the coefficient of coancestry of its parents, which he defined as the probability that the two alleles from the same locus chosen randomly, one from the sire and the other from the dam, are identical by descent. This probability is equivalent to the correlation between uniting gametes, hence the equivalence of [2.10] to Malécot's inbreeding coefficient.

Both Wright and Malécot assumed a reference population comprised of individuals that are unrelated. Thus, the inbreeding coefficient is expressed as a deviation from this reference population and is a measure of the amount of homozygosity that has accrued since the reference population (Wright, 1922b). From a biological point of view, inbreeding coefficient is a measure of the quality of an individual relative to the foundation population (Wright and McPhee, 1925).

For identity by descent to hold as a measure of homozygosity due to inbreeding, all other forces that change the gene frequency are assumed nonsignificant. This is valid

when an infinitesimal model is assumed since selection, for example, will not influence gene frequency. That is, expected inbreeding coefficient is an unbiased estimate of observed inbreeding coefficient. However, under a finite model, selection may be a significant force in changing gene frequency and thus pedigree-based inbreeding coefficient may not be useful as a measure of homozygosity. Groen et al. (1995) has shown through simulation assuming a finite locus model that inbreeding coefficients based on pedigree information were biased compared to those derived using genotypic frequencies at loci under selection.

2.4. Genotypic frequencies

In a population in Hardy-Weinberg equilibrium, gene frequencies are expected to remain constant from one generation to the next. In addition, the genotypic frequencies can be determined from the gene frequencies. For a single locus with two alleles, *A* and *a*, at frequencies *p* and *q* respectively, the genotypic frequencies at this locus can be represented as follows:

$$p^2\ AA : 2pq\ Aa : q^2\ aa$$

When the population under consideration is of finite size and hence inbreeding occurs, there are more homozygotes than expected under Hardy-Weinberg conditions. The effect of inbreeding on genotypic frequencies is accounted for as follows:

$$\left(p^2 + pqF\right) AA : \left(2pq - 2pqF\right) Aa : \left(q^2 + pqF\right) aa$$

where F is the degree of inbreeding of the population at a particular time. The second term in each of the above expressions represents the change in genotypic frequencies due to inbreeding. For instance, if there is no inbreeding in the population ($F = 0$) and other forces that change the genotypic frequency are negligible, the genotypic frequencies reduce to those in the population in equilibrium. On the other hand, under complete inbreeding ($F = 1$) the entire population will be comprised of one or other of the two homozygotes and the population is considered fixed at this particular locus.

2.5. Effect of inbreeding on mean and variance of a trait

In the preceding section it has been shown that under inbreeding, genotypic frequencies deviate from those in a population in Hardy-Weinberg equilibrium. This deviation may lead to a change in the mean and variance in an inbred population as observed for egg-to-pupa viability in *Drosophila melanogaster* (López-Fanjul and Villaverde, 1989). Prediction equations for the mean and variance in inbred populations are available (Wright, 1951). The derivations were carried out by assuming a population divided into isolated subpopulation of small sizes where mating within subpopulations occurs at random. Furthermore, it was assumed that all active evolutionary forces such as selection and migration were absent and the average inbreeding coefficient in the entire population at generation t was $\overline{F}_t$. For a single locus with two alleles, A and a, at average

frequencies, $\overline{p}_t$ and $\overline{q}_t$, respectively, and genotypic values $+a$, d and $-a$ for genotypes AA, Aa, and aa respectively, the mean in generation t is given by (Falconer and Mackay, 1996)

$$\begin{aligned}\overline{Y}_t &= a(\overline{p}_t - \overline{q}_t) + 2d\overline{p}_t\overline{q}_t(1-\overline{F}_t) \\ &= \overline{Y}_0 - 2d\overline{p}_t\overline{q}_t\overline{F}_t\end{aligned} \qquad [2.11]$$

where $\overline{Y}_0$ is the population mean at generation zero (i.e. base population).

It is clear from [2.11] that for the population mean to change with inbreeding, d should not equal zero. It then follows that under an additive genetic model the mean is independent of the level of inbreeding since $d = 0$. Therefore, the mean will only change if the locus under consideration exhibits dominant gene action.

An extension of [2.11] to multiple loci assuming no epistasis leads to

$$\overline{Y}_t = \overline{Y}_0 - 2\overline{F}_t\Sigma d\overline{p}_t\overline{q}_t \qquad [2.12]$$

Equation [2.12] shows that the presence of dominance does not necessarily result in the change of the mean under inbreeding as implied above. For the mean to change under inbreeding, directional dominance is required otherwise the effects from different loci might cancel out.

When the loci influencing the trait exhibit additive gene action within and between loci, change in variance due to inbreeding (in the absence of selection and

mutation) is predictable (Wright, 1951, 1952). The total variance in a population (in a steady state i.e. $F_{t+1} \sim F_t$) divided into lines or strains is redistributed into the variance within and between the strains as shown below.

$$\begin{aligned} \sigma_W^2 &= (1-F)\sigma_O^2 \\ \sigma_B^2 &= 2F\sigma_O^2 \\ \sigma_T^2 &= (1+F)\sigma_O^2 \end{aligned} \qquad [2.13]$$

where F is the inbreeding of the entire population at time t and σ^2_O is the additive genetic variance in the base population; σ^2_T, σ^2_B and σ^2_W are the total, between- and within-strain(s) variances at time t, respectively.

Equations [2.13] show that as inbreeding increases the variance within the strain declines while that between strains increases. When all the loci influencing the trait are fixed, the total variance is double the original variance in the base population and the total variance comes from the differences between strains. As the strains become highly inbred, they become genetically distinct. For traits that are moderate to highly heritable, predictions from [2.13] are in agreement with empirical evidence (Whitlock and Flower, 1999; Wade et al., 1996).

The derivation of [2.13] refers strictly to variance arising from loci exhibiting additive gene action. When some loci influencing a trait are under dominant gene action, the total variance in the population can no longer be decomposed into between- and within-strain(s) variances (Wright, 1952). A theoretical investigation by Robertson (1952) has shown that the variance also depends on the contribution from loci exhibiting

dominance; with their contribution dependent on the gene frequency in the base population. He considered a situation of recessive alleles at low frequency in the base population and found that as inbreeding increases, within population variance initially increased peaking just before inbreeding was 0.5 and gradually decreased until it was zero at complete inbreeding. This result was consistent with that obtained by López-Fanjul and Villaverde (1989) who observed a superior response in egg-to-pupa viability in *Drosophila melanogaster* from within-line selection in inbred compared to non-inbred lines. They attributed high response in inbred lines to initial increase in additive genetic variance. Wang et al. (1998) concluded from a theoretical investigation that dominance is the main cause for the increased genetic variance in fitness-related traits after a population bottleneck or inbreeding.

2.6. Inbreeding depression

Inbreeding depression refers to a reduction in performance in traits related to fitness in inbred compared to outcrossed offspring (Falconer and Mackay, 1996). There are two competing hypotheses for the genetic basis of inbreeding depression: the partial dominance hypothesis which supposes that inbreeding depression is a result of homozygosity of recessive or partially recessive alleles, and the overdominance hypothesis which argues that the fitness of heterozygotes exceeds that of both homozygotes such that as inbreeding increases, the proportion of homozygotes increase resulting in decline in the fitness of the population. Wright seem to subscribe to the dominance hypothesis as can be gathered from a quote in his 1922 paper "The best

explanation of the decrease in vigor is dependent on the view that Mendelian factors unfavorable to vigor in any respect are more frequently recessive than dominant, a situation which is the logical consequence of the two propositions that mutations are more likely to injure than improve the complex adjustments within an organism and that injurious dominant mutations will be relatively promptly weeded out, leaving the recessive ones to accumulate, especially if they happen to be linked with favorable dominant factors".

These two hypotheses are in agreement in explaining lower fitness of inbred individuals. However, they differ in the possibility of purging inbreeding depression from breeding populations (Kärkkäinen et al., 1999). For example, inbreeding depression due to partially recessive alleles can be purged by selection while that caused by overdominance cannot be removed without lowering the mean fitness of the population (Dudash and Carr, 1998; Wang et al., 1999; Wang, 2000). Wang et al. (1999) demonstrated through simulation that purging is most effective when recessive alleles are lethal and less effective when they have a small effect on viability.

Also, crossing of inbred lines should result in performance similar to that of the outbreds under the overdominance hypothesis while it is expected under the dominance hypothesis that performance of progeny from crossed inbred lines will exceed that of the outbred progeny (Roff, 2002). The distinction between the two hypotheses can also be made based on the performance of the inbred lines relative to that of parental stock under certain conditions (Falconer and Mackay, 1996). Under the partial dominance hypothesis, it is possible for some of the inbred lines to outperform the outbred base population. On the contrary, no inbred line can perform better than the parental stock when the

overdominance locus is at its equilibrium gene frequency in the base population (Hill and Robertson, 1968). However, Minvielle, (1979) demonstrated theoretically that overdominance can produce inbred lines that perform as good or even better than the base population under certain conditions e.g. when the gene frequency of the better homozygote is below its equilibrium frequency in the base population.

Empirical evidence in support of both hypotheses is available. Dudash and Carr (1998) investigated the genetic basis of inbreeding depression for life-history traits (e.g., total flower production, above ground biomass, ovule production, pollen viability and pollen production) in annual plants, selfing *Mimulus micranthus* and the mixed-mating *M. guttatus*. They observed that inbreeding depression for life-history traits was due to partially recessive deleterious alleles for all traits but pollen production, where overdominance loci were responsible for inbreeding depression. Kärkkäinen et al. (1999) obtained results in support of both hypotheses in a self-incompatible herb *Arabi petraea.* Roff (2002) crossed seven inbred lines of sand cricket, *Gryllus firmus* and measured nymphal weights at different ages and early fecundity in the linecrosses and an outbred population. The linecrosses outperformed the outcross in support of the partial dominance hypothesis. In a review of the genetic basis of inbreeding depression, Charlesworth and Charlesworth (1999) concluded that inbreeding depression in most fitness-related traits is due to partial recessivity of deleterious alleles maintained in the population through mutation-selection balance, in support of the dominance rather than overdominance hypothesis. However, the importance of the overdominance hypothesis was not ruled out for other traits.

2.7. Inbreeding and selection

When inbreeding is practiced in a finite population assuming no selection, mutation and migration, the only evolutionary force affecting the genetic properties of the population is random genetic drift – the random sampling of individuals that become parents of the next generation. Though migration and mutation can be safely ignored under controlled breeding, selection is a force to be reckoned with. For example, most of the inbreeding experiments in beef cattle have been conducted concurrently with selection (Brinks, 1971; Clarke, 1988; MacNeil et al., 1989; Pariacote et al., 1998). When inbreeding occurs simultaneously with selection, two forces are involved, random genetic drift and selection i.e. the change in gene frequency at loci under selection is determined by the interaction between the two forces (Hill and Robertson, 1968).

The interaction between inbreeding and selection has been studied extensively for loci exhibiting heterozygote advantage (Hayman and Mather, 1953; Reeve, 1955; Hill and Robertson, 1968; Minvielle, 1979). Results from these studies show that selection may delay inbreeding depression. According to Hill and Robertson (1968) the delay is achieved in two ways: 1) when the equilibrium gene frequency is 0.5, selection reduces inbreeding depression by maintaining high levels of heterozygosity, and 2) when the equilibrium frequency is 0.1, selection reduces heterozygosity but also reduces inbreeding depression by reducing the probability of fixation of the poor homozygote.

According to Wang et al. (1999) the outcome of the joint effect of inbreeding and selection depends on, 1) the nature of mutational load i.e. the mutation rate and the distributions of mutant selective effects and dominance coefficients, 2) the rate of

inbreeding, 3) the reproductive capacity, and 4) the organization of the genome (number of chromosomes and their map lengths) of the species. They recommended that in order to eliminate recessive mutations from a population, inbreeding should be followed by strong selection because weak selection will allow fixation of some of the recessives.

2.8. Empirical results of inbreeding depression in beef cattle

The effect of inbreeding on performance in beef cattle has been evaluated under varied conditions (e.g. management and breeding policy, selection, rate of inbreeding) and thus comparison of the results from different studies is difficult. Also, response to inbreeding might be confounded with the environment (Keller and Brinks, 1978a,b) and in some cases selection could overcome inbreeding depression (Pariacote et al., 1998). Nevertheless, results from different studies provide indications about the magnitude of inbreeding depression in beef cattle.

Studies on the effect of inbreeding in beef cattle mainly focused on traits such as fertility, survival and growth. For ease of presentation, comparisons of different studies are given under two broad classifications: reproduction (including survival) and growth. Estimates of the effect of inbreeding (partial regressions) on traits under consideration are summarized in Tables 2.1 through 2.7. In addition, information on the population (e.g. number of records and average level of inbreeding) from which the estimates were obtained is also provided. Regression coefficients given in the tables are expressed using the same unit of measurement for each trait and therefore might be different from those reported in the original studies since, in some cases, different units were used for the

same trait in different studies (e.g. weaning weight expressed in pounds and kilograms in different studies).

2.8.1. Reproductive traits

Prenatal survival. MacNeil et al. (1989) evaluated the effect of inbreeding of the dam on prenatal survival of the noninbred fetus in heifers bred to calve as 2-year olds. Four groups (inbred, linecross, topcross and a control) of heifers were considered. Average inbreeding levels of the heifers were 26.5, 6.9, 0.0 and 0.0% for inbreds, controls, linecross and topcross, respectively. Linecross heifers were reciprocal crosses of inbred lines and topcross heifers were produced by mating inbred bulls to control line cows. There was a significant difference in the survival of the fetus between the inbred and control group reflecting the difference in average inbreeding between the two groups. Fetuses from heifers in the inbred group were 15.2% less likely to survive until calving relative to those from the control group and within the inbred and control lines, an increase in inbreeding of the dam was associated with a reduction (-1.67%/%) in fetal survival indicating that a percentage increase in inbreeding resulted in 1.67% less calves born alive (Table 2.1). On the contrary, Brinks and Knapp (1975) reported no increase in the percentage of calves born dead due to inbreeding of the dam but inbreeding of the fetus significantly increased (0.08%/%) the percentage of calves born dead. However, the effect of inbreeding in this study might have not been accurately estimated because inbreeding of the fetus and dam may be confounded.

Table 2.1. Effect of inbreeding on pre- (PRS) and postnatal (POS) survival.

Trait evaluated	Population (No. of lines)	Years studied	No. Records	Average inbreeding (%)		Partial regression coefficient (%/%)		Source
				Fetus/Calf	Dam	F_f / F_c	F_d	
PRS								
• % survived	SDAES Hereford (5)	1971-78	130	0	27/7[a]		-1.67	MacNeil et al. (1989)
• % dead	Lines from different stations in U.S. (48)					0.08	0.03	Brinks and Knapp (1975)
POS								
• % survived	SDAES Hereford (5)	1971-78	130	0	27/7[a]		-0.57	MacNeil et al. (1989)
	MSFL at Reno Hereford (2 inbred lines and linecross progeny)	1969-73	314	~13	8		-0.36	Bailey et al. (1977)
• % dead	Lines from 8 stations in U.S. (44)					0.08	0.07	Brinks (1971)
	Lines from different stations in U.S. (48)					0.08	0.07	Brinks and Knapp (1975)

SDAES - South Dakota Agricultural Experiment Station beef cattle breeding project, MSFL - Main Station Field Laboratory, F_f - inbreeding of the fetus, F_c - inbreeding of the calf, F_d - inbreeding of the dam. [a] Inbreeding of the inbred and control groups.

Calving rate. Dinkel et al. (1972) assessed calving rate of four one-sire inbred lines and a single four-sire control line where heifers were bred to calve at 3-years of age. Fertility was defined as the presence or absence of a calf at birth irrespective of the condition of the calf. The difference in calving rate between the control and inbred groups was significant with cows in the control group averaging 8% more calves than those in the inbred group. However, the partial regression of calving rate on inbreeding of the fetus and dam were small and nonsignificant, 0.03%/% and 0.15%/%, respectively (Table 2.2). This result might have arisen from the fact that regression coefficients were computed within small subclasses of year-age of cow-line. In 2-year old heifers originating from the same population considered by Dinkel et al. (1972), MacNeil et al. (1989) observed that heifers from the control group had 11% more calves at birth than those from the inbred group but the difference was not significant. The dams were inbred and fetus not inbred in this study. However, there was a significant decline (-1.17 %/%) in calving rate associated with the inbreeding of the dam within the inbred and control lines.

Krehbiel et al. (1969) found no difference in calving rate of Shorthorn inbred, outbred and selection lines but the outbred group outperformed the inbred and selection lines in Angus cattle. There was a tendency for a decline (-0.2 vs. -0.4 %/%) in calving rate of both Shorthorn and Angus as inbreeding of the dam increased. The magnitude of the effect of inbreeding of the calf was similar in both breeds but opposite in direction with a tendency for a decline in Angus and an increase in Shorthorn. Bailey et al. (1977) also observed a tendency for a decline in calving rate with inbreeding of the dam in

Table 2.2. Effect of inbreeding on calving rate (CR).

Trait evaluated	Population (No. of lines)	Years studied	No. Records	Average inbreeding (%)		Partial regression coefficient (%/%)		Source
				Fetus	Dam	F_f	F_d	
CR								
• % born alive	SDAES Hereford (5)	1971-78	182	0	27/7[a]		-1.17	MacNeil et al. (1989)
	SDAES Hereford (8)	1956-68	701	~23	~13	0.03	0.15	Dinkel et al. (1972)
	MSFL at Reno Hereford	1969-73	403	~13	8		-0.05	Bailey et al. (1977)
• % born alive or stillborn	FRBCRS Shorthorn and Angus (18)	1950-62	741	~13	~4	Angus -0.40	-0.20	Krehbiel et al. (1969) ,,
			672	~20	~14	Shorthorn 0.40	-0.40	,,

SDAES - South Dakota Agricultural Experiment Station beef cattle breeding project, MSFL - Main Station Field Laboratory, FRBCRS – Front Royal Beef Cattle Research Station, F_f - inbreeding of the fetus, F_d - inbreeding of the dam. [a] Average inbreeding of dams from the inbred and control lines.

linecross Hereford calves. Davenport et al. (1965) compared the percentage calf crop between inbred and linecross groups of cows of ages ranging from 2 to 10 years and reported a significant difference in the calf crop between the two groups. However, when heifer cows (2-year olds) were excluded from the analysis, the difference between the two groups disappeared. It was concluded that the effect of inbreeding on fertility was more pronounced in heifers than in older cows. In general, results from these studies suggest a decline in calving rate with an increase in inbreeding particularly that of the dam. Furthermore, differences in the regression coefficient emphasize the fact that the effect of inbreeding is unique to the population.

Survival from birth to weaning. Stonaker (1954) reported a 30% difference in the number of calves raised to weaning between outbred and inbred matings in favor of outbred matings. Likewise, Bailey et al. (1977) reported a significant decline (-0.36%/%) in survival to weaning associated with inbreeding of the dam (Table 2.1). Calves in this study were inbred but inbreeding of the calf was not included in the model. Therefore, the regression on inbreeding of the dam represents the overall effect of inbreeding of the calf and dam. In contrast, MacNeil et al (1989) observed a nonsignificant reduction (-0.57%/%) in the percentage of calves surviving to weaning due to inbreeding of the dam. This result is suggestive of the unfavorable effects of inbreeding of the dam on calf survival to weaning. Brinks (1971), in an extensive study of 44 inbred lines from eight stations in the U.S., also reported a nonsignificant increase (0.071%/%) in percentage of calves dead at weaning associated with inbreeding of the dam. However, inbreeding of the calf significantly increased (0.079%/%) the percentage of calves dead at weaning. Brinks and Knapp (1975), in a study involving 48 inbred lines, found evidence for an

increase and decline in the percentage of calf surviving to weaning in different lines but generally inbreeding of the calf and dam were unfavorable across different lines (i.e. 56% and 63% of the regression coefficients on inbreeding of the calf and dam, respectively, were unfavorable). Results from these studies show that there is evidence for and against the detrimental effect of inbreeding of the dam on calf survival to weaning.

2.8.2. Growth traits

Birth weight. Alexander and Bogart (1961) evaluated the effects of inbreeding on birth weight in three closed lines of Herefords (Lionheart, Prince and David) and a single line of Aberdeen Angus. Most of the calves (43%) had inbreeding coefficients between 10 and 20%. The herds had been selected on an index based on suckling gains, gains during feed test, feed consumption per unit gain and score for type and conformation. Neither the inbreeding of the calf or the dam significantly affected birth weight. However, a tendency for birth weight to decrease (-0.04kg/%) and increase (0.01kg/%) with inbreeding the calf and dam, respectively, was observed (Table 2.3). Swiger et al. (1961) reported similar results on inbred lines located at Fort Robinson Nebraska though the size of the effect of inbreeding of the dam was about five times larger than that obtained by Alexander and Bogart (1961). Similarly, Nelms and Stratton (1967) found a nonsignificant effect of inbreeding of either the calf or the dam on birth weight. MacNeil et al. (1989) investigated the effect of inbreeding of the dam on birth weight in noninbred calves and observed a nonsignificant reduction (-0.05kg/%) in birth weight due to inbreeding.

In contrast to the previous studies where no effect of inbreeding was found, significant effect of inbreeding of the calf and dam has been reported in several studies (Swiger et al., 1961; Brinks et al., 1963; Anderson et al., 1972; Willis and Wilson, 1974; Clarke, 1988; MacNeil et al., 1992; Snelling et al., 1996; Pariacote et al., 1998). Swiger et al. (1961) considered data from inbred lines located at Lincoln Nebraska and reported a significant reduction (-0.17kg/%) in birth weight associated with the inbreeding of the calf. MacNeil et al. (1992) and Snelling et al. (1996) analyzed subsets of data from the same inbred line and reported a significant reduction in birth weight due to inbreeding of the calf (-0.043 vs. -0.084kg/%). Likewise, Pariacote et al. (1998) found that each percentage increase in inbreeding of the calf resulted in a 0.058kg decrease in birth weight. Willis and Wilson (1974) assessed the effects of inbreeding on birth weight in Santa Gertrudis cattle in the Oriente Province of Cuba. Only inbreeding of the calf was considered because most of the cows were minimally inbred - less than 1%. A significant decline (-0.12 kg/%) in birth weight as the inbreeding of the calf increased was observed.

In the preceding paragraphs studies where response to inbreeding was assumed to be similar across sex of the calf were considered. The following paragraphs focuses on studies where the effect of inbreeding was evaluated within sex. Brinks et al. (1963) studied the effects of inbreeding of the calf and dam on birth weight of bull and heifers calves in Line 1 Hereford cattle. The average inbreeding of the calf and dam were essentially the same for bull and heifer calves at 16 and 11% respectively. An increase in inbreeding of the calf was associated with a reduction in its birth weight. However, the response to inbreeding was variable between the two sexes; the depression was more pronounced in heifer calves (-0.181kg/%) than bull calves (-0.059kg/%). The effect of

inbreeding of the dam was similar to that observed by Alexander and Bogart (1961) though the impact was five times more in heifer calves (0.045kg/%) than in bull calves (0.005kg/%). Opposite estimates of the effect of inbreeding of the dam on birth weight have been reported by Anderson et al. (1972) on their study of three inbred lines of Hereford cattle located at Northern Agricultural Research Center, Havre, Montana. They found a highly significant positive effect (0.168kg/%) of inbreeding of the dam on birth weight of bull calves and a nonsignificant effect (0.057kg/%) on heifer calves. The positive effect of inbreeding of the dam on birth weight was thought to be due to selection pressure exceeding the depressing effect of inbreeding of the dam (Anderson et al., 1972). Contrary to Brinks et al. (1963), Clarke (1988) reported more severe effects of inbreeding of calf on bull than heifer calves in a study involving 23 inbred lines of Hereford cattle from San Juan Basin Research Center, Hesperus Colorado. Partial regressions of birth weight on inbreeding of calf were -0.04kg/% for bull calves and -0.02kg/% for heifer calves. Inbreeding of the dam was more detrimental on heifer calves (-0.03kg/%) than bull calves (-0.01kg/%).

From Table 2.3, 88% and 69% of the regressions on inbreeding of the calf and dam, respectively, were negative. These figures suggest that, in general, inbreeding of the calf and dam has a detrimental effect on birth weight. However, the effect of inbreeding of the dam seems to be much less than that of the calf.

Table 2.3. Effect of inbreeding on birth weight (kg/%).

Population	Breed	Years studied	No. Records	Average Inbreeding (%)		Partial regression coefficient				Source
						F_c		F_d		
				Calf	Dam	BC/HC[a]	HC	BC/HC[a]	HC	
OAES, Corvallis OR	H, A	1952-57	280	~15		-0.041		0.014		Alexander and Bogart (1961)
FRBCRS, Crawford NE	H, A, S	1951-55	677	13	10	-0.027		0.059		Swiger et al. (1961)
NAES, Lincoln NE	H, A	1951-55	283	5	3	-0.172		-0.014		Swiger et al. (1961)
USRLES, Miles City MO	H	1934-59	2,027	~16	~11	-0.059	-0.181	0.005	0.045	Brinks et al. (1963)
UWS, Gillette WY	H	1953-64	302	11	5	-0.015		-0.026		Nelms and Stratton (1967)
NARC, Havre MO	H	1950-63	650	~13	~7	0.009	-0.025	0.168	0.057	Anderson et al. (1972)
OPC	SG	1956-70	568	3		-0.116				Willis and Wilson (1974)
SJBRC, Hesperus CO	H	1970-86	2,885	36/4[b]		-0.040	-0.020	-0.010	-0.030	Clarke (1988)
SDAES	H	1971-78	576	0	27			-0.050		MacNeil et al. (1989)
USRLES, Miles City MO	H	1935-89	4,716	~20	~20	-0.043		-0.006		MacNeil et al. (1992)
USRLES, Miles City MO	H	1935-91	5,346			-0.084		-0.018		Snelling et al. (1996)
NARC	H	1964-91	1,648			-0.069		-0.025		Snelling et al. (1996)
CHR	H	1978-91	1,166			0.010		-0.012		Snelling et al. (1996)
HH	H	1979-91	1,428			-0.114		-0.090		Snelling et al. (1996)
USRLES, Miles City MO	H	1931-75	8,065	10	8	-0.058		-0.047		Pariacote et al. (1998)

OAES - Oregon Agricultural Experiment Station, NARC - Northern Agricultural Research Center, UWS - University of Wyoming Substation, USRLES - United States Range Livestock Experiment Station, SJBRC - San Juan Basin Research Center, FRBCRS - Fort Robinson Beef Cattle Research Station, NAES - Nebraska Agricultural Experiment Station, CHR - Cooper Hereford Ranch, HH - Holden Herefords, SDAES - South Dakota Agricultural Experiment Station, OPC - Oriente Province of Cuba, H - Hereford, A - Angus, S - Shorthorn, SG - Santa Gertrudis, BC - bull calf, HC - heifer calf, F_c and F_d - inbreeding of calf and dam. [a] regression of birth weight on inbreeding for bull calves for within sex analysis or for bull and heifer calves together for across sex analysis. [b]Average inbreeding level of the inbred group and linecross calves.

Preweaning daily gain. A study by Brinks et al. (1963) in an inbred line of Herefords showed a reduction in preweaning gain as a result of inbreeding of either the calf or the dam (Table 4). The effect of inbreeding of the calf was more pronounced on heifer calves (-4.3g/d/%) than bull calves (-1.2g/d/%). In contrast, inbreeding of the dam was more detrimental to bull calves (-4.8g/d/%) compared to heifer calves (-1.1g/%). Clarke (1988) also found evidence of detrimental effects of inbreeding of the calf and dam on preweaning average daily gain. Contrary to Brinks et al. (1963) the response to inbreeding of the calf and dam was similar for bull and heifer calves. Similarly, Alexander and Bogart (1961) observed inbreeding depression of preweaning daily gain associated with inbreeding of the calf. Each percentage increase in inbreeding of the calf resulted in 4.5g reduction in preweaning daily gain. Animals in this study were selected on an index including preweaning daily gain and thus the inbreeding depression observed suggests the ineffectiveness of selection in increasing the frequency of genes for preweaning growth (Alexander and Bogart, 1961). Pariacote et al. (1998) also found significant effects of inbreeding of the calf and dam though inbreeding of the dam was more important than that of the calf. MacNeil et al. (1992) reported no evidence of inbreeding depression on preweaning daily gain in Line 1 Hereford. However, inbreeding of the dam had a significant effect on preweaning daily gain though the size of the effect was considerably small (-0.003g/%) compared to similar studies. The correlation between inbreeding of the calf and dam was high (r = 0.83) and therefore interpretation of this result may be difficult. Studies considered here provide overwhelming evidence of the detrimental effect of inbreeding on preweaning growth.

Table 2.4. Effect of inbreeding on preweaning gain (g/d/%).

Population	Breed	Years studied	No. Records	Average inbreeding (%)		Partial regression coefficient				Source
						F_c		F_d		
				Calf	Dam	BC/HC[a]	HC	BC/HC[a]	HC	
OAES, Corvallis OR	H, A	1952-57	280	~15		-4.5360		-0.0000		Alexander and Bogart (1961)
USRLES, Miles City MO	H	1934-59	2,027	~16	~11	-1.1610	-4.3111	-4.7611	-1.0833	Brinks et al. (1963)
SJBRC, Hesperus CO	H	1946-86	5,420	36/4[b]		-1.7237	-1.0433	-1.3608	-1.5422	Clarke (1988)
USRLES, Miles City MO	H	1935-89	4,427	~20	~20	-0.0004		-0.0030		MacNeil et al. (1992)
USRLES, Miles City MO	H	1931-75	7,380	10	8	-1.8900		-2.5200		Pariacote et al. (1998)

OAES - Oregon Agricultural Experiment Station, USRLES - United States Range Livestock Experiment Station, SJBRC - San Juan Basin Research Center, H - Hereford, A - Angus, BC - bull calf, HC - heifer calf, F_c and F_d - inbreeding of calf and dam. [a] regression of birth weight on inbreeding for bull calves for within sex analysis or for bull and heifer calves together for across sex analysis. [b]Average inbreeding level of the inbred group and linecross calves.

Weaning weight. Keller and Brinks (1976) evaluated the effect of inbreeding of the calf and dam on weaning weight separately for bull calves and heifer calves (Table 2.5). Inbreeding of the calf significantly reduced (-0.33kg/%) weaning weight of bull calves and not heifer calves (-0.15kg/%). Inbreeding of the dam had an opposite effect on bull and heifer calves with heifer calves experiencing more depression (-0.19kg/%) compared to bull calves (-0.04kg/%). Dinkel et al. (1968) reported inbreeding depression more than twice that observed by Keller and Brinks (1976). Inbreeding of the calf significantly reduced weaning weight of bull calves by 0.61kg/% and the depression of heifer calves was not significant (-0.36kg/%). On the contrary, inbreeding of the dam was significant in heifer calves (-0.73kg/%) and unimportant in bull calves (-0.23kg/%). These results were contradictory to an earlier report by Brinks et al. (1963) in Line 1 Hereford. Brinks et al. (1963) observed a more pronounced effect of inbreeding of the calf on weaning weight of heifer calves (-0.96kg/%) than bull calves (-0.27kg/%). On the other hand, the effect of inbreeding of the dam was more drastic on bull calves (-0.85kg/%) compared to heifer calves (-0.15kg/%). Their interpretation was that since bull calves have greater growth potential compared to heifers, reduction in milk production of the dams due to inbreeding will hinder the expression of the potential more in bulls than in heifer calves. In a study involving 23 inbred lines of Hereford cattle, Clarke (1988) observed that the effect of inbreeding of the calf was significant and similar in bull calves (-0.33kg/%) and heifer calves (-0.32kg/%). While the effect of inbreeding of the dam was detrimental to both sexes, the impact was double on heifer calves compared to bull calves (-0.40 vs. -0.20 kg/%).

Table 2.5. Effect of inbreeding on weaning weight (kg/%).

Population	Breed	Years studied	No. Records	Average Inbreeding (%)		Partial regression coefficient				Source
						F_c		F_d		
				Calf	Dam	BC/HC[a]	HC	BC/HC[a]	HC	
FRBCRS, Crawford NE	H, A, S	1951-55	677	13	10	-0.650		-0.070		Swiger et al. (1961)
USRLES, Miles City MO	H	1934-59	2,027	~16	~11	-0.268	-0.957	-0.853	-0.150	Brinks et al. (1963)
SJBRC, Hesperus CO	H	1946-51	546			-0.800		-0.520		Burgess et al. (1964)
UWS, Gillette WY	H	1953-64	302	11	5	-0.465		0.265		Nelms and Stratton (1967)
SDAES	H	1953-64	860	20	9	-0.612	-0.364	-0.231	-0.731	Dinkel et al. (1968)
NARC, Havre MO	H	1950-63	650	~13	~7	-0.118	-0.049	-0.777	-0.066	Anderson et al. (1972)
SJBRC, Hesperus CO	H	1950-71	1,234	37	27	-0.330	-0.150	-0.040	-0.190	Keller and Brinks (1976)
MSFL, Reno NV	H	1969-73	303	~13	8			-0.523		Bailey et al. (1977)
NARC, Havre MO	H	1976-83	611	15	14	0.390	-0.840	-0.650	0.200	Nevins et al. (1985)
SJBRC, Hesperus CO	H	1946-86	5,420	36/4[b]		-0.330	-0.320	-0.200	-0.400	Clarke (1988)
SDAES	H	1971-78	514	0	27			-0.470		MacNeil et al. (1989)
USRLES, Miles City MO	H	1931-75	7,380	10	8	-0.202		-0.255		Pariacote et al. (1998)

NARC - Northern Agricultural Research Center, UWS - University of Wyoming Substation, USRLES - United States Range Livestock Experiment Station, SJBRC - San Juan Basin Research Center, FRBCRS - Fort Robinson Beef Cattle Research Station, SDAES - South Dakota Agricultural Experiment Station, MSFL - Main Station Field Laboratory, H - Hereford, A - Angus, S - Shorthorn, BC - bull calf, HC - heifer calf, F_c and F_d - inbreeding of calf and dam.

[a]regression of birth weight on inbreeding for bull calves for within sex analysis or for bull and heifer calves together for across sex analysis. [b]Average inbreeding level of the inbred group and linecross calves.

Contrary to the approach employed in the other studies of estimating the effects of inbreeding by sex, Swiger (1961) obtained estimates of effects of inbreeding of the calf and dam across sex. In general, their results were consistent with other studies. Inbreeding of the calf resulted in a significant decrease (-0.65kg/%) in weaning weight. Also, a marginal reduction (-0.07kg/%) in weaning weight due to inbreeding of the dam was observed. Likewise, Burgess et al. (1954) demonstrated that weaning weight of calves declined with inbreeding in purebred Herefords at San Juan Basin Experiment Station in Hesperus, Colorado. They reported a highly significant effect (-0.80kg/%) of inbreeding of the calf on weaning weight. Weaning weight was also significantly affected (-0.52kg/%) by the inbreeding of the dam. Nelms and Stratton (1967) observed a 0.47kg reduction in weaning weight with a 1% increase in inbreeding of the calf in a closed line of Hereford cattle. Inbreeding of the dam did not affect weaning weight. Recently, Pariacote et al. (1998) also found that weaning weight of the calf was reduced significantly with the inbreeding of the calf (-0.20kg/%) and dam (-0.26kg/%). Likewise, Bailey et al. (1977) found a 0.52 kg reduction in weaning weight per percentage increase in inbreeding of the dam. MacNeil et al. (1989) reported effects of inbreeding of the dam on weaning weight of similar magnitude (-0.47kg/%).

Regression coefficients given in Table 2.5 indicate that reduction in weaning weight due to inbreeding of the calf and dam could be as high as about 1kg for each percentage increase in inbreeding. In the majority of the studies reviewed, inbreeding was practiced concurrently with selection. Therefore, the general observation of inbreeding depression may be suggestive of the ineffectiveness of selection in overcoming inbreeding depression in weaning weight.

Postweaning daily gain. Alexander and Bogart (1961) assessed the effect of inbreeding on postweaning daily gain in lines of Hereford and Aberdeen Angus cattle (Table 2.6). Replacement animals were selected based on an index including postweaning daily gain. Postweaning daily gain tended to decline (-4.5g/d/%) with inbreeding of the calf. However, inbreeding of the dam significantly increased postweaning gain (4.5g/d/%). The increase in postweaning gain with inbreeding of the dam was counterintuitive. Their explanation for these results was that calves from highly inbred cows were genetically superior to calves from lowly inbred cows due to confounding of inbreeding and genetic level; therefore an increase in postweaning gain with inbreeding of the dam was an indication of the effectiveness of selection in overcoming inbreeding depression which also explained the nonsignificant reduction due to inbreeding of the calf. These results were consistent with those reported by Swiger et al. (1961) who found a negative and positive effect of inbreeding of the calf and dam respectively, for inbred lines at Fort Robinson Nebraska. The direction of the effects of inbreeding of the calf and dam on postweaning gain was similar to that observed by Dinkel et al. (1968) in a study involving four inbred lines of Hereford cattle. However, the magnitude of the effects was not significant in bull or heifer calves. The decline in postweaning gain with increased inbreeding of the calf was also observed by Swiger et al. (1961) in inbred lines located at Lincoln, Nebraska. Contrary to other studies, they found a decline in postweaning gain with increasing levels of inbreeding of the dam.

Clarke (1988) investigated the effect of inbreeding of the calf and dam in only bull calves. Inbreeding of the calf significantly reduced (-1.7g/day/%) postweaning while inbreeding of the dam led to a nonsignificant decline (-0.5g/day/%). Snelling et al. (1996)

also reported significant reduction in postweaning gain with an increase in inbreeding of the calf in three of the four herds considered. Nelms and Stratton (1967) found no effect of inbreeding of calf or dam on postweaning gain.

There is general agreement among studies that inbreeding of the calf reduces postweaning daily gain while inbreeding of the dam has a minimal effect at least for the range of inbreeding considered.

Final weight on test. Dinkel et al. (1968) reported a significant reduction in final weight of bull and heifer calves in Hereford cattle associated with inbreeding of the calf and only heifers were affected by inbreeding of the dam (Table 2.6). The reduction due to inbreeding of the calf was more severe on bull calves (-1.09 kg/%) than heifer calves (-0.53 kg/%) and the effect of inbreeding of the dam was switched around, with the impact being minimal and nonsignificant on bull calves (-0.004 kg/%) and severe on heifer calves (-0.61kg/%). The results reported by Brinks (1971) were in agreement with those of Dinkel et al. (1968) regarding the effects of inbreeding of the calf. However, the magnitude of the effects were about half in this study (Brinks (1971). Brinks et al. (1971) found no effect of inbreeding of the dam on final weight. Detrimental effects of inbreeding on final weight was also reported by Stonaker (1954) who observed 7% superiority in final weight of outbred compared to inbred bull calves. On the contrary, Anderson et al. (1972) observed no significant reduction in final weight due to inbreeding of either the calf or the dam though there was a tendency for final weight to decline with inbreeding. The effects of inbreeding of the calf and dam were similar for bulls (-0.61 vs. -0.65 kg/%) and heifer calves (-0.17 vs. -0.19 kg/%). Nelms and Stratton (1967) also found a nonsignificant effect of inbreeding of the calf and dam on final weight in a closed

Table 2.6. Effect of inbreeding on average daily gain and final weight on test.

Population	Breed	Years studied	No. Records	Average F (%)		Partial regression coefficient				Source
						F_c		F_d		
				Calf	Dam	BC/HC[a]	HC	BC/HC[a]	HC	
Average daily gain (g/d)										
OAES, Corvallis OR	H, A	1952-57	280	~15		-4.50		4.50		Alexander and Bogart (1961)
FRBCRS, Crawford NE	H, A, S	1951-55	677	13	10	-1.50		1.60		Swiger et al. (1961)
NAES, Lincoln NE	H, A	1951-55	283	5	3	-3.40		-2.50		Swiger et al. (1961)
UWS, Gillette WY	H	1953-64	302	11	5	1.10		1.30		Nelms and Stratton (1967)
SDAES	H	1953-64	713	20	9	-1.00	-0.50	0.40	0.70	Dinkel et al. (1968)
NARC, Havre MO	H	1950-63	650	~13	~7	-3.20	-0.50	-1.40	0.50	Anderson et al. (1972)
SJBRC, Hesperus CO	H	1950-85	1,873	36/4[b]		-1.70		-0.50		Clarke (1988)
USRLES, Miles City MO	H	1935-90	4,025			-2.10				Snelling et al. (1996)
NARC	H	1976-90	990			-5.40				,,
CHR	H	1974-90	1,811			-2.40				,,
HH	H	1967-90	1,642			-1.00				,,
Final weight (kg)										
UWS, Gillette WY	H	1953-64	302	11	5	-0.147		0.500		Nelms and Stratton (1967)
SDAES	H	1953-64	713	20	9	-1.086	-0.525	-0.004	-0.607	Dinkel et al. (1968)
USRS	H			4 to 41		-0.440	-0.252	-0.308	0.134	Brinks (1971)
NARC, Havre MO	H	1950-63	650	~13	~7	-0.611	-0.168	-0.650	-0.187	Anderson et al. (1972)

OAES - Oregon Agricultural Experiment Station, NARC - Northern Agricultural Research Center, UWS - University of Wyoming Substation, USRLES - United States Range Livestock Experiment Station, SJBRC - San Juan Basin Research Center, FRBCRS - Fort Robinson Beef Cattle Research Station, SDAES - South Dakota Agricultural Experiment Station, USRS – United State Research Stations (44 lines), NAES - Nebraska Agricultural Experiment Station, CHR - Cooper Hereford Ranch, HH - Holden Herefords, H - Hereford, A - Angus, S - Shorthorn, BC - bull calf, HC - heifer calf, F_c and F_d - inbreeding of calf and dam. [a] regression of birth weight on inbreeding for bull calves for within sex analysis or for bull and heifer calves together for across sex analysis. [b] Average inbreeding level of the inbred group and linecross calves.

herd of Hereford undergoing selection.

Results of studies considered here indicate that inbreeding of the calf and dam reduces final weight on test. Based on results reported by Brinks (1971), the maximum reduction in an individual's final weight can be as high as 18kg (i.e. for the maximum inbreeding coefficient of 41% reported in this study).

Mature weight and rate of maturing. McCurley et al. (1984) determined the influence of inbreeding of the calf and dam on estimated fall mature weight and rate of maturing (growth rate relative to mature weight) in inbred and noninbred lines developed from Angus, Hereford and Shorthorn. The average inbreeding level across lines ranged from 15 to 23% for Angus, 6 to 10% for Hereford, and 16 to 32% for Shorthorn. Inbreeding levels of the noninbred lines ranged from 2 to 5% across breeds. They reported a significant decline in estimated fall mature weight for all the breeds with increasing level of inbreeding of the individual and inbreeding of the dam increased fall mature weight in all the breeds though the increase was more pronounced and significant in Shorthorn (Table 2.7). Brinks et al. (1965) also reported a decline in spring and fall mature weight as the level of inbreeding of the individual increased in Line 1 Hereford. The size of the effects of inbreeding was slightly lower in this study compared to those reported by McCurley et al. (1984). Unlike mature weight, rate of maturing was not affected by inbreeding of the individual in Hereford, Shorthorn and Angus (McCurley et al., 1984). However, inbreeding of the dam reduced rate of maturing in Angus and Shorthorn.

Table 2.7. Effect of inbreeding on cow mature weight (kg/%) and rate of maturing per mature weight.

Population	Breed	Years studied	No. Records	Average inbreeding (%)		Partial regression coefficient		Source
				Individual	Dam	F_i	F_d	
Mature Fall Weight								
USRLES, Miles City MO	H	1934-59	536	16	12	-1.486		Brinks et al. (1965)
BCRS, Front Royal VA	H	1954-68	140	~4		-2.263	0.940	McCurley et al. (1984)
BCRS, Front Royal VA	A	1954-68	283	~12		-2.029	0.881	,,
BCRS, Front Royal VA	S	1954-68	280	~15		-1.951	1.225	,,
Mature Spring Weight								
USRLES, Miles City MO	H	1934-59	536	16	12	-1.280		Brinks et al. (1965)
Rate of maturing								
BCRS, Front Royal VA	H	1954-68	140	~4		0.00013	-0.00016	McCurley et al. (1984)
BCRS, Front Royal VA	A	1954-68	283	~12		0.00006	-0.00035	,,
BCRS, Front Royal VA	S	1954-68	280	~15		0.00002	-0.00032	,,

BCRS – Beef Cattle Research Station, USRLES - United States Range Livestock Experiment Station, H - Hereford, A - Angus, S - Shorthorn, F_i – inbreeding of the cow, F_d – inbreeding of the dam.

2.9. Conclusions

The results of studies reviewed in this chapter varied from depression to enhancement in performance associated with the increase in the level of inbreeding of the calf and dam. For example, superior and inferior inbred lines were observed in different experiments. The magnitude of the impact of inbreeding was also variable among studies reflecting the differences in foundation stocks, environmental conditions, management policy, rate of inbreeding and selection practiced within particular line. However, the general observation is that increases in the level of inbreeding of either the calf or the dam depress fertility, survival, preweaning and postweaning growth in beef cattle.

2.10. Literature Cited

Alexander, G. I., and R. Bogart. 1961. Effect of inbreeding and selection on performance characteristics of beef cattle. J. Anim. Sci. 20:702-707.

Anderson, D. C., A. E. Flower, F. S. Willson, and C. Windecker. 1972. Factors affecting production in beef cattle. Proc. Western Section, American Society of Animal Science 23:6-11.

Bailey, C. M., J. A. Edwards, and Y. O. Koh. 1977. F_1 cross between mildly inbred Hereford selection lines of common genetic origin. J. Anim. Sci. 44:23-29.

Brinks, J. S. 1971. Inbreeding, linecrossing and topcrossing results from inbred lines of beef cattle. Symposium on Development stocks and Their Use. Colorado State University, Fort Collins, CO.

Brinks, J. S. 1975. Origin and History of cattle lines at the San Juan Basin Research Center, Hesperus, Colorado. 26th Ann. Beef Cattle Improvement Rep. Colorado State Univ. Exp. Sta. Rep. 940.

Brinks, J. S., and B. W. Knapp. 1975. Effects of inbreeding on performance traits in beef cattle in the western region. Colorado State Univ. Exp. Sta. Tech. Bull. 123.

Brinks, J. S., B. W. Knapp, J. J. Urick, and O. F. Pahnish. 1972. Heterosis in preweaning maternal traits among lines of Hereford cattle. J. Anim. Sci. 34:14-20.

Brinks, J. S., R. T. Clark, and N. M. Kieffer. 1963. Sex differences in response to inbreeding in a line of Hereford cattle. Proc. Western Section, American Society of Animal Science ppV1-V6.

Brinks, J. S., R. T. Clark, and N. M. Kieffer. 1965. Evaluation of response to selection and inbreeding in a closed line of Hereford cattle. Tech. Bull. 1323. ARS, USDA, Washington, DC.

Burgess, J. B., N. L. Landblom, and H. H. Stonaker. 1954. Weaning weights of Hereford calves as affected by inbreeding, sex, and age. J. Anim. Sci. 13:843-851.

Burrow, H. M. 1993. The effects of inbreeding in beef cattle. Anim. Breed. Abstr. 61:737-751.

Caballero, A. 1994. Developments in the prediction of effective population size. Heredity 73:657-679.

Charlesworth, B., and D. Charlesworth. 1999. The genetic basis of inbreeding depression. Genet. Res. (Camb.) 74:329-340.

Clarke, L. S. 1988. Effect of inbreeding on performance traits in Hereford cattle. Ph.D. Dissertation. Colorado State University.

Davenport, R. L., H. H. Stonaker, K. Riddle, and T. M. Sutherland. 1965. Heritability of reproductive performance in inbred and linecross beef cows. J. Anim. Sci. 24:434-437.

Dinkel, C. A., D. A. Busch, J. A. Minyard, and W. R. Trevillyan. 1968. Effects of inbreeding on growth and conformation of beef cattle. J. Anim. Sci. 27:313-322.

Dinkel, C. A., L. M. Anderson, W. R. Parker, and W. R. Trevillyan. 1972. Effects of inbreeding on fertility and livability in beef cattle. J. Anim. Sci. 35:725-729.

Dudash, M. R., and D. E. Carr. 1998. Genetics underlying inbreeding depression in *Mimulus* with contrasting mating systems. Nature 393:682-684.

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to Quantitative genetics. Longman Group, Essex, UK.

Fisher, R. A. 1930. The genetical theory of natural selection. Clarendon Press, Oxford.

Flower, A. E., J. S. Brinks, J. J. Urick, and F. S. Willson. 1963. Comparisons of inbred lines and linecrosses for performance traits in Hereford range cattle. J. Anim. Sci. 22:914-918.

Grapevine, P. W., J. S. Brinks, and G. V. Richardson. 1975. General and specific combining abilities of inbred lines of Hereford cattle. J. Anim. Sci. 41:527-533.

Groen, A. F., B. W. Kennedy, and J. J. Eissen. 1995. Potential bias in inbreeding depression estimate when using pedigree relationships to assess the degree of homozygosity for loci under selection. Theor. Appl. Genet. 91:665-671.

Hayman B. I., and K. Mather. 1953. The progress of inbreeding when homozygotes are at a disadvantage. Heredity 7:165-183.

Hill, W. G., and A. Robertson. 1968. The effects of inbreeding at loci with heterozygote advantage. Genetics 60:615-628.

Kärkkäinen, K., H. Kuittinen, R. Van Treuren, C. Vogl, S. Oikarinen, and O. Savolainen. 1999. Genetic basis of inbreeding depression in *Arabis petraea*. Evolution 53:1354-1365.

Keller, D. G., and J. S. Brinks. 1976. Response to inbreeding in Hereford lines. Proc. Western Section, American Society of Animal Science pp 38-40.

Keller, D. G., and J. S. Brinks. 1978a. Inbreeding by environment interactions for weaning weight in Hereford cattle. J. Anim. Sci. 46:48-53.

Keller, D. G., and J. S. Brinks. 1978b. Mating system by environment interactions for weaning weight in Hereford cattle. J. Anim. Sci. 46:54-59.

Krehbiel, E. V., R. C. Carter, K. P. Bovard, J. A. Gaines, and B. M. Priode. 1969. Effects of inbreeding and environment on fertility of beef cattle matings. J. Anim. Sci. 29:528-533.

Kress, D. D., D. E. Doornbos, D. C. Anderson, and D. Rossi. 1992. Performance of crosses among Hereford, Angus, and Simmental cattle with different levels of Simmental breeding: VI. Maternal heterosis of 3- to 8-year-old dams and dominance model. J. Anim. Sci. 70:2682-2687.

Kress, D. D., P. J. Burfening, D. C. Anderson, and R. L. Blackwell. 1979. Heterosis among closed lines of Hereford cattle. I. Preweaning growth and survival. J. Anim. Sci. 49:950-957.

López-Fanjul, C., and A. Villaverde. 1989. Inbreeding increases genetic variance for viability in *Drosophila melanogaster*. Evolution 43:1800-1804.

MacNeil, M. D., D. D. Dearborn, L. V. Cundiff, C. A. Dinkel, and K. E. Gregory. 1989. Effects of inbreeding and heterosis in Hereford females on fertility, calf survival and preweaning growth. J. Anim. Sci. 67:895-901.

MacNeil, M. D., J. J. Urick, S. Newman, and B. W. Knapp. 1992. Selection for postweaning growth in inbred Hereford cattle: The Fort Keogh, Montana Line 1 Example. J. Anim. Sci. 70:723-733.

Malécot, G. 1948. Les Mathematiques de l heredite. Masson, Paris.

McCurley, J. R., W. T. Butts Jr., and K. P. Bovard. 1984. Growth patterns of Angus, Hereford and Shorthorn cattle. I. Comparison of inbreed and noninbred lines, changes in patterns over time and effects of level of inbreeding and reproductive performance. J. Anim. Sci. 59:1194-1204.

Minvielle, F. 1979. Comparing the means of inbred lines with the base population: a model with overdominant loci. Genet. Res. 33:89-92.

Nelms, G. E., and P. O. Stratton. 1967. Selection practiced and phenotypic change in a closed line of beef cattle. J. Anim. Sci. 26:274-277.

Nevins, D. L., D. D. Kress, D. C. Anderson, D. E. Doornbos, and P. J. Burfening. 1985. Effect of inbreeding on weaning weight and pregnancy rate in a closed line of Hereford cattle. Proc. Western Section, American Society of Animal Science 36:54-56.

Pariacote, F., L. D. Van Vleck, and M. D. MacNeil. 1998. Effects of inbreeding and heterozygosity on preweaning traits in a closed population of Herefords under selection. J. Anim. Sci. 76:1303-1310.

Pearl, R. 1913. A contribution towards an analysis of the problem of inbreeding. Amer. Nat. XLVII:577-614.

Pearl, R. 1914a. On the results of inbreeding a Mendelian population: A correction and extension of previous conclusions. Amer. Nat. 48:57-62.

Pearl, R. 1914b. Studies on inbreeding. V. Inbreeding and relationship coefficients. Amer. Nat. XLVIII:513-523.

Pearl, R. 1917a. Studies on inbreeding. VII. Some further considerations regarding the measurement and numerical expression of degrees of kinship. Amer. Nat. 51:545-559.

Pearl, R. 1917b. Studies on inbreeding. VIII. A single numerical measure of the total amount of inbreeding. Amer. Nat. 51:636-639.

Reeve, E. C. R. 1955. Inbreeding with homozygotes at a disadvantage. Ann. Hum. Genet. 19:332-346.

Robertson, A. 1952. The effect of inbreeding on variation due to recessive genes. Genetics 37:189-207.

Roff, D. A. 2002. Inbreeding depression: tests of the overdominance and partial dominance hypotheses. Evolution 56:768-775.

Snelling, W. M., M. D. MacNeil, D. D. Kress, D. C. Anderson, and M. W. Tess. 1996. Factors influencing genetic evaluations of linebred Hereford cattle in diverse environments. J. Anim. Sci. 74:1499-1510.

Stonaker, H. H. 1954. Observations on reproduction, growth, feed utilization and grades of inbred and outbred Hereford cattle. J. Anim. Sci. 13:963 (Abstr.).

Swiger, L. A., K. E. Gregory, R. M. Koch, and V. A. Arthaud. 1961. Effect of inbreeding on performance traits of beef cattle. J. Anim. Sci. 20:626-630.

Urick, J. J., J. S. Brinks, R. T. Clark, O. F. Pahnish, and F. S. Willson. 1966. History and performance of inbred lines of Hereford cattle developed at the United State Range Livestock Experiment Station. Montana Agriculture Experiment Station, Montana State University, Bozeman, Bulletin 602:3-22.

Wade, M. J., S. M. Schuster, and L. Stevens. 1996. Inbreeding: its effect on response to selection for pupal weight and the heritability variance in fitness in the flour beetle, *Tribolim castaneun*. Evolution 50:723-733.

Wang, J. 2000. Effects of population structures and selection strategies on the purging of inbreeding depression due to deleterious mutations. Genet. Res., (Camb.) 76:75-86.

Wang, J. A., Caballero, P. D. Keightley, and W. G. Hill. 1998. Bottleneck effect on genetic variance: A theoretical investigation of the role of dominance. Genetics 150:435-447.

Wang, J., W. G. Hill, D. Charlesworth, and B. Charlesworth. 1999. Dynamics of inbreeding due to deleterious mutations in small populations: mutation parameters and inbreeding rate. Genet. Res., (Camb.) 74:165-178.

Whitlock, M. C., and K. Fowler. 1999. The changes in genetic and environmental variance with inbreeding in *Drosophila melanogaster*. Genetics 152:345-353.

Willis, M. B., and A. Wilson. 1974. Factors affecting birth weight of Santa Gertrudis calves. Anim. Prod. 18:231-236.

Wright, S. 1922a. The effects of inbreeding and crossbreeding on guinea pigs. III. Crosses between highly inbred families. U. S. Department of Agriculture Report, Washington D. C. Bulletin No. 1121. pp1-60.

Wright, S. 1922b. Coefficients of inbreeding and relationship. Amer. Nat. 61:330-338.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97-159.

Wright, S. 1938. Size of population and breeding structure in relation to evolution. Science 87:430-431.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics 15:323-354.

Wright, S. 1952. The theoretical variance within and among subdivisions of a population that is in a steady state. Genetics 37:312-321.

Wright, S., and H. C. McPhee. 1925. An approximate method of calculating coefficient of inbreeding and relationships from livestock pedigree. J. Agric. Res. 31:377-383.

CHAPTER III

ASSESSING LEVEL OF INBREEDING IN RED ANGUS CATTLE

3.1. Abstract

The objective of this study was to assess the level of inbreeding in Red Angus cattle. Pedigree records (n = 829,882) of animals registered with the Red Angus Association of America (RAAA) were obtained from Colorado State University Center for Genetic Evaluation of Livestock. The proportion of individuals with one or both parents unknown was 11% while individuals with both parents known comprised 89% of the data. Total registrations in 2001 was approximately 50 000 (50 times the registrations in 1960). The average inbreeding of the entire population was 3.1% with the minimum and maximum of 0 and 51%, respectively. The proportion of inbred individuals increased over time and 90% of the population were inbred in 2001. The major source of inbreeding (98%) in 2001 was due to mating among distant relatives. The apparent rate of inbreeding fluctuated over time due to incomplete pedigree. Considering the period 1992-2001, where pedigrees were more complete, the rate of inbreeding was 0.08%±0.002 per year. The results from this study suggest that inbreeding is inevitable in this population. However, the rate at which inbreeding is occurring is still minimal at a breed level. The

fact that individuals exist with inbreeding coefficients as high as 51% may be suggestive that occasional intense inbreeding is practiced by some breeders.

3.2. Introduction

The Red Angus Association of America (RAAA) was established in 1954 with its main focus on the registration of pedigree and recording of performance information for Red Angus cattle. Since its inception, various changes have occurred regarding the criteria for registration. Furthermore, reproductive and selection technologies have been increasingly adopted and may have impacted on the genetic structure of the population.

There has been extensive use of artificial insemination (AI) among Red Angus breeders since the late 1950's. This technology made it easier for breeders who are geographically separated from each other to exchange genetic material. This exchange created genetic ties among herds, which was essential for the implementation of the national genetic evaluation program for the breed. Injudicious use of AI (e.g. no limit on the number of matings from a sire) may lead to high genetic relationships in the population, which results in high levels of inbreeding.

For the past fifteen to twenty years, expected progeny differences (EPD) obtained from Best Linear Unbiased Prediction (BLUP) have been available as a selection tool in Red Angus. The BLUP methodology uses the numerator relationship matrix to enhance the accuracy of genetic prediction. However, when the heritability of the trait is low, BLUP places considerable emphasis on information from relatives. This increases the correlation between EPD of relatives, increasing the probability of truncation selection

identifying related individuals for breeding purpose. As a result, the selected group is likely to have, on average, a higher relationship coefficient compared to the population from which they were selected. Research has shown that truncation selection using BLUP EPD without any constraint on the level of inbreeding can lead to high rates of inbreeding in subsequent generations (Belonsky and Kennedy, 1988; Leitch et al., 1993; Klieve et al., 1994; Sanchez et al., 1999).

In order to minimize inbreeding while achieving appreciable genetic response, optimization approaches have been developed recently (Villanueva and Woolliams, 1997; Grundy et al., 1998; Meuwissen and Sonesson, 1998; Shepherd and Kinghorn, 1998; Bijma et al., 2000a; Weigel and Lin, 2000; Bijma et al., 2001; Tozer and Stokes, 2002). Given the number of candidates required for selection, these programs identify the set of individuals that provide compromise between the level of inbreeding and the genetic response. Such software can maximize the genetic response at a given ceiling for inbreeding. The order of importance with regard to minimizing inbreeding and rapid genetic response may differ from one breeder to the other. Tozer and Stokes (2002) using multiple-objective programming, found that breeders who ranked net merit first and minimizing inbreeding second selected different set of sires from those breeders who had these objectives in reverse order.

In view of these developments, it is likely that the rate of increase of inbreeding in the Red Angus population may have changed over time. The objective of this study was to assess the level of inbreeding in the Red Angus cattle. Other aspects related to inbreeding such as pedigree completeness and type of mating practiced, were considered.

3.3. Materials and Methods

The RAAA uses a category-based program of registration initiated in 1980. The program allows registration of purebred and upgraded Red Angus cattle. Progeny of animals registered with the Canadian Red Angus Association (CRAA) and the American Angus Association (AAA) are also eligible for registration. In 1995, the RAAA introduced a Total Herd Reporting (THR) program, whereby members are required to collect annual production and performance information on all animals. For example, records on stillborn calves and cows that failed to get pregnant also form part of the database.

A total of 829,882 pedigree records of registered Red Angus cattle born between 1930 and 2001 were obtained from Colorado State University Center for Genetic Evaluation of Livestock (CSU-CGEL). Each record consisted of the following information: animal ID, sire ID, dam ID and date of birth (ID – refers to identification number). Information about the knowledge of the pedigree for the entire dataset is summarized in Table 3.1. About nine out of ten animals had both their parents known. Therefore, inbreeding coefficients calculated in this study reflect a lower limit for actual inbreeding.

Table 3.1. Proportion of animals with known or unknown paternity in Red Angus cattle population.

No. Parents Known	No. Animals	Proportion of total (%)
0	15,149	2
1	72,788	9
2	741,945	89
Total	829,882	100

Inbreeding coefficients were computed using the Animal Breeder's Tool Kit (ABTK: Golden et al., 1992): an implementation of an algorithm proposed by Meuwissen and Luo (1992). The ABTK requires animals to be ordered chronologically and that an input file with pedigree information has three fields: animal, sire and dam identification numbers. Missing or unknown parents should be denoted by a period. An output file with inbreeding coefficients for inbred individuals is created.

Three sets of analyses were performed using the computed inbreeding coefficients. In the first analysis, inbreeding coefficients were classified into categories (of 1% interval) to assess the distribution of individuals based on animal's inbreeding coefficient. The second analysis was based on individuals born between 1960 and 2001 inclusive, amounting to 96% of the entire dataset. Inbreeding trend was established by calculating average inbreeding level per year of birth. Furthermore, an estimate of the rate of inbreeding was obtained by regressing inbreeding coefficient on year of birth using the REG procedure of SAS (SAS, Inc. Cary, NC). The third analysis was aimed at determining whether inbreeding was intentional or accidental. This was achieved by

classifying whether inbreeding was due to mating of close relatives or to distant relationships. Inbreeding was considered close when mating occurred between parent-offspring (PO), full sibs (FS) and half sibs (HS) while any other matings that resulted in inbred individuals were considered distant inbreeding and classified as other relationships (OR).

3.4. Results

3.4.1. Annual registration and proportion of inbred animals

The annual registration and proportion of inbred individuals from 1960 to 2001 are given in Figure 3.1. It is apparent that the annual registration in this population can be broken down into two phases: prior to 1985 and post 1985, i.e. the periods of mild and rapid increase, respectively. The period prior to 1985 represents the formative stage of the breed with few breeders participating in the association. The rapid increase in the post 1985 period seems to coincide with the introduction of the category-based registration in 1980. Initiation of the Total Herd Reporting program in 1995 might have also contributed in sustaining the positive trend. In 2001 the registration was about 50 times that of 1960.

In contrast to yearly registrations, the period prior to 1985 was characterized by rapid increase in the proportion of inbred individuals while the period after 1985 showed a mild increase. The linear increase in the proportion of inbred animals prior to 1985 may be due to increased knowledge about the pedigree information and small gene pool from which breeding animals had to be selected. The trend post 1985 is likely to be due to registration of animals whose genetic relationships to the registered population were

unknown. In 2001, the proportion of inbred individuals was 90%. It seems, from this result, that the majority of the breeding individuals in the population are related.

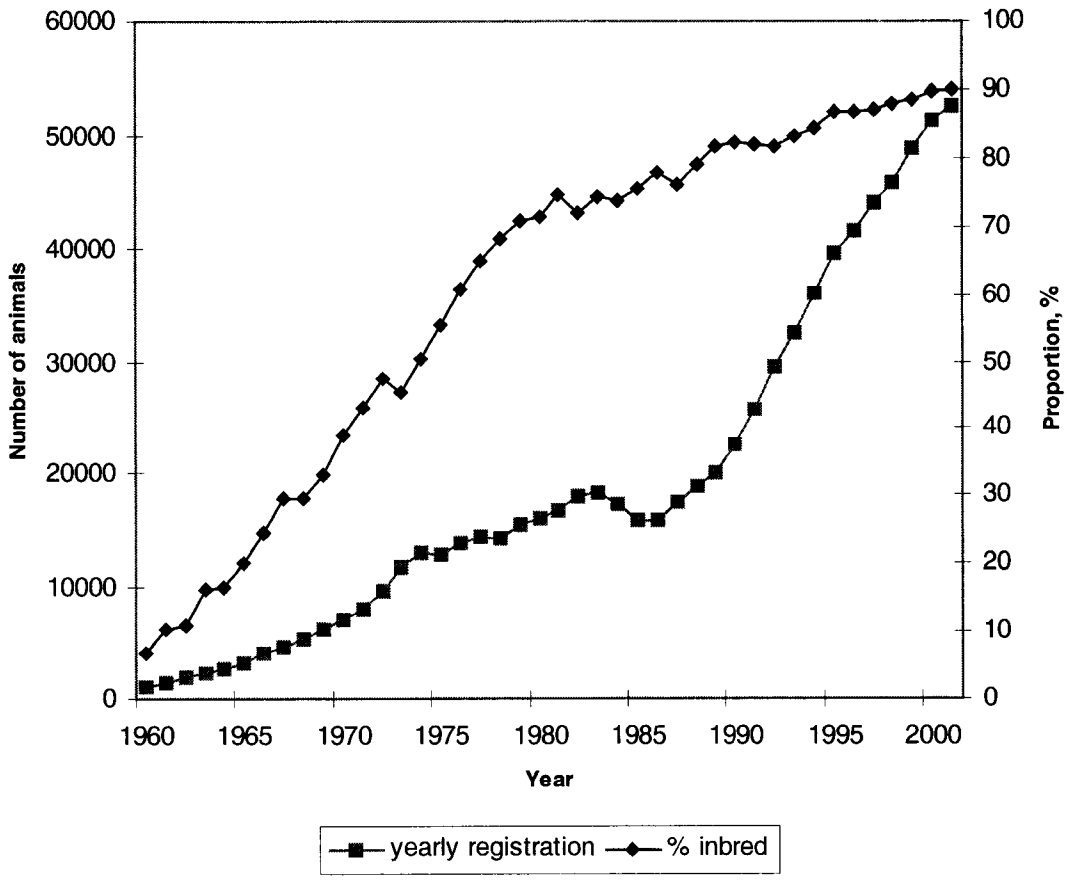


Figure 3.1. Yearly registration and proportion of inbred individuals.

3.4.2. Pedigree completeness

The proportion of individuals with known and unknown paternity between the period 1960 and 2001 is given in Figure 3.2. The inbreeding level of these individuals is therefore underestimated. The trend seems to follow a wave-like pattern that lasts about fifteen years. The proportion of individuals of unknown paternity declined from about

40% in 1960 to less than 10% in 2001. Partitioning the proportion of animals with unknown paternity into single and both parent(s) unknown, suggest that pedigree incompleteness in this population is mainly due to animals with a single parent unknown. This result was expected because most breeders follow a multi-sire breeding system where an individual's sire cannot be ascertained without DNA paternity testing of the potential sires.

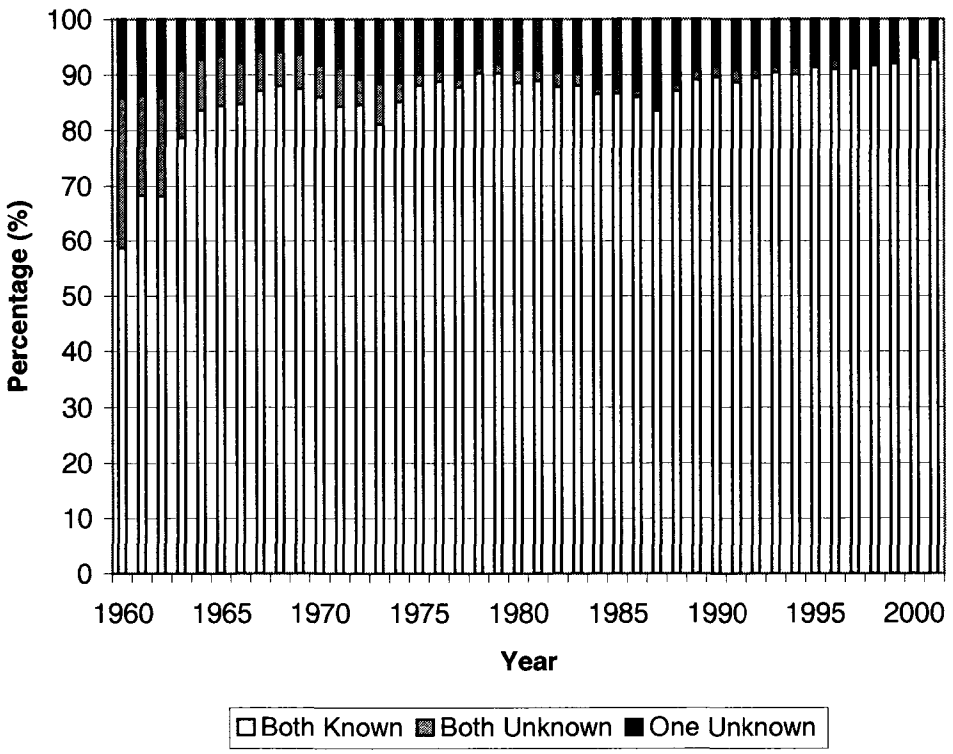


Figure 3.2. Proportion of individuals born each year with known and unknown paternity.

3.4.3. Types of mating practiced

Figure 3.3 shows a rapid increase in OR for the period 1960 to 1970 which was followed by a gradual increase that seem to plateau in the last ten years of the study. An opposite trend was observed for PO and HS, which accounted for about 20 to 40% of the mating that resulted in inbred progeny over the same period. The FS matings were rare and essentially constant over the entire period. The relatively low proportion of FS mating might have resulted from avoidance of this type of mating by breeders or due to low frequency of full sibs in beef cattle, which reduces the chance of an accidental random mating between full sibs (Bijma et al., 2001).

In the year 2001 about 98% of the mating were OR and the close inbreeding comprised only about 2%. This result suggests that the general tendency among breeders is to avoid close inbreeding.

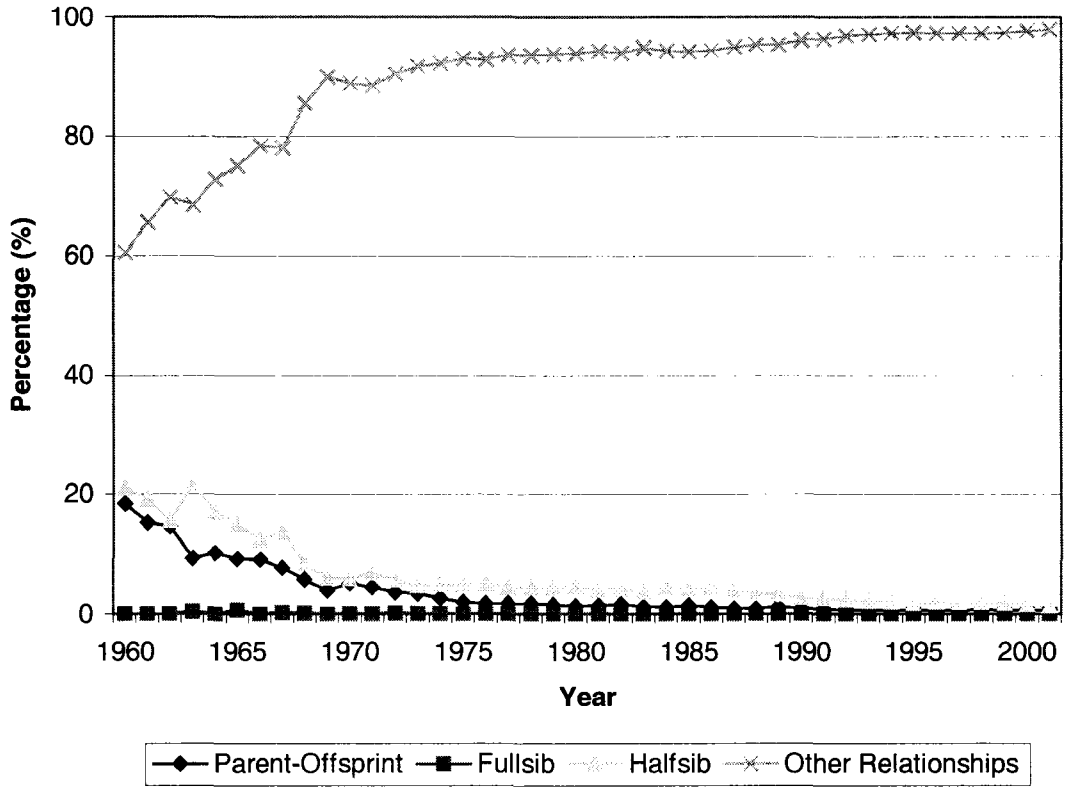


Figure 3.3. Types of mating practiced between parents of inbred individuals.

3.4.4. Distribution of individual inbreeding coefficients

The distribution of individuals based on their inbreeding coefficients is highly skewed, with about 93% of the population falling between the interval 0 and 10% inbreeding (Figure 3.4). Individual inbreeding coefficients ranged from 0 to 51% and the population average was low at 3.1%.

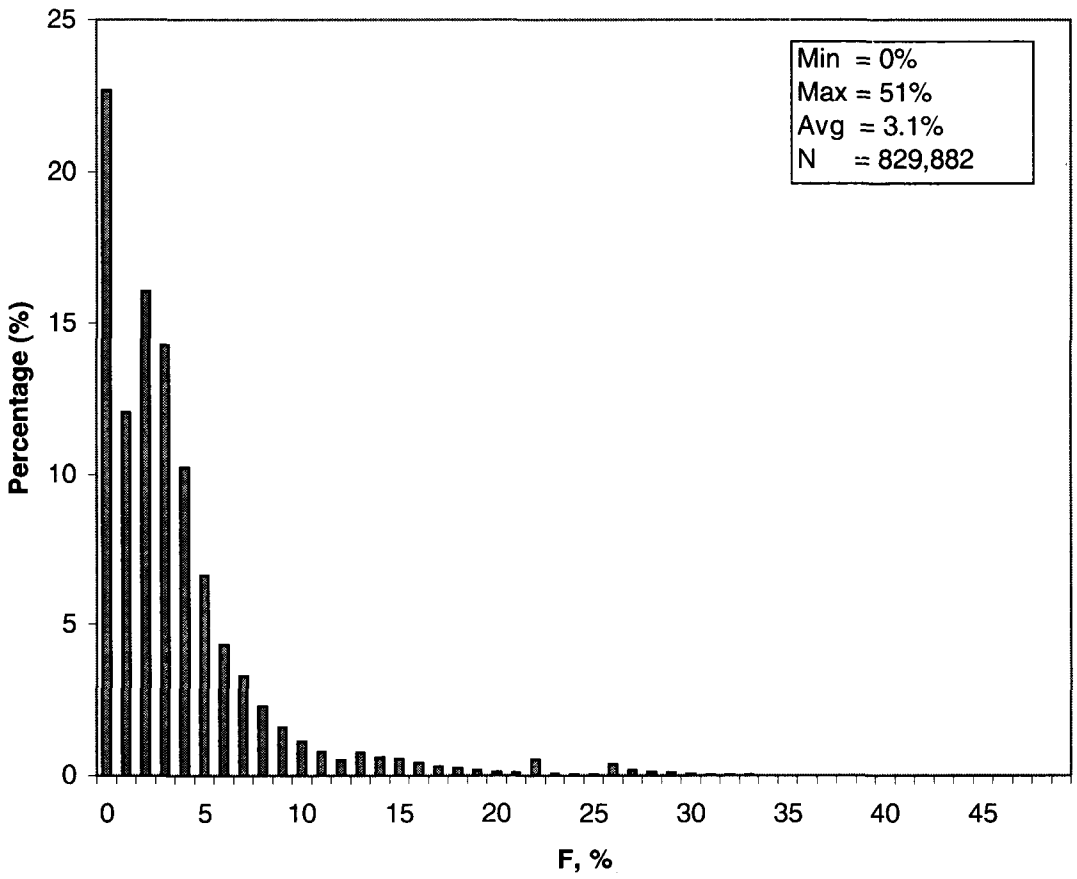


Figure 3.4. The distribution of individual inbreeding coefficients (F).

To obtain the time trend in the distribution of individual inbreeding levels, animals were grouped into four time intervals and their distribution computed within each class as shown in Figure 3.5. There was a considerable decline in the proportion of individuals that were non-inbred from 76% to 15% over the entire period. This trend was due to a general decline in the number of individuals with unknown paternity over time as shown earlier in Figure 3.2. The proportion of individuals within the 0 to 6% group was a reverse of the non-inbred group. Most of the individuals (72%) in the recent population (1990-2001) were within the 0 to 6% inbreeding group.

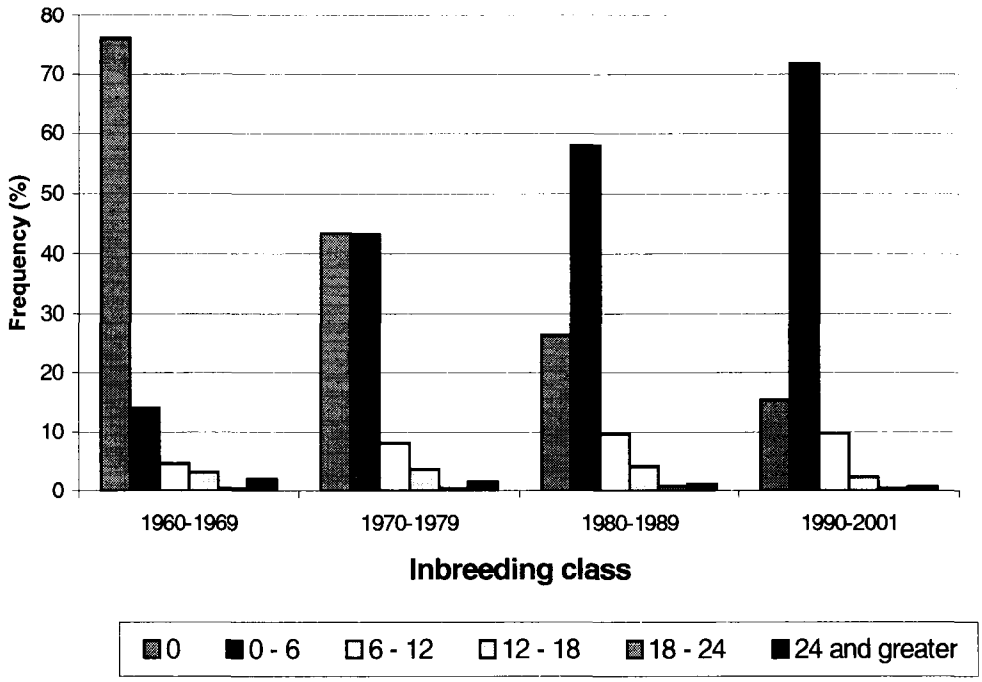


Figure 3.5. The distribution of individuals based on classes of inbreeding level (F%) over for year of birth time periods.

3.4.5. Level and rate of inbreeding

The inbreeding trend for the entire population was episodic (Figure 3.6). The period prior to 1980 was characterized by an increase in the level of inbreeding after which it was essentially constant for the next ten years followed by mild increases in recent years. This result was in agreement with observations in Figures 3.1 and 3.2. The average inbreeding level in 2001 was 3.6%.

There was considerable discrepancy in the mean level of inbreeding between the inbred group and the entire population prior to 1980. The most likely reason is that as the

entire population becomes inbred, the population inbreeding level and the inbreeding level of the inbred individuals converges.

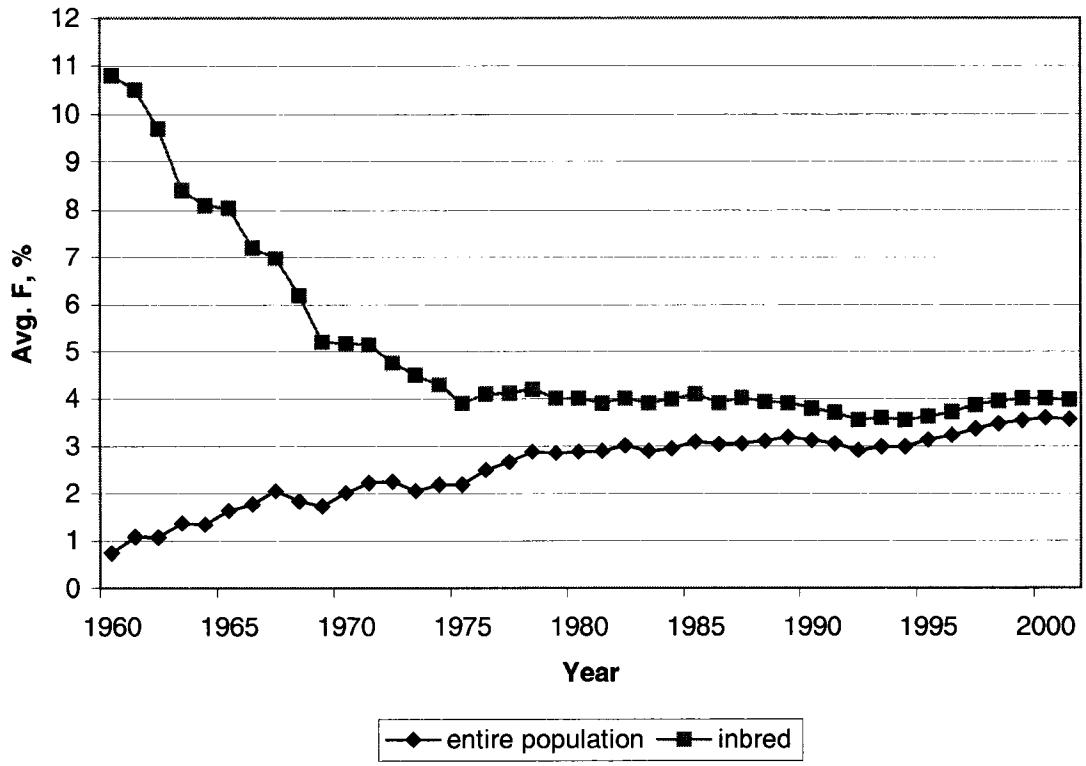


Figure 3.6. Average inbreeding level of the population.

Due to episodic behavior of the changes in the level of inbreeding, the rate of inbreeding (ΔF) was computed for each period: 1960-1977 (ΔF_{60}), 1978-1991 (ΔF_{78}), 1992-2001 (ΔF_{92}) and the entire period (ΔF_{pop}). The rate of accumulation of inbreeding was similar and highly significant ($P < 0.0001$) at 0.08% per year for ΔF_{60} and ΔF_{92} (Table 3.2). The estimate for ΔF_{78} was lower (0.01% per year) compared to the other two periods but significant ($P = 0.0005$). Comparisons among periods showed that ΔF_{78} was

different ($P < 0.0001$) from the other two periods. Considering the whole population, the estimate for rate of inbreeding (ΔF_{pop}) was highly significant at 0.04% per year.

Table 3.2. Estimates of the rate of inbreeding for different time periods.

Rate of inbreeding	Annual Change±SE	P-value
ΔF_{pop}	0.04±.0005	< 0.0001
ΔF_{60}	0.08±.003	< 0.0001
ΔF_{78}	0.01±.002	0.0005
ΔF_{92}	0.08±.002	< 0.0001

ΔF_{60}, ΔF_{78}, ΔF_{92}, and ΔF_{pop} are the rates of inbreeding for the time periods 1960 to 1977, 1978 to 1991, 1992 to 200, and 1960 to 2001, respectively. P-value is the probability associated with observed F-value given the null hypothesis that the annual change in inbreeding is not significantly different from zero.

3.5. Discussion

The results showed that the proportion of inbred individuals increased over time. The increase was mainly due to increased knowledge of the pedigree information as shown by the dramatic decline in the proportion of individuals with unknown paternity over time. The impact of pedigree incompleteness in computation of level of inbreeding has been investigated in other studies (Te Braake et al., 1994; Boichard et al., 1997, Goyache et al., 2003). These studies showed that the level of inbreeding in the population depends, to a large extent, on the number of individuals with unknown parentage.

In a study on genetic diversity in Swiss sheep breeds, Hagger (2002) observed a curvilinear relationship between ancestor generation and inbreeding level. That is, as the

pedigree was traced more generations further back, so was the increase in the level of inbreeding. In horses, MacCluer et al. (1983) reported similar results where the inbreeding levels of the population increased substantially with pedigree depth until 12 generations. However, when horses with the same degree of pedigree completeness were considered, they found that average inbreeding level decreased over time. Lutaaya et al. (1999) used a different approach to investigate the effects of pedigree incompleteness on the level of inbreeding. They randomly deleted varying proportions of dams from a pedigree to simulate unknown paternity. Their results showed an increase in the magnitude of underestimation with increase in the proportion of unknown dams up to 50%.

The observations above render ΔF_{60}, ΔF_{78}, and ΔF_{pop} as inaccurate measures of the rate of inbreeding in this population and thus of little, if any, biological importance. However, the ΔF_{92} can be considered a reliable but conservative estimate of the rate of inbreeding in this population because only a small proportion of individuals had unknown parents (7%). However, the observation that individuals with inbreeding as high as 51% exist in this population, suggest that some breeders may be practicing matings between highly related individuals.

3.6. Conclusions

The results from this study demonstrated an upward trend in the level of inbreeding over time in the registered Red Angus cattle population. Examination of the level of pedigree information revealed that the inbreeding trend might have been

influenced by incomplete knowledge about pedigree information on some individuals, particularly during the early years. However, the rate of inbreeding computed based on individuals with minimal pedigree incompleteness provided evidence in support of a gradual increase in the level of inbreeding. The low accumulation of inbreeding seem to be due to avoidance of matings between close relatives and also the initiative by the breed association to allow registration of individuals whose relationship to the breed was unknown because their parents were from other breed associations. A cause for concern, however, is the high proportion of inbred individuals in 2001 indicating that majority of the animals bred in 2000 were related. These relationships should be monitored to ensure long-term diversity in this population. Finally, it is important to note that the rate of inbreeding reported in this study is a conservative estimate and thus, the actual rate of inbreeding may be higher than observed in this study.

3.7. Literature Cited

Belonsky, G. M., and B. W. Kennedy. 1988. Selection on individual phenotype and best linear unbiased predictor of breeding values in a closed swine herd. J. Anim. Sci. 66:1124-1131.

Bijma, P., J. A. M. Van Arendonk, and J. A. Woolliams. 2000a. A general procedure to predict rates of inbreeding in populations undergoing mass selection. Genetics 154:1865-1877.

Bijma, P., J. A. M. Van Arendonk, and J. A. Woolliams. 2001. Predicting rates of inbreeding for livestock improvement schemes. J. Anim. Sci. 79:840-853.

Boichard, D., L. Maignel, and E. Verrier. 1997. The value of using probabilities of origin to measure genetic variability in a population. Gen. Sel. Evol. 29:5-23.

Golden, B. L., W. M. Snelling, and C. H. Mallinckrodt. 1992. Animal breeder's tool kit user's guide and reference manual. Colorado State Univ. Agric. Exp. Sta. Tech. Bull. LTB92-2.

Goyache, F., J. P. Gutiérrez, I. Fernández, E. Gomez, I. Alvarez, J. Díez, and L. J. Royo. 2003. Using pedigree information to monitor genetic variability of endangered populations: the Xalda sheep breed of Asturias as an example. J. Anim. Breed. Genet. 120:95-105.

Grundy, B., B. Villanueva, and J. A. Woolliams. 1998. Dynamic selection procedures for constrained inbreeding and their consequence for pedigree development. Genet. Res. Camb. 72:159-168.

Hagger, C. 2002. Genetic variability of two Swiss sheep breeds derived from pedigree information. 7th World Congr. Genet. Appl. Livest. Prod. Aug. 19-23, Montpellier, France.

Klieve, H. M., B. P. Kinghorn, and S. A. Barwick. 1994. The joint regulation of genetic gain and inbreeding under mate selection. J. Anim. Breed. Genet. 111:81-88.

Leitch, H. W., C. Smith, E. B. Burnside, and M. Quinton. 1993. Genetic response and inbreeding with different selection methods and mating designs for nucleus breeding programs of dairy cattle. J. Dairy Sci. 77:1702-1718.

Lutaaya, E., I. Misztal, J. K. Bertrand, and J. W. Mabry. 1994. Inbreeding in populations with incomplete pedigrees. J. Anim. Breed. Genet. 116:475-480.

MacCluer, J. W., A. J. Boyce, B. Dyke, L. R. Wertkamp, D. W. Pfennig, and C. J. Parsons. 1983. Inbreeding and pedigree structure in standardbred horses. J. Hered. 74:394-399.

Meuwissen, T. H. E., and A. K. Sonesson. 1998. Maximizing the Response of Selection with a predefined rate of inbreeding: overlapping generations. J. Anim. Sci. 76:2575-2583.

Meuwissen, T. H. E., and Z. Luo. 1992. Computing inbreeding coefficients in large populations. Genet. Sel. Evol. 24:305-313.

Sanchez, L., M. A. Toro, and C. Garcia. 1999. Improving the efficiency of artificial selection: more selection pressure with less inbreeding. Genet. 151:1103-1114.

Shepherd, R. K., and B. P. Kinghorn, 1998. A tactical approach to the design of crossbreeding programs. 6th World Congr. Genet. Appl. Livest. Prod. Armidale, 11-16 January. 25:431-438.

Te Braake, M. F. H., A. F. Groen, and A. W. van der Lugt. 1994. Trends in inbreeding in Dutch Black and White dairy cattle. J. Anim. Breed. Genetics 111:356-366.

Tozer, P. R., and J. R. Stokes. 2002. Producer breeding objectives and optimal sire selection. J. Dairy Sci. 85:3518-3525.

VanRaden, P. M., and L. A. Smith. 1999. Selection and mating considering expected inbreeding of future progeny. J. Dairy Sci. 82:2771-2778.

Weigel, K. A., and S. W. Lin. 2000. Use of computerized mate selection programs to control inbreeding of Holstein and Jersey cattle in the next generation. J. Dairy. Sci. 83:822-828.

CHAPTER IV

ESTIMATES OF INBREEDING DEPRESSION FOR FEMALE REPRODUCTION AND LONG-TERM PRODUCTION IN RED ANGUS CATTLE FROM FIELD DATA

4.1. Abstract

The objective of this study was to estimate the effect of inbreeding on early female reproduction and long-term productivity in beef cattle. Performance and pedigree information on Red Angus cattle registered with the Red Angus Association of America were obtained from the Center for Genetic Evaluation of Livestock at Colorado State University. Three categorical traits considered were: heifer calving rate (HCR, n = 1,197); heifer calving ease (HCE, n = 636); and cow stayability (CS, n = 14,268). Inbreeding coefficients for the individual (F_d) and its dam (F_m) were computed using the Animal Breeder's Tool Kit (ABTK). Solutions for the effects of F_d and F_m on categorical performance were obtained by fitting single trait threshold models. The fixed effects in the model were contemporary group fitted as a class variable and F_d and F_m as covariates for all the traits. An additional covariate of age of the heifer at breeding was included for HCR. The random direct additive genetic effect of the animal was fitted for all the traits.

An additional correlated random maternal additive genetic effect was included for HCE. The partial regression of trait on F_d and F_m expressed on a probability scale were 0.75 and -0.40%/% for HCR, -1.44 and 0.20%/% for HCE, and -0.56 and 0.12%/% for CS. All partial regression coefficients were nonsignificant ($P > 0.10$) except for inbreeding of the cow for CS ($P = 0.0007$). The partial regression coefficient of CS on inbreeding of the cow indicated that a percentage increase in inbreeding of the cow was associated with a 0.56% decline in the ability of the cow to maintain productivity to or beyond six years of age. Regression estimates for F_d and F_m obtained in this study provide little evidence of unfavorable relationships between inbreeding and performance at least for the levels of inbreeding investigated, except for CS. However, further research is required to corroborate the evidence found in this study.

4.2. Introduction

The effect of inbreeding on performance in beef cattle has been extensively studied (Alexander and Bogart, 1961; Swiger et al., 1961; Brinks et al., 1963; Davenport et al., 1965; Dinkel et al., 1968; Krehbiel et al., 1969; Anderson et al., 1972; Dinkel et al., 1972; Brinks and Knapp, 1975; Bailey et al., 1977; MacNeil et al., 1989; Clarke, 1988; Snelling et al., 1996; Pariacote et al., 1998). The majority of these studies were conducted under experimental conditions whereby mating of closely related individuals were carried out to obtain highly inbred lines for future linecrossing to exploit hybrid vigor. In general, the results from these studies agree that inbreeding has a detrimental effect on fitness related traits in beef cattle. However, little has been done to investigate the impact of

inbreeding on performance in beef cattle under practical breeding conditions using field data. One explanation for this attitude is that breeding objectives are more variable in beef cattle such that the genetic pool is diverse and that the effects of inbreeding are limited to the seedstock which comprise a small proportion of the beef cattle industry. Also, the use of technologies such as artificial insemination is minimal in commercial beef cattle compared to the dairy cattle industry for instance.

For the past twenty to twenty-five years, best linear unbiased prediction (BLUP) of breeding values based on animal models have been available to breeders. The BLUP procedure uses information on relatives to compute the genetic merit of an individual by incorporating genetic relationships among animals. For traits that are lowly heritable, information from relatives plays an important role in the genetic evaluation of an individual. This results in correlation between estimated breeding values (EBV) of relatives. When these EBV are used for selection, the probability of co-selection of close relatives is increased.

The objective of this study was to obtain estimates of inbreeding depression for early female reproduction and long-term productivity in the seedstock sector of the beef cattle industry in the U.S. using the registered population of Red Angus cattle.

4.3. Materials and Methods

Data and traits: Performance and pedigree records of cattle registered with the Red Angus Association of America (RAAA) were obtained from the Center for Genetic Evaluation of Livestock at Colorado State University. The RAAA requires breeders to

report performance and production information on all animals in the herd. This provides an opportunity for reliable evaluation of the economically important trait, such as reproduction.

Performance data included information on early female reproduction and long-term productivity from ten prominent breeders. The traits analyzed were heifer calving rate (HCR), heifer calving ease (HCE), and cow stayability (CS). These traits were expressed on binary scales with success denoted as a "1" and failure a "0". Heifer calving rate was defined as the presence or absence of a calf at birth with heifers giving birth to a calf assigned a score of "1" and "0" otherwise. This trait is a measure of the potential of the heifer to conceive, maintain pregnancy and give birth to a live calf. Heifer calving ease referred to whether a heifer that calved required assistance or not during calving. Heifers experiencing no calving problem were given a score of "1" and those that required minor to major assistance were given a score of "0". Cow stayability was defined as an observation of the cow being in the herd at or beyond six years of age. This trait is a measure of the long-term productivity of the cow under the prevailing production conditions.

Information on contemporary group and age at breeding were available. The contemporary group for HCR was a concatenation of heifer yearling and breeding management group while that for HCE consisted of birth contemporary group and sex of the calf. The contemporary group for CS comprised of herd(s) where each of the cow's calves were born. Based on the definition of stayability in this study, the cow will have a maximum of five calves when the stayability observation is assigned. Therefore, the maximum number of herds in a contemporary group is five if the cow calved every year

until six years of age. Thus, the creation of the contemporary group for CS takes into account the fact that a cow may have calves in more than one herd. Records from contemporary groups with no variation in observed scores contribute nothing to a threshold analysis and were excluded. The records remaining for analyses were 1,197, 636, and 14,268 for HCR, HCE, and CS, respectively. These records were collected during the periods 1989-2000, 1994-2001, and 1950-1996 for HCR, HCE, and CS, respectively.

Descriptive statistics about the data are given in Table 4.1. The incidence of heifer calving success was 72% and that of calving ease was 71%. The proportion of cows that were still productive by the age of six was 50%. On average, there were 11 individuals within a contemporary group for HCR. The mean contemporary group sizes were higher for HCE and CS, 23 and 35 individuals per contemporary group respectively, than HCR. The average age at breeding of the heifer was 402 days or about 13 months of age. There was minimal variation in age at which heifers were bred.

Inbreeding coefficients for the individuals with performance records were computed based on the whole breed pedigree records (n = 829,882) using the Animal Breeder's Tool Kit (ABTK: Golden et al., 1992). As expected, the average inbreeding of the heifer or cow tended to be higher than that of their dam across the traits. The ranges of inbreeding coefficients for heifer and dam were similar for HCR and HCE at about 0-20%. Inbreeding coefficients were more variable for CS compared to HCE and HCR.

Table 4.1. Descriptive statistics of the data.

Trait	Variable	N	Mean	Min	Max	SD	CV(%)
HCR,%		1,197	72				
	age,d		402	341	484	22	6
	F_d,%		5	0	20	3	62
	F_m,%		4	0	19	4	108
	CG	107	11	2	83	17	155
HCE,%		636	71				
	F_d,%		6	0	18	3	51
	F_m,%		5	0	20	3	59
	CG	28	23	3	45	13	57
CS,%		14,268	50				
	F_d,%		3	0	32	4	132
	F_m,%		2	0	34	4	160
	CG	413	35	2	385	58	167

HCR, HCE and CS refer to heifer calving rate, heifer calving ease and cow stayability, respectively. F_d and F_m denote inbreeding coefficient of the individual and its dam, respectively, CG – contemporary group.

Statistical Analyses: The traits were analyzed using a single trait maximum *a posteriori* probit (MAP) threshold model (Gianola and Foulley, 1983; Harville and Mee, 1984). The threshold model assumes a conceptual underlying normally distributed variable influenced by genetic and environmental effects (Wright, 1934). The unobservable underlying normal variable (y^*) relates to the observed categories (y) through the thresholds (τ_k) that divides the underlying variable into discrete classes hence threshold model. In other words, the underlying normal variable has a joint

distribution with the observed discrete variable (Gianola, 1982). For a binary case, where $y = 0$ or 1 ("failure" or "success") with a single threshold (τ) separating the two categories, the relationship between the observed and unobserved variable for the jth animal or experimental unit can be represented mathematically as follows:

$$y_j = \begin{cases} 1 & \text{if } y_j^* > \tau \\ 0 & \text{otherwise} \end{cases}$$

In addition, it is assumed that $y_j^* \sim N(\eta_j, \sigma_e^2)$ such that the probability that $y_j = 1$ given η_j (the mean of the jth animal on the underying scale) is:

$$\begin{aligned} \Pr(y_j = 1) &= \Pr(y_j^* > \tau \mid \eta_j) \\ &= 1 - \Pr(y_j^* < \tau \mid \eta_j) \\ &= 1 - \Phi\left(\frac{\tau - \eta_j}{\sigma_e}\right) \\ &= \Phi(\eta_j) \text{ if } \tau = 0 \text{ and } \sigma_e^2 = 1 \end{aligned}$$

where $\Phi(.)$ is a standard normal cumulative distribution function and η_j is a linear combination of fixed and random effects such that $y_j^* = \eta_j + e_j = \mathbf{x}'\mathbf{b} + \mathbf{z}'\mathbf{u} + e_j$ (Gianola and Foulley, 1983) where $\mathbf{x}$ and $\mathbf{z}$ are the first columns of the usual $\mathbf{X}$ and $\mathbf{Z}$ incidence matrices relating the fixed and random effects to $\mathbf{y}^*$, respectively; $\mathbf{b}$ and $\mathbf{u}$ are vectors of fixed and random effects, respectively and e_j is a residual for the jth animal or experimental unit independently and identically distributed (i.e. $e_j \sim N(0, \sigma_e^2)$).

In this study, where maternal genetic effects were included, the unobservable underlying variable was described using the following model equation:

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{m} + \mathbf{e}$$

where $\mathbf{y}^*$ is a vector of the unobserved underlying normal variable, $\boldsymbol{\beta}$ is a vector of unknown fixed effects, $\mathbf{u}$ is a vector of unobservable random direct additive genetic effect of the animal, $\mathbf{m}$ is a vector of unobservable random maternal additive genetic effect of the animal, $\mathbf{X}$ is an incidence matrix relating fixed effects in $\boldsymbol{\beta}$ to the underlying variable in $\mathbf{y}^*$, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are the incidence matrices relating the direct and maternal additive genetic effects, respectively, to the observations in $\mathbf{y}^*$, $\mathbf{e}$ is a vector of unobservable random residual effects on $\mathbf{y}^*$. The first moments of the random effects in the model were assumed to be:

$$\mathrm{E}(\mathbf{u}) = \mathbf{0},\ \mathrm{E}(\mathbf{m}) = \mathbf{0},\ \mathrm{E}(\mathbf{e}) = \mathbf{0}$$

and as a consequence

$$\mathrm{E}(\mathbf{y}^*) = \mathbf{X}\boldsymbol{\beta}$$

The second moments were:

$$\text{Var}\begin{pmatrix}\mathbf{u}\\ \mathbf{m}\\ \mathbf{e}\end{pmatrix}=\begin{pmatrix}\mathbf{A}\sigma_u^2 & \mathbf{A}\sigma_{u,m} & \mathbf{0}\\ \mathbf{A}\sigma_{m,u} & \mathbf{A}\sigma_m^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}\end{pmatrix}$$

where $\mathbf{A}$ is the Wright's numerator relationship matrix, σ_u^2 and σ_m^2 are the variances associated with the direct and maternal additive genetic effects of the animal and $\sigma_{u,m}$ is the covariance between direct and maternal additive genetic effects of the animal. The additive genetic effects were assumed to be uncorrelated with the residual effects as shown in the variance-covariance matrix (i.e. $\text{cov}(\mathbf{u},\mathbf{e}) = \mathbf{0}$ and $\text{cov}(\mathbf{m},\mathbf{e}) = \mathbf{0}$). The covariance matrix of the residual effects is an identity ($\mathbf{I}$) because the residual effects were assumed to be independently and identically distributed and the residual variance is arbitrarily set to 1 under the MAP threshold model (i.e. $\mathbf{I}\sigma_e^2 = \mathbf{I}$).

The variance-covariance components (expressed on the underlying scale) required to set-up and solve the mixed model equations were those used in the Red Angus national genetic evaluation as shown below:

Trait	σ_u^2	σ_m^2	$\sigma_{u,m}$	σ_e^2
HCS	0.37	-	-	1.00
HCE	0.19	0.20	-0.03	1.00
CS	0.11	-	-	1.00

The vector of the unobserved underlying variate ($\mathbf{y}^*$) is computed at each round of iteration from previous estimates of fixed and random effects plus an estimate of the vector of residual effects ($\hat{\mathbf{e}}$). The variance of the estimate is a function of the fixed and

random effects and must therefore be computed at each round of iteration. Since the underlying variance of σ_e^2 is arbitrarily set to 1, the variance of estimates is really a weight matrix ($\mathbf{W}$). The left-hand side of the system of mixed model equations must be updated at each round of iteration to reflect the current estimate of this weight. Explicit expressions for $\mathbf{y}^*$ and $\hat{\mathbf{e}}$ are given in Hoeschele (1988). Briefly, the $\mathbf{y}^*$ at the k^{th} round of iteration is given by $\mathbf{y}_k^* = \mathbf{X}\hat{\mathbf{b}}_{k-1} + \mathbf{Z}\hat{\mathbf{u}}_{k-1} + \hat{\mathbf{e}}_k$ where $\hat{\mathbf{e}}_k = \mathbf{W}_k\mathbf{v}_k$, $\mathbf{W}_k$ is a diagonal matrix of weights at the k^{th} round whose jth diagonal element is

$$w_{j,j} = n_j\left[\frac{\phi^2(\mu_j)}{\Phi(\mu_j)\left(1-\Phi(\mu_j)\right)}\right]$$

and the jth element of $\mathbf{v}_k$ is

$$v_j = \phi(\mu_j)\left[\frac{n_{j1} - n_j\Phi(\mu_j)}{\Phi(\mu_j)\left(1-\Phi(\mu_j)\right)}\right]$$

and $\mu_j = \tau - \eta_j = \mathbf{x}'\hat{\mathbf{b}} + \mathbf{z}'\hat{\mathbf{u}}$; $\phi(.)$ is the standard normal density function, n_{j1} and n_j are the number of observations in category 1 and total number of observations for the jth experimental unit, respectively. The numerator of v_j is the deviation of the observations in category one from the expected number given the mean of jth experimental unit on the underlying scale (η_j).

The fixed effects included in the model for HCR were contemporary group, age of the heifer at calving and inbreeding of the individual and its dam. The effects of age and inbreeding were fitted as covariates. Contemporary group and inbreeding coefficient of the individual and dam were included in the model for HCE and CS. Random effects fitted in the model for HCR and CS were direct additive genetic effects of the heifer/cow while an additional correlated random maternal additive genetic effect was considered for HCE.

Pedigree data included individuals with performance information, their parents and grandparents. The pedigree information was used in creating the inverse of the numerator relationship matrix required in assembling the mixed model equations. Solutions of the effects in the model were obtained by solving the nonlinear mixed model equations using the *dscat* tool of the ABTK (Golden et al., 1992). Solutions were expressed as a deviation from the threshold. For a binary trait it is convenient to constrain the threshold solution to zero since there is only one threshold dividing the two categories. For ease of interpretation, the solution for partial regression coefficients on age and inbreeding were transformed to a probability scale. This was achieved by calculating the probabilities associated with the partial regression coefficients using the "probnorm" function of SAS (SAS Inst. In., Cary, NC) and expressing these probabilities as a deviation from 50%, a probability associated with the threshold solution of zero obtained from a standardized normal distribution.

Test of significance for the fixed effects were conducted following the procedure described by Boik et al. (1993). This procedure requires a solution to the mixed model equation to compute a likelihood ratio F-test. Relationships among animals can be

incorporated in this procedure. The F-value was calculated using the solutions of the mixed model including and excluding the variable of interest. The probabilities (p-values) associated with the calculated F-value were obtained from the F-distribution using the "probf" function of SAS (1999). Standard errors of the partial regression coefficients were obtained using the diagonal elements of the inverse of the coefficient matrix.

4.5. Results and Discussion

Estimates of the effect of inbreeding and age for the traits evaluated and test of significance for the fixed effects are presented in Tables 4.2, 4.3, and 4.4. The contemporary group effect was highly significant ($P < 0.0001$) for all the traits. Given that the traits considered are lowly to moderately heritable, it was expected that the differences in environmental factors (e.g. availability of feed and management) as confounded in the contemporary group will account for a major proportion of the observed variation in reproduction and long-term productivity. For example, heifers raised under "good" nutritional conditions are likely to have "better" reproductive performance relative to those raised under "poor" nutritional regimes.

The effect of age of the heifer at breeding did not significantly ($P = 0.1049$) influence HCR. However, there was a tendency for an increase (0.28%/day) in HCR as the age of the heifer increased. A physiological explanation for these results is that older heifers are more sexually mature than young heifers at a given age. That is, older heifers tend to have more opportunities for conception compared to young heifers. Based on the range in the age at breeding in this study, the oldest heifer was 82 days older than the

average age in this population. According to the partial regression of HCR on age, this heifer was 20% more likely to calve compared to the average of the population. Thus, the magnitude of the influence of age on HCR may be biologically important though it was found to be statistically unimportant. Evans (1996) reported a significant increase in the probability of conceiving and maintaining pregnancy to palpation at 120d in Hereford cattle. The estimate of the effect was about double that observed in this study.

Table 4.2. Partial regressions of heifer calving rate on age and inbreeding of the heifer (F_d) and its dam (F_m) expressed on the probability and underlying scale with the associated standard errors.

Variable	Partial regression coefficient		P-value
	Probability scale (%)	Underlying scale±SE (RSD)	
F_d, %	0.750	0.019±0.020	0.5661
F_m, %	-0.403	-0.010±0.016	0.3980
Age, d	0.275	0.007±0.004	0.1049
CG			<0.0001

CG – contemporary group, RSD – residual standard deviation, P-value is the probability associated with the observed F-value.

Inbreeding of the heifer and its dam did not significantly ($P > 0.10$) influence HCR (Table 4.2). However, the partial regressions suggest an increase in HCR with inbreeding of the heifer and a decline with the inbreeding of the dam. A non-significant effect of inbreeding of the heifer was consistent with the results reported by Dinkel et al. (1972). In contrast, MacNeil et al. (1989) observed a significant decline in calving rate with increasing level of inbreeding of the dam. However, the average inbreeding of the dam was considerably higher than in this study. A tendency for decline in HCR with

inbreeding of the dam has also been reported in other studies (Davenport et al., 1965; Bailey et al., 1977; Krehbiel et al., 1996).

Table 4.3. Partial regressions of heifer calving ease on inbreeding of the heifer (F_d) and its dam (F_m) expressed on the probability and underlying scale with the associated standard errors.

Variable	Partial regression coefficient		P-value
	Probability scale (%)	Underlying scale±SE (RSD)	
F_d, %	-1.436	-0.036±0.026	0.1603
F_m, %	0.200	0.005±0.022	0.8615
CG			<0.0001

CG – contemporary group, RSD – residual standard deviation, P-value is the probability associated with the observed F-value.

Results in Table 4.3 show that the inbreeding of the heifer and dam did not significantly ($P > 0.10$) affect HCE. The partial regression of HCE on inbreeding of the heifer can be interpreted to mean that each percentage increase in inbreeding of the heifer was associated with a 1.44% increase in the probability of experiencing calving problems. A heifer that is 20% inbred is 29% more likely to experience calving problems compared to a noninbred heifer. This indicates that an increase in the probability of calving difficulty may be substantial for individual heifers. When the entire population is considered (i.e. average F_d of 6% in this study), the incidence of dystocia associated with inbreeding is increased by about 9% in comparison to non-inbred population.

Inbreeding of the cow significantly ($P = 0.0007$) reduced long-term productivity of the cow. On the other hand, inbreeding of the dam did not significantly ($P > 0.10$) influence CS. The most inbred cow had an inbreeding coefficient of 32%. Based on the

partial regression coefficient of CS on inbreeding of the cow, the probability of this cow to maintain productivity until or beyond six years of age was reduced by 18% compared to a non-inbred cow. When the entire population was considered, the reduction in CS was about 2%. Even though the effect of inbreeding of the cow on CS was found to be statistically important, the magnitude of this effect is too small to be of major importance from a practical standpoint. The larger sample size for CS relative to HCR and HCE may have contributed in the significant results for inbreeding of the cow on CS.

Table 4.4. Partial regressions of cow stayability on inbreeding of the cow (F_d) and its dam (F_m) expressed on the probability and underlying scale with the associated standard errors.

Variable	Partial regression coefficient		P-value
	Probability scale (%)	Underlying scale±SE (RSD)	
F_d, %	-0.559	-0.014±0.005	0.0007
F_m, %	0.120	0.003±0.004	0.5198
CG			<0.0001

CG - contemporary group, RSD - residual standard deviation, P-value is the probability associated with the observed F-value.

4.6. Conclusions

The results obtained in this study provide little evidence of unfavorable effects of inbreeding on early reproduction. However, there was a significant reduction in long-term productivity associated with the inbreeding of the cow. It was shown that the effects of inbreeding could be substantial at an individual level. However, when the entire

population was considered the impact of inbreeding tended to be minimal at current levels of inbreeding in the Red Angus population.

4.7. Literature Cited

Alexander, G. I., and R. Bogart. 1961. Effect of inbreeding and selection on performance characteristics of beef cattle. J. Anim. Sci. 20:702-707.

Anderson, D. C., A. E. Flower, F. S. Willson, and C. Windecker. 1972. Factors affecting production in beef cattle. Proc. Western Section, American Society of Animal Science 23:6-11.

Bailey, C. M., J. A. Edwards, and Y. O. Koh. 1977. F_1 cross between mildly inbred Hereford selection lines of common genetic origin. J. Anim. Sci. 44:23-29.

Brinks, J. S., and B. W. Knapp. 1975. Effects of inbreeding on performance traits in beef cattle in the western region. Colorado State Univ. Exp. Sta. Tech. Bull. 123.

Brinks, J. S., R. T. Clark, and N. M. Kieffer. 1963. Sex differences in response to inbreeding in a line of Hereford cattle. Proc. Western Section, American Society of Animal Science ppV1-V6.

Clarke, L. S. 1988. Effect of inbreeding on performance traits in Hereford cattle. Ph.D. Dissertation. Colorado State University.

Davenport, R. L., H. H. Stonaker, K. Riddle, and T. M. Sutherland. 1965. Heritability of reproductive performance in inbred and linecross beef cows. J. Anim. Sci. 24:434-437.

Dinkel, C. A., D. A. Busch, J. A. Minyard, and W. R. Trevillyan. 1968. Effects of inbreeding on growth and conformation of beef cattle. J. Anim. Sci. 27:313-322.

Dinkel, C. A., L. M. Anderson, W. R. Parker, and W. R. Trevillyan. 1972. Effects of inbreeding on fertility and livability in beef cattle. J. Anim. Sci. 35:725-729.

Evans, J. L. 1996. Genetic parameters estimates of yearling heifer pregnancy and yearling bull scrotal circumference in Hereford cattle. MSc Thesis. Colorado State University.

Gianola, D. 1982. Theory and analysis of threshold characters. J. Anim. Sci. 54:1079-1095.

Gianola, D., and J. L. Foulley. 1983. Sire evaluation for ordered categorical data with a threshold model. Genet. Sel. Evol. 15:201-223.

Harville, D. A., and R. W. Mee. 1984. A mixed-model procedure for analyzing ordered categorical data. Biometrics 40:393-408.

Hoeschele, I. 1988. Comparison of "Maximum A-Posteriori Estimation" and "Quasi Best Linear Unbiased Prediction" with threshold characters. J. Anim. Breed. Genet. 105:337-361.

Krehbiel, E. V., R. C. Carter, K. P. Bovard, J. A. Gaines, and B. M. Priode. 1969. Effects of inbreeding and environment on fertility of beef cattle matings. J. Anim. Sci. 29:528-533.

MacNeil, M. D., D. D. Dearborn, L. V. Cundiff, C. A. Dinkel, and K. E. Gregory. 1989. Effects of inbreeding and heterosis in Hereford females on fertility, calf survival and preweaning growth. J. Anim. Sci. 67:895-901.

MacNeil, M. D., J. J. Urick, S. Newman, and B. W. Knapp. 1992. Selection for postweaning growth in inbred Hereford cattle: The Fort Keogh, Montana Line 1 Example. J. Anim. Sci. 70:723-733.

Pariacote, F., L. D. Van Vleck, and M. D. MacNeil. 1998. Effects of inbreeding and heterozygosity on preweaning traits in a closed population of Herefords under selection. J. Anim. Sci. 76:1303-1310.

Snelling, W. M., M. D. MacNeil, D. D. Kress, D. C. Anderson, and M. W. Tess. 1996. Factors influencing genetic evaluations of linebred Hereford cattle in diverse environments. J. Anim. Sci. 74:1499-1510.

Swiger, L. A., K. E. Gregory, R. M. Koch, and V. A. Arthaud. 1961. Effect of inbreeding on performance traits of beef cattle. J. Anim. Sci. 20:626-630.

Wright, S. 1934. An analysis of variability in number of digits in an inbred strain of guinea pigs. Genetics 19:506-536.

CHAPTER V

ESTIMATES OF INBREEDING DEPRESSION FROM FIXED AND MIXED MODELS FOR CARCASS TRAITS IN RED ANGUS CATTLE

5.1. Abstract

The objectives of this study were to 1) obtain estimates of inbreeding depression for carcass traits in beef cattle, and 2) to evaluate the effect of ignoring the random additive genetic effect of the animal on estimates of inbreeding depression. Pedigree and performance records of cattle registered with the Red Angus Association of America were used in the analysis. Carcass data were available on 951 animals collected between the years 1997 and 2002. Pedigree information of these animals dated as far back as the early 1930s. Inbreeding coefficients were computed for each animal using the Animal Breeder's Tool Kit. Traits analyzed were hot carcass weight (HCW, n = 951), rib-eye area (REA, n = 947), backfat thickness (BFT, n = 767), and marbling score (MRB, n = 947). Two models were used to analyze the data. The first analysis involved fitting a fixed effects model to carcass observations. The fixed effects included in the model were age at slaughter, inbreeding of the individual (F_d) and its dam (F_m), and contemporary group. In the second analysis, a linear mixed model was used to analyze the data. The random effect considered was the animal additive genetic effect. The variance

components used in the analysis were those assumed in the Red Angus national genetic evaluation. The partial regressions on F_d and F_m were essentially the same from fitting a fixed and mixed model. This result suggests that there was no confounding between level of inbreeding and additive genetic merit of the individuals in the population studied. The estimates of the effect of F_d and F_m from the mixed model were -0.71 and -0.45 kg/%, -0.04 and -0.12 cm^2/%, 0.0005 and -0.0012 cm/%, and -0.011 and -0.003/% for HCW, REA, BFT, and MRB respectively. The only significant partial regression coefficients were those of HCW on F_d ($P = 0.0238$) and REA on F_m ($P = 0.0111$). Results from this study indicate that there is little evidence of unfavorable relationships between inbreeding of either the individual or its dam with carcass traits for the levels of inbreeding investigated except for HCW and REA.

5.2. Introduction

Research on the effects of inbreeding in beef cattle has been conducted mainly in traits associated with growth and reproduction (Swiger et al. 1961; Brinks et al., 1963; Dinkel et al. 1968; Dinkel et al., 1972; Brinks and Knapp, 1975; MacNeil et al., 1989; Clarke, 1988; Snelling et al., 1996; Pariacote et al., 1998). In general, results from these studies are in agreement in estimating unfavorable effects of inbreeding on growth and reproduction. To the author's knowledge no information is available on the effects of inbreeding on carcass traits in beef cattle. The reason for this is that commercial animals in the beef cattle industry are typically crossbred and therefore the impact of inbreeding is minimal if present at all in the beef industry. According to Koch et al. (1986) about 70%

of the animals slaughtered for beef in the U.S. were crossbreds. Gama and Smith (1993) investigated the role of inbreeding in livestock production systems from an industry-wide perspective and concluded that the impact of inbreeding in beef cattle is minimal. Although the effect of inbreeding on the carcass may be avoided by crossbreeding, knowledge of the effect of inbreeding may be useful at the seedstock level to allow for adjustment of this effect in the genetic evaluation process.

Except for current studies (MacNeil et al., 1989; Snelling et al., 1996; Pariacote et al., 1998) estimates of the effects of inbreeding have been obtained within sire by year subclasses. Current developments in statistical methodology and computing power allows for estimates of inbreeding to be obtained from complex models (e.g. mixed models).

The objectives of this study were to 1) estimate the effect of inbreeding on carcass traits in beef cattle using Red Angus cattle data, and 2) to evaluate the effect of ignoring the random additive genetic effect of the animal on estimates of inbreeding depression.

5.3. Materials and Methods

Data and traits. Performance and pedigree records of cattle registered with the Red Angus Association of America (RAAA) were obtained from the Center for Genetic Evaluation of Livestock (CGEL) at Colorado State University. Performance data were collected between the years 1997 and 2002 while pedigree records dated as far back as the early 1930s. The pedigree information on each animal included the animal, its sire and dam identification numbers and was used to compute inbreeding coefficients for

individuals with performance records using the ainv tool of Animal Breeder's Tool Kit (ABTK: Golden et al., 1992).

Traits evaluated were hot carcass weight (HCW), rib-eye area (REA), backfat thickness (BFT), and marbling score (MRB). These traits were defined as follows: HCW is the weight of the carcass measured 24-h after slaughter; REA is a cross-sectional area of the longissimus muscle at the 12^{th}-13^{th} rib interface; BFT is a measurement taken at a point ¾ of the length of the rib-eye from its medial edge; and MRB is a measure of the degree of marbling at the 12^{th}-13^{th} rib interface expressed as a score from 1 to 10. Phenotypic measurements on these traits were available on 951 individuals from ten prominent breeders that were intensively fed prior to slaughter at the end of the feeding period. However, observations for REA, BFT and MRB were not available on all individuals as shown in Table 5.1.

Descriptive statistics of the data is given in Table 5.1. The mean age at slaughter was about 16 months of age with a range of 12 to 20 months. The average inbreeding of the individual and its dam were 6 and 5% respectively. Inbreeding level was more variable than age of the individual at slaughter. On average, there were 63 individuals within a contemporary group. Contemporary group consisted of individuals from the same breeder that were slaughtered on the same date.

Table 5.1. Descriptive statistics of the data.

Trait	N	Mean	Min.	Max.	SD	CV(%)
HCW, kg	951	339	187	447	38	11
REA, cm^2	947	79	54	115	8	10
BFT, cm	767	1.2	0.2	2.3	0.4	31
MRB	947	5.0	1.0	8.9	0.9	17
CG	15	63	12	159	41	65
age, d		479	366	609	42	9
F_d, %		6	0	25	4	68
F_m, %		5	0	20	4	69

HCW, REA, BFT, and MRB refer to hot carcass weight, rib-eye area, backfat thickness, and marbling score, respectively. F_d and F_m – inbreeding coefficient of the individual and its dam, respectively, CG – contemporary group.

Statistical analyses. Two sets of analyses were performed to obtain estimates of inbreeding depression. In the first analysis a fixed effects model was fitted to the carcass observations. The fixed effects model equation was as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}$$

where $\mathbf{y}$ is a vector of trait phenotypes, $\mathbf{b}$ is a vector of unknown fixed effects of age, inbreeding of the individual and its dam, and contemporary group, $\mathbf{X}$ is an incidence matrix relating the fixed effects to trait phenotypes, $\mathbf{e}$ is a vector of random residual effects unique to each animal assumed to be independently and identically normally distributed i.e. $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$. As a consequence the first moment of the observations is $\mathrm{E}(\mathbf{y}) = \mathbf{Xb}$.

In the second analysis traits were analyzed using a single trait linear mixed model. The difference between the fixed effects model and the mixed model is that a random additive genetic effect is included in the mixed model and that relationships between animals were taken into account such that $\text{var}(\mathbf{y}) = \mathbf{ZGZ}' + \mathbf{I}\sigma_e^2$ in the mixed model instead of $\text{var}(\mathbf{y}) = \mathbf{I}\sigma_e^2$ in the fixed effects model. The model equation describing the trait phenotypes can be expressed in matrix notation as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

where $\mathbf{y}$, $\mathbf{X}$, $\mathbf{b}$ and $\mathbf{e}$ are as defined before, $\mathbf{u}$ is a vector of unobservable random direct additive genetic effects of the animal, $\mathbf{Z}$ is an incidence matrix relating the random animal additive genetic effects to trait phenotypes. The assumptions are the same as that of the fixed effects model except that the first moment of the random genetic effect is:

$$\text{E}(\mathbf{u}) = \mathbf{0}$$

The second moments are:

$$\text{var}\begin{pmatrix}\mathbf{u}\\ \mathbf{e}\end{pmatrix} = \begin{pmatrix}\mathbf{A}\sigma_u^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\sigma_e^2\end{pmatrix}$$

where **A** is the Wright's numerator relationship matrix, **I** is an identity matrix representing the relationship between residual effects, σ_u^2 and σ_e^2 are the variances of the random additive genetic and residual effects, respectively. The variances of additive genetic and residual effects assumed in this study were those used in Red Angus national genetic evaluation. These variances were as follows:

Trait	σ_u^2	σ_e^2
HCW	261.84	671.80
REA	16.16	37.09
BFT	0.02	0.08
MRB	0.23	0.39

Pedigree records used to obtain the inverse of the numerator relationship matrix required in setting up the mixed model equations included the individual, its parents and grandparents totaling 2,874 for HCW, 2,858 for REA, 2,479 for BFT and 2,858 for MRB. The inverse of the numerator relationship matrix was obtained using the ABTK. The mixed model equations were assembled and solved iteratively using the ABTK software. The convergence criterion was the absolute value of the difference of the average of the solutions between successive rounds and the solutions were considered converged when the sum of squares of the difference was less than 1.0e-12.

The effects of age and inbreeding of the individual and its dam were fitted as covariates while contemporary group was considered a class variable. Tests of significance for fixed effects were conducted following the procedure described by Boik et al. (1993). This procedure requires a solution to the mixed model equations to compute

likelihood ratio F-test. This procedure was preferred because the significance test for fixed effects is conducted taking into account the random effects in the model. In addition, relationships among animals are incorporated in the mixed model equations, which should increase the accuracy of predicting the random genetic effects and thereby reduce the standard error associated with the fixed effects solutions. The probabilities (p-values) associated with the calculated F-values were obtained from the F-distribution using the "probf" function of SAS (SAS Inst. In., Cary, NC). Standard errors of the partial regression coefficients of age and inbreeding were obtained by squaring the product of the diagonal elements of the inverse of the coefficient matrix and the mean square error.

5.4. Results and Discussion

The estimates of the effects of age at slaughter and inbreeding of the individual and its dam from the fixed and mixed models are given Tables 5.2 and 5.3.

Effect of model on estimates. Estimates of the effects of inbreeding of the individual and its dam obtained from fitting the fixed effects model were essentially the same as those from the mixed model for all the traits. This result suggests that there was no association between inbreeding coefficient and the additive genetic merit of the individual in this study. However, the standard errors of the estimates from the mixed model tended to be lower than those from the fixed effects model. Therefore, accounting for the random additive genetic effect results in the reduction in the sampling variance of the estimates.

Age at slaughter. Results in Table 5.2 provide evidence of an association between age at slaughter and HCW but not REA, BFT and MRB. An increase in age was associated with a significant ($P = 0.0238$) increase in HCW. The partial regression of HCW on age was 0.12kg/d indicating that an increase in age at slaughter of ten days was associated with a 1.2kg increase in HCW. The significant effect of age at slaughter on HCW and not other carcass traits is in agreement with the results reported by Pariacote et al. (1998) in Shorthorn cattle.

Inbreeding of the individual. Inbreeding of the individual did not have a significant effect on carcass traits with the exception of HCW ($P = 0.0111$) (Tables 5.2 and 5.3). A percentage increase in the inbreeding of the individual was associated with 0.71kg reduction in HCW. Dinkel et al. (1968) reported a significant reduction in final weight on test of bull and heifer calves in Hereford cattle associated with inbreeding of the individual. For the purpose of comparison with the results in the current study, partial regression coefficients from Dinkel et al. (1968) were expressed on a carcass weight basis. Assuming a dressing percentage of 60%, the partial regression on inbreeding of the individual for bull calves was similar (-0.66kg/%) to the estimate obtained in this study and that for heifer was considerably lower at -0.30kg/%. Anderson et al. (1972) found no evidence of the effect of inbreeding of the individual on final weight. The partial regressions on inbreeding of the individual obtained by Anderson et al. (1972) were lower (-0.37 and -0.10kg/% in bull and heifer calves respectively on a carcass weight basis) than that of Dinkel et al. (1968) and estimates from the current study.

Table 5.2. Partial regressions of hot carcass weight (HCW) and rib-eye area (REA) on age or inbreeding of the individual (F_d) and its dam (F_m) from the fixed and mixed models.

Trait	Variable	Partial regression coefficient		P-value
		Fixed Model	Mixed Model	
HCW				
	age, d	0.119±0.049	0.117±0.044	0.0238
	F_d, %	-0.703±0.266	-0.707±0.239	0.0111
	F_m, %	-0.446±0.282	-0.449±0.246	0.1168
REA				
	age, d	0.011±0.012	0.011±0.010	0.3491
	F_d, %	0.003±0.062	-0.035±0.055	0.5852
	F_m, %	-0.114±0.066	-0.117±0.057	0.0766

P-value is the probability associated with the observed F-value for the mixed model.

Inbreeding of the dam. Inbreeding of the dam significantly ($P = 0.0766$) reduced REA and not HCW, BFT, and MRB (Tables 5.2 and 5.3). However, there was a tendency for HCW to decrease with inbreeding of the dam. Similarly, Anderson et al. (1972) reported a tendency of a decline in final weight associated with inbreeding of the dam. Dinkel et al. (1968) observed a significant reduction in final weight of heifer calves but not bull calves in Hereford cattle.

The partial regression coefficient of REA on inbreeding of the dam indicates that a 1% increase in the inbreeding of the dam was associated with a 0.12 cm^2 centimeters decrease in REA. For an individual that is 20% inbred, this represents a 2.4 cm^2 decrease in REA. Given that the mean inbreeding level of the dam in the current study is 5%, the

mean reduction in REA due to inbreeding of the dam amounts to 0.6 cm^2. From a practical point of view given the REA of 79 cm^2 the reduction in REA associated with inbreeding at a population level may not be of much importance.

Table 5.3. Partial regressions of backfat thickness (BFT) and marbling score (MRB) on age or inbreeding of the individual (F_d) and its dam (F_m) from the fixed and mixed models.

Trait	Variable	Partial regression coefficient		P-value
		Fixed Model	Mixed Model	
BFT				
	age, d	-0.0002±0.0005	-0.0002±0.0005	0.7483
	F_d, %	0.0007±0.0028	0.0005±0.0026	0.8631
	F_m, %	-0.0014±0.0029	-0.0012±0.0026	0.6978
MRB				
	age, d	0.0008±0.0013	0.0010±0.0012	0.4457
	F_d, %	-0.0175±0.0132	-0.0107±0.0057	0.1246
	F_m, %	-0.0054±0.0137	-0.0033±0.0058	0.6413

P-value is the probability associated with the observed F-value for the mixed model.

5.5. Conclusions

The results from the current study provide evidence of an association between inbreeding of the individual and its dam on HCW and REA for the levels of inbreeding occurring in the population under consideration. However, no evidence of the effect of inbreeding of either the individual or the dam was found for BFT and MRB. Therefore,

for traits such as HCW and REA inbreeding of the individual and its dam should be taken into account in genetic evaluation of animals.

5.6. Literature Cited

Anderson, D. C., A. E. Flower, F. S. Willson, and C. Windecker. 1972. Factors affecting production in beef cattle. Proc. Western Section, American Society of Animal Science 23:6-11.

Boik, R. J., M. W. Tess, and C. Todd. 1993. Technical Note: Computing tests of fixed effects in a restricted class of mixed models. J. Anim. Sci. 71:51-56.

Brinks, J. S., and B. W. Knapp. 1975. Effects of inbreeding on performance traits in beef cattle in the western region. Colorado State Univ. Exp. Sta. Tech. Bull. 123.

Clarke, L. S. 1988. Effect of inbreeding on performance traits in Hereford cattle. Ph.D. Dissertation. Colorado State University.

Dinkel, C. A., D. A. Busch, J. A. Minyard, and W. R. Trevillyan. 1968. Effects of inbreeding on growth and conformation of beef cattle. J. Anim. Sci. 27:313-322.

Dinkel, C. A., L. M. Anderson, W. R. Parker, and W. R. Trevillyan. 1972. Effects of inbreeding on fertility and livability in beef cattle. J. Anim. Sci. 35:725-729.

Gama, L. T., and C. Smith. 1993. The role of inbreeding depression in livestock production systems. Livest. Prod. Sci. 36: 203-211.

Golden, B. L., W. M. Snelling, and C. H. Mallinckrodt. 1992. Animal breeder's tool kit user's guide and reference manual. Colorado State Univ. Agric. Exp. Sta. Tech. Bull. LTB92-2.

Koch, R.M., J. B. Gibb, and J. A. Gosey. 1986. Evaluation of beef cattle industry breeding programs: breeders and breed associations. 3rd World Congress on Genetics Applied to Livestock production. July 16-22. pp398-409.

MacNeil, M. D., D. D. Dearborn, L. V. Cundiff, C. A. Dinkel, and K. E. Gregory. 1989. Effects of inbreeding and heterosis in Hereford females on fertility, calf survival and preweaning growth. J. Anim. Sci. 67:895-901.

Pariacote F., L. D. Van Vleck, and R. E. Hunsley. 1998. Genetic and phenotypic parameters for carcass traits of American Shorthorn beef cattle. J. Anim. Sci. 76:2584-2588.

Pariacote, F., L. D. Van Vleck, and M. D. MacNeil. 1998. Effects of inbreeding and heterozygosity on preweaning traits in a closed population of Herefords under selection. J. Anim. Sci. 76:1303-1310.

Snelling, W. M., M. D. MacNeil, D. D. Kress, D. C. Anderson, and M. W. Tess. 1996. Factors influencing genetic evaluations of linebred Hereford cattle in diverse environments. J. Anim. Sci. 74:1499-1510.

Swiger, L. A., K. E. Gregory, R. M. Koch, and V. A. Arthaud. 1961. Effect of inbreeding on performance traits of beef cattle. J. Anim. Sci. 20:626-630.

CHAPTER VI

LITERATURE REVIEW

INCORPORATING GENETIC MARKERS IN GENETIC PREDICTION

6.1. Introduction

Estimated breeding values (EBV) are typically made available to assist producers in the livestock industry in selection of individuals for breeding purposes. Breeding values are predicted using trait phenotypic information, measured on an individual and/or its relatives, most commonly through animal model Best Linear Unbiased Prediction (BLUP) procedures (Henderson, 1984). For traits that are moderate to highly heritable such as growth traits, selection employing BLUP-EBV has been successful. However, genetic progress has been limited for traits that are lowly heritable, sex-limited or expressed late in life primarily due to low accuracy associated with genetic predictions at the time selection decisions are being made.

Advances in molecular genetics have allowed the use of genetic markers that detect polymorphisms at certain locations on the genome. These genetic markers may be responsible for phenotypic variation in traits of economic importance (e.g. single nucleotide polymorphisms) or may by proximity be associated with genes that play a role

in the expression of a trait (e.g. microsatellite genetic markers). Genetic markers have been instrumental in the construction of the bovine linkage map (Bishop et al., 1994; Kappes et al., 1997), which has provided an opportunity to search for quantitative trait loci (QTL) that explain some of the genetic differences among individuals in traits of economic importance in cattle. Several studies have reported evidence of significant associations between markers and QTL in dairy cattle (Georges et al., 1995; Spelman et al., 1996; Ashwell et al., 2004; Van Tassel et al., 2004). This QTL information may be valuable in livestock breeding programs especially for those traits that are lowly heritable, sex-limited, difficult to measure or expressed late in life provided these statistical associations or the QTL themselves are confirmed in outbreeding populations.

Exclusive use of QTL information has been argued against (Lande and Thompson, 1990) even though information on QTL promises to revolutionize selection of breeding individuals in animal breeding. Reasons for this contention are 1) major genes usually have pleiotropic effects, and thus injudicious selection on these genes may be accompanied by undesirable genetic response in correlated traits, 2) most traits of economic importance are influenced by many genes of small effect that may be difficult to detect (Lande and Thompson, 1990). Lande and Thompson (1990) proposed an integration of phenotypic and QTL or genetic marker information in a selection index which appropriately weights the two different sources of information.

Two approaches of incorporating genetic marker or QTL information into the current genetic evaluation (BLUP) procedures have been considered in animal breeding. In the first approach, the numerator relationship matrix is replaced by the relationship matrix computed conditional on pedigree and genotypic information at the QTL locus or

genetic markers linked to the QTL in the genetic evaluation model (Nejati-Javaremi et al., 1997). No QTL effects are explicitly fitted in the model in this approach. In the second approach, marker information is used in the estimation of the marked QTL allelic effects following computation of their covariance matrix under the mixed-inheritance framework (Fernando and Grossman, 1989).

This chapter discusses the two approaches in detail. Special consideration is given to the genetic model under which these approaches are implemented. Attention is paid to the computation of the relationship matrix using pedigree and/or marker information required for each approach. The different approaches to compute genetic relationships are illustrated using a small example pedigree. The methods to build the conditional gametic relationship matrix are discussed with respect to their similarities and differences in the discussion section. Concluding remarks are given at the end of the chapter.

6.2. Best linear unbiased prediction and the mixed model equations

The BLUP procedure allows simultaneous estimation and prediction of fixed and random effects, respectively. A general mixed linear model equation, for a single trait, can be written in matrix notation as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \qquad [6.1]$$

where $\mathbf{y}$ is a vector of observations of order N (number of observations); $\mathbf{b}$ is a vector of unknown fixed effects; $\mathbf{u}$ is a vector of unobservable random effects of order

s (number of animals); $\mathbf{X}$ and $\mathbf{Z}$ are the incidence matrices relating fixed and random effects to observations, respectively; and $\mathbf{e}$ is a vector of random residual effects. It should be pointed out that equation [6.1] can accommodate multiple traits and more than one random effect other than the random residual effect.

The first moments of the random effects in the model are assumed to be:

$$E[\mathbf{u}] = \mathbf{0}, \qquad E[\mathbf{e}] = \mathbf{0}$$

and as a consequence

$$E[\mathbf{y}] = \mathbf{Xb}$$

The second moments are:

$$\mathrm{Var}\begin{bmatrix}\mathbf{u}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix},$$

so that

$$\mathrm{Var}[\mathbf{y}] = \mathbf{ZGZ}' + \mathbf{R}$$

where $\mathbf{G}$ and $\mathbf{R}$ are the variance-covariance matrices of order $s \times s$ and $N \times N$ for the unobservable random and residual effects, respectively.

The mixed model equations (MME) for a nonsingular $\mathbf{G}$ are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \qquad [6.2]$$

where $\mathbf{K}'\hat{\mathbf{b}}$ is a vector of best linear unbiased estimates (BLUE) of $\mathbf{K}'\mathbf{b}$ when $\mathbf{K}'\mathbf{b}$ is estimable and $\hat{\mathbf{u}}$ is a vector of best linear unbiased predictor (BLUP) of the unobservable random genetic effects in $\mathbf{u}$ given the observations in $\mathbf{y}$.

Solving equation [6.2] gives the following results

$$\hat{\mathbf{b}} = \left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right)^{-} \mathbf{X}'\mathbf{R}^{-1}\left(\mathbf{y}-\hat{\mathbf{u}}\right)$$

and

$$\hat{\mathbf{u}} = \left(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\right)^{-1} \mathbf{Z}'\mathbf{R}^{-1}\left(\mathbf{y}-\mathbf{X}\hat{\mathbf{b}}\right)$$

The solutions above indicate that fixed effects solutions depend on random effects and random effects solutions depend on fixed effects.

When the random residual effects are assumed to be identically and independently distributed such as in a single trait setting, $\mathbf{R} = \mathbf{I}\sigma_e^2$, equation [6.2] simplifies to:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\sigma_e^2\mathbf{G}^{-1} \end{bmatrix}\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix}=\begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix} \qquad [6.3]$$

where $\mathbf{I}$ is an identity matrix and σ_e^2 is the variance associated with the random residual effects. The variance-covariance matrix for the unobservable random effects, $\mathbf{G}$, when the relationships between individuals are computed from pedigree information is $\mathbf{G} = \mathbf{A}\sigma_u^2$ where $\mathbf{A}$ is Wright's numerator relationship matrix and σ_u^2 is the variance associated with the unobservable random effects. When $\mathbf{G}^{-1} = \frac{1}{\sigma_u^2}\mathbf{A}^{-1}$, then the mixed model equations can be written:

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z}+\lambda\mathbf{A}^{-1} \end{bmatrix}\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix}=\begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \end{bmatrix} \qquad [6.4]$$

where $\lambda=\frac{\sigma_e^2}{\sigma_u^2}$.

The MME can equivalently be written:

$$\begin{bmatrix} \mathbf{X'R}^{-1}\mathbf{X} & \mathbf{X'R}^{-1}\mathbf{Z} \\ \mathbf{GZ'R}^{-1}\mathbf{X} & \mathbf{GZ'R}^{-1}\mathbf{Z}+\mathbf{I} \end{bmatrix}\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix}=\begin{bmatrix} \mathbf{X'R}^{-1}\mathbf{y} \\ \mathbf{GZ'R}^{-1}\mathbf{y} \end{bmatrix} \qquad [6.5]$$

The equation above yields the same solutions as equation [6.2] but has the added advantage of solving problems when **G** is singular (Henderson, 1984). However, the coefficient matrix of equation [6.5] is nonsymmetric when a singular **G** is used. A modified version of equation [6.5] with a symmetric coefficient matrix is given below (presented in Henderson, 1984).

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{G} \\ \mathbf{G}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{G}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{G}+\mathbf{G} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\boldsymbol{\alpha}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{G}\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \qquad [6.6]$$

and a solution to the random effects can be obtained from $\hat{\mathbf{u}} = \mathbf{G}\hat{\boldsymbol{\alpha}}$.

6.3. Additive genetic variance-covariance matrix (G)

In the preceding section a form of **G** (i.e. $\mathbf{G} = \mathbf{A}\sigma_u^2$) was briefly introduced. The σ_u^2 can be estimated from phenotypic records while **A** can be obtained from pedigree information and its elements are the additive genetic relationships between animals in a population. Individuals must share a common descendent in their ancestry to be related and are termed identical-by-descent. In the absence of inbreeding, relationships between a parent and offspring, full sibs, and half sibs are ½, ½, and ¼, respectively. These relationships are an average or expected relationship given pedigree information and therefore the actual or observed relationships vary around the mean. For example, the actual relationships for a group of half sibs take values from 0 to 50% with an average of

25%. Pedigree-based relationships do not account for this variation. Pedigree relationships are expressed relative to an arbitrary base population. All animals in the base population are assumed to be unrelated and noninbred.

Genetic relationships can also be obtained from genotypic information at the QTL (Nejati-Javaremi, 1995). Relationship based on QTL information refers to identity-by-state at the QTL and is not dependent on pedigree information. The QTL based relationships are the actual or observed relationships at the QTL and therefore account for the variation in relatedness between individuals. When the QTL information is not available, relationship at the QTL can be approximated using two approaches. Firstly, relationship at the QTL can be estimated using pedigree and marker information (Nejati-Javaremi, 1995). Secondly, relationship at the QTL can be inferred from the pedigree, marker genotypes and recombination rate between the marker and QTL (Fernando and Grossman, 1989). Similar to pedigree-based relationship, pedigree- and marker-based relationships are computed based on the assumption that QTL carried by individuals in the base population are unique and relationship represents identity-by-descent at the QTL. Methods to calculate genetic relationships using pedigree, QTL and marker information will be covered in subsequent sections.

6.4. Conventional BLUP with QTL or marker-based relationships

Let the matrix of relationships based on QTL information be denoted by $\boldsymbol{\Theta}$. Then, the additive genetic variance-covariance matrix for the unobservable random effects, $\mathbf{G}$, is given by

$$\mathbf{G} = \boldsymbol{\Theta}\sigma_u^2 \qquad [6.7]$$

The mixed model equations, when assuming $\boldsymbol{\Theta}$ is nonsingular and not explicitly fitting the QTL effects in the model are

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \boldsymbol{\Theta}^{-1}\sigma_u^{-2} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \qquad [6.8]$$

Denoting the matrix of relationships at the QTL estimated using pedigree and marker information by $\tilde{\boldsymbol{\Theta}}$, the additive genetic variance-covariance matrix is

$$\mathbf{G} = \tilde{\boldsymbol{\Theta}}\sigma_u^2 \qquad [6.9]$$

Then, the mixed model equations for a nonsingular $\tilde{\boldsymbol{\Theta}}$ are:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \tilde{\boldsymbol{\Theta}}^{-1}\sigma_u^{-2} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \qquad [6.10]$$

6.5. BLUP under mixed-inheritance model

Fernando and Grossman (1989) showed that the unobservable random additive genetic effect of the i^{th} individual, u_i, can be partitioned into the QTL gametic effects and the residual polygenic effects as follows:

$$u_i = a_i + v_i^p + v_i^m \qquad [6.11]$$

where a_i is the residual polygenic effect assumed to be uncorrelated with the QTL gametic effects; v_i^p and v_i^m are gametic effects of the paternal and maternal marked QTL (MQTL) alleles, respectively.

A modified version of equation [6.1] taking into account the partitioning of the unobservable random effects can be written as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Wv} + \mathbf{e} \qquad [6.12]$$

where $\mathbf{a}$ is a vector of residual polygenic effects of order s; $\mathbf{v}$ is a vector of gametic effects of order $2s$; $\mathbf{W}$ is an incidence matrix relating observations to gametic effects; $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{b}$ and $\mathbf{e}$ are as defined in equation [6.1]. Equation [6.12] can be expanded to include gametic effects of other QTL.

The variance-covariance matrix of the random effects is

$$\text{var}\begin{bmatrix}\mathbf{a}\\ \mathbf{v}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{G} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{P} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{R}\end{bmatrix}=\begin{bmatrix}\mathbf{A}\sigma_a^2 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}\sigma_v^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{I}\sigma_e^2\end{bmatrix} \qquad [6.13]$$

where **A** is the numerator relationship matrix; **Q** is the conditional gametic relationship matrix of order $2s\times 2s$ obtained from pedigree and marker information; **P** is the covariance matrix of the QTL allelic effects of order $2s\times 2s$; **I** is an identity matrix of order N^2; σ_a^2, σ_v^2 and σ_e^2 are the variances associated with the residual polygenic, gametic and random residual effects, respectively.

The mixed model equations, when **G** and **P** are nonsingular, are presented below:

$$\begin{bmatrix}\mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W}\\ \mathbf{Z'X} & \mathbf{Z'Z}+\sigma_e^2\mathbf{G}^{-1} & \mathbf{Z'W}\\ \mathbf{W'X} & \mathbf{W'Z} & \mathbf{W'W}+\sigma_e^2\mathbf{P}^{-1}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{b}}\\ \hat{\mathbf{a}}\\ \hat{\mathbf{v}}\end{bmatrix}=\begin{bmatrix}\mathbf{X'y}\\ \mathbf{Z'y}\\ \mathbf{W'y}\end{bmatrix} \qquad [6.14]$$

6.6. Computing the numerator relationship matrix (NRM) and its inverse

Numerator relationship matrix. The numerator relationship matrix, denoted **A** in the previous section, is obtained using pedigree information. Each element in this matrix represents the average or expected relationship between the same or different individual(s) average over the entire genome and takes values between 0 and 2 inclusive.

The diagonal elements represent the relationship between an individual and itself and is given by $a_{ii} = 1 + f_i$ where a_{ii} is the i^{th} diagonal element of **A** and f_i is Wright's inbreeding coefficient of individual *i*. A non-inbred animal has a diagonal of one and a completely inbred animal has a diagonal of two. The off-diagonal elements are the relationships between any two individuals. For example, the relationship between individuals *i* and *j* is expressed as $a_{ij} = r_{ij}\sqrt{a_{ii}a_{jj}}$ where r_{ij} is Wright's coefficient of relationship between individuals *i* and *j*.

Henderson (1976) presented simple rules to compute the NRM recursively using pedigree information. He presented two approaches that differ in terms of computational efficiency. In the first method, each element of the NRM or **A**, are obtained from the pedigree information. This method requires that individuals are ordered chronologically and indexed 1,2,…,n. The NRM is built one row at a time. For example, elements of the i^{th} row (or individual indexed *i*) are computed as follows:

(i) Both parents (indexed *p* and *q*) known

$$\begin{cases} a_{it} = a_{ti} = 0.5\left(a_{tp} + a_{tq}\right) & \text{for } t = 1,\ldots,i\text{-}1 \\ a_{ii} = 1 + 0.5a_{pq} \end{cases}$$

(ii) One parent is known, denoted p

$$\begin{cases} a_{it} = a_{ti} = 0.5a_{tp} \text{ for } t = 1,...,i\text{-}1 \\ a_{ii} = 1 \end{cases}$$

(iii) Neither parent is known

$$\begin{cases} a_{it} = a_{ti} = 0 \text{ for } t = 1,...,i\text{-}1 \\ a_{ii} = 1 \end{cases}$$

The second method is based on matrix decomposition of **A** such that $\mathbf{A} = \mathbf{LL}'$. The matrix **L** is a lower triangular matrix whose elements are computed from pedigree information as follows. Individuals have to be ordered chronologically and elements of **L** are calculated row at time. Elements of the i^{th} row (index of animal i) of **L** are obtained using the following rules:

(i) Both parents (indexed p and q) known

$$\begin{cases} l_{it} = 0.5\left(l_{pt} + l_{qt}\right) & \text{for } t = 1,...,p \\ \quad = 0.5l_{qt} & \text{for } t = p+1,...,q \\ \quad = 0 & \text{for } t = q+1,...,i-1 \\ \\ l_{ii} = \sqrt{1 + 0.5\sum_{t=1}^{p} l_{pt}l_{qt} - \sum_{t=1}^{q} l_{it}^2} = \sqrt{0.5 - 0.25\left(f_p + f_q\right)} \end{cases}$$

where f_p and f_q are the inbreeding coefficients of individual i's parents, p and q, respectively.

An equivalent equation to obtain l_{ii} is (Quaas, 1976):

$$l_{ii} = \sqrt{1 - 0.25\sum_{t=1}^{p} l_{pt}^2 - 0.25\sum_{t=1}^{q} l_{qt}^2}$$

(ii) One parent is known, denoted p

$$\begin{cases} l_{it} = 0.5 l_{pt} & \text{for } t = 1,\ldots, p \\ \quad = 0 & \text{for } t = p+1,\ldots, i\text{-}1 \\ \\ l_{ii} = \sqrt{1 - \sum_{t=1}^{p} l_{it}^2} = \sqrt{0.75 - 0.25 f_p} \end{cases}$$

Equivalently (Quaas,1976)

$$l_{ii} = \sqrt{1 - 0.25\sum_{t=1}^{p} l_{pt}^2}$$

(iii) Neither parent is known

$$\begin{cases} l_{it} = 0 & \text{for } t = 1,...,i\text{-}1 \\ l_{ii} = 1 \end{cases}$$

For illustrative purposes, a simple pedigree example of six animals shown in Table 6.1 is used to compute the NRM. In this example, two individuals of unknown ancestry were mated to produce a son and a daughter who in turn were inter se mated to produce two full-sib offspring.

Table 6.1. Pedigree information.

Animal	Sire	Dam
1	-	-
2	-	-
3	1	2
4	1	2
5	3	4
6	3	4

A dash refers to unknown parent.

The NRM constructed using the first method is given in Table 6.2. For example, the relationship between individuals 5 and 6 was computed as follows:

$$a_{5,6} = 0.5(a_{6,3} + a_{6,4}) = 0.5(0.75 + 0.75) = 0.75$$

The relationship between individual 6 and itself is

$$a_{6,6} = 1 + 0.5a_{3,4} = 1 + (0.5 \times 0.5) = 1.25$$

indicating that individual 6 is 25% inbred.

Table 6.2. Wright's numerator relationship matrix (**A**).

Animal	1	2	3	4	5	6
1	1.00	0.00	0.50	0.50	0.50	0.50
2	0.00	1.00	0.50	0.50	0.50	0.50
3	0.50	0.50	1.00	0.50	0.75	0.75
4	0.50	0.50	0.50	1.00	0.75	0.75
5	0.50	0.50	0.75	0.75	1.25	0.75
6	0.50	0.50	0.75	0.75	0.75	1.25

The lower triangular matrix, **L**, of the NRM was constructed using the rules given in the second method and is shown in Table 6.3. The product of **L** and its transpose gave the NRM presented in Table 6.2 confirming the equivalence between the two methods. Computation of the third diagonal element of **L** is given below:

$$l_{3,3} = \sqrt{1 + 0.5(l_{1,1} \times l_{2,1}) - (l_{3,1}^2 + l_{3,2}^2)}$$

$$= \sqrt{1 + 0.5(1 \times 0) - (0.5^2 + 0.5^2)}$$

$$= 0.71$$

Table 6.3. The lower triangular matrix (**L**).

Animal	1	2	3	4	5	6
1	1.00	0.00	0.00	0.00	0.00	0.00
2	0.00	1.00	0.00	0.00	0.00	0.00
3	0.50	0.50	0.71	0.00	0.00	0.00
4	0.50	0.50	0.00	0.71	0.00	0.00
5	0.50	0.50	0.35	0.35	0.71	0.00
6	0.50	0.50	0.35	0.35	0.00	0.71

Elements of the NRM in Table 6.2 are interpreted as follows. Animals 1 and 2 had unknown parents and were assumed unrelated and noninbred. The full sib grand progeny are inbred due to the fact that their parents were full sibs. The full sib grand progeny were more related to their grandparents than the expected ¼ in the absence of inbreeding. Most of the elements in the numerator relationship matrix were nonzero since all the individuals in the pedigree were related except for the two base animals. This indicates that performance information will be "shared" among these individuals in the prediction of their breeding values. That is, records on any one animal contribute information to the evaluation of the other animals. On the other hand, if there was no knowledge of the parents on the six individuals the numerator relationship matrix would be an identity matrix and phenotypic information on any one animal would not contribute to prediction of merit of other animals.

Inverse of the NRM. The inverse of the NRM can be derived using direct inversion methods available in most computer packages. However, this method may be

prohibitive for large datasets common in animal breeding. Henderson (1976) discovered simple rules to obtain the inverse of the NRM directly from pedigree information without creating the NRM itself.

Henderson (1976) showed that the lower triangular matrix of the NRM, **L**, can be expressed as $\mathbf{L} = \mathbf{TD}$ where $\mathbf{D}$ is a diagonal matrix whose diagonal elements are identical to those of **L** and **T** is computed using the same method as for **L** except that its diagonal elements are equal to 1. Therefore, the inverse of the NRM can be written as follows:

$$\mathbf{A}^{-1} = (\mathbf{LL}')^{-1} = (\mathbf{TDDT}')^{-1} = (\mathbf{T}^{-1})'(\mathbf{D}^{-1})^2\mathbf{T}^{-1} \qquad [6.15]$$

The matrix $\mathbf{T}^{-1}$, has a simple structure with diagonal elements equal to 1 and all other elements in the i^{th} row equal to zero except for -0.5 for the ij^{th} element where j is the parent of individual i. The $\mathbf{T}^{-1}$ can be computed recursively, row at a time, as follows:

$$\mathbf{T}_i^{-1} = \begin{bmatrix} \mathbf{T}_{i-1}^{-1} & \mathbf{0} \\ -\mathbf{a}' & 1 \end{bmatrix} \qquad [6.16]$$

where $-\mathbf{a}'$ is an $i-1$ vector whose elements are zero except for -0.5 corresponding to the known parents of individual i. The $\mathbf{T}^{-1}$ was computed for the pedigree example in Table 6.1 and results are shown in Table 6.4.

Table 6.4. The inverse of **T**.

Animal	1	2	3	4	5	6
1	1.0	0.0	0.0	0.0	0.0	0.0
2	0.0	1.0	0.0	0.0	0.0	0.0
3	-0.5	-0.5	1.0	0.0	0.0	0.0
4	-0.5	-0.5	0.0	1.0	0.0	0.0
5	0.0	0.0	-0.5	-0.5	1.0	0.0
6	0.0	0.0	-0.5	-0.5	0.0	1.0

It was obvious to Henderson (1976), based on the structure of $\mathbf{T}^{-1}$ and equation [6.15], that computation of the inverse of the NRM required only the list of parents and the diagonal elements of the **L** matrix. An algorithm to compute the inverse of the NRM directly from pedigree information is given below (Henderson, 1976):

(i) Define $\left(\mathbf{D}^{-1}\right)^2$ as $\boldsymbol{\Gamma}$

(ii) Initialize $\mathbf{A}^{-1}$ to $\boldsymbol{\Gamma}$

a. If both parents (p and q) of individual i are known, add the following elements to $\mathbf{A}^{-1}$

$-0.5\boldsymbol{\Gamma}_{ii}$ to (p,i), (i,p), (q,i), and (i,q) elements;

$0.25\boldsymbol{\Gamma}_{ii}$ to (p,p), (p,q), (q,p), and (q,q) elements.

b. If one parent (p) is known, add

$-0.5\boldsymbol{\Gamma}_{ii}$ to (p,i) and (i,p) elements;

$0.25\boldsymbol{\Gamma}_{ii}$ to (p,p) element.

The rules given above were applied in obtaining the inverse of **A** from the knowledge of **L** and the pedigree information given in Table 6.1. The initial diagonal elements of $\mathbf{A}^{-1}$ are $diag(\Gamma) = [1,1,2,2,2,2]$ and contributions of each individual to $\mathbf{A}^{-1}$ are presented in Table 6.5. The sums of the contributions within each cell are the elements of $\mathbf{A}^{-1}$ as shown in Table 6.6. This inverse was equal to the inverse computed using equation [6.15] and direct inversion routine using a computer package.

Table 6.5. Individual contributions to $\mathbf{A}^{-1}$.

Animal	1	2	3	4	5	6
1	**1.00** 0.50 (3) 0.50 (4)	**0.00** 0.50 (3) 0.50 (4)	**0.00** -1.00 (3)	**0.00** -1.00 (4)	**0.00**	**0.00**
2	**0.00** 0.50 (3) 0.50 (4)	**1.00** 0.50 (3) 0.50 (4)	**0.00** -1.00 (3)	**0.00** -1.00 (4)	**0.00**	**0.00**
3	**0.00** -1.00 (3)	**0.00** -1.00 (3)	**2.00** 0.50 (5) 0.50 (6)	**0.00** 0.50 (5) 0.50 (6)	**0.00** -1.00 (5)	**0.00** -1.00 (6)
4	**0.00** -1.00 (4)	**0.00** -1.00 (4)	**0.00** 0.50 (5) 0.50 (6)	**2.00** 0.50 (5) 0.50 (6)	**0.00** -1.00 (5)	**0.00** -1.00 (6)
5	**0.00**	**0.00**	**0.00** -1.00 (5)	**0.00** -1.00 (5)	**2.00**	**0.00**
6	**0.00**	**0.00**	**0.00** -1.00 (6)	**0.00** -1.00 (6)	**0.00**	**2.00**

Values in bold are the elements of the Γ matrix or initial values of A^{-1}. Numbers in parenthesis are the identification numbers of the individual who made the contribution.

Table 6.6. The $\mathbf{A}^{-1}$ matrix.

Animal	1	2	3	4	5	6
1	2	1	-1	-1	0	0
2	1	2	-1	-1	0	0
3	-1	-1	3	1	-1	-1
4	-1	-1	1	3	-1	-1
5	0	0	-1	-1	2	0
6	0	0	-1	-1	0	2

It is apparent from Henderson (1976)'s algorithm that the only major task in the computation of $\mathbf{A}^{-1}$ is obtaining the diagonal elements of the $\mathbf{L}$ matrix in order to form $\mathbf{D}^{-1}$ and Γ. Quaas (1976) presented an algorithm to compute the diagonal elements and the inverse of the numerator relationship matrix. This algorithm is efficient because it does not create the whole $\mathbf{L}$ matrix.

6.7. Computing the allelic relationship matrix

Procedures to construct the relationship matrix using genotypic information at either QTL or genetic marker loci were proposed by Nejati-Javaremi (1995). The relationship matrix derived from the QTL genotypes refers to identity-by-state (IBS) at a given QTL locus between two individuals and is independent of the pedigree information. It refers to the observed genetic relationships between individuals averaged over loci under consideration. A term "total allelic (TA) relationship matrix" was coined for this kind of relationship matrix (Nejati-Javaremi et al., 1997). The IBS relationships

can deviate considerably from the expected relationships when few loci are considered. However, IBS relationships based on a large or infinite number of loci are expected to be equivalent to the numerator relationships. Alternatively, relationships can be computed based on genotypes at more than one genetic marker loci in the absence of QTL genotypic information. The marker-based relationships refer to identity-by-descent (IBD) of the chromosomal segments flanked by markers. The two kinds of relationships are described in detail separately in the next sections.

Identity-by-state (IBS) relationships. The IBS relationships matrix was denoted by $\boldsymbol{\Theta}$ earlier in equation [6.7]. The IBS relationship between two individuals (*a* and *b*) at a given locus, say *l*, is computed as follows:

$$\Theta_{a,b}(l) = 2 \times \frac{\sum_{i=1}^{2}\sum_{j=1}^{2} \Pr(Q_i = Q_j)}{4} = \frac{\sum_{i=1}^{2}\sum_{j=1}^{2} \Pr(Q_i = Q_j)}{2} \qquad [6.17]$$

where $\Pr(Q_i = Q_j)$ is the probability that i^{th} QTL allele of the first individual, *a*, is identical-by-state with the j^{th} QTL allele of the second individual, *b*. The $\Pr(Q_i = Q_j)$ takes on two values; a one if the two alleles are identical-by-state otherwise a zero. The coefficient of 2 indicates that IBS relationship is equivalent to the numerator relationship of Wright (Wright, 1922). The IBS relationship for individuals *a* and *b* averaged over *k* QTL is:

$$\Theta_{a,b} = \frac{\sum_{l=1}^{k} \Theta_{a,b}(l)}{k} = \frac{\sum_{l=1}^{k} \left(\frac{\sum_{i=1}^{2} \sum_{j=1}^{2} \Pr(Q_i = Q_j)_{(l)}}{2} \right)}{k} \qquad [6.18]$$

There is an implicit assumption in equation [6.18] that each QTL contributes equally to the additive genetic variance.

Table 6.7. Pedigree and QTL genotypes.

			QTL Genotypes	
Animal	**Sire**	**Dam**	**QTL 1**	**QTL 2**
1	-	-	13	13
2	-	-	23	12
3	1	2	13	12
4	1	2	12	13
5	3	4	12	23
6	3	4	13	12

Assume that genotypic information on two QTL loci was available on six individuals in the pedigree example shown in Table 6.7. Equation [6.18] was used to compute the IBS relationships using the QTL genotypes and results are presented in Table 6.8. Let us consider the IBS relationship between animals 2 and 3 for illustrative purposes. Based on the first QTL locus, individual 2 had alleles 2 and 3 while individual 3 had 1 and 3. The identity-by-state, $\Pr(Q_i = Q_j)$, for the four alleles carried by

individuals 2 and 3 at the first QTL locus are presented below; columns refer to alleles carried by individual 3 and the rows are the alleles from individual 2:

$$\begin{array}{cc} \text{Allele} & \begin{array}{cc} 1 & 3 \end{array} \\ \begin{array}{c} 2 \\ 3 \end{array} & \begin{bmatrix} \Pr(Q_1 = Q_1) & \Pr(Q_1 = Q_2) \\ \Pr(Q_2 = Q_1) & \Pr(Q_2 = Q_2) \end{bmatrix} \end{array}$$

$$\begin{array}{cc} \text{Allele} & \begin{array}{cc} 1 & 3 \end{array} \\ \begin{array}{c} 2 \\ 3 \end{array} & \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \end{array}$$

Then, the IBS relationship between individuals 2 and 3 at the first QTL is

$$\Theta_{2,3}(1) = \frac{\Pr(Q_1 = Q_1) + \Pr(Q_1 = Q_2) + \Pr(Q_2 = Q_1) + \Pr(Q_2 = Q_2)}{2}$$

$$= \frac{0+0+0+1}{2} = 0.5$$

The IBS relationship for individuals 2 and 3 for the two QTL is an average of the IBS relationship at each QTL locus as shown below:

$$\Theta_{2,3} = \frac{\sum_{l=1}^{2} \Theta_{2,3}(l)}{2} = \frac{0.5+1.0}{2} = 0.75$$

Comparing the IBS relationship matrix (Θ) with the numerator relationship matrix (**A**) presented in Table 6.2 shows that pedigree-based relationships represent an underestimate of the genetic relationship between individuals. For example, individuals 1 and 2 were unrelated when pedigree information was used. However, the same individuals were 0.5 related when the QTL genotypes were used. In general, IBS relationships are greater than the numerator relationships. This is because the IBS relationships are based on the identity-by-state of the QTL alleles rather than only considering those assumed identity-by-descent. The fundamental difference in the two approaches is in terms of the reference population relative to which the relationships are expressed. For instance, IBS relationship is relative to the beginning of the breed or when the mutation of the QTL occurred while the pedigree-derived relationship is implicitly relative to the point where no information on the individual's pedigree is available or some generation in the individual's pedigree.

Table 6.8. The IBS relationship matrix (Θ).

Animal	**1**	**2**	**3**	**4**	**5**	**6**
1	1.00	0.50	0.75	0.75	0.50	0.75
2	0.50	1.00	0.75	0.50	0.50	0.75
3	0.75	0.75	1.00	0.50	0.50	1.00
4	0.75	0.50	0.50	1.00	0.75	0.50
5	0.50	0.50	0.50	0.75	1.00	0.50
6	0.75	0.75	1.00	0.50	0.50	1.00

Identity-by-descent (IBD) relationships. The IBD relationships are computed for the segments within flanking markers. A marker segment is defined as an interval between two flanking markers thought to be containing the QTL as shown in Figure 6.1. When the two flanking markers are closely linked, the inheritance of the segment within them will follow Mendelian laws of segregation i.e. each segment will be inherited as if it were an allele at a single locus.

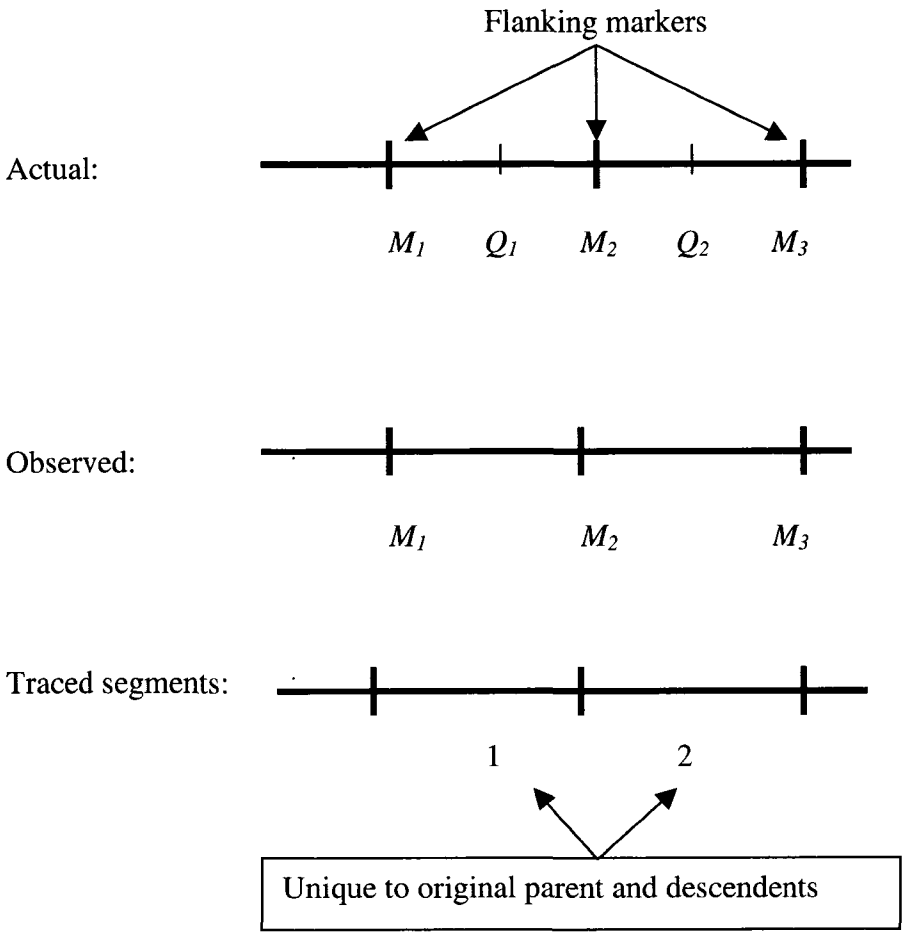


Figure 6.1. An illustration of the marker segments upon which IBD relationships are based (from Nejati-Javaremi, 1995).

The concept of IBD relationship was developed by Nejati-Javaremi (1995). Each segment is assumed to be unique in the base population because information about the QTL flanked by the markers is not available implying that base individuals are unrelated and non-inbred with respect to these segments. This is a similar assumption to that made in forming pedigree-based relationships. The total number of unique segments within each pair of flanking markers is therefore twice the number of individuals in the base population. Parental origin of each segment is assigned based on the marker genotypes of parents and their descendents. This in turn allows tracing of each segment to its origin. Segments that did not undergo recombination can be traced to their origin in the base population. On the other hand, the origin of segments that underwent recombination is the generation at which recombination occurred. Using the knowledge of the origin of each segment, the IBD relationship between two individuals can be expressed as the proportion of the segment in common or identical-by-descent. Given that each individual has two segments, there are four possible ways that two individuals could be IBD. Therefore, the IBD relationship between two individuals, *a* and *b*, can be obtained from:

$$\tilde{\Theta}_{a,b} = 2 \times \frac{\sum_{i=1}^{2} \sum_{j=1}^{2} \mathrm{LS}_{ij}}{4} \qquad [6.19]$$

where LS_{ij} is the proportion of the i^{th} linkage segment from the first individual, *a*, in common with the j^{th} linkage segment from individual, *b*. In this case, LS_{ij} is an estimate of the probability of identity-by-state, $\Pr(Q_i = Q_j)$, given earlier in the section

on IBS relationships and thus LS_{ij} range from 0 to 1. Therefore, the IBD relationships range from 0 to 2 similar to IBS and numerator relationships.

Computations of the IBD relationships are illustrated using an example pedigree in Table 6.9. This is the same pedigree given in Table 6.7 except that here we assume that only genotypic information at the marker loci flanking the two QTL is available. There were a total of eight unique segments identified in the base generation. Twelve more recombinations occurred after the base generation resulting in a total of 20 segments. The high recombination rate depicted here represents markers that are far apart from each other. The parental origin of the haplotypes for animals in the base population was assigned arbitrarily because their parents were unknown. Each segment in the descendents (animals 4 to 6) was traced to its origin by observing the crossover events that occurred between consecutive generations as suggested by markers. For example, a crossover occurred between the first and second marker of individual 1 and thus, individual 3 inherited a recombinant segment from individual 1 (the sire). On average, a crossover event occurs at the center of the segment and thus the recombinant segment is expected to consist of half of segment 1 and half of segment 3. This recombinant segment is denoted $1_{m1}\_1_{p2}$ in Table 6.9. Other segments were identified using the same procedure. If no recombination occurred only eight segments from the base population would have been passed on to the descendents.

Table 6.9. Pedigree, marker genotypes and segments within flanking markers.[a]

			Parental	Marker / Segments				
Animal	**Sire**	**Dam**	**Origin**	**M1**	**Seg1**	**M2**	**Seg2**	**M3**
1	-	-	p	1	$\mathbf{1}_{p1}\_\mathbf{1}_{p2}$	2	$\mathbf{1}_{p2}\_\mathbf{1}_{p3}$	1
			m	2	$\mathbf{1}_{m1}\_\mathbf{1}_{m2}$	3	$\mathbf{1}_{m2}\_\mathbf{1}_{m3}$	2
2	-	-	p	1	$\mathbf{2}_{p1}\_\mathbf{2}_{p2}$	1	$\mathbf{2}_{p2}\_\mathbf{2}_{p3}$	1
			m	3	$\mathbf{2}_{m1}\_\mathbf{2}_{m2}$	2	$\mathbf{2}_{m2}\_\mathbf{2}_{m3}$	3
3	1	2	p	2	$\mathbf{1}_{m1}\_\mathbf{1}_{p2}$	2	$\mathbf{1}_{p2}\_\mathbf{1}_{m3}$	2
			m	3	$\mathbf{2}_{m1}\_\mathbf{2}_{p2}$	1	$\mathbf{2}_{p2}\_\mathbf{2}_{m3}$	3
4	1	2	p	1	$\mathbf{1}_{p1}\_\mathbf{1}_{m2}$	3	$\mathbf{1}_{m2}\_\mathbf{1}_{p3}$	1
			m	3	$\mathbf{2}_{m1}\_\mathbf{2}_{p2}$	1	$\mathbf{2}_{p2}\_\mathbf{2}_{m3}$	3
5	3	4	p	3	$\mathbf{2}_{m1}\_\mathbf{1}_{p2}$	2	$\mathbf{1}_{p2}\_\mathbf{1}_{m3}$	2
			m	3	$\mathbf{2}_{m1}\_\mathbf{1}_{m2}$	3	$\mathbf{1}_{m2}\_\mathbf{2}_{m3}$	3
6	3	4	p	3	$\mathbf{2}_{m1}\_\mathbf{2}_{p2}$	1	$\mathbf{2}_{p2}\_\mathbf{1}_{m3}$	2
			m	1	$\mathbf{1}_{p1}\_\mathbf{1}_{m2}$	3	$\mathbf{1}_{m2}\_\mathbf{1}_{p3}$	1

[a] p – paternal, m – maternal, $1_{p1}\_1_{p2}$ denotes the first segment of individual 1 flanked by the paternal marker alleles from the first marker on the left and second marker on the right. The underscore is the point where crossover occurs; the part to the left of the underscore represents the first half of the segment and that on the right is the second half of the segment.

The IBD relationship of animals 3 and 5 at the paternal segments $1_{m1}\_1_{p2}$ and $2_{m1}\_1_{p2}$, respectively is equal to the proportion in common at the first segment which is 0.5 divided by 2 giving 0.25. The IBD relationship of animals 3 and 5 averaged across linkage segments is

$$\tilde{\Theta}_{3,5} = \frac{\sum_{i=1}^{2}\sum_{j=1}^{2} LS_{ij}}{2} = \frac{(0.75)+(0.00)+(0.25)+(0.50)}{2} = 0.75$$

Table 6.10. The IBD relationship matrix ($\tilde{\Theta}$).

Animal	1	2	3	4	5	6
1	1.00	0.00	0.50	0.50	0.63	0.63
2	**0.00**	1.00	0.50	0.50	0.38	0.38
3	**0.50**	**0.50**	1.00	0.50	0.75	0.50
4	**0.50**	0.50	0.50	1.00	0.63	0.88
5	**0.63**	**0.38**	**0.75**	**0.63**	**1.25**	0.63
6	**0.63**	**0.38**	**0.50**	**0.88**	**0.63**	1.00

The IBD relationship matrix is given in Table 6.10. Elements of the IBD matrix that are different from those obtained using the QTL genotypes (IBS relationships) are shown in bold; only elements on the lower triangle were bolded since the IBD matrix is symmetric. It is apparent that in majority of the cases, IBD relationships under- or over-predicted the IBS genetic relationships. In comparison, with the pedigree-based relationships, the IBD relationship performed worst. This result might have been due to large distance between the flanking markers such that inheritance at the QTL was not accurately inferred using pedigree and marker information. Therefore, genetic predictions based on relationships derived from markers that are located far apart along the genome might be worse than those obtained using relationships calculated from pedigree information alone.

6.8. Computing the conditional gametic relationship matrix (GRM)

The gametic relationship represents the relationships between two individuals at a gametic rather than a genomic level. The matrix of these relationships is computed using pedigree information. Here we focus on conditional gametic relationships computed based on pedigree, genetic marker information and the knowledge of the recombination rate between the marker and a QTL. The usage of the term gametic relationship hereafter refers to conditional gametic relationship. The gametic relationship between gametes of two individuals at a QTL is given below; the QTL alleles are numbered 1 to 4 with the first two being the alleles for individual 1 and the last two for individual 2.

QTL alleles	**1**	**2**	**3**	**4**
1	$\Pr(Q_1 \equiv Q_1)$	$\Pr(Q_1 \equiv Q_2)$	$\Pr(Q_1 \equiv Q_3)$	$\Pr(Q_1 \equiv Q_4)$
2	$\Pr(Q_2 \equiv Q_1)$	$\Pr(Q_2 \equiv Q_2)$	$\Pr(Q_2 \equiv Q_3)$	$\Pr(Q_2 \equiv Q_4)$
3	$\Pr(Q_3 \equiv Q_1)$	$\Pr(Q_3 \equiv Q_2)$	$\Pr(Q_3 \equiv Q_3)$	$\Pr(Q_3 \equiv Q_4)$
4	$\Pr(Q_4 \equiv Q_1)$	$\Pr(Q_4 \equiv Q_2)$	$\Pr(Q_4 \equiv Q_3)$	$\Pr(Q_4 \equiv Q_4)$

The block diagonal elements (highlighted boxes) represent gametic relationships between the two gametes of the same individual and the block offdiagonal elements are the gametic relationships between gametes of two different individuals. The $\Pr(Q_i \equiv Q_j)$

refers to the probability that the QTL allele carried by gamete i is identical-by-descent to the QTL allele from gamete j conditional on marker genotypes and recombination rate between the marker and QTL. Note that the subscripts i and j now refers to the i^{th} and j^{th} gametes whereby gametes are numbered from 1 to $2n$; where n is the number of the animals. The probability, $\Pr(Q_i \equiv Q_j)$, typically has values between 0 and 1 (rather than 0 or 1) because information on QTL genotypes is unavailable. That is, only probability statements can be made about the gametic identity at the QTL. An illustration of the marker and QTL location on the chromosome is shown in Figure 6.2.

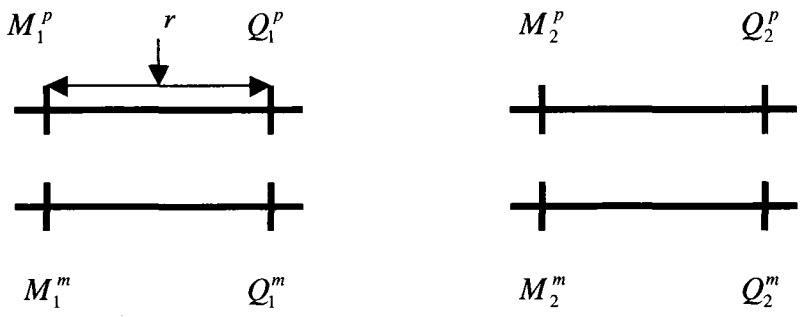


Figure 6.2. Two individuals and their haplotypes, M and Q are the marker locus and QTL, respectively and r is the recombination rate between the marker locus and QTL.

Several algorithms have been developed to obtain $\Pr(Q_i \equiv Q_j)$ and construct the gametic relationship matrix when genotypes on genetic marker loci are available and the recombination rate between genetic marker locus and QTL is known or its estimate is available (Fernando and Grossman, 1989; Goddard 1992; Van Arendonk et al., 1994;

Wang et. al., 1995; Abdel-Azim and Freeman, 2001; Liu et al. al., 2002). These algorithms have distinct features, as a result of simplifying assumptions invoked, which restrict their application to certain scenarios. Computation of the gametic relationship matrix and its inverse will be shown separately for each algorithm in the next sections.

6.8.1. Fernando and Grossman (1989) approach

Fernando and Grossman (1989) showed that $\Pr(Q_i \equiv Q_j)$ can be computed recursively using genetic markers and knowledge of the recombination rate between the marker and QTL. Their derivation of $\Pr(Q_i \equiv Q_j)$ will be shown using animals 1 and 2 with a common father as shown in Figure 6.3. Assume that only marker genotypes are observed. Let the index for the QTL alleles be of the form:

$$\begin{bmatrix} Q_s^p \\ Q_s^m \\ Q_1^p \\ Q_1^m \\ Q_2^p \\ Q_2^m \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}$$

The paternal alleles of individuals 1 and 2 are used to illustrate the recursive formula to compute $\Pr(Q_i \equiv Q_j)$. Given that animal 1 and 2 are not direct descendants of each other as shown in Figure 6.3 and that the parental origin of the marker allele is

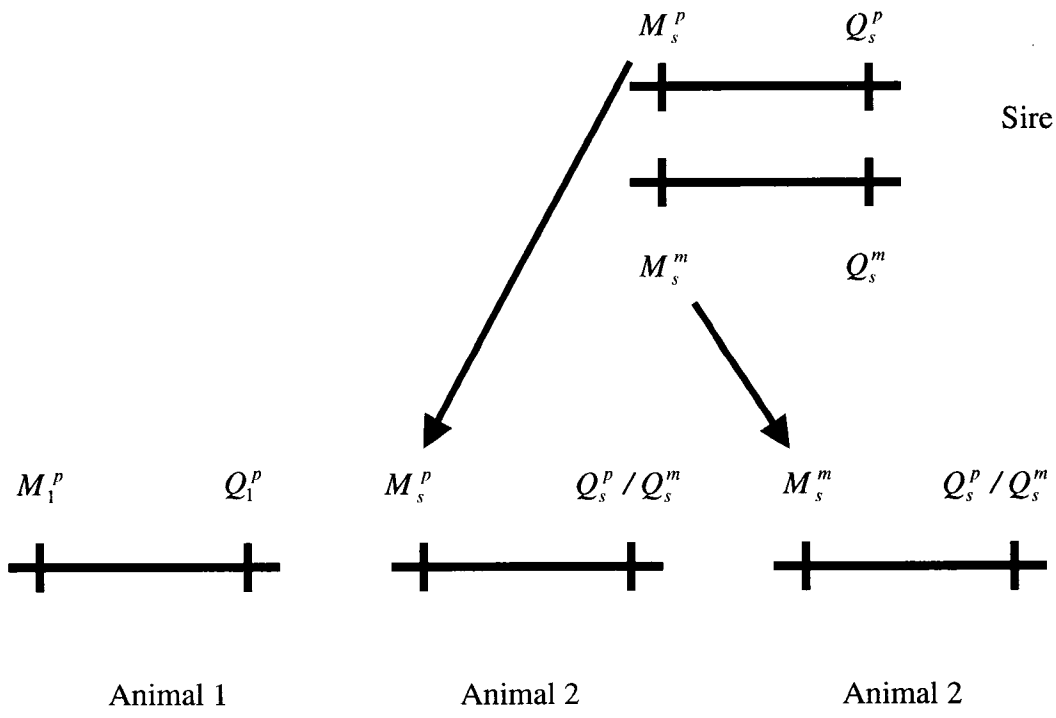


Figure 6.3. Diagram of the animal haplotypes of known parental origin.

known, there are two possibilities that the paternal QTL alleles (Q_1^p and Q_2^p) of animals 1 and 2 are identical-by-descent as shown below:

	$Q_1^p \equiv Q_s^p$	$Q_1^p \equiv Q_s^m$
$Q_2^p \Leftarrow Q_s^p$	$Q_1^p \equiv Q_2^p$ IBD	not IBD
$Q_2^p \Leftarrow Q_s^m$	not IBD	$Q_1^p \equiv Q_2^p$ IBD

The notation $Q_2^p \Leftarrow Q_s^p$ refers to the event that individual 2 inherited its sire's paternal QTL allele. Using the probabilities above, the general form for calculating the probability of identity-by-descent between the paternal QTL alleles of two individuals, *i* and *j*, is:

$$\begin{aligned}\Pr\left(Q_i^p \equiv Q_j^p\right) = \Pr\left(Q_i^p \equiv Q_s^p\right).\Pr\left(Q_j^p \Leftarrow Q_s^p\right) \\ + \Pr\left(Q_i^p \equiv Q_s^m\right).\Pr\left(Q_j^p \Leftarrow Q_s^m\right)\end{aligned} \quad [6.20]$$

where *s* is the sire of individual *j* and not necessarily *i*. The probability, $\Pr\left(Q_i^p \equiv Q_j^m\right)$, can be derived in the similar manner.

In the absence of the QTL information (as assumed in this example) but presence of marker information, inheritance at the QTL locus can be inferred from inheritance at the marker locus and the recombination rate (*r*) between the marker and QTL. Consider animal 1 and 2 in our example. The conditional probability that animal 2 inherits Q_s^p given that it inherited M_s^p is (1-*r*) i.e. $\Pr\left(Q_2^p \Leftarrow Q_s^p \mid M_2^p \Leftarrow M_s^p\right) = (1-r)$. Therefore, the probability of identity-by-descent of Q_1^p (indexed 3) and Q_2^p (indexed 5) given that animal 2 inherits M_s^p can be obtained as follows

$$\begin{aligned}\Pr\left(Q_1^p \equiv Q_2^p\right) = \Pr\left(Q_1^p \equiv Q_s^p\right).(1-r) \\ + \Pr\left(Q_1^p \equiv Q_s^m\right).r\end{aligned} \quad [6.21]$$

Alternatively, if animal 2 inherited the maternal marker allele from its sire (M_s^m), the $\Pr\left(Q_1^p \equiv Q_2^p\right)$ is

$$\Pr\left(Q_1^p \equiv Q_2^p\right) = \Pr\left(Q_1^p \equiv Q_s^p\right).r + \Pr\left(Q_1^p \equiv Q_s^m\right).(1-r) \quad [6.22]$$

These probabilities of identity-by-descent are the elements of the GRM i.e. $\Pr\left(Q_1^p \equiv Q_2^p\right)$ is the $\mathbf{Q}_{1^p,2^p}$ element of the GRM. When marker information is not available or the marker and QTL are far apart, there is an equal probability of either of the sire's QTL alleles (Q_s^p and Q_s^m) to have been transmitted to animal 2 (i.e. equations [6.21] and [6.22] can be used as if $r = 0.5$).

Equations [6.21] and [6.22] lead to a tabular method (similar to the one used to construct the numerator relationship matrix by Henderson (1976)) to construct the gametic relationship matrix (denoted **Q** in equation [6.13]). The gametic relationship matrix is twice the size of the numerator relationship matrix. The GRM is computed as follows: The gametes are ordered such that parental gametes precede the progeny gametes and indexed from 1 to $2n$; where n is the number of animals. Let QTL gametes for animal t (Q_t^p and Q_t^m) be indexed i_t^p and i_t^m, its sire (Q_s^p and Q_s^m) be i_s^p and i_s^m and its dam (Q_d^p and Q_d^m) be i_d^p and i_d^m. The offdiagonal elements of the gametic relationship matrix, **Q**, can be obtained as follows. The first $2(t-1)$ elements of row i_t^p up to the diagonal can be obtained by

$$\mathbf{Q}_{i_t^p,j} = \left(1-\rho_t^p\right).\mathbf{Q}_{i_s^p,j} + \rho_t^p.\mathbf{Q}_{i_s^m,j} \qquad [6.23]$$

for $j = 1...i_t^p - 1$, where $\rho_t^p = r$ if individual t inherits M_s^p or $\rho_t^p = (1-r)$ if individual t inherits M_s^m. Due to the fact that $\mathbf{Q}$ is symmetric, elements of column i_t^p above the diagonal are the corresponding elements of row i_t^p.

Similarly, the first $2(t-1)$ elements of row i_t^m up to the diagonal can be obtained using

$$\mathbf{Q}_{i_t^m,j} = \left(1-\rho_t^m\right).\mathbf{Q}_{i_d^p,j} + \rho_t^m\mathbf{Q}_{i_d^m,j} \qquad [6.24]$$

for $j = 1...i_t^m - 1$, where $\rho_t^m = r$ if individual t inherits M_d^p or $\rho_t^m = (1-r)$ if individual t inherits M_d^m. The corresponding elements above the diagonal are obtained by symmetry. Diagonal elements of the GRM are equal to 1 indicating that each gamete is identical-by-descent to itself at the QTL.

Table 6.11. Pedigree and marker genotypes.

Animal	Sire	Dam	Genotype	Marker parental origin	
				Paternal	Maternal
1	-	-	1,1[a]	-	-
2	-	-	2,2	-	-
3	-	-	1,2	-	-
4	1	2	1,2	1	2
5	3	4	1,1	1	1
6	1	4	1,2	1	2
7	5	6	1,2	1	2

[a] The first and second markers allele of an individual (pedigree example from Abdel-Azim and Freeman, 2001).

The gametic relationship matrix was computed using an example pedigree in Table 6.11 assuming a recombination rate of 0.1. The QTL alleles were numbered from 1 to 14, i.e. $Q_1^1, Q_1^2, Q_2^1, ..., Q_7^2$ were indexed $1, 2, 3, ..., 14$. Gametic relationships among the base individuals could not be determined given available pedigree and marker information and were thus assumed to be zero. The gametic relationships of the three base animals are in the top left 6×6 submatrix of the GRM in Table 6.12. That is, each QTL allele of each gamete was considered to be unique and therefore not identical-by-descent. The computation of the elements of the GRM is shown using the gametic relationship between the paternal QTL alleles of animal 1 (Q_1^1) and 4 (Q_4^1) i.e. QTL alleles indexed 1 and 7. It could not be determined if the paternal marker allele of animal 4 was inherited from its sire's (animal 1) paternal or maternal allele given the marker

information. Therefore, either of the sire's QTL alleles (Q_1^1 or Q_1^2) have an equal chance of being transmitted to animal 4 (i.e. $r = 0.5$). The gametic relationship between the paternal QTL alleles of animals 1 (Q_1^1) and 4 (Q_4^1) is computed as follows:

$$\begin{aligned}\mathbf{Q}_{7,1} &= \mathbf{Q}_{1,1}(1-r) + r.\mathbf{Q}_{2,1} \\ &= (1)(1-0.5) + (0.5)(0) \\ &= 0.5\end{aligned}$$

where $\mathbf{Q}_{1,1}$ is the IBD of the paternal allele of animal 1 (Q_1^1) and itself which is equal to 1, and $\mathbf{Q}_{2,1}$ is the IBD of the maternal (Q_1^2) and paternal (Q_1^1)alleles of animal 1, which by assumption was zero. The gametic relationship of 0.5 is the same as the expected gametic relationship conditional on pedigree information. This is not surprising because it was not clear whether the paternal marker allele of animal 4 was the paternal or maternal marker allele of its sire. Thus, the marker information provided no more information than that from the pedigree.

The gametic relationship matrix for all the animals in the example pedigree is in Table 6.12. Unlike the relationships at the animal level, gametic relationships range from 0 to 1 i.e. they are the probability that QTL alleles are identical-by-descent. The offdiagonal elements in the 2×2 block diagonal submatrices are the gametic relationship between two QTL alleles carried by the same individual, which is the inbreeding coefficient at the QTL. To differentiate this inbreeding from Wright's inbreeding coefficient (denoted f_i earlier), the inbreeding at the QTL for individual i is denoted by

F_i. From Table 6.12 individuals 6 and 7 were inbred with inbreeding coefficients at the QTL of $F_6 = \mathbf{Q}_{11,12} = 0.050$ and $F_7 = \mathbf{Q}_{13,14} = 0.104$, respectively.

Table 6.12. Fernando-Grossman gametic relationship matrix ($\mathbf{Q}$).

QTL	Q_1^1	Q_1^2	Q_2^1	Q_2^2	Q_3^1	Q_3^2	Q_4^1	Q_4^2	Q_5^1	Q_5^2	Q_6^1	Q_6^2	Q_7^1	Q_7^2
Q_1^1	1.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.450	0.500	0.050	0.225	0.095
Q_1^2	0.000	1.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.450	0.500	0.050	0.225	0.095
Q_2^1	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.500	0.000	0.050	0.000	0.450	0.025	0.405
Q_2^2	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.500	0.000	0.050	0.000	0.450	0.025	0.405
Q_3^1	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.900	0.000	0.000	0.000	0.450	0.000
Q_3^2	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.100	0.000	0.000	0.000	0.050	0.000
Q_4^1	0.500	0.500	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.900	0.500	0.100	0.450	0.140
Q_4^2	0.000	0.000	0.500	0.500	0.000	0.000	0.000	1.000	0.000	0.100	0.000	0.900	0.050	0.810
Q_5^1	0.000	0.000	0.000	0.000	0.900	0.100	0.000	0.000	1.000	0.000	0.000	0.000	0.500	0.000
Q_5^2	0.450	0.450	0.050	0.050	0.000	0.000	0.900	0.100	0.000	1.000	0.450	0.180	0.500	0.207
Q_6^1	0.500	0.500	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.450	1.000	0.050	0.225	0.145
Q_6^2	0.050	0.050	0.450	0.450	0.000	0.000	0.100	0.900	0.000	0.180	0.050	1.000	0.090	0.905
Q_7^1	0.225	0.225	0.025	0.025	0.450	0.050	0.450	0.050	0.500	0.500	0.225	0.090	1.000	0.104
Q_7^2	0.095	0.095	0.405	0.405	0.000	0.000	0.140	0.810	0.000	0.207	0.145	0.905	0.104	1.000

The superscript refers to the QTL allele (1 or 2) and the subscript refers to animal number.

The tabular method can be expressed in matrix notation as follows (Fernando and Grossman, 1989)

$$\mathbf{Q} = (\mathbf{I} - \mathbf{J})^{-1}\mathbf{V}(\mathbf{I} - \mathbf{J}')^{-1} \qquad [6.25]$$

where $\mathbf{I}$ is an identity matrix, $\mathbf{J}$ is simple matrix with each row containing at most two non-zero elements (QTL transmission probabilities), $\mathbf{V}$ is a diagonal matrix with diagonal elements equal to $2(1-\rho_i^p)\rho_i^p(1-F_s)$ for paternal QTL if marker information is available and individual *i*'s sire is known or 1 if the sire is unknown; the diagonal for the maternal QTL equals $2(1-\rho_i^m)\rho_i^m(1-F_d)$ if marker information is available and the dam is known otherwise it is equal to 1.

The matrix $\mathbf{J}$ for the example pedigree is

$$\mathbf{J} = \begin{bmatrix}
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.9 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.9 & 0.1 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.9 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.1 & 0.9 & 0.0 & 0.0
\end{bmatrix}$$

The matrix $\mathbf{V}$ is $\mathbf{V} = diag\{1\ 1\ 1\ 1\ 1\ 1\ 0.5\ 0.5\ 0.18\ 0.18\ 0.5\ 0.18\ 0.5\ 0.17\}$. The GRM computed using equation [6.25] was the same as that in Table 6.12.

From equation [6.25], the inverse of $\mathbf{Q}$ is

$$\mathbf{Q}^{-1} = (\mathbf{I} - \mathbf{J}')\mathbf{V}^{-1}(\mathbf{I} - \mathbf{J}) \qquad [6.26]$$

Equation [6.25] was derived following the approach used by Quaas et al. (1984) and Quaas (1988) to invert the additive relationship matrix. Thus, the rules to obtain $\mathbf{Q}^{-1}$ given pedigree and marker information are analogous to the rules for $\mathbf{A}^{-1}$. The rules are as follows:

Let *k*, *s* and *d* denote individual *k*, its sire and dam respectively and $g_{i_k^p}$ be the diagonal element of $\mathbf{V}^{-1}$ for the paternal allele of individual *k*.

1) Compute diagonals of $\mathbf{V}^{-1}$

   - The diagonal is equal to $\frac{1}{2(1-\rho_k^p)\rho_k^p(1-F_s)}$ for paternal QTL allele

     or $\frac{1}{2(1-\rho_k^m)\rho_k^m(1-F_d)}$ for maternal QTL allele if the parent is known

     otherwise the diagonal is equal to 1

2) Initialize $\mathbf{Q}^{-1}$ to a null matrix

The contributions to $\mathbf{Q}^{-1}$ from individual k are as follows:

- If sire is known, add
  - i. $\left(1-\rho_k^p\right)^2 g_{i_k^p}$ to diagonal element i_s^p, i_s^p;
  - ii. $\left(1-\rho_k^p\right)\rho_k^p g_{i_k^p}$ to elements i_s^p, i_s^m and i_s^m, i_s^p;
  - iii. $-\left(1-\rho_k^p\right) g_{i_k^p}$ to elements i_s^p, i_k^p and i_k^p, i_s^p;
  - iv. $\left(\rho_k^p\right)^2 g_{i_k^p}$ to diagonal element i_s^m, i_s^m;
  - v. $-\rho_k^p g_{i_k^p}$ to elements i_s^m, i_k^p and i_k^p, i_s^m
- If dam is known, add
  - i. $\left(1-\rho_k^m\right)^2 g_{i_k^m}$ to diagonal element i_d^p, i_d^p;
  - ii. $\left(1-\rho_k^m\right)\rho_k^m g_{i_k^m}$ to elements i_d^p, i_d^m and i_d^m, i_d^p;
  - iii. $\left(1-\rho_k^m\right) g_{i_k^m}$ to elements i_d^p, i_k^m and i_k^m, i_d^p;
  - iv. $\left(\rho_k^m\right)^2 g_{i_k^m}$ to diagonal element i_d^m, i_d^m;
  - v. $-\rho_k^m g_{i_k^m}$ to elements i_d^m, i_k^m and i_k^m, i_d^m
- Always, add
  - i. $g_{i_k^p}$ to element i_k^p, i_k^p;

ii. $g_{i_k^m}$ to element i_k^m, i_k^m

The inverse of the GRM, $\mathbf{Q}^{-1}$, obtained using the rules above is

$$\mathbf{Q}^{-1} = \begin{bmatrix}
2.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & -1.00 & 0.00 & 0.00 & 0.00 & -1.00 & 0.00 & 0.00 & 0.00 \\
1.00 & 2.00 & 0.00 & 0.00 & 0.00 & 0.00 & -1.00 & 0.00 & 0.00 & 0.00 & -1.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 1.50 & 0.50 & 0.00 & 0.00 & 0.00 & -1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.50 & 1.50 & 0.00 & 0.00 & 0.00 & -1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 5.50 & 0.50 & 0.00 & 0.00 & -5.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.50 & 1.06 & 0.00 & 0.00 & -0.56 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
-1.00 & -1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 6.56 & 1.00 & 0.00 & -5.00 & 0.00 & -0.56 & 0.00 & 0.00 \\
0.00 & 0.00 & -1.00 & -1.00 & 0.00 & 0.00 & 1.00 & 6.56 & 0.00 & -0.56 & 0.00 & -5.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & -5.00 & -0.56 & 0.00 & 0.00 & 6.06 & 0.50 & 0.00 & 0.00 & -1.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -5.00 & -0.56 & 0.50 & 6.06 & 0.00 & 0.00 & -1.00 & 0.00 \\
-1.00 & -1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 2.06 & 0.53 & 0.00 & -0.59 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -0.56 & -5.00 & 0.00 & 0.00 & 0.53 & 10.29 & 0.00 & -5.26 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -1.00 & -1.00 & 0.00 & 0.00 & 2.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -0.59 & -5.26 & 0.00 & 5.85
\end{bmatrix}$$

6.8.2. Van Arendonk et al. (1994) approach

The theory to construct the GRM and its inverse was developed using partitioned matrix theory (Van Arendonk et al., 1994). Some of its desirable features include: its ability to handle multiple markers and marked QTL (MQTL), incomplete marker data, inbred populations and large data sets.

Gametes (or MQTL alleles) should be ordered chronologically to compute the GRM. For a single MQTL, the GRM for gametes 1 to k, $\mathbf{Q}_k$ is given by (Van Arendonk et al., 1994)

$$\mathbf{Q}_k = \begin{bmatrix} \mathbf{Q}_{k-1} & \mathbf{Q}_{k-1}\mathbf{n}_k \\ \mathbf{n}_k'\mathbf{Q}_{k-1} & \mathbf{Q}_{k,k} \end{bmatrix} \qquad [6.27]$$

where $\mathbf{n}_k$ is a column vector of order $(k-1)\times 1$ containing non-zero elements relating gamete k to gametes in the parent (if known) and zeroes elsewhere; $\mathbf{Q}_{k-1}$ is the GRM for gametes 1 to $(k-1)$; $\mathbf{Q}_{k,k}$ is the gametic relationship between the k^{th} gamete and itself which is equal to one i.e. $\mathbf{Q}_{k,k}$ is a scalar.

Equation [6.27] is equivalent to equation [6.25] of Fernando and Grossman (1989). The derivation of equation [6.27] from equation [6.25] proceeds as follows:

$$\mathbf{Q} = (\mathbf{I}-\mathbf{J})^{-1}\mathbf{V}(\mathbf{I}-\mathbf{J}')^{-1}$$

The matrix $(\mathbf{I}-\mathbf{J})_k^{-1}$ can be partitioned as follows:

$$(\mathbf{I}-\mathbf{J})_k^{-1} = \begin{bmatrix} (\mathbf{I}-\mathbf{J})_{k-1}^{-1} & \mathbf{0} \\ \mathbf{n}_k'(\mathbf{I}-\mathbf{J})_{k-1}^{-1} & 1 \end{bmatrix} \qquad [6.28]$$

Similarly,

$$(\mathbf{I}-\mathbf{J}')_k^{-1} = \begin{bmatrix} (\mathbf{I}-\mathbf{J}')_{k-1}^{-1} & (\mathbf{I}-\mathbf{J}')_{k-1}^{-1}\mathbf{n}_k \\ \mathbf{0}' & 1 \end{bmatrix} \qquad [6.29]$$

where $(\mathbf{I}-\mathbf{J}')_k^{-1}$ and $(\mathbf{I}-\mathbf{J}')_{k-1}^{-1}$ are submatrices for gametes 1 to k, and 1 to $(k-1)$, respectively; $\mathbf{0}$ is a column vector of order $(k-1)$.

The matrix $\mathbf{V}$ can be partitioned as

$$\mathbf{V}_k = \begin{bmatrix} \mathbf{V}_{k-1} & \mathbf{0} \\ \mathbf{0}' & \mathbf{V}_{k,k} \end{bmatrix} \qquad [6.30]$$

where $\mathbf{V}_{k,k}$ is a scalar.

Then,

$$\mathbf{Q}_k = \begin{bmatrix} (\mathbf{I}-\mathbf{J})_{k-1}^{-1} & \mathbf{0} \\ \mathbf{n}_k'(\mathbf{I}-\mathbf{J})_{k-1}^{-1} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{V}_{k-1} & \mathbf{0} \\ \mathbf{0}' & \mathbf{V}_{k,k} \end{bmatrix} \begin{bmatrix} (\mathbf{I}-\mathbf{J}')_{k-1}^{-1} & (\mathbf{I}-\mathbf{J}')_{k-1}^{-1}\mathbf{n}_k \\ \mathbf{0}' & 1 \end{bmatrix} \qquad [6.31]$$

Upon simplification, equation [6.31] becomes

$$\mathbf{Q}_k = \begin{bmatrix} (\mathbf{I}-\mathbf{J})_{k-1}^{-1}\mathbf{V}_{k-1}(\mathbf{I}-\mathbf{J}')_{k-1}^{-1} & (\mathbf{I}-\mathbf{J})_{k-1}^{-1}\mathbf{V}_{k-1}(\mathbf{I}-\mathbf{J}')_{k-1}^{-1}\mathbf{n} \\ \mathbf{n}_k'(\mathbf{I}-\mathbf{J})_{k-1}^{-1}\mathbf{V}_{k-1}(\mathbf{I}-\mathbf{J}')_{k-1}^{-1} & \mathbf{n}_k'(\mathbf{I}-\mathbf{J})_{k-1}^{-1}\mathbf{V}_{k-1}(\mathbf{I}-\mathbf{J}')_{k-1}^{-1}\mathbf{n}+\mathbf{V}_{k,k} \end{bmatrix} \qquad [6.32]$$

Since $(\mathbf{I}-\mathbf{J})_{k-1}^{-1}\mathbf{V}_{k-1}(\mathbf{I}-\mathbf{J}')_{k-1}^{-1} = \mathbf{Q}_{k-1}$, equation [6.32] takes the following form:

$$\mathbf{Q}_k = \begin{bmatrix} \mathbf{Q}_{k-1} & \mathbf{Q}_{k-1}\mathbf{n} \\ \mathbf{n}_k'\mathbf{Q}_{k-1} & \mathbf{n}_k'\mathbf{Q}_{k-1}\mathbf{n} + \mathbf{V}_{k,k} \end{bmatrix} \quad [6.33]$$

The element $\mathbf{n}_k'\mathbf{Q}_{k-1}\mathbf{n} + \mathbf{V}_{k,k}$ is the diagonal of the matrix $\mathbf{Q}_k$ and by definition is equal to one i.e. the gametic relationship between a gamete and itself. Thus, equation [6.33] becomes

$$\mathbf{Q}_k = \begin{bmatrix} \mathbf{Q}_{k-1} & \mathbf{Q}_{k-1}\mathbf{n} \\ \mathbf{n}_k'\mathbf{Q}_{k-1} & \mathbf{Q}_{k,k} \end{bmatrix} \quad [6.34]$$

Therefore, equations [6.25] and [6.27] are equivalent indicating that the approach of Fernando and Grossman (1989) and that of Van Arendonk et al. (1994) are equivalent.

For illustrative purpose, equation [6.34] was applied to the example pedigree in Table 6.11. The computation for $\mathbf{Q}_7$ (the paternal QTL of animal 4) is used as an example.

$$\mathbf{Q}_7 = \begin{bmatrix} \mathbf{Q}_6 & \mathbf{Q}_6\mathbf{n}_7 \\ \mathbf{n}_7'\mathbf{Q}_6 & \mathbf{Q}_{7,7} \end{bmatrix}$$

where the elements of $\mathbf{Q}_6$ are the gametic relationships between base animals and is an identity matrix of order 6, $\mathbf{Q}_{7,7}$ is a scalar always equal to 1. Based on marker information (i.e. the sire was homozygous at the marker locus), either of the two alleles

of the sire's QTL could have been transmitted to animal 4. The QTL transmission vector for the 7^{th} gamete is

$$\mathbf{n}_7 = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.0 \\ 0.0 \\ 0.0 \\ 0.0 \end{bmatrix}$$

and

$$\mathbf{n}_7'\mathbf{Q}_6 = \begin{bmatrix} 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix} \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}$$

Then,

$$\mathbf{Q}_7 = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}$$

As expected, the gametic relationship matrix for all animals in the pedigree example constructed using the Van Arendonk approach was the same as that obtained using the Fernando-Grossman approach given in Table 6.12.

From equation [6.34] the inverse of $\mathbf{Q}_k$ can be obtained recursively from

$$\mathbf{Q}_k^{-1} = \begin{bmatrix} \mathbf{Q}_{k-1}^{-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix} + \left(\mathbf{Q}_{k,k} - \mathbf{n}_k' \mathbf{Q}_{k-1} \mathbf{n}_k\right)^{-1} \begin{bmatrix} \mathbf{n}_k \mathbf{n}_k' & -\mathbf{n}_k \\ -\mathbf{n}_k' & 1 \end{bmatrix} \qquad [6.35]$$

Using equation [6.35] $\mathbf{Q}_7^{-1}$ is computed as follows

$$\mathbf{Q}_7^{-1} = \begin{bmatrix} \mathbf{Q}_6^{-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} + \left(\mathbf{Q}_{7,7} - \mathbf{n}_7' \mathbf{Q}_6 \mathbf{n}_7\right)^{-1} \begin{bmatrix} \mathbf{n}_7 \mathbf{n}_7' & -\mathbf{n}_7 \\ -\mathbf{n}_7' & 1 \end{bmatrix}$$

$$\mathbf{Q}_7^{-1} = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix} +$$

$$2\times \begin{bmatrix} 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & -1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & -1.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 0.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 0.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 0.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 0.00 \\ -1.00 & -1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 2.00 \end{bmatrix}$$

$$= \begin{bmatrix} 2.00 & 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & -1.00 \\ 1.00 & 2.00 & 1.00 & 1.00 & 1.00 & 1.00 & -1.00 \\ 1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 1.00 & 0.00 \\ 1.00 & 1.00 & 1.00 & 2.00 & 1.00 & 1.00 & 0.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 2.00 & 1.00 & 0.00 \\ 1.00 & 1.00 & 1.00 & 1.00 & 1.00 & 2.00 & 0.00 \\ -1.00 & -1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 2.00 \end{bmatrix}$$

The $\mathbf{Q}^{-1}$ for all the animals is the same as that obtained using the Fernando-Grossman approach as shown below:

$$\mathbf{Q}^{-1} = \begin{bmatrix}
2.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & -1.00 & 0.00 & 0.00 & 0.00 & -1.00 & 0.00 & 0.00 & 0.00 \\
1.00 & 2.00 & 0.00 & 0.00 & 0.00 & 0.00 & -1.00 & 0.00 & 0.00 & 0.00 & -1.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 1.50 & 0.50 & 0.00 & 0.00 & 0.00 & -1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.50 & 1.50 & 0.00 & 0.00 & 0.00 & -1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 5.50 & 0.50 & 0.00 & 0.00 & -5.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.50 & 1.06 & 0.00 & 0.00 & -0.56 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
-1.00 & -1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 6.56 & 1.00 & 0.00 & -5.00 & 0.00 & -0.56 & 0.00 & 0.00 \\
0.00 & 0.00 & -1.00 & -1.00 & 0.00 & 0.00 & 1.00 & 6.56 & 0.00 & -0.56 & 0.00 & -5.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & -5.00 & -0.56 & 0.00 & 0.00 & 6.06 & 0.50 & 0.00 & 0.00 & -1.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -5.00 & -0.56 & 0.50 & 6.06 & 0.00 & 0.00 & -1.00 & 0.00 \\
-1.00 & -1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 2.06 & 0.53 & 0.00 & -0.59 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -0.56 & -5.00 & 0.00 & 0.00 & 0.53 & 10.29 & 0.00 & -5.26 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -1.00 & -1.00 & 0.00 & 0.00 & 2.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & -0.59 & -5.26 & 0.00 & 5.85
\end{bmatrix}$$

6.8.3. Wang et al. (1995) approach

This approach is an extension of the Fernando and Grossman's (1989) approach. It accommodates unknown parental origin of marker alleles implying that for two individuals, say i and j, that are not descendents of each other (Figure 6.4), there are four possible ways that the k^{th} (for $k = 1, 2$) QTL allele from individual i is identical-by-descent to the m^{th} (for $m = 1, 2$) QTL allele in individual j:

	$Q_s^1 \equiv Q_j^m$	$Q_s^2 \equiv Q_j^m$	$Q_d^1 \equiv Q_j^m$	$Q_d^2 \equiv Q_j^m$
$Q_i^1 \Leftarrow Q_s^1$	$Q_i^1 \equiv Q_j^m$ IBD	not IBD	not IBD	not IBD
$Q_i^1 \Leftarrow Q_s^2$	not IBD	$Q_i^1 \equiv Q_j^m$ IBD	not IBD	not IBD
$Q_i^1 \Leftarrow Q_d^1$	not IBD	not IBD	$Q_i^1 \equiv Q_j^m$ IBD	not IBD
$Q_i^1 \Leftarrow Q_d^2$	not IBD	not IBD	not IBD	$Q_i^1 \equiv Q_j^m$ IBD

The conditional probability that Q_i^k is identical-by-descent to Q_j^m given marker information is

$$\begin{aligned}\Pr\left(Q_i^k \equiv Q_j^m\right) = \Pr\left(Q_s^1 \equiv Q_j^m\right).\Pr\left(Q_i^1 \Leftarrow Q_s^1\right) + \Pr\left(Q_s^2 \equiv Q_j^m\right).\Pr\left(Q_i^1 \Leftarrow Q_s^2\right) \\ + \Pr\left(Q_d^1 \equiv Q_j^m\right).\Pr\left(Q_i^1 \Leftarrow Q_d^1\right) + \Pr\left(Q_d^2 \equiv Q_j^m\right).\Pr\left(Q_i^1 \Leftarrow Q_d^2\right)\end{aligned} \quad [6.36]$$

where s^1 and s^2, and d^1 and d^2 are the first and second QTL alleles of the sire and dam, respectively; $\Pr\left(Q_i^1 \Leftarrow Q_s^1\right)$ is the probability that allele Q_i^1 of individual *i* was inherited from allele Q_s^1 of its sire. Equation [6.36] reduces to equations [6.21] and [6.22] given earlier in the section on the approach by Fernando and Grossman (1989) when the parental origin of the marker allele is known.

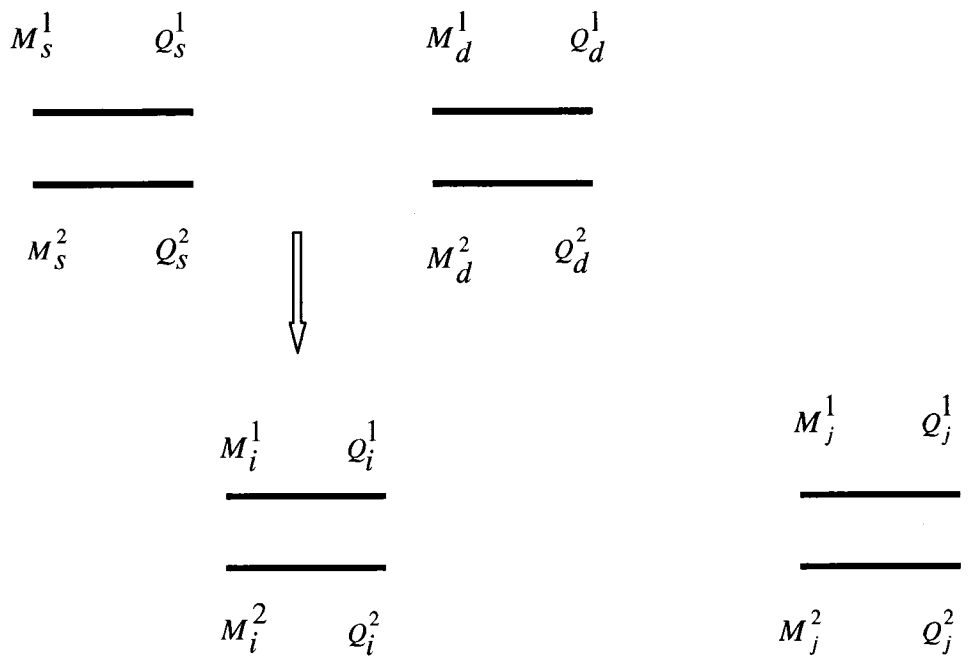


Figure 6.4. Haplotypes at the marker (M) and QTL (Q) locus for individuals i, j, s and d.

The conditional QTL transmission probabilities from s and d to individual i given the marker information can be computed as follows:

$$\mathbf{B}_i = \mathbf{S}_i\mathbf{E} \qquad [6.37]$$

where

$$\mathbf{B}_i = \begin{bmatrix} \Pr\left(Q_i^1 \Leftarrow Q_s^1\right) & \Pr\left(Q_i^1 \Leftarrow Q_s^2\right) & \Pr\left(Q_i^1 \Leftarrow Q_d^1\right) & \Pr\left(Q_i^1 \Leftarrow Q_d^2\right) \\ \Pr\left(Q_i^2 \Leftarrow Q_s^1\right) & \Pr\left(Q_i^2 \Leftarrow Q_s^2\right) & \Pr\left(Q_i^2 \Leftarrow Q_d^1\right) & \Pr\left(Q_i^2 \Leftarrow Q_d^2\right) \end{bmatrix}$$

$$\mathbf{S}_i = \begin{bmatrix} \Pr\left(M_i^1 \Leftarrow M_s^1\right) & \Pr\left(M_i^1 \Leftarrow M_s^2\right) & \Pr\left(M_i^1 \Leftarrow M_d^1\right) & \Pr\left(M_i^1 \Leftarrow M_d^2\right) \\ \Pr\left(M_i^2 \Leftarrow M_s^1\right) & \Pr\left(M_i^2 \Leftarrow M_s^2\right) & \Pr\left(M_i^2 \Leftarrow M_d^1\right) & \Pr\left(M_i^2 \Leftarrow M_d^2\right) \end{bmatrix}$$

and

$$\mathbf{E} = \begin{bmatrix} 1-r & r & 0 & 0 \\ r & 1-r & 0 & 0 \\ 0 & 0 & 1-r & r \\ 0 & 0 & r & 1-r \end{bmatrix}$$

where $Q_i^1 \Leftarrow Q_s^1$ and $M_i^1 \Leftarrow M_s^1$ are the events that Q_s^1 and M_s^1 were transmitted to individual i (Q_i^1 and M_i^1) given the observed marker genotypes of individual i, s and d, and r is as defined before.

Therefore, equation [6.36] can be expressed as

$$\begin{aligned} \Pr\left(Q_i^k \equiv Q_j^m\right) = \Pr\left(Q_s^1 \equiv Q_j^m\right).\mathbf{B}_{k,1} + \Pr\left(Q_s^2 \equiv Q_j^m\right).\mathbf{B}_{k,2} \\ + \Pr\left(Q_d^1 \equiv Q_j^m\right).\mathbf{B}_{k,3} + \Pr\left(Q_d^2 \equiv Q_j^m\right).\mathbf{B}_{k,4} \end{aligned} \qquad [6.38]$$

where $\mathbf{B}_{k,1}$ is the $(k,1)$ element of the matrix $\mathbf{B}_i$.

From equation [6.38] a tabular method can be used to build the gametic relationship matrix ($\mathbf{Q}$) row at a time. The row elements corresponding to the two QTL alleles of individual i (Q_i^k where $k=1,2$) is given by:

$$\mathbf{Q}_{\delta_i^k,j} = \mathbf{Q}_{\delta_s^1,j}\mathbf{B}_{k,1} + \mathbf{Q}_{\delta_r^2,j}\mathbf{B}_{k,2} + \mathbf{Q}_{\delta_d^1,j}\mathbf{B}_{k,3} + \mathbf{Q}_{\delta_d^2,j}\mathbf{B}_{k,4} \qquad [6.39]$$

where $\delta_i^1 = 2(i-1)+1$, $\delta_i^2 = 2(i-1)+2$, and $j=1,...,\delta_i^1-1$; element $\mathbf{Q}_{\delta_i^2,\delta_i^1} = F_i$ where F_i is as defined before.

The tabular method can be represented in matrix notation following Van Arendonk et al. (1994) as follows:

$$\mathbf{Q}_i = \begin{bmatrix} \mathbf{Q}_{i-1} & \mathbf{Q}_{i-1}\mathbf{N}_i \\ \mathbf{N}_i'\mathbf{Q}_{i-1} & \mathbf{Q}_{i,i} \end{bmatrix} \qquad [6.40]$$

where $\mathbf{N}_i$ is a $2(i-1)\times 2$ matrix with the maximum of eight non-zero elements located in rows of the parental gametes of individual i (if the parents are known); these are the elements of $\mathbf{B}_i$, $\mathbf{Q}_{i,i}$ is 2×2 matrix of the IBD between two homologous QTL alleles of individual i which can be written as

$$\mathbf{Q}_{i,i} = \begin{bmatrix} 1 & F_i \\ F_i & 1 \end{bmatrix} \qquad [6.41]$$

and

$$F_i = \sum_{k=1}^{2}\sum_{m=1}^{2} \mathbf{Q}_{\delta_s^k,\delta_d^m} \Pr\left(T_{Q_s^k,Q_d^m}\right) \qquad [6.42]$$

where $\mathbf{Q}_{\delta_s^k,\delta_d^m}$ is the probability that Q_s^k is identical-by-descent to Q_d^m, $T_{Q_s^k,Q_d^m}$ is the event that homologous alleles in individual i descended from the parental pair Q_s^k and Q_d^m conditional on marker information and is given by

$$\Pr\left(T_{Q_s^k,Q_d^m}\right) = \frac{\Pr\left(Q_i^1 \Leftarrow Q_s^k\right)\Pr\left(Q_i^2 \Leftarrow Q_d^m\right)}{\Pr\left(Q_i^1 \Leftarrow Q_s^1\right)+\Pr\left(Q_i^1 \Leftarrow Q_s^2\right)} + \frac{\Pr\left(Q_i^1 \Leftarrow Q_d^m\right)\Pr\left(Q_i^2 \Leftarrow Q_s^k\right)}{\Pr\left(Q_i^1 \Leftarrow Q_d^1\right)+\Pr\left(Q_i^1 \Leftarrow Q_d^2\right)} \qquad [6.43]$$

The QTL transmission probabilities in equation [6.43] are the elements of $\mathbf{B}_i$ and therefore $\Pr\left(T_{Q_s^k,Q_d^m}\right)$ can be expressed as a function of $\mathbf{B}_i$. For example, the probability of an event that Q_s^1 and Q_d^1 descended to individual i is

$$\Pr\left(T_{Q_s^1,Q_d^1}\right) = \frac{\mathbf{B}_{1,1}.\mathbf{B}_{2,3}}{\mathbf{B}_{1,1}+\mathbf{B}_{2,2}} + \frac{\mathbf{B}_{1,3}.\mathbf{B}_{2,1}}{\mathbf{B}_{1,3}+\mathbf{B}_{2,4}}$$

Individual 7 is used as an example to illustrate the computation of $\mathbf{Q}_7$ using equation [6.40]. The matrix $\mathbf{Q}_6$ is

$$\mathbf{Q}_6 = \begin{bmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.50 & 0.00 & 0.23 & 0.23 & 0.50 & 0.05 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.50 & 0.00 & 0.23 & 0.23 & 0.50 & 0.05 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.50 & 0.03 & 0.03 & 0.00 & 0.45 \\ 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.50 & 0.03 & 0.03 & 0.00 & 0.45 \\ 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.45 & 0.45 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.05 & 0.05 & 0.00 & 0.00 \\ 0.50 & 0.50 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.45 & 0.45 & 0.50 & 0.10 \\ 0.00 & 0.00 & 0.50 & 0.50 & 0.00 & 0.00 & 0.00 & 1.00 & 0.05 & 0.05 & 0.00 & 0.90 \\ 0.23 & 0.23 & 0.03 & 0.03 & 0.45 & 0.05 & 0.45 & 0.05 & 1.00 & 0.00 & 0.23 & 0.09 \\ 0.23 & 0.23 & 0.03 & 0.03 & 0.45 & 0.05 & 0.45 & 0.05 & 0.00 & 1.00 & 0.23 & 0.09 \\ 0.50 & 0.50 & 0.00 & 0.00 & 0.00 & 0.00 & 0.50 & 0.00 & 0.23 & 0.23 & 1.00 & 0.05 \\ 0.05 & 0.05 & 0.45 & 0.45 & 0.00 & 0.00 & 0.10 & 0.90 & 0.09 & 0.09 & 0.05 & 1.00 \end{bmatrix}$$

The matrix $\mathbf{B}_7$ is

$$\mathbf{B}_7 = \mathbf{S}_7\mathbf{E}$$

$$= \begin{bmatrix} 0.5 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.00 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 & 0.0 & 0.0 \\ 0.1 & 0.9 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.9 & 0.1 \\ 0.0 & 0.0 & 0.1 & 0.9 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.1 & 0.9 \end{bmatrix}$$

Then,

$$\mathbf{N}_7' = \begin{bmatrix} 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.50 & 0.50 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.10 & 0.90 \end{bmatrix}$$

Multiplying $\mathbf{N}_7'$ by $\mathbf{Q}_6$ gives

$$\mathbf{N}_7'\mathbf{Q}_6 = \begin{bmatrix} 0.225 & 0.225 & 0.025 & 0.025 & 0.450 & 0.050 & 0.450 & 0.050 & 0.500 & 0.500 & 0.225 & 0.090 \\ 0.095 & 0.095 & 0.405 & 0.405 & 0.000 & 0.000 & 0.140 & 0.810 & 0.104 & 0.104 & 0.145 & 0.905 \end{bmatrix}$$

The inbreeding of individual 7 at the QTL locus is computed as follows:

$$\Pr\left(T_{Q_s^1,Q_d^1}\right) = \frac{\mathbf{B}_{1,1}.\mathbf{B}_{2,3}}{\mathbf{B}_{1,1}+\mathbf{B}_{2,2}} + \frac{\mathbf{B}_{1,3}.\mathbf{B}_{2,1}}{\mathbf{B}_{1,3}+\mathbf{B}_{2,4}} = \frac{(0.5)(0.1)}{(0.5+0.5)} + \frac{(0.0)(0.0)}{(0.0+0.9)} = 0.05$$

$$\Pr\left(T_{Q_s^1,Q_d^2}\right) = \frac{\mathbf{B}_{1,1}.\mathbf{B}_{2,4}}{\mathbf{B}_{1,1}+\mathbf{B}_{1,2}} + \frac{\mathbf{B}_{2,4}.\mathbf{B}_{2,1}}{\mathbf{B}_{1,3}+\mathbf{B}_{2,4}} = \frac{(0.5)(0.9)}{(0.5+0.5)} + \frac{(0.0)(0.0)}{(0.0+0.9)} = 0.45$$

$$\Pr\left(T_{Q_s^2,Q_d^1}\right) = \frac{\mathbf{B}_{1,2}.\mathbf{B}_{2,3}}{\mathbf{B}_{1,1}+\mathbf{B}_{1,2}} + \frac{\mathbf{B}_{1,3}.\mathbf{B}_{2,2}}{\mathbf{B}_{1,3}+\mathbf{B}_{1,4}} = \frac{(0.5)(0.1)}{(0.5+0.5)} + \frac{(0.0)(0.0)}{(0.0+0.0)} = 0.05$$

$$\Pr\left(T_{Q_s^2,Q_d^2}\right) = \frac{\mathbf{B}_{1,2}.\mathbf{B}_{2,4}}{\mathbf{B}_{1,1}+\mathbf{B}_{1,2}} + \frac{\mathbf{B}_{1,4}.\mathbf{B}_{2,2}}{\mathbf{B}_{1,3}+\mathbf{B}_{1,4}} = \frac{(0.5)(0.9)}{(0.5+0.5)} + \frac{(0.0)(0.0)}{(0.0+0.0)} = 0.45$$

Then,

$$F_7 = \mathbf{Q}_{\delta_s^1,\delta_d^1}\Pr\left(T_{Q_s^1,Q_d^1}\right)+\mathbf{Q}_{\delta_s^1,\delta_d^2}\Pr\left(T_{Q_s^1,Q_d^2}\right)+\mathbf{Q}_{\delta_s^2,\delta_d^1}\Pr\left(T_{Q_s^2,Q_d^1}\right)+\mathbf{Q}_{\delta_s^2,\delta_d^2}\Pr\left(T_{Q_s^2,Q_d^2}\right)$$

$$=(0.23\times0.05)+(0.09\times0.45)+(0.23\times0.05)+(0.09\times0.45)$$

$$=0.104$$

Substituting F_7 into equation [6.41] gives the conditional gametic relationship between the two gametes (Q_i^1 and Q_i^2) carried by individual 7

$$\mathbf{Q}_{7,7}=\begin{bmatrix}1.000 & F_i\\ F_i & 1.000\end{bmatrix}=\begin{bmatrix}1.000 & 0.104\\ 0.104 & 1.000\end{bmatrix}$$

Adding the contribution of individual 7 to $\mathbf{Q}_6$ yields

$$\mathbf{Q}_7 = \begin{bmatrix}
1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.500 & 0.000 & 0.225 & 0.225 & 0.500 & 0.050 & 0.225 & 0.095\\
0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.500 & 0.000 & 0.225 & 0.225 & 0.500 & 0.050 & 0.225 & 0.095\\
0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.500 & 0.025 & 0.025 & 0.000 & 0.450 & 0.025 & 0.405\\
0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.500 & 0.025 & 0.025 & 0.000 & 0.450 & 0.025 & 0.405\\
0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.450 & 0.450 & 0.000 & 0.000 & 0.450 & 0.000\\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.050 & 0.050 & 0.000 & 0.000 & 0.050 & 0.000\\
0.500 & 0.500 & 0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.450 & 0.450 & 0.500 & 0.100 & 0.450 & 0.140\\
0.000 & 0.000 & 0.500 & 0.500 & 0.000 & 0.000 & 0.000 & 1.000 & 0.050 & 0.050 & 0.000 & 0.900 & 0.050 & 0.810\\
0.225 & 0.225 & 0.025 & 0.025 & 0.450 & 0.050 & 0.450 & 0.050 & 1.000 & 0.000 & 0.225 & 0.090 & 0.500 & 0.104\\
0.225 & 0.225 & 0.025 & 0.025 & 0.450 & 0.050 & 0.450 & 0.050 & 0.000 & 1.000 & 0.225 & 0.090 & 0.500 & 0.104\\
0.500 & 0.500 & 0.000 & 0.000 & 0.000 & 0.000 & 0.500 & 0.000 & 0.225 & 0.225 & 1.000 & 0.050 & 0.225 & 0.145\\
0.050 & 0.050 & 0.450 & 0.450 & 0.000 & 0.000 & 0.100 & 0.900 & 0.090 & 0.090 & 0.050 & 1.000 & 0.090 & 0.905\\
0.225 & 0.225 & 0.025 & 0.025 & 0.450 & 0.050 & 0.450 & 0.050 & 0.500 & 0.500 & 0.225 & 0.090 & 1.000 & 0.104\\
0.095 & 0.095 & 0.405 & 0.405 & 0.000 & 0.000 & 0.140 & 0.810 & 0.104 & 0.104 & 0.145 & 0.905 & 0.104 & 1.000
\end{bmatrix}$$

The GRM given above is different from that obtained using Fernando-Grossman approach which was expected due to the difference in the assumption of the knowledge of the parental origin of marker.

The inverse of $\mathbf{Q}_i$ is

$$\mathbf{Q}_i^{-1} = \begin{bmatrix} \mathbf{Q}_{i-1}^{-1} & \mathbf{0} \\ \mathbf{0}' & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{N}_i \mathbf{V}_{i,i}^{-1} \mathbf{N}_i' & -\mathbf{V}_{i,i}^{-1} \mathbf{N}_i \\ -\mathbf{N}_i' \mathbf{V}_i^{-1} & \mathbf{V}_{i,i}^{-1} \end{bmatrix} \qquad [6.43]$$

where $\mathbf{V}_i = \mathbf{Q}_{i,i} - \mathbf{N}_i' \mathbf{Q}_{i-1} \mathbf{N}_i$ is a 2×2 matrix and the second part in the right-hand side of equation [6.43] is the contribution of individual i to $\mathbf{Q}_i^{-1}$ with at most 36 non-zero elements.

The relationship between $\mathbf{N}_i$ and $\mathbf{B}_i$ shown earlier in equation [6.40] allows $\mathbf{V}_{i,i}$ to be expressed as follows:

$$\mathbf{V}_{i,i} = \mathbf{Q}_{i,i} - \mathbf{B}_i \mathbf{Q}_{s,d} \mathbf{B}_i' \qquad [6.44]$$

where $\mathbf{Q}_{s,d}$ is a 4×4 gametic relationship matrix for the parents of individual i (i.e. s and d). If individual i and its parents are not inbred (i.e. $F_i = F_s = F_d = 0$) equation [6.44] reduces to

$$\mathbf{V}_{i,i} = \mathbf{I}_2 - \mathbf{B}_i\mathbf{B}_i' \qquad [6.45]$$

where $\mathbf{I}_2$ is a 2×2 identity matrix. Equation [6.44] indicates that the gametic relationship matrix is required to obtain its inverse in an inbred population (with respect to the QTL locus) while equation [6.45] shows that the gametic relationship matrix is not required to obtain its inverse in a non-inbred population.

Therefore, the contribution of individual *i* (the second term on the right-hand side of [6.43]) is given by a 6×6 matrix

$$\mathbf{W}_i = \begin{bmatrix} \mathbf{B}_i'\mathbf{V}_{i,i}^{-1}\mathbf{B}_i & -\mathbf{B}'\mathbf{V}_{i,i}^{-1} \\ -\mathbf{V}_{i,i}^{-1}\mathbf{B}_i & \mathbf{V}_{i,i}^{-1} \end{bmatrix} \qquad [6.46]$$

The position of contribution of element $\mathbf{W}_i(l,k)$ (for $l = 1,...,6$ and $k = 1,...,6$) to $\mathbf{Q}_i^{-1}$ is contained in the matrix

$$\Pi_i = \begin{bmatrix} (\delta_s^1,\delta_s^1) & (\delta_s^1,\delta_s^2) & (\delta_s^1,\delta_d^1) & (\delta_s^1,\delta_d^2) & (\delta_s^1,\delta_i^1) & (\delta_s^1,\delta_i^2) \\ (\delta_s^2,\delta_s^1) & (\delta_s^2,\delta_s^2) & (\delta_s^2,\delta_d^1) & (\delta_s^2,\delta_d^2) & (\delta_s^2,\delta_i^1) & (\delta_s^2,\delta_i^2) \\ (\delta_d^1,\delta_s^1) & (\delta_d^1,\delta_s^2) & (\delta_d^1,\delta_d^1) & (\delta_d^1,\delta_d^2) & (\delta_d^1,\delta_i^1) & (\delta_d^1,\delta_i^2) \\ (\delta_d^2,\delta_s^1) & (\delta_d^2,\delta_s^2) & (\delta_d^2,\delta_d^1) & (\delta_d^2,\delta_d^2) & (\delta_d^2,\delta_i^1) & (\delta_d^2,\delta_i^2) \\ (\delta_i^1,\delta_s^1) & (\delta_i^1,\delta_s^2) & (\delta_i^1,\delta_d^1) & (\delta_i^1,\delta_d^2) & (\delta_i^1,\delta_i^1) & (\delta_i^1,\delta_i^2) \\ (\delta_i^2,\delta_s^1) & (\delta_i^2,\delta_s^2) & (\delta_i^2,\delta_d^1) & (\delta_i^2,\delta_d^2) & (\delta_i^2,\delta_i^1) & (\delta_i^2,\delta_i^2) \end{bmatrix} \qquad [6.47]$$

where $\delta_a^b = 2(a-1)+b$ for $a = s, d,$ or i and $b = 1$ or 2. Elements of $\prod_i$ are not defined if the parent associated with that element is unknown.

Equations [6.43 to 6.47] lead to the following algorithm to compute $\mathbf{Q}_i^{-1}$:

1. Initialize $\mathbf{Q}^{-1}$ to a null matrix
2. For individual i, $i = 1,...,n$:
    a. If both parents are unknown, add 1s to $\mathbf{Q}^{-1}_{\delta_i^1,\delta_i^1}$ and $\mathbf{Q}^{-1}_{\delta_i^2,\delta_i^2}$
    b. If at least 1 parent is known:
        i. Compute $\mathbf{B}_i$, $\mathbf{V}_{i,i}$ and $\mathbf{W}_i$
        ii. For each defined element in $\prod_i$ add element $\mathbf{W}_i(l,k)$ to $\mathbf{Q}^{-1}$ at position $\prod_i(l,k)$

The contribution of animal 7 to $\mathbf{Q}_6^{-1}$ is used to illustrate the algorithm given above. Since both parents of animal 7 are known, the matrices $\mathbf{B}_7$, $\mathbf{V}_{7,7}$, $\mathbf{W}_7$ and $\prod_7$ need to be computed. The matrix $\mathbf{B}_7$ was given earlier in the section on computing the gametic relationship. Equation [6.44] is used to compute $\mathbf{V}_{7,7}$ because animal 7 is inbred:

$$\mathbf{V}_{7,7} = \mathbf{Q}_{7,7} - \mathbf{B}_7\mathbf{Q}_{5,6}\mathbf{B}_7' \qquad [6.48]$$

$$\mathbf{V}_{7,7} = \begin{bmatrix} 1.000 & 0.104 \\ 0.104 & 1.000 \end{bmatrix} - \begin{bmatrix} 0.500 & 0.500 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.100 & 0.900 \end{bmatrix}$$

$$\begin{bmatrix} 1.000 & 0.000 & 0.225 & 0.090 \\ 0.000 & 1.000 & 0.225 & 0.090 \\ 0.225 & 0.225 & 1.000 & 0.050 \\ 0.090 & 0.090 & 0.050 & 1.000 \end{bmatrix} \begin{bmatrix} 0.500 & 0.000 \\ 0.500 & 0.000 \\ 0.000 & 0.100 \\ 0.000 & 0.900 \end{bmatrix}$$

$$= \begin{bmatrix} 0.500 & 0.001 \\ 0.001 & 0.171 \end{bmatrix}$$

and

$$\mathbf{V}_{7,7}^{-1} = \begin{bmatrix} 2.000 & -0.006 \\ -0.006 & 5.848 \end{bmatrix}$$

The matrix $\mathbf{W}_7$ is obtained as follows

$$\mathbf{W}_7 = \begin{bmatrix} \mathbf{B}_7' \mathbf{V}_{7,7}^{-1} \mathbf{B}_7 & -\mathbf{B}_7' \mathbf{V}_{7,7}^{-1} \\ -\mathbf{V}_{7,7}^{-1} \mathbf{B}_7 & \mathbf{V}_{7,7}^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{W}_{1,1} & \mathbf{W}_{1,2} \\ \mathbf{W}_{2,1} & \mathbf{W}_{2,2} \end{bmatrix}$$

Then,

$$\mathbf{W}_{1,1} = \mathbf{B}_7' \mathbf{V}_{7,7}^{-1} \mathbf{B}_7$$

$$= \begin{bmatrix} 0.500 & 0.000 \\ 0.500 & 0.000 \\ 0.000 & 0.100 \\ 0.000 & 0.900 \end{bmatrix} \begin{bmatrix} 2.000 & -0.006 \\ -0.006 & 5.848 \end{bmatrix} \begin{bmatrix} 0.500 & 0.500 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.100 & 0.900 \end{bmatrix}$$

$$= \begin{bmatrix} 0.500 & 0.500 & 0.000 & -0.003 \\ 0.500 & 0.500 & 0.000 & -0.003 \\ 0.000 & 0.000 & 0.058 & 0.526 \\ -0.003 & -0.003 & 0.526 & 4.737 \end{bmatrix}$$

$$\mathbf{W}_{1,2} = -\mathbf{B}_7' \mathbf{V}_{7,7}^{-1}$$

$$= \begin{bmatrix} -0.500 & 0.000 \\ -0.500 & 0.000 \\ 0.000 & -0.100 \\ 0.000 & -0.900 \end{bmatrix} \begin{bmatrix} 2.000 & -0.006 \\ -0.006 & 5.848 \end{bmatrix}$$

$$= \begin{bmatrix} -1.000 & 0.003 \\ -1.000 & 0.003 \\ 0.001 & -0.585 \\ 0.005 & -5.263 \end{bmatrix}$$

$$\mathbf{W}_{1,2} = \mathbf{W}_{1,2}' \text{ and } \mathbf{W}_{2,2} = \mathbf{V}_{7,7}^{-1}$$

The contribution of individual 7 to $\mathbf{Q}^{-1}$ is given by

$$\mathbf{W}_7 = \begin{bmatrix} 0.500 & 0.500 & 0.000 & -0.003 & -1.000 & 0.003 \\ 0.500 & 0.500 & 0.000 & -0.003 & -1.000 & 0.003 \\ 0.000 & 0.000 & 0.580 & 0.526 & 0.001 & -0.585 \\ -0.003 & -0.003 & 0.526 & 4.737 & 0.005 & -5.263 \\ -1.000 & -1.000 & 0.001 & 0.005 & 2.000 & -0.006 \\ 0.003 & 0.003 & -0.585 & -5.263 & -0.006 & 5.848 \end{bmatrix}$$

and the contribution of $\mathbf{W}_7$ are in the following position of $\mathbf{Q}^{-1}$

$$\mathbf{\Pi}_7 = \begin{bmatrix} (9,9) & (9,10) & (9,11) & (9,12) & (9,13) & (9,14) \\ (10,9) & (10,10) & (10,11) & (10,12) & (10,13) & (10,14) \\ (11,9) & (11,10) & (11,11) & (11,12) & (11,13) & (11,14) \\ (12,9) & (12,10) & (12,11) & (12,12) & (12,13) & (12,14) \\ (13,9) & (13,10) & (13,11) & (13,12) & (13,13) & (13,14) \\ (14,9) & (14,10) & (14,11) & (14,12) & (14,13) & (14,14) \end{bmatrix}$$

Then, the inverse of the conditional gametic relationship matrix ($\mathbf{Q}^{-1}$) for the pedigree example is

$$\mathbf{Q}_7^{-1} = \mathbf{Q}^{-1} = \begin{bmatrix} 2.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -1.000 & 0.000 & 0.000 & 0.000 & -1.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 2.000 & 0.000 & 0.000 & 0.000 & 0.000 & -1.000 & 0.000 & 0.000 & 0.000 & -1.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 1.500 & 0.000 & 0.000 & 0.000 & 0.000 & -1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 1.500 & 0.000 & 0.000 & 0.000 & -1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 3.250 & 0.250 & 2.250 & 0.250 & -2.500 & -2.500 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.250 & 1.028 & 0.250 & 0.028 & -0.278 & -0.278 & 0.000 & 0.000 & 0.000 & 0.000 \\ -1.000 & -1.000 & 0.000 & 0.000 & 2.250 & 0.250 & 4.306 & 0.750 & -2.500 & -2.500 & 0.000 & -0.556 & 0.000 & 0.000 \\ 0.000 & 0.000 & -1.000 & -1.000 & 0.250 & 0.028 & 0.750 & 6.528 & -0.278 & -0.278 & 0.000 & -5.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.000 & 0.000 & -2.500 & -0.278 & -2.500 & -0.278 & 3.778 & 2.778 & 0.000 & -0.003 & -1.000 & 0.003 \\ 0.000 & 0.000 & 0.000 & 0.000 & -2.500 & -0.278 & -2.500 & -0.278 & 2.778 & 3.778 & 0.000 & -0.003 & -1.000 & 0.003 \\ -1.000 & -1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 2.058 & 0.526 & 0.001 & -0.585 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.556 & -5.000 & -0.003 & -0.003 & 0.526 & 10.292 & 0.005 & -5.263 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -1.000 & -1.000 & 0.001 & 0.005 & 2.000 & -0.006 \\ 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.003 & 0.003 & -0.585 & -5.263 & -0.006 & 5.848 \end{bmatrix}$$

6.8.4. Abdel-Azim and Freeman (2001) approach

The conditional gametic relationship matrix, $\mathbf{Q}$, was obtained using Henderson (1976)'s approach of matrix decomposition used in building the numerator relationship matrix in section 6.3. Abdel-Azim and Freeman (2001) showed that the matrix $\mathbf{Q}$, obtained using the approach of Wang et al. (1995), can be decomposed as follows:

$$\mathbf{Q} = \mathbf{LVL}' \qquad [6.49]$$

where $\mathbf{L}$ is a lower triangular matrix and $\mathbf{V}$ is as defined before and can be obtained recursively using equations [6.44 and 6.45] given earlier. The $\mathbf{L}$ matrix can be obtained recursively for each individual, say *i*, as shown below

$$\mathbf{L}_i = \begin{bmatrix} \mathbf{L}_{i-1} & \mathbf{0} \\ \mathbf{N}_i'\mathbf{L}_{i-1} & \mathbf{I}_2 \end{bmatrix} \qquad [6.50]$$

where $\mathbf{N}_i'$ and $\mathbf{I}_2$ are as defined earlier.

Equation [6.50] was applied to the pedigree example in Table 6.11. Animal 4 is used as an example. The matrices $\mathbf{N}_4'$ and $\mathbf{I}_2$ are

$$\mathbf{N}_4' = \begin{bmatrix} 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.5 & 0.0 & 0.0 \end{bmatrix}$$

and

$$\mathbf{I}_2 = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}$$

The nonzero elements in the first and second rows of $\mathbf{N}_4'$ indicate that either of parental QTL alleles might have been transmitted to animal 4. The **L** matrix computed using equation [6.50] is given below:

$$\mathbf{L} = \begin{bmatrix}
1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.50 & 0.50 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.50 & 0.50 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.23 & 0.23 & 0.03 & 0.03 & 0.45 & 0.45 & 0.45 & 0.45 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.23 & 0.23 & 0.03 & 0.03 & 0.45 & 0.05 & 0.45 & 0.05 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.50 & 0.50 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\
0.05 & 0.05 & 0.45 & 0.45 & 0.00 & 0.00 & 0.10 & 0.90 & 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\
0.23 & 0.23 & 0.03 & 0.03 & 0.45 & 0.05 & 0.45 & 0.05 & 0.50 & 0.50 & 0.00 & 0.00 & 1.00 & 0.00 \\
0.10 & 0.10 & 0.41 & 0.41 & 0.00 & 0.00 & 0.09 & 0.81 & 0.00 & 0.00 & 0.10 & 0.90 & 0.00 & 1.00
\end{bmatrix}$$

The **V** matrix (obtained using equations [6.44 and 6.45]) is

$$\mathbf{V} = \begin{bmatrix}
1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.500 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.500 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.500 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.590 & -0.410 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & -0.410 & 0.590 & 0.000 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.500 & 0.000 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.180 & 0.000 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.500 & 0.001 \\
0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.001 & 0.171
\end{bmatrix}$$

Multiplication of $\mathbf{LVL}'$ gave the same conditional gametic relationship matrix obtained using the approach of Wang et al. (1995).

From equation [6.49], the inverse of $\mathbf{Q}$ is

$$\mathbf{Q}^{-1} = (\mathbf{L}')^{-1}\mathbf{V}^{-1}\mathbf{L}^{-1} \qquad [6.51]$$

The inverse of the lower triangular matrix ($\mathbf{L}_i^{-1}$) for individual i is

$$\mathbf{L}_i^{-1} = \begin{bmatrix} \mathbf{L}_{i-1}^{-1} & \mathbf{0} \\ -\mathbf{N}_i' & \mathbf{I}_2 \end{bmatrix} \qquad [6.52]$$

The conditional gametic relationship between Mendelian sampling within an individual can be decomposed as follows:

$$\mathbf{V}_{i,i} = \mathbf{T}_i \mathbf{T}_i' \qquad [6.53]$$

where $\mathbf{T}_i$ is a 2×2 lower triangular matrix.

The inverse of $\mathbf{T}_i$ can be obtained from the elements of $\mathbf{V}_{i,i}$ as follows:

Let the elements of $\mathbf{V}_{i,i}$ be denoted as

$$\mathbf{V}_{i,i} = \begin{bmatrix} p & k \\ k & q \end{bmatrix} \text{ and } c = \sqrt{q - (k^2 / p)}$$

Then

$$\mathbf{T}_i^{-1} = \begin{bmatrix} \frac{1}{\sqrt{p}} & 0 \\ \frac{-k}{pc} & \frac{1}{c} \end{bmatrix} \qquad [6.54]$$

Equation [6.54] shows that the inverse of $\mathbf{T}_i$ can be obtained from the elements of $\mathbf{V}_{i,i}$ without decomposing $\mathbf{V}_{i,i}$ and inverting $\mathbf{T}_i$. When all animals are considered, the decomposition in equation [6.53] can be written as

$$\mathbf{V} = \mathbf{T}\mathbf{T}' \qquad [6.55]$$

Substituting the inverse of $\mathbf{V}$ into [6.51] gives

$$\mathbf{Q}^{-1} = (\mathbf{L}')^{-1}(\mathbf{T}')^{-1}\mathbf{T}^{-1}\mathbf{L}^{-1} \qquad [6.56]$$

Since $\mathbf{L}_i^{-1}$ and $\mathbf{T}_i^{-1}$ can be computed recursively from equations [6.52 and 6.54], the contribution of each individual to $\mathbf{Q}^{-1}$ can be computed separately. The contribution matrix for individual i is given by

$$\mathbf{W}_i = \boldsymbol{\Omega}'\boldsymbol{\Omega} = \begin{bmatrix} (-\mathbf{T}_i')^{-1}\mathbf{B}_i' \\ (-\mathbf{T}_i')^{-1}\mathbf{I}_2 \end{bmatrix} \begin{bmatrix} -\mathbf{T}_i^{-1}\mathbf{B}_i & \mathbf{T}_i^{-1}\mathbf{I}_2 \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{B}'\mathbf{T}_i'^{-1}\mathbf{T}_i^{-1}\mathbf{B}_i & -\mathbf{B}'\mathbf{T}_i'^{-1}\mathbf{T}_i^{-1} \\ -\mathbf{T}_i'^{-1}\mathbf{T}_i^{-1}\mathbf{B}_i & \mathbf{T}_i'^{-1}\mathbf{T}_i^{-1} \end{bmatrix} \qquad [6.57]$$

where $\boldsymbol{\Omega}$ is a 2×6 matrix and $\mathbf{B}_i$ is as defined before. Given that $\mathbf{V}_{i,i}^{-1} = \mathbf{T}_i'^{-1}\mathbf{T}_i^{-1}$ equation [6.57] can be written:

$$\mathbf{W}_i = \begin{bmatrix} \mathbf{B}'\mathbf{V}_{i,i}^{-1}\mathbf{B}_i & -\mathbf{B}'\mathbf{V}_{i,i}^{-1} \\ -\mathbf{V}_{i,i}^{-1}\mathbf{B}_i & \mathbf{V}_{i,i}^{-1} \end{bmatrix} \qquad [6.58]$$

Equation [6.58] is equivalent to equation [6.43] derived by Wang et al. (1995).

The positions of the elements affected by individual i in the $\mathbf{Q}^{-1}$ matrix are obtained using equation [6.44].

An algorithm to compute $\mathbf{W}_i$ and add contribution to $\mathbf{Q}^{-1}$ is as follows:

1. Initialize $\boldsymbol{\Omega}$ to $\mathbf{0}$.
2. Create a 1×6 vector, say $\boldsymbol{\psi}$, with elements 1 to 6 equal to $2s-1$, $2s$, $2d-1$, $2d$, $2i-1$, and $2i$.
3. Compute $\mathbf{V}_{i,i}$ and assign $\mathbf{T}_i^{-1}(1,1)$ to $\boldsymbol{\Omega}(1,5)$, $\mathbf{T}_i^{-1}(1,2)$ to $\boldsymbol{\Omega}(1,6)$, $\mathbf{T}_i^{-1}(2,1)$ to $\boldsymbol{\Omega}(2,5)$, and $\mathbf{T}_i^{-1}(2,2)$ to $\boldsymbol{\Omega}(2,6)$.
4. Assign $-\mathbf{B}_i(1,)/\sqrt{p}$ to elements 1 to 4 of $\boldsymbol{\Omega}(1,)$ where $\boldsymbol{\Omega}(1,)$ refers to the first row of $\boldsymbol{\Omega}$.
5. Assign $(k\mathbf{B}(1,)/pc-\mathbf{B}(2,)/c)$ to elements 1 to 4 of $\boldsymbol{\Omega}(2,)$.
6. For $x=1$ to 6

   For $y=1$ to 6

   Add $(\boldsymbol{\Omega}(1,x)\boldsymbol{\Omega}(1,y)+\boldsymbol{\Omega}(2,x)\boldsymbol{\Omega}(2,y))$ to $\mathbf{Q}^{-1}(\boldsymbol{\psi}(x),\boldsymbol{\psi}(y))$.

The contribution of individual 7 to $\mathbf{Q}^{-1}$ is used as an example. The matrix $\mathbf{V}_{7,7}$ was computed earlier and is given by

$$\mathbf{V}_{7,7} = \begin{bmatrix} 0.500 & 0.001 \\ 0.001 & 0.171 \end{bmatrix}$$

From equation [6.46],

$$\mathbf{T}_{7,7}^{-1} = \begin{bmatrix} 1.4140 & 0.0000 \\ -0.0004 & 2.4183 \end{bmatrix}$$

The QTL transmission probabilities matrix ($\mathbf{B}_7$) was also computed earlier

$$\mathbf{B}_7 = \begin{bmatrix} 0.500 & 0.500 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.100 & 0.900 \end{bmatrix}$$

The contribution matrix is given by

$$\mathbf{W}_7 = \begin{bmatrix} 0.500 & 0.500 & 0.000 & -0.003 & -1.000 & 0.003 \\ 0.500 & 0.500 & 0.000 & -0.003 & -1.000 & 0.003 \\ 0.000 & 0.000 & 0.580 & 0.526 & 0.001 & -0.585 \\ -0.003 & -0.003 & 0.526 & 4.737 & 0.005 & -5.263 \\ -1.000 & -1.000 & 0.001 & 0.005 & 2.000 & -0.006 \\ 0.003 & 0.003 & -0.585 & -5.263 & -0.006 & 5.848 \end{bmatrix}$$

and the contribution of $\mathbf{W}_7$ are in the following position of $\mathbf{Q}^{-1}$

$$\mathbf{\Pi}_7 = \begin{bmatrix} (9,9) & (9,10) & (9,11) & (9,12) & (9,13) & (9,14) \\ (10,9) & (10,10) & (10,11) & (10,12) & (10,13) & (10,14) \\ (11,9) & (11,10) & (11,11) & (11,12) & (11,13) & (11,14) \\ (12,9) & (12,10) & (12,11) & (12,12) & (12,13) & (12,14) \\ (13,9) & (13,10) & (13,11) & (13,12) & (13,13) & (13,14) \\ (14,9) & (14,10) & (14,11) & (14,12) & (14,13) & (14,14) \end{bmatrix}$$

The matrices $\mathbf{W}_7$ and $\mathbf{\Pi}_7$ are the same as those computed using Wang approach in the previous section. The $\mathbf{Q}^{-1}$ obtained using the algorithm above was also the same as that of Wang et al. (1995).

6.9. Discussion

Different approaches to calculate genetic relationships were considered. The main difference amongst these approaches is the information upon which the relationships are computed. In an ideal scenario where information on QTL is available, genetic relationships at the QTL can be obtained easily using the identity-by-state formula given earlier (equations 6.17 and 6.18). In practice information on QTL is often unavailable and therefore genetic relationships have to be inferred from other sources of information such as pedigree, marker information and knowledge about the recombination rate between the marker and QTL. The pedigree-based relationships are more useful for genetic evaluation if the trait under consideration is controlled by infinite number of genes. When a finite number of genes influence a trait, pedigree-based genetic relationships may not be useful because the observed genetic relationships deviate considerably from expectation.

Pedigree and marker-based provide a better estimate of the genetic relationships when the markers are in close proximity with the QTL and are informative. However, when markers are located far from the QTL or not informative, pedigree-based relationships should be preferred to pedigree- and marker-based relationships.

It has been shown that conditional gametic relationships can be computed based on marker information and knowledge of the recombination rate between the marker and QTL. Four different procedures to compute the GRM were presented. The procedures can be classified into two groups based on the assumption of the knowledge about the origin of the parental haplotype. The approaches of Fernando and Grossman (1989) and Van Arendonk et al. (1994) belong to the same group (i.e. they assume knowledge of the parental origin of the marker alleles). On the other hand, the approaches of Wang et al. (1995) and Abdel-Azim and Freeman (2001) are in the same class (i.e. they assume no knowledge of the parental origin of the marker allele. The equivalence between the approaches has been shown. Essentially, these approaches are different matrix representations of the same thing. Their difference lies in the efficiency with which the gametic relationship matrix and its inverse can be built.

6.10. Conclusions

This chapter focused on how different information can be used to compute the genetic relationships required in mixed model equations for prediction of genetic merit. The differences between various approaches were demonstrated using a simple pedigree example. It has been shown that marker information can be easily incorporated in genetic

evaluation programs. However, the usefulness of the marker information will depend on how close they are located to the QTL. When markers are not in close proximity with the QTL, pedigree-based relationships should be preferred to pedigree- and marker-based relationships. However, when QTL information is available accuracy of genetic prediction can be enhanced since information from relatives will be appropriately weighted.

6.11. Literature Cited

Abdel-Azim, G., and A. E. Freeman. 2001. A rapid method for computing the inverse of the gametic covariance matrix between relatives for a marked quantitative trait locus. Genet. Sel. Evol. 33:153-173.

Ashwell, M. S., D. W. Heyen, T. S. Sonstegard, C. P. Van Tassel, Y. Da, P. M. VanRaden, M. Ron, J. I. Weller, and H. A. Lewin. 2004. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. J. Dairy Sci. 87:468-475.

Bishop, M. D., S. M. Kappes, J. W. Keele, R. T. Stone, S. L. F. Sunden, G. A. Hawkins, S. S. Toldo, R. Fries, M. D. Grosz, J. Yoo, and C. W. Beattie. 1994. A genetic linage map for cattle. Genetics 136:619-639.

Fernando, R. L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21:467-477.

Georges, M., D. Nielson, M. Mackinnon, A. Mishra, R. Okimoto, A. T. Pasquino, L. S. Sargeant, A. Sorensen, M. R. Steele, X. Zhao, J. E. Womack, and I. Hoeschele.

1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139:907-920.

Henderson, C. R. 1976. A simple method for computing the inverse of the numerator relationship matrix used in prediction of breeding values. Biometrics 32:69-83.

Henderson, C. R. 1984. Applications of Linear Models in Animal Breeding. University of Guelph, Ontario, Canada.

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. 7:235-249.

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

Liu, Y., G. B. Jansen, and C. Y. Lin. 2002. The covariance between relatives conditional on genetic markers. Genet. Sel. Evol. 34:657-678.

Nejati-Javaremi, A. 1995. Alternative methods for defining relationship, assigning haplotypes and measuring linkage in animal breeding. Ph.D. Dissertation. University of Guelph.

Nejati-Javaremi, A. C. Smith, and J. P. Gibson. 1997. Effect of total allelic relationship on accuracy of evaluation and response to selection. J. Anim. Sci. 75:1738-1745.

Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics 32:949-953.

Quaas, R. L. 1988. Additive genetic model with groups and relationships. J. Dairy. Sci. 71:1338-1345.

Quaas, R. L., R. D. Anderson, and A. R. Gilmour. 1984. BLUP School Handbook; Use of Mixed Models for Prediction and for Estimation of (Co)variance Components. Animal Breeding and Genetics Unit, University of New England, NSW 2351, Australia.

Spelman, R. J., and H. Bovenhuis. 1996. Genetic response from marker assisted selection in an outbred population for differing marker bracket sizes and with two identified quantitative trait loci. Genetics 148:1389-1396.

Van Arendonk, J. A. M., B. Tier, and B. P. Kinghorn. 1994. Use of multiple genetic markers in prediction of breeding values. Genetics 137:319-329.

Van Tassel, C. P., T. S. Sonstegard, and M. S. Ashwell. 2004. Mapping quantitative trait loci affecting dairy conformation to chromosome 27 in two Holstein grandsire families. J. Dairy Sci. 87:450-457.

Wang, T., R. L. Fernando, S. Van der Beek, M. Grossman, and J. A. M. Van Arendonk. 1995. Covariance between relatives for a marked quantitative trait locus. Genet. Sel. Evol. 27:251-274.

Wright, S. 1922. Coefficients of inbreeding and relationship. Am. Nat. 56:330-338.

CHAPTER VII

WEIGHTING OF INFORMATION FROM RELATIVES WHEN ESTIMATING TRANSMITTING ABILITIES USING THE STANDARD OR MARKER-BASED INVERSE OF THE NUMERATOR RELATIONSHIP MATRIX

7.1. Abstract

Best linear unbiased prediction uses information from relatives when evaluating the genetic merit of an animal by accounting for additive genetic relationships among animals. Relationships among animals are usually computed using pedigree information ($\mathbf{A}_P$). Relationships among animals can now be computed using genetic markers ($\mathbf{A}_M$). The objective of this study was to derive weights for an individual and its relatives when the $\mathbf{A}_P$ or $\mathbf{A}_M$ numerator relationship matrices were inverted and used in the mixed model equations. The data set for a trait measured on females including one grandsire with half sib sons was used in the analysis. The grandsire and his sons had several hundred daughters with yield deviations (DYD). The predicted transmitting abilities (PTA) were obtained using a sire model. The PTA for each son was expressed as a linear function of his DYD if available and PTA of his relatives. The weights for a half-sib bull using $\mathbf{A}_P$ were always zero except for his DYD and the grandsire's PTA. The relative emphasis on

the DYD increased as the number of daughters increased. The zero weight on half-sibs was a feature of the inverse of the $\mathbf{A}_P$. In any non-inbred population, the $\mathbf{A}_P$ uses a relationship of 0.25 between all half-sibs. A bull with no daughters was evaluated as half its sires' PTA. Its reliability is limited because there can be no accounting for Mendelian sampling. In contrast, the relationship between half sibs can vary from 0 to 0.5 in $\mathbf{A}_M$. That resulted in an inverse with more non-zero coefficients than existed when $\mathbf{A}_P$ was used. Accordingly, the evaluation of any particular half-sib had non-zero weights for all half-sibs with the weight on each half-sib varying with the proportions of alleles shared in common. A bull with no daughters could be more reliably assessed using $\mathbf{A}_M$ rather than $\mathbf{A}_P$ if he had half sibs with daughters because the inferiority or superiority of his Mendelian sampling could be assessed to some extent. The use of relationships based on markers may be advantageous for young bulls.

7.2. Introduction

Current genetic evaluation procedures allow utilization of all available sources of information in predicting the genetic merit of an animal (Henderson, 1975). The common sources of information in livestock include performance records on the animal, its progeny, sibs and ancestors. These different sources carry different amounts of information and thus, they need to be weighted appropriately. The weight given to each source depends primarily on the heritability of the trait and the additive genetic relationship between the animal and its relatives providing the information. For instance, when the heritability of a trait is moderate to high, more emphasis is placed on the

animal's phenotypic observation relative to other sources of information. On the other hand, when the heritability of the trait is low individual performance receives little weighting compared to progeny performance. Moreover, the weights change when the quantities of information on various sources alters. The expected breeding value of young animals without individual performance is the mean of the evaluations of their parents. As the animal obtains individual or progeny performance information, the emphasis placed on pedigree information diminishes.

The use of information from relatives in predicting the genetic merit of animals is readily achieved by incorporating the inverse of the standard numerator relationship matrix in the mixed model equations (Henderson, 1975). This has been made possible for large-scale genetic evaluation programs through discovery of efficient methods to obtain the inverse of the numerator relationship matrix directly from pedigree information without the need for the numerator relationship matrix itself (Henderson, 1976; Quaas, 1976). The mixed model equations incorporating all known pedigree relationships among animals have been used to derive weighting factors to account for the amount of information provided by relatives in the evaluation of the animal's genetic merit (VanRaden and Wiggans, 1991).

Recent developments in molecular genetics have generated interest in the use of genetic markers in the computation of relationships at specific locations on the genome or quantitative trait loci (Fernando and Grossman, 1989; van Arendonk et al., 1994; Perez-Enciso et al., 2000; Abdel-Azim, 2001; Wang, 2002). Nejati-Javaremi (1995) proposed an approach to compute additive genetic relationships based on genetic marker and pedigree information. In this approach segments between consecutive markers are

compared between individuals to establish the additive genetic relationship. However, no attempt has been made so far to determine how the information is shared among relatives in the presence of genetic marker information. The objective of this study was to derive the weight given to each source of information when the inverse of the standard or marker-based numerator relationship matrices were used in the mixed model equation.

7.3. Materials and Methods

A simple data set comprising of a single grandsire family with 9 sons and 13,648 daughters was used in this study. Each son had different number of daughters ranging from 37 to 99,999. Performance records were only available on daughters. The numbers of daughters per sire are given in Table 7.1.

Table 7.1. Pedigree and the number of daughters per sire.

Animal	Sire[a]	Dam[a]	No. Daughters
1	0	0	13,648
2	1	0	1,372
3	1	0	99,999
4	1	0	348
5	1	0	16,677
6	1	0	2,122
7	1	0	37
8	1	0	766
9	1	0	12,039
10	1	0	77,300

[a]A zero refers to unknown parent.

The response variable considered in this study was the daughter yield deviations (DYD) weighted by the square root of the number of daughters contributing to the DYD for each sire. The DYD is an average of the daughters' yields adjusted for fixed and non-genetic random effects of the daughters and genetic effects of the dam. According to VanRaden and Wiggans (1991) the DYD is an unregressed measure of the progeny performance. The DYD used in this study were those published by the Animal Improvement Programs Laboratory of the U.S. Agricultural Research Service. Data analysis proceeded by fitting a sire model to weighted daughter yield deviations as shown below:

$$\mathbf{y} = \mathbf{Z}\mathbf{s} + \mathbf{e} \qquad [7.1]$$

where **y** is a vector of weighted daughter yield deviations; **s** is a vector of half of the random additive genetic effects of the sires or transmitting abilities (TA); **Z** is an incidence matrix (whose nonzero elements are the square root of the number of daughters for each sire) relating the TA of the sire to the observations; **e** is a vector of residuals assumed to be independent and identically distributed. The expectations of the observations and the random variables were mutually zero. The variance-covariance structure of the random effects was assumed to be:

$$var\begin{bmatrix}\mathbf{s}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{A}\sigma_a^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\sigma_e^2\end{bmatrix}$$

where **A** is the numerator relationship matrix, **I** is an identity matrix, σ_a^2 and σ_e^2 are the additive genetic and residual variances, respectively.

The system of equations from fitting equation [7.1] to the data is

$$\left[\mathbf{Z'Z} + \lambda \mathbf{A}^{-1}\right][\hat{\mathbf{s}}] = [\mathbf{Z'y}] \qquad [7.2]$$

where $\lambda = \frac{4\text{-}h^2}{h^2}$, h^2 is the heritability of the trait assumed to be 0.30, **Z' Z** is a diagonal matrix whose elements are the number of daughters contributing to the sire's DYD, **Z'y** is a vector whose elements are the product of the number of daughters and the DYD (i.e. the sum of progeny deviation). These equations are identical to those that would have been formed if a sire model was fitted to individual observations on all progeny.

The weight given to different sources of information to obtain predictions of TA (PTA) for each sire were obtained by rearranging the equation corresponding to the sire to be evaluated in equation [7.2]. That is, the sire's PTA was expressed as a function of its DYD and the PTA of its sons. Two sets of weights were calculated for the sire of interest. The first set was obtained by solving [7.2] incorporating the inverse of the standard pedigree-based numerator relationship matrix ($\mathbf{A}_P^{-1}$). The second set was calculated by using the inverse of the marker-based numerator relationship matrix ($\mathbf{A}_M^{-1}$).

The marker-based relationships were computed at the segments within flanking markers following the procedure proposed by Nejati-Javaremi (1995). A marker segment

was defined as an interval between two flanking markers thought to be containing the QTL as shown in Figure 7.1.

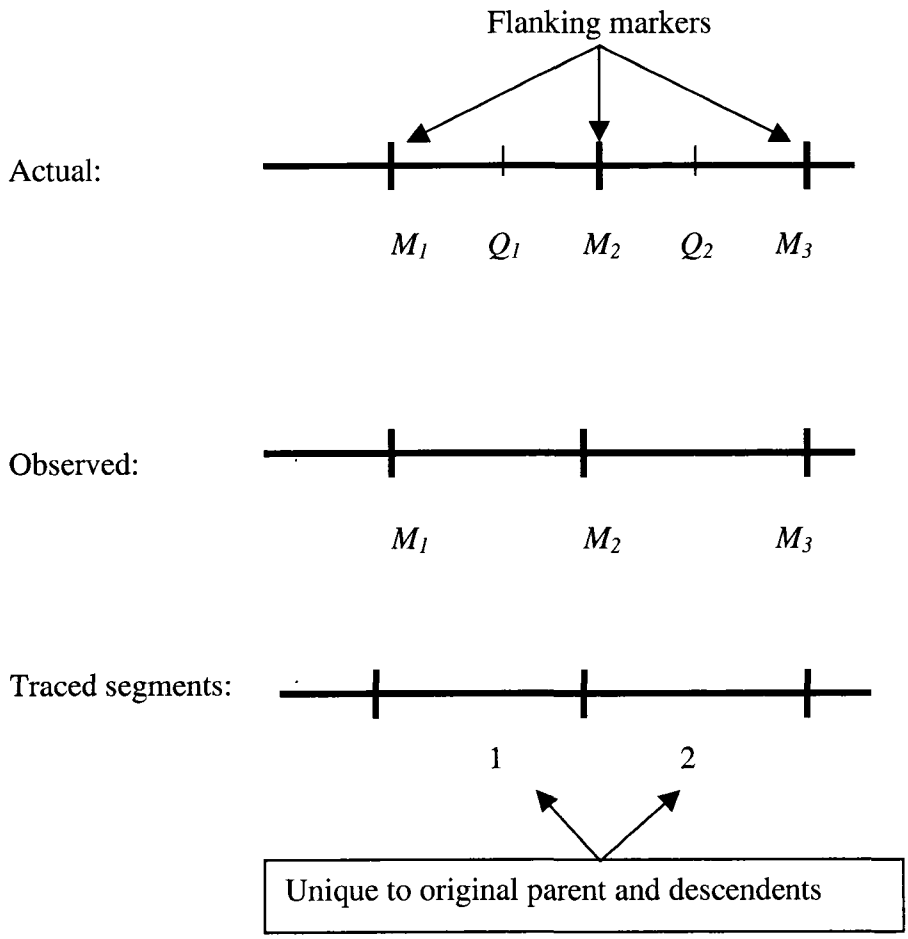


Figure 7.1. An illustration of the marker segments upon which marker relationships are based (from Nejati-Javaremi, 1995).

The concept of marker-based relationship was developed as follows (Nejati-Javaremi, 1995). Each segment is assumed to be unique in the base population because information about the QTL flanked by the markers is not available implying that base

individuals are unrelated and non-inbred with respect to these segments. This is a similar assumption to that made in forming pedigree-based relationships. The total number of unique segments within each pair of flanking markers is therefore twice the number of individuals in the base population. Parental origin of each segment is assigned based on the marker genotypes of parents and their descendents. This in turn allows tracing of each segment to its origin. Segments that did not undergo recombination can be traced to their origin in the base population. On the other hand, the origin of segments that underwent recombination is the generation at which recombination occurred. Using the knowledge of the origin of each segment, the relationship between two individuals can be expressed as the proportion of the segments in common or identical-by-descent. Given that each individual has two segments, there are four possible ways that two individuals could be identical-by-descent (IBD). Therefore, the relationship between two individuals, *a* and *b*, can be obtained from:

$$\mathbf{A}_{\mathrm{M}(a,b)} = 2\times\frac{\sum_{i=1}^{2}\sum_{j=1}^{2}\mathrm{LS}_{ij}}{4} \qquad [7.3]$$

where LS_{ij} is the proportion of the i^{th} linkage segment from the first individual, *a*, in common with the j^{th} linkage segment from individual, *b*. The LS_{ij} range from 0 to 1. Therefore, the relationships range from 0 to 2 similar to numerator relationships.

Genotypic information was available on fifty-two genetic marker locations distributed over six bovine chromosomes (BTA1, BTA3, BTA9, BTA10, BTA14, and

BTA20). The genotypic information was used to infer the linkage segments (LS) for the ten sires and marker-based relationships were computed using equation [7.3].

7.4. Results and discussion

Pedigree-based numerator relationship matrix. The $\mathbf{A}_P$ computed using the pedigree in Table 1 is presented below. The elements of $\mathbf{A}_P$ are the usual expected additive genetic relationships among relatives. For instance, all half sibs are assumed to be 0.25 related whereas in reality it is known that this value could vary form 0 to ½. This uncertainty in the knowledge of the sample half that each half sib inherited from the sire has an impact on the weighting of information which will become more obvious later.

$$\mathbf{A}_P = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.5 & 0.25 & 1 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 & 1 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 & 0.25 & 1 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 & 0.25 & 0.25 & 1 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 1 & 0.25 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 1 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 1 & 0.25 \\ 0.5 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & 1 \end{bmatrix}$$

Below is the inverse of $\mathbf{A}_P$,

$$\mathbf{A}_P^{-1} = \begin{bmatrix} 4.00 & -0.67 & -0.67 & -0.67 & -0.67 & -0.67 & -0.67 & -0.67 & -0.67 & -0.67 \\ -0.67 & 1.33 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.67 & 0 & 1.33 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.67 & 0 & 0 & 1.33 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.67 & 0 & 0 & 0 & 1.33 & 0 & 0 & 0 & 0 & 0 \\ -0.67 & 0 & 0 & 0 & 0 & 1.33 & 0 & 0 & 0 & 0 \\ -0.67 & 0 & 0 & 0 & 0 & 0 & 1.33 & 0 & 0 & 0 \\ -0.67 & 0 & 0 & 0 & 0 & 0 & 0 & 1.33 & 0 & 0 \\ -0.67 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.33 & 0 \\ -0.67 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1.33 \end{bmatrix}$$

The $\mathbf{Z'Z}$ is

$$\mathbf{Z'Z} = \begin{bmatrix} 13648 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1372 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 99999 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 348 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 16677 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2122 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 37 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 766 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 12039 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 77300 \end{bmatrix}$$

After augmenting the $\mathbf{Z'Z}$ with $\lambda\mathbf{A}_P^{-1}$ (λ=12.33) the left-hand and right-hand side of the model equations are:

$$\begin{bmatrix} 13697.33 & -8.22 & -8.22 & -8.22 & -8.22 & -8.22 & -8.22 & -8.22 & -8.22 & -8.22 \\ -8.22 & 1388.44 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -8.22 & 0 & 100015.44 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -8.22 & 0 & 0 & 364.44 & 0 & 0 & 0 & 0 & 0 & 0 \\ -8.22 & 0 & 0 & 0 & 16690.44 & 0 & 0 & 0 & 0 & 0 \\ -8.22 & 0 & 0 & 0 & 0 & 2138.44 & 0 & 0 & 0 & 0 \\ -8.22 & 0 & 0 & 0 & 0 & 0 & 53.44 & 0 & 0 & 0 \\ -8.22 & 0 & 0 & 0 & 0 & 0 & 0 & 782.44 & 0 & 0 \\ -8.22 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 12325.44 & 0 \\ -8.22 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 77316.44 \end{bmatrix} \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{s}_4 \\ \hat{s}_5 \\ \hat{s}_6 \\ \hat{s}_7 \\ \hat{s}_8 \\ \hat{s}_9 \\ \hat{s}_{10} \end{bmatrix} = \begin{bmatrix} 13648 \times DYD_1 \\ 1372 \times DYD_2 \\ 99999 \times DYD_3 \\ 348 \times DYD_4 \\ 16677 \times DYD_5 \\ 2122 \times DYD_6 \\ 37 \times DYD_7 \\ 766 \times DYD_8 \\ 12039 \times DYD_9 \\ 77300 \times DYD_{10} \end{bmatrix}$$

The weighting factors derived using the pedigree-based numerator relationship matrix are given in Table 7.2. For illustrative purpose, derivation of weights is presented for the grandsire (sire 1) and a representative son (sire 2). Consider the grandsire. The transmitting ability of the grandsire can be obtained by rearranging equation 1 of the system of equations above as follows:

$$\hat{s}_1 = b_1 \times DYD_1 + \sum_{j=2}^{10} \left[b_j \times 2\hat{s}_j \right] \qquad [7.4]$$

where $b_1 = \left[\frac{n_1}{n_1 + \sum_{j=2}^{10} q_j} \right]$, $b_j = \left[\frac{q_j}{n_1 + \sum_{j=2}^{10} q_j} \right]$, n_1 is the number of daughters contributing to the grandsire's DYD, $q_j = -0.5 \times \mathbf{C}_{1,j}$ and $\mathbf{C}_{1,j}$ is the 1,jth off-diagonal element of the coefficient matrix.

The PTA of the grandsire can be obtained from equation [4] as follows:

$$\hat{s}_1 = \left[\frac{13648}{13685}\right] \times DYD_1 + \sum_{j=2}^{10}\left[\frac{4.11}{13685}\right] \times 2\hat{s}_j$$

$$= 0.9973 \times DYD_1 + \sum_{j=2}^{10}[0.0003] \times 2\hat{s}_j$$

Table 7.2. The weighting factors for the grandsire (sire 1) and his son (sire 2) derived from the model equations including the pedigree-based relationships ($\mathbf{A}_P$).

Sire	DYD	Weighting Factors									
		b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
1	ALL[a]	0.9973	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
2	ALL	0.0118	0.9882								
2	NO[b]	1.0000	0.0000								

[a] DYD available on all sires, [b] DYD not available on sire 2 (as would occur prior to his progeny test).

The weighting factors for sire 2 were obtained by rearranging the second equation of the model equations. When the DYD information on sire 2 was available, the weighting factors were as follows:

$$\hat{s}_2 = \left[\frac{q_1}{d}\right] \times 0.5\hat{s}_1 + \sum_{j=3}^{10}\left[\frac{q_j}{d}\right]\hat{s}_j + \left[\frac{n_2}{d}\right] \times DYD_2 \qquad [7.5]$$

where $q_1 = -2 \times \mathbf{C}_{2,1}$, $q_j = -\mathbf{C}_{2,j}$, n_2 is the total number of daughters for sire 2 contributing to its DYD, $d = n_2 + q_1 + \sum_{j=3}^{10} q_j$.

In this case q_j is zero and therefore PTA for sire 2 simplifies to

$$= \left[\frac{16.44}{1388.44}\right] \times 0.5\hat{s}_1 + \left[\frac{1372}{1388.44}\right] \times DYD_2$$

$$= 0.0118(0.5\hat{s}_1) + 0.9882 \times DYD_2$$

All the ten weighting factors for the grandsire were nonzero (Table 7.2). This indicates that daughters' performance on all the ten sires contributed in the evaluation of sire 1 or the grandsire. However, the daughters of sire 1 contributed predominantly (99.7%) to its evaluation while the contribution from the daughters of its sons (sire 2 through 10) was minimal. This result was expected since there were far more daughters than sons and the fact that the trait under consideration was measured only in females. All sons' evaluations contributed equally to their sires' evaluation.

There were only two weighting factors for sire 2's evaluation (Table 7.2). The weighting factors were for the PTA of the grandsire and sire 2's daughter performance. This implies that only information on the bull's father and its daughters contributed to its evaluation. However, it was observed earlier that the grandsire's PTA is a linear function of its daughters' performance and the PTA of sires 2 through 10. Therefore, the contribution of the half sibs to each other's evaluation was indirect through their common sire or grandsire. When sire 2 has no daughters, its PTA is half the PTA of its sire.

Marker-based numerator relationship matrix. The numerator relationship matrix computed using pedigree and genetic marker information is presented below.

$$\mathbf{A}_M = \begin{bmatrix} 1 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 & 0.5 \\ 0.5 & 1 & 0.37 & 0.35 & 0.14 & 0.30 & 0.17 & 0.26 & 0.19 & 0.26 \\ 0.5 & 0.37 & 1 & 0.35 & 0.14 & 0.28 & 0.29 & 0.28 & 0.25 & 0.19 \\ 0.5 & 0.35 & 0.35 & 1 & 0.16 & 0.30 & 0.23 & 0.36 & 0.25 & 0.23 \\ 0.5 & 0.14 & 0.14 & 0.16 & 1 & 0.23 & 0.24 & 0.25 & 0.32 & 0.30 \\ 0.5 & 0.30 & 0.28 & 0.30 & 0.23 & 1 & 0.07 & 0.35 & 0.16 & 0.28 \\ 0.5 & 0.17 & 0.29 & 0.23 & 0.24 & 0.07 & 1 & 0.16 & 0.37 & 0.17 \\ 0.5 & 0.26 & 0.28 & 0.36 & 0.25 & 0.35 & 0.16 & 1 & 0.20 & 0.30 \\ 0.5 & 0.19 & 0.25 & 0.25 & 0.32 & 0.16 & 0.37 & 0.20 & 1 & 0.25 \\ 0.5 & 0.26 & 0.19 & 0.23 & 0.30 & 0.28 & 0.17 & 0.30 & 0.25 & 1 \end{bmatrix}$$

Contrary to the elements of $\mathbf{A}_P$ which represented expected relationships among relatives, elements of $\mathbf{A}_M$ are the realized additive genetic relationship among relatives. The relationships among the nine half sibs ranged from 0.07 to 0.37.

The inverse of $\mathbf{A}_M$ is

$$\mathbf{A}_M^{-1} = \begin{bmatrix} 4.14 & -0.73 & -0.55 & -0.51 & -0.85 & -0.79 & -0.97 & -0.57 & -0.62 & -0.70 \\ -0.73 & 1.43 & -0.20 & -0.15 & 0.16 & -0.04 & 0.14 & 0.04 & 0.08 & -0.02 \\ -0.55 & -0.20 & 1.42 & -0.13 & 0.16 & -0.06 & -0.10 & -0.05 & -0.02 & 0.09 \\ -0.51 & -0.15 & -0.13 & 1.42 & 0.13 & -0.06 & 0.01 & -0.20 & -0.04 & 0.03 \\ -0.85 & 0.16 & 0.16 & 0.13 & 1.41 & 0.01 & 0.04 & -0.03 & -0.13 & -0.07 \\ -0.79 & -0.04 & -0.06 & -0.06 & 0.01 & 1.45 & 0.31 & -0.14 & 0.11 & -0.02 \\ -0.97 & 0.14 & -0.10 & 0.01 & 0.04 & 0.31 & 1.49 & 0.12 & -0.19 & 0.13 \\ -0.57 & 0.04 & -0.05 & -0.20 & -0.03 & -0.14 & 0.12 & 1.41 & 0.06 & -0.08 \\ -0.62 & 0.08 & -0.02 & -0.04 & -0.13 & 0.11 & -0.19 & 0.06 & 1.40 & -0.02 \\ -0.70 & -0.02 & 0.09 & 0.03 & -0.07 & -0.02 & 0.13 & -0.08 & -0.02 & 1.37 \end{bmatrix}$$

The left-hand side of the model equations is as follows:

$$\begin{bmatrix} 13657.67 & -1.69 & -1.30 & -1.18 & -1.97 & -1.84 & -2.27 & -1.32 & -1.45 & -1.64 \\ -1.69 & 1375.34 & -0.46 & -0.34 & 0.37 & -0.08 & 0.34 & 0.09 & 0.18 & -0.04 \\ -1.30 & -0.46 & 100002.32 & -0.31 & 0.37 & -0.13 & -0.24 & -0.11 & -0.06 & 0.22 \\ -1.18 & -0.34 & -0.31 & 351.31 & 0.30 & -0.13 & 0.01 & -0.46 & -0.10 & 0.08 \\ -1.97 & 0.37 & 0.37 & 0.30 & 16677.30 & 0.03 & 0.10 & -0.07 & -0.30 & -0.17 \\ -1.84 & -0.08 & -0.13 & -0.13 & 0.03 & 2125.38 & 0.72 & -0.32 & 0.26 & -0.05 \\ -2.27 & 0.34 & -0.24 & 0.01 & 0.10 & 0.72 & 40.47 & 0.27 & -0.43 & 0.30 \\ -1.32 & 0.09 & -0.11 & -0.46 & -0.07 & -0.32 & 0.28 & 769.28 & 0.15 & -0.19 \\ -1.45 & 0.17 & -0.06 & -0.10 & -0.30 & 0.26 & -0.43 & 0.15 & 12312.26 & -0.05 \\ -1.64 & -0.04 & 0.22 & 0.08 & -0.17 & -0.05 & 0.29 & -0.19 & -0.05 & 77303.2 \end{bmatrix} \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{s}_4 \\ \hat{s}_5 \\ \hat{s}_6 \\ \hat{s}_7 \\ \hat{s}_8 \\ \hat{s}_9 \\ \hat{s}_{10} \end{bmatrix}$$

The weighting factors derived from using the marker-based relationship matrix are given in Table 7.3. The weighting factors for the different sources of information for the grandsire and sire 2 were calculated using equations [7.4] and [7.5], respectively. The notable difference between the current weights for the grandsire and those obtained using $\mathbf{A}_P$ was that the information from sires 2 to 10 were no longer weighted equally. Sire 7 contributed more than the other sires. This was a surprising result because all of the nine sires were progeny of sire 1 hence equally related to him as shown in the first column of $\mathbf{A}_M$. However, sires 2 through 10 were not equally related to each other.

In contrast to only two weights obtained in the evaluation of sire 2 using the pedigree-based relationships, all the ten sires contributed in the evaluation of sire 2 when using the marker-based relationships. This implies that each of the half sibs contribute to sire 2's evaluation indirectly through their contribution to sire 1's PTA and directly through their PTA. This latter contribution is associated with the covariance between Mendelian sampling effects.

Table 7.3. The weighting factors for the grandsire (sire 1) and sire 2 derived from the model equations including the marker-based relationships ($\mathbf{A}_P$).

Sire	DYD	Weighting Factors									
		b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}
1	ALL[1]	0.99946	0.00006	0.00005	0.00004	0.00007	0.00007	0.00008	0.00005	0.00005	0.00006
2	ALL	0.00062	0.99942	0.00033	0.00025	-0.0003	0.00006	-0.0002	-0.0001	-0.00013	0.00003
2	NO[2]	1.07197	0.0000	0.57543	0.42877	-0.4678	0.10177	-0.4238	-0.1120	-0.22101	0.04685

[1] DYD available on all sires, [2] DYD not available on sire 2.

To ascertain the potential source of unequal weighting of the information from sires 2 to 10 in the evaluation of sire 1, the average relationships of sires 2 through 10 to the remaining eight sires were calculated (Figure 7.2). For example, the average relationship of sire 2 to sires 3 to 10 was calculated as follows:

$$\overline{r}_2 = \frac{\sum_{i=3}^{10} a_{i,2}}{n-1}$$

where $a_{i,2}$ is the relationship between sire i and sire 2, n is the number of half sibs. Inspection of these averages revealed that sires 5 and 7 were the least related to the group (Figure 7.2). Based on these observation it is apparent that the weights assigned to each of the nine half sibs were based on their average relationship to their siblings.

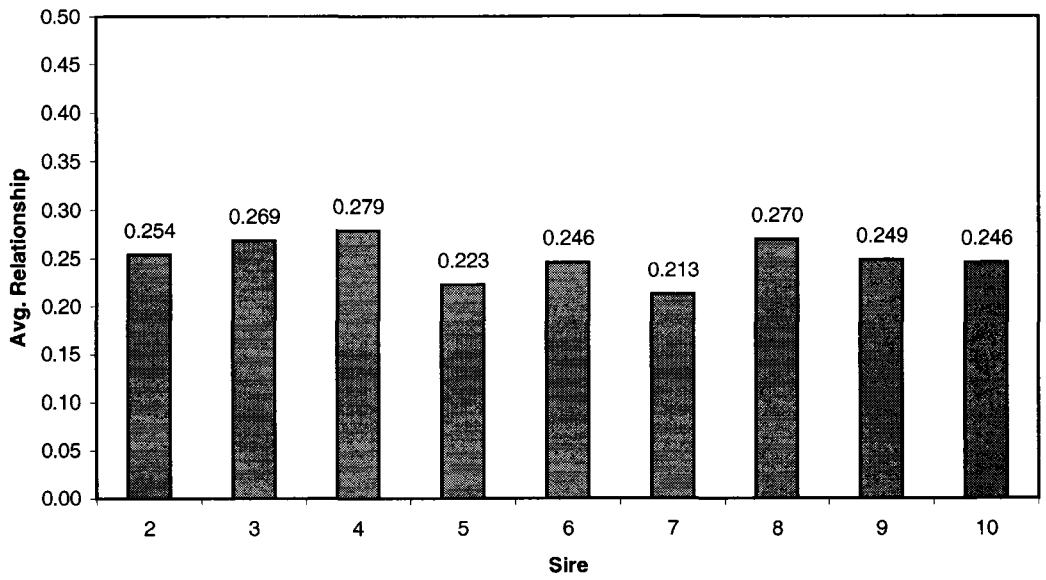


Figure 7.2. Average additive genetic relationship of sires 2 through 10 to the remaining 8 sires.

7.5. Conclusions

Results from this study showed that information from relatives is weighed differently depending on the nature of the numerator relationship matrix. The marker-based relationship matrix allows for appropriate weighting of information from relatives. A bull with no daughters could be more reliably assessed using $\mathbf{A}_M$ rather than $\mathbf{A}_P$ if he had half sibs with daughters because the inferiority or superiority of his Mendelian sampling could be partially to some extent. The use of relationships based on markers may be advantageous for young bulls without their own or progeny performance particularly for sex-limited traits.

7.6. Literature Cited

Abdel-Azim, G., and A. E. Freeman. 2001. A rapid method for computing the inverse of the gametic covariance matrix between relatives for a marked quantitative trait locus. Genet. Sel. Evol. 33:153-173.

Fernando, R. L., and M. Grossman. 1989. Marker assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21:467-477.

Henderson, C. R. 1975. Use of relationship among sires to increase accuracy of sire evaluation. J. Dairy Sci. 58:1731-1738.

Henderson, C. R. 1976. A simple method for computing the inverse of the numerator relationship matrix used in prediction of breeding values. Biometrics 32:69-83.

Nejati-Javaremi, A. 1995. Alternative methods for defining relationship, assigning haplotypes and measuring linkage in animal breeding. Ph.D. Dissertation. University of Guelph.

Perez-Enciso, M., L. Varona, and M. F. Rothschild. 2000. Computation of identity by descent probabilities conditional on DNA markers via a Monte Carlo Markov Chain Method. Genet. Sel. Evol. 32:467-482.

Quaas, R. L. 1976. Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics 32:949-953.

Van Arendonk, J. A. M., B. Tier, and B. P. Kinghorn. 1994. Use of multiple genetic markers in prediction of breeding values. Genetics 137:319-329.

VanRaden, P. M., and G. R. Wiggans. 1991. Derivation, calculation, and use of national animal model information. J. Dairy Sci. 74:2737-2746.

Wang, J. 2002. An estimator for pairwise relatedness using molecular markers. Genetics 160:1203-1215.

CHAPTER VIII

THE ACCURACY OF GENETIC PREDICTION OBTAINED USING ADDITIVE GENETIC RELATIONSHIPS BASED ON PEDIGREE OR GENETIC MARKER INFORMATION

8.1. Abstract

The objective of the current study was to evaluate the change in accuracy of genetic prediction when the inverse of the marker-based numerator relationship matrix (A_M) was substituted for the inverse of the standard numerator relationship matrix (A_P) in the mixed model equations using empirical data. Records on 1,811 male descendents from 8 Holstein sires' families were obtained from the Animal Improvement Programs Laboratory of the USDA-ARS Beltsville Research Center. Each record comprised of daughter yield deviations for milk, fat and protein yields and genotypic information on 52 microsatellite genetic markers. The daughter yield deviations were used to reconstruct the sire breeding value model. The same heritability of 0.3 was assumed for all the traits. The mixed model equations were assembled and solved using the Animal Breeder's Tool Kit. Three sets of analysis were conducted to obtain breeding values. The first set of breeding values (EBV-ALL) was obtained using all sources of information (e.g. own DYD and those of all relatives) incorporating the inverse of A_P in the mixed model equations. The

second set of breeding values (EBV-PED) were computed as in the first analysis except that the sire's own DYD (but not those of its relatives) was excluded when predicting its breeding value. The third set of breeding values (EBV-MRK) were computed as in the second analysis except that the inverse of A_M was substituted for the inverse of A_P. Pearson and Spearman rank order correlations were computed between EBV-ALL and EBV-PED or EBV-MRK to evaluate the change in accuracy and ranking of sires when A_P was replaced by A_M. Considering all sires without sons in the data set, the accuracy of prediction increased by 4.3% for milk yield but did not change for fat and protein yields when A_P was replaced by A_M. Considering A_M computed within chromosomes, the use of marker information resulted in at least an improvement in accuracy of prediction for milk and protein yield and sometimes a decline in accuracy for fat yield. These results suggest that different A_M may be required for different traits. The rank order correlations were consistently higher across traits when A_M was used suggesting that use of markers provide better ranking of sires compared to use pedigree information only. Results from the current study suggest that marker information may be used successfully to enhance the accuracy of genetic prediction in routine genetic evaluation particularly for young animals without own performance information.

8.2. Introduction

The animal model best linear unbiased prediction (BLUP) is the methodology of choice for prediction of genetic merit in most livestock species. The BLUP procedure accounts for the covariance among relatives which determines the extent of "sharing" of

information among relatives. For animals without own or progeny performance (e.g. young animals) information from relatives plays a crucial role in the genetic prediction of the animal. In fact, the genetic merit of young animals without own performance is the average genetic merit of the parents. In this case, the accuracy of genetic prediction is limited since no information is available to estimate the Mendelian sampling effect of the animal of interest. Therefore, for large half-sib families selection among progeny is effectively family selection.

For the past fifteen to twenty years, discovery of genetic markers has been increasing at an exponential rate resulting in linkage maps with high marker density (e.g. bovine linkage map; Kappes et al., 1997). These genetic markers may be useful for calculating additive genetic relationships among animals. A theoretical study using simulation by Nejati-Javaremi (1995) demonstrated that accuracy of genetic prediction could be enhanced through use of genetic relationships calculated from pedigree and marker information. In that study, increase in accuracy from replacing standard relationships with marker-based relationships ranged from 0 to 74%. The increase in accuracy was more dramatic under low heritability ($h^2 = 0.1$) compared to moderate heritability ($h^2 = 0.3$). Nejati-Javaremi (1995) assumed a scenario where there was complete genome marker coverage with a QTL within each marker interval. In practice, some intervals are likely not to have any QTL and therefore lead to reduction in accuracy of prediction. Each QTL contributed equally to the variability of the trait. Some of these assumptions may not hold under practical conditions. Therefore, it is necessary that the utility of using genetic markers to calculate genetic relationships be investigated using empirical data.

The objective of this study was to assess the change in accuracy when the inverse of the standard numerator relationship matrix was replaced by the inverse of genetic marker-based relationship matrix in the mixed model equations. A new method to calculate additive genetic relationships using pedigree and marker information is also presented.

8.3. Materials and Methods

8.3.1. Data description

Animals and pedigree information. A total of eight registered Holstein sire families were considered. Records on 1,811 male descendents from these families were obtained from the Animal Improvement Programs Laboratory (AIPL) of the USDA-ARS Beltsville Research Center. Each record consisted of the animal's registration number assigned by the Holstein Association USA and genotypic information based on microsatellite genetic markers. Ancestry information, performance and date of birth for each animal were downloaded from the official bull evaluation files of the AIPL site (ftp://aipl.arsusda.gov/pub/bulls). The pedigree structure of the families is given in Figure 8.1. All the 1,811 animals sired different number of sons and daughters. However, for convenience sires that did not have sons in the data set were referred to as cow fathers as shown in Figure 8.1. The family with the most extensive pedigree had sires in all four generations while the smallest family had descendents only in generation 1. The minimum and maximum number of sires within each sire family per generation were 1

and 6 respectively. The number of sons among the 26 sires used in generations 1 to 4 ranged from 3 to 233 with an average of 67 progeny per sire.

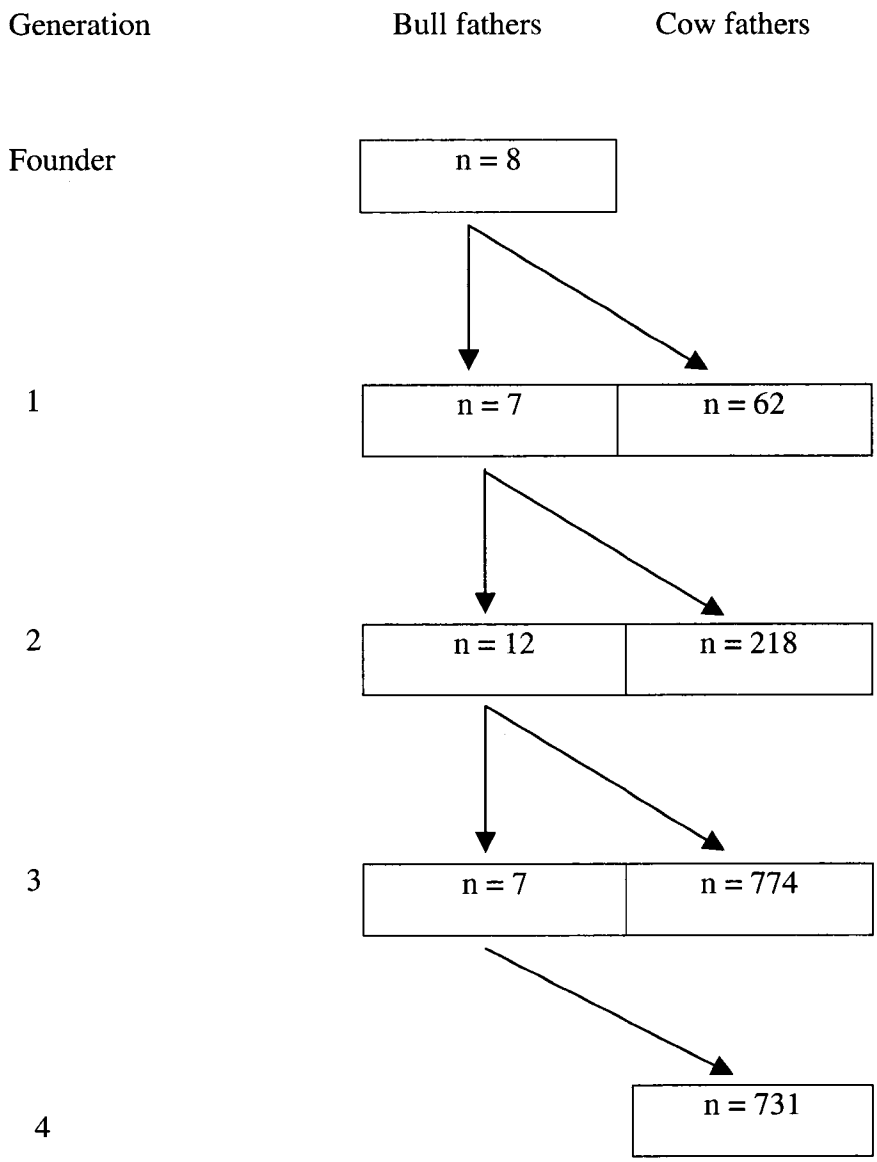


Figure 8.1. The pedigree structure of the families. Bull and cow fathers refer to progeny-tested sires that have or do not have sons in the data, respectively.

Genotypic information. Microsatellite genetic marker information was available on sires in all generations 1 to 4 i.e. 1,811 animals. Each sire was genotyped at fifty-two marker loci located on six autosomal bovine chromosomes representing about 20% of the genome. That is, the markers were in interesting QTL regions and not randomly spread around the genome. Information about the relative position of each marker locus on the linkage map was obtained from the USDA-ARS Meat Animal Research Center, Nebraska (http://sol.marc.usda.gov/). The location of each marker locus on the six chromosomes is shown in Figure 8.2. Markers were evenly distributed throughout these six chromosomes. On average, there were nine markers per chromosome. The average distance between adjacent markers across chromosomes was 13 cM. Markers closest to each other were BMS574 and BMS2321 located together on chromosome 1 while markers BL41 and BMS1266 on chromosome 3 were the furthest from each with a distance of 32.3 cM between them.

The total number of alleles for all the markers was 476. The minimum and maximum allele sizes across markers were 81 and 263 base pairs, respectively and the average allele size was 155 base pairs. Markers BMS4048 on chromosome 1 and BMS896 on chromosome 3 had the least number of alleles with 4 alleles per marker while markers BL1035 and CSSM46 on chromosome 10 had the most alleles totaling 20 alleles per marker. The average number of alleles per marker across the chromosomes was nine.

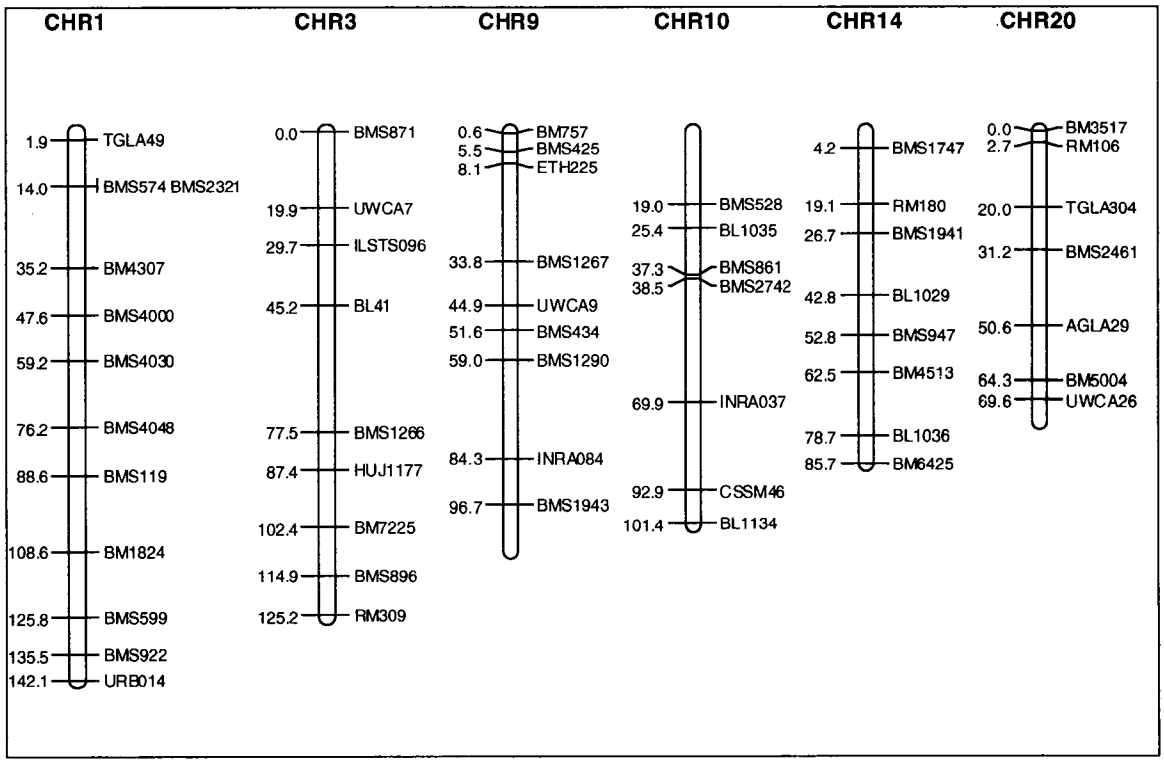


Figure 8.2. The relative position of each marker locus on the chromosome expressed in centiMorgans.

Marker information was not available on every locus for each sire. Figure 8.3 shows the number of marker loci available on each sire. The percentage of sires with marker information available on all 52 markers was about 70%. Most of the sires (97%) had no more than three markers missing. Considering sires with genotypic information on all marker loci, marker BMS1266 expressed the least heterozygosity (47%) on chromosome 3 and marker BMS528 on chromosome 10 was the most heterozygous locus (84%). In general, the marker loci considered expressed high levels of polymorphism with the average within marker heterozygosity of 71%. The marker with the minimum heterozygosity was BMS1266 and this could be explained by the fact that a single marker allele had a frequency of 70%. About 66% of the alleles had frequency greater that 1%.

The rarest allele had a frequency of 0.04%. Across marker loci, individual heterozygosity ranged from 44 to 89% with an average of 71%.

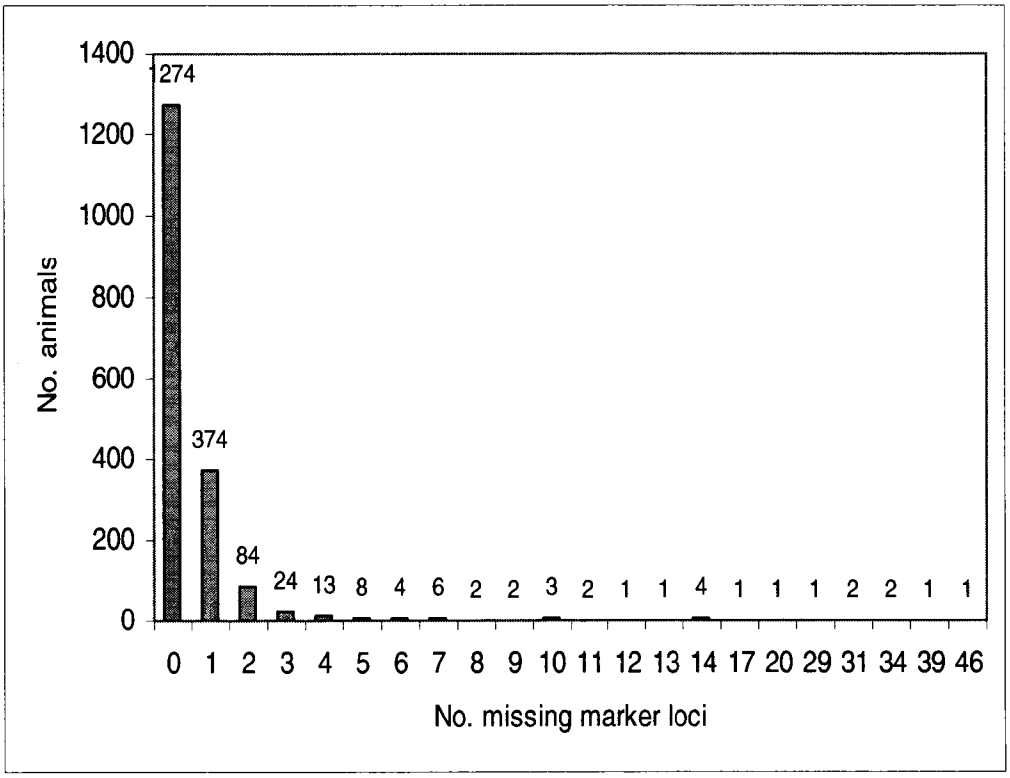


Figure 8.3. Distribution of the number of missing markers.

Performance information. Traits evaluated were milk, fat and protein yield expressed as daughter yield deviations (DYD). The DYD is an average of the daughters' performance adjusted for fixed and non-genetic random effects of the daughters and additive genetic effects of their dams. Therefore, the DYD is an unregressed measure of the performance of the sire's daughters (VanRaden and Wiggans, 1991). A total of 873 sires were used in the analysis. Table 8.1 provides the summary information for all the traits. The number of daughters per sire ranged from 11 to 99,999.

Table 8.1. Descriptive statistics for milk, fat and protein Daughter Yield Deviations.

Trait	Avg. No. Daughters per sire	Min	Max	Mean
Milk (lbs)	1,973	-3,283	2,154	40
Fat (lbs)	1,973	-82	80	8
Protein (lbs)	1,952	-72	65	6

8.3.2. Statistical Analyses

The general model used to analyze the DYD observations for each trait was a single trait sire breeding value model. The model equation can be represented in matrix notation as:

$$\mathbf{y} = \mathbf{Zs} + \mathbf{e} \qquad [8.1]$$

where **y** is a vector of milk DYD multiplied by the square root of the number of daughters for each sire, **Z** is an incidence matrix relating the breeding value of the sire to the observations in **y**, **s** is a vector of the sire's breeding value, **e** is a vector of unobservable random residual effect unique to each observation. The expectation of the sire's breeding value and the residual effects were assumed to be zero resulting in the expectation of the observations being also zero. The variance-covariance structure of the sire's breeding values and residual effects were:

$$\text{var}\begin{bmatrix}\mathbf{s}\\ \mathbf{e}\end{bmatrix}=\begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}=\begin{bmatrix}\mathbf{A}*0.25\sigma_a^2 & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\sigma_e^2\end{bmatrix}$$

where σ^2_a and σ^2_e are the additive genetic and residual variances respectively, **A** is the additive genetic relationship matrix typically called the numerator relationship matrix and **I** is an identity matrix indicating that the residual effects are independent. Given the assumed variance-covariance structure of the random effects, the variance-covariance of the observations was $\text{var}[\mathbf{y}]=\mathbf{ZGZ}'+\mathbf{R}$.

The set of equations derived from equation [8.1] can be written as follows:

$$\left[\mathbf{Z}'\mathbf{Z}+\lambda\mathbf{A}^{-1}\right][\hat{\mathbf{s}}]=[\mathbf{Z}'\mathbf{y}] \qquad [8.2]$$

where $\lambda=\frac{4\text{-}h^2}{4h^2}$ and h^2 is the heritability of the trait.

The system of equations given in equation [8.2] was assembled and solved iteratively using the Animal Breeder's Tool Kit (ABTK: Golden et al., 1992). Solutions were considered to have reached convergence when the average absolute difference between subsequent solutions was less than 10^{-12}. In addition, $\mathbf{A}^{-1}$ was obtained directly using ABTK owing to its small dimension due to a relatively small number of sires considered. The same heritability of 0.30 was assumed for all the three traits.

Three sets of solutions for the estimates of the breeding values (EBV) of the sires were obtained using equation [8.2]. The first set of EBV was computed incorporating the

inverse of the numerator relationship matrix in equation [8.2] and using all available sources of information on the sire i.e. own DYD and that of the relatives. These EBV were referred to as EBV-ALL. The second set of EBV, referred to as EBV-PED, was obtained in the same way as in the first analysis except that the individual sire's DYD was excluded from the analysis but the DYDs on all relatives were included. This analysis simulates young sires without own DYD whereby the sire's EBV is half the genetic merit of its sire. The third set of EBV was computed using the same approach as in the second analysis except that the marker-based relationship matrix replaced the pedigree-based relationship matrix. This set of EBV was referred to as EBV-MRK. The EBV-ALL prediction was expected to be more accurate compared to EBV-PED and EBV- MRK because of the inclusion of all daughters on each sire and also the fact that all available sources of information were used in predicting EBV-ALL.

The Pearson correlation coefficient between EBV-ALL with EBV-PED or EBV-MRK were used to evaluate the change in accuracy when information from relatives is used to predict the sire's genetic merit when the relationship between sires were computed using pedigree or genetic marker information. The Spearman rank order correlation coefficient was also computed to assess the effect of using pedigree or marker-based relationships in ranking sires based on EBV when own performance is not available.

8.3.3. Computation of the additive genetic relationships

Two additive genetic relationship matrices were constructed. The first relationship matrix was the usual numerator relationship matrix computed using pedigree information hereafter referred to as pedigree-based (PB) relationship matrix, denoted $\mathbf{A_P}$. The second relationship matrix was computed using the pedigree and genotypic information at marker loci hereafter referred to as the marker-based (MB) relationship matrix ($\mathbf{A_M}$).

The PB relationships were computed following the procedure presented in Henderson (1976) whereby the relationship between two individuals is half the sum of the relationship between one of the individual with the parents of the second individual and the relationship between an individual and itself is one plus its inbreeding coefficient or half the relationship of its parents. The MB relationship was calculated using the modification of the procedure proposed by Nejati-Javaremi (1995). Nejati-Javaremi (1995) presented the following formula to compute the relationship between two individuals, say *a* and *b*, conditional on pedigree and genetic marker information:

$$\mathbf{A}_{\mathrm{M_N}(a,b)} = 2\times\frac{\sum_{i=1}^{2}\sum_{j=1}^{2}\mathrm{LS}_{ij}}{4} \qquad [8.3]$$

where LS_{ij} is the proportion of the i^{th} linkage segment from the first individual, *a*, in common with the j^{th} linkage segment from individual, *b*. The LS_{ij} takes values between 0 and 1 and thus the MB relationships range from 0 to 2 similar to PB relationships. A schematic representation of an animal's haplotypes or linkage segments

is given in Figure 8.4. Essentially, the relationship computed using equation [8.3] is based on the proportion of the traced segment (Figure 8.4) in common between two individuals.

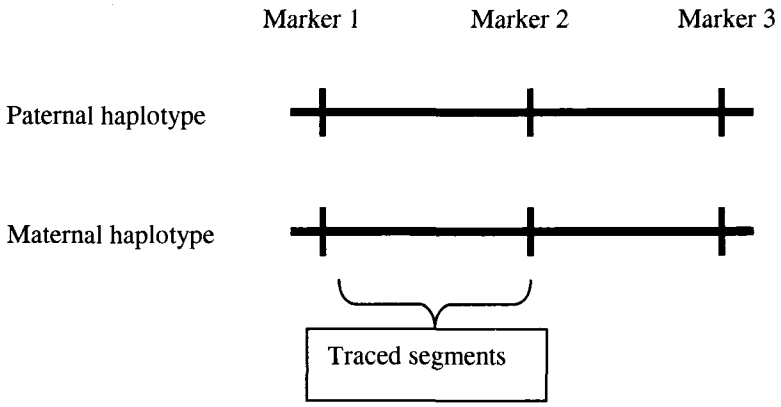


Figure 8.4. An illustration of an animal's haplotypes or linkage segments.

Equation [8.3] gives reliable estimates of the MB relationships when the entire genome has marker coverage. Often in practice informative markers may only be available at certain genomic locations and therefore genome marker coverage may be limited. In the current study, a modification of equation [8.3], which account for the relationship at locations or intervals without marker coverage is presented. That is, MB relationship computed using equation [8.3] is combined with expected or PB relationships using:

$$\mathbf{A}_{\mathrm{M_M}(a,b)} = \frac{r.\mathbf{A}_{\mathrm{P}(a,b)} + s.\mathbf{A}_{\mathrm{M_N}(a,b)}}{l} \qquad [8.4]$$

where s and r are the genomic intervals with and without marker coverage respectively, s and r sum to l, l is the total genome size in centiMorgans (cM), $\mathbf{A}_{P(a,b)}$ is the PB relationships between individuals a and b, $\mathbf{A}_{M_N(a,b)}$ is the MB relationship computed using equation [8.3]. Equation [8.4] reduces to equation [8.3] when there is full marker coverage of the genome. Alternatively, when no marker information is available equation [8.4] reduces to the PB relationships.

Calculation of MB relationships using equation [8.4] proceeded by assuming that each marker allele in the founder generation was unique although they may be identical in-state and thus founder animals were assumed unrelated similar to the PB relationships. Each marker allele on the animal's haplotypes was recoded so that the number of alleles in the founder generation was twice the number of the founder animals. Then, each descendant's haplotypes were expressed in terms of the founder alleles. Using the recoded animal haplotypes, the inheritance at each marker locus could be traced from one generation to the next. A java program was written to compute the MB relationship using equation [8.4].

The data considered in this study included genotypic information on sires and their sons. This genotypic information was used to infer the linkage segments or haplotypes for the sire and its sons. Alleles contributed by the dam were also inferred since marker information on the dam was not available. The construction of the linkage haplotypes for the sire and its sons proceeded using the simple rules of logic assuming that the linkage order of markers has already been established. For illustrative purpose, the rules were applied to the hypothetical example in Table 8.2. Given that the genotypes

of the sire and son 1 at the first marker locus in Table 8.2 are 1,2 and 2,4 respectively, it can be determined that the paternal and maternal allele of son 1 were 2 and 4, respectively. Thus, construction of the linkage haplotype proceeds as follows:

Step 1: Assignment of the son's paternal and maternal haplotypes

This step is straightforward when markers are informative i.e. the parental origin of the two alleles carried by the son can be ascertained based on the genotypes of the sire and its son at the marker locus of interest as is the case at the first marker locus in Table 8.2. However, when markers are not informative the parental origin of the sons' alleles could not be determined. Therefore, at the end of this step parental origins of the alleles carried by the son at certain marker loci were unresolved. An example of an unresolved linkage haplotype is shown in Table 8.2. The unresolved locus was marker 2 for son 1.

Step 2: Construct the sire haplotypes

The sire haplotypes were assigned based on the sons' paternal haplotype constructed in step 1, i.e. the P haplotype in Table 8.2. Assignment of the sire alleles at the first marker locus into paternal or maternal was arbitrary because the sire's parental genotypes were unavailable. Assignment of alleles at the remaining marker loci is based on the frequency of occurrence of the sire haplotypes already constructed from the previous markers and the two alleles contributed by the sire to the sons at the current marker locus. For example, consider the second marker locus. At this stage the sire

haplotype comprise of alleles 1 and 2 at the first marker locus. Allele 1 occurred with allele 3 at the second marker locus twice and once with allele 1. Therefore, the sire haplotypes including marker 2 are 13 and 21. This process was repeated until the entire sire haplotypes were constructed.

Table 8.2. The sire and its sons' genotypes and haplotypes at four marker loci.

Animal	**M1**	**M2**	**M3**	**M4**	**Allele origin**	**Haplotypes**			
Sire	1,2[a]	1,3	2,3	2,2	?[b]	1	3	2	2
					?[b]	2	1	3	2
Son 1	2,4	1,3	1,3	1,2	P	2	1/3[c]	3	2
					M	4	1/3[c]	1	1
Son 2	1,1	2,3	1,2	2,2	P	1	3	2	2
					M	1	2	1	2
Son 3	2,4	1,5	3,4	1,2	P	2	1	3	2
					M	4	5	4	1
Son 4	1,4	1,4	2,4	2,3	P	1	1	2	2
					M	4	4	4	3
Son 5	1,3	3,5	2,5	2,6	P	1	3	2	2
					M	3	5	5	6

[a] Two alleles carried by the sire at the first marker, [b] parental origin of the sire alleles unknown, [c] allele origin is unknown (unresolved) before the sire haplotype is constructed after which the paternal and maternal alleles for son 1 are more likely to be 1 and 3, respectively, M1 through M4 are marker loci 1 to 4, P and M denotes the paternal and maternal haplotype respectively.

Step 3: Update the sons' haplotypes at the unresolved marker locus

This was achieved by identifying the closest informative locus to the left or right of the unresolved marker locus on the son's paternal haplotype. From Table 8.2, the origin of the two alleles at the second marker locus was unresolved for son 1. Based on the sire haplotypes it can be concluded that the paternal and maternal alleles for son 1 was 1 and 3 respectively. Therefore, if the allele carried by the son at the informative marker locus is of the sire's paternal origin, the allele at the unresolved locus is assigned to the sire's paternal allele and vice verse. The decision assumes that no recombination occurred between the sire's closest informative and unresolved marker locus. The error rate associated with this decision is a function of the distance between the informative marker and the marker of interest.

From the example in Table 8.2 it was demonstrated that parental haplotypes could be determined with limited number of progeny. With large number of progeny available in large half-sib families parental haplotypes should be assigned accurately. Based on the procedure outlined above, a java program was written to assign haplotypes when genotypic information is available on one parent and several progeny. This program accommodates limited missing genotype of the parent or progeny at certain marker loci. Missing genotypes for the common parent at a particular marker locus are inferred using progeny genotypes by considering the two most frequent alleles as the parent's genotype at the marker locus of interest, which is reasonable for half-sib families where the two most frequent alleles are likely to be from the common parent. This program has an added capability to check for compatibility between parent and progeny genotypes. When

the parent and progeny genotypes are incompatible, the progeny genotype is ignored in the haplotype construction process.

8.4. Results

Additive genetic relationships. The number of nonzero elements in the PB and MB relationship matrices was the same. However, the MB relationships expressed deviation from the expected relationships computed based on pedigree information alone i.e. the PB relationships. The deviation was however small because the marker information only accounted for 20% of the genome such that pedigree information still played a major role in MB relationships. For example, in one of the half-sib families with 79 siblings the MB relationships ranged from 0.231 to 0.278 with an average relationship of 0.256 which is close to the expected relationships among half-sibs of 0.25. The deviation of MB from PB provided some knowledge about the actual sample of genes shared between relatives i.e. the covariance between Mendelian sampling effects. The accuracy associated with the estimate of the covariance between Mendelian sampling effects has a bearing on the accuracy of the genetic prediction.

Distribution of EBV. Results about the distribution of the three sets of EBV are presented in Table 8.3. The notable difference among the three sets of EBV was on their variability. Across traits, EBV-ALL exhibited the most variability followed by EBV-MRK and then EBV-PED. These results were expected since the variability of the prediction depends on its accuracy. The accuracy of prediction is the correlation between true breeding value and its prediction (i.e. $r_{g,\hat{g}} = \frac{\mathrm{cov}(g,\hat{g})}{\sigma_g.\sigma_{\hat{g}}}$ where g is the true breeding

value). But for best linear prediction $\text{cov}(g, \hat{g}) = \sigma^2_{\hat{g}}$ and thus $r_{g,\hat{g}} = \frac{\sigma_{\hat{g}}}{\sigma_g}$. Therefore, $\sigma_{\hat{g}} = r_{g,\hat{g}}\sigma_g$ and since σ_g is a constant, the differences in the variability of the predictions reflect the accuracy of the prediction ($r_{g,\hat{g}}$).

Table 8.3. Descriptive statistics for EBV-ALL, EBV-PED and EBV-MRK for milk, fat and protein yield in lbs for 849 sons.

Trait	EBV	Mean	SD.	Min.	Max.	CV.
Milk yield						
	EBV-ALL[a]	137	1192	-6465	4267	871
	EBV-PED[b]	230	455	-948	1567	197
	EBV-MRK[c]	224	476	-1273	1652	213
Fat yield						
	EBV-ALL	16	45	-145	159	283
	EBV-PED	13	17	-44	56	136
	EBV-MRK	13	18	-47	60	137
Protein yield						
	EBV-ALL	13	36	-120	130	283
	EBV-PED	13	15	-38	57	114
	EBV-MRK	13	16	-42	59	120

[a] EBV computed using all the DYDs and the A_P^{-1}, [b] EBV computed using all DYDs except for the sire to be evaluated and A_P^{-1}, [c] EBV computed similar as EBV-PED except that A_P^{-1} was replaced by A_M^{-1}

Pearson linear correlation. Table 8.4 provides evidence of significant correlations between EBV-ALL and EBV-PED or EBV-MRK within and across chromosomes. There was an increase in accuracy of genetic prediction for milk yield

when pedigree and marker information (from all chromosomes) were considered in computing the relationships. The enhancement in accuracy was 4.3%. On the other hand, use of marker information lead to no change in accuracy for fat and protein yields. Increases and decrease in accuracy occurred when relationships computed from pedigree and marker information within chromosome were used. For example, supplementing pedigree with marker information resulted in the same or higher accuracy than use of pedigree information alone for milk and protein yields while predictions of less, same or higher accuracy were observed for fat yield. Markers on chromosomes 1, 14 and 20 for milk yield and 3 and 10 for protein led to more accurate prediction. On the other hand, markers on chromosome 3 provided useful information for computation of relationships. These, results are in agreement with QTL studies that show that QTL for different traits are located at different regions of the genome. For instance, there is increase in evidence of association between markers and QTL on chromosome 1 for milk yield as shown in Appendix I. This may explain the increase in accuracy observed from using markers on chromosome 1 for milk yield.

Considering all the traits, the correlations were consistently higher for protein yield compared to milk and fat yields. In fact, fat yield had the lowest correlations i.e. the least accurate. Given that the same heritability was assumed for the three traits, it is unclear why the protein yield was more accurate compared to milk and fat yields.

Table 8.4. The Pearson correlation between EBV-ALL and EBV-PED ($r_{ALL,PED}$) or EBV-MRK ($r_{ALL,MRK}$) within and across chromosome for 849 sons.

Trait	$r_{ALL,PED}$	$r_{ALL,MRK}$						
		GENOME[a]	CHR1	CHR3	CHR9	CHR10	CHR14	CHR20
Milk	0.47	**0.49**[i]	**0.48**[i]	0.47	0.47	0.47	**0.48**[i]	**0.48**[i]
Fat	0.37	0.37	0.37	**0.38**[i]	**0.36**[d]	0.37	0.37	**0.36**[d]
Protein	0.62	0.62	0.62	**0.63**[i]	0.62	**0.63**[i]	0.62	0.62

[a] refers to all the six chromosomes, [i] increase, [d] decrease.

Spearman rank order correlations. Rank correlations were consistently higher for EBV-MRK compared to EBV-PED indicating that use of marker information provides better ranking between sires compared to use of pedigree information alone irrespectively of the accuracy of prediction. Similar to the linear correlations, rank correlations were higher for milk and protein yield relative to fat yield. Markers on chromosome 1 provided the same rank correlations as using the markers on all the chromosomes.

Table 8.5. The Spearman rank order correlation between EBV-ALL and EBV-PED ($r_{ALL,PED}$) or EBV-MRK ($r_{ALL,MRK}$) within and across chromosome for 849 sons.

Trait	$r_{ALL,PED}$	$r_{ALL,MRK}$						
		GENOME[a]	CHR1	CHR3	CHR9	CHR10	CHR14	CHR20
Milk	0.38	0.44	0.42	0.40	0.40	0.41	0.42	0.41
Fat	0.27	0.30	0.29	0.31	0.30	0.30	0.30	0.29
Protein	0.54	0.59	0.59	0.58	0.56	0.57	0.57	0.56

[a] refers to all the six chromosomes.

8.5. Discussion

Results from this study demonstrated using empirical data that accuracy of genetic prediction can be enhanced through joint use of pedigree and genetic marker information. However, the improvement achieved here was not dramatic primarily due to relatively small number of markers considered. The markers used were only restricted to six chromosomes and represented only 20% of the genome. In addition, the average distance between consecutive markers was 13 cM, which may play a major role in determining the crossover events and assignment of haplotypes. In a simulation study, Nejati-Javaremi (1995) observed change in accuracy of genetic prediction from using marker information (to calculate genetic relationships) ranging from 4.4 to 34.6% for a trait with a heritability of 0.3 and the recombination rate between subsequent markers of 0.02. When the recombination rate was increased to 0.20, the change in accuracy ranged from 0.0 to 8.0%. Therefore, the change in accuracy ranging from 0.0 to 4.3% observed in the current study is within theoretical expectations.

The change in accuracy also depended on the trait considered. For example, in some cases the accuracy declined instead of increasing when MB relationships were substituted for PB relationships for fat yield. This observation suggests that marker information should be use cautiously and perhaps different MB relationships matrices should be used for different traits, which may be less practical for large-scale genetic evaluation. However, an alternative may be to use similar MB relationships matrices for traits that are highly correlated which will reduce the number of relationship matrices required.

Markers used in this study were not chosen based on any knowledge of their association with the traits evaluated but were in interesting QTL regions. When there is evidence of an association between markers and trait phenotypes, the accuracy of genetic prediction may be further enhanced by using these markers. In the current study, the proportion of the segments in common was weighted by the size (in cM) of the segment. This may not be optimal since the length of the segment does not necessarily imply the magnitude of its influence on trait phenotype. Alternately, when information on the genomic regions influencing performance is available, the proportion of variance accounted by the region may be used to weight the relationships.

8.6. Conclusions

The results obtained in the current study provide first line of empirical evidence of an increase in accuracy when marker information was used in conjunction with pedigree information to compute the additive relationships between animals. The use of marker information led to better ranking of sires irrespective of the accuracy associated with predictions. However, the increase in accuracy depended on the trait evaluated implying that different relationships matrices should be considered for different traits. These results suggest that marker information may be used successfully to increase accuracy of genetic prediction particularly for young sires without own performance. For the dairy industry in particular, this will allow for selection among candidate sons to be placed under progeny testing. Further research should focus on evaluation of trait specific marker-based relationships.

8.7. Literature Cited

Henderson, C. R. 1976. A simple method for computing the inverse of the numerator relationship matrix used in prediction of breeding values. Biometrics 32:69-83.

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. 7:235-249.

Nejati-Javaremi, A. 1995. Alternative methods for defining relationship, assigning haplotypes and measuring linkage in animal breeding. Ph.D. Dissertation. University of Guelph.

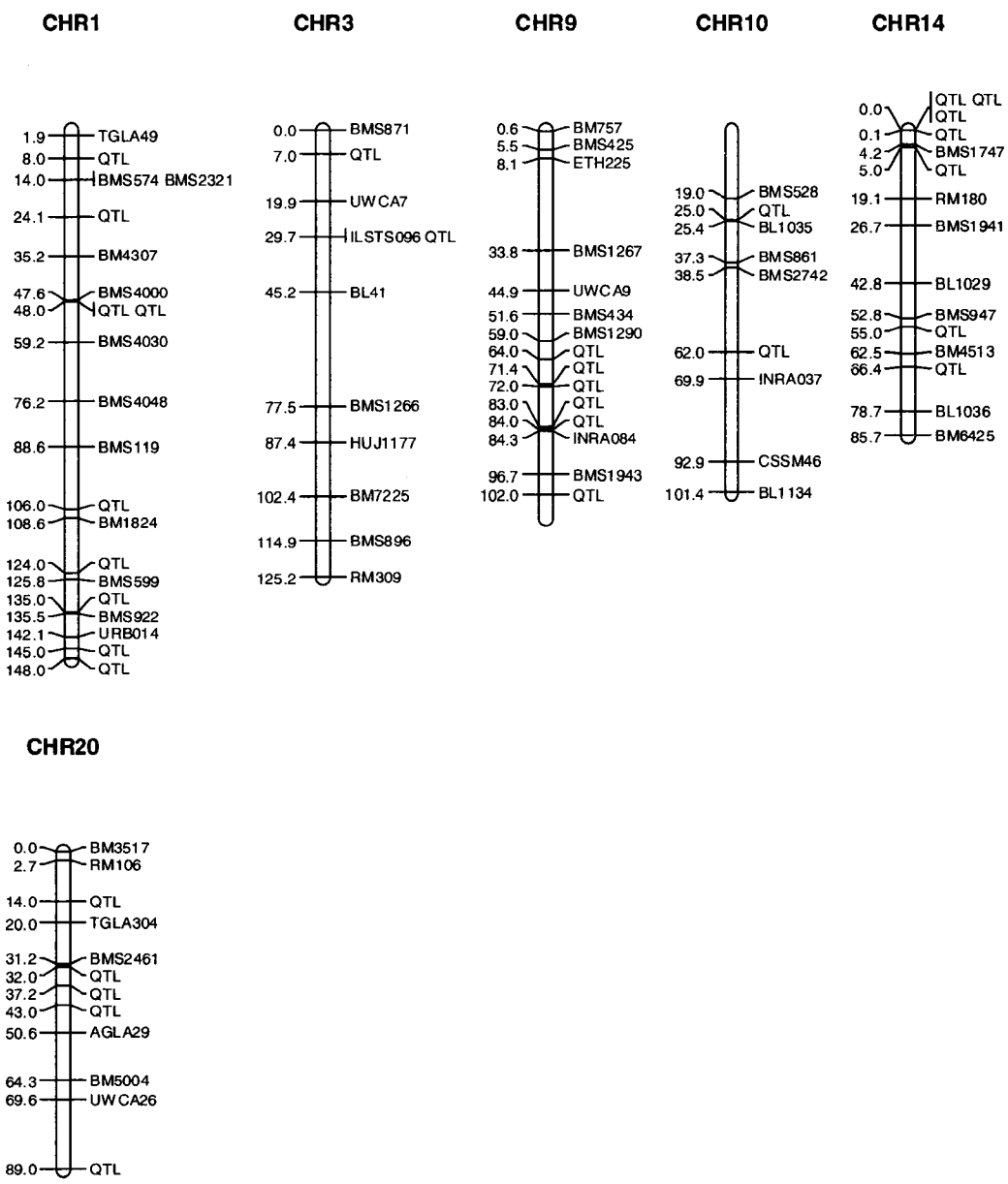
APPENDIX

This section provides diagrams of the chromosomal linkage groups depicting relative marker and QTL positions for milk, fat and protein yields. The QTL shown here were obtained from a meta-analysis of QTL studies in dairy cattle by Khatkar et al. (2004; Genet. Sel. Evol. 36:163-190). The online combined QTL map is accessible at http://www.vetsci.usyd.edu.au/reprogen/QTL_Map/?QTL=Yes. In addition, this section provides *awk* scripts and java programs that were created and used in the analyses of data in this study. A detailed description and source code for each script or program is also given.

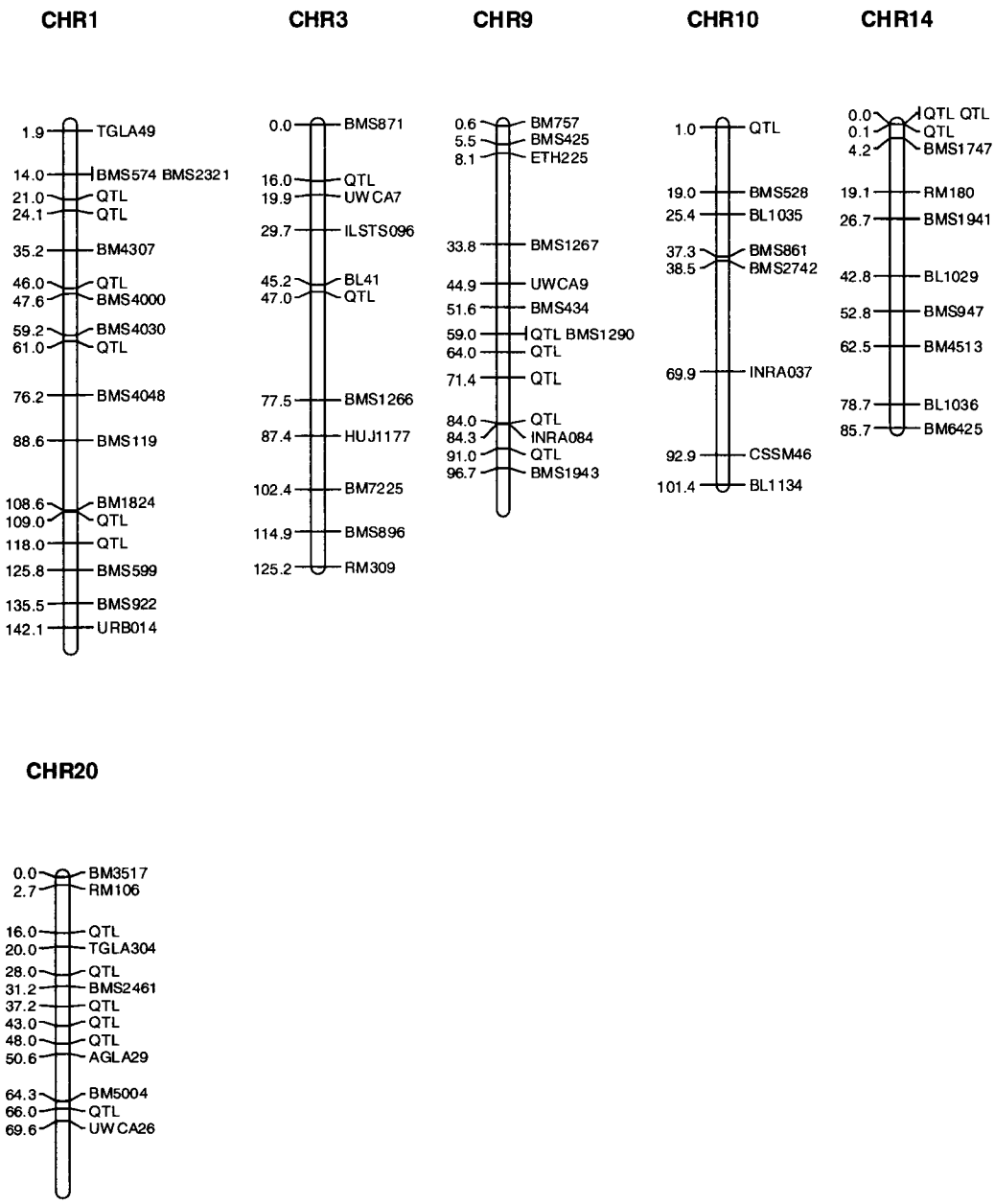
APPENDIX I

CHROMOSOMAL LINKAGE GROUPS WITH PUTATIVE QTL FOR MILK, FAT AND PROTEIN YIELDS

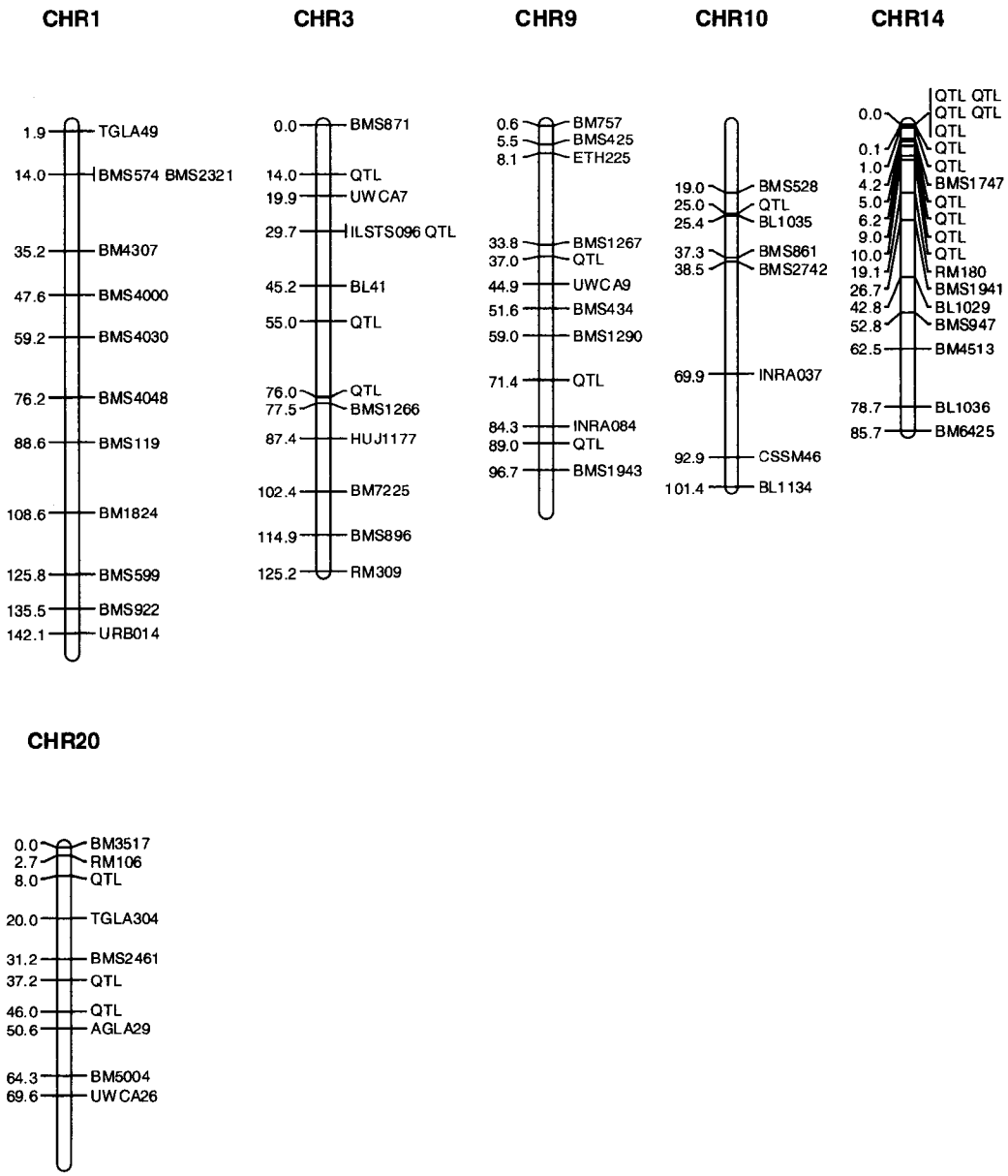
Below are the chromosomal linkage groups with QTL for milk yield



Below are the chromosomal linkage groups with QTL for protein yield



Below are the chromosomal linkage groups with QTL for fat yield



APPENDIX II

AWK SCRIPT TO COMPUTE THE LIKELIHOOD RATIO F-VALUE STATICTIC

FOLLOWING THE PROCEDURE OF Boik et al. (1993)

```
##############################################################################################
#
#       SCRIPT          :       ./CalcBoikFvalue
#       AUTHOR          :       Maiwashe Azwihangwisi
#       DATE            :       02/27/03
#       PURPOSE         :
#                       This script computes the F-value following Boik et al. (1993) approach
#
#       SYNTAX          :
#                       execute command : ./CalcBoikFvalue
#
#       NOTES           :
#
#       This script compute the F-value for the test of significance for fixed effects when a mixed model is
#       fitted to the data. Boik et al. (1993: JAS, 71:51-56) proposed the following Likelihood Ratio
#       statistic for testing the  significance of fixed effects:
#
#               F = [SS_1/(r - r_1)]/[SSE/(n - r)]
#
#       where SS_1 is the difference between the error sums of squares of the full and reduced models, r
#       and r_1 are the ranks for the incidence matrices of the full and reduced models, respectively, n is the
#       number of observation, SSE is the sums of squares of error of the full model. The solution to the
#       mixed model equations is required to obtain SS_1 and SSE. That is,
#
#               SS_1 = y'R^-1(Xb*_hat - X_1b_1*_hat ) + y'R^-1Z(u_hat - u_1_hat)
#
#       and
#
#               SSE = y'R^-1y - y'R^-1Xb*_hat - y'R^-1Zu_hat
#
#       where  y is a vector of observations, b*_hat and b_1*_hat are the solutions for the fixed effects from
#       the full and reduced models, respectively, u_hat  and u_1_hat are the solution for the random effects
#       from the full and reduced models, respectively, R^-1 is the inverse of the variance-covariance matrix
#       of the residual effects.
#
#       This procedure requires the knowledge of the variance ratio of the random effects.
#
##############################################################################################


# extract B_r and u_r from the reduced model

awk 'NR<=138 {print $1}' beta > B_r
awk 'NR>138 {print $1}' beta > u_r

# extract B_f and u_f from the full model

awk 'NR<=139 {print $1}' ../ALL/beta > B_f
awk 'NR>139 {print $1}' ../ALL/beta > u_f

cp X X_r
cp ../ALL/X X_f
cp ../ALL/Z Z
cp ../ALL/Y Y.s
```

```
#####################################
# DON'T CHANGE ANYTHING FROM HERE
#####################################

# calculate SS1

mult X_f B_f X_f.B_f
mult X_r B_r X_r.B_r
awk 'NR>1 {print $3}' X_f.B_f >  X_f.B_f.col
awk 'NR>1 {print $3}' X_r.B_r > X_r.B_r.col
subtract X_f.B_f.col X_r.B_r.col XB
subtract u_f u_r u
d2s u u.s
mult Z u.s Zu
d2s XB XB.s
mult -tY.s XB.s Y_XB
mult -tY.s Zu Y_Zu
sadd Y_XB Y_Zu SS1.s
awk 'NR>1 {print $3}' SS1.s > SS1


# calculate SSE

mult -tY.s Y.s Y.tot
mult -tY.s X_f.B_f SS.fixed
mult Z u_f Z.u_f
mult -tY.s Z.u_f SS.random
awk 'NR==1 {$1=1} {print $1,$2,$3}' Y.tot > Y.tot.s
awk 'NR==1 {$1=1} {print $1,$2,$3}' SS.fixed > SS.fixed.s
awk 'NR==1 {$1=1} {print $1,$2,$3}' SS.random > SS.random.s

subtract Y.tot.s SS.fixed.s Y_adjFixed
awk 'NR>1 {print $1,$2,$3}' Y_adjFixed > Y_adjFixed.s
subtract Y_adjFixed.s SS.random.s SSE.s
awk 'NR==3 {print $3}' SSE.s > SSE

# rank of reduced and full model

r_r=`awk '{print $1}' rank.r`
r_f=`awk '{print $1}' rank.f`
N=`awk '{print $1}' NumRec`

awk 'NR>1 {print "SSTot = ",$3}' Y.tot.s
awk '{print "SS1 = ",$1}' SS1
awk 'NR>1{print "SSFixed = ",$3}' SS.fixed.s
awk 'NR>1{print "SSRand = ",$3}' SS.random.s
awk '{print "SSE = ",$1}' SSE

echo "Rank FullModel = " $r_f
echo "Rank ReducModel = " $r_r
echo "Num Records = " $N

paste SS1 SSE NumRec rank.f rank.r > f1

# calculate F value
```

```
awk -f calc_F.awk f1 > File
F=`awk '{print $1}' File`

echo "F-value = " $F

cat Y.tot.s SS1 SS.fixed.s SS.random.s SSE NumRec rank.f rank.r > out.file
```

APPENDIX III

AWK SCRIPT TO COMPUTE EBV FOR SIRES EXCLUDING INFORMATION ON DAUGHTER PERFORMANCE (DYD)

```
#############################################################################################
#
#       SCRIPT          :       ./CalcEBV
#       AUTHOR          :       Maiwashe Azwihangwisi
#       DATE            :       03/05/04
#       PURPOSE         :
#                       This script calculates EBV for the sire excluding its Daughter
#                       Yield Deviation information
#       SYNTAX          :
#                       execute command : ./CalcEBV numRec
#                       numRec is the number of records (sires) in the data file
#
#       DESCRIPTION     :
#
#       The EBV for each sire is computed as follows:
#
#       Step 1:
#               The DYD for the ith sire is zeroed
#
#       Step 2:
#               Assemble the mixed model equations and solve iteratively
#
#       Step 3:
#               Store the EBV for the ith sire in the solution file
#
#       Step 4:
#               Repeat step 1 to 3 until all the sires have been processed
#
#
#       NOTES           :
#
#       The MME are as follows:
#
#       Z'Z+lambda*Ainv = Z'y
#
#       where lambda = (4 – h2)/4h2, Z'Z is a diagonal matrix whose elements are 0.25*n_i, Z'y is a vector
#       whose ith element is 0.5* n_i*DYD_i, n_i is the number of daughters for the ith sire, Ainv is the
#       inverse of the relationship matrix.
#############################################################################################

# Initialize the solution file (EBVstart) to zero

awk 'NR==1 {print 0}' EBVmilk > EBVstart

#################################################################
# Compute the inverse of the relationship matrix (pedigree or marker-based)
#################################################################

# The relationship matrix is sparse stored

mprint -p A > A.sq
invert -i A.sq -o Ainv
d2s Ainv Ainv.s
```

```
i=1
NUMREC=$1  # number of animals in the data file

lambda=`awk 'NR==1 {print (4-0.3)/(4*0.3)}' EBVstart`
echo $lambda

#####################
# Beginning of the loop
#####################

while
        test $i -le $NUMREC

        do

                ###########################################################
                # Zero the DYD (column 2 in data) and number of daughters (column 3)
                # for the ith sire
                ###########################################################

                java ZeroRecords data data1 $i 2 3

                ###################################
                # Create LEFT-HAND SIDE of MME
                ###################################

                awk '{print $3*(1/4)}' data1 > Nvec
                awk -f diagonal.awk -v NumRec=$NUMREC Nvec > diag
                cat headZtZ diag | d2s - ZtZ

                sadd ZtZ Ainv.s*$lambda LHS


                ###################################
                # Create RIGHT-HAND SIDE of MME
                ###################################

                awk '{print $2*$3*(1/2)}' data1 > Zty
                d2s Zty RHS

                ###################################
                # Solve the MME iteratively
                ###################################

                halfit -C LHS -R RHS -c 1.0e-12 -m 10000 -B EBVmilk -o -

                awk -f GetRec.awk -v REC=$i EBVmilk > EBVtemp

                echo $i

                cat EBVstart EBVtemp > EBVmilkCorr

                cp EBVmilkCorr EBVstart
                rm ZtZ LHS RHS diag

                i=`expr $i + 1`
```

```
done

######################
# END of the loop
######################

rm EBVtemp
echo END
```

APPENDIX IV

AWK SCRIPT TO COMPUTE EBV FOR SIRES INCLUDING INFORMATION ON DAUGHTER PERFORMANCE (DYD)

```
#############################################################################################
#
#       SCRIPT          :       ./CalcEBVPed
#       AUTHOR          :       Maiwashe Azwihangwisi
#       DATE            :       03/05/04
#       PURPOSE         :
#                       This script calculates EBV for the sires including own Daughter Yield Deviation
#
#       SYNTAX          :
#                       execute command : ./CalcEBVPed
#
#       NOTES           :
#
#       Similar to ./CalcEBVPedMarker
#
#############################################################################################

# The data file has 3 columns i.e. animID, DYD, n where n is the number of daughters

awk '{print $1}' data > relFile    # relFile contains the ID of the animals whose relationships are required

lambda=`awk 'NR==1 {print (4-0.3)/(4*0.3)}' EBVmilk1`
echo $lambda

#################################################################
# Compute the inverse of the numerator relationship matrix
#################################################################

mprint -p Aped > A
invert -i A -o Ainv
d2s Ainv Ainv.s
NUMREC=`wc data | awk '{print $1}'`
echo $NUMREC

##############################
# CREATE LHS
##############################

awk '{print $3*(1/4)}' data > Nvec
awk -f diagonal.awk -v NumRec=$NUMREC Nvec > diag
cat headZtZ diag | d2s - ZtZ
sadd ZtZ Ainv.s*$lambda LHS

##############################
# CREATE RHS
##############################

awk '{print $2*$3*(1/2)}' data > Zty
d2s Zty  RHS

##############################
# Solve the MME
##############################

halfit -C LHS -R RHS -c 1.0e-12 -m 10000 -B EBVmilk1 -o -
```

APPENDIX V

SOURCE CODE FOR A JAVA PROGRAM TO COMPUTE MARKER-BASED NUMERATOR RELATIONSHIP MATRIX

```
/*****************************************************************************
/       CLASS       :       MarkerBasedRelationship.java
/
/       AUTHOR      :       Maiwashe Azwihangwisi
/
/       DATE        :       Created 12/29/03
/                           Modified 07/31/04
/
/       PURPOSE     :
/                           Primary – calculate the numerator relationships conditional on pedigree
/                           and DNA marker information
/                           Secondary – calculate Wright's numerator relationship matrix
/
/       SYNTAX      :
/                           Compile command:
/                           javac MarkerBasedRelationship.java
/
/                           Execute command:
/                           java MarkerBasedRelationship inputFile founderFile relationFile
/                           MarkerMatrixFile PedMatrixFile numRec1 numRec2 numMarkers
/                           markerIntervalFile
/
/       ARGUMENTS DESCRIPTION:
/
/                   inputFile – file with animal records. Each animal has two records with the first
/                   record containing the pedigree and paternal haplotype and the second record has
/                   pedigree and maternal haplotype
/                   e.g.        animID  sireID  damID   marker1 marker2 marker3
/                       record1 10      5       3       2       1       4
/                       record2 10      5       3       1       3       1
/
/                   founderFile – file with a list of founder animals (a single column file)
/                   relationFile – file with animals whose relationships are require
/                   MarkerMatrixFile – output file with marker relationships in sparse storage
/                   PedMatrixFile – output file with standard numerator relationship matrix
/                   numRec1 – number of animals in inputFile
/                   numRec2 – number of animals in relationFile
/                   numMarkers – number of marker loci
/                   markerInterval – file with between marker distance within a chromosome. If
/                   there are more than one chromosome, the distance between the last marker locus
/                   on the first chromosome and the first marker on the second chromosome is a
/                   zero
/
/       NOTES       :
/
/       This program calculates genetic relationships conditional on pedigree and DNA marker
/       information using the modification of the approach proposed by Nejati-Javaremi (1995,
/       Dissertation). As in the pedigree-based method, founder animals are assumed to be unrelated. That
/       is, alleles carried by the founder animals are assumed to be unique. Essentially, the relationships
/       are computed at the segments between two consecutive DNA markers. These segments are joined
/       together to form a linkage group or haplotype such that chromosomes are joined together into a
/       continuum. Since each individual has two linkage segments (i.e. the paternal and maternal linkage
/       segments (LS)), there are four possible ways for two individual to be related. Nejati-Javaremi
/       (1995) proposed the following formula to compute marker-based relationships assuming complete
/       genome marker coverage:
/
```

```
/                RELmarker(a,b) = [LS(p,p)+LS(p,m)+LS(m,p)+LS(m,m)]/2
/
/        where RELmarker(a,b) is the marker-based relationship for individuals a and b, LS(p,p) is the
/        proportion of the paternal linkage segment in common between individuals a and b. These
/        relationships range from 0 to 2 similar to the numerator relationships. The equation above was
/        modified to accommodate a scenario of incomplete genome marker coverage as follows:
/
/                RELmodified(a,b) = [r.RELped(a,b) + s.RELmarker(a,b)]/l
/
/        where r and s represents the interval of the genome with and without marker coverage,
/        respectively (r and s should sum to l), l is the total genome length, RELped(a,b) is the pedigree
/        based relationships between individuals a and b, RELmarker(a,b) is Nejati-Javaremi's relationship
/        at interval with genome marker coverage from the first equation.
/
/        The proportion of the LS in common is determined as follows:
/
/        The segment between two consecutive DNA markers is partitioned into two halves, with the
/        marker allele from the marker on the left representing 50% of the segment and that on the right
/        representing the remaining 50%. The length of the linkage segment is explicitly taken into
/        account.
/
/        The program is structured as follows:
/
/        1. The founder alleles are recoded so that the number of alleles in the founder population is twice
/        the number of the founder animals.
/
/        2. Using the paternal and maternal haplotype of the descendents the origin of the descendents' two
/        alleles is established and this is continued until all the origin of each allele has been ascertained.
/        When the parent is homozygous at a particular locus, the origin of the descendant's allele is based
/        on the closest heterozygous marker of the parent. On the other hand, when the parent is
/        heterozygous, the origin of the descendant's alleles is straightforward.
/
/        3. The relationships are computed using the recoded marker alleles.
/
/*****************************************************************************/

import java.util.*;
import java.io.*;
import java.lang.*;
import java.text.*;

public class MarkerBasedRelationship{
    public static void main(String[] args) throws IOException{

        /*-------------------------------------------------------------------------------------------------
        /
        / READ RECORDS FROM EXTERNAL FILE AND STORE THEM IN AN ARRAY
        /
        /-------------------------------------------------------------------------------------------------*/

        BufferedReader inFile = new BufferedReader(new FileReader(args[0]));
        BufferedReader founderFile = new BufferedReader(new FileReader(args[1]));
        BufferedReader relationFile = new BufferedReader(new FileReader(args[2]));
        PrintWriter outFile = new PrintWriter(new BufferedWriter(new FileWriter(args[3])));
        PrintWriter outAmatrix = new PrintWriter(new BufferedWriter(new FileWriter(args[4])));
```

```
int NUMANIM = Integer.parseInt(args[5]), NUMRELATEDANIM = Integer.parseInt(args[6]);
int NUMMARKERS = Integer.parseInt(args[7]), VECTORSIZE = 5000;
final int PEDCOLS = 3, NUMROWS = 2, COUNTROWS = 0;
String[][] pedArray = new String[NUMANIM][PEDCOLS];
String[][][] haploArray = new String[NUMANIM][NUMROWS][NUMMARKERS];

int countAnim = 0; String linePat = inFile.readLine(), lineMat = inFile.readLine();
StringTokenizer tokenPat, tokenMat;

while(linePat != null){

  tokenPat = new StringTokenizer(linePat); tokenMat = new StringTokenizer(lineMat);
  tokenMat.nextToken(); tokenMat.nextToken(); tokenMat.nextToken();

  // store the pedigree

  for(int i = 0; i < PEDCOLS; i++){
    pedArray[countAnim][i] = tokenPat.nextToken();
  }


  for(int i = 0; i < NUMMARKERS; i++){
    haploArray[countAnim][COUNTROWS][i] = tokenPat.nextToken();
    haploArray[countAnim][COUNTROWS + 1][i] = tokenMat.nextToken();
  }

  linePat = inFile.readLine(); lineMat = inFile.readLine(); countAnim++;
} // close the WHILE loop

inFile.close();

/*-------------------------------------------------------------------------------------------------------------
/
/ RECODE THE MARKER ALLELES SO THAT EACH ALLELE CAN BE TRACED
/ TO THE FOUNDER ANIMAL. THE NUMBER OF ALLELES IN THE FOUNDER
/ POPULATION ARE TWICE THE NUMBER OF FOUNDER ANIMALS. EACH ALLELE IS
/ UNIQUE.
/-------------------------------------------------------------------------------------------------------------*/

Vector parents = new Vector(VECTORSIZE); Vector progeny = new Vector(VECTORSIZE);

/*-----------------------------------------------------
/
/  PROCESS THE FOUNDER POPULATION
/
/-----------------------------------------------------*/
String lineFounder = founderFile.readLine();

while(lineFounder != null){
  parents.addElement(lineFounder);
  lineFounder = founderFile.readLine();
}

final int NUMFOUNDER = parents.size();
 founderFile.close();
String[][][] haploArrayNew = new String[NUMANIM][NUMROWS][NUMMARKERS];
```

```
String founderAllele1, founderAllele2, nextHeteroLocusAllele1, nextHeteroLocusAllele2, sex;

int countPed = 0, countFound, countNextHeteroLocus, newCountNextHeteroLocus;
int checkIfHetero, countProgenyAdded = 0, countProg, numParents = parents.size();

for(int i = 0; i < numParents; i++){

  // find the ith founder in pedArray

  countFound = 0; countPed = 0;

  while(!pedArray[countPed][0].equals((String)parents.elementAt(i)) && countFound == 0){

    if(pedArray[countPed][0].equals((String)parents.elementAt(i))){
      countFound++;
    }
    countPed++;
  }

  if(countFound != 0){
    countPed--;
  }


  // find the progeny of the ith founder in pedArray

  for(int j = 0; j < pedArray.length; j++){

    sex = null;

    if(pedArray[j][1].equals(pedArray[countPed][0])){
      sex = "M";
     }
     else
       if(pedArray[j][2].equals(pedArray[countPed][0])){
         sex = "F";
        }

        if(pedArray[j][1].equals(pedArray[countPed][0]) || pedArray[j][2].equals(pedArray
        [countPed][0])){

          // check progeny that become parents and store it in the progeny vector

          countProgenyAdded = 0; countProg = 0;

          while(countProg < pedArray.length && countProgenyAdded == 0){
            if(pedArray[countProg][1].equals(pedArray[j][0]) || pedArray[countProg] [2].equals
            (pedArray[j][0])){
              if(!progeny.contains(pedArray[j][0])){
                progeny.addElement(pedArray[j][0]);
                countProgenyAdded++;
              }
             }

             countProg++;
          }
```

```
/*-----------------------------------------------------------------------------------------------
/ Recode marker alleles for the current marker for each individual. This allows alleles to
/ be traced from ancestor to the progeny. Tracing of alleles is straightforward when
/ informative meiosis occurs. When uninformative meiosis coccurs the closest
/ heterozygous loci are used to trace the inheritance.
/-----------------------------------------------------------------------------------------------*/

for(int m = 0; m < NUMMARKERS; m++){

  // recode the founder alleles

  haploArrayNew[countPed][COUNTROWS][m] = pedArray[countPed][0] + "_" + m +
  "_" + 1;
  haploArrayNew[countPed][COUNTROWS + 1][m] = pedArray[countPed][0] + "_" +
  m + "_" + 2;
  founderAllele1 = haploArray[countPed][COUNTROWS][m];
  founderAllele2 = haploArray[countPed][COUNTROWS + 1][m];

  /*-----------------------------------------------------------------------------------------------
  / The ancestor is homozygous at current marker locus so we cannot tell whether the
  / progeny allele is the parental paternal or maternal allele
  /-----------------------------------------------------------------------------------------------*/

  nextHeteroLocusAllele1 = null;
  nextHeteroLocusAllele2 = null;

  if(founderAllele1.equals(founderAllele2)){

    /*-----------------------------------------------------------------------------------------------
    /  Two scenarios exist :
    /    - a heterozygous marker exists on the right of  side of the current marker and
    /    - a heterozygous marker exists to the left
    /-----------------------------------------------------------------------------------------------*/

    // by default we start searching  for the heterozygous marker locus to the right of the
    // current locus

    countNextHeteroLocus = m;

    if(m < NUMMARKERS - 1){ // there are still marker locus to the right
      countNextHeteroLocus = m + 1;
      checkIfHetero = 0;
      nextHeteroLocusAllele1 = haploArray[countPed][COUNTROWS]
      [countNextHeteroLocus];
      nextHeteroLocusAllele2 = haploArray[countPed][COUNTROWS +
      1][countNextHeteroLocus];

      if(!nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2)){
        checkIfHetero++;
      }

      while(nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2) &&
      (checkIfHetero == 0 && countNextHeteroLocus < (NUMMARKERS - 1))){

        // This allows the search for a heterozygous marker locus to go beyond the
```

```
        // current marker
         countNextHeteroLocus++;
         nextHeteroLocusAllele1 = haploArray[countPed][COUNTROWS]
         [countNextHeteroLocus];
         nextHeteroLocusAllele2 = haploArray[countPed][COUNTROWS + 1]
         [countNextHeteroLocus];

            if(!nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2)){
              checkIfHetero++;
            }
       }

       // a heterozygous to the left

       if(countNextHeteroLocus >= m && checkIfHetero == 0){
         newCountNextHeteroLocus = m - 1;
         nextHeteroLocusAllele1 = haploArray[countPed][COUNTROWS]
         [newCountNextHeteroLocus];
         nextHeteroLocusAllele2 = haploArray[countPed][COUNTROWS + 1]
         [newCountNextHeteroLocus];

         if(!nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2)){
           checkIfHetero++;
         }

         while(nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2) &&
         checkIfHetero == 0 && newCountNextHeteroLocus > 0){

           newCountNextHeteroLocus--;
           nextHeteroLocusAllele1 = haploArray[countPed][COUNTROWS]
           [newCountNextHeteroLocus];
           nextHeteroLocusAllele2 = haploArray[countPed][COUNTROWS + 1]
           [newCountNextHeteroLocus];

           if(!nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2)){
             checkIfHetero++;
           }
         }

         countNextHeteroLocus = newCountNextHeteroLocus;
       }
     }
     else
       // the current marker locus is the last marker, so we start searching for the
       // heterozygous locus to the left of the current marker locus

       if(m == NUMMARKERS - 1){
         checkIfHetero = 0;
         newCountNextHeteroLocus =m - 1;
         nextHeteroLocusAllele1 = haploArray[countPed][COUNTROWS]
         [newCountNextHeteroLocus];
         nextHeteroLocusAllele2 = haploArray[countPed][COUNTROWS + 1]
         [newCountNextHeteroLocus];

         if(!nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2)){
           checkIfHetero++;
```

```
    }
    while(nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2) &&
    checkIfHetero == 0 && newCountNextHeteroLocus > 0){

      newCountNextHeteroLocus--;
      nextHeteroLocusAllele1 = haploArray[countPed][COUNTROWS]
      [newCountNextHeteroLocus];
      nextHeteroLocusAllele2 = haploArray[countPed][COUNTROWS + 1]
      [newCountNextHeteroLocus];

      if(!nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2)){
        checkIfHetero++;
       }
     }

    countNextHeteroLocus = newCountNextHeteroLocus;
 } // close the ELSE loop

 // recode the progeny alleles

if(sex.equals("M")){
  if(haploArray[j][COUNTROWS][countNextHeteroLocus].equals(haploArray
  [countPed][COUNTROWS][countNextHeteroLocus])){
    haploArrayNew[j][COUNTROWS][m] =pedArray[countPed][0] + "_"
    + m + "_" + 1;
  }
  else
    if(haploArray[j][COUNTROWS][countNextHeteroLocus].equals
    (haploArray[countPed][COUNTROWS + 1][countNextHeteroLocus])){
      haploArrayNew[j][COUNTROWS][m] =pedArray[countPed][0]
      + "_" + m + "_" + 2;
    }
}
else
  if(sex.equals("F")){
    if(haploArray[j][COUNTROWS + 1][countNextHeteroLocus].equals
    (haploArray[countPed][COUNTROWS][countNextHeteroLocus])){
      haploArrayNew[j][COUNTROWS + 1][m] = pedArray[countPed][0]  + "_" +
      m + "_" + 1;
    }
    else
      if(haploArray[j][COUNTROWS + 1][countNextHeteroLocus].equals
      (haploArray[countPed][COUNTROWS + 1][countNextHeteroLocus])){
        haploArrayNew[j][COUNTROWS + 1][m] = pedArray[countPed][0] + "_"
        + m + "_" + 2;
      }
    }
  } // close the homozygous IF loop
  else
  /*-------------------------------------------------------------------------------------------
  / The ancestor is heterozygous at the current marker locus such that it can be
  / determined if the progeny allele is the paternal or maternal allele
  /-------------------------------------------------------------------------------------------*/

    if(!founderAllele1.equals(founderAllele2)){
```

```
                    // recode the progeny alleles
                    if(sex.equals("M")){
                      if(haploArray[j][COUNTROWS][m].equals(haploArray[countPed]
                      [COUNTROWS][m])){
                        haploArrayNew[j][COUNTROWS][m] = pedArray[countPed][0] + "_" + m
                        + "_" + 1;
                      }
                      else
                        if(haploArray[j][COUNTROWS][m].equals(haploArray[countPed]
                        [COUNTROWS + 1][m])){
                          haploArrayNew[j][COUNTROWS][m] = pedArray[countPed][0]
                          + "_" + m + "_" + 2;
                        }
                    }
                    else
                      if(sex.equals("F")){
                        if(haploArray[j][COUNTROWS + 1][m].equals(haploArray[countPed]
                        [COUNTROWS][m])){
                          haploArrayNew[j][COUNTROWS + 1][m] = pedArray[countPed][0] +
                          "_" + m + "_" + 1;
                        }
                        else
                          if(haploArray[j][COUNTROWS + 1][m].equals(haploArray[countPed]
                          [COUNTROWS + 1][m])){
                            haploArrayNew[j][COUNTROWS + 1][m] = pedArray[countPed][0] +
                            "_" + m + "_" + 2;
                          }
                      } // close the IF sex equals F loop
                    } // close the heterozgous IF loop
                  } // close the FOR loop
                } // close the IF loop
              } // close the j FOR loop

            countPed++;
} // close the i (main) FOR loop


/*-----------------------------------------------------------------------------
/
/ PROCESS THE DESCENDENTS OF THE FOUNDER POPULATION
/
/-----------------------------------------------------------------------------*/

// transfer the records in the progeny vector to the parents vector

int countAddedProgs = 0;
parents.removeAllElements();

while(countAddedProgs < progeny.size()){
  parents.addElement((String)progeny.elementAt(countAddedProgs));
  countAddedProgs++;
}

progeny.removeAllElements();

while(parents.size() != 0){
```

```
numParents = parents.size();

for(int i = 0; i < numParents; i++){

  // find the ith parent in pedArray

  countFound = 0;
  countPed = 0;

  while(!pedArray[countPed][0].equals((String)parents.elementAt(i)) && countFound == 0){

    if(pedArray[countPed][0].equals((String)parents.elementAt(i))){
      countFound++;
    }
    countPed++;
  }

  if(countFound != 0){
    countPed--;
  }

  // find the progeny of the ith parent in pedArray

  for(int j = 0; j < pedArray.length; j++){

    sex = null;

    if(pedArray[j][1].equals(pedArray[countPed][0])){
      sex = "M";
    }
    else
      if(pedArray[j][2].equals(pedArray[countPed][0])){
        sex = "F";
      }

    if(pedArray[j][1].equals(pedArray[countPed][0]) || pedArray[j][2].equals
    (pedArray[countPed][0])){

      // check if the progeny became a parent and store it in the progeny vector

      countProgenyAdded = 0;
      countProg = 0;

      while(countProg < pedArray.length && countProgenyAdded == 0){
        if(pedArray[countProg][1].equals(pedArray[j][0]) || pedArray[countProg][2].equals
        (pedArray[j][0])){
          if(!progeny.contains(pedArray[j][0])){
            progeny.addElement(pedArray[j][0]);
            countProgenyAdded++;
          }
        }
        countProg++;
      }
```

```
/*-----------------------------------------------------------------------------------------------
/ RECODE MARKER ALLELES FOR THE CURRENT MARKER FOR EACH
/ INDIVIDUAL.
/-----------------------------------------------------------------------------------------------*/

for(int m = 0; m < NUMMARKERS; m++){

  // recode the founder alleles

  founderAllele1 = haploArray[countPed][COUNTROWS][m];
  founderAllele2 = haploArray[countPed][COUNTROWS + 1][m];
  nextHeteroLocusAllele1 = null;
  nextHeteroLocusAllele2 = null;
  checkIfHetero = 0;
  countNextHeteroLocus = m;

  // parental alleles homozygous

  if(founderAllele1.equals(founderAllele2)){

    // there are still marker locus to the right

    if(m < NUMMARKERS - 1){
      countNextHeteroLocus = m + 1; checkIfHetero = 0;
      nextHeteroLocusAllele1 = haploArray[countPed][COUNTROWS]
      [countNextHeteroLocus];
      nextHeteroLocusAllele2 = haploArray[countPed][COUNTROWS + 1]
      [countNextHeteroLocus];

      if(!nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2)){
        checkIfHetero++;
      }

      while(nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2) &&
      checkIfHetero == 0 && countNextHeteroLocus < (NUMMARKERS - 1)){
        countNextHeteroLocus++;
        nextHeteroLocusAllele1 = haploArray[countPed][COUNTROWS]
        [countNextHeteroLocus];
        nextHeteroLocusAllele2 = haploArray[countPed][COUNTROWS + 1]
        [countNextHeteroLocus];

        if(!nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2)){
          checkIfHetero++;
        }
      }

      // a heterozygous to the left

      if(countNextHeteroLocus >= m && checkIfHetero == 0){

        newCountNextHeteroLocus = m - 1;
        nextHeteroLocusAllele1 = haploArray[countPed][COUNTROWS]
        [newCountNextHeteroLocus];
        nextHeteroLocusAllele2 = haploArray[countPed][COUNTROWS + 1]
        [newCountNextHeteroLocus];
```

```
      if(!nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2)){
        checkIfHetero++;
      }

      while(nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2) &&
      checkIfHetero == 0 && newCountNextHeteroLocus > 0){
        newCountNextHeteroLocus--;
        nextHeteroLocusAllele1 = haploArray[countPed][COUNTROWS]
        [newCountNextHeteroLocus];
        nextHeteroLocusAllele2 = haploArray[countPed]
        [COUNTROWS + 1][newCountNextHeteroLocus];

        if(!nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2)){
          checkIfHetero++;
        }
      }
      countNextHeteroLocus = newCountNextHeteroLocus;
    }
  }
  else
  // heterozygous to the left of the last marker
    if(m == NUMMARKERS - 1 && checkIfHetero == 0){
      newCountNextHeteroLocus = m - 1;
      nextHeteroLocusAllele1 = haploArray[countPed][COUNTROWS]
      [newCountNextHeteroLocus];
      nextHeteroLocusAllele2 = haploArray[countPed][COUNTROWS + 1]
      [newCountNextHeteroLocus];

      if(!nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2)){
        checkIfHetero++;
      }

      while(nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2) &&
      checkIfHetero == 0 && newCountNextHeteroLocus > 0){
        newCountNextHeteroLocus--;
        nextHeteroLocusAllele1 = haploArray[countPed][COUNTROWS]
        [newCountNextHeteroLocus];
        nextHeteroLocusAllele2 = haploArray[countPed][COUNTROWS + 1]
        [newCountNextHeteroLocus];

        if(!nextHeteroLocusAllele1.equals(nextHeteroLocusAllele2)){
          checkIfHetero++;
        }
      }

      countNextHeteroLocus = newCountNextHeteroLocus;
    } // close the ELSE loop

   // recode the progeny alleles

   if(sex.equals("M")){
     if(haploArray[j][COUNTROWS][countNextHeteroLocus].equals(
     haploArray[countPed][COUNTROWS][countNextHeteroLocus])){
       haploArrayNew[j][COUNTROWS][m] = haploArrayNew[countPed]
       [COUNTROWS][m];
```

```
    }
    else
      if(haploArray[j][COUNTROWS][countNextHeteroLocus].equals
      (haploArray[countPed][COUNTROWS + 1][countNextHeteroLocus])){
        haploArrayNew[j][COUNTROWS][m] = haploArrayNew[countPed]
        [COUNTROWS + 1][m];
      }
  }
  else
    if(sex.equals("F")){
      if(haploArray[j][COUNTROWS + 1][countNextHeteroLocus].equals
      (haploArray[countPed][COUNTROWS][countNextHeteroLocus])){
        haploArrayNew[j][COUNTROWS + 1][m] = haploArrayNew[countPed]
        [COUNTROWS][m];
      }
      else
        if(haploArray[j][COUNTROWS + 1][countNextHeteroLocus].equals(
        haploArray[countPed][COUNTROWS + 1][countNextHeteroLocus])){
          haploArrayNew[j][COUNTROWS + 1][m] = haploArrayNew[countPed]
          [COUNTROWS + 1][m];
        }
    }
  } // close the homozygous IF loop
  else // heterozygous case
    if(!founderAllele1.equals(founderAllele2)){

      // recode the progeny alleles

      if(sex.equals("M")){
        if(haploArray[j][COUNTROWS][m].equals(haploArray[countPed]
        [COUNTROWS][m])){
          haploArrayNew[j][COUNTROWS][m] = haploArrayNew[countPed]
          [COUNTROWS][m];
        }
        else
          if(haploArray[j][COUNTROWS][m].equals (haploArray[countPed]
          [COUNTROWS + 1][m])){
            haploArrayNew[j][COUNTROWS][m] = haploArrayNew[countPed]
            [COUNTROWS + 1][m];
          }
      }
      else
        if(sex.equals("F")){
          if(haploArray[j][COUNTROWS + 1][m].equals(haploArray[countPed]
          [COUNTROWS][m])){
            haploArrayNew[j][COUNTROWS + 1][m] = haploArrayNew[countPed]
            [COUNTROWS][m];
          }
          else
            if(haploArray[j][COUNTROWS + 1][m].equals(haploArray[countPed]
            [COUNTROWS + 1][m])){
              haploArrayNew[j][COUNTROWS + 1][m] = haploArrayNew[countPed]
              [COUNTROWS + 1][m];
            }
        } // close the IF sex equals F loop
    } // close the heterozygous IF loop
```

```
            } // close the marker loop
          } // close the IF loop
        } // close the progeny loop

        countPed++;

      } //close the parent loop

      // transfer the records in the progeny vector to the parents vector

      countAddedProgs = 0;
      parents.removeAllElements();

      while(countAddedProgs < progeny.size()){
        parents.addElement((String)progeny.elementAt(countAddedProgs));
        countAddedProgs++;
      }

      progeny.removeAllElements();

    } // close the generation loop

    /* DELETE OBJECTS THAT ARE NO LONGER NECESSARY */

    parents = null;
    progeny = null;
    haploArray = null;

    /-----------------------------------------------------------------------------
    / RECODE PEDIGREE FILE
    /-----------------------------------------------------------------------------*/

    int[][] pedRecode = new int[NUMANIM][NUMANIM];

    for(int i = 0; i < NUMANIM; i++){

      pedRecode[i][0] = i + 1;

      for(int n = 1; n <= 2; n++){
        if(pedArray[i][n].equals(".")){
          pedRecode[i][n] = 0;
        }
      }

      for(int j = 1; j <= 2; j++){
        for(int k = i + 1; k < NUMANIM; k++){
          if(pedArray[k][j].equals(pedArray[i][0])){
            pedRecode[k][j] = pedRecode[i][0];
          }
        }
      }

    } // close MAIN FOR loop

    // get the list of the individuals whose relationships are required
```

```
String lineRelation = relationFile.readLine();
String[] relatedAnim = new String[NUMRELATEDANIM];
int countRelation = 0;

while(lineRelation != null){
  relatedAnim[countRelation] = (String)lineRelation;
  lineRelation = relationFile.readLine();
  countRelation++;
}

relationFile.close();

int[] relatedAnimRecode = new int[NUMRELATEDANIM];

for(int i = 0; i < NUMRELATEDANIM; i++){

  countRelation = 0;

  while(!relatedAnim[i].equals(pedArray[countRelation][0]) && countRelation <
  pedArray.length){
    countRelation++;
  }

  relatedAnimRecode[i] = countRelation + 1;
} // close FOR loop

pedArray = null;
relatedAnim = null;
System.gc(); // garbage collection

/*-------------------------------------------------------------------------------
/          CALCUTATING THE NUMERATOR RELATIONSHIPS
/-------------------------------------------------------------------------------*/

System.out.println("CREATING NUMERATOR MATRIX");

double[][] numeratorMatrix = new double[pedRecode.length][pedRecode.length];
int isire, jsire, idam, jdam; double rel_iparents, rel_iandjsire = 0., rel_iandjdam = 0.;

// creates the identity matrix for the relationship between founder animals

for(int i = 0; i < NUMFOUNDER; i++){
  numeratorMatrix[i][i] = 1.0;
}

// CALCULATE THE UPPER TRIANGULAR ELEMENTS

for(int i = 0; i < pedRecode.length; i++){

  System.out.println("anim : " + i);

  for(int j = NUMFOUNDER; j < pedRecode.length; j++){
    if(i == j){
      isire = pedRecode[i][1];
      idam = pedRecode[i][2];
      rel_iparents = numeratorMatrix[isire-1][idam-1];
```

```
      numeratorMatrix[i][j] = 1.0 + (0.5*rel_iparents);
    }
    else
      if(j > i){

        // get the index of the parents of individual j from pedArray

        jsire = pedRecode[j][1];

        if(jsire == 0){
          rel_iandjsire = 0.;
        }
        else{
          if(i < (jsire-1) || i == (jsire-1)){
            rel_iandjsire = numeratorMatrix[i][jsire-1];
          }
          else{
            rel_iandjsire = numeratorMatrix[jsire-1][i];
          }
        } // close the ELSE loop

        jdam = pedRecode[j][2];

        if(jdam == 0){
          rel_iandjdam = 0.;
        }
        else{
          if(i < (jdam-1) || i == (jdam-1)){
            rel_iandjdam = numeratorMatrix[i][jdam-1];
          }
          else{
            rel_iandjdam = numeratorMatrix[jdam-1][i];
          }
        } // close the ELSE loop

        // Calculate the relationship between individual i and j

        numeratorMatrix[i][j] = 0.5*(rel_iandjsire + rel_iandjdam);
        numeratorMatrix[j][i] = numeratorMatrix[i][j];

      } // the OUTER IF loop
  } // close the INNER FOR loop
} // close the OUTER FOR loop

pedRecode = null;
System.gc();

/*-------------------------------------------------------------------------------------------------
/
/          CALCUTATING THE GENETIC MARKER BASED RELATIONSHIPS
/          AMONG ANIMALS OF INTEREST
/
/------------------------------------------------------------------------------------------------*/
```

```
/*--------------------------------------------------------------------------
/        CREATE AN ARRAY TO STORE INTERVALS
/        BETWEEN MARKERS
/-------------------------------------------------------------------------*/


BufferedReader intervalFile = new BufferedReader(new FileReader(args[8]));
double[] markerInterval = new double[NUMMARKERS-1];
String lineInterval = intervalFile.readLine();
int countInterval = 0;
double mapLength = 0.;

while(lineInterval != null){
  markerInterval[countInterval] = Double.parseDouble(lineInterval);
  mapLength = mapLength + markerInterval[countInterval];
  lineInterval = intervalFile.readLine();
  countInterval++;
}

intervalFile.close();

double relationMatrix;
double[][] pedMarkerRelation = new double[NUMRELATEDANIM][NUMRELATEDANIM];
double halfSegment1, halfSegment2, intervalSegInCommonPP,intervalSegInCommonPM,
double intervalSegInCommonM, intervalSegInCommonMM, propPP, propPM, propMP;
double propMM;
final double bovineGenomeSize = 3000; // cM
int COUNTNONZERO = 0;

System.out.println("CREATING MARKER RELATIONSHIP MATRIX");

for(int i = 0; i < NUMRELATEDANIM; i++){

  System.out.println("Anim  " + i);

  for(int j = 0; j < NUMRELATEDANIM; j++){
    intervalSegInCommonPP = 0.;
    intervalSegInCommonPM = 0.;
    intervalSegInCommonMP = 0.;
    intervalSegInCommonMM = 0.;
   propPP = 0.; propPM = 0.; propMM = 0.; propMP = 0.;

    for(int m = 0; m < NUMMARKERS - 1; m++){

     halfSegment1 = 0.; halfSegment2 = 0.;

     /*-----------------------------------------------------------------------------------------------------
     / The interval in common between the paternal segments of individual i and j (i.e. between
     / markers m and m+1)
     /----------------------------------------------------------------------------------------------------*/

     if(haploArrayNew[relatedAnimRecode[i]-1][COUNTROWS][m].equals(haploArrayNew
     [relatedAnimRecode[j]-1][COUNTROWS][m])){
       halfSegment1 = 0.5*markerInterval[m];
     }
```

```
if(haploArrayNew[relatedAnimRecode[i]-1][COUNTROWS][m+1].equals(haploArrayNew
[relatedAnimRecode[j]-1][COUNTROWS][m+1])){
  halfSegment2 = 0.5*markerInterval[m];
}

intervalSegInCommonPP = intervalSegInCommonPP + halfSegment1 + halfSegment2;

/*-----------------------------------------------------------------------------------------------------------
/ The interval in common between the paternal segment of individual i and the maternal
/ segment of individual j (i.e. between markers m and m+1)
/----------------------------------------------------------------------------------------------------------*/

halfSegment1 = 0.; halfSegment2 = 0.;

if(haploArrayNew[relatedAnimRecode[i]-1][COUNTROWS][m].equals(haploArrayNew
[relatedAnimRecode[j]-1][COUNTROWS + 1][m])){
   halfSegment1 = 0.5*markerInterval[m];
}

if(haploArrayNew[relatedAnimRecode[i]-1][COUNTROWS][m+1].equals(haploArrayNew
[relatedAnimRecode[j]-1][COUNTROWS + 1][m+1])){
  halfSegment2 = 0.5*markerInterval[m];
}

intervalSegInCommonPM = intervalSegInCommonPM + halfSegment1 + halfSegment2;

/*-----------------------------------------------------------------------------------------------------------
/ The interval in common between the maternal segment of individual i and the paternal
/ segment of individual j (i.e. between markers m and m+1)
/----------------------------------------------------------------------------------------------------------*/

halfSegment1 = 0.; halfSegment2 = 0.;

if(haploArrayNew[relatedAnimRecode[i]-1][COUNTROWS + 1][m].equals(haploArrayNew
[relatedAnimRecode[j]-1][COUNTROWS + 1][m])){
  halfSegment1 = 0.5*markerInterval[m];
}

if(haploArrayNew[relatedAnimRecode[i]-1][COUNTROWS + 1][m+1].equals
(haploArrayNew[relatedAnimRecode[j]-1][COUNTROWS + 1][m+1])){
  halfSegment2 = 0.5*markerInterval[m];
}

intervalSegInCommonMP = intervalSegInCommonMP + halfSegment1 + halfSegment2;

/*-----------------------------------------------------------------------------------------------------------
/ The interval in common between the maternal segments of individual i and j (i.e. between
/ markers m and m+1)
/----------------------------------------------------------------------------------------------------------*/

halfSegment1 = 0.; halfSegment2 = 0.;

if(haploArrayNew[relatedAnimRecode[i]-1][COUNTROWS + 1][m].equals(haploArrayNew
[relatedAnimRecode[j]-1][COUNTROWS][m])){
```

```
        halfSegment1 = 0.5*markerInterval[m];
      }

      if(haploArrayNew[relatedAnimRecode[i]-1][COUNTROWS + 1][m+1].equals(
       haploArrayNew[relatedAnimRecode[j]-1][COUNTROWS][m+1])){
        halfSegment2 = 0.5*markerInterval[m];
      }

      intervalSegInCommonMM = intervalSegInCommonMM + halfSegment1 + halfSegment2;


    } // close immediate for loop

    propPP = intervalSegInCommonPP/mapLength;
    propPM = intervalSegInCommonPM/mapLength;
    propMP = intervalSegInCommonMP/mapLength;
    propMM = intervalSegInCommonMM/mapLength;
    relationMatrix = (propPP + propPM + propMP + propMM)/2.0;

    // calculate the pedigree and marker-based relationship

    pedMarkerRelation[i][j] = ((numeratorMatrix[relatedAnimRecode[i]-1][relatedAnimRecode
    [j]-1]*(bovineGenomeSize - mapLength)) + (relationMatrix*mapLength))/bovineGenomeSize;

    if(pedMarkerRelation[i][j] != 0){
      COUNTNONZERO++;
    }

  } // close the FOR j loop
} // close the main FOR i loop

outFile.println("SPARSE " + NUMRELATEDANIM + " " + NUMRELATEDANIM + " " +
COUNTNONZERO);

/*-------------------------------------------------------------------------------------
/ PRINT THE SPARSE STORED MARKER RELATIONSHIP MATRIX
/-------------------------------------------------------------------------------------*/

for(int i = 0; i < NUMRELATEDANIM; i++){
  for(int j = 0; j < NUMRELATEDANIM; j++){
    if(pedMarkerRelation[i][j] != 0){
      outFile.println((i + 1) + " " + (j + 1) + " " + pedMarkerRelation[i][j]);
    }
  }
}

/*-------------------------------------------------------------------------------------
/ PRINT THE SPARSE STORED NUMERATOR RELATIONSHIP MATRIX
/-------------------------------------------------------------------------------------*/

/*-----------------------------------------------------------------------------
/ DETERMINE THE NUMBER OF NONZERO
/ ELEMENTS IN THE NUMERATOR RELATIONSHIP MATRIX
/-----------------------------------------------------------------------------*/

int countNonZeroNumMatrix = 0;
```

```
        for(int i = 0; i < NUMRELATEDANIM; i++){
          for(int j = 0; j < NUMRELATEDANIM; j++){
            if(numeratorMatrix[relatedAnimRecode[i]-1][relatedAnimRecode[j]-1] != 0){
              countNonZeroNumMatrix++;
            }
          }
        }

        outAmatrix.println("SPARSE" + " " + NUMRELATEDANIM + " " + NUMRELATEDANIM + "
        " + countNonZeroNumMatrix);

        for(int i = 0; i < NUMRELATEDANIM; i++){
          for(int j = 0; j < NUMRELATEDANIM; j++){
            if(numeratorMatrix[relatedAnimRecode[i]-1][relatedAnimRecode[j]-1] != 0){
              outAmatrix.println((i + 1) + " " + (j + 1) + " " + numeratorMatrix[relatedAnimRecode
              [i]-1][relatedAnimRecode[j]-1]);
            }
          }
        }

        outFile.close();
        outAmatrix.close();

        /*-------------------------------
        /          END !!!!!!!
        /--------------------------------*/

        } // close MAIN method
} // close PUBLIC class
```

APPENDIX VI

THE SOURCE CODE FOR A JAVA PROGRAM TO ASSIGN HAPLOTYPES FROM GENOTYPIC INFORMATION AT GENETIC MARKER LOCI

```
/*****************************************************************************
/
/       CLASS       :       AssigLinkageHaplotype.java
/
/       AUTHOR      :       Maiwashe Azwihangwisi
/
/       DATE        :       Created 11/13/03
/
/       PURPOSE     :
/
/                   Primary - assign the parent and progeny haplotypes based on the parent and
/                   progeny genotypes
/
/                   Secondary - detect erroneous genotypes based on the incompatibility
/                   between the parent and progeny genotypes
/
/       SYNTAX      :
/                   Compile command : javac AssignLinkageHaplotype.java
/
/                   Execute command : java AssignLinkageHaplotype inputFile1 inputFile2
/                   outputFile numRec1 numRec2 numMarkers
/
/                   ~ inputFile1 - file containing genotyped sires
/                   ~ inputFile2 - file containg all the genotyped and ungenotyped animals,
/                   including the sires in inputFile1 i.e. a pedigree for each individual should be
/                   included (founder animals should be assigned unknown parents (".") and
/                   missing genotypes (0)
/                   ~ outputFile - file containing the paternal and maternal haplotypes
/                   ~ numRec1 - number of records in inputFile1
/                   ~ numRec2 - number of records in inputFile2
/                   ~ numMarkers - number of markers that have been genotyped
/
/                   An example of the structure of the input files for animals genotyped at two
/                   marker loci :
/
/                   animID    sireID    damID    MRK1_allele1    MRK1_allele2
/                     3         1         2          20              21
/
/                   WARNING: Do not include the file header
/
/
/       NOTES :
/
/       This program is designed to assign haplotypes for data with marker genotypes observed on one
/       parent (sire/dam) and several progeny. However, few missing parent and progeny genotypes are
/       accommodated. The procedure is based on the logic presented by Nejati-Javaremi (1995). The
/       program proceeds as follows:
/
/       1. The progeny haplotypes are assigned first using the parent and the progeny genotypes
/
/       2. Parent haplotypes are assigned using progeny haplotypes derived in step 1
/       i.e. the two most common haplotypes are considered to be the parent haplotypes
/
/       3. Unresolved loci in the progeny paternal haplotypes are resolved based on the parent haplotypes
/       constructed in step 2. The closest heterozygous locus (on the parent haplotype) to the right or left
/       of the unresolved locus (in the progeny) is used to infer the allele at the missing locus in the
```

```
/          progeny's paternal haplotype. The assumption here is that no cross-over event occurred between
/          the marker to be resolved and the closest heterozygous marker. The accuracy of the haplotyping
/          procedure used here depends to a large extent on the markers being in close proximity to each
/          other for uninformative markers. When the proportion of uninformative markers is low, this
/          procedure is reasonably accurate.
/
/          4. Infer the haplotypes for the other parent (e.g., dam) with unknown genotypes
/
/*****************************************************************************/

import java.util.*;
import java.io.*;
import java.lang.*;
import java.text.*;

public class AssignLinkageHaplotype{
     public static void main(String[] args) throws IOException{

          // access parent file

          FileReader frs = new FileReader(args[0]);
          BufferedReader brs = new BufferedReader(frs);

          // access the file with all animals

          FileReader fri = new FileReader(args[1]);
          BufferedReader bri = new BufferedReader(fri);

          // link to output file

          FileWriter fout = new FileWriter(args[2]);
          BufferedWriter outF = new BufferedWriter(fout);
          PrintWriter outFile = new PrintWriter(outF);


          final int numSire = Integer.parseInt(args[3]);
          final int numAnim = Integer.parseInt(args[4]);
          final int numMarkers = Integer.parseInt(args[5]);

          // an array to store sire records

          String[] inputSire = new String[numSire];

          // read sire records and store them in inputSire array

          String lineSire = brs.readLine();
          int count = 0;

          while(lineSire != null){
            inputSire[count] = lineSire;
            lineSire = brs.readLine();
            count++;
          }

          brs.close();
```

```
// array to store anim records

String[] inputAnim = new String[numAnim];

// read anim records and store in inputAnim array

String lineAnim = bri.readLine();
count = 0;

while(lineAnim != null){
  inputAnim[count] = lineAnim;
  lineAnim = bri.readLine();
   count++;
 }

bri.close();


/*-----------------------------------------------------------------
/
/          STEP 1:
/
/          ASSIGN PROGENY HAPLOTYPES
/
/-----------------------------------------------------------------*/

int animAllele1, animAllele2, sireAllele1, sireAllele2, progAllele, alleleMissing1, alleleMissing2;
int progNum, int countMissing;
int[][] progAlleleSireArray;
int[][] progAlleleDamArray;
int countVec;
int[][] alleleArray;

String anim, animSire, animDam, sire, anim1, anim1Sire, anim1Dam, haplotype1, haplotype2;
String sireSireHaplo, sireDamHaplo, prog, progSire, sireSireMissing = " ";
String sireMissingRecord = " ", sireSireInterest = " ";
String[] progSireHaplo, progDamHaplo, progIDvec;

Vector alleleVec = new Vector();
Vector siresAlreadyStored = new Vector(20000);

StringTokenizer tokenProg, tokenProgNo, tokenSire, tokenAnim, tokenMissingSireGeno;

System.out.println("Warning : Incompatible progeny alleles are set to zero");

for(int s = 0; s < numSire; s++){

  System.out.println();
  System.out.println("Sire Family : " + (s + 1));

  tokenSire = new StringTokenizer(inputSire[s]);
  sire = tokenSire.nextToken();
  sireSireInterest = tokenSire.nextToken();
  tokenSire.nextToken();
  sireAllele1 = Integer.parseInt(tokenSire.nextToken());
  sireAllele2 = Integer.parseInt(tokenSire.nextToken());
```

```
// determine the number of progeny for a sire

progNum = 0;

for(int i = 0; i < numAnim; i++){
  tokenAnim = new StringTokenizer(inputAnim[i]);
  anim = tokenAnim.nextToken();
  animSire = tokenAnim.nextToken();

  if(sire.equals(animSire)){
    progNum++;
  }
}

progAlleleSireArray = new int[progNum][numMarkers];
progAlleleDamArray = new int[progNum][numMarkers];
progIDvec = new String[progNum];
progSireHaplo = new String[progNum];
progDamHaplo = new String[progNum];
progNum = 0;  countMissing = 0; countVec = 0;

// get all the sire 's progeny and infer their haplotypes, marker at a time

for(int p = 0; p < numAnim; p++){

  tokenAnim = new StringTokenizer(inputAnim[p]);
  anim = tokenAnim.nextToken(); animSire = tokenAnim.nextToken();
  animDam = tokenAnim.nextToken();

  if(sire.equals(animSire)){
    progIDvec[countVec] = anim;
    countVec++;
    progSireHaplo[progNum] = anim + " " + animSire + " " + animDam;
    progDamHaplo[progNum] = anim + " " + animSire + " " + animDam;

    for(int m = 0; m < numMarkers; m++){

      animAllele1 = Integer.parseInt(tokenAnim.nextToken());
      animAllele2 = Integer.parseInt(tokenAnim.nextToken());

      /*-----------------------------------------------------------------------------------------------------
      / INFER THE MISSING SIRE GENOTYPE BY CALCULATING THE FREQUENCY
      /  OF EACH ALLELE. USING THE CALCULATED FREQUENCY, THE TWO MOST
      /  FREQUENT ALLELES ARE BELIEVED TO HAVE ORIGINATED FROM THE
      /  SIRE BECAUSE THE PROGENY ARE HALF-SIBS
      /-----------------------------------------------------------------------------------------------------*/

      int allele1SireNoGeno = 0, allele2SireNoGeno = 0, countAlleleVec = 0;
      String alleleString1, alleleString2;

      if(sireAllele1 == 0 && sireAllele2 == 0){

        // determine the total number of alleles for the current locus based on progeny
        // genotypes
```

```
// create and store the alleles in the array

for(int progNo = 0; progNo < numAnim; progNo++){

  tokenProgNo = new StringTokenizer(inputAnim[progNo]);
  anim1 = tokenProgNo.nextToken(); anim1Sire = tokenProgNo.nextToken();
  anim1Dam = tokenProgNo.nextToken();

  if(sire.equals(anim1Sire)){
    countAlleleVec = 0;

    while(countAlleleVec < m){
     tokenProgNo.nextToken(); tokenProgNo.nextToken(); countAlleleVec++;
    }

    allele1SireNoGeno = Integer.parseInt(tokenProgNo.nextToken());
    allele2SireNoGeno = Integer.parseInt(tokenProgNo.nextToken());
    alleleString1 = allele1SireNoGeno + ""; alleleString2 = allele2SireNoGeno + "";

    if(!alleleVec.contains(alleleString1) && !alleleString1.equals("0")){
      alleleVec.addElement(alleleString1);
    }

    if(!alleleVec.contains(alleleString2) && !alleleString2.equals("0")){
      alleleVec.addElement(alleleString2);
    }
  } // close the if loop
} // close the for loop

alleleArray = new int[alleleVec.size()][2];

for(int progArrayCount = 0; progArrayCount < alleleVec.size(); progArrayCount++){
  alleleArray[progArrayCount][0] = Integer.parseInt((String)
  alleleVec.elementAt(progArrayCount));
}

// determine the frequency of each allele

for(int progNo = 0; progNo < numAnim; progNo++){

  tokenProgNo = new StringTokenizer(inputAnim[progNo]);
  anim1 = tokenProgNo.nextToken(); anim1Sire = tokenProgNo.nextToken();
  anim1Dam = tokenProgNo.nextToken();

  if(sire.equals(anim1Sire)){
    countAlleleVec = 0;

    while(countAlleleVec < m){
      tokenProgNo.nextToken(); tokenProgNo.nextToken(); countAlleleVec++;
    }

    allele1SireNoGeno = Integer.parseInt(tokenProgNo.nextToken());
    allele2SireNoGeno = Integer.parseInt(tokenProgNo.nextToken());

    for(int progFreq = 0; progFreq < alleleVec.size(); progFreq++){
```

```
          if(allele1SireNoGeno == alleleArray[progFreq][0]){
            alleleArray[progFreq][1] = alleleArray[progFreq][1] + 1;
          }

          if(allele2SireNoGeno == alleleArray[progFreq][0]){
            alleleArray[progFreq][1] = alleleArray[progFreq][1] + 1;
          }

        } // close the immediate for loop
      } // close the if loop
    } // close the main for loop

    // identify the two most common alleles and assign them to the sire

    // get the most common allele

    int MAXCOUNT = 0, countHighest = 1;

    for(int i = 0; i < alleleArray.length(); i++){
      if(alleleArray[i][1] > MAXCOUNT){
        MAXCOUNT = alleleArray[i][1];
      }
    }

    for(int countmax = MAXCOUNT; countmax > 0; countmax--){
      for(int alleleFreq = 0; alleleFreq < alleleVec.size(); alleleFreq++){
        if(countmax == alleleArray[alleleFreq][1] && countHighest == 1){
          sireAllele1 = alleleArray[alleleFreq][0]; countHighest++;
        }
        else
          if(countmax == alleleArray[alleleFreq][1] && (countHighest > 1 &&
          countHighest <= 2)){
            sireAllele2 = alleleArray[alleleFreq][0];
            countHighest++;
          }
      } // close the inner for loop
    } //close the main for loop
  } // close the main if loop

  alleleVec.removeAllElements();  // empty the allele vector

  /*-------------------------------------------------------------------------
  / DETERMINE THE PHASE OF THE TWO ALLELES
  /------------------------------------------------------------------------*/

  // 1st allele comes from the sire and second from dam

  if((animAllele1 != 0 && animAllele2 != 0) && (animAllele1 == sireAllele1 ||
  animAllele1 == sireAllele2) && (animAllele2 != sireAllele1 && animAllele2 !=
  sireAllele2)){
    progSireHaplo[progNum] = progSireHaplo[progNum] + " " + animAllele1 + " ";
    progDamHaplo[progNum] = progDamHaplo[progNum] + " " + animAllele2 + " ";
  }
  else    // 2nd allele comes from sire and 1st from dam
    if((animAllele1 != 0 && animAllele2 != 0) && (animAllele1 != sireAllele1
     && animAllele1 != sireAllele2) && (animAllele2 == sireAllele1 || animAllele2 ==
```

```
sireAllele2)){
  progSireHaplo[progNum] = progSireHaplo[progNum] + " " + animAllele2 + " ";
  progDamHaplo[progNum] = progDamHaplo[progNum] + " " + animAllele1 + " ";
}
else      // either of the two alleles may have originated from the sire or the dam
  if((animAllele1 != 0 && animAllele2 != 0) && (animAllele1 == sireAllele1 ||
  animAllele1 == sireAllele2) && (animAllele2 == sireAllele1 || animAllele2 ==
  sireAllele2)){
    if(animAllele1 == animAllele2){ // homozygous case
      progSireHaplo[progNum] = progSireHaplo[progNum] + " " + animAllele1 + " ";
      progDamHaplo[progNum] = progDamHaplo[progNum] + " "+ animAllele2 + " ";
     }
    else // heterozygous case - 0 refers to not assigned or unresolved
      if(animAllele1 != animAllele2){
        progSireHaplo[progNum] = progSireHaplo[progNum] + " " + 0 + " ";
        progDamHaplo[progNum] = progDamHaplo[progNum] + " " + 0 + " ";
      }
    }
    else  // erronous genotype i.e. none of the progeny alleles came from the sire
      if((sireAllele1 != 0 && sireAllele2 != 0) && (animAllele1 != 0 &&
      animAllele2 != 0)){
        if((animAllele1 != sireAllele1 && animAllele1 != sireAllele2) && (animAllele2
        != sireAllele1 && animAllele2 != sireAllele2)){
          System.out.println("Anim " + anim + " " + animAllele1 + " and  " +
          animAllele2 + " | sire "+ sire + " " + sireAllele1 + " and " + sireAllele2
          + " " + " at marker " + (m + 1));
          progSireHaplo[progNum] = progSireHaplo[progNum]+ " " + 0 + " ";
          progDamHaplo[progNum] = progDamHaplo[progNum]+ " " + 0 + " ";
        }
      } // close the if loop
      else  // the progeny and sire genotypes are missing
        if(animAllele1 == 0 || animAllele2 == 0){
          progSireHaplo[progNum] = progSireHaplo[progNum]+ " " + 0 + " ";
          progDamHaplo[progNum] = progDamHaplo[progNum] + " " + 0 + " ";
        }

        // change the value of the sire alleles

        if(m < (numMarkers - 1)){
          sireAllele1 = Integer.parseInt(tokenSire.nextToken());
          sireAllele2 = Integer.parseInt(tokenSire.nextToken());
        }
  } // close the marker for loop

  progNum++;

} // close the if loop

tokenSire = new StringTokenizer(inputSire[s]);
sire = tokenSire.nextToken(); tokenSire.nextToken(); tokenSire.nextToken();
sireAllele1 = Integer.parseInt(tokenSire.nextToken());
sireAllele2 = Integer.parseInt(tokenSire.nextToken());

} // close the anim for loop

// Store the progeny paternal and maternal haplotype into an array
```

```
StringTokenizer tokenProgForSire, tokenProgForDam;

int countTokenParents = 0;

for(int e = 0; e < progNum; e++){

  tokenProgForSire = new StringTokenizer(progSireHaplo[e]);
   tokenProgForDam = new StringTokenizer(progDamHaplo[e]);
   countTokenParents = 0;

   while(countTokenParents < 3){
     tokenProgForSire.nextToken(); tokenProgForDam.nextToken(); countTokenParents++;
   }

   for(int f = 0; f < numMarkers; f++){
     progAlleleSireArray[e][f] = Integer.parseInt(tokenProgForSire.nextToken());
     progAlleleDamArray[e][f] = Integer.parseInt(tokenProgForDam.nextToken());
  }
}

/*-------------------------------------------------------------------------------------------------------------
/
/          STEP 2:
/
/ INFER THE SIRE HAPLOTYPES BASED ON PROGENY HAPLOTYPES.
/ WITHOUT THE SIRE'S PARENTAL GENOTYPE THE PHASE OF THE SIRE
/ HAPLOTYPE IS UNCERTAIN. THUS, THE USE OF THE TERMS SIRE PATERNAL AND
/ MATERNAL HAPLOTYPE IS ABITRARY
/-------------------------------------------------------------------------------------------------------------*/


sireSireHaplo = " "; sireDamHaplo = " ";
int firstNonZeroAllele = 10, countFirst, countSecond, countParentalComb1, countParentalComb2,
int countMaternalComb, countMaternalComb2, countRecomb1, countRecomb2, countSireFirst;
int countDamFirst, countSireSecond, countDamSecond, countToken, countRemainHaplo;
int countPrint, countInitialise1, countInitialise2, parentalComb1, parentalComb2;
int previousAllele1, previousAllele2, recomb, recomb1, recomb2, alleleSirePaternal = 0;
int alleleDamPaternal = 0, firstAllele = 0, secondAllele = 0, alleleStatus = 0;
int countAlleleAssign = 0;
int[] sireDadHaplo,  sireMomHaplo;

String tempHaplo, sireSireHaploUpDate = "",  sireDamHaploUpDate = "";
String UpDatedProgSireAllele, sireOriginal = " ", sireRecomb = " ", damOriginal = " ";
String damRecomb = " ", lastNoneZero, animLastNoneZero;
StringTokenizer tokenUpDateProg1, tokenUpDateSireHaplo;
StringTokenizer tokenUpDateDamHaplo, tokenLociUpDateProg;

for(int j = 0; j < numMarkers; j++){

  countFirst = 0; countSecond = 0; countSireFirst = 0; countDamFirst = 0; countSireSecond = 0;
  countDamSecond = 0; countParentalComb1 = 0; countParentalComb2 = 0;
  countMaternalComb1 = 0; countMaternalComb2 = 0; countRecomb1 = 0; countRecomb2 = 0;
  countInitialise1 = 0; countInitialise2 = 0;      countAlleleAssign = 0; lastNoneZero = " ";
   animLastNoneZero = " "; firstAllele = 0; secondAllele = 0; parentalComb1 = 0;
```

```
parentalComb2 = 0; previousAllele1 = 0; previousAllele2 = 0; recomb1 = 0; recomb2 = 0;
countPrint = 0;

if(j == 0){  // assign the sire's paternal and maternal allele for the first marker

  for(int t = 0; t < progNum; t++){
    tokenProg = new StringTokenizer(progSireHaplo[t]);
    tokenProg.nextToken(); tokenProg.nextToken(); tokenProg.nextToken(); countToken = 0;
    tempHaplo = " ";
    progAllele = Integer.parseInt(tokenProg.nextToken());

    if(progAllele != 0){
      if(countFirst == 0){
        firstNonZeroAllele = progAllele;
        sireSireHaplo = sireSireHaplo + firstNonZeroAllele;
        sireSireHaploUpDate = sireSireHaploUpDate + " " + firstNonZeroAllele;
        countFirst++;
      }
      else
        if(firstNonZeroAllele != progAllele && countSecond == 0){
          sireDamHaplo = sireDamHaplo + progAllele;
          sireDamHaploUpDate = sireDamHaploUpDate + " " + progAllele;
          countSecond++;
        }
        else
          if(sireDamHaplo.equals(" ") && t == (progNum -1)){
            // the sire is homozygous for the current marker

            sireDamHaplo = sireDamHaplo + progAllele;
            sireDamHaploUpDate = sireDamHaploUpDate + " " + progAllele;
          }
    } // close inner if loop
  } //close the progeny for loop
} // close the if loop
else{ // assign the sire haplotypes for the rest of the markers i.e. j > 0

  // determine whether the sire was homozygous or heterozygous for the current marker

  for(int c = 0; c < progNum; c++){

    tokenProg = new StringTokenizer(progSireHaplo[c]);
    tokenProg.nextToken();  tokenProg.nextToken(); tokenProg.nextToken();
    countToken = 0;

    while(countToken < j){
      tokenProg.nextToken(); countToken++;
    }

    progAllele = Integer.parseInt(tokenProg.nextToken());

    if(progAllele != 0){
      if(countAlleleAssign == 0){
        firstAllele = progAllele; countAlleleAssign++;
      }
      else
        if(firstAllele != progAllele){
```

```
        secondAllele = progAllele;
      }
  } // close if progAllele != 0 loop
} //close the for c progNum for loop

if(secondAllele == 0){
  alleleStatus = 1;  // 1 represents homozygous
}
else{
  alleleStatus = 2;  // 2 represents heterozygous
} // close the else loop

// START CHECKING THE PROGENY HAPLOTYPES

for(int t = 0; t < progNum; t++){
  tokenProg = new StringTokenizer(progSireHaplo[t]);
  animLastNoneZero = tokenProg.nextToken();
  tokenProg.nextToken(); tokenProg.nextToken();
  countToken = 0; tempHaplo = " ";

  // create an array to store the progeny paternal haplotype

  int[] progAllele1 = new int[j + 1];

  while(countToken <= j){
    progAllele1[countToken] = Integer.parseInt(tokenProg.nextToken());
    countToken++;
  }

  // create tokens to access previous sire paternal and maternal allele

  StringTokenizer tokenSireParental = new StringTokenizer(sireSireHaploUpDate);
  StringTokenizer tokenSireMaternal = new StringTokenizer(sireDamHaploUpDate);

  // search for a previous marker that is heterozgous to assign the sire alleles

  int markerCount = 0;

  // create an array to store the previous sire paternal and maternal alleles

  int[] sirePaternalArray = new int[j]; int[] sireMaternalArray = new int[j];

  while(markerCount < j){
    sirePaternalArray[markerCount] = Integer.parseInt(tokenSireParental.nextToken());
    sireMaternalArray[markerCount] = Integer.parseInt(tokenSireMaternal.nextToken());
    markerCount++;
  }

  int countReverse = j - 1;

  if(progAllele1[j] != 0){

    // moving in a backward direction until a heterozygous marker is found

    while(countReverse >= 0 && (sirePaternalArray[countReverse] ==
    sireMaternalArray[countReverse])){
```

```
  countReverse--;
}

// the paternal and maternal alleles are the same from the current marker to the first
// marker

if(countReverse < 0){   // a heterozygous marker not found

  // assign paternal allele
  if(countSireFirst == 0){
    sireSireHaplo = sireSireHaplo + progAllele1[j];
    sireSireHaploUpDate = sireSireHaploUpDate + " " + progAllele1[j];
    parentalComb1 = progAllele1[j]; countSireFirst++;
  }
  else   // assign maternal allele - for a heterozygous case
    if(countDamFirst == 0 && parentalComb1 != progAllele1[j]){
      sireDamHaplo = sireDamHaplo + progAllele1[j];
      sireDamHaploUpDate = sireDamHaploUpDate + " " + progAllele1[j];
      countDamFirst++;
    }
    else  // assign maternal allele - for a homozygous case
      if(countDamFirst == 0 && (t == progNum - 1)){
        sireDamHaplo = sireDamHaplo + progAllele1[j];
        sireDamHaploUpDate = sireDamHaploUpDate + " "+ progAllele1[j];
        countDamFirst++;
      } // close the last if loop
} // close the countReverse < 0 loop
else{   // a heterozygous marker found to the left of the current marker
   if(countReverse >= 0 && (alleleStatus == 2 && progAllele1[countReverse] != 0)){

     /*-------------------------------------------------------------------------------------------
     /
     / DECISION ON THE SIRE PATERNAL AND MATERNAL HAPLOTYPE WAS
     / BASED ON THE FREQUENCY OF OCCURENCE OF THE PROGENY
     / PATERNAL ALLELE WITH THE SIRE PATERNAL AND MATERNAL
     / HAPLOTYPE ALREADY CREATED FOR THE PREVIOUS MARKERS
     / EXCEPT FOR THE CURRENT MARKER LOCUS TO BE ASSIGNED
     /  e.g. if the SIRE PATERNAL haplotype for previous markers is 12 and
     /  the SIRE MATERNAL haplotype is 23 with the PROGENY PATERNAL alleles
     / being 2 and 3 for the current marker. Say PROGENY PATERNAL allele 2
     / occurred 10 and 2 times with the SIRE PATERNAL and MATERNAL
     / HAPLOTYPES, respectively. Then, the UPDATED SIRE PATERNAL AND
     / MATERNAL HAPLOTYPES are 122 and 233, respectively
     /-------------------------------------------------------------------------------------------*/

     // the current marker is heterozygous

     // sire paternal haplotype based decision

     if(countInitialise1 == 0){
       previousAllele1 = sirePaternalArray[countReverse];
       parentalComb1 = sirePaternalArray[countReverse] + firstAllele;
       countInitialise1++;
     }

     if(parentalComb1 == progAllele1[countReverse] + progAllele1[j]){
```

```
                countParentalComb1++;
              }
              else
                if((progAllele1[countReverse] == previousAllele1) && (progAllele1[j] ==
                secondAllele)){
                  countParentalComb2++;
                }

              // sire maternal haplotype based decision
              if(countInitialise2 == 0){
                previousAllele2 = sireMaternalArray[countReverse];
                parentalComb2 = sireMaternalArray[countReverse] + firstAllele;
                countInitialise2++;
              }

              if(parentalComb2 == progAllele1[countReverse] + progAllele1[j]){
                countMaternalComb1++;
              }
              else
                if((progAllele1[countReverse] == previousAllele2) && (progAllele1[j] ==
                secondAllele)){
                  countMaternalComb2++;
                }
            } // close the if countReverse >= 0 loop
            else
              if(countReverse >= 0 && (alleleStatus == 1 && countPrint == 0)){

                // the current marker is homozygous - this loop is processed once

                // assign the sire paternal and maternal haplotype

                sireSireHaplo = sireSireHaplo + progAllele1[j];
                sireSireHaploUpDate = sireSireHaploUpDate + " "+ progAllele1[j];
                sireDamHaplo = sireDamHaplo + progAllele1[j];
                sireDamHaploUpDate = sireDamHaploUpDate + " "+ progAllele1[j];
                countPrint++;
              } // close the if alleleStatus loop
          } // close the else loop
        } // close the progAllele[j] != 0 if loop
    } // close the progeny count for loop
} // close the else loop

// check if the sire haplotype has been updated already

if(alleleStatus == 2){ // upddate the sire haplotypes for a heterozygous loci

  // decision based on sire paternal haplotype

  if(countParentalComb1 != 0 || countParentalComb2 != 0){

    if(countParentalComb1 > countParentalComb2){
      sireSireHaplo = sireSireHaplo + firstAllele;
      sireSireHaploUpDate = sireSireHaploUpDate + " " + firstAllele;
      sireDamHaplo = sireDamHaplo + secondAllele;
      sireDamHaploUpDate = sireDamHaploUpDate + " " + secondAllele;
    }
```

```
    else
      if(countParentalComb1 < countParentalComb2){
        sireSireHaplo = sireSireHaplo + secondAllele;
        sireSireHaploUpDate = sireSireHaploUpDate + " " + secondAllele;
        sireDamHaplo = sireDamHaplo + firstAllele;
        sireDamHaploUpDate = sireDamHaploUpDate + " " + firstAllele;
      }
      else
        if(countParentalComb1 == countParentalComb2){
          sireSireHaplo = sireSireHaplo + firstAllele;
          sireSireHaploUpDate = sireSireHaploUpDate + " " + firstAllele;
          sireDamHaplo = sireDamHaplo + secondAllele;
          sireDamHaploUpDate = sireDamHaploUpDate + " " + secondAllele;
        }
  }
  else    // decision based on sire maternal haplotype
    if((countParentalComb1 == 0 && countParentalComb2 == 0) && (countMaternalComb1 != 0
    || countMaternalComb2 != 0)){
      if(countMaternalComb1 > countMaternalComb2){
        sireSireHaplo = sireSireHaplo + secondAllele;
        sireSireHaploUpDate = sireSireHaploUpDate + " " + secondAllele;
        sireDamHaplo = sireDamHaplo + firstAllele;
        sireDamHaploUpDate = sireDamHaploUpDate + " " + firstAllele;
      }
      else
        if(countMaternalComb1 < countMaternalComb2){
          sireSireHaplo = sireSireHaplo + firstAllele;
          sireSireHaploUpDate = sireSireHaploUpDate + " "+ firstAllele;
          sireDamHaplo = sireDamHaplo + secondAllele;
          sireDamHaploUpDate = sireDamHaploUpDate + " "+ secondAllele;
        }
        else
          if(countParentalComb1 == countParentalComb2){
            sireSireHaplo = sireSireHaplo + firstAllele;
            sireSireHaploUpDate = sireSireHaploUpDate + " "+ firstAllele;
            sireDamHaplo = sireDamHaplo + secondAllele;
            sireDamHaploUpDate = sireDamHaploUpDate + " "+ secondAllele;
          }
    } // close the first if after the esle loop
  } // close the if loop
} // close numMarker for loop


/*-------------------------------------------------------------------------------------------------
/
/         STEP 3:
/
/                   RESOLVE THE UNRESOLVED AND MISSING
/                   PROGENY HAPLOTYPES
/-------------------------------------------------------------------------------------------------*/

// store the sire paternal and maternal haplotype in an array

sireDadHaplo = new int[numMarkers];
sireMomHaplo = new int[numMarkers];
```

```
StringTokenizer tokenDad = new StringTokenizer(sireSireHaploUpDate);
StringTokenizer tokenMom = new StringTokenizer(sireDamHaploUpDate);

for(int r = 0; r < numMarkers; r++){
  sireDadHaplo[r] = Integer.parseInt(tokenDad.nextToken());
  sireMomHaplo[r] = Integer.parseInt(tokenMom.nextToken());
}


int countSteps = 0, pos = 0;

for(int q = 0; q < progNum; q++){

  for(int v = 0; v < numMarkers; v++){   // update the progeny haplotype marker at a time

    /*-------------------------------------------------------------------------------------------------------
    /
    / TWO CASES EXISTS HERE.
    /
    / THE SIRE HAPLOTYPES ARE HOMOZYGOUS OR HETEROZYGOUS AT THE
    / LOCUS WHERE THE PROGENY HAS A MISSING GENOTYPE.
    /-------------------------------------------------------------------------------------------------------*/

    if(progAlleleSireArray[q][v] == 0){
      if(sireDadHaplo[v] != sireMomHaplo[v]){  // a hereozygous case

        /*-------------------------------------------------------------------------------------------------
        / search for the none-zero progeny allele to the right of the missing progeny allele
        /-------------------------------------------------------------------------------------------------*/

        countSteps = 0; pos = 0;

        for(int w = v + 1; w < numMarkers; w++){
          if(progAlleleSireArray[q][w] != 0 && (sireDadHaplo[w] != sireMomHaplo[w]) &&
          countSteps == 0){
            pos = w;
            countSteps++;
          }
        }

        /*-------------------------------------------------------------------------------------------------
        / if the none-zero progeny allele not found to the right of the missing progeny allele
        / search for the none-zero progeny allele to the left of the missing allele
        /-------------------------------------------------------------------------------------------------*/

        if(countSteps == 0){
          for(int w = v -1; w >= 0; w--){
            if(progAlleleSireArray[q][w] != 0 && (sireDadHaplo[w] != sireMomHaplo[w])
            && countSteps == 0){
              pos = w; countSteps++;
            }
          }
        }

        if(progAlleleSireArray[q][pos] == sireDadHaplo[pos]){
          progAlleleSireArray[q][v] = sireDadHaplo[v];
```

```
          }
          else
            if(progAlleleSireArray[q][pos] == sireMomHaplo[pos]){
              progAlleleSireArray[q][v] = sireMomHaplo[v];
            }
        } // close the heterozygous case
        else
          if(sireDadHaplo[v] == sireMomHaplo[v]){ // a homozygous case
            progAlleleSireArray[q][v] = sireDadHaplo[v];
          } // close the homozygous case
      } // close the if = 0 case
    } // close the inner for loop
  } // close the progNum for loop

/*-------------------------------------------------------------------------------------------------------------
/         STEP4:
/
/         UPDATE THE PROGENY MATERNAL HAPLOTPYE AND SET THE DAM
/         MATERNAL HAPLOTYPE TO PROGENY MATERNAL HAPLOTYPE AND THE
/         DAM PATERNAL HAPLOTYPE TO UNKNOWN (UNK) TO FACILITATE
/         CALCULATION OF THE IBD PROBABILITIES.
/-------------------------------------------------------------------------------------------------------------*/

StringTokenizer progToDamToken;
String[] damPaternalHaplo = new String[progNum];
String[] damMaternalHaplo = new String[progNum];
String DAMID[] = new String[progNum];
String progID;
int count1, progForDamAllele1, progForDamAllele2;

for(int g = 0; g < progNum; g++){

  // find the progeny genotype from the genotype file

  count1 = 0;
  progToDamToken = new StringTokenizer(inputAnim[count1]);
  progID = progToDamToken.nextToken();

  while(!progIDvec[g].equals(progID) && count1 < numAnim){
    progToDamToken = new StringTokenizer(inputAnim[count1]);
    progID = progToDamToken.nextToken();
    count1++;
  }

  // skip the sire and assign the dam ID

  progToDamToken.nextToken();
  DAMID[g] = progToDamToken.nextToken();
  damPaternalHaplo[g] = DAMID[g] + " ";
  damMaternalHaplo[g] = DAMID[g] + " ";

  // processing the marker at a time

  for(int d = 0; d < numMarkers; d++){

    progForDamAllele1 = Integer.parseInt(progToDamToken.nextToken());
```

```
    progForDamAllele2 = Integer.parseInt(progToDamToken.nextToken());

    // the progeny should be heterozygous at this locus

    if(progAlleleDamArray[g][d] == 0){
      if((progForDamAllele1 != 0 && progForDamAllele2 != 0) && (progForDamAllele1 !=
       progForDamAllele2)){
        if(progAlleleSireArray[g][d] == progForDamAllele1){
          progAlleleDamArray[g][d] = progForDamAllele2;
        }
        else
          if(progAlleleSireArray[g][d] == progForDamAllele2){
            progAlleleDamArray[g][d] = progForDamAllele1;
          }
      }
    }

    damPaternalHaplo[g] = damPaternalHaplo[g] + " " + "UNK";
    damMaternalHaplo[g] = damMaternalHaplo[g] + " " + progAlleleDamArray[g][d];
  } // close the d for loop
} // close the progeny for loop

// store the animal haplotypes in a vector for printing later

// a vector to store all the haplotypes

Vector haplotypes = new Vector(20000);

// get the sire pedigree and its haplotypes

StringTokenizer sireTok;
String sireAsAnim, sirePed, sirePaternalFinHaplo, sireMaternalFinHaplo;

int countSire = 0;
sireTok = new StringTokenizer(inputAnim[countSire]);
sireAsAnim = sireTok.nextToken();
countSire++;

while(!sire.equals(sireAsAnim) && countSire < numAnim){
  sireTok = new StringTokenizer(inputAnim[countSire]);
  sireAsAnim = sireTok.nextToken();
  countSire++;
}

sirePed = sireAsAnim + " " + sireTok.nextToken() + " " + sireTok.nextToken() + " ";
sirePaternalFinHaplo = sirePed; sireMaternalFinHaplo = sirePed;
int countMarkers = 0;

while(countMarkers < numMarkers){
  sirePaternalFinHaplo = sirePaternalFinHaplo + " " + sireDadHaplo[countMarkers] + " ";
  sireMaternalFinHaplo = sireMaternalFinHaplo + " " + sireMomHaplo[countMarkers] + " ";
  countMarkers++;
 }

if(!siresAlreadyStored.contains(sirePed)){
  haplotypes.addElement(sirePaternalFinHaplo);
```

```
  haplotypes.addElement(sireMaternalFinHaplo);
}

// get the progeny and dam pedigree and haplotypes

StringTokenizer animToken, damTok, getDamToken, damSireHaploToken;
StringTokenizer damDamHaploToken;
String damPaternalFinHaplo, damMaternalFinHaplo, animPaternalFinHaplo;
String animMaternalFinHaplo, dam, damAsAnim, animPed, damPed;
int countAnim, countDam;

for(int y = 0; y < progNum; y++){

  // get the dam pedigree and its haplotypes

  getDamToken = new StringTokenizer(progSireHaplo[y]);
  getDamToken.nextToken(); getDamToken.nextToken();
  dam = getDamToken.nextToken();
  countDam = 0;
  damTok = new StringTokenizer(inputAnim[countDam]);
   damAsAnim = damTok.nextToken();
   countDam++;

  if(!dam.equals(".")){  // known dams
     while(!dam.equals(damAsAnim) && countDam < numAnim){
       damTok = new StringTokenizer(inputAnim[countDam]);
       damAsAnim = damTok.nextToken();
       countDam++;
     }

     damPed = damAsAnim + " " + damTok.nextToken() + " " +
     damTok.nextToken() + " ";
     damPaternalFinHaplo = damPed;
     damMaternalFinHaplo = damPed;
     countMarkers = 0;
     damSireHaploToken = new StringTokenizer(damPaternalHaplo[y]);
     damDamHaploToken = new StringTokenizer(damMaternalHaplo[y]);

     // skip the damID

     damSireHaploToken.nextToken();
     damDamHaploToken.nextToken();

      while(countMarkers < numMarkers){
        damPaternalFinHaplo = damPaternalFinHaplo + " " +
       damSireHaploToken.nextToken() + " ";
       damMaternalFinHaplo = damMaternalFinHaplo + " " +
        damDamHaploToken.nextToken() + " ";
        countMarkers++;
      }

      haplotypes.addElement(damPaternalFinHaplo);
      haplotypes.addElement(damMaternalFinHaplo);

  } // close the if dam known loop
```

```
            animToken = new StringTokenizer(progSireHaplo[y]);
            animPed = animToken.nextToken() + " " + animToken.nextToken() + " " +
            animToken.nextToken() + " ";
            animPaternalFinHaplo = animPed;
            animMaternalFinHaplo = animPed;

             for(int w = 0; w < numMarkers; w++){
               animPaternalFinHaplo = animPaternalFinHaplo + " "+ progAlleleSireArray[y][w] + " ";
               animMaternalFinHaplo = animMaternalFinHaplo + " "+ progAlleleDamArray[y][w] + " ";
             }

            if(!haplotypes.contains(animPaternalFinHaplo)){
              haplotypes.addElement(animPaternalFinHaplo);
              haplotypes.addElement(animMaternalFinHaplo);
              siresAlreadyStored.addElement(animPed);
            }
          } // close the y for loop

            String  haploString;

            // print the haplotypes

            for(int i = 0; i < haplotypes.size(); i++){
              haploString = (String)haplotypes.elementAt(i);
              outFile.println(haploString);
            }
          }  // close the numSire for loop

          outFile.close();

      } // close the MAIN method
} // close the PUBLIC class
```