THE RELEVANCE OF MILK PROTEIN POLYMORPHISMS FOR DAIRY CATTLE BREEDING

Promotor: dr. ir. E. W. Brascamp,
hoogleraar in de veefokkerij

Co-promotor: dr. ir. S. Korver,
directeur produktstrategie en research & development,
Euribrid B.V.

Henk Bovenhuis

THE RELEVANCE OF MILK PROTEIN POLYMORPHISMS FOR DAIRY CATTLE BREEDING

Proefschrift
ter verkrijging van de graad van doctor
in de landbouw- en milieuwetenschappen
op gezag van de rector magnificus,
dr. H.C. van der Plas,
in het openbaar te verdedigen
op maandag 28 september 1992
des namiddags te vier uur in de Aula
van de Landbouwuniversiteit te Wageningen.

Voorwoord

De totstandkoming van een proefschrift is zelden het werk van een éénling. Hoewel mijn naam op de omslag prijkt, vormt dit proefschrift hierop geen uitzondering. Siem Korver heeft aan de wieg van dit project gestaan. In de eerste jaren is Siem mij bijzonder behulpzaam geweest bij het op het juiste tijdstip doorhakken van de juiste knopen. Ook na zijn vertrek bij de vakgroep veefokkerij is Siem nauw betrokken gebleven bij dit project. Hiervoor mijn hartelijke dank. Na het vertrek van Siem is de directe begeleiding overgenomen door Johan van Arendonk. Johan, bedankt voor de uitstekende begeleiding. Prof. dr. ir. R. D. Politiek was in de beginperiode betrokken bij het project. De rol van promotor is in een vroeg stadium door Prof. dr. ir. Pim Brascamp overgenomen. Pim wil ik bedanken voor zijn bijdrage aan de discussies in de begeleidingscommissie. Het laatste hoofdstuk van dit proefschrift zou niet tot stand zijn gekomen zonder de hulp van Imke de Boer.

Een gedeelte van de werkzaamheden zijn in het laboratorium uitgevoerd. In eerste instantie nogal een omschakeling voor een veefokker. Gelukkig werd ik liefdevol opgevangen en ingewijd in de geheimen van het lab door Rosilde Dijkhof, Jan van der Poel en Martien Groenen. Veruit de belangrijkste bijdrage aan het totstandkomen van dit proefschrift is geleverd door Esther Verstege. Zij heeft duizenden melkmonsters geannalyseerd. Ik kan me levendig voorstellen dat het op een gegeven moment letterlijk melk *monsters* waren die je annalyseerde. Esther, heel hartelijk bedankt voor je inzet, je doorzettingsvermogen en de prettige samenwerking. Wanneer er weer een stapel kratten met honderden melkmonsters werd afgeleverd bekroop mij zo nu dan het gevoel om zwaaiend aan één van de vleeshaken het lab te ontvluchten. Gelukkig was er dan de hulp van Tjalling de Jong en Chris Schrooten. Jongens, ook jullie bedankt.

Een groot aantal mensen hebben geholpen bij het verkrijgen van melkmonsters van de juiste koeien. Henk Vos heeft geholpen bij de organisatie. De centrale melkcontrole diensten in Oost- en Zuid-Nederland, de vele VVB's en veehouders zijn bijzonder behulpzaam geweest bij het verzamelen van de melkmonsters. Mijn hartelijke dank voor de goede samenwerking.

Ook gaat mijn dank uit naar mijn kamergenoot gedurende deze vier jaar, Ron Hovenier. Ondanks dat "ham" en "kaas" zeer verschillend zijn blijken ze uitstekend samen te gaan. Bedankt voor de goede werksfeer, de hulp bij diverse computerproblemen en bij het bouwen van luchtkastelen die hun plaats vonden in de geitenweide. Verder heeft Ron een belangrijk aandeel in de totstandkoming van de in dit proefschrift vermelde resultaten, door het veelvuldig verstrekken van random getallen die als "seeds" zijn gebruikt bij de simulaties. I would also

Stellingen

1. Het κ-caseïne gen vertoont een sterke relatie met het eiwitgehalte en het ß-lactoglobuline gen met het vetgehalte in de melk. In beide gevallen kan dit een effect zijn van het gen zelf of van een nauw gekoppeld gen.
   Dit proefschrift

2. Het vetgehalte in de melk wordt naast een aantal genen met een onbekend effect en milieufactoren, beïnvloed door één of enkele genen met een zeer groot effect.
   Dit proefschrift

3. Additionele selectie op κ-caseïne en ß-lactoglobuline genotypen biedt interessante mogelijkheden voor het verhogen van de selectierespons.
   Dit proefschrift

4. Gezien de in de huidige studie gevonden effecten van inkruising met Holstein Friesians op de ß-caseïne genfrequenties, lijkt het waarschijnlijk dat de door Bech en Kristiaansen (1990) gepresenteerde effecten van ß-caseïne varianten op melkproductiekenmerken herleid kunnen worden tot raseffecten.
   Bech and Kristiaansen (1990), J. Dairy Res. 57:53-62.

5. Het succes waarmee selectie op enkelvoudige genen in de rundveefokkerij kan worden toegepast neemt toe naarmate een groter aantal nakomelingen uit één paring kan worden verkregen.

6. De gevolgen van de éénzijdige selectie op melkproduktiekenmerken voor de secundaire productiekenmerken worden op dit moment gemaskeerd door de inkruising van de Nederlandse rundveepopulatie met Holstein Friesians.

7. In plaats van het selecteren van ouderdieren waarvan wordt verwacht dat ze in de volgende *generatie* de beste nakomelingen geven dient de veefokkerij zich te richten op het selecteren van ouderdieren waarvan wordt verwacht dat ze in volgende *generaties* de best nakomelingen geven.

8. Een drastische vermindering van de krachtvoergift in de rundveehouderij is schadelijk voor het milieu.

9. Bij fusies van onderzoeksinstellingen, met als doel het verkrijgen van een grotere kritische massa, dient rekening gehouden te worden met de traagheid van massa.

10. Runderen met een oormerk hebben een streepje voor.

11. De door de opkomst van de biotechnologie geboden mogelijkheid tot het bestuderen van effecten van individuele genen zal leiden tot een toenemend toeschrijven van diverse fenomenen aan erfelijke factoren waarbij de rol van milieu factoren naar de achtergrond wordt gedrongen.

12 Het bepalen van een DNA sequentie vergt niet de voor het octrooirecht vereiste inventiviteit.

13. De met het voortschrijden van de wetenschap aan het licht komende complexiteit van het leven werkt in het voordeel van de schepping als verklaring ervan.

Henk Bovenhuis
The relevance of milk protein polymorphisms for dairy cattle breeding.
Wageningen, 28 september 1992.

like to thank my other roommate Ming Wei. Ming, thanks for the pleasant working atmosphere and the interesting discussions.

I thank Dr. Joel Weller, Margaret Mackinnon, Dr. Micha Ron, Ariel and Amira for making my stay in Israel pleasant and very useful. Further, I would like to thank Prof. Dr. Mike Grossman, Prof. Dr. Brian Kennedy and Dr. Chris Davies for their corrections and suggestions of parts of the thesis.

Tot slot wil ik mijn ouders bedanken. Zij hebben mij altijd gesteund en gestimuleerd bij mijn studie.

Henk

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG
ISBN 90-5485-024-8

Bovenhuis, H., 1992. The relevance of milk protein polymorphisms for dairy cattle breeding (Het belang van melkeiwitvarianten voor de rundveefokkerij). An isoelectric focusing method is described that is suitable for phenotyping a large number of milk samples due to its short separation time and its high capacity. To estimate milk protein gene frequencies for breeds represented in the crossbred population and to estimate the fraction of animals misclassified a maximum likelihood model was developed. The estimates revealed that Dutch Friesians and Holstein Friesians differ for ß-casein and ß-lactoglobulin gene frequencies. Associations between milk protein genotypes and milk production traits were estimated using an animal model. Results from the present study and from literature indicate that the κ-casein gene or a very closely linked gene affects protein percentage, and the ß-lactoglobulin gene or a very closely linked gene affects fat percentage. A maximum likelihood model was constructed to estimate effects of both a marker gene and linked quantitative trait locus (QTL) on quantitative traits in segregating populations. Significant effects of QTLs linked to ß-casein, κ-casein and ß-lactoglobulin on fat percent were found. The effects of selection for κ-casein and ß-lactoglobulin genotypes were studied by using stochastic simulation of a closed adult MOET nucleus breeding scheme. Results showed that selection for κ-casein and ß-lactoglobulin genotypes has the potential to increase selection response. *Doctoral thesis, Department of Animal Breeding, Wageningen Agricultural University, P.O. Box 338, 6700 AH Wageningen, The Netherlands.*

CONTENTS

Chapter 1.	General introduction.	1
Chapter 2.	Milk proteins and milk protein genetic variants.	9
Chapter 3.	Improved method for phenotyping milk protein variants by isoelectric focusing using PhastSystem.	23
Chapter 4.	Estimation of milk protein gene frequencies in crossbred cattle by maximum likelihood method.	31
Chapter 5.	Associations between milk protein polymorphisms and milk production traits.	49
Chapter 6.	Mapping and analysis of dairy cattle quantitative trait loci by maximum likelihood methodology using milk protein genes as genetic markers.	69
Chapter 7.	The value of selection for κ-casein and ß-lactoglobulin genotypes in dairy cattle breeding.	101
Summary		127
Samenvatting		133
Curriculum vitae		139

Chapter 1

GENERAL INTRODUCTION

In the late fifties, protein content of milk became part of the Dutch milk recording and payment system, because of the increasing importance of cheese production (Politiek, 1957). Figure 1 shows the total amount of milk processed by the Dutch dairy industry, the amount of milk manufactured into cheese and the fraction of milk manufactured into cheese, from 1950 to 1990 (Jaarverslagen PZ, 1950 - 1990). In 1950 about 21 percent of the milk was manufactured into cheese. This fraction increased steadily from 1950 to 1990 and in 1990 more than

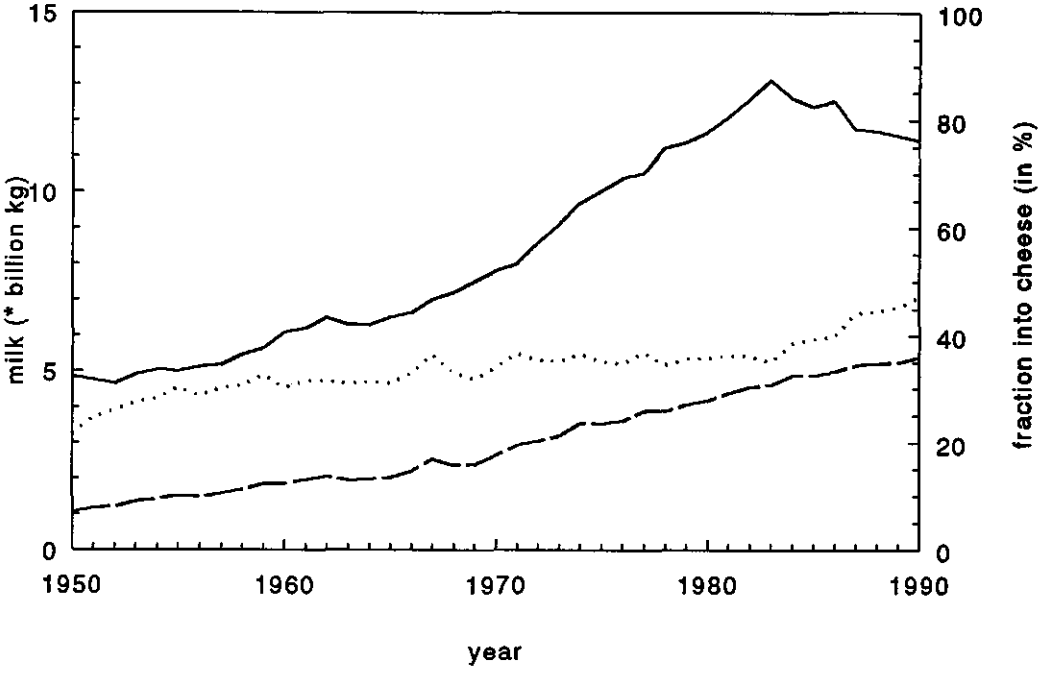


Figure 1. The total amount of milk processed by Dutch dairy industry (——), the amount of milk manufactured into cheese (-------) and the fraction of milk manufactured into cheese (·····), from 1950 to 1990.

47 percent of the total amount of milk was manufactured into cheese. Since the consideration of milk protein content in the late fifties, cheese production has gained increasing importance. Consequently, the importance of milk protein increased further during this period.

Selection of dairy cattle in The Netherlands is mainly based on a Net-profit-index (INET), i.e. an index combining breeding values for milk yield, fat yield and protein yield. Dutch dairy cattle breeding organisations have anticipated on changes in payment for fat and protein content by changing the weighing factors for the three production traits in the INET accordingly (Dommerholt and

Wilmink, 1986). In 1980 a fat to protein price ratio of 1:1 was used to calculate the weighing factors for the INET, in 1988 a price ratio of 1:1.2 was used and in 1989 the fat to protein price ratio changed to 1:3. This shows that in recent years Dutch dairy cattle breeding has placed more emphasis on increasing protein yield relative to increasing fat yield.

Several studies have reported relations between milk protein genetic variants and cheese manufacturing properties of milk (reviewed by Grosclaude, 1988). The main results of these studies are that κ-casein genotypes are associated with renneting time of milk and ß-lactoglobulin genotypes are associated with casein number. Milk protein genotypes might be useful additional selection criteria to improve the quality of milk for cheese production, especially for countries where a large fraction of the milk is manufactured into cheese.

At present, breeding values of animals are estimated assuming an infinitesimal model, i.e. assuming an infinite number of genes, each with a small effect and estimation of breeding values is limited to situations where phenotypic observations are available on the animal or on its relatives. For milk production traits differences in phenotypic observations between animals are only partly due to genetic differences. Environmental effects contribute to phenotypic differences as well. Further, segregation of genes takes place each time they are transmitted from parent to offspring. As a result of these two factors an accurate estimation of the breeding value of an animal is possible only if a large number of records on its offspring or on the individual itself are available. In general, the requirement of a large number of records postpones the age at which the animal is selected and therefore restricts attainable annual genetic progress. When the genes and their effects on traits of interest are known, typing of animals at DNA level can lead to estimation of breeding values independent of phenotypic observations, and can increase annual genetic progress (Smith and Simpson, 1986). However, first those genes need to be identified and their effects on milk production traits need to be quantified. Milk protein genes might be genes that are involved in the regulation of milk production traits. Further, milk protein genes can be used for the mapping of genes, affecting milk production traits, linked to the milk protein genes. As such, milk protein genes might be useful in dairy cattle breeding.

AIM OF STUDY

Cheese production is of great importance for Dutch dairy industry, and milk protein genotypes affect cheese manufacturing properties. Consequently, the question can be raised whether, in addition to traditional selection, selection should be for milk protein genotypes to improve the quality of milk for cheese production. In addition to the effect on cheese manufacturing properties, the economic value of milk protein genotypes also depends on their associations with milk production traits. Furthermore, mapping and the analysis of genes affecting milk production traits has a potential use in animal breeding.

The mapping of genes linked to milk protein genes in a segregating population requires the typing of large numbers of animals (Soller and Genizi, 1978). Separation of milk protein genetic variants has traditionally been performed by electrophoresis using urea containing starch- or polyacrylamide gels (Swaisgood, 1975). These methods, however, require a long separation time (4-18 h) and two separate runs (alkaline and acid conditions) to distinguish all genetic variants. Chapter 3 describes a method that, considering restrictions set by time and money, makes it possible to type a large number of animals.

The possibilities for genetically improving the population for a single gene are determined in part by the frequencies of the alleles (Falconer, 1989). Therefore, information is required on the gene frequencies of milk protein genetic variants in the Dutch dairy cattle population. A complicating factor, when estimating gene frequencies in the Dutch cattle population, is that semen from Holstein Friesian sires has frequently been used. The fraction of Holstein Friesian genes in the population has not yet stabilised, and is expected to increase. Therefore, the gene frequency at present reflects the gene frequency given the fraction of Holstein Friesian genes currently in the population. To overcome this problem, a maximum likelihood model to estimate gene frequencies in a segregating crossbred population was developed. The method accounts for relations between animals due to a common sire, for breed differences in gene frequencies and for the probability of misclassification of genotypes. Estimates of milk protein gene frequencies for the breeds represented in the population and of the probability of misclassification for the population are provided by the method. The method and the results of the calculations are described in chapter 4.

The evaluation of selection for milk protein genotypes to improve the quality of milk for cheese production requires the estimation of relations between milk protein genotypes and milk production traits. These relations, together with the value of milk protein genotypes for manufacturing properties, determine the

value of milk protein genotypes for selection purposes. Chapter 5 describes the estimation of effects of milk protein genes on milk production traits. To obtain an exact test of associated hypotheses and unbiased estimates of genotype effects an animal model was used (Kennedy et al., 1992). Because the casein genes are closely linked, results for a model accounting for the effects of one milk protein gene at a time and those from a model accounting for the effects of all milk protein genes simultaneously were compared to disentangle the effects of the different genes.

Milk protein genes can be used as genetic markers for mapping quantitative trait loci affecting milk production traits. By using contrasts between daughters of a sire, that are grouped according to the milk protein allele that was inherited from the sire, it is possible to estimate effects of genes linked to the milk protein genes (Neimann-Sorensen and Robertson, 1961). The methods that have been described to date assume no effect of the genetic marker itself. However, milk protein genes are functional genes and therefore the milk protein gene itself or a closely linked region (promoter or enhancer) might be involved in the regulation of milk protein production. To account for this a model was constructed to estimate effects of both the marker gene and a linked QTL on a quantitative trait. This model was used to estimate direct and linked effects of milk protein genes on milk production traits. In chapter 6 the model and the results of the calculations are described.

The value of incorporating knowledge on κ-casein or ß-lactoglobulin genotypes in a selection programme was studied for a closed MOET (Multiple Ovulation and Embryo Transfer) nucleus breeding scheme. In this study information on gene frequencies (chapter 4), associations with milk production traits (chapter 5) and the value of milk protein genotypes with respect to manufacturing properties of milk was used. Chapter 7 describes the simulation model and the results of the model calculations.

REFERENCES

Dommerholt, J. and J.B.M. Wilmink. 1986. Optimal selection responses under varying milk prices and margins for milk production. Livest. Prod. Sci. 14: 109.

Falconer, D. S. 1989. Introduction to quantitative genetics. Longman Scientific & Technical, Essex, Engl. p. 1-438.

Grosclaude, F. 1988. Le polymorphisme génétique des principales lactoprotéines bovines. INRA Prod. Anim. 1: 5.

Kennedy, B. W., M. Quinton and J. A. M. van Arendonk. 1992. Estimation of effects of single genes on quantitative traits. J. Anim. Sci. 70: 1999.

Neimann-Sorensen, A. and A. Robertson. 1961. The associations between blood groups and several production characteristics in three Danish cattle breeds. Acta Agric. Scan. 11: 163.

Produktschap voor zuivel. Statistisch jaaroverzicht 1950 - 1990.

Politiek, R.D. 1957. De invloed van erfelijkheid en milieu op de samenstelling van de melk bij Friese koeien en de practische mogelijkheid van selectie op het eiwitgehalte. Ph.D. Thesis. Wageningen Agric. Univ., Neth. p. 1-174.

Smith, C. and S. P. Simpson. 1986. The use of genetic polymorphisms in livestock improvement. J. Anim. Breed. and Genet. 103: 205.

Soller, M. and A. Genizi. 1978. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. Biometrics 34: 47.

Swaisgood, H.E. (Ed.). 1975. Methods of gel electrophoresis of milk proteins, American Dairy Science Association, p. 1-30.

Chapter 2

MILK PROTEINS AND MILK PROTEIN GENETIC VARIANTS

Milk proteins have been subject of numerous studies. These studies were undertaken from a wide variety of scientific disciplines. Therefore, this thesis might be of potential interest for a diverse group of researchers. The aim of this chapter is to provide the reader with a brief overview of what is known about milk proteins and milk protein genes.

MILK PROTEIN COMPOSITION

Table 1 shows the approximate bovine milk protein composition. About 80 percent of the milk protein consists of caseins while the other 20 percent are whey proteins. The caseins coagulate at pH 4.6 while whey proteins remain in solution. Caseins can be represented by four gene products: α_{s1}-casein, α_{s2}-casein, ß-casein and κ-casein. α_{s1}-Casein and ß-casein are with some 30 percent each, the main milk protein components. In table 1, the concentration of ß-casein includes γ-casein, a casein formed by the cleavage of ß-casein by the enzyme plasmin. α_{s2}-Casein and κ-casein contribute at about 10 percent each to the total milk protein content. Almost all the caseins in milk are present in casein micelles. κ-casein plays an important role in preventing the precipitation of the other caseins. Because the enzyme chymosin (rennin) cleaves κ-casein, treatment of milk with chymosin results in the formation of a curd. The function of caseins is to provide the progeny with a source of amino acids, phosphate and calcium (Swaisgood, 1982, Holt and Sawyer, 1988, Mercier et al.,1990). Recently it was suggested that caseins also might have a role in the cytotoxic T lymphocytes-mediated cytolysis (Grusby et al., 1990).

The main whey proteins are ß-lactoglobulin and α-lactalbumin. At about 10 percent of the milk protein consists of ß-Lactoglobulin. α-Lactalbumin is with 3.7 percent a minor milk protein component. α-Lactalbumin is involved in the synthesis of lactose. The function of ß-lactoglobulin is less clear but it is suggested that ß-lactoglobulin is involved in the transport of vitamin A (Grosclaude, 1989, Mercier et al., 1990).

GENETIC VARIANTS

Milk protein genes can be subject to a deletion, a duplication or an insertion. In some cases this leads to a change in the amino acid composition of the protein, resulting in a new genetic variant. Asschaffenburg and Drewry (1955) were the first to discover that the ß-lactoglobulin gene is polymorphic. Later on, polymorphism of other milk protein genes was discovered. Table 1 shows the

Table 1. Approximate concentration of the major bovine milk proteins and their genetic variants (Walstra and Jenness, 1984, Eigel et al., 1984).

	Concentration in milk ($g \cdot kg^{-1}$)	Percentage of Total Protein (w/w)	Genetic Variants
Total Protein	33.0	100.0	
Total Casein	26.0	79.5	
α_{s1}-Casein	10.0	30.6	A, B, C, D, E
α_{s2}-Casein	2.6	8.0	A, B, C, D
ß-Casein[1]	10.1	30.8	A^1, A^2, A^3, B, C, D, E
κ-Casein	3.3	10.1	A, B
Whey Proteins	6.3	19.3	
ß-Lactoglobulin	3.2	9.8	A, B, C, D, E, F, G
α-Lactalbumin	1.2	3.7	A, B
Other	1.9	5.7	
Fat Globule Membrane Proteins	0.4	1.2	

[1)] including γ-casein

genetic variants of the different milk proteins (Eigel et al., 1984). Not all genetic variants in table 1 appear in Western dairy breeds, e.g. in most Western dairy breeds only the α_{s2}-casein A allele appears whereas the α_{s2}-casein C variant was observed in yaks and the α_{s2}-casein B variant was found in zebu cattle (Eigel et al.,1984). In addition to genetic variants of caseins, variation which results from post translational modification, such as phosphorylation or glycosylation, occurs (Larson, 1979, Swaisgood, 1982).

Genetic variants differ from breed to breed in their occurrence and frequency of occurrence. Table 2 shows that in Western dairy breeds α_{s1}-casein B and C, ß-casein A^1, A^2 and B, κ-casein A and B and ß-lactoglobulin A and B are the most common genetic variants. In most Western dairy breeds the α_{s2}-casein and the α-lactalbumin genes are fixated at the A and the B allele, respectively, and therefore they were not included in table 2. In most breeds the α_{s1}-casein gene

Table 2. Milk protein gene frequency estimates for different western dairy cattle breeds in various countries.

Breed Country	HF[1] USA	HF[2] Canada	DF[3] Netherl.	MRIJ[3] Netherl.	Jersey[4] Australia	Fleckvieh[5] Germany	Braunvieh[6] Germany	Guernsey[7] USA
				α_{s1} - Casein				
A	.003	.003	-	-	-	-	-	-
B	.957	.970	.980	.980	.628	.900	.920	.737
C	.040	.027	.020	.020	.372	.100	.080	.263
No.	6874	2045	693	272	308	2262	2139	3888
				ß - Casein				
A^1	.415	.561			.074			.008
A^2	.532	.421			.564			.962
A^3	.028	.011			-			-
A^*	.975	.993	.910	.920	.638	.876	.659	.970
B	.025	.007	.090	.080	.362	.081	.282	.016
C	-	-	-		-	.043	.059	.014
No.	6575	2045	693	272	308	2262	2139	3888
				κ - Casein				
A	.800	.744	.660	.540	.227	.687	.442	.730
B	.200	.256	.340	.460	.773	.313	.558	.270
No.	6531	2045	164	92	308	2262	2139	3888
				ß - Lactoglobulin				
A	.526	.387			.329	.469	.421	.385
B	.474	.613			.565	.515	.574	.615
C	-	-			.106	-	-	-
D	-	-			-	.012	.005	-
No.	6465	3870	-	-	308	2262	2139	3888

A^* ; A^1, A^2 and A^3 were not distinguished.

1) Gonyon et al., (1987), 2) Ng-Kwai-Hang et al., (1984), 3) Schmidt, (1966) (Calculated from genotype frequencies), 4) Mclean et al., (1984), 5) Graml et al., (1984^a), 6) Graml et al., (1984^b), 7) Haenlein et al. (1987)

Table 3. Total number amino acids and amino acid substitutions of the main genetic variants of milk proteins (reviewed by Grosclaude, 1988).

Milk Protein	Number of amino acid residues	Comparison of genetic variants	Change in amino acid composition
α_{s1} - Casein	199	C ⟶ B	192[1]: Gly ⟶ Glu
α_{s2} - Casein	207		
ß - Casein	209	A^2 ⟶ A^1	67 : Pro ⟶ His
		A^2 ⟶ B	67 : Pro ⟶ His 122 : Ser ⟶ Arg
		A^2 ⟶ A^3	106 : His ⟶ Gln
κ - Casein	169	A ⟶ B	136 : Thr ⟶ Ile 148 : Asp ⟶ Ala
ß - Lactoglobulin	162	B ⟶ A	64 : Gly ⟶ Asp 118 : Ala ⟶ Val
α - Lactalbumin	123		

1) number of amino acid that has been substituted

is almost fixated at the B-allele. For the ß-casein gene the A^* allele is predominant. Studies that distinguished the A^1, A^2 and A^3 alleles show that the ß-casein A^2-allele is the most common ß-casein allele. In most breeds the κ-casein A allele has a much higher frequency than the κ-casein B allele. The A and B-allele of ß-lactoglobulin appear at intermediate frequencies.

CHARACTERISTICS OF GENETIC VARIANTS

Table 3 shows the differences in amino acid composition between the main milk protein genetic variants (reviewed by Grosclaude, 1988). The α_{s1}-casein C and B variant differ in amino acid composition at position 192 which is glycine in α_{s1}-casein C and glutamine in α_{s1}-casein B. Table 3 shows that the difference between genetic variants is limited to one or two changes in amino acid composition. In the case of α_{s1}-casein A, however (not in table 3), compared with α_{s1}-casein B, amino acids 14 up to 26 are deleted. Detection of genetic variants

is usually by means of electrophoresis (Swaisgood, 1975). Because separation of proteins by electrophoresis is based on differences in electric charge, only those substitutions of amino acids will be detected that cause a change in the net charge of the protein. Because most amino acids are specified by more than one codon, not all mutations at the DNA-level will lead to the production of another protein. Therefore, it should be realised that the polymorphism detected at the product level by means of electrophoresis reflects only part of the polymorphism at the DNA level (Grosclaude, 1988). Recently, the use of high resolution techniques for the separation of proteins has lead to the detection of new genetic variants of milk proteins (e.g. Erhardt, 1989).

GENETIC ORGANISATION OF MILK PROTEIN GENES

By using classical linkage analysis, it has been shown that the casein genes are closely linked (Hines et al., 1981). Recently, this has been confirmed by using pulse field gel electrophoresis (Ferretti et al., 1990, Threadgill and Womack, 1990). Threadgill and Womack (1990) concluded that the four casein genes reside on less than 200 kb of DNA in the order α_{s1}-casein - ß-casein - α_{s2}-casein - κ-casein. The four bovine casein genes were assigned to bovine chromosome 6 (Threadgill and Womack, 1990). α-Lactalbumin was assigned to the bovine syntenic group U3, which has been assigned to the bovine chromosome 5 (Fries et al. cited by Threadgill and Womack, 1990). The ß-lactoglobulin gene was located on syntenic group U16. The chromosome carrying syntenic group U16 has not yet been identified.

The structure of most of the milk protein genes has been elucidated by complete or partial sequencing of the genes. Table 4 shows the number of coding regions and the length of the main milk protein genes. It shows that the two α-casein genes are similar in length and number of exons. When compared to the α-casein genes, the number and the length of the intron sequences is lower for the ß-casein gene. The nucleotide sequence of the κ-casein gene suggests that this gene is unrelated to the other casein genes (Alexander et al., 1988). Further research in this area is aimed at the isolation of regulatory sequences involved in the regulation of casein gene expression (e.g. Groenen et al., 1990).

The knowledge about the DNA sequence of milk protein genes together with the development of restriction fragment length polymorphism (RFLP) and polymerase chain reaction (PCR) techniques has created the possibility for identifying milk protein genotypes directly at the DNA-level (e.g. Medrano, 1990, Rando et al. 1990, Savva et al., 1990, Skidmore et al., 1990, Zadworny et al.,

Table 4. Number of coding regions and the total length of the main bovine milk protein genes.

Milk Protein gene	Number of exons	Total gene length (*10^3 base pairs)
α_{s1}-casein[1]	19	17.5
α_{s2}-casein[2]	18	18.5
ß-casein[3]	9	8.6
κ-casein[4]	5	~ 13
ß-lactoglobulin[5]	7	~ 3.5
α-lactalbumin[6]	4	~ 2

1) Koczan et al. (1991), 2) Groenen et al. (1992), 3) Gorodetsky et al. (1988), 4) Alexander et al. (1988), 5) Silva et al. (1990), 6) Vilotte et al. (1987).

1990). This enables the determination of milk protein genotypes of non-lactating animals, i.e. animals at a young age and males.

GENETIC VARIANTS AND MANUFACTURING PROPERTIES

Several studies investigated relations between milk protein genetic variants and manufacturing properties of milk (e.g. Schaar et al. 1985, Marziali and Ng-Kwai-Hang, 1986[a,b], McLean et al. 1987, Aaltonen and Antila, 1987, Mariani et al. 1988, Corradini et al., 1988, McLean and Schaar, 1989, Rahali and Ménard, 1991, Van den Berg et al., 1992). Relations between genetic variants of milk proteins and manufacturing properties of milk can be due to (Grosclaude, 1988):

- The differences in amino acid composition between genetic variants of milk proteins that results in different physical or chemical properties. In this explanation the genetic variants differ qualitatively with respect to manufacturing properties.
- The amount of a certain milk protein fraction produced is related to its genetic variant. In this explanation a quantitative difference is the underlying cause for the qualitative effects of genetic variants on manufacturing properties.

Table 5. Main effects of milk protein genotypes on manufacturing properties of milk (Van den Berg et al., 1992).

Milk protein genotype	Renneting time (min)	Conversion of total N into cheese-N (in %)
	κ - Casein	ß - Lactoglobulin
AA	29.9	71.0
AB	25.8	72.8
BB	22.7	73.9

Further, it is possible that not the milk protein gene but (a) gene(s) linked to the milk protein gene affect manufacturing properties of milk. A possible effect of linked genes on manufacturing properties might again have a qualitative or a quantitative origin.

The main results of an extensive study carried out in the Netherlands, are in table 5. This indicates that genetic variants of κ-casein are related with renneting time of milk while genetic variants of ß-lactoglobulin are related with cheese yield. These results agree with previous findings (reviewed by Grosclaude, 1988). The association between the κ-casein B allele and a shorter renneting time might be explained from the increased κ-casein B content that was found in milk of cows carrying the κ-casein B allele (Van den Berg et al., 1992, Van Eenennaam and Medrano, 1991). A higher κ-casein content results in smaller micelles, which in turn result in better renneting properties. Table 5 shows that ß-Lactoglobulin genotypes have a considerable effect on the conversion of milk nitrogen into cheese nitrogen. Van den Berg et al. (1992) indicated that this effect can be explained by the association between the ß-lactoglobulin alleles and casein number. The B allele of ß-lactoglobulin is related to a lower ß-lactoglobulin content and a higher casein content (Grosclaude, 1988, Van den Berg et al., 1992).

It seems that the main effects of κ-casein and ß-lactoglobulin on manufacturing properties can be explained by the relation between the genetic variant and the quantity of the corresponding protein fraction. This phenomena has been observed for several milk protein genes in different species (Grosclaude et

al.,1987, Mahé et al., 1989, Graml et al., 1989, Brignon et al., 1990, Van Eenennaam and Medrano, 1991, Van den Berg et al., 1992).

REFERENCES

Aaltonen, M.L. and V. Antila. 1987. Milk renneting properties and the genetic variants of proteins. Milchwissenschaft 42 (8): 490.

Alexander, L. J., A. F. Stewart, A. G. Mackinlay, T. V. Kapelinskaya, T. M. Tkach and S. I. Gorodetsky. 1988. Isolation and characterization of the bovine κ-casein gene. Eur. J. Biochem. 178: 395.

Aschaffenburg, R., and J. Drewry. 1955. Occurrence of different beta-lactoglobulins in cow's milk. Nature 176: 218.

Brignon, G., M.-F. Mahé, B. Ribadeau-Dumas, J.-C. Mercier and F. Grosclaude. 1990. Two of the three genetic variants of goat α_{s1}-casein which are synthesized at a reduced level have an internal deletion possible due to altered RNA splicing. Eur. J. Biochem. 193: 237.

Corradini, C., P. Vecchia and G. Rossi. 1988. Effects of protein genetic polymorphism on renneting ability of low acidity individual milks. Scienza e Techn. Latt.-Cas 39: 423.

Eigel, W.N., J.E. Butler, C.A. Ernstrom, H.M. Farrel Jr, V.R. Harwalker, R. Jenness and R. Mcl. Whitney. 1984. Nomenclature of proteins of cow's milk: fifth revision. J. Dairy Sci. 67: 1599.

Erhardt, G. 1989. κ-Kaseine in Rindermilch-Nachweis eines weiteren allels (κ-Cn^E) in verschiedenen rassen. J. Anim. Breed. Genet. 106: 225.

Ferretti, L., P. Leone, G. Rognoni and V. Sgaramella. 1990. Linkage of the four bovine casein genes as demonstrated by pulsed field gel electrophoresis. Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh, Scotland. XIII: 75.

Gonyon, D. S., R. E. Mather, H. C. Hines, G. F. W. Haenlein, C. W. Arave, and S. N. Gaunt. 1987. Associations of bovine blood and milk polymorphisms with lactation traits: Holsteins. J. Dairy Sci. 70: 2585.

Gorodetsky, S. I., T. M. Tkach and T. V. Kapelinskava. 1988. Isolation and characterization of the Bos taurus ß-casein gene. Gene 66: 87.

Graml, R., J. Buchberger, O. Kirchmeier, F. Kiermeier and F. Pirchner. 1984[a]. Genfrequenzschätzung bei Milchproteinen des bayerischen Fleckviehs. Züchtungskunde 56: 73.

Graml, R., J. Buchberger, H. Klostermeyer and F. Pirchner. 1984[b]. Untersuchungen über die Genfrequenzen der Caseine und ß-

Lactoglobuline bei der bayerischen Braunviehpopulation. Züchtungskunde 56: 221.

Graml, R., G. Weiss, J. Buchberger and F. Pirchner. 1989. Different rates of synthesis of whey protein and casein by alleles of the ß-lactoglobulin and α_{s1}-casein locus in cattle. Genet. Sél. Evol. 21: 547.

Groenen, M. A. M., R. J. M. Dijkhof and J. J. Van der Poel. 1990. Organization and regulation of expression of the bovine α_{s2}-casein gene. Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh, Scotland. XIII: 79.

Groenen, M. A. M., R. J. M. Dijkhof, A. J. M. Verstege and J. J. Van der Poel. The complete nucleotide sequence of the bovine α_{s2}-casein gene. In Preparation.

Grosclaude, F. 1987. A Mendelian polymorphism underlying quantitative variations of goat α_{s1}-casein. Genet. Sél. Evol. 19: 399.

Grosclaude, F. 1988. Le polymorphisme génétique des principales lactoprotéines bovines. INRA Prod. Anim. 1: 5.

Grusby, M.J., S.C. Mitchell, N.Nabavi and L.H. Glimcher. 1990. Casein expression in cytotoxic T lymphocytes. Proc. Natl. Acad. Sci. USA 87: 6897.

Haenlein, G. F. W., D. S. Gonyon, R. E. Mather, and H. C. Hines. 1987. Associations of bovine blood and milk polymorphisms with lactation traits: Guernseys. J. Dairy Sci. 70: 2599.

Hines, H.C., G.F.W. Haenlein, J.P. Zikakis and H.C. Dickey. 1977. Blood antigen, serum protein, and milk protein gene frequencies and genetic interrelationships in Holstein cattle. J. Dairy Sci. 60: 1143.

Hines, H.C., J.P. Zikakis, G.F.W. Haenlein, C.A. Kiddy and, C.L. Trowbridge. 1981. Linkage relationships among loci of polymorphisms in blood and milk of cattl. J. Dairy Sci. 64: 71.

Holt, C. and L. Sawyer. 1988. Primary and predicted structures of the caseins in relation to their biological functions. Protein Engineering 4: 251.

Larson, B. L. 1979. Biosynthesis and secretion of milk proteins : a review. J. Dairy Res. 46: 161.

Koczan, D., G. Hobom and H. -M. Seyfert. 1991. Genomic organization of the bovine alpha-S1 casein gene. Nucleic Acids Res. 20: 5591.

Mahé, M.F. and F. Grosclaude. 1989. α_{s1}-Cn^{D}, another allele associated with a decreased synthesis rate at the caprine α_{s1}-casein locus. Genet. Sél. Evol. 21: 127.

Mariani, P., P. Bonatti and M. Pecorari. 1988. Rennet coagulation properties of cow milk in relation to α_{s1}-casein genotypes. Scienza e Techn. Latt.-Cas.

39: 431

Marziali, A.S. and K.F. Ng-Kwai-Hang. 1986[a]. Effects of milk composition and genetic polymorphism on coagulation properties of milk. J. Dairy Sci. 69: 1793.

Marziali, A.S. and K.F. Ng-Kwai-Hang. 1986[b]. Effects of milk composition and genetic polymorphism on cheese composition. J. Dairy Sci. 69: 2533.

McLean, D.M., E.R.B. Graham, R.W. Ponzoni and H.A. McKenzie. 1984. Effects of milk protein genetic variants on milk yield and composition. J. Dairy Res. 51: 531.

Mclean, D.M., E.R.B. Graham, R.W. Ponzoni and H.A. McKenzie. 1987. Effects of milk protein genetic variants and composition on heat stability of milk. J. Dairy Res. 54: 219.

Mclean, D.M. and J. Schaar. 1989. Effects of ß-lactoglobulin and κ-casein genetic variants and concentrations on syneresis of gels from renneted heated milk. J. Dairy Res. 56: 297.

Medrano, J. V. 1990. Application of the polymerase chain reaction procedure for genetic evaluation in cattle. Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh, Scotland. XIII: 71.

Mercier, J.-C., J.-L. Vilotte and C. Provot. 1990. Structure and function of milk protein genes. In: H. Geldermann (ed.): Genome analysis in domestic animals. VCH, Weinheim, Germany.

Ng-Kwai-Hang, K.F., J.F. Hayes, J.E. Moxley and H.G. Monardes. 1984. Association of genetic variants of casein and milk serum proteins with milk, fat, and protein production by dairy cattle. J. Dairy Sci. 67: 835.

Rahali, V. and J.L. Ménard. 1991. Influence des variants génétiques de la ß-lactoglobuline et de la κ-caséine sur la composition du lait et son aptitude fromagère. Lait 71: 275.

Rando, A., P. Di Gregorio, R. Davoli, S. Dall'Olio, P. Masina and V. Russo. 1990. Identification of the two common alleles of the bovine α_{s1}-casein locus by means of RFLP's. Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh, Scotland. XIV: 241.

Ribadeau-Dumas, R. 1991. Physicochimie et biochimie des protéines du lait. Données récentes. Lait 71: 133.

Savva, D., S. J. Pinder and C. J. Skidmore. 1990. Genotyping the ß-casein locus in cattle using PCR. Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh, Scotland. XIV: 245.

Schaar J., B. Hansson, and H. -E. Pettersson. 1985. Effects of genetic variants of κ-casein and ß-lactoglobulin on cheesemaking. J. Dairy Res. 52: 429.

Schmidt, D.G. 1966. Genetische varianten van caseïne. Het vóórkomen in Nederland. Veeteelt en Zuivelberichten 9: 431.

Silva, M. C., D. W. S. Wong and C. A. Batt. 1990. Cloning and partial nucleotide sequence of the genomic bovine ß-lactoglobulin gene. Nucleic Acids Res. 18: 3051.

Skidmore, C. J., S. J. Pinder, B. N. Perry and D. Savva. 1990. Genotyping the κ-casein locus in cattle using PCR. Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh, Scotland. XIV: 248.

Swaisgood, H.E. (Ed.). 1975. Methods of gel electrophoresis of milk proteins, American Dairy Science Association, p. 1-30.

Swaisgood, H.E. 1982. Chemistry of milk protein. In: P.F. Fox (Ed.): Developments in dairy chemistry - I proteins. Applied Science Publ., London. p. 1-59.

Threadgill, D. W. and J. E. Womack. 1990. Genomic analysis of the major bovine milk protein genes. Nucleic Acids Res. 18: 6935.

Van den Berg, G., J. T. M. Escher, P. J. de Koning and H. Bovenhuis. 1992. Genetic polymorphism of κ-casein and ß-lactoglobulin in relation to milk composition and processing properties. Neth. Milk Dairy J. 46: 145.

Vilotte, J. -L., S. Soulier, J. -C. Mercier, P. Gaye, D. Hue-Delahaie and J. -P. Furet. 1987. Complete nucleotide sequence of bovine α-lactalbumin gene: comparison with its rat counterpart. Biochimie 69: 609.

Walstra, P. and R. Jenness. 1984. Dairy chemistry and physics. John Wiley & Sons, Inc., New York. p 1-467.

Zadworny, D. U. Kuhnlein and K. F. Ng-Kwai-Hang. 1990. Determination of kappa-casein alleles in Holstein dairy cows and bulls using the polymerase chain reaction. Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh, Scotland. XIV: 251.

Chapter 3

IMPROVED METHOD FOR PHENOTYPING MILK PROTEIN VARIANTS BY ISOELECTRIC FOCUSING USING PHASTSYSTEM

Henk Bovenhuis and Esther A.J.M. Verstege
Department of Animal Breeding
Wageningen Agricultural University
P.O. Box 338
6700 AH Wageningen
The Netherlands

Published in: Netherlands Milk and Dairy Journal 43: 447 (1989)

ABSTRACT

A rapid method for the phenotyping of milk protein variants has been described. The method is based on the separation of milk protein variants by isoelectric focusing using PhastSystem. The method is suitable for phenotyping a large number of samples due to its short separation time and its high capacity.

Key words: milk protein polymorphism, isoelectric focusing, PhastSystem

INTRODUCTION

In all major milk protein fractions, α_{s1}-casein (α_{s1}-Cn), α_{s2}-casein (α_{s2}-Cn), β-casein (β-Cn), κ-casein (κ-Cn), β-lactoglobulin (β-Lg) and α- lactalbumin (α-La), polymorphism has been detected (Eigel et al., 1984). Because the polymorphism of some of the milk proteins influences processing properties of milk, these genetic variants might be of importance for the dairy industry and thereby for animal breeding (Schaar et al., 1985, McLean et al., 1987). However, before selection for certain favourable genetic variants is put into practice, information must be obtained about associations between genetic variants and milk production traits. This kind of study requires an extensive set of data and therefore a method which is suitable for phenotyping a large number of milk samples. Separation of genetic variants of milk proteins has traditionally been performed by electrophoresis in urea containing starch- or polyacrylamide gels (Swaisgood, 1975). The main disadvantages of these methods are the long separation time (4-18 h.) and the need for two separate runs (alkaline and acid conditions) to distinguish all described genetic variants. Seibert et al. (1985) reported simultaneous separation of casein and whey protein genetic variants by isoelectric focusing in ultrathinlayer polyacrylamide gels. This paper describes a method, which is suitable for phenotyping milk protein genetic variants in one single run with PhastSystem (Pharmacia, Uppsala, Sweden). This improved method of phenotyping by isoelectric focusing has the advantage of standardization, short separation time and high resolution.

MATERIAL AND METHODS

Sample preparation. 300 μl of a 8M urea solution containing 3% 2-ß mercaptoethanol was added to 100 μl of whole milk, mixed and incubated for 15

min at room temperature.

Isoelectric focusing. Isoelectric focusing was performed by PhastSystem (Pharmacia, Uppsala, Sweden). Before using, PhastGels IEF 4-6.5 were incubated overnight in a 100 ml 8M urea (BRL, Gaithersburg, USA) solution containing 1% Triton X-100 (Serva, Heidelberg, FRG). After incubation the gels were soaked for 15 min in a 8M urea solution containing 0.8% Triton X-100 and 16% carrier ampholytes, pH ranges 4.2 - 4.9; 4.5 - 5.4 (Pharmacia, Uppsala, Sweden) and 3.5 - 5 (LKB, Bromma, Sweden) mixed in the ratio 1:3:1 (v/v/v). The excess of liquid on the surface of the gel was removed by compressed air.

The following focusing conditions were employed. In all cases 0.3 μl of sample per lane was applied automatically at the anodic end of the gel at 2000 V, 25 mA, 2.0 W, 20° C for 40 Vh; final focusing was at 2000 V, 25 mA, 4.0 W, 20° C for 540 Vh.

The gel was stained automatically in the Development Unit of PhastSystem. The development conditions are noted below: Each gel was fixed for 10 min in a 20% trichloroacetic acid (Merck) solution at 20° C and washed for 2 min in a 30% methanol and 10% acetic acid (p.a Merck) solution at 20° C. Staining took place in a 0.03% PhastGel Blue R (Coomassie R-350) (Pharmacia, Uppsala, Sweden), 30% methanol, 10% acetic acid and 0.1% (w/v) $CuSO_4$ solution for 10 min at 37° C. Destaining was carried out in 30% methanol and 10% acetic acid solution for 25 min at 37° C.

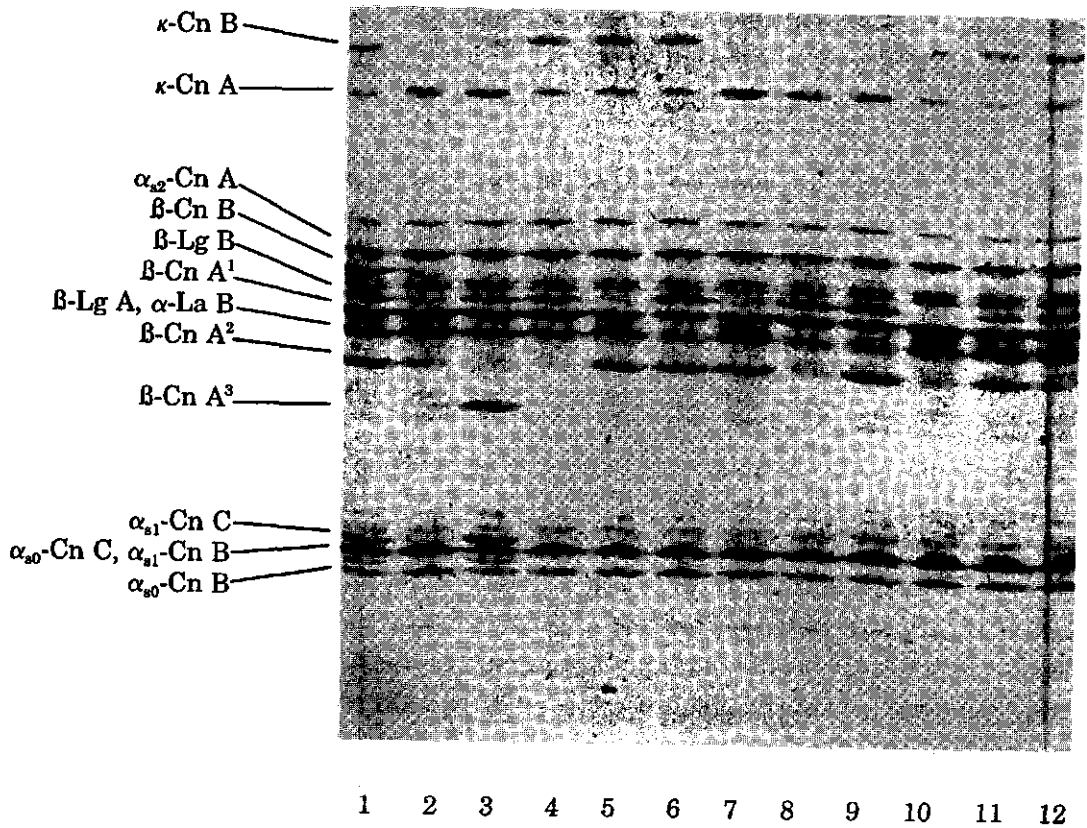
RESULTS

Identification of protein fractions was carried out by running side by side milk samples and purified α-Cn, ß-Cn, κ-Cn, ß-Lg and α-La (Sigma, St. Louis, MO, USA). The genetic variants were identified by comparing results with Seibert et al. (1985). The separation pattern of milk samples of 12 different cows is shown in figure 1. The genetic variants formed sharp distinct bands except for α-lactalbumin B and ß-lactoglobulin A variants which focused together.

DISCUSSION

In Western dairy breeds α_{s1}-Cn B and C, α_{s2}-Cn A, ß-Cn A^1, A^2 and B, κ-Cn A and B, ß-Lg A and B and α-La B are the most common genetic variants of the different milk protein fractions (Li and Gaunt, 1972). All these variants could be distinguished in one run with the method described except the α-La B.

The starting point of the analyses was the information on preparation of urea



κ-Cn	AB	AA	AA	AB	AB	AB	AA	AA	AA	AB	AB	AB
ß-Cn	$A^{1,2}B$	A^1A^1	A^1A^3	A^1A^1	A^1A^2	A^1A^2	A^1A^2	A^1A^1	A^1A^2	A^1A^1	A^1A^2	A^1A^1
α_{s2}-Cn	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA
α_{s1}-Cn	BC	BB	BC	BB	BB	BB	BB	BB	BB	BB	BB	BB
ß-Lg	AB	AA	AB	AB	AA	BB	AA	BB	BB	AA	AA	AB
Milk Sample	1	2	3	4	5	6	7	8	9	10	11	12

Figure 1. Separation patterns of a mixed milk sample (1) and 11 other whole milk samples after isoelectric focusing.

modified gels, separation and development conditions provided by PhastSystem instruction manual.

Preparation of urea modified PhastGels was first done according to the instruction manual. This implied the consecutive soaking of four gels in a urea-

ampholyte mixture. In our hands this procedure gave good results only for the first PhastGel. The next gels showed diffuse bands for the ß-lactoglobulin variants. An explanation of this phenomenon was not found. Probably some unknown component(s) present in the gel diffused out of the gel and changed the composition of the urea-Pharmalyte mixture. Soaking the gels overnight in a 100 ml solution of 8M urea + 1% Triton X-100 before soaking in the urea-Pharmalyte mixture gave a good separation of ß-lactoglobulin variants. In this way diffuse components present in the gel are washed out and urea together with the correct mixture of ampholytes can enter the gel.

After soaking the gel in the urea solution excess of liquid should be removed. Blotting the gel with nitrocellulose damaged the gel in some cases. A better approach was to blow off the excess of liquid by compressed air.

Fixation of the PhastGel was first done at 20°C, staining and destaining at 50°C. It appeared that after fixation the κ-casein was visible as a white band but disappeared during staining. After lowering the staining and destaining temperature from 50°C to 37°C this problem was solved. It was concluded that κ-casein was washed out from the gel when the temperature was too high.

Separation of genetic variants of milk protein fractions by isoelectric focusing with Pharmacia PhastSystem gave good results. Separation time was about 40 min, development of the gels took 45 min. With twelve milk samples on one gel it is possible for one person to analyse 240 milk samples a day. This high capacity is mainly due to the ready-to-use PhastGels and the short separation time. Compared to other methods (Seibert et al., 1985, Bech and Munk, 1988) prefocusing is not necessary and no electrode wicks nor electrode fluid are used. Cathodic drift, which is a frequently observed problem with IEF (Bech and Munk, 1988), was minimal. In our laboratory thousands of milk samples have been analysed successfully using this method.

REFERENCES

Bech, A. M. and K. S. Munk. 1988. Studies on bovine milk protein polymorphism by electrofocusing in agarose gels containing 7 M urea. Milchwissenschaft 43: 230.

Eigel, W. N., J. E. Butler, C. A. Ernstrom, H. M. Farrel Jr, V. R. Harwalker, R. Jenness and R. Mcl. Whitney. 1984. Nomenclature of proteins of cow's milk: fifth revision. J. Dairy Sci. 67: 1599.

Li, F. H. F. and S. N. Gaunt. 1972. A study of genetic polymorphisms of milk ß-lactoglobulin, α_{s1}-casein, ß-casein and, κ-casein in five dairy breeds. Biochem.

genetics 6: 9.

McLean, D.M., E.R.B. Graham, R.W. Ponzoni and H.A. McKenzie. 1987. Effects of milk protein genetic variants and composition on heat stability of milk. J. Dairy Res. 54: 219.

Schaar J., B. Hansson, and H. -E. Pettersson. 1985. Effects of genetic variants of κ-casein and ß-lactoglobulin on cheesemaking. J. Dairy Res. 52: 429.

Seibert, B., G. Erhardt and B. Senft. 1985. Procedure for simultaneous phenotyping of genetic variants in cow's milk by isoelectric focusing. Animal blood groups and Biochem. Genetics 16: 183.

Swaisgood, H. E. (ed.), Methods of gel electrophoresis of milk proteins, American Dairy Science Association, (1975) p. 1-30.

Chapter 4

ESTIMATION OF MILK PROTEIN GENE FREQUENCIES IN CROSSBRED CATTLE BY MAXIMUM LIKELIHOOD METHOD

Henk Bovenhuis and Johan A.M. van Arendonk
Department of Animal Breeding
Wageningen Agricultural University
P.O. Box 338
6700 AH Wageningen
The Netherlands

Published in: Journal of Dairy Science 74: 2728 (1991)

ABSTRACT

A maximum likelihood method is presented to estimate the fraction of animals misclassified and breed effects for milk protein gene frequencies based on crossbred data. A simulation study indicates that the method provides estimates of gene frequencies that agree closely with the true values. Gene frequencies in the Dutch Black and White and the Dutch Red and White crossbred populations, based on data on 10,151 and 580 animals respectively, were estimated. Dutch Friesian and Holstein Friesian breeds differ in gene frequencies for β-casein and β-lactoglobulin. Estimates for fractions misclassified are zero for α_{s1}-casein, .09 for β-casein and β-lactoglobulin and .12 for κ-casein. Differences between Dutch Red and Whites and Red Holstein Friesian breeds are small, and estimates for fractions misclassified are high but have high approximate standard errors. Compared with the Black and White breeds, the Red and Whites have a high κ-casein B gene frequency.

Key words: milk protein, gene frequencies, maximum likelihood, crossbred.

INTRODUCTION

Genetic polymorphism have been detected in all major milk proteins of dairy cows (Eigel et al., 1984). Genetic variants of some milk proteins show relationships with manufacturing properties of milk, especially for cheese (reviewed by Grosclaude, 1988). Therefore, it is important to know frequencies of milk protein alleles. Many studies have reported milk protein gene frequency estimates (Graml et al., 1984[a], Graml et al., 1984[b], Grosclaude, 1988, Hines et al., 1977, McLean et al., 1984, Ng-Kwai-Hang et al., 1984, Schmidt, 1966). Most studies used the method of gene counting to estimate gene frequencies, because all milk proteins are controlled by codominant genes and genotypes can be determined by phenotyping milk samples. This method does not account for relations between animals (Cotterman, 1954). When estimating gene frequencies from a sample containing related individuals, however, some genes are counted repeatedly. In situations of large numbers of randomly sampled family groups this unequal weighting of observations would not be expected to bias gene frequency estimates but would bias the precision of the estimate (Cotterman, 1954). If the sample of animals that are analyzed, is selected so that a few sires have large progeny groups and so that the counting of alleles is as if they were

independent, estimates are expected to be biased. Bias in estimating gene frequencies can be avoided by accounting for relations between animals due to a common sire (Cotterman, 1954).

Since 1980, semen from Holstein Friesian sires has been used extensively in the Dutch Black and White population and in recent years in the Dutch Red and White population although to a lesser extent. Gene frequencies for milk protein variants might differ between Holstein Friesian (HF) and Dutch breeds, Dutch Friesians (DF) and Dutch Red and Whites, Meuse-Rhine Yssel (MRY). These gene frequencies are expected to differ more between breeds than between strains of the same breed (Nei, 1987). In this paper, cattle populations of different origins will be referred to as breeds, and the term crossbred will be used to indicate a cross of Holstein Friesian and Dutch cattle.

To examine the extent to which gene frequencies are influenced by crossbreeding, gene frequencies can be estimated for groups of animals with similar breed composition, i.e. the estimation of gene frequencies in groups of animals carrying for example 0, 50 and 100% genes of an immigrant breed (e.g., Graml et al., 1984[b]). This method uses the available information in an inefficient way, however, because estimates within the different groups are not combined to obtain a gene frequency estimate for the immigrant and the native breed, and there might be genetic links between groups.

The procedure for collecting, analyzing and phenotyping samples implies the possibility of misclassification of an individual's genotype, which might affect estimates of gene frequencies. To overcome this problem, it is possible to phenotype samples repeatedly (e.g., Graml et al., 1984[a], Ng-Kwai-Hang et al., 1984). Misclassifications due to errors in pedigree, however, cannot be eliminated by repeated typing. Also, if large numbers of samples are involved, then repeated typing is cumbersome. Information on the fraction misclassified is valuable when data are used for other analyses such as studies to examine linkage between milk protein genes or to estimate relationships between milk protein variants and milk production traits.

In this paper, a maximum likelihood model to estimate gene frequencies in a segregating crossbred population is described. The method accounts for relations between animals due to a common sire, for breed differences in gene frequencies, and for the probability of misclassification of genotypes. Estimates of gene frequencies for the breeds represented in the population and of the probability of misclassification for the population are provided by the method. A simulation study was carried out to verify the method. Milk protein gene frequencies are estimated for the Dutch Black and White and the Dutch Red and White crossbred

cow populations to reveal the influence of crossbreeding Dutch breeds with Holstein Friesians on milk protein gene frequencies. Results of the maximum likelihood model are compared with results obtained by the method of gene counting.

MATERIALS AND METHODS

Data

The Dutch Black and White crossbred cow population will be further referred to as the Dutch Black and White population and consists of Dutch Friesians with varying amounts of Holstein Friesian genes. For the Dutch Black and White population, heifers of 39 proven bulls and 79 young bulls were selected from the database of the milk recording service of the Royal Dutch Cattle Syndicate. The proven bulls had on average 153 daughters (with a standard error of 170): the young bulls had on average 56 daughters (with a standard error of 19). Data were collected to analyze relationships between milk protein variants and milk production traits. To adjust records for herd-year-season (HYS) effects, HYS classes with at least one daughter of a young bull and in total at least three daughters of the selected proven or young bulls were required; a year consisted of two seasons. Milk samples of 10,151 first lactation animals were collected routinely by milk recording associations on 2618 herds. Pedigree information and information on the proportion of HF genes of heifers, sires and dams were obtained from the registration files of the Royal Dutch Cattle Syndicate. The average fraction of HF genes of the heifers was 0.63, whereas this fraction was 0.35 for dams and 0.87 for sires.

The Dutch Red and White crossbred cow population consists of MRY cattle with varying amounts of Red Holstein Friesian (RHF) genes. The material of the Dutch Red and White population consisted of 580 heifers of 41 young bulls. On average 14 daughters (with a standard error of 6) of each sire were randomly selected. Milk samples were collected routinely by milk recording associations on 520 herds. Pedigree information and information on the proportion of RHF genes of heifers, sires and dams were obtained from the registration files of the Royal Dutch Cattle Syndicate. The average fraction of RHF genes was 0.28, whereas this fraction was 0.20 for dams and 0.40 for sires.

Milk protein genetic variants were phenotyped using isoelectric focusing (Bovenhuis and Verstege, 1989).

Maximum likelihood model

Purebred population, accounting for misclassifications.

The probability for an individual's genotype includes a misclassified fraction (M), and a correctly classified fraction (1-M). The distribution of the correctly classified fraction is conditional on the sire's genotype. For misclassification, however, the probabilities of genotypes are assumed to be the population genotype frequencies. This means that in the case of misclassification the genotype of the individual is assumed to be replaced by the genotype of an average animal. When the population is in Hardy-Weinberg equilibrium, population genotype frequencies are p^2, 2pq and q^2, where p and q are the probabilities of the occurrence of alleles A and B. The probability of the offspring given the genotype of sire s (P_s(O | genotype of sire)) for a single locus with two alleles (A and B) can be written as:

$$P_s(O \mid AA) = C \cdot [Mp^2 + (1-M)p]^x \cdot [2Mpq + (1-M)q]^y \cdot [Mq^2]^z$$

$$P_s(O \mid AB) = C \cdot [Mp^2 + \tfrac{1}{2}(1-M)p]^x \cdot [2Mpq + \tfrac{1}{2}(1-M)]^y \cdot [Mq^2 + \tfrac{1}{2}(1-M)q]^z$$

$$P_s(O \mid BB) = C \cdot [Mp^2]^x \cdot [2Mpq + (1-M)p]^y \cdot [Mq^2 + (1-M)q]^z$$

where C = N! / (x! y! z!); x, y and z are the number of offspring with genotype AA, AB and BB; N is the total number of offspring of sire s.

The prior probabilities of a sire being AA, AB or BB are p^2, 2pq or q^2, respectively, when assuming Hardy-Weinberg equilibrium. By combining the prior probability, with the probability of the offspring given the genotype of the sire we obtain the likelihood for sire s with unknown genotype (L_s).

$$L_s = p^2 P_s(O \mid AA) + 2pq P_s(O \mid AB) + q^2 P_s(O \mid BB)$$

The likelihood for the complete data set is obtained by multiplication of the likelihood functions of the n_s sires, assuming sires to be independent with respect

to their genotypes:

$$L = \prod_{s=1}^{n_s} L_s \qquad (5)$$

Including breed effects.

When an animal is correctly classified, breed effects on gene frequencies were incorporated by replacing p by

$$p = (1\text{-}HF_j)p_n + HF_j p_i$$

where

p_n = the gene frequency of allele A in the native population (DF or MRY),

p_i = the gene frequency of allele A in the immigrant population (HF or RHF), and

HF_j = the fraction Holstein Friesian genes (HF or RHF) of animal j.

For the correctly classified fraction in equations (1), (2) and (3), p and q are the probabilities of a heifer of obtaining an A or B allele from her dam. These probabilities depend on the breed composition of the dam, and HF_j, therefore, refers to the fraction Holstein Friesian genes in the dams.

The genotypes for the fraction misclassified in equations (1), (2) and (3) are assumed to be distributed according Hardy-Weinberg equilibrium. When gene frequencies differ between the native and the immigrant breed, however, differences in breed composition between sires and dams will cause different gene frequencies in the sire and dam populations. This will cause a departure from Hardy-Weinberg equilibrium in the daughter population. Therefore in equations (1), (2) and (3) p^2, 2pq, and q^2 are replaced by $p_s p_d$, $(p_s(1\text{-}p_d)+(1\text{-}p_s)p_d)$ and $(1\text{-}p_s)(1\text{-}p_d)$, where p_s and p_d are the frequencies of the A allele in the sire and dam populations. Subsequently p_s was replaced by $(1\text{-}HF_{\overline{sires}})p_n + HF_{\overline{sires}}p_i$, where $HF_{\overline{sires}}$ is the weighted (by number of progeny) Holstein Friesian fraction of the sire population; p_d was replaced by $(1\text{-}HF_{\overline{dams}})p_n + HF_{\overline{dams}}p_i$, where $HF_{\overline{dams}}$ was the average Holstein Friesian fraction of the dam population. After substitution, equation (1), (2) and (3) can be written as:

$$P_s(O \mid AA) = C \cdot \prod_{j=1}^{x}\left[M p_s p_d + (1-M)\{(1-HF_j)p_n + HF_j p_i\}\right]$$

$$\cdot \prod_{j=1}^{y}\left[M(p_s(1-p_d)+(1-p_s)p_d) + (1-M)\{1-((1-HF_j)p_n + HF_j p_i)\}\right] \quad (6$$

$$\cdot \prod_{j=1}^{z}\left[M(1-p_s)(1-p_d)\right]$$

$$P_s(O \mid AB) = C \cdot \prod_{j=1}^{x}\left[M p_s p_d + \tfrac{1}{2}(1-M)\{(1-HF_j)p_n + HF_j p_i\}\right]$$

$$\cdot \prod_{j=1}^{y}\left[M(p_s(1-p_d)+(1-p_s)p_d) + \tfrac{1}{2}(1-M)\right] \quad (7$$

$$\cdot \prod_{j=1}^{z}\left[M(1-p_s)(1-p_d) + \tfrac{1}{2}(1-M)\{1-((1-HF_j)p_n + HF_j p_i)\}\right]$$

$$P_s(O \mid BB) = C \cdot \prod_{j=1}^{x}\left[M p_s p_d\right]$$

$$\cdot \prod_{j=1}^{y}\left[M(p_s(1-p_d)+(1-p_s)p_d) + (1-M)\{(1-HF_j)p_n + HF_j p_i\}\right] \quad (8$$

$$\cdot \prod_{j=1}^{z}\left[M(1-p_s)(1-p_d) + (1-M)\{1-((1-HF_j)p_n + HF_j p_i)\}\right]$$

where: M = *misclassified fraction*

p_n = *gene frequency of allele A in the native population*

p_i = *gene frequency of allele A in the immigrant population*

$p_d = (1-HF_{\overline{dams}})p_n + HF_{\overline{dams}}\, p_i$

$p_s = (1-HF_{\overline{sires}})p_n + HF_{\overline{sires}}\, p_i$

HF_j = *fraction of Holstein Friesian genes of dam j*

$HF_{\overline{dams}}$ = *average Holstein Friesian fraction of dams*

$HF_{\overline{sires}}$ = *average(weighted by number of progeny) Holstein Friesian fraction of sire*

x = *number of offspring of sire s having genotype AA*

y = *number of offspring of sire s having genotype AB*

z = *number of offspring of sire s having genotype BB*

C = *a constant*

The HF $\overline{\text{sires}}$ and HF $\overline{\text{dams}}$ were respectively .91 and .35 for the Dutch Black and White and .35 and .20 for the Dutch Red and White population. Note that (.91+.35) / 2 and (.35+.20) / 2 are respectively .63 and .28 which are the average fractions of HF and RHF genes of the heifers.

After substitution of p by $(1\text{-}HF_s)p_n + HF_s p_i$, equation (4) can be written as:

$$\begin{aligned} L_s = \quad & [(1-HF_s)p_n+HF_s p_i]^2 \; P_s(O \mid AA) \\ + \; & 2[(1-HF_s)p_n+HF_s p_i][1-((1-HF_s)p_n+HF_s p_i)] \; P_s(O \mid AB) \\ + \; & [1-((1-HF_s)p_n+HF_s p_i)]^2 \; P_s(O \mid BB) \end{aligned} \qquad (9)$$

where HF_s is the HF fraction of sire s.

Estimates of p_n, p_i and M were obtained by maximization of the likelihood function with respect to these parameters. For a situation with more then two alleles the number of possible genotypes and the number of gene frequencies increases. The approach, however, is similar and equations 6, 7, 8 and 9 can be expanded easily to account for this.

Maximization of the likelihood function was by the simplex method (Nelder and Mead, 1965). Convergence was declared when differences between simplex points were less than .0001. Approximate standard errors were calculated using a quadratic approximation of the likelihood function based on simplex points close to the maximum (Nelder and Mead, 1965). The points used to determine quadratic approximation were chosen in such a way that the approximated maximum of the likelihood function and the corresponding approximated parameters agreed closely with the estimated values. These points, rather than those in the final simplex, were used to obtain a better approximation of the likelihood function.

Simulation

A simulation study was carried out to verify the estimation procedure. For a single locus with two alleles, A and B, a crossbred population, reconstituted from a native and an immigrant breed differing in gene frequencies, was simulated. Dams and sires had 0, 50, or 100% of their genes from the immigrant breed. Two population structures were distinguished, one to reflect the Dutch Black and White data (I) and the other to reflect the Dutch Red and White data (II). In population structure I, ninety percent of the sires were purebred immigrant (class 100%), whereas the remaining ten percent were crossbred (class 50%). Of

the dams, 40% were purebred native (class 0%), 50% were crossbred (class 50%), and 10% were purebred immigrant (class 100%). One third of the 110 sires simulated had 150 daughters, whereas the remainder had 50 daughters. The total number of animals simulated was 9200.

In population structure II, 40, 40, and 20% of the sires were in classes 0, 50 and 100%. Of the dams, 65, 30, and 5% were in class 0, 50 and 100%, respectively. The 40 sires each had 14 daughters; thus, the total number of animals simulated was 560.

With respect to the gene frequencies of the native and immigrant breeds, three different data sets were simulated: 1, 2 and 3. Gene frequencies of the native and immigrant breed were .80 and .90 for data set 1, .50 and .60 for 2 and .10 and .90 for data set 3. In each data set, the fraction misclassified was .08 and simulated to be distributed according the population genotype frequencies, i.e., $p_s p_d$ for AA, $(p_s(1-p_d)+(1-p_s)p_d)$ for AB and $(1-p_s)(1-p_d)$ for BB genotypes.

Empirical standard errors were calculated as sample standard deviations from the replicate estimates to check the method used to approximate standard errors.

RESULTS

Simulation

Table 1 shows the results of the simulation study. Estimates of gene frequencies, the fraction misclassified and the approximated standard errors (A-SE) were averages over 15 replicates, whereas the empirical standard errors (E-SE) were sample standard deviations over the 15 replicates. The method proposed to estimate gene frequencies in two breeds based on crossbred data provides estimates of the gene frequencies that agree closely with their true values; differences between the average values of the estimates and the true values were between -.0040 and .0060 for population structure I and between -.0070 and .0059 for population structure II. Gene frequency estimates for the immigrant breed differ more from their true values than those for the native breed. This is caused by the smaller fraction of immigrant genes than native genes present in the simulated populations. Gene frequency estimates seem not to be influenced by the differences in structure of the data (population structure I versus II).

Estimates of M agreed well with their true values when gene frequencies of the population were intermediate (data set 2 and 3). At extreme gene frequencies (data set 1) differences between the true and the estimated fraction misclassified

Table 1. Maximum likelihood estimates of gene frequencies[1], p_n and p_i, for two breeds involved in a crossbred and the fraction misclassified (M) with empirical[2] (E-SE) and approximated[1] (A-SE) standard errors for two population structures.

		Population structure I[3]			Population structure II[4]		
	True	Estimate	E-SE	A-SE	Estimate	E-SE	A-SE
				Data set 1			
p_n	.80	.7999	.0088	.0099	.7993	.0190	.0358
p_i	.90	.8960	.0092	.0117	.8970	.0550	.0604
M	.08	.0658	.0208	.0262	.1239	.0997	.1371
				Data set 2			
p_n	.50	.4966	.0114	.0183	.5059	.0313	.0620
p_i	.60	.6060	.0114	.0239	.5959	.0548	.1206
M	.08	.0811	.0096	.0182	.0650	.0324	.1108
				Data set 3			
p_n	.10	.0999	.0055	.0116	.1007	.0178	.0354
p_i	.90	.8968	.0132	.0175	.8930	.0398	.0718
M	.08	.0754	.0122	.0202	.0896	.0288	.0646

[1] Average of 15 replicates. Frequencies for native (p_n) and immigrant (p_i) populations.
[2] Sampling standard deviations based on estimates of 15 replicates.
[3] Population structure I: 9200 animals, 37 sires with 150 daughters, and 73 sires with 50 daughters.
[4] Population structure II: 560 animals and 40 sires with 14 daughters.

were larger. Differences between the average values of the estimates and the true values were between -.0142 and .0011 for population structure I. For population structure II these differences were larger; between -.0150 and .0439.

Empirical standard errors were smaller than the approximated standard errors. This was especially the case for population structure II, indicating that the approximation of the standard error was less accurate for this situation. The

smaller number of animals involved under population structure II caused increased standard errors. Standard errors for the estimates of M were high for data set 1, especially under population structure II.

Field data

Table 2 shows the milk protein gene frequency estimates for the Dutch Black and White and the Dutch Red and White populations. In the Dutch Black and White population small differences in α_{s1}-casein gene frequencies were observed between the DF and the HF breed. The imported Holstein Friesian population tended to have a higher α_{s1}-casein C gene frequency. Larger differences between the two breeds were found for ß-casein gene frequencies. In the Dutch Friesians the A^1 allele is predominant (.766) and the A^2 allele appears at a relatively low frequency (.147), whereas in the Holstein Friesians the A^1 and A^2 alleles appear at about equal frequencies. Further, the B-allele of ß-casein is more frequent in the Dutch Friesian breed (.073) than in the Holstein Friesian breed (.026). There are small differences in gene frequency estimates for κ-casein between Dutch Friesians and imported Holstein Friesians. The ß-lactoglobulin B allele has a higher frequency (.575) in the imported Holstein Friesian population than in the Dutch Friesian population (.475). Estimates for fractions misclassified are zero for α_{s1}-casein, .09 for ß-casein and for ß-lactoglobulin, and .12 for κ-casein.

Differences between gene frequency estimates for MRY and RHF were small; however, estimates have high approximated standard errors, especially for the RHF population. Differences between DF and MRY were small for most gene frequencies. As in the Dutch Friesian breed, the ß-casein A^1 allele is predominant in the MRY breed. The ß-casein B and κ-casein A alleles appear at a lower frequency (.009 and .490) in the MRY breed then in the DF breed (.073 and .831). Further, in the Dutch Red and White population the rare κ-casein C allele, reported by Erhardt (1989) in the German Brown Swiss and Simmental breeds, was observed. Estimates for fractions misclassified in the Dutch Red and Whites are high but have high approximated standard errors.

DISCUSSION

Milk protein gene frequency estimates, reported by Schmidt (1966), were higher for κ-casein B in the DF population (.34) and higher for ß-casein B frequency in the MRY population (.08). The small number of animals in the experiment of Schmidt (1966), however, might partly explain the observed differences. Gene

Table 2. Maximum Likelihood estimates of milk protein gene frequencies and fractions misclassified (M) (and approximate standard errors) for the Dutch Black and White and the Dutch Red and White population.

	Dutch Friesians	Imported Holstein Friesian	MRY[1]	Imported Red Holstein Friesian
Allele		α_{s1}-CN[2]		
B	.980 (.006)	.950 (.005)	1.000 (.070)	.936 (.167)
C	.020 (.006)	.050 (.005)	.000 (.070)	.064 (.167)
M	.000 (.122)		.797 (.243)	
Allele		ß-CN		
A^1	.766 (.018)	.462 (.027)	.794 (.064)	.661 (.152)
A^2	.147 (.016)	.498 (.027)	.193 (.063)	.284 (.146)
A^3	.014 (.003)	.014 (.004)	.004 (.016)	.024 (.043)
B	.073 (.010)	.026 (.011)	.009 (.010)	.031 (.033)
M	.088 (.017)		.196 (.153)	
Allele		κ-CN		
A	.831 (.012)	.847 (.015)	.490 (.073)	.512 (.159)
B	.169 (.012)	.153 (.015)	.510 (.082)	.468 (.174)
C	....	....	.000 (.093)	.020 (.114)
M	.120 (.020)		.129 (.106)	
Allele		ß-LG		
A	.525 (.019)	.425 (.025)	.484 (.081)	.377 (.173)
B	.475 (.019)	.575 (.025)	.516 (.081)	.623 (.173)
M	.086 (.018)		.156 (.108)	

[1]MRY = Meuse-Rhine-Yssel cattle.
[2]CN = Casein; LG = lactoglobulin.

frequencies estimated for the Holstein Friesian population are comparable with those of Hines et al. (1977).

Bovenhuis et al. (1990) presented a method by which the average fraction of Holstein Friesian genes of dams instead of Holstein Friesian fractions of individual dams, was used to estimate breed effects. In this case estimates of breed effects depend on differences between the average fraction of Holstein Friesian genes of dams between sires. If sires are mated at random to dams with respect to their Holstein Friesian fraction, these differences are likely to disappear with large progeny groups. The method presented in the present study is more accurate because it accounts for information on the fraction of Holstein Friesian genes of individual animals. Gene frequency estimates presented earlier for the Dutch Black and White population by Bovenhuis et al. (1990), differ only slightly (.01-.03) from results in Table 2.

Bovenhuis et al. (1990) assumed the fraction of animals misclassified to be distributed according to Hardy-Weinberg equilibrium. Differences in gene frequencies between breeds and different breed compositions of sires and dams, however, will cause an excess of heterozygotes which will lead to underestimation of the fraction misclassified. In the Black and White data, a departure from Hardy-Weinberg equilibrium was observed for the ß-casein and ß-lactoglobulin genotypes. Estimates for the fraction misclassified reported by Bovenhuis et al. (1990) were .080 for ß-casein and .083 for ß-lactoglobulin which are slightly lower than the values in Table 2.

Results in Table 1 show that inaccurate estimates of the fraction misclassified were obtained when gene frequencies were extreme. One reason for this is that a fraction of the animals misclassified by chance obtained the correct genotype. This is the case if a change of milk samples occurs but the exchanged milk sample has the same genotype as the original milk sample. It can be shown that the correctly classified proportion of M, when assuming Hardy-Weinberg equilibrium for a gene with two alleles, is $p^4 + 4p^2q^2 + q^4$. This fraction should be subtracted from M to obtain the effective fraction misclassified, i.e., the fraction misclassified that leads to a change of genotype of the individual. Figure 1 shows that the effective fraction misclassified is small at extreme gene frequencies, causing difficulties in estimating M. More accurate estimates of M were obtained at intermediate gene frequencies, which is where the effective fraction misclassified is maximal. Another reason for inaccurate estimates of M at extreme frequencies is that most of the information about M comes from the "visible part" of M, which includes the animals that can be eliminated on pedigree, i.e., AA daughters of a BB sire or BB daughters of an AA sire. The

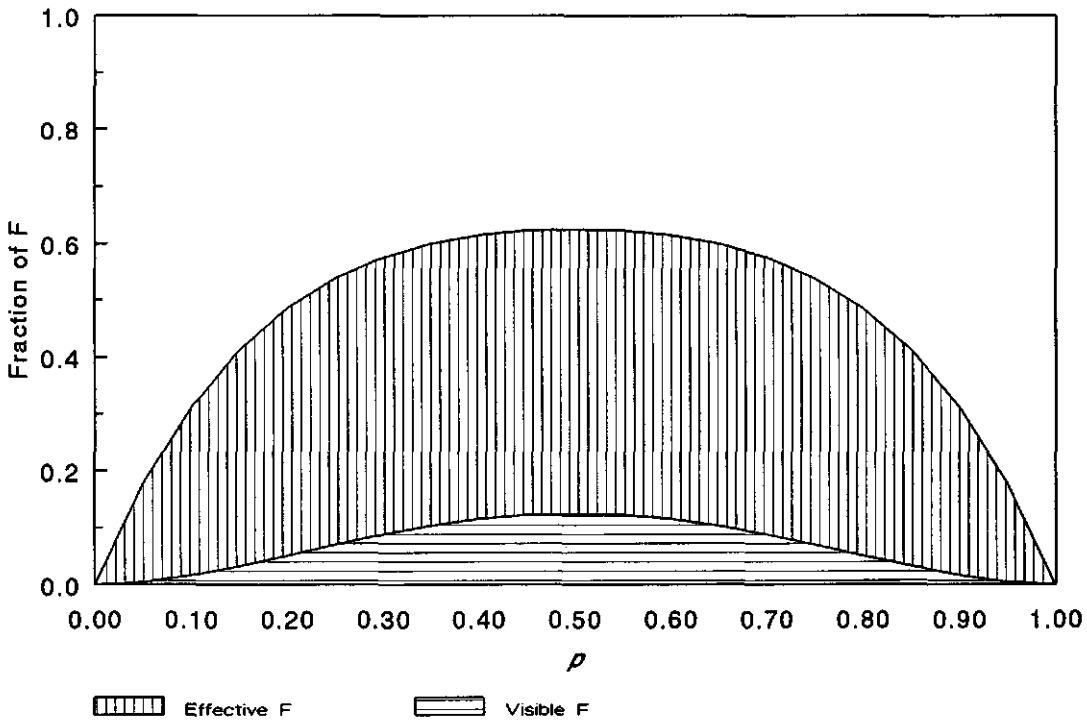


Figure 1. The effective fraction misclassified (M) and the misclassified animals that can be eliminated based on their pedigree for a locus with two alleles in a population in Hardy-Weinberg equilibrium.

number of animals that can be eliminated on pedigree, i.e., $2p^2q^2$ assuming Hardy-Weinberg equilibrium for a gene with two alleles, is small for extreme frequencies (Figure 1). For these reasons it is difficult to obtain accurate estimates for M at extreme gene frequencies, as is the case for α_{s1}-casein. In general, the effective fraction misclassified and the visible part of M will be increased for a gene with more than two alleles, giving more accurate estimates of M.

The fraction misclassified is assumed to be distributed according population genotype frequencies. To test this assumption, 95 milk samples of Black and White cattle were collected and analyzed twice. The number of animals that had different genotypes on the first and second analysis, were 1 for α_{s1}-casein, 8 for ß-casein, 11 for κ-casein and 9 for ß-lactoglobulin. The number of animals misclassified for more than one genotype are the minimum number of animals subject to errors in pedigree or changing of milk samples. It seems reasonable to

Table 3. Estimates of gene frequencies by maximum likelihood (ML) and by the method of gene counting for the Dutch Black and White and for the Dutch Red and White population.

	Dutch Black and White		Dutch Red and White	
Alleles	ML[1]	Gene counting	ML[2]	Gene counting
		α_{s1}-CN[2]		
B	.961	.982	.982	.982
C	.039	.018	.018	.018
		ß-CN		
A^1	.575	.560	.757	.751
A^2	.368	.353	.218	.234
A^3	.014	.008	.010	.006
B	.043	.079	.015	.010
		κ-CN		
A	.841	.805	.496	.492
B	.159	.195	.498	.505
C	...	...	.006	.003
		ß-LG		
A	.462	.444	.454	.446
B	.538	.556	.546	.554

[1] Calculated from Table 2 : .371 p_{DF} + .629 p_{HF}; DF = Dutch Friesian, HF = Holstein Friesian.
[2] Calculated from Table 2 : .724 p_{MRY} + .276 p_{RHF}; MRY = Meuse-Rhine-Yssel; RHF = Red Holstein Friesian.
[3] CN = Casein, LG = lactoglobulin.

assume that if these types of errors occur, then the misclassified fraction is distributed according to the population genotype frequencies. This assumption might be incorrect, however, when systematic errors occur, due to the method used for analyzing samples, for example. The maximum number due to

systematic errors is assumed to be equal to the number of animals misclassified for one genotype only i.e., 0 for α_{s1}-casein, 2 for ß-casein, 7 for κ-casein and 3 for ß-lactoglobulin. This indicates that, for κ-casein more than for the other milk protein genes, errors might be due to the method of analysis. Sometimes it was difficult to type samples for κ-casein.

The method of gene counting does not account for relations between animals due to a common sire. In the Black and White data, a few sires have large progeny groups, which might influence gene counting estimates of frequencies. Table 3 shows the gene counting and the calculated maximum likelihood estimates. Maximum likelihood gene frequency estimates were calculated from estimates given in Table 2 with HF=.629 for Black and White and RHF=.276 for Red and White cattle. Differences between the two estimates for the Black and White population can be observed for the ß-casein B gene frequency (.043 versus .079), the κ-casein (.841 versus .805) and ß-lactoglobulin (.462 versus .444) gene frequencies. A possible cause for the observed differences in gene frequencies is the variation in numbers of progeny.

CONCLUSIONS

Compared with the method of gene counting, the maximum likelihood method offers the framework for estimating additional parameters. The maximum likelihood method was used to estimate the fraction of animals misclassified and of breed effects on milk protein gene frequencies. Results of the simulation study indicate that the method presented provides a powerful tool to estimate differences in gene frequencies between breeds based on crossbred data. Accurate estimates of the fraction misclassified can be obtained at intermediate gene frequencies.

ACKNOWLEDGMENTS

The authors wish to thank Esther Verstege for phenotyping the milk samples and the Royal Dutch Cattle Syndicate for supplying the data and financial support.

REFERENCES

Bovenhuis, H. and A.J.M. Verstege 1989. Improved method for phenotyping milk protein variants by isoelectric focusing using PhastSystem. Neth. Milk

Dairy J. 43: 447.

Bovenhuis, H., S. Korver and A.J.M. Verstege 1990. Genetic polymorphism of milk protein variants in crossbred populations. 4th World Congress on Genetics Applied to Livestock Production, Edinburgh 14: 136.

Cotterman, C.W. 1954. Estimation of gene frequencies in nonexperimental populations. *In* Statistics and mathematics in biology. O. Kempthorne, T.A. Bancroft, J.W. Gowen and J.L. Lush, ed. The Iowa State College Press, Ames, Iowa.

Eigel, W.N., J.E. Butler, C.A. Ernstrom, H.M. Farrel Jr, V.R. Harwalker, R. Jenness and R. Mcl. Whitney, 1984. Nomenclature of proteins of cow's milk: fifth revision. J. Dairy Sci. 67: 1599.

Erhardt, G. 1989. κ-Kaseine in Rindermilch-Nachweis eines weiteren allels (κ-Cn^{E}) in verschiedenen rassen. J. Anim. Breed. Genet. 106: 225.

Graml, R., J. Buchberger, O. Kirchmeier, F. Kiermeier and F. Pirchner, 1984[a]. Genfrequenzschätzung bei Milchproteinen des bayerischen Fleckviehs. Züchtungskunde 56: 73.

Graml, R., J. Buchberger, H. Klostermeyer and F. Pirchner, 1984[b]. Untersuchungen über die Genfrequenzen der Caseine und ß-Lactoglobuline bei der bayerischen Braunviehpopulation. Züchtungskunde 56: 221.

Grosclaude, F. 1988. Le polymorphisme génétique des principales lactoprotéines bovines. INRA Prod. Anim. 1: 5.

Hines, H.C., G.F.W. Haenlein, J.P. Zikakis and H.C. Dickey, 1977. Blood antigen, serum protein, and milk protein gene frequencies and genetic interrelationships in Holstein cattle. J. Dairy Sci. 60: 1143.

McLean, D.M., E.R.B. Graham, R.W. Ponzoni and H.A. McKenzie, 1984. Effects of milk protein genetic variants on milk yield and composition. J. Dairy Res. 51: 531.

Nelder, J.A. and R. Mead, 1965. A simplex method for function minimization. Computer J. 7: 308.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press. New York Guildford, Surrey.

Ng-Kwai-Hang, K.F., J.F. Hayes, J.E. Moxley and H.G. Monardes, 1984. Association of genetic variants of casein and milk serum proteins with milk, fat, and protein production by dairy cattle. J. Dairy Sci. 67: 835.

Schmidt, D.G. 1966. Genetische varianten van caseïne. Het vóórkomen in Nederland. Veeteelt en Zuivelberichten 9: 431.

Chapter 5

ASSOCIATIONS BETWEEN MILK PROTEIN POLYMORPHISMS AND MILK PRODUCTION TRAITS

Henk Bovenhuis, Johan A.M. van Arendonk, and Siem Korver[1]
Department of Animal Breeding
Wageningen Agricultural University
PO Box 338
6700 AH Wageningen
The Netherlands

Accepted for publication in Journal of Dairy Science

---

[1]Present address: Euribrid B.V., PO Box 30, 5830 AA Boxmeer, The Netherlands.

ABSTRACT

Associations between milk protein genotypes and milk production traits were estimated from 6803 first lactation records. Exact tests of associated hypotheses and unbiased estimates of genotype effects were from an animal model. Milk protein genotype effects were estimated using a model in which each milk protein gene was analyzed separately (single-gene analysis) and a model in which all milk protein genes were analyzed simultaneously (multigene analysis).

The results of the two models indicate that some of the effects ascribed to certain milk protein genes in the single-gene analysis are not effects of the milk protein gene itself but of linked genes. Results from this study and from literature indicate that the κ-casein gene or a very closely linked gene affects protein percentage, and the β-lactoglobulin gene or a very closely linked gene affects fat percentage. Furthermore, effects of β-casein genotypes on milk production, fat percentage, and protein yield were significant, and β-lactoglobulin genotypes had significant effects on milk production and protein yield. It is less clear whether those effects are due to effects of milk protein genes themselves or to effects of linked genes.

Key words: milk protein, genetic variants, milk production traits.

INTRODUCTION

Several authors have reported associations of milk protein genetic variants with manufacturing properties of milk (McLean et al., 1984, Schaar et al., 1985, Van den Berg et al., 1992). Research has focused on the relationship of milk protein loci with cheese production. Two milk protein genes, κ-casein and ß-lactoglobulin, have been intensively studied in this context. Several studies (reviewed by Grosclaude, 1988) indicated that κ-casein variants are associated with renneting time, whereas ß-lactoglobulin variants are associated with casein number. In both cases, the B-variants are favorable. Because of these associations, interest in selecting for favorable milk protein genotypes is considerable. This is especially true for countries in which a significant portion of the milk is manufactured into cheese. However, the economic importance of milk protein genetic variants also depends on the association of these genes with milk production traits. Therefore, before selection for milk protein genetic variants is incorporated into the breeding program, relationships between milk protein genetic variants and milk production traits should be studied.

Several studies (Aleandri et al., 1990, Gonyon et al., 1987, Graml et al., 1985, Graml et al., 1986, Haenlein et al., 1987, Ng-Kwai-Hang et al., 1984, Ng-Kwai-Hang et al., 1986) examined the effects of milk protein genetic variants on milk production traits. However, results of those studies conflict with respect to the significance and the size of genotype effects. The reason for conflicting results might be that the associations are due to the effects of linked genes rather than to the effects of the milk protein loci themselves. In this case, the associations in different populations might differ. Another reason could be the statistical model used to analyze the data. Kennedy et al. (1992) concluded that using ordinary least squares analysis when relations between animals exist results in an inflated *F* test. Consequently, an excess of spurious significant genotype effects are found when, actually, no genotype effect exists. Some studies (Aleandri et al., 1990, Ng-Kwai-Hang et al., 1984, Ng-Kwai-Hang et al., 1986) did not take relationships between animals into account. Other studies (Gonyon et al., 1987, Graml et al., 1985, Graml et al., 1986, Haenlein et al., 1987) only accounted for relationships between animals according to sires. Furthermore, Kennedy et al. (1992) showed that if the gene truly has an effect and if directional selection for the trait of interest has been practiced, then ordinary least squares estimates are potentially biased. In these situations, an animal model that treats genotype effects as fixed effects can provide an exact test of associated hypotheses and unbiased estimates of genotype effects. In most dairy cattle populations, relationships between animals exist, and selection has been practiced. Therefore, the animal model is the method of choice for estimating genotype effects.

The aim of the present study was to estimate associations between milk protein genotypes and milk production traits. For this purpose, a large data set was used. To obtain an exact test of associated hypotheses and unbiased estimates of genotype effects, an animal model was used.

MATERIALS AND METHODS

For 10,151 first lactation cows in 2618 herds, one additional milk sample was collected routinely by the Dutch milk recording associations. This additional sample was used to determine the genotype for α_{s1}-casein, α_{s2}-casein, ß-casein, κ-casein, and ß-lactoglobulin, using isoelectric focusing (Bovenhuis and Verstege, 1989). Samples were collected from February 1989 until October 1989. Milk protein gene frequencies were described by Bovenhuis and Van Arendonk (1991). Pedigree information, 305-d milk production records, and information on the proportion of Holstein-Friesian genes of heifers were obtained from the milk

Table 1. Means and standard deviations of 305-d milk production traits of 6803 first lactation cows.

Trait	$\bar{x}$	SD
Milk, kg	6138	1010
Fat, %	4.56	.44
Protein, %	3.46	.19
Fat, kg	278.5	43.2
Protein, kg	211.9	32.8

recording and registration files of the Royal Dutch Cattle Syndicate. The 305-d milk production records were based on three or four weekly test day observations for milk production, fat percentage and, protein percentage by the milk recording associations. Analysis for fat and protein content was in one laboratory by Infra Red using the Milko Scan 605 (Foss Electric, Hillerod, Denmark). Incomplete lactations of 90 d or more were extended to 305 d, taking into account the herd level, age of calving, and year-season of calving, according to Wilmink (1987). Means and standard deviations of the milk production traits are in Table 1. About 60% of the heifers were daughters of 39 proven bulls; the remaining 40% were daughters of 75 young bulls. All animals were crosses between Dutch-Friesians and Holstein-Friesians; the average fraction of Holstein-Friesian genes of the heifers was .63. Herd-year-season classes of calving were required to have at least three records, leaving 6803 records and 1711 herd-year-season classes.

The following animal model was used to estimate effects of milk protein genotypes on milk production traits:

$$y_{ijklm} = hys_i + b_1(c_{ijklm} - \bar{c}) + b_2(c_{ijklm} - \bar{c})^2 + m_j + mp_k + g_l + u_m + e_{ijklm} \quad [1]$$

where

y_{ijklm} = observation ijklm,

hys_i = the fixed effect of herd-year-season class of calving i (i = 1....1711). Three year-season classes were distinguished:

February 1988 to August 1988, September 1988 to January 1989 and February 1989 to August 1989,

c_{ijklm} = the calving age in months of cow ijklm,
$\bar{c}$ = the mean calving age,
b_1 = the linear regression coefficient of age at calving,
b_2 = the quadratic regression coefficient of age at calving,
m_j = the fixed effect of month of calving class j (j = 1,..,12),
mp_k = fixed effect of milk protein genotype k,
g_l = effect of genetic group l (l = 1,..21),
u_m = random effect of animal m, and
e_{ijklm} = the random residual effect of cow ijklm,

Because each milk protein gene in this model is analyzed separately, this model is referred to as the single-gene model. Solutions were obtained by Gauss-Seidel iteration as described by Schaeffer and Kennedy (1986). The relative difference between consecutive solutions was used as a convergence criterion (Misztal et al., 1987). In matrix notation,

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

where **b**, **g**, and **a** are vectors of solutions of fixed effects, genetic group effects and animal effects, respectively, and **X**, **W**, and **Z** are the corresponding design matrices. To obtain solutions heritability estimates from Van der Werf and De Boer (1989) for the Dutch Black and White crossbred cattle population were used. Heritabilities were .38, .80, .70, .36, and .33 for milk production, fat percentage, protein percentage, fat yield, and protein yield, respectively. Relationships between animals because of sires, dams, paternal grandsires, paternal granddams, and maternal grandsires were considered. Number of animals involved in the analyses, including ancestors without records, was 14,592.

Criteria used to assign phantom parents to genetic groups (Westell et al., 1988) consisted of selection path, fraction of Holstein-Friesian genes, and year of birth of paternal granddams, paternal grandsires, maternal grandsires, or dams. Selection paths were sires to breed sires, dams to breed sires, sires to breed dams, and dams to breed dams. Criteria to assign animals, based on fraction of Holstein-Friesian genes and year of birth, varied within selection paths in order to attain a minimum of 10 animals in each genetic group. This resulted in 21 genetic groups.

To test the significance of genotype effects, the quadratic Q was calculated as

$$Q = (K'\tilde{\beta} - m)'(K'C_{11}K)^{-1}(K'\tilde{\beta} - m)$$

where C_{11} is the part of the generalized inverse of the coefficient matrix corresponding to **X'X**, and $(K'\beta - m) = 0$ is the hypothesis to be tested. Because solutions are obtained by iterating on the data, the coefficient matrix is not set up explicitly. To obtain the matrix $K'C_{11}K$ without computing the generalized inverse **C**, the strategy as described by Groeneveld, Kovac, Wang, and Fernando (1991, unpublished data) was used. In that strategy, the right-hand sides were substituted by a vector $(K_i', 0 \cdots 0)'$, where K_i is column i of **K** representing contrast i between fixed effects. The solutions obtained by iterating on the data represent

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} K_i \\ 0 \end{bmatrix} = \begin{bmatrix} C_{11}K_i \\ C_{21}K_i \end{bmatrix}.$$

Multiplying the vector of solutions with $(K_i', 0 \cdots 0)$ gives

$$\begin{bmatrix} K_i' & 0 \end{bmatrix} \begin{bmatrix} C_{11}K_i \\ C_{21}K_i \end{bmatrix} = K_i' C_{11} K_i .$$

Error variance, $\hat{\sigma}_e^2$ was estimated as

$$\hat{\sigma}_e^2 = (y'y - \tilde{\beta}'X'y - \tilde{u}'Z'y) / (n - r)$$

where n is the number of observations, $\tilde{\beta}$ is the vector with estimated fixed effects, $\tilde{u}$ is the vector with estimated animal effects and r is the rank of **X**. Now $Q/(s\hat{\sigma}_e^2)$ has a central *F* distribution with s and (n - r) degrees of freedom, where s is the rank of the matrix with vectors K_i. The elements of a K_i vector are all zero except for the element corresponding to milk protein genotype (i+1):

$$K_1' \quad (\cdots 0_1\ 1_2 \cdots 0_{MP} \cdots)$$
$$\vdots$$
$$K_{MP-1}' \quad (\cdots 0_1\ 0_2 \cdots 1_{MP} \cdots)$$

where MP is the number of milk protein genotypes. Because in each analysis the effect of the first milk protein genotype was set equal to zero, i = 1, MP-1. Standard errors of genotype effect i were estimated as

$$SE_i = \sqrt{((K_i' C_{11} K_i)\ \hat{\sigma}_{e^2}}$$

Because casein loci are closely linked (Grosclaude et al., 1973) casein genes might be in linkage disequilibrium. Therefore, estimates of casein genotype effects obtained using the single-gene model might be affected by effects of linked casein genes. To account for this, a second model was used in which the milk protein genotype effects were adjusted for effects of all other milk protein genes:

$$y_{ijklmnop} = hys_i + b_1(c_{ijklmnop} - \bar{c}) + b_2(c_{ijklmnop} - \bar{c})^2 + m_j + \alpha_{s1}\text{-}Cn_k + ß\text{-}Cn_l + \kappa\text{-}Cn_m + ß\text{-}Lg_n + g_o + u_p + e_{ijklmnop} \quad [2]$$

where

α_{s1}-Cn_k	=	fixed effect of α_{s1}-casein genotype k (k = 1, 2),
ß-Cn_l	=	fixed effect of ß-casein genotype l (l = 1,.. 9),
κ-Cn_m	=	fixed effect of κ-casein genotype m (m = 1,.. 3), and
ß-Lg_n	=	fixed effect of ß-lactoglobulin genotype n (n = 1,.. 3).

Other effects in Model [2], the procedures used to obtain solutions, and the calculation of *F* values and standard errors were as described for Model [1]. This model will be referred to as the multigene model. Statistical models that adjusted milk protein genotype effects for effects of other milk protein genes have been used previously (Aleandri et al., 1990, Gonyon et al., 1987, Graml et al., 1985, Graml et al., 1986, Haenlein et al., 1987, Ng-Kwai-Hang et al., 1984).

RESULTS

Genotype Distribution

For α_{s1}-casein, the BB and BC genotypes were observed. As in most Western dairy cattle breeds, for α_{s2}-casein, only the AA genotype was found. For ß-casein, the genotypes A^1A^1, A^1A^2, A^2A^2, A^1B, A^2B, BB, A^1A^3, A^2A^3, and A^3B were found. For both κ-casein and ß-lactoglobulin, the AA, AB, and BB genotypes were observed. Table 2 shows the α_{s1}-casein, ß-casein, and κ-casein genotype frequencies and the deviation of observed from expected frequencies of casein genotype combinations for the 6803 animals included in the analysis. Expected

Table 2. Frequencies for α_{s1}-casein, ß-casein, and κ-casein genotypes and the deviation of observed from expected frequencies of casein genotype combinations (percentage).

		α_{s1}-Casein		κ-Casein		
		BB	BC	AA	AB	BB
ß-Casein	Genotype Frequencies[2]	96.38	3.62	63.63	32.62	3.75
A^1A^1	29.33	+ 1.01[1]	- 1.02	+ .45	- .43	- .03
A^1A^2	42.01	+ .46	- .46	+ 5.67	- 4.45	- 1.21
A^2A^2	11.25	- .12	+ .12	+ 3.26	- 2.86	- .41
A^1B	10.05	+ .36	- .36	- 6.10	+ 4.89	+ 1.21
A^2B	5.01	- .05	+ .06	- 3.04	+ 3.06	- .01
BB	.51	+ .02	- .02	- .32	- .17	+ .49
A^1A^3	.91	- .82	+ .82	+ .10	- .06	- .03
A^2A^3	.65	- .62	+ .61	+ .16	- .14	- .02
A^3B	.28	- .26	+ .25	- .17	+ .16	.00
α_{s1}-Casein						
BB	96.38			- .31	+ .25	+ .06
BC	3.62			+ .31	- .25	- .07

[1](Observed - Expected) frequencies. Expected frequencies were calculated assuming equilibrium.
[2]Genotype frequencies are given for both the row and the column.

frequencies were calculated as if the casein genes were in linkage equilibrium, i.e., multiplying genotype frequencies. Because of rounding errors, the summation of the deviations was not always exactly equal to zero. To test whether the observed deviations were significant a *Chi-square* test was performed. To be valid the *Chi-square* test requires a minimum expected number of observations per cell of 5; therefore, animals carrying ß-casein BB, A^1A^3, A^2A^3, and A^3B were excluded from these analysis, leaving 6643 observations. Significant linkage disequilibria existed between ß-casein and α_{s1}-casein ($P < 0.001$), ß-casein and

κ-casein ($P < 0.001$), and α_{s1}-casein and κ-casein ($P = 0.033$).

Table 2 shows that the ß-casein A^1A^1 genotype appeared more often in combination with α_{s1}-casein BB than with α_{s1}-casein BC. The distribution of the ß-casein A^1A^1 genotype over κ-casein genotypes approximated the distribution as would have been expected if the genes were segregating independently. The ß-casein A^2 allele, however, appeared more often in combination with the κ-casein A allele. ß-Casein A^2A^2 was indifferent with respect to α_{s1}-casein genotypes. The ß-casein BB genotype appeared only in combination with the κ-casein BB genotype, which resulted in a positive deviation for the ß-casein BB and κ-casein BB genotype combination (+.49). The frequency distributions of the ß-casein A^1A^3 and A^2A^3 genotypes clearly indicated that the ß-casein A^3 allele appeared more frequently together with the α_{s1}-casein C allele than with the α_{s1}-casein B allele. The deviations as observed for the other heterozygote ß-casein genotypes were a mixture of the effects observed for both homozygotes. The distribution of α_{s1}-casein genotypes over κ-casein genotypes deviated slightly, although significantly, from the expected distribution.

Single-Gene Analysis

Table 3 shows the significance and the estimates of milk protein genotype effects on milk production traits for a model in which each milk protein was analyzed separately (Model [1]). Effects of α_{s1}-casein genotypes on protein percentage ($P = .003$) and protein yield ($P = .025$) were significant. Compared with α_{s1}-casein BB, the BC genotype of α_{s1}-casein was associated with a .04 higher protein percentage and a 4.09 kg higher protein yield.

ß-Casein genotypes had significant effects on milk production ($P = .001$), fat percentage ($P = .017$), protein percentage ($P < .001$), and protein yield ($P = .007$). This was especially due to effects associated with the A^3 and the B alleles of ß-casein. ß-Casein A^1B and ß-casein BB cows produced 155 and 320 kg less milk, respectively, than ß-casein A^1A^1 cows. However, ß-casein A^1B and BB cows produced milk with a higher fat and protein content than that of ß-casein A^1A^1 cows. Estimates for the ß-casein A^2B and A^3B genotypes did not confirm these trends. However, standard errors for these estimates were large. The A^3 allele of ß-casein was associated with a lower fat content and a higher protein content of milk. ß-Casein A^1A^3, A^2A^3, and A^3B cows, respectively, produced milk with .05, .05 and .09% lower fat content and .05, .06 and .03% higher protein content. Because of high standard errors, effects of the ß-casein A^3 allele on milk production were less clear.

Table 3. Estimates of milk protein genotype effects with standard errors on 305-d milk production traits and the significance of milk protein genotype effects for a statistical model in which each milk protein gene is analyzed separately.

	Milk, kg		Fat, %		Protein, %		Fat, kg		Protein, kg	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
					α_{s1} - Casein					
BB	.0		.0		.0		.0		.0	
BC	53.	61.	-.04	.03	.04	.01	-.31	2.57	4.09	1.84
P	.383		.157		.003		.870		.025	
					ß - Casein					
A^1A^1	.0		.0		.0		.0		.0	
A^1A^2	49.	25.	-.04	.01	-.00	.01	-.37	1.17	1.55	.85
A^2A^2	21.	42.	-.00	.02	.01	.01	.45	1.77	.98	1.27
A^1B	-155.	43.	.03	.02	.04	.01	-5.47	1.81	-2.70	1.30
A^2B	-5.	55.	-.03	.03	.02	.01	-1.80	2.31	1.50	1.66
BB	-320.	150.	.08	.07	.05	.03	-10.00	6.28	-8.34	4.51
A^1A^3	74.	113.	-.05	.06	.05	.02	.15	4.72	5.71	3.40
A^2A^3	-48.	137.	-.05	.07	.06	.03	-8.53	5.72	-1.03	4.11
A^3B	222.	202.	-.09	.10	.03	.04	5.53	8.46	9.36	6.08
P	.001		.017		< .001		.066		.007	
					κ - Casein					
AA	.0		.0		.0		.0		.0	
AB	-51.	16.	.01	.01	.03	.01	-2.05	1.06	.32	.77
BB	-173.	61.	.05	.03	.08	.01	-5.28	2.54	-1.23	1.82
P	< .001		.325		< .001		.039		.663	
					ß - Lactoglobulin					
AA	.0		.0		.0		.0		.0	
AB	-15.	28.	.04	.01	-.00	.01	1.34	1.25	-.80	.90
BB	-93.	35.	.11	.02	.01	.01	2.50	1.45	-2.86	1.04
P	.019		< .001		.191		.224		.009	

κ-Casein genotypes had a significant effect on milk production ($P < .001$). κ-Casein BB cows produced 173 kg of milk less than κ-casein AA cows. Furthermore, κ-casein genotypes had a highly significant effect on protein content ($P < .001$); the κ-casein BB cows produced milk with a .08% higher protein content than that of the AA cows. The effect of κ-casein genotypes on fat percentage was not significant. However, the effect of κ-casein genotypes on kilograms of fat was significant; here, the B allele was associated with a lower fat yield. Protein yield was not significantly affected by κ-casein genotypes.

The effect of ß-lactoglobulin genotypes on milk production was significant ($P = .019$). Cows carrying the ß-lactoglobulin BB genotype produced 93 kg of milk less than ß-lactoglobulin AA cows. Furthermore, was the effect of ß-lactoglobulin genotypes on fat percentage highly significant ($P < .001$). Cows carrying the BB genotype had .11% higher fat content. Effects of ß-lactoglobulin genotypes on protein content and fat yield were not significant. ß-Lactoglobulin genotypes had a significant effect on protein yield; BB cows had lower protein yields than AA or AB cows.

Multigene Analysis

Table 4 shows the results of the analysis in which all milk protein genes were analyzed simultaneously (Model [2]). In general, compared with the results of the single-gene model (Table 3), estimates of casein genotype effects decreased, and standard errors of the estimates increased. After accounting for the effects of other milk proteins, effects of α_{s1}-casein genotypes on protein percentage and protein yield were not significant. The effect of α_{s1}-casein genotypes on protein percentage was only slightly reduced, but, because of the larger standard error, the effect was no longer significant ($P = .089$).

In the multigene analysis, the effect of ß-casein genotypes on protein percentage was no longer significant. Data Correction for the other milk protein genes especially affected the estimated ß-casein BB effect on protein percentage. The ß-casein BB genotype effect changed from +.05% (Table 3) to -.02 (Table 4). Furthermore, positive effects of the ß-casein A^1B, A^2B, A^1A^3, and A^2A^3 genotypes on protein percentage were reduced by .02: estimates changed from .04, .02, .05, and .06 (Table 3) to .02, .00, .03, and .04 (Table 4), respectively. Effects of ß-casein genotypes on milk production, fat percentage, and protein yield remained significant.

Because the data were adjusted for effects of other milk protein genotypes, the κ-casein genotype effects on milk production and fat yield were no longer

Table 4. Estimates of milk protein genotype effects (with standard errors) on 305-d milk production traits and the significance of milk protein genotype effects for a statistical model in which all milk protein genes are analyzed simultaneously.

	Milk, kg		Fat, %		Protein, %		Fat, kg		Protein, kg	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
					α_{s1} - Casein					
BB	.0		.0		.0		.0		.0	
BC	55.	84.	-.02	.04	.03	.02	.55	3.59	3.37	2.57
P	.519		.625		.089		.852		.186	
					ß - Casein					
A^1A^1	.0		.0		.0		.0		.0	
A^1A^2	47.	29.	-.04	.01	.00	.01	-.59	1.20	1.70	.86
A^2A^2	17.	37.	-.01	.02	.02	.01	.04	1.82	1.35	1.30
A^1B	-145.	46.	.03	.02	.02	.01	-4.62	2.01	-4.03	1.44
A^2B	1.	53.	-.03	.03	.00	.01	-1.28	2.48	0.23	1.78
BB	-264.	159.	.07	.08	-.02	.03	-7.70	6.83	-10.54	4.90
A^1A^3	11.	132.	-.02	.07	.03	.03	-0.26	5.74	2.52	4.11
A^2A^3	-208.	148.	-.02	.08	.04	.03	-9.32	6.71	-3.91	4.81
A^3B	201.	215.	-.10	.11	-.02	.04	5.11	9.15	5.79	6.56
P	.005		.021		.266		.339		.005	
					κ - Casein					
AA	.0		.0		.0		.0		.0	
AB	-5.	31.	-.01	.02	.03	.01	-.97	1.31	1.81	.94
BB	-60.	64.	.01	.03	.08	.01	-2.68	2.92	2.73	2.10
P	.679		.793		< .001		.602		.129	
					ß - Lactoglobulin					
AA	.0		.0		.0		.0		.0	
AB	-22.	30.	.04	.01	.00	.01	1.12	1.26	-.88	.90
BB	-100.	25.	.12	.02	.01	.01	2.32	1.45	-2.99	1.04
P	.003		< .001		.199		.265		.006	

significant. The estimated κ-casein BB genotype effect on milk production changed from -173 kg (Table 3) to -60 kg (Table 4) of milk. The effect of κ-casein genotypes on protein percentage was not affected by correcting the data for effects of other milk protein genes. Although κ-casein genotypes had no significant effect on fat percentage, the estimated genotype effect of κ-casein BB changed from .05% (Table 3) to .01% (Table 4). The reduced effect of κ-casein genotypes on milk production caused the effect of κ-casein BB on protein yield to change from -1.23 kg (Table 3) to +2.73 kg (Table 4).

ß-Lactoglobulin genotype effects were not affected by correcting the data for casein genotype effects. ß-Lactoglobulin genotypes had significant effects on milk production, fat percentage, and protein yield.

DISCUSSION

The Model

In a simulation study, Kennedy et al. (1992) showed that an animal model should be used to estimate genotype effects in a population containing related animals. Milk protein genotypes of ancestors, however, are not known; therefore, only records of one generation of animals are available. Consequently, Kennedy et al. (1992) also simulated a situation in which only one generation of data and all relationships were used. Results indicated that, in that situation, selection bias was small and the test of hypothesis was not influenced. In the present study, only records of heifers were used, and relationships between animals were traced back two generations. Therefore, the model used in the present study is expected to give estimates with only a small selection bias, and the test of hypothesis is expected to be exact.

Causes of Associations

Associations between milk production traits and milk protein genes can be caused by a pleiotropic effect of the milk protein gene on the milk production trait. In the terminology used by Geldermann (1975), the milk protein gene in this case is a quantitative trait locus (QTL). Another possible cause for finding associations between milk protein genes and milk production traits is linkage between a QTL and a milk protein gene. In this situation, the milk protein gene is a marker gene (Geldermann, 1975). The probability of finding significant milk protein genotype effects that are, in fact, due to the effect of linked QTL depends

on the amount of linkage disequilibrium between the milk protein genes and the linked QTL. Linkage disequilibrium can be due to selection, the intermixture of populations, and random factors. However, linkage disequilibrium is reduced by recombination (Falconer, 1989). Reduction of linkage disequilibrium is small when the recombination fraction is small, and, therefore, it is difficult to distinguish between effects of closely linked genes. The main factor producing linkage disequilibrium in the present study is probably the crossbreeding of Dutch-Friesians with Holstein-Friesians. Therefore, if linked QTL play a role, estimates and significance of milk protein genotype effects in the present study might be increased compared with those from purebred populations.

Single-Gene versus Multigene Analysis

Table 2 shows that the casein genes are in linkage disequilibrium. Therefore, estimates of, for example, κ-casein genotype effects might be affected by ß-casein genotype effects. Using a model in which all milk proteins are analyzed simultaneously, estimates of genotype effects can be adjusted for the effects of other milk protein genes.

Comparison of the results in Tables 3 and 4 indicates that some of the effects ascribed to certain milk protein genes in the single-gene analysis are not effects of the milk protein gene itself, but effects of linked genes. According to the model in which one milk protein gene was analyzed at a time, κ-casein genotypes had a significant effect on milk production, whereas, in the model in which all milk protein genes were analyzed together, the effect of κ-casein genotypes on milk production was not significant. An explanation for this phenomenon is that the B allele of ß-casein often appeared with the κ-casein B allele (Table 2). As a consequence, the negative effect on milk production associated with the ß-casein B allele is reflected in the estimated κ-casein AB and BB genotype effects (Table 3). Simultaneous estimation of κ-casein and ß-casein genotype effects removes this negative effect. Also, the positive effect associated with the ß-casein B allele on fat percentage is reflected in the estimated genotype effects of κ-casein AB and BB (Table 3). Simultaneous estimation of κ-casein genotype effects and ß-casein genotypes effects reduced the estimates of the effect of the κ-casein AB and BB genotypes on fat percentage from .01% and .05% (Table 3) to -.01% and .01%, respectively (Table 4). The same argument can be used to explain why, after accounting for κ-casein genotype effects, ß-casein genotypes no longer have a significant effect on protein percentage. The positive effect of the ß-casein B allele on protein percentage, found in the single-gene analysis, can be explained

by the effect of the linked κ-casein B allele. ß-Lactoglobulin is not linked to the casein genes. Therefore, single-gene and multigene analyses are expected to give similar estimates, as confirmed in Tables 3 and 4.

Literature

The previous section showed how estimates of genotype effects can be influenced by genotype effects of a linked gene. The single-gene and multigene analyses indicate that a gene with an effect on protein percentage is more closely linked to the κ-casein gene than to the ß-casein gene. However, the results of the present study alone do not necessarily indicate that the κ-casein gene itself has an effect on protein percentage. If κ-casein itself has an effect on protein percentage, then estimates of κ-casein genotype effects are expected to be consistent in different populations or in different breeds. However, if effects are due to linked genes, then estimates are likely to be contradictory.

Table 5 gives an overview of significant effects and of favorable genotypes in six recent studies on associations between milk protein genotypes and milk production traits. There are some difficulties in comparing results from different studies, e.g., statistical models differed, and not all of the studies distinguished among A^1, A^2, and A^3 alleles of ß-casein. All the results in table 5 are based on statistical analysis using the multigene model.

Four out of the six studies in Table 5 report a significant effect of ß-lactoglobulin genotypes on fat percentage. In all cases, cows carrying the BB genotype had higher fat percentage. These results agree with the highly significant effect of ß-lactoglobulin genotypes on fat percentage in the present study. The consistency between results from different studies strongly indicates that ß-lactoglobulin itself affects fat percentage.

Table 5 shows that three studies reported significant effects of κ-casein genotypes on protein percentage. In all cases, cows carrying the κ-casein BB genotype produced milk with a higher protein content. These results agree with the present study. Remarkably, the effect was only significant in studies of the Holstein- Friesian breed. Three studies using breeds found no significant effect of κ-casein genotypes on protein percentage, and, in two of these studies, the BB genotype was not the most favorable genotype, i.e., the genotype with the highest protein percentage. This suggests that a gene very closely linked to the κ-casein gene affects protein percentage. Recent results indicated that the positive association of κ-casein B with protein percentage in Holstein-Friesians is due to an increased κ-casein B content (Van den Berg et al., 1992, Van Eenennaam and

Table 5. Literature overview of favorable milk protein genotypes and significance of milk protein genotype effects on milk production traits.

Reference	Milk, kg		Fat, %		Protein, %		Fat, kg		Protein, kg	
					α_{s1} - Casein					
1	**[3]	BB	ns[1]	BC	ns	BC	**	BB	**	BB
2	ns	-[5]	ns	-	ns	-	-[6]		-	
3	*[2]	BB	ns	BB	*	BC	**	BB	ns	BB
4	ns	BB	*	CC	*	BC/CC	ns	BB	ns	BB
5	ns	BB	ns	CC	ns	CC	ns	BB	ns	CC
6	ns	-	ns	-	*	CC	-		-	
					ß - Casein					
1	*	A^2A^3	*	A^1A^1	ns	A^1B	*	A^1A^3	*	A^2A^3
2	ns	-	ns	-	**	A^1A^3	-		-	
3	ns	A^2A^3	ns	A^2A^3	ns	A^1B	ns	A^2A^3	ns	A^2A^3
4	ns	BC	ns	BB	***[4]	BC	ns	BC	*	BC
5	*	BC	**	BB	ns	CC	*	BC	*	BC
6	ns	-	ns	-	ns	-	-		-	
					κ - Casein					
1	ns	BB	ns	AA	*	BB	ns	BB	*	BB
2	ns	AA	ns	-	*	BB	-		-	
3	ns	AB	ns	AB	**	BB	ns	AB	*	BB
4	ns	AB	ns	AA	ns	AA	ns	AB	ns	BB
5	ns	AA	ns	AA	ns	AB	ns	AA	ns	AA
6	ns	-	ns	-	ns	-	-		-	
					ß - Lactoglobulin					
1	ns	AA	*	BB	**	AA	ns	BB	*	AA
2	ns	-	ns	-	ns	-	-		-	
3	ns	AA	**	BB	ns	AA	ns	BB	*	AA
4	ns	AA	*	BB	*	AA	ns	BB	ns	AA
5	ns	BD	ns	BB	ns	AD	ns	BD	ns	BD
6	ns	-	**	BB	ns	-	-		-	

[1] ns : not significant ; [2] * : $P < .050$; [3] ** : $P < .010$; [4] *** : $P < .001$; [5] ns - : effect not significant, no estimates of genotype effects.; [6] - : not included in the analysis.

References;

1 Ng-Kwai-Hang et al., (1984) - Holstein-Friesians - n = 2045 (casein genes), n = 3870 (whey proteins)
2 Gonyon et al., (1987) - Holstein-Friesians - n = 3571 (milk production and fat percentage), n = 3111 (protein percentage)
3 Aleandri et al., (1990) - Holstein-Friesians - n = 1383
4 Graml et al. (1985) and Graml et al. (1986) - Braunvieh - n = 2139
5 Graml et al. (1985) and Graml et al. (1986) - Fleckvieh - n = 2262
6 Haenlein et al. (1987) - Guernseys - n = 3888 (milk production and fat percentage), n = 1732 (protein percentage)

Medrano, 1991). Possible explanations for a higher expression of the κ-casein B allele could be a higher stability of the κ-casein B mRNA or the linkage of different promoter regions to the A and B variants of κ-casein. The literature overview in table 5 suggests that the latter explanation is more likely.

Three studies found significant effects of α_{s1}-casein genotypes on protein percentage. In all cases, the most favorable effect was associated with the C allele. Also, in the present study, the C allele was associated with a higher protein percentage. However, in the multigene analysis, this effect was not significant ($P = .089$). The low frequency of the α_{s1}-casein C allele in most breeds might explain the absence of significant effects in some studies. The consistency of the different studies on the effect of α_{s1}-casein genotypes on protein percentage indicates that α_{s1}-casein itself or a closely linked gene affects protein percentage.

For ß-casein, three studies found a significant effect on protein yield. In the present study, this effect was also significant. In several studies, rare genotypes were the most favorable. However, standard errors of estimates of effects for these rare genotypes were high, and, therefore, it is difficult to draw a conclusion about the consistency of the results of the different studies. Furthermore, some studies did not distinguish between A^1, A^2, and A^3 alleles. Associations in the present study are mainly due to effects associated with the B and the A^3 alleles of ß-casein. The low frequencies of these alleles in most populations might explain the absence of significant ß-casein genotype effects in other studies.

Selection for Manufacturing Properties

Because of the effect of κ-casein genetic variants on renneting time, interest in selecting animals with the favorable κ-casein B allele is considerable. Unfortunately, because of the association between the κ-casein B allele and milk production (Table 3), selection for this κ-casein allele would have a negative effect on milk yield. This could, however, be avoided by taking ß-casein genotypes into account. Selection would then be for animals carrying the favorable κ-casein B allele in combination with the ß-casein A^1, A^2, or A^3 allele.

CONCLUSIONS

The results of the present study indicate that κ-casein genetic variants are associated with protein percentage, whereas ß-lactoglobulin genetic variants are associated with fat percentage. In both cases, there are strong indications that

these effects are due to the gene itself or due to a very closely linked gene. Furthermore, effects of ß-casein genotypes on milk production, fat percentage, and protein yield were significant, whereas ß-lactoglobulin genotypes had significant effects on milk production and protein yield. It is less clear whether those effects are due to effects of the milk protein genes or to effects of linked genes. This will be addressed in a subsequent study (Bovenhuis and Weller, 1992, unpublished data).

ACKNOWLEDGMENTS

The authors wish to thank Esther Verstege for phenotyping the milk samples and the Royal Dutch Cattle Syndicate for supplying the data and financial support.

REFERENCES

Aleandri, R., L. G. Buttazzoni, J. C. Schneider, A. Caroli, and R. Davoli. 1990. The effects of milk protein polymorphisms on milk components and cheese-producing ability. J. Dairy Sci. 73: 241.

Bovenhuis, H. and J. A. M. Van Arendonk. 1991. Estimation of milk protein gene frequencies in crossbred cattle by maximum likelihood. J. Dairy Sci. 74: 2728.

Bovenhuis, H., and A. J. M. Verstege. 1989. Improved method for phenotyping milk protein variants by isoelectric focusing using PhastSystem. Neth. Milk Dairy J. 43: 447.

Falconer, D. S. 1989. Introduction to quantitative genetics. Longman Scientific & Technical, Essex, Engl. p. 1-438.

Geldermann, H. 1975. Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods. Theor. Appl. Genet. 46: 319.

Gonyon, D. S., R. E. Mather, H. C. Hines, G. F. W. Haenlein, C. W. Arave, and S. N. Gaunt. 1987. Associations of bovine blood and milk polymorphisms with lactation traits: Holsteins. J. Dairy Sci. 70: 2585.

Graml, R., J. Buchberger, H. Klostermeyer, and F. Pirchner. 1985. Pleiotrope wirkungen von ß-lactoglobulin- und casein genotypen auf milchfett- und milchproteinmengen des Bayerischen Fleckviehs und Braunviehs. Z. Tierz. Zuechtgsbiol. 103: 33.

Graml, R., J. Buchberger, H. Klostermeyer, and F. Pirchner. 1986. Pleiotrope wirkungen von ß-lactoglobulin- und casein genotypen auf milchinhaltsstoffe des Bayerischen Fleckviehs und Braunviehs. Z. Tierz. Zuechtgsbiol. 102: 355.

Grosclaude, F. 1988. Le polymorphisme génétique des principales lactoprotéines bovines. Inst. Nat. Rech. Agron. Prod. Anim. 1: 5.

Grosclaude, F., J. C. Mercier and B. R. Ribadeau Dumas, 1973. Genetic aspects of cattle casein research. Neth. Milk Dairy J. 27: 328.

Haenlein, G. F. W., D. S. Gonyon, R. E. Mather, and H. C. Hines. 1987. Associations of bovine blood and milk polymorphisms with lactation traits: Guernseys. J. Dairy Sci. 70: 2599.

Kennedy, B. W., M. Quinton and J. A. M. van Arendonk. 1992. Estimation of effects of single genes on quantitative traits. J. Anim. Sci. 70: 1999.

McLean, D. M., E. R. B. Graham, R. W. Ponzoni, and H. A. McKenzie. 1984. Effects of milk protein genetic variants on milk yield and composition. J. Dairy Res. 51: 531.

Misztal, I., D. Gianola, and L. R. Schaeffer. 1987. Extrapolation and convergence criteria with Jacobi and Gauss-Seidel iteration in animal models. J. Dairy Sci. 70: 2577.

Ng-Kwai-Hang, K. F., J. F. Hayes, J. E. Moxley, and H. G. Monardes. 1984. Association of genetic variants of casein and milk serum proteins with milk, fat, and protein production by dairy cattle. J. Dairy Sci. 67: 835.

Ng-Kwai-Hang, K. F., J. F. Hayes, J. E. Moxley, and H. G. Monardes. 1986. Relationships between milk protein polymorphisms and major milk constituents in Holstein Friesian cows. J. Dairy Sci. 69: 22.

Schaar J., B. Hansson, and H. -E. Pettersson. 1985. Effects of genetic variants of κ-casein and ß-lactoglobulin on cheesemaking. J. Dairy Res. 52: 429.

Schaeffer, L. R. and B. W. Kennedy. 1986. Computing solutions to mixed model equations. Proc. 3rd World Congr. Genet. Appl. Livest. Prod., Lincoln, Nebraska. XII: 382.

Van den Berg, G., J. T. M. Escher, P. J. de Koning and H. Bovenhuis. 1992. Genetic polymorphism of κ-casein and ß-lactoglobulin in relation to milk composition and processing properties. Neth. Milk Dairy J. 46: 145.

Van der Werf, J. H. J. and W. De Boer. 1989. Estimation of genetic parameters in a crossbred population of black and white dairy cattle. J. Dairy Sci. 72: 2615.

Van Eenennaam, A. L., and J. F. Medrano. 1991. Differences in allelic protein expression in the milk of heterozygous κ-casein cows. J. Dairy Sci. 74: 1496.

Westell, R. A., R. L. Quaas and L. D. Van Vleck. 1988. Genetic groups in an animal model. J. Dairy Sci. 71: 1310.

Wilmink, J. B. M. 1987. Studies on testday and lactation milk, fat and protein yield of dairy cows. Ph.D. Diss. Wageningen Agric. Univ., Neth. p. 1-123.

Chapter 6

MAPPING AND ANALYSIS OF DAIRY CATTLE QUANTITATIVE TRAIT LOCI BY MAXIMUM LIKELIHOOD METHODOLOGY USING MILK PROTEIN GENES AS GENETIC MARKERS.

Henk Bovenhuis*[1] and Joel I. Weller†

*Department of Animal Breeding
Wageningen Agricultural University
The Netherlands

†A. R. O., The Volcani Center
Bet Dagan
Israel

Submitted for publication to Genetics

[1] Prepared during a leave at A. R. O., The Volcani Center, Bet Dagan, Israel.

ABSTRACT

Maximum likelihood methodology was used to estimate effects of both a marker gene and a linked quantitative trait locus (QTL) on quantitative traits in a segregating population. Two alleles were assumed for the QTL. In addition to the effects of genotypes at both loci on the mean of the quantitative trait, recombination frequency between the loci, frequency of the QTL alleles, and the within sire standard deviation for a given marker-QTL genotype of the daughter, were also estimated. Thus six parameters were estimated in addition to the marker genotype means. The statistical model was tested on simulated data, and used to estimate direct and linked effects of the milk protein genes, β-lactoglobulin, κ-casein, and β-casein, on milk, fat, and protein production and fat and protein percent in the Dutch dairy cattle population. β-Lactoglobulin had a significant direct effect on fat percent. κ-Casein had a significant direct effect on protein percent. β-Casein had significant direct effects on fat and protein percent. Linked QTL with significant effects on fat percent were found for all three loci. Since the β-casein and κ-casein genes are closely linked, it is likely that the same QTL was detected for those two markers.

Key words: Mapping, quantitative trait loci, dairy cattle.

INTRODUCTION

Numerous studies have shown that quantitative trait loci (QTLs) affecting traits of economic interest in agricultural species can be detected and mapped by linkage to known genetic markers (reviewed by Soller, 1990, 1991). Most studies have analyzed crosses between inbred lines. This is a viable option for many plant species, but not for fruit trees and most livestock species, especially dairy cattle. In these species, as in humans, it is necessary to analyze existing segregating populations. In commercial dairy cattle populations elite sires often have hundreds or even thousands of daughters produced by artificial insemination. Thus, a segregating QTL can be detected by analyzing the progeny of a sire heterozygous for a genetic marker (Neimann-Sorensen and Robertson, 1961). If the marker is linked to a QTL heterozygous in the sire, then daughters inheriting the different sire marker alleles should also display a difference for the quantitative trait. Even if the sire is heterozygous for the genetic marker, he may still be homozygous for the QTL. Furthermore, even if the sire is heterozygous

for both loci, linkage relationships between the marker and QTL alleles may be different for different sires. Thus, progeny of several sires should be analyzed jointly, and it is necessary to analyze marker effects within sires. This can be done either by analysis of variance (Soller and Genizi, 1978) or Chi-squared analysis (Gelderman, 1975, Neimann-Sorensen and Robertson, 1961, Weller et al., 1990). Because of differing linkage relationships among sires, this "daughter design" has less statistical power than crosses between inbred lines. To detect a QTL with a substitution effect of 0.1-0.3 phenotypic standard deviation, it is necessary to determine genotypes of thousands of daughters (Soller and Genizi, 1978, Weller et al., 1990). With inbred lines, the same power can be obtained by determining the genotype of less than 1000 progeny (Soller et al., 1976).

If genetic linkage between the QTL and the genetic marker is incomplete, the estimated QTL effects on the quantitative trait will be biased by recombination. Several studies have shown that, for crosses between inbred lines, maximum likelihood (ML) methodology can be used to obtain unbiased estimates for both recombination frequency and QTL effects (Jensen, 1989, Knapp et al., 1990, Lander and Botstein, 1989, Weller, 1986, Weller, 1987). Weller (1990) constructed a likelihood function for estimating parameters related to a linked QTL for the daughter design. In addition to the genotype effects, residual variance, and recombination frequency between the genetic marker and the QTL, it is also necessary to estimate allele frequencies for the QTL.

Milk protein content is a major criterion for selection of dairy cattle. Milk protein consists of different fractions of which α_{s1}-casein, α_{s2}-casein, ß-casein, κ-casein, α-lactalbumin and ß-lactoglobulin are the most important. ß-Casein, κ-casein, and ß-lactoglobulin are polymorphic in commercial dairy cattle breeds (Eigel et al. 1984). Several studies have investigated the effects of milk protein genes on production traits in dairy cattle. Most of these analysis estimated the genotype effect on economic traits (e.g. Graml et al. 1985, 1986, Gonyon et al. 1987, Haenlein et al. 1987). Therefore, only effects of milk protein genes itself, or very closely linked genes in linkage disequilibrium with milk protein genes, could be detected. Although variants of the daughter design have been employed by a few studies (Geldermann et al., 1985, Gonyon et al. 1987, Haenlein et al. 1987), it is not clear whether the effects found are due to linked QTLs or direct effects of the milk protein alleles.

Unlike RFLP or VNTR markers, milk protein markers are functional genes. Therefore, it is likely that the milk protein gene itself or a closely linked region (promoter or enhancer) is involved in the regulation of milk protein production. Although models have been described to locate multiple markers and QTLs in

inbred populations (Knapp, 1991), no model has been described yet that accounts for both an effect of the marker and a linked QTL in segregating populations.

The goals of this study were: 1) to construct a statistical model that is able to estimate effects of both a marker gene and a linked QTL on a quantitative trait in a segregating population, 2) to test this model on simulated data, and 3) use the model to estimate direct and linked effects of milk protein genes on milk production traits in the Dutch dairy cattle population.

MATERIAL AND METHODS

Likelihood model.

In order to be able to estimate direct and linked effects of a marker, a likelihood model was constructed for a marker locus with alleles M^1 and M^2, and a linked QTL with alleles Q^1 and Q^2. Marker genotypes of sires and their daughters were assumed to be known without error. Assuming Hardy-Weinberg equilibrium in the sire population, the prior probability of the sire being Q^1Q^1, Q^1Q^2 or Q^2Q^2 for the QTL is p^2, 2pq or q^2, respectively, where frequencies of QTL alleles Q^1 and Q^2 are p and q, respectively. A Q^1Q^1 genotype sire can only transmit the Q^1 allele. The probability of a Q^1Q^1 daughter is then p, *i.e.* the probability of obtaining a Q^1-allele from the dam. Similarly, the probability of a Q^1Q^2 daughter from that sire is equal to q, i.e. the probability of obtaining a Q^2 allele from the dam. Assuming additive gene action, normal distribution of the observations within marker-QTL genotypes and linkage equilibrium between the marker and the QTL, a fraction p of the M^1M^1 daughters of a M^1M^1/Q^1Q^1 sire will be normally distributed with a mean determined by the effect of the Q^1Q^1 and the M^1M^1 genotypes on the quantitative trait. Similarly a fraction q will be normally distributed with a mean determined by the effect of the Q^1Q^2 and the M^1M^1 genotypes on the quantitative trait. The likelihood function of M^1M^1-sire s can be described as follows;

$$L_{M^1M^1\text{-}s} =$$

$$\left[p^2 \prod_{i=1}^{N_{M^1M^1,s}} p \cdot f\left(x, \mu_{Q^1Q^1} + \mu_{M^1M^1}\right) + q \cdot f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^1}\right) \right.$$

$$+2pq \prod_{i=1}^{N_{M^1M^1,s}} \frac{1}{2} p \cdot f\left(x, \mu_{Q^1Q^1} + \mu_{M^1M^1}\right) + \frac{1}{2} \cdot f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^1}\right) + \frac{1}{2} q \cdot f\left(x, \mu_{Q^2Q^2} + \mu_{M^1M^1}\right)$$

$$\left. + q^2 \prod_{i=1}^{N_{M^1M^1,s}} p \cdot f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^1}\right) + q \cdot f\left(x, \mu_{Q^2Q^2} + \mu_{M^1M^1}\right) \right]$$

$$\cdot \left[p^2 \prod_{i=1}^{N_{M^1M^2,s}} p \cdot f\left(x, \mu_{Q^1Q^1} + \mu_{M^1M^2}\right) + q \cdot f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^2}\right) \right.$$

$$+2pq \prod_{i=1}^{N_{M^1M^2,s}} \frac{1}{2} p \cdot f\left(x, \mu_{Q^1Q^1} + \mu_{M^1M^2}\right) + \frac{1}{2} \cdot f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^2}\right) + \frac{1}{2} q \cdot f\left(x, \mu_{Q^2Q^2} + \mu_{M^1M^2}\right)$$

$$\left. + q^2 \prod_{i=1}^{N_{M^1M^2,s}} p \cdot f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^2}\right) + q \cdot f\left(x, \mu_{Q^2Q^2} + \mu_{M^1M^2}\right) \right]$$

The likelihood function shows that differences between the M^1M^1 and M^1M^2 daughters of a M^1M^1 sire are due to differences between the M^1M^1 and M^1M^2 marker genotype effects. Therefore, daughters of this sire are only informative with respect to marker genotype effects.

Only sires that are heterozygous for both the marker and the QTL are informative with respect to the QTL. This is because the transfer of a marked chromosome section can only be traced for a heterozygous sire. If the sire is heterozygous for the QTL, the homozygous daughters will differ not only because they differ with respect to their marker genotypes, but also because they have different QTL genotypes. If the two loci are in coupling phase (M^1 and Q^1 are on the same chromosome) then, except for recombinants, daughters that inherit the M^1-allele will also inherit the Q^1- allele. The opposite is true if the two loci are in repulsion. Under the assumption of linkage equilibrium between the marker and the QTL, the prior probabilities for a M^1Q^1/M^2Q^2 and M^1Q^2/M^2Q^1 sire are pq. The likelihood of M^1M^2-sire s can now be written as;

$$
\begin{aligned}
L_{M^1M^2,s} \;=\; & \\
& \Bigg[p^2 \prod_{i=1}^{N_{M^1M^1,s}} p \bullet f\left(x, \mu_{Q^1Q^1} + \mu_{M^1M^1}\right) + q \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^1}\right) \\
& + pq \prod_{i=1}^{N_{M^1M^1,s}} (1-r) p \bullet f\left(x, \mu_{Q^1Q^1} + \mu_{M^1M^1}\right) + (1-r) q \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^1}\right) \\
& \qquad + (r) p \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^1}\right) + (r) q \bullet f\left(x, \mu_{Q^2Q^2} + \mu_{M^1M^1}\right) \\
& + pq \prod_{i=1}^{N_{M^1M^1,s}} (1-r) p \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^1}\right) + (1-r) q \bullet f\left(x, \mu_{Q^2Q^2} + \mu_{M^1M^1}\right) \\
& \qquad + (r) p \bullet f\left(x, \mu_{Q^1Q^1} + \mu_{M^1M^1}\right) + (r) q \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^1}\right) \\
& + q^2 \prod_{i=1}^{N_{M^1M^1,s}} p \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^1}\right) + q \bullet f\left(x, \mu_{Q^2Q^2} + \mu_{M^1M^1}\right) \Bigg] \\
\bullet & \Bigg[p^2 \prod_{i=1}^{N_{M^1M^2,s}} p \bullet f\left(x, \mu_{Q^1Q^1} + \mu_{M^1M^2}\right) + q \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^2}\right) \\
& + 2pq \prod_{i=1}^{N_{M^1M^2,s}} \frac{1}{2} p \bullet f\left(x, \mu_{Q^1Q^1} + \mu_{M^1M^2}\right) + \frac{1}{2} \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^2}\right) + \frac{1}{2} q \bullet f\left(x, \mu_{Q^2Q^2} + \mu_{M^1M^2}\right) \\
& + q^2 \prod_{i=1}^{N_{M^1M^2,s}} p \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^2}\right) + q \bullet f\left(x, \mu_{Q^2Q^2} + \mu_{M^1M^2}\right) \Bigg] \\
\bullet & \Bigg[p^2 \prod_{i=1}^{N_{M^2M^2,s}} p \bullet f\left(x, \mu_{Q^1Q^1} + \mu_{M^2M^2}\right) + q \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^2M^2}\right) \\
& + pq \prod_{i=1}^{N_{M^2M^2,s}} (1-r) p \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^2M^2}\right) + (1-r) q \bullet f\left(x, \mu_{Q^2Q^2} + \mu_{M^2M^2}\right) \\
& \qquad + (r) p \bullet f\left(x, \mu_{Q^1Q^1} + \mu_{M^2M^2}\right) + (r) q \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^2M^2}\right) \\
& + pq \prod_{i=1}^{N_{M^2M^2,s}} (1-r) p \bullet f\left(x, \mu_{Q^1Q^1} + \mu_{M^2M^2}\right) + (1-r) q \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^2M^2}\right) \\
& \qquad + (r) p \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^2M^2}\right) + (r) q \bullet f\left(x, \mu_{Q^2Q^2} + \mu_{M^2M^2}\right) \\
& + q^2 \prod_{i=1}^{N_{M^2M^2,s}} p \bullet f\left(x, \mu_{Q^1Q^2} + \mu_{M^2M^2}\right) + q \bullet f\left(x, \mu_{Q^2Q^2} + \mu_{M^2M^2}\right) \Bigg]
\end{aligned}
$$

The likelihood function for M^2M^2-sire s can be described similarly to the likelihood for a M^1M^1-sire;

$$L_{M^2M^2\text{-}s} =$$

$$\left[p^2 \prod_{i=1}^{N_{M^1M^2,\,s}} p \cdot f\left(x, \mu_{Q^1Q^1} + \mu_{M^1M^2}\right) + q \cdot f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^2}\right) \right.$$

$$+2pq \prod_{i=1}^{N_{M^1M^2,\,s}} \frac{1}{2} p \cdot f\left(x, \mu_{Q^1Q^1} + \mu_{M^1M^2}\right) + \frac{1}{2} \cdot f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^2}\right) + \frac{1}{2} q \cdot f\left(x, \mu_{Q^2Q^2} + \mu_{M^1M^2}\right)$$

$$\left. + q^2 \prod_{i=1}^{N_{M^1M^2,\,s}} p \cdot f\left(x, \mu_{Q^1Q^2} + \mu_{M^1M^2}\right) + q \cdot f\left(x, \mu_{Q^2Q^2} + \mu_{M^1M^2}\right) \right]$$

$$\cdot \left[p^2 \prod_{i=1}^{N_{M^2M^2,\,s}} p \cdot f\left(x, \mu_{Q^1Q^1} + \mu_{M^2M^2}\right) + q \cdot f\left(x, \mu_{Q^1Q^2} + \mu_{M^2M^2}\right) \right.$$

$$+2pq \prod_{i=1}^{N_{M^2M^2,\,s}} \frac{1}{2} p \cdot f\left(x, \mu_{Q^1Q^1} + \mu_{M^2M^2}\right) + \frac{1}{2} \cdot f\left(x, \mu_{Q^1Q^2} + \mu_{M^2M^2}\right) + \frac{1}{2} q \cdot f\left(x, \mu_{Q^2Q^2} + \mu_{M^2}\right.$$

$$\left. + q^2 \prod_{i=1}^{N_{M^2M^2,\,s}} p \cdot f\left(x, \mu_{Q^1Q^2} + \mu_{M^2M^2}\right) + q \cdot f\left(x, \mu_{Q^2Q^2} + \mu_{M^2M^2}\right) \right]$$

where;

$N_{M^1M^1,\,s}$ = *number of* M^1M^1 *- daughters of sire s*

$N_{M^1M^2,\,s}$ = *number of* M^1M^2 *- daughters of sire s*

$N_{M^2M^2,\,s}$ = *number of* M^2M^2 *- daughters of sire s*

p = *frequency of the QTL - allele* Q^1

q = *frequency of the QTL - allele* Q^2

r = *recombination fraction between marker and QTL*

$$f(x, \mu_{Q^1Q^1} + \mu_{M^1M^1}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x - \mu_{Q^1Q^1} - \mu_{M^1M^1})^2}{2\sigma^2}}$$, *i.e. a normal distribution*

with mean $\mu_{Q^1Q^1} + \mu_{M^1M^1}$ *and standard deviation* σ

Assuming that sires are independent, the natural logarithm of the likelihood for the total population is obtained by summation of the natural logarithm of likelihoods for all sires;

$$\ln(L) = \sum_{s=1}^{N(1)} \ln(L_{M^1M^1-s}) + \sum_{s=1}^{N(2)} \ln(L_{M^1M^2-s}) + \sum_{s=1}^{N(3)} \ln(L_{M^2M^2-s})$$

where; $N(1)$ = *number of sires with* M^1M^1 *genotype*

$N(2)$ = *number of sires with* M^1M^2 *genotype*

$N(3)$ = *number of sires with* M^2M^2 *genotype*

Since this model includes two fixed effects, the genetic marker and the QTL, there is no unique solution for the main effects. A solution was obtained by setting the mean of the M^1M^1-genotype equal to zero. The function was maximized with respect to the remaining five genotype means and with respect to p, r and σ, using the downhill simplex method (Nelder and Mead, 1965). Convergence was assumed when the difference between the simplex point with the highest and the lowest likelihood was less than 0.001. Maximization was repeated using different starting values in order to ascertain whether the global maximum was reached.

Test statistic.

Significance of a linked QTL effect was tested by using the following likelihood ratio test statistic:

$$\textit{Likelihood ratio test statistic} = \ln(L \mid r = 0.5) - \ln(L \mid r = \hat{r})$$

where;

$\ln(L \mid r = 0.5)$ = *the natural logarithm of the likelihood with r fixed at 0.5 and maximised for all other parameters.*

$\ln(L \mid r = \hat{r})$ = *the natural logarithm of the likelihood maximised for all parameters.*

Under the null hypothesis of r=0.5, i. e. no linkage between the marker and a QTL, the test statistic is asymptotically distributed as -½ (χ^2_1) (Simpson, 1989).

Significance of marker genotype effect was also tested by a likelihood ratio test, with the null hypothesis that all marker genotypes have the same effect on the production trait. The denominator of the test statistic was computed by maximizing the likelihood for all parameters, while the numerator was computed by maximizing the likelihood function under the constraint that all marker genotype effects are equal, i.e. set equal to zero. Under the null hypothesis, this test statistic is asymptotically distributed as $-\frac{1}{2}(\chi^2_{df})$, where df, the degrees of freedom, is equal the difference in the number of parameters maximized by the full model and the alternative hypothesis. Because for both hypotheses M^1M^1 was fixed at zero, the degrees of freedom in this case are one less than the number of marker genotypes.

Following Lander and Botstein (1989), "support intervals" for r were computed by maximizing the likelihood function for the other parameters with r fixed at a range of values; from r = 0.0 to r = 0.5, with increments of 0.05.

Simulation study.

Simulation was used to verify the estimation procedure. The additive genetic value of sire i due to many background QTLs, each with a small effect, was simulated as a random deviate from a normal distribution with a mean of zero and a standard deviation of σ_a. The additive genetic value due to background QTLs for daughter j of sire i, a_{ij}, was simulated as follows;

$$a_{ij} = \tfrac{1}{2} \cdot a_i + \sqrt{(3/4)} \cdot x_{ij} \cdot \sigma_a$$

where; a_i = additive genetic value of sire i.

x_{ij} = random number from a standard normal distribution for daughter j of sire i.

σ_a = the additive genetic standard deviation.

Marker- and QTL-genotypes of sires were simulated by sampling from a uniform distribution using the population gene frequencies. Marker and QTL genotypes of daughters were simulated by sampling from uniform distributions, accounting for the genotype of the sire, the recombination fraction between the marker and the linked QTL and the genotype frequencies in the population. Daughter quantitative trait values were then simulated as;

$$y_{ij} = a_{ij} + MARKER_k + QTL_l + w_{ij} \cdot \sigma_e$$

where; y_{ij} = quantitative trait value of daughter j of sire i.
a_{ij} = additive genetic value due to background QTL for daughter j of sire i.
$MARKER_k$ = effect of marker genotype k.
QTL_l = effect of QTL genotype l.
w_{ij} = random number from a standard normal distribution for daughter j of sire i.
σ_e = the residual standard deviation.

Five thousand daughter records were simulated, divided into 100 half-sib families of 50. The frequency of the M^1 and Q^1 alleles in the sire and dam population was set at 0.5. σ_a^2 was 160,000, and σ_e^2 was 373,333. Ignoring QTL- and marker-effects, heritability = 0.3, which is similar to the heritability of milk production in the Dutch dairy population. The simulated within sire standard deviation, for a given marker-QTL genotype of the daughter, was computed as the within sire additive genetic variance due to background genes and residual variance, i.e. $\sqrt{¾(160,000) + 373,333} = 702.4$.

Field data.

Milk samples of 10,151 first lactation cows on 2,618 herds were collected from February 1989 until October 1989 by the Dutch milk recording associations. About 60% of the heifers were daughters of 39 proven bulls; the remaining 40% were daughters of 75 young bulls. All animals were crosses between Dutch-Friesians and Holstein-Friesians; the mean fraction of Holstein-Friesian genes of heifers was 0.63.

Of one milk sample of each cow the genotype for α_{s1}-casein, ß-casein, κ-casein, and ß-lactoglobulin was determined by isoelectric focusing (Bovenhuis and Verstege, 1989). Milk protein gene frequencies were as described by Bovenhuis and Van Arendonk (1991). Pedigree information, heifer breed, and 305-day milk production records for milk, fat and protein yield, and fat and protein percent were obtained from the registration files of the Royal Dutch Cattle Syndicate. Milk production records were based on test-day observations every 3 or 4 weeks as part of the national milk recording scheme.

To avoid selection bias, only first lactation records where analyzed. Incomplete lactations with at least 90 days in milk were extended to expected 305-day production by the methods of Wilmink (1987), accounting for herd level, calving age, and year-season of calving.

Adjustment for fixed effects.

Milk, fat, and protein yield, and fat and protein percent were adjusted individually for fixed effects using the following fixed model;

$$y_{ijkl} = hys_i + b_1(c_{ijkl} - \bar{c}) + b_2(c_{ijkl} - \bar{c})^2 + m_j + HF_k + e_{ijkl}$$

where:

y_{ijkl}	=	the production record for each trait for a cow from the i^{th} herd-year-season, with calving age ijkl and j^{th} calving month, of the k^{th} Holstein Friesian genetic group.
hys_i	=	The i^{th} herd-year-season of calving. Three year-seasons were defined; February 1988 - August 1988, September 1988 - January 1989, February 1989 - August 1989.
c_{ijkl}	=	The calving age in months of cow ijkl.
$\bar{c}$	=	The mean calving age.
b_1	=	The linear regression coefficient of age at calving.
b_2	=	The quadratic regression coefficient of age at calving.
M_j	=	The j^{th} calendar month of calving.
HF_k	=	The k^{th} Holstein Friesian genetic group. Nine groups were defined.
e_{ijkl}	=	The random residual associated with each record.

Hys with <3 records were deleted, leaving 6803 cows and 1711 hys. Solutions were obtained by Gauss-Seidel iteration as described by Schaeffer and Kennedy (1986). The data, after adjusting for fixed effects, were used in subsequent analysis.

Because the likelihood functions described above assume underlying normal distributions, the distributions of the adjusted production records for each trait were tested for deviation from normality using the Kolmogorov D statistic (SAS, 1982). Further, for the distributions of the adjusted production records for each trait the skewness and kurtosis were calculated (SAS, 1982).

Milk protein genotypes of sires.

Milk protein genotypes of sires were not known, and therefore genotypes of sires were inferred from information on 10,151 daughters, as described by Bovenhuis and Van Arendonk (1991). The model accounted for differences in milk protein gene frequencies between Dutch Friesian and Holstein Friesian breeds and for the probability of misclassifications. Given the genotypes of the

daughters and the parameters, for each sire the likelihoods for all possible genotypes were calculated. Sires were assigned the genotype with the highest likelihood.

RESULTS

Simulation study.

Mean ML parameter estimates and empirical standard deviations of estimates computed from 25 replicates of simulated data are given in Table 1. The simulated marker genotype effects of -75 and -200 for the M^1M^2 and M^2M^2 genotypes, respectively, agree closely with the mean estimated values of -60 and -209. Empirical standard deviations of the estimated M^1M^2 and M^2M^2 genotype effects were 41 and 69, respectively. The M^2M^2 genotype effect was estimated less accurately because there were less M^2M^2 observations than M^1M^2 observations. The difference between the estimated Q^1Q^1 and Q^2Q^2 genotype effects was 1609, which was larger than the simulated difference of 1400. The empirical standard errors of the estimates were 75, 86, and 90 for Q^1Q^1, Q^1Q^2 and Q^2Q^2 genotypes, respectively. The estimated within sire standard deviation, for a given marker-QTL genotype of the daughter, was 676. This was slightly lower than the simulated value of 702.4. The mean estimated frequency of the Q^1-allele was 0.48, as compared to the simulated value of 0.5. The mean estimate for the recombination fraction was 0.223, which is close to the simulated recombination fraction of 0.20. The empirical standard deviation of the estimate was 0.10.

Descriptive statistics.

Observed milk protein genotype frequencies for daughters and estimated genotypes for sires are given in Table 2. Linked QTL effects were not computed for α_{s1}-casein because frequency of the rare allele, C, was very low. No sires were homozygous for this allele, and only one sire was heterozygous for α_{s1}-casein. Frequencies for ß-casein BB, A^1A^3, A^2A^3 and A^3B genotypes were also very low. Nearly 97% of the cows had A^1A^1, A^1A^2, A^2A^2, A^1B and A^2B ß-casein genotypes. Only these genotypes were included in the ML analysis. More than 50% of the sires were heterozygous for ß-casein. The κ-casein A allele was the most common, and 28% of the sires were heterozygous. Frequencies for the A and B ß-lactoglobulin alleles were similar, and 45% of

Table 1. Mean maximum likelihood estimates and empirical standard deviations of the estimates[1] from a simulation study over 25 replicates.

Parameter	Simulated	Estimated	Standard dev.
Marker genotype M^1M^1	0.0	0.0	0.0
Marker genotype M^1M^2	-75.	-60	41.
Marker genotype M^2M^2	-200.	-209.	69.
QTL genotype Q^1Q^1	700.	925.	75.
QTL genotype Q^1Q^2	75.	163.	86.
QTL genotype Q^2Q^2	-700.	-684.	90.
Standard dev. (σ)	702.4	676.5	18.1
Frequency QTL allele Q^1	0.500	0.484	0.043
Recomb. fraction (r)	0.200	0.223	0.104

[1)]Sampling standard deviation based on estimates of 25 replicates.

the sires were heterozygous.

The basic statistics of the production traits before and after adjustment for fixed effects are given in Table 3. As expected, the means after adjustment for fixed effects are equal to zero, and the standard deviations are lower. Skewness was close to zero, but kurtosis was positive for all traits. The Kolmogorov D statistic indicated that the distributions for milk and protein yield and protein percent differed significantly from normality ($p < 0.05$). Thus these traits may not accurately fit the likelihood functions used to estimate linked QTL-effects.

Maximum likelihood estimation.

The ML estimates for effects on production traits of milk protein genotypes and QTLs linked to the milk protein genes are given in Tables 4, 5, and 6 for ß-lactoglobulin, κ-casein, and ß-casein, respectively. Significant effects for both the ß-lactoglobulin locus ($p = 0.04$), and a linked QTL ($p = 0.017$) were found for fat percent. No significant effects were found for the other traits. AB and

Table 2. Genotype frequencies (in %) of α_{s1}-casein, ß-casein, κ-casein and ß-lactoglobulin for 6803 daughters and 114 sires (inferred from information on 10,151 daughters).

Milk Protein Genotype		Daughters	Sires
α_{s1}-Casein	BB	96.4	99.1
	BC	3.6	0.9
ß-Casein	A^1A^1	29.3	22.8
	A^1A^2	42.0	40.3
	A^2A^2	11.2	24.6
	A^1B	10.1	6.1
	A^2B	5.0	5.3
	BB	0.5	0.9
	A^1A^3	0.9	0.0
	A^2A^3	0.6	0.0
	A^3B	0.3	0.0
κ-Casein	AA	63.6	68.4
	AB	32.6	28.1
	BB	3.7	3.5
ß-Lactoglobulin	AA	19.2	12.3
	AB	50.5	45.6
	BB	30.2	42.1

BB cows produced milk with 0.02% and 0.05% higher fat percent than the AA cows. The estimated QTL genotype difference in fat percent between the homozygotes was 0.52%. The ML estimate for r was 0.227, and the frequency of the positive Q^1 allele was 0.457. ML estimates for r within the parameter space of 0 to 0.5 were also found for milk and protein percent.

κ-Casein had a significant direct effect only on protein percent ($p < 0.001$),

Table 3. The mean and the standard deviation before adjusting the data for fixed effects and the mean, the standard deviation, the skewness and the kurtosis of the distribution after adjusting for fixed effects based on the 6803 records included in the analysis.

	Before		After			
Trait	Mean	σ	Mean	σ	Skewness	Kurtosis
Kg milk	6137.8	1009.6	-0.05	686.9	0.089	0.547
Fat %	4.56	0.445	-0.00	0.352	0.034	0.160
Protein %	3.46	0.191	0.00	0.147	0.060	0.087
Kg fat	278.5	43.16	-0.00	28.14	-0.024	0.334
Kg protein	211.9	32.81	-0.00	20.38	-0.015	0.639

and a significant linked QTL was found only for fat percent ($p = 0.028$). κ-Casein AB and BB cows had a 0.02% and a 0.07% higher protein content, respectively, than κ-casein AA cows. The ML estimate for r between κ-casein and the QTL was 0.266. The homozygotes QTL genotypes differed 0.54% in fat content. The frequency of the positive allele was 0.42. ML estimates for r were within the parameter space for all traits, except for protein percent.

ß-Casein genotypes had significant effects on fat and protein percent, $p = 0.014$ and $p < 0.001$, respectively. Cows with the ß-casein A^1B genotype had a 0.05 higher fat percent than ß-casein A^1A^1 cows, whereas ß-casein A^1A^2 and A^2B had a lower fat percent than ß-casein A^1A^1 cows. Cows with the ß-casein A^2 and B-allele had a higher protein percent than cows with the ß-casein A^1-allele. A significant linked QTL effect was found only for fat percent ($p = 0.012$). The estimated difference between the homozygote QTL genotypes was 0.53% fat. The likelihood was maximum for a recombination fraction 0.232, and the estimated frequency of the Q^1-allele was 0.42. Protein percent and kg fat also had ML estimates for r within the parameter space.

Likelihood as a function of r and maximized for the other parameters is shown for the five traits analyzed in Figures 1, 2, and 3 for ß-lactoglobulin, κ-casein and ß-casein, respectively. The likelihood values for significant difference from the null hypothesis of $r=0.5$ for $p < 0.1$ and $p < 0.05$ are denoted in the figures by horizontal lines. For all three significant effects, the ML for r is above the 0.05 significance level, but not for the non-significant effects. Ten percent support intervals are about $\pm$ 0.15 recombination frequency for all three significant QTL's.

DISCUSSION

The model.

Verification of the model by simulation showed that the model provides good estimates, although QTL-effects were somewhat overestimated in the simulation study. The estimated difference between the homozygotes was 1609, whereas the simulated difference was 1400. Information on the QTL genotype effects comes from the difference between homozygous daughter groups of sires heterozygous for the marker. The expectation of this difference for complete additivity is: (1-2r)a, where a is half the difference between the means of the two QTL homozygotes. Based on the simulated parameters, a difference between M^1M^1 and M^2M^2 daughters of 420 would be expected. The estimated

Table 4. Maximum likelihood estimates of ß-lactoglobulin genotypes and parameters that refer to a linked QTL on milk production traits.

Parameters	Milk production trait				
	Kg milk	Fat %	Protein %	Kg Fat	Kg Protein
ß-lactoglobulin AA	0.0	0.0	0.0	0.0	0.0
ß-lactoglobulin AB	-21.7	0.019	0.012	1.48	-0.60
ß-lactoglobulin BB	-73.9	0.049	0.012	1.77	-0.93
QTL genotype Q^1Q^1	953.1	0.248	0.109	15.59	24.71
QTL genotype Q^1Q^2	227.9	0.007	-0.012	-0.99	5.19
QTL genotype Q^2Q^2	-203.4	-0.275	-0.118	-16.43	-6.03
Standard deviation (σ)	609.7	0.298	0.123	25.79	18.50
Frequency QTL allele Q^1	0.261	0.457	0.474	0.494	0.272
Recombination fraction (r)	0.225	0.227	0.237	0.500	0.501
Likelihood at maximum	-53550.82	-2329.32	3540.79	-32061.38	-29847.14
Significance marker effect	p = 0.065	p = 0.040	p = 0.252	p = 0.411	p = 0.560
Significance QTL effect	p = 0.093	p = 0.017	p = 0.145	p = 1.000	p = 1.000

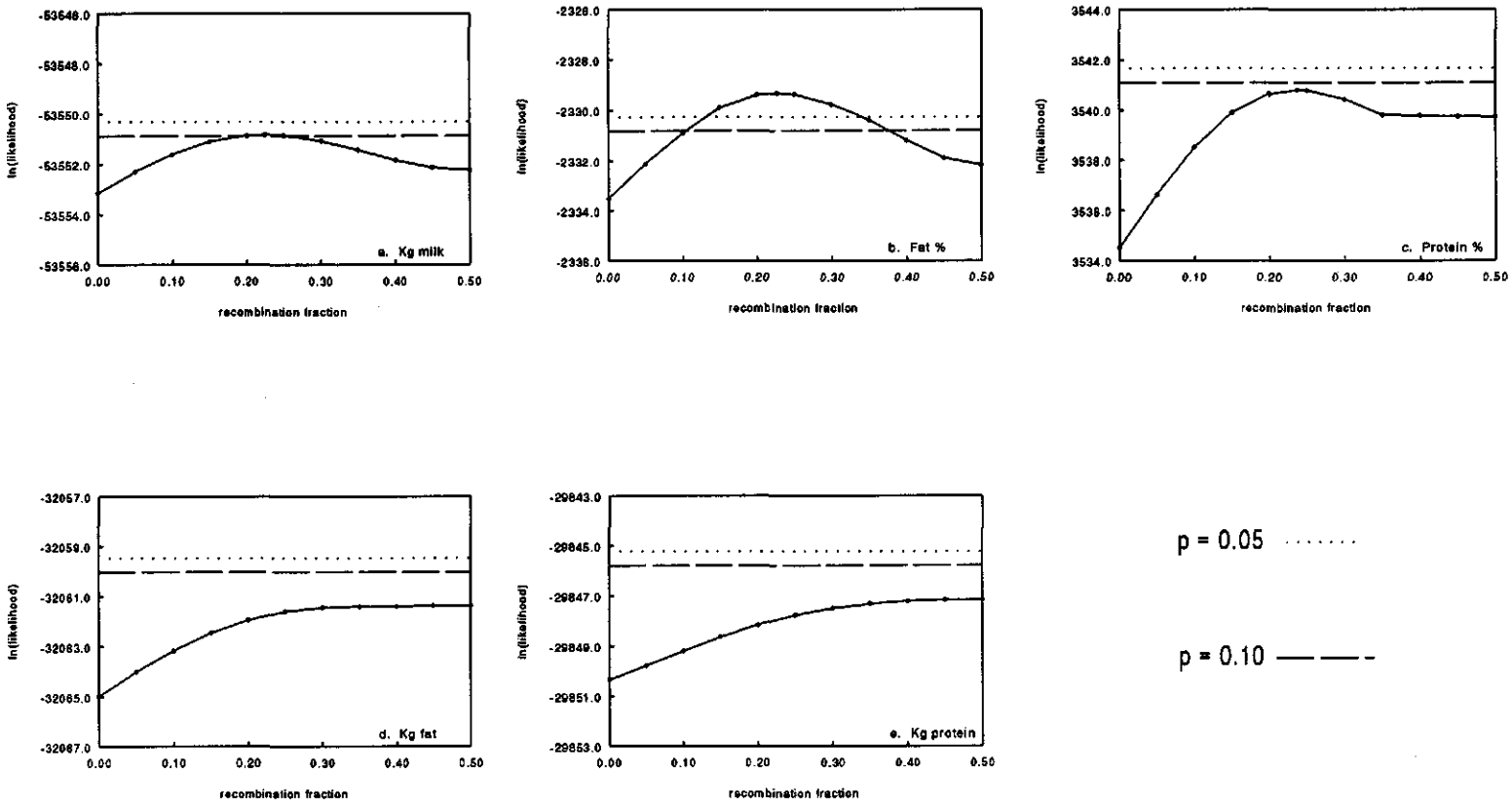


Figure 1. The likelihood as a function of r and maximized for the other parameters for the five traits analyzed when ß-lactoglobulin was used as a genetic marker. The likelihood values for significant differences from the null hypothesis of r=0.5 for $p < 0.1$ and $p < 0.05$ are denoted by horizontal lines.

Table 5. Maximum likelihood estimates of κ-casein genotypes and parameters that refer to a linked QTL on milk production traits.

Parameters	Milk production trait				
	Kg milk	Fat %	Protein %	Kg Fat	Kg Protein
κ-casein AA	0.0	0.0	0.0	0.0	0.0
κ-casein AB	-5.1	0.006	0.017	-1.45	0.42
κ-casein BB	-67.5	-0.017	0.069	-2.89	0.43
QTL genotype Q^1Q^1	1058.6	0.302	0.118	18.04	25.72
QTL genotype Q^1Q^2	252.3	0.050	-0.004	0.23	4.88
QTL genotype Q^2Q^2	-211.0	-0.239	-0.108	-12.66	-5.75
Standard deviation (σ)	609.1	0.296	0.123	25.97	18.72
Frequency QTL allele Q^1	0.218	0.423	0.446	0.480	0.241
Recombination fraction (r)	0.339	0.266	0.500	0.461	0.365
Likelihood at maximum	-53710.53	-2345.72	3593.62	-32148.41	-29938.78
Significance marker effect	$p = 0.423$	$p = 0.741$	$p < 0.001$	$p = 0.215$	$p = 0.794$
Significance QTL effect	$p = 0.365$	$p = 0.028$	$p = 1.000$	$p = 0.889$	$p = 0.750$

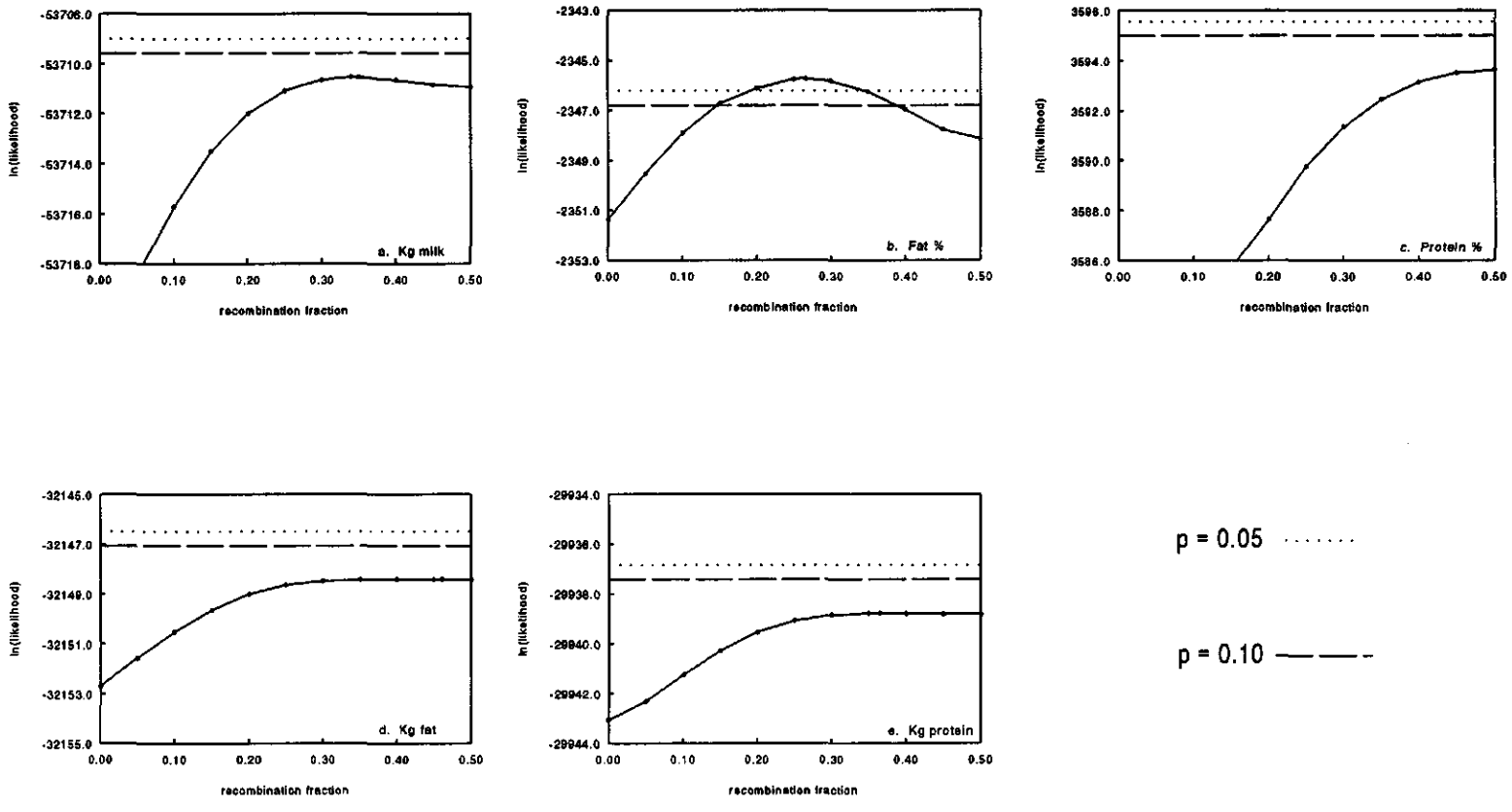


Figure 2. The likelihood as a function of r and maximized for the other parameters for the five traits analyzed when κ-casein was used as a genetic marker. The likelihood values for significant differences from the null hypothesis of r=0.5 for $p < 0.1$ and $p < 0.05$ are denoted by horizontal lines.

Table 6. Maximum likelihood estimates of ß-casein genotypes and parameters that refer to a linked QTL on milk production traits.

Parameters	Milk production trait				
	Kg milk	Fat %	Protein %	Kg Fat	Kg Protein
ß-casein A^1A^1	0.0	0.0	0.0	0.0	0.0
ß-casein A^1A^2	22.7	-0.028	0.011	-1.44	1.36
ß-casein A^2A^2	-0.3	0.009	0.019	0.33	0.86
ß-casein A^1B	-39.3	0.051	0.043	-5.18	-0.12
ß-casein A^2B	39.3	-0.016	0.030	-0.69	2.70
QTL genotype Q^1Q^1	1018.0	0.303	0.121	18.35	27.73
QTL genotype Q^1Q^2	221.0	0.066	0.003	4.57	4.88
QTL genotype Q^2Q^2	-222.7	-0.224	-0.105	-12.79	-6.17
Standard deviation (σ)	611.4	0.299	0.124	25.82	18.64
Frequency QTL allele Q^1	0.225	0.420	0.405	0.428	0.217
Recombination fraction (r)	0.499	0.232	0.385	0.265	0.500
Likelihood at maximum	-52169.59	-2289.11	3468.43	-31212.59	-29065.37
Significance marker effect	p = 0.310	p = 0.014	p < 0.001	p = 0.166	p = 0.133
Significance QTL effect	p = 1.000	p = 0.012	p = 0.439	p = 0.371	p = 1.000

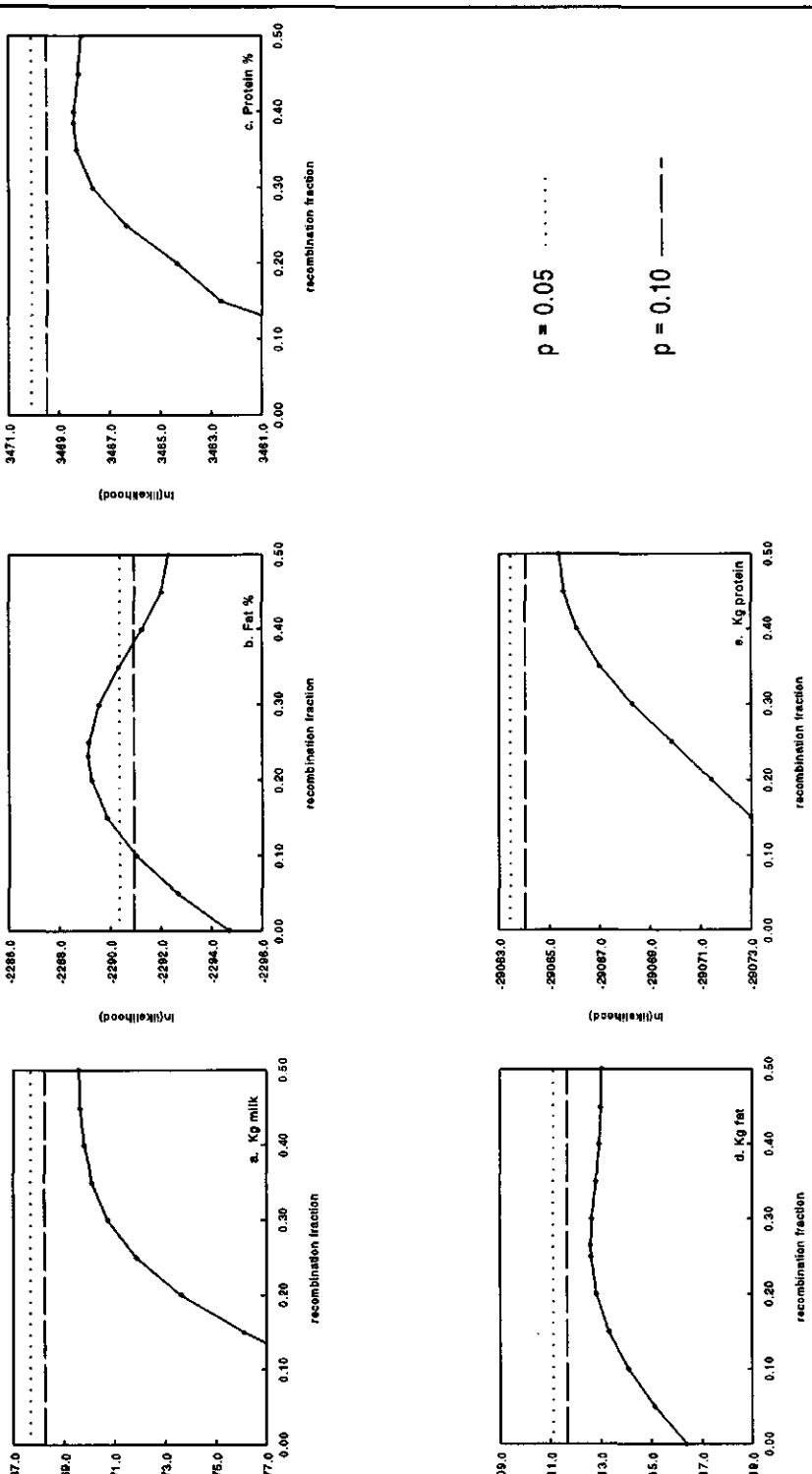


Figure 3. The likelihood as a function of r and maximized for the other parameters for the five traits analyzed when ß-casein was used as a genetic marker. The likelihood values for significant differences from the null hypothesis of r=0.5 for $p < 0.1$ and $p < 0.05$ are denoted by horizontal lines.

parameters suggest a difference of 446. This differs only slightly from the value calculated based on simulated parameters. It illustrates, however, that estimates for the recombination fraction and the QTL genotype effects are correlated, and that both are sensitive to small changes in the difference between M^1M^1 and M^2M^2 daughters of a double heterozygous sire. This is also reflected in the high empirical standard errors found for these parameters in the simulation study. The 95% confidence interval for recombination fraction included nearly the entire possible parameter space. The frequency of the QTL allele was accurately estimated, and with a low standard deviation. This is the first time that this parameter has been estimated by ML.

Robustness of the model against violation of the assumptions was not tested. The model assumed only a single QTL with two alleles linked to the genetic marker. To violate this assumption for a single QTL, it would be necessary to postulate a locus with at least three alleles, all of relatively high frequencies, and that at least four of the possible genotypes would have measurably different effects on the quantitative trait. However, a very similar situation, with respect to the ML model, could arise by two closely linked loci.

The assumption that the sire population is in Hardy-Weinberg equilibrium with respect to the genotype frequencies of the QTL becomes less important as the number of daughter records increases. With many daughter records, the prior information on the QTL genotype of the sire is overwhelmed by the information from the daughters. The model further assumed linkage equilibrium between the marker and the QTL. If this assumption is incorrect, the model will not be able to separate accurately marker and QTL genotype effects. This is because estimated QTL genotype effects are affected by marker genotype effects if coupling and repulsion phases do not appear at equal frequencies. It is expected that in this case marker genotype effects will be overestimated and QTL genotype effects will be underestimated, regardless of the number of records analyzed.

The model further assumed that the data analyzed is a random sample from the population. Proven sires are highly selected, and therefore the frequency of economically favorable QTL alleles is expected to be higher. This should not result in serious problems with many daughter records. The prior sire estimates will be overwhelmed by the posterior information of daughters.

Direct Effects of the Milk protein loci.

The significant effect of ß-lactoglobulin on fat percent found in this study is

consistent with previous results (Aleandri et al., 1990, Bovenhuis et al., 1992, Graml et al., 1986, Haenlein et al., 1987, Ng-Kwai-Hang et al., 1984). In all these studies ß-lactoglobulin BB genotype cows had a higher fat percent. A significant direct effect of κ-casein on protein percent was reported in other studies (Aleandri et al., 1990, Bovenhuis et al., 1992, Gonyon et al., 1987, Ng-Kwai-Hang et al., 1984). As found in the present study, κ-casein BB genotype cows produced milk with a higher protein percent than κ-casein AA cows. Some studies indicate that the κ-casein B allele is associated with a higher κ-casein production which would explain the higher protein content (Van Eenennaam and Medrano, 1991, Van den Berg et al., 1992).

Significant effects of ß-casein genotypes on fat % and protein % were also found. The ß-casein and κ-casein genes are closely linked (Grosclaude et al., 1973, Hines et al., 1981, Threadgill and Womack, 1990). Therefore, if these casein genes are in linkage disequilibrium, effects found for ß-casein genes might be due to the linked κ-casein gene, or *visa versa*. Bovenhuis et al. (1992) showed that in the present data set ß-casein and κ-casein are in linkage disequilibrium, and that the effect of ß-casein genotypes on protein % is in fact an effect of the closely linked κ-casein gene. However, the effect of ß-casein genotypes on fat % could not be explained by linkage disequilibrium to κ-casein.

Effects of linked QTL in Field Data.

For 15 analyses, it is expected that one analysis should display significance at the five percent level purely by chance (Weller et al., 1988). In fact, three significant linked QTL-effects were found, but all with respect to fat percent. Two of these effects were nearly significant at the one percent level. Thus it is justified to assume that these are "true" effects. For the non-significant effects, r was outside or on the edge of the parameter space in five cases, but always at $r = 0.5$. No cases were found with ML at $r = 0$, all estimates of r were > 0.2. Apparently when both a direct marker, and linked QTL effect are postulated, the model cannot distinguish between a closely linked QTL and a direct marker effect. All ML estimates for QTL allele frequency were > 0.2.

Significant effects of QTL's linked to κ-casein and ß-casein, which are closely linked, were found for fat percent. The estimates of r, 0.266 and 0.232, for κ-casein and ß-casein, respectively, suggest that the same QTL may have been detected by both markers. The estimated frequencies of the favorable Q^1-allele, 0.423 and 0.420 for κ-casein and ß-casein, respectively; and the estimates for the QTL genotype effects presented in Tables 5 and 6 are consistent with this

conclusion. Because ß-casein is more polymorphic, it is a better genetic marker and estimates of the QTL are expected to be more accurate. This is reflected in the higher significance level for the QTL when using ß-casein as a genetic marker; p= 0.014 versus p= 0.028 when using κ-casein as a genetic marker. This is the second known case in which location of a QTL by ML in field data was independently confirmed (Weller, 1986).

A limited number of studies considered effects of genes linked to milk protein genes. Gonyon et al. (1987) used ANOVA, where marker groups were nested within sires, to detect effects of QTL linked to milk protein loci on milk production traits. Significant associations between κ-casein and fat percent (p = 0.001) and protein percent (p = 0.047) were found. However, these results could not be confirmed when using the closely linked ß-casein as a genetic marker, instead of κ-casein. Haenlein et al. (1987) performed a similar analysis in another breed, but could not confirm the results found by Gonyon et al. (1987). Significant effects by ANOVA, with marker genotypes nested within sires, could also occur due to a direct effect of the marker gene.

Geldermann et al. (1985) tested for differences for milk production traits between homozygous daughters of a heterozygous sire. Daughters of three sires were considered. Significant differences were found in milk yield and fat percent for two sires that were heterozygous for ß-lactoglobulin. For both sires the daughters that inherited the ß-lactoglobulin B allele from the sire had a lower milk production and a higher fat percent. This agrees with the effects of the ß-lactoglobulin found in the present study, although the effect on milk yield was not significant.

Cowan et al. (1992) used κ-casein and ß-lactoglobulin genetic variants to study chromosome substitution effects among 103 sons of two Holstein sires. The analysis were based on the predicted transmitting abilities of the sons for milk production traits. Among the offspring of one sire, the κ-casein B-allele was associated with a 0.079 percent decreased transmitting ability for fat percent. For ß-lactoglobulin a significant chromosome substitution effect was found on fat percent and milk protein yield. The effects of both genes found on fat percent agree with the results of the present study.

The analysis of previous studies (Cowan et al, 1992, Geldermann et al., 1985, Gonyon et al., 1987, Haenlein et al., 1987) did not distinguish between direct effects and effects of linked QTL. Further, these studies did not account for recombination between the marker and the QTL.

Besides a significant linkage relationship between the casein genes ($r < 0.06$), results by Hines et al. (1981) suggest a weak linkage relationship

between ß-lactoglobulin and κ-casein (r = 0.4). However, Threadgill and Womack (1990) assigned the casein genes and ß-lactoglobulin to different syntenic groups. If ß-lactoglobulin and the casein genes are linked, then the QTL affecting fat percent linked to ß-lactoglobulin may be the same QTL linked to the casein genes. Estimates for p and genotype effects are consistent with this possibility, and the sum of the recombination distances between the two markers and the QTL, about 0.5, is not inconsistent with the recombination frequency between the milk protein loci given above.

Based on the estimated parameters, the additive genetic variance accounted for by the QTL linked to the casein genes is 0.035. For the QTL linked to ß-lactoglobulin this value is 0.034. The additive genetic variance for fat percent in the Dutch cattle population is 0.096 (Van der Werf and De Boer, 1989). Consequently this QTL would account for about 35% of the additive genetic variation, or 28% of the phenotypic variance. Genes with effects of this magnitude or greater have been detected in commercial animal populations (Hanset, 1982), and it may be possible to detect this locus even without the use of genetic markers (Hoeschele, 1988). Boichard et al. (1990) used methods derived from segregation analysis to investigate the presence of one major gene affecting fat content. Although the null hypothesis of pure polygenic additive inheritance was always rejected, the estimated transmission probabilities for the major gene allele differed strongly from Mendelian rules.

The J-blood group system is linked to ß-lactoglobulin, with r = 0.18 (Hines et al. 1981). The effect of the J-blood group on milk production traits has also been studied (e.g. Gonyon et al., 1987, Haenlein et al., 1987, Niemann-Sorensen and Robertson, 1961). Gonyon et al. (1987) found a significant effect of the J-blood group system on protein % (P = 0.006). At a population level, associations were found between the J-blood group and fat percent (Haenlein et al., 1987; Niemann-Sorensen and Robertson, 1961). However, Gonyon et al. (1987) did not find a significant effect for fat percent in a within family analysis. It is possible that the QTL affecting fat % linked to ß-lactoglobulin could also be detected by an ML analysis using the J-blood group as a genetic marker.

CONCLUSIONS

The maximum likelihood model used in this study allowed for direct effects of a marker gene and effects of a linked QTL on a quantitative trait in a segregating population. Verification of the model by simulation showed that the

model provides good estimates. High empirical standard errors were found for the recombination fraction and the QTL genotype effects. The significant direct effects of milk protein genotypes found in this study are in agreement with results reported in literature. The significant linked QTL effects found in the present study seem to be supported by results in literature. However, the analysis of previous studies did not distinguish between direct effects and effects of linked QTL. Further these studies did not account for recombination between the marker and the QTL. The estimates of the recombination fraction, the QTL allele frequency and the estimated QTL genotype effects when using κ-casein or ß-casein as genetic markers indicate that the same linked QTL has been detected.

ACKNOWLEDGMENTS

The authors wish to express their thanks to Margaret Mackinnon for the many useful discussions, Esther Verstege for phenotyping the milk samples and the Royal Dutch Cattle Syndicate for supplying the data and financial support.

REFERENCES

Aleandri, R., L.G. Buttazzoni, J.C. Schneider, A. Caroli and R. Davoli. 1990. The effects of milk protein polymorphisms on milk components and cheese-producing ability. J. Dairy Sci. 73: 241.

Boichard, D., J.M. Elsen, P. Le Roy and B. Bonaiti. 1990. Segregation analysis of fat content data in Holstein * European Friesian crossbred cattle. Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh, Scotland. XIV: 167.

Bovenhuis, H., and A.J.M. Verstege. 1989. Improved method for phenotyping milk protein variants by isoelectric focusing using PhastSystem. Neth. Milk Dairy J. 43: 447.

Bovenhuis, H., and J.A.M. Van Arendonk. 1991. Estimation of milk protein gene frequencies in crossbred cattle by maximum likelihood. J. Dairy Sci. 74: 2728.

Bovenhuis, H., J.A.M. Van Arendonk and S. Korver. 1992. Associations between milk protein polymorphisms and milk production traits. J. Dairy Sci. (in press).

Cowan, C.M., M.R. Dentine and T. Coyle. 1992. Chromosome substitution

effects associated with κ-casein and ß-lactoglobulin in Holstein cattle. J. Dairy Sci. 75: 1097.

Eigel, W.N., J.E. Butler, C.A. Ernstrom, H.M. Farrel JR, V.R. Harwalker, R. Jenness and R. MCL. Whitney. 1984. Nomenclature of proteins of cow's milk: fifth revision. J. Dairy Sci. 67: 1599.

Geldermann, H. 1975. Investigations on inheritance of quantitative characters in animals by gene markers. I. Methods. Theor. Appl. Genet. 46: 319.

Geldermann, H., U. Pieper and B. Roth. 1985. Effects of marked chromosome sections on milk performance in cattle. Theor. Appl. Genet. 70: 138.

Gonyon, D.S., R.E. Mather, H.C. Hines, G.F.W. Haenlein, C.W. Arave and S.N. Gaunt. 1987. Associations of bovine blood and milk polymorphisms with lactation traits: Holsteins. J. Dairy Sci. 70: 2585.

Graml, R., J. Buchberger, H. Klostermeyer and F. Pirchner. 1985. Pleiotrope wirkungen von ß-lactoglobulin- und casein genotypen auf milchfett- und milchproteinmengen des Bayerischen Fleckviehs und Braunviehs. Z. Tierzüchtg. Züchtgsbiol. 103: 33.

Graml, R., J. Buchberger, H. Klostermeyer and F. Pirchner. 1986. Pleiotrope wirkungen von ß-lactoglobulin- und casein genotypen auf milchinhaltsstoffe des Bayerischen Fleckviehs und Braunviehs. Z. Tierzüchtg. Züchtgsbiol. 102: 355.

Grosclaude, F., J.C. Mercier and B.R. Ribadeau Dumas. 1973. Genetic aspects of cattle casein research. Neth. Milk Dairy J. 27: 328.

Haenlein, G.F.W., D.S. Gonyon, R.E. Mather and H.C. Hines. 1987. Associations of bovine blood and milk polymorphisms with lactation traits: Guernseys. J. Dairy Sci. 70: 2599.

Hanset, R. 1982. Major genes in animal production, examples and perspectives: cattle and pigs. Proc. 2nd World Congr. Genet. Appl. Livest. Prod., Madrid. VI: 439.

Hines, H.C., J.P. Zikakis, G.F.W. Haenlein, C.A. Kiddy and C.L. Trowbridge. 1981. Linkage relationships among loci of polymorphisms in blood and milk of cattle. J. Dairy Sci. 64: 71.

Hoeschele, I. 1988. Statistical techniques for detection of major genes in animal breeding data. Theor. Appl. Genet. 76: 311.

Jensen, J. 1989. Estimation of recombination parameters between a quantitative trait locus (QTL) and two marker gene loci. Theor. Appl. Genet. 78: 613.

Knapp, S. J. 1991. Using molecular markers to map multiple quantitative loci: models for backcross, recombinant inbred, and doubled haploid progeny.

Theor. Appl. Genet. 81: 333.

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

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

Neimann-Sorensen, A. and A. Robertson. 1961. The associations between blood groups and several production characteristics in three Danish cattle breeds. Acta Agric. Scan. 11: 163.

Nelder, J.A. and R. Mead. 1965. A simplex method for function minimization. Computer J. 7: 308.

Ng-Kwai-Hang, K.F., J.F. Hayes, J.E. Moxley and H.G. Monardes. 1984. Association of genetic variants of casein and milk serum proteins with milk, fat, and protein production by dairy cattle. J. Dairy Sci. 67: 835.

SAS, 1982 SAS users Guide: Statistics. Statistical Analysis System Institute, Inc., Cary, North Carolina.

Schaeffer, L.R. and B.W. Kennedy. 1986. Computing solutions to mixed model equations. Proc. 3rd World Congr. Genet. Appl. Livest. Prod., Lincoln, Nebraska. XII: 382.

Simpson, S. P. 1989. Detection of linkage between quantitative trait loci and restriction fragment length polymorphisms using inbred lines. Theor. Appl. Genet. 77: 815.

Soller, M. 1990. Genetic mapping of the bovine genome using DNA-level markers with particular attention to loci affecting quantitative traits of economic importance. J. Dairy Sci. 73: 2628.

Soller, M. 1991. Mapping quantitative trait loci affecting traits of economic importance in animal populations using molecular markers. In: Gene mapping: Strategies, Techniques and Applications. L. B. Schook, H. A. Lewin, and D. G. McLaren (Ed.) Marcel Dekker, Inc. New York. pp 21-50.

Soller, M., T. Brody and A. Genizi. 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet. 47: 35.

Soller, M. and A. Genizi. 1978. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. Biometrics 34: 47.

Threadgill, D. W. and J. E. Womack. 1990. Genomic analysis of the major bovine milk protein genes. Nucleic Acids Res. 18: 6935.

Van den Berg, G., J.T.M. Escher, P.J. De Koning and H. Bovenhuis. 1992. Genetic polymorphism of κ-casein and ß-lactoglobulin in relation to milk

composition and processing properties. Neth. Milk Dairy J. 46: 145.

Van der Werf, J.H.J. and W. De Boer. 1989. Estimation of genetic parameters in a crossbred population of black and white dairy cattle. J. Dairy Sci. 72: 2615.

Van Eenennaam, A. L., and J. F. Medrano. 1991. Differences in allelic protein expression in the milk of heterozygous κ-casein cows. J. Dairy Sci. 74: 1491.

Weller, J. I. 1986. Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. Biometrics 42: 627.

Weller, J. I. 1987. Mapping and analysis of quantitative trait loci in *Lycopersicon*. Heredity 59: 413.

Weller, J. I. 1990. Experimental designs for mapping quantitative trait loci in segregating populations. Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh, Scotland. XIII: 113.

Weller, J. I., M. Soller and T. Brody. 1988. Linkage analysis of quantitative traits in an interspecific cross of tomato (*L. esculentum* x *L. pimpinellifolium*) by means of genetic markers. Genetics 118: 329.

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

Wilmink, J.B.M. 1987. Studies on testday and lactation milk, fat and protein yield of dairy cows. Ph.D. Thesis. Wageningen Agric. Univ., Neth. p. 1-123.

Chapter 7

THE VALUE OF SELECTION FOR κ-CASEIN AND ß-LACTOGLOBULIN GENOTYPES IN DAIRY CATTLE BREEDING

Henk Bovenhuis and Imke J. M. de Boer
Department of Animal Breeding
Wageningen Agricultural University
P.O. Box 338
6700 AH Wageningen
The Netherlands

ABSTRACT

Milk protein genetic variants have been shown to be associated with milk production traits and manufacturing properties of milk. Therefore, milk protein genotypes might be of value as selection criteria. The aim of the present study was to quantify the potential effects of selection for κ-casein and β-lactoglobulin genotypes by using stochastic simulation of a closed adult MOET nucleus breeding scheme. The results show that selection for κ-casein or β-lactoglobulin genotypes can increase selection response in a MOET nucleus breeding scheme. For κ-casein, the gain was 9 to 22 units of INET depending on wether κ-casein genotype effects needed to be estimated from the simulated data or κ-casein genotype effects were known without error. When selection was for κ-casein genotypes, the annual genetic progress was increased by 2.4 to 4.8% in the first seven generations. The additional response obtained by including β-lactoglobulin genotypes was 16 units of INET. In the first seven generations the annual genetic progress was increased by 3.9%. For β-lactoglobulin, genotype effects were not estimated and were assumed to be known without error. If estimates of β-lactoglobulin genotype effects are inaccurate, the additional genetic response is expected to be lower. Additional gain for both milk protein genes was due to an increased selection intensity on the male side. For a progeny testing scheme, it would be expected that increased genetic gain could be obtained if including milk protein genotypes results in an increased selection intensity.

Key words: κ-casein, ß-lactoglobulin, selection, MOET nucleus scheme.

INTRODUCTION

Several studies have shown milk protein genetic variants to be associated with manufacturing properties of milk. The main findings were that κ-casein genetic variants affect renneting time of milk and ß-lactoglobulin genetic variants are associated with casein number (reviewed by Grosclaude, 1988). These two characteristics are of importance for cheese production, and for countries where a large fraction of the milk is manufactured into cheese, milk protein genotypes might be of value as additional selection criteria to improve the quality of milk for cheese production.

There are reports also of associations between milk protein genetic variants and milk production traits (e.g. Ng-Kwai-Hang et al., 1984, Graml et al., 1985, Graml

et al., 1986, Gonyon et al., 1987, Haenlein et al., 1987, Aleandri at al., 1990, Bovenhuis et al., 1992). Results from these studies indicate that κ-casein genotypes are associated with protein content and ß-lactoglobulin genotypes are related to fat content (reviewed by Bovenhuis et al., 1992). Therefore, κ-casein and ß-lactoglobulin genotypes can be used as a source of information for the estimation of breeding values for protein and fat content, respectively, and use of this information can increase genetic progress in selection for milk production traits (e.g. Smith and Simpson, 1986).

Recently, it has become possible to type animals for κ-casein and ß-lactoglobulin genotypes at the DNA-level (e.g. Medrano, 1990), which allows genotyping of both males and females at a young age. In dairy cattle breeding, accuracy of breeding value estimation is low for young animals because phenotypic records on the animal or its progeny are not available. Therefore, knowledge of milk protein genotypes will be of special use in MOET (Multiple Ovulation and Embryo Transfer) nucleus breeding schemes where selection is at a young age.

The aim of the present study was to quantify the potential effects of selection for κ-casein and ß-lactoglobulin genotypes by using stochastic simulation of a closed adult MOET nucleus breeding scheme.

MATERIAL AND METHODS

For this study, two effects of milk protein genotypes were distinguished; the effect of milk protein genotypes on milk production traits and the effects of milk protein genotypes on manufacturing properties of milk. In the first case milk protein genotypes give additional information about a trait that is currently selected for and in the second case milk protein genotypes give information about an additional trait, which is not currently selected for. The former and latter cases are denoted as quantitative and qualitative effects, respectively.

Quantitative effects of milk protein genotypes

Associations between milk protein genotypes and milk production traits were described by Bovenhuis et al. (1992). However, selection of dairy cattle in the Netherlands is on a Net-profit-index (INET), combining breeding values for milk, fat and protein yield (Dommerholt and Wilmink, 1986). To reflect this a quantitative trait was simulated using parameters that apply to INET. The current economic weights used in the INET are -0.14, 1.39 and 11.68 (in Dutch Guilders per kg) for milk, fat and protein yields, respectively. Genetic (C) and

phenotypic (**P**) variances and covariances used to estimate genetic parameters for INET were those obtained for the Dutch Black and White population by Van der Werf and De Boer (1987). Genetic variance (σ_u^2), phenotypic variance (σ_p^2) and heritability (h^2) of INET were calculated as;

$$= \boldsymbol{v}'\boldsymbol{C}\,\boldsymbol{v} = \begin{bmatrix} -0.14 & 1.39 & 11.68 \end{bmatrix} \begin{bmatrix} 186212. & 4353.1 & 4610.3 \\ 4353.1 & 299.4 & 153.8 \\ 4610.3 & 153.8 & 151.5 \end{bmatrix} \begin{bmatrix} -0.14 \\ 1.39 \\ 11.68 \end{bmatrix} = 13117$$

$$\boldsymbol{v}'\boldsymbol{P}\,\boldsymbol{v} = \begin{bmatrix} -0.14 & 1.39 & 11.68 \end{bmatrix} \begin{bmatrix} 489973. & 16037.2 & 13957.7 \\ 16037.2 & 828.6 & 527.1 \\ 13957.7 & 527.1 & 460.8 \end{bmatrix} \begin{bmatrix} -0.14 \\ 1.39 \\ 11.68 \end{bmatrix} = 39294$$

$$h^2 = \frac{13117}{39294} = 0.334$$

Effects of milk protein genotypes on INET were estimated by combining milk-, fat- and protein yield to one trait according to their economic values. This trait was analyzed assuming a heritability of 0.334 and using the data and methods as described by Bovenhuis et al. (1992). Bovenhuis et al. (1992) used a model in which each milk protein gene was analyzed separately (single gene analysis) and a model in which all milk protein genes were analyzed simultaneously (multigene analysis). When using the single gene analysis estimates for κ-casein and ß-lactoglobulin genotype effects on INET were not significant. Estimates for κ-casein and ß-lactoglobulin AA, AB and BB genotypes were 0.0, 8.0 and 2.5 and 0.0, -5.0 and -16.5, respectively. The estimated κ-casein and ß-lactoglobulin genotype effects on INET when adjusting for effects of other milk protein genes using a multigene model are in table 1. Here the effects of κ-casein genotypes were significant. Bovenhuis et al. (1992) showed that the difference in estimates of κ-casein genotype effects obtained when using the single gene and the multigene model could be explained from the linkage disequilibrium between κ-casein and ß-casein in the present Dutch Black and White dairy cattle population. The appearance of ß-casein B together with κ-casein B, caused that the negative effect of the first on yield traits was reflected in the single gene estimates for the κ-casein genotypes. Therefore, in the this study it was assumed that both κ-casein and ß-casein genotypes of animals were known, and selection was for κ-casein B

Table 1. Estimates of κ-casein and ß-lactoglobulin genotype effects on INET when using the multigene model (in Dutch Guilders).

Milk protein genotype	κ - Casein	ß - Lactoglobulin
AA	0.0	0.0
AB	20.4	-5.4
BB	36.4	-17.5
Significance	p = 0.034	p = 0.123

in combination with ß-casein A^1, A^2 or A^3. In this study estimates for κ-casein genotypes obtained from the multigene analysis were used. Because selection is additionally against ß-casein B, these effects slightly underestimate the real difference. To account for the combined selection for κ-casein and ß-casein the gene frequency was adjusted to 0.11, i.e. the frequency of the favourable κ-casein B-allele in combination with ß-casein A^1, A^2 or A^3 (Bovenhuis et al., 1992). Effects of ß-lactoglobulin genotypes on INET were not significant in the single and multigene analysis, and therefore, were not considered in the present study.

Qualitative effects of milk protein genotypes

Several studies have shown that milk protein genotypes are associated with manufacturing properties of milk (reviewed by Grosclaude, 1988). Most studies investigated effects of κ-casein or ß-lactoglobulin genotypes. Although there are some indications that α_{s1}-casein and ß-casein genotypes also affect manufacturing properties, they were not included in the present study.

Table 2 shows the effect of κ-casein and ß-lactoglobulin genotypes on the conversion of total nitrogen content of milk into cheese nitrogen, i.e. a measurement for the efficiency of cheese production (Van den Berg et al., 1992). The κ-casein B allele is associated with a slightly higher efficiency. Although κ-casein B is associated with a higher casein number (reviewed by Grosclaude, 1988), this is not clearly reflected in an association with efficiency for cheese production, because the κ-casein B allele is also associated with higher losses of

Table 2. Effects of milk protein genotypes on the conversion of total nitrogen content of milk into cheese nitrogen (in %) (Van den Berg et al., 1992).

Milk protein genotype	κ - Casein	ß - Lactoglobulin
AA	72.4	71.0
AB	72.6	72.8
BB	72.9	73.9

glycomacropeptide in the whey (Van den Berg et al., 1992). ß-lactoglobulin genotypes have a considerable effect on the conversion of milk nitrogen into cheese nitrogen (table 2), and consequently cheese yield. Van den Berg et al. (1992) explained this effect by the association between the ß-lactoglobulin alleles and casein number. The B allele of ß-lactoglobulin is related to a higher casein content and a lower whey protein content. Overall, however, this resulted in no significant effect of ß-lactoglobulin genotypes on protein content (Bovenhuis et al.,1992).

Economic values of ß-lactoglobulin genotypes were determined as follows. The starting point for calculation of the economic value of milk protein genotypes was the economic value of protein yield as it is used at present in the INET. Given the genotype frequencies in the population (Bovenhuis et al. ,1992), the average conversion (in %) for the population was;

$$0.192 \cdot 71.0 + 0.506 \cdot 72.8 + 0.302 \cdot 73.9 = 72.8$$

The economic value for milk protein was then apportioned to a part that relates to the value of protein that ends up in cheese and a part that remains in the whey;

$$v_{protein} = 11.68 = 0.728 \cdot v_{protein\ in\ cheese} + 0.272 \cdot v_{protein\ in\ whey}$$

A higher conversion results in a higher fraction of the protein that ends up in the cheese. This protein gets the value of cheese. Therefore, the cheese to whey price ratio is a factor determining the economic value of ß-lactoglobulin genotypes. In 1991 the net values of cheese and whey for industry were in the proportion of 172 to 1 (Leunis, Produktschap voor Zuivel, personal communication, 1992).

Values for $v_{protein\ in\ cheese}$ and $v_{protein\ in\ whey}$ were obtained by substitution in the equation above and were;

$$v_{protein\ in\ cheese} = 16.01 \quad \text{and} \quad v_{protein\ in\ whey} = 0.09$$

Based on these values and on the conversion of milk nitrogen into cheese nitrogen (table 2) the economic value of milk protein for ß-lactoglobulin AA, AB and BB genotypes was;

$$v_{AA;protein} = 0.710 \cdot 16.01 + 0.290 \cdot 0.09 = 11.39$$
$$v_{AB;protein} = 0.728 \cdot 16.01 + 0.272 \cdot 0.09 = 11.68$$
$$v_{BB;protein} = 0.739 \cdot 16.01 + 0.261 \cdot 0.09 = 11.85$$

This calculation of the value of milk protein for ß-lactoglobulin AA, AB and BB genotypes assumes that the value of cheese or whey does not change if it contains more or less protein. The conversion of milk nitrogen into cheese nitrogen is important only for the fraction of the milk that is manufactured into cheese. It is assumed that for other dairy products the value of casein and whey protein is equal. In 1990, in the Netherlands, over 47% of the milk was manufactured into cheese (Jaarverslag PZ, 1990). Because this fraction is still increasing, it was assumed for this study that 50% of the milk is manufactured into cheese. Therefore, on the basis of an average 305-day production per cow of 210 kg protein, 105 kg of protein would be used for cheese production. The value of milk protein produced by an average cow carrying the ß-lactoglobulin AA, AB or BB genotype was calculated as;

ß-lactoglobulin AA; $105 \cdot 11.39 = 1196.0$
ß-lactoglobulin AB; $105 \cdot 11.68 = 1226.4$
ß-lactoglobulin BB; $105 \cdot 11.85 = 1244.3$

The relative economic value of ß-lactoglobulin genotypes in Dutch guilders per cow per year was;

$$v_{AA} = -24.2 \quad v_{AB} = +6.3 \quad v_{BB} = +24.2$$

The frequency of the ß-lactoglobulin B allele in the Dutch Black and White population is 0.54 (Bovenhuis and Van Arendonk, 1991).

Economic values of κ-casein genotypes were calculated in a similar way as described for ß-lactoglobulin. There was a difference between the homozygous κ-casein genotypes of 8.4. This effect is small and was neglected in the present study.

Better curd properties associated with κ-casein B, result in a lower amount of curd fines in the whey (Van den Berg et al., 1992). This effect of κ-casein

genotypes is not included in the conversion of milk nitrogen into cheese nitrogen. However, the economic importance of this effect is negligible (Van den Berg, 1992, personal communication) and was not accounted for in the present study.

Van den Berg et al. (1992) found that κ-casein genetic variants are associated with renneting time of milk. However, these effects could to a great extent be compensated by the addition of calcium chloride. Therefore, the economic importance of the effect of κ-casein genotypes on renneting time is negligible and was not accounted for in the present study.

Simulation study

Previously it was indicated that two effects of κ-casein and ß-lactoglobulin genes need to be considered;

1) A quantitative effect of κ-casein genotypes on INET.

2) A qualitative effect of ß-lactoglobulin genotypes for cheese production.

Table 3 summarizes the effects and the gene frequencies for κ-casein and ß-lactoglobulin that were used in the simulation study. In the simulation study it was assumed that the milk protein gene itself had an effect. Effects of genes linked to milk protein genes (Bovenhuis and Weller, 1992) were not considered.

To quantify the potential effect of using information on milk protein genotypes, the selection response for a base situation, where selection is for estimated breeding values for the quantitative trait, assuming an infinitesimal model, without specific knowledge of milk protein genotypes was compared with a situation were milk protein genotype information was incorporated. The two breeding schemes where information on genotypes is used for selection were;

Situation I; selection is with knowledge of κ-casein genotypes of both males and females. Selection is for a breeding value combining information about the single gene and an infinite genetic part, both affecting the quantitative trait.

Situation II; selection is with knowledge of ß-lactoglobulin genotypes of both males and females. Selection is for a breeding value combining information about the single gene and an infinite genetic part where the former has a qualitative effect and the latter affects the quantitative trait.

Because the effects of the single gene in both situations are different (quantitative and qualitative), for each situation the results were compared with a different base situation.

Data simulation

The additive genetic value of animal i, due to an infinite number of genes each with a small effect, was simulated as;

$$u_i = x_i \cdot \sigma_u$$

where; u_i = additive genetic value of individual i due to an infinite number of genes;
x_i = i^{th} random number from a standard normal distribution;
σ_u = additive genetic standard deviation.

A milk protein genotype for each animal was simulated based on the gene frequencies in the population. Because a milk production trait was simulated, only female individuals had records. For each female one record was simulated. In situation I, a phenotypic record of female individual j was simulated according to the following model;

$$Y_j = G_k + u_j + x_j \cdot \sigma_e$$

where; Y_j = the phenotypic record of female individual j;
G_k = the fixed effect of genotype k (k=1,3; AA, AB and BB);
u_j = the additive genetic value of animal j;
x_j = j^{th} random number from a standard normal distribution ($x_i \neq x_j$);
σ_e = residual standard deviation.

For the base situation, were selection was without κ-casein genotype knowledge, the simulation of data was identical to that of situation I. The additive genetic variance (σ_u^2) in situation I and the corresponding base situation was the additive genetic variance of INET adjusted for the variance due to the single gene.

In situation II, the single gene had no effect on the quantitative trait and a phenotypic record of female individual j was simulated as;

$$Y_j = u_j + x_j \cdot \sigma_e$$

The simulation of data was the same for the base situation. In situation II and the corresponding base situation σ_u^2 was the additive genetic variance of INET.

Additive genetic variance before selection was $h^2\sigma_p^2$ while the residual variance was $(1\text{-}h^2)\sigma_p^2$. The infinite additive genetic value of animals in later generations was simulated as:

$$u_i = \frac{1}{2} u_{s_i} + \frac{1}{2} u_{d_i} + \phi_i$$

Table 3. Summary of the milk protein genotype effects on economic value (Dutch guilders) on a lactation basis and the gene frequencies used in the simulation study.

Situation I	frequency κ-Casein B-allele = 0.11	
κ-Casein AA	κ-Casein AB	κ-Casein BB
-18.2	2.2	18.2
Situation II	frequency ß-Lactoglobulin B-allele = 0.54	
ß-Lactogl. AA	ß-Lactogl. AB	ß-Lactogl. BB
-24.2	6.3	24.2

where u_s and u_d are the additive genetic values of the sire and dam of individual i and ϕ_i is the Mendelian sampling term. The variance of the Mendelian sampling term was;

$$var(\phi) = \frac{1}{2}\left[1 - \frac{1}{2}(F_s + F_d)\right] \bullet \sigma_u^2$$

where F_s and F_d are inbreeding coefficients of sire and dam, respectively and σ_u^2 is the additive genetic variance in generation 0. Inbreeding coefficients were calculated as described by Tier (1990). The transmission probability of milk protein alleles from parents to progeny was ½.

Breeding value estimation

Breeding values were estimated using an animal model accounting for all genetic relations between animals and using all records. The following model was used:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e}$$

with

$$\mathbf{E(y)} = \mathbf{X\beta}$$

$$\mathrm{Var(y)} = \mathbf{ZAZ'}\sigma_u^2 + \mathbf{I}_n\,\sigma_e^2$$

where **u** was a vector of additive genetic effects of animals and **e** was a vector of random residuals. **X** and **Z** were the design matrices for fixed and animal effects,

respectively and **A** was the additive genetic relationship matrix. For situation I, ß was a vector with estimated genotype effects whereas in the corresponding base situation ß was the population mean. In situation II and the corresponding base situation, ß was also the population mean. Best Linear Unbiased Prediction of breeding values and Best Linear Unbiased Estimates of fixed effects were obtained by solving Henderson's mixed model equations;

$$\begin{bmatrix} \tilde{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\lambda \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

where λ was calculated as:

$$\lambda = \frac{(1 - h^2)}{h^2}$$

The heritability used to calculate λ in situation I, was the heritability free from the additive genetic variance due to the single gene. In the corresponding base population, however, the heritability used to calculate λ included the additive genetic variance explained by the single gene;

Situation I;

$$h^2 = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} = 0.332$$

Base situation;

$$h^2 = \frac{\sigma_u^2 + 2p(1-p)[a+d(1-2p)]^2}{\sigma_u^2 + 2p(1-p)[a+d(1-2p)]^2 + \sigma_e^2} = 0.334$$

where 2a was the difference between the two homozygotes, d was the dominance deviation and p was the frequency of the allele with the positive effect. In situation II the single gene did not affect the quantitative trait and therefore the heritability for situation II and the corresponding base population was the same, i.e. 0.334. Solutions for the mixed model equations were obtained by Gauss-Seidel iteration as described by Schaeffer and Kennedy (1986).

Selection criteria

In both base situations selection was for $\hat{\mathbf{u}}$, using a model where ß was the population mean. When genotypes where known and the single gene had an effect on the quantitative trait, i.e. situation I, breeding values for the single gene in generation t were calculated as ;

$$V_{t;AA} = 2\left[\left[\tfrac{1}{2}f_{t;AB} + f_{t;AA}\right]\cdot GENOT_{t;AA} + \left[\tfrac{1}{2}f_{t;AB} + f_{t;BB}\right]\cdot GENOT_{t;AB} - [\mu_{pop}]\right]$$

$$V_{t;AB} = 2\left[\tfrac{1}{2}\cdot\left[\tfrac{1}{2}f_{t;AB} + f_{t;AA}\right]\cdot GENOT_{t;AA} + \left[\tfrac{1}{2}\right]\cdot GENOT_{t;AB} + \tfrac{1}{2}\cdot\left[\tfrac{1}{2}f_{t;AB} + f_{t;BB}\right]\cdot GENOT_{t;BB} - [\mu_{pop}]\right]$$

$$V_{t;BB} = 2\left[\left[\tfrac{1}{2}f_{t;AB} + f_{t;AA}\right]\cdot GENOT_{t;AB} + \left[\tfrac{1}{2}f_{t;AB} + f_{t;BB}\right]\cdot GENOT_{t;BB} - [\mu_{pop}]\right]$$

where $GENOT_{t;AA}$, $GENOT_{t;AB}$ and $GENOT_{t;BB}$ were the estimated genotype effects in generation t obtained from the model as described above, $f_{t;AA}$, $f_{t;AB}$, and $f_{t;BB}$ were the genotype frequencies in generation t and μ_{pop} was equal to;

$$\text{pop.} = (f_{0;AA}\cdot GENOT_{t;AA}) + (f_{0;AB}\cdot GENOT_{t;AB}) + (f_{0;BB}\cdot GENOT_{t;BB})$$

where $f_{0;..}$ was the genotype frequency in generation 0, i.e the base population. The breeding value of animal i was calculated by adding $\hat{\mathbf{u}}_i$ and $BV_{t;AA}$, $BV_{t;AB}$ or $BV_{t;BB}$.

In situation II, qualitative differences between genotypes were accounted for. The breeding values for the single gene were calculated in a similar way as described for situation I. However, $GENOT_{t;AA}$, $GENOT_{t;AB}$ and $GENOT_{t;BB}$ in the formula's described above were replaced by V_{AA}, V_{AB} and V_{BB}, respectively, representing the economic value of genotypes due to qualitative differences. The breeding value of an animal was calculated by addition of the breeding value for the quantitative trait and the breeding value for the single gene where $\hat{\mathbf{u}}$ was calculated using a model where ß was the population mean.

Breeding scheme

A closed adult MOET (Multiple Ovulation and Embryo Transfer) nucleus breeding scheme was simulated. A base population (generation 0) of unrelated and unselected animals consisting of 64 donor cows and 16 sires was simulated.

Each sire was mated at random to 4 donor cows resulting in 8 progeny; 4 males and 4 females. Each subsequent generation 64 donor cows and 16 sires were selected. Selection of males and females was after the 256 females had one phenotypic observation. To restrict inbreeding, the number of males selected per full sib group was restricted to one. In situations I and II it was assumed that genotypes of males and females were known at the time of selection.

RESULTS

Situation I: κ-casein genotypes

The results, which are averages over 350 replicates, are shown in figures 1-7. Figure 1 shows the change of κ-casein B frequency for the base situation and where κ-casein genotype information was used in selection, i.e. situation I. Because κ-casein B was associated with a higher INET, κ-casein B frequency increased in the base situation from 0.11 at generation 1 to 0.23 at generation 11. When genotype information was used in selection, κ-casein B frequency increased to a value of 0.77 by generation 11. Figure 1 suggested that the κ-casein B frequency asymptotically approached a value of 0.8, as a result of fixation of the A-allele in 60 out of 350 replicates by generation 11. Consequently, the κ-casein B frequency that could be attained at maximum was 0.83. For the base situation, fixation of the A-allele occurred in 83 of the replicates.

Figure 2 shows the difference in genetic level between situation I and the base situation where a value of zero reflects the genetic progress in the base situation. Difference in total genetic level was apportioned into a difference due to the single gene and a part that could be attributed to the infinite genetic part. Differences between the two breeding schemes were expressed as a percentage of the maximum, which was defined as the difference between the population genetic level due to the single gene in generation 1 ($p = 0.11$) and that when the single gene was fixed at the favourable allele ($p = 1.0$). In the case of κ-casein, the maximum (100%) was 32 units of INET. Figure 2 shows that the total genetic level was higher when information on κ-casein genotypes was used in selection. By generation 11, typing of animals for κ-casein genotypes resulted in a genetic level that was 27% of the maximum higher, as a result of a higher genetic level for the κ-casein gene (+60%) but a lower genetic level for the infinite genetic part (-33%). The genetic level due to the single gene did not reach the 100%, i.e. the level where the κ-casein gene for all replicates would be fixed for the B allele, partly because of the 60 replicates where the κ-casein gene was

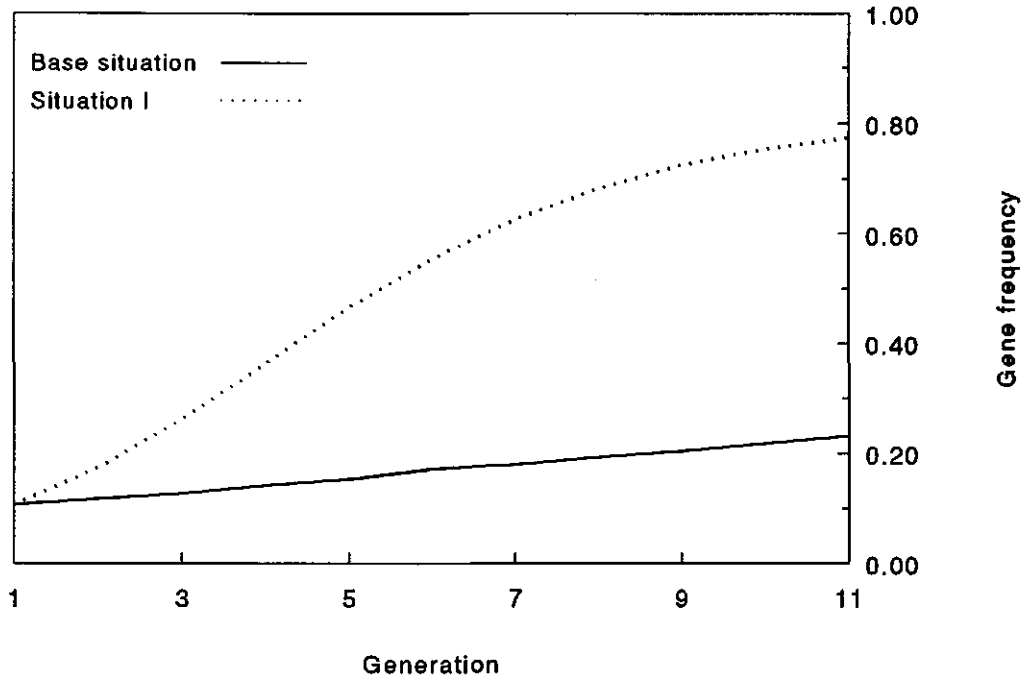


Figure 1. The κ-casein B frequency for the base situation, where selection is for INET, and for situation I, where additionally selection is for κ-casein genotypes.

fixed by drift at the A allele. For the 290 replicates where the κ-casein gene was not fixed at the A allele, κ-casein B frequency was 0.93. For those replicates the average difference due to the single gene was 79% of the maximum, i.e. 19% higher than in figure 2. In the case of complete fixation of the κ-casein gene at the B allele the κ-casein population average would become 86% of the maximum, with the still remaining 14% as a consequence of the increase in the κ-casein gene frequency in the base population.

The results are for a situation were κ-casein genotypes are estimated from the simulated data. However, estimates of genotype effects might be inaccurate due to the small amount of data, especially for κ-casein BB genotype effects in the first generations. An alternative was studied where the real κ-casein genotype effects were assumed to be known. In this alternative, $\hat{\mathbf{u}}$ was adjusted for the real genotype effects. Figure 3 shows the average differences with the base situation. At generation 11, the total genetic level for the scheme where information on κ-casein genotypes was used, was 70% of the maximum higher, i.e. 22 units of INET which was considerably higher than the value of 27% obtained when κ-

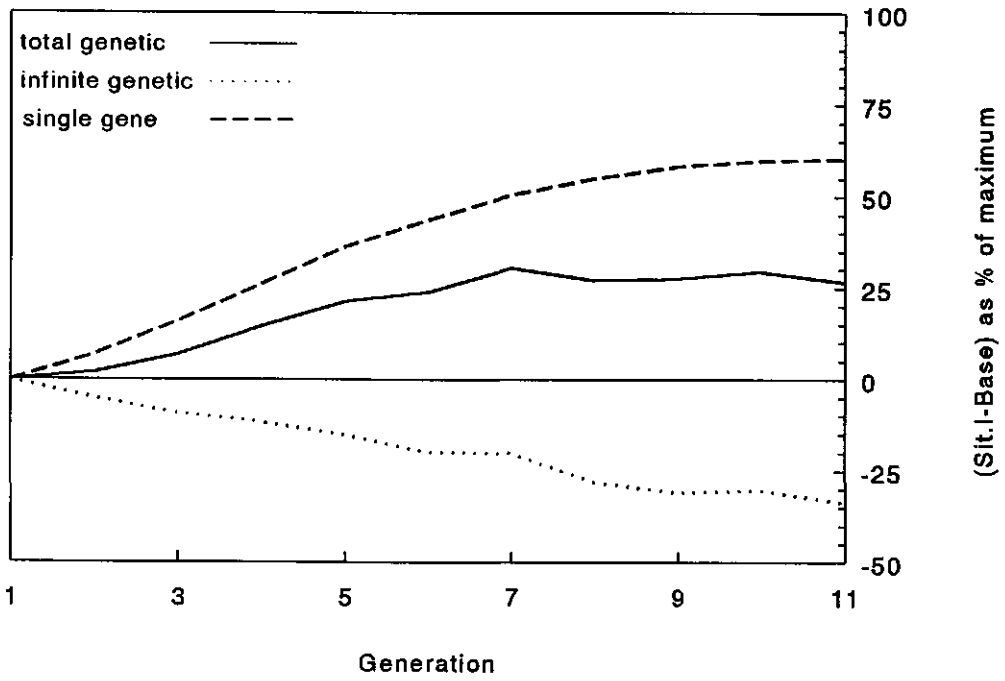


Figure 2. The differences in total genetic level, infinite genetic level and the genetic level due to the κ-casein gene, between situation I and the base situation expressed as a percentage of the maximum. The maximum is defined as the difference between the population genetic level due to the κ-casein gene in generation 1 (p = 0.11) and that when the single gene was fixed at the favourable allele (p = 1.0). For κ-casein the maximum (100%) was 32 units of INET. A value of zero reflects the genetic progress in the base situation.

casein genotype effects were estimated from the simulated data (figure 2). Because by generation 11 the κ-casein gene was fixed for the A allele only for 3 out of 350 replicates, the difference in genetic level due to the single gene reached a value of 82% of the maximum. The κ-casein B frequency at generation 11 was 0.97. In the breeding scheme where additional selection was for κ-casein genotypes, the infinite genetic level was -12% of the maximum, compared to the value of -33% in figure 2. Knowledge about the real genotype effects can decrease the reduction in progress for the infinite genetic part.

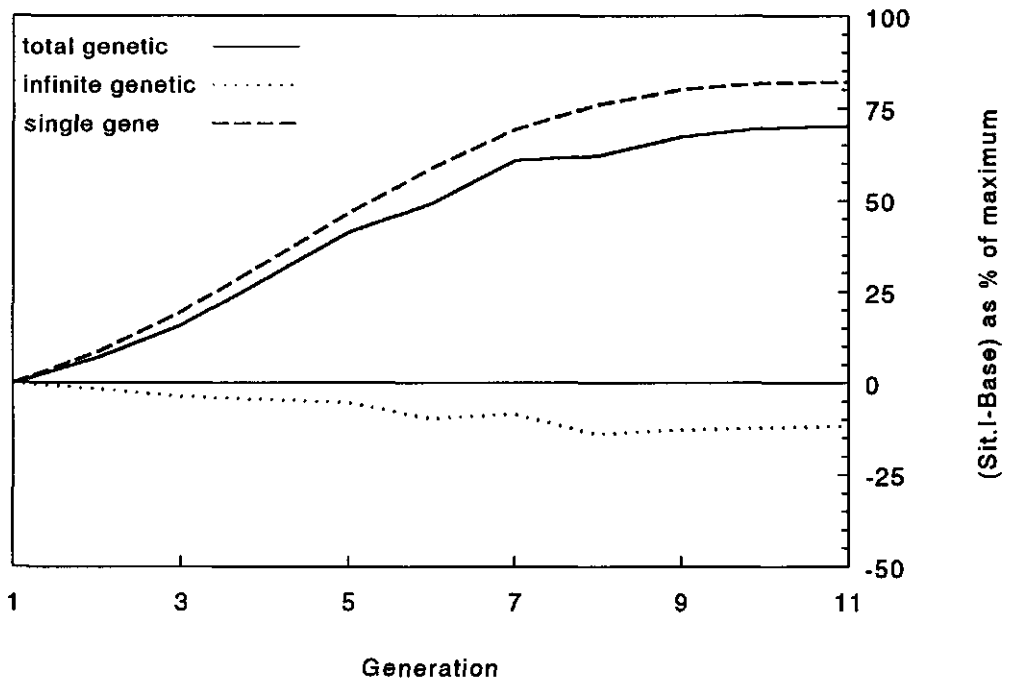


Figure 3. The differences in total genetic level, infinite genetic level and the genetic level due to the κ-casein gene, between situation I, when assuming that real κ-casein genotype effects were known, and the base situation expressed as a percentage of the maximum. The maximum is defined as the difference between the population genetic level due to the κ-casein gene in generation 1 (p = 0.11) and that when the single gene was fixed at the favourable allele (p = 1.0). For κ-casein the maximum (100%) was 32 units of INET. A value of zero reflects the genetic progress in the base situation.

Situation II: ß-lactoglobulin genotypes

Figure 4 shows the ß-lactoglobulin B gene frequency for the base situation and for situation II. In the base situation ß-lactoglobulin had no effect on the trait that was selected for and the ß-lactoglobulin B frequency did not change. In situation II, however, the economic effects of ß-lactoglobulin were selected for and the B frequency increased. At generation 4 the frequency of the B allele was 0.91 and at generation 7 the ß-lactoglobulin gene was close to fixation of the B allele.

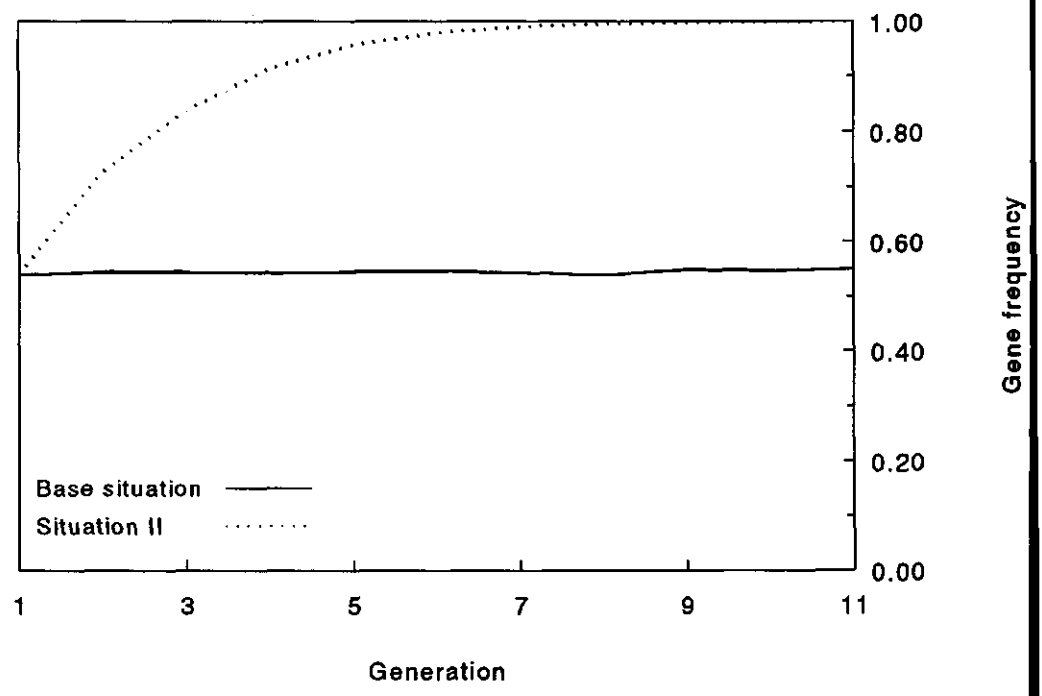


Figure 4. The ß-lactoglobulin B frequency for the base situation, where selection is for INET, and for situation II, where additionally selection is for ß-lactoglobulin genotypes.

Figure 5 shows the differences between situation II and the base situation for the ß-lactoglobulin population averages, the infinite genetic part and the sum of these two, i.e. the total economic level. The difference between the ß-lactoglobulin population level at generation 1 ($p = 0.54$) and that when the ß-lactoglobulin gene was fixed for the B allele, i.e. the maximum, is 19 units of INET. At generation 11 the total economic level was higher when the ß-lactoglobulin genotypes were taken into account. In the base population the ß-lactoglobulin B gene frequency did not change and in none of the replicates there was fixation of the B allele, and the difference in population mean due to the single gene approached the maximum. At generation 11 the difference due to the single gene was 102% which was due to a decrease in population mean for ß-lactoglobulin in the base population. Although the gene frequency was stable, inbreeding caused a reduction of the number of heterozygotes which resulted in a decreased genotype mean. Selection for ß-lactoglobulin genotypes resulted in a lower infinite genetic level (-17%) relative to the base situation.

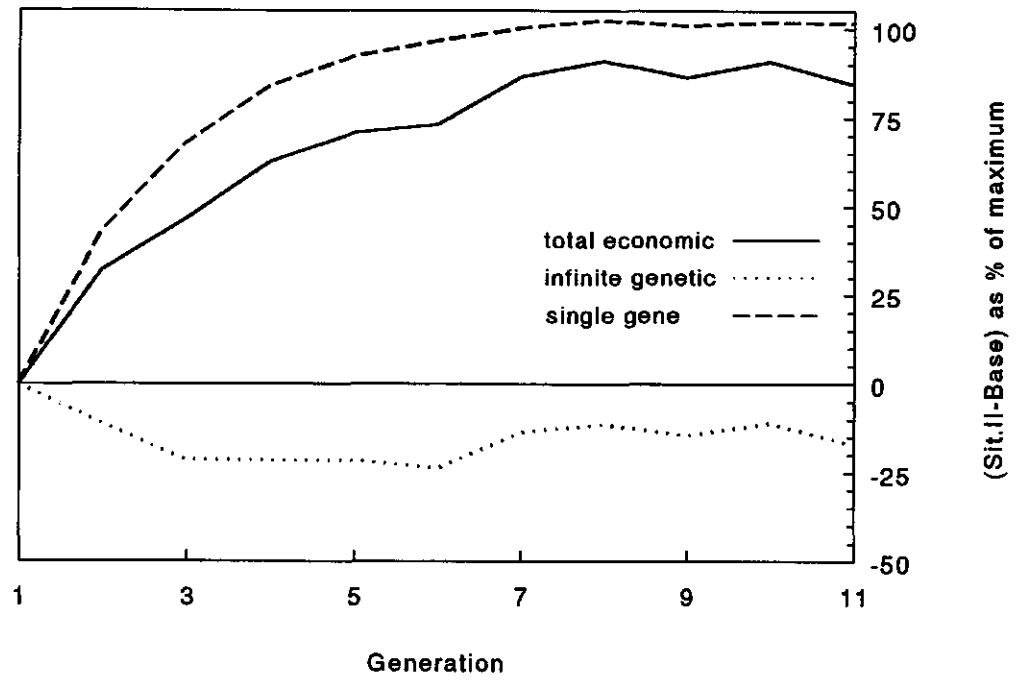


Figure 5. The differences in total economic level, infinite genetic level and the level of the ß-lactoglobulin gene, between situation II and the base situation expressed as a percentage of the maximum. The maximum is defined as the difference between the population genetic level due to the ß-lactoglobulin gene in generation 1 (p = 0.54) and that when the single gene was fixed at the favourable allele (p = 1.0). For ß-lactoglobulin the maximum (100%) was 19 units of INET. A value of zero reflects the genetic progress in the base situation.

Selection intensity

Selection for κ-casein and ß-lactoglobulin genotypes resulted in increased genetic improvement. Selection intensities, accuracies of selection, inbreeding and genetic variance in the base situations and situation I and II were examined (see de Boer and van Arendonk, 1992). In situation I, the average accuracy of selection over subsequent rounds of selection was 0.6143 and 0.4425 for females and males, respectively. In situation I accuracies were slightly lower; 0.6122 and 0.4398 for females and males, respectively. When real κ-casein genotypes were known accuracies were 0.6161 and 0.4457 for females and males, respectively, i.e.

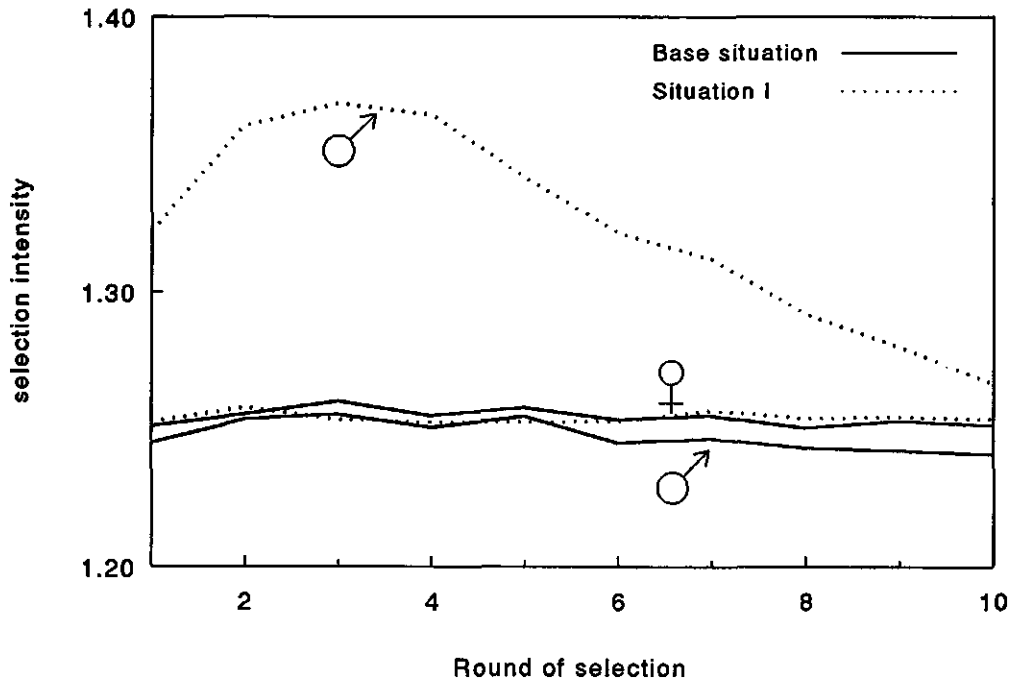


Figure 6. Realised selection intensities for male and female animals in the base situation and in situation I.

slightly higher than in the base situation. However, differences in accuracy of selection were small and did not explain observed differences in genetic progress. Additive genetic variance and inbreeding were also similar for the base situation and situation I; in generation 11 the additive genetic variance was 8448 and the average inbreeding coefficient was 0.23. Figure 6 shows the realised selection intensities for males and females for situation I and for the base situation. The realised selection intensity was calculated as the standardised difference between the estimated breeding values of the selected animals and the estimated breeding values of the selection candidates. In the base situation, selection intensities of males and females were similar; 64 females were selected out of 256 tested and for the males effectively 16 males were selected out of 64 full sib groups. Selection of males was restricted to one male per full sib group. In the base situation male full sibs had the same breeding value and therefore one male was chosen randomly. In situation I, breeding values of full sib males differed if animals had different κ-casein genotypes, and this permitted the increased selection intensity of males. Male full sibs only had different breeding values if

they had different genotypes. Therefore, selection intensities of males depended on the κ-casein gene frequency. In later rounds of selection, when the κ-casein frequency had increased, selection intensity decreased.

Figure 7 shows selection intensities for situation II and the corresponding base situation. As with κ-casein, the gain for ß-lactoglobulin was the result of increased selection intensity of males which depended on gene frequency.

DISCUSSION

Selection for κ-casein or ß-lactoglobulin genotypes has the potential to increase selection response in a MOET nucleus breeding scheme. For κ-casein the gain was 9 to 22 units of INET depending on whether κ-casein genotype effects needed to be estimated from the simulated data or κ-casein genotype effects were known without error. This corresponds with the genetic progress that can be achieved in 0.4-1 year, assuming selection after 90 days in lactation and thereby a generation interval of 3 years. Most of the gain was in the first seven generations. In this period an increased annual genetic progress of 2.4-4.8% was achieved when selection was for κ-casein genotypes. Gains from selection for a single gene are greatest if the genotype effects can be estimated accurately or if accurate estimates of genotype effects are known from previous studies. If effects of linked genes play a role information from other populations might be of less value because estimates of genotype effects might differ between populations.

The additional response obtained by including ß-lactoglobulin genotypes was 16 units of INET, which corresponds to the genetic progress attainable in 0.7 year. In the first seven generations the annual genetic progress was increased by 3.9%. For ß-lactoglobulin, genotype effects were not estimated and were assumed to be known without error. If estimates of ß-lactoglobulin genotype effects are inaccurate, the additional genetic response is expected to be lower.

The effects of selection for κ-casein or ß-lactoglobulin genotypes were studied here for an adult MOET nucleus breeding scheme. Additional gain was due to an increased selection intensity on the male side. The effect on accuracy of selection was negligible. For a progeny testing scheme, it would be expected that increased genetic gain could be obtained if including milk protein genotypes results in an increased selection intensity. When considering a situation where MOET is used to obtain young bulls, and only a limited number of males of each full sib family is progeny tested, milk protein genotypes can be used to increase selection intensity. However, if only one full sib male is available or if all full sibs are allowed to be progeny tested, no additional genetic gain from selection for milk

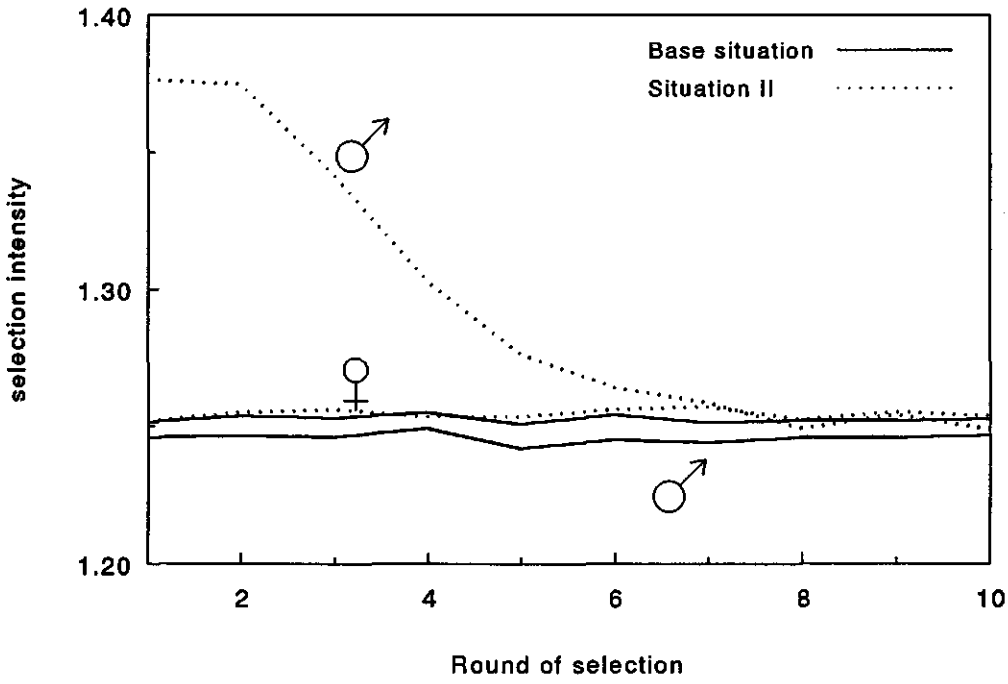


Figure 7. Realised selection intensities for male and female animals in the base situation and in situation II.

protein genotypes is expected. In a juvenile MOET nucleus breeding scheme selection on the female side is restricted to one female per full sib group and selection for κ-casein or ß-lactoglobulin genotypes will also give an increased selection intensity on the female side. In the juvenile scheme, the change of the gene frequency will be increased relative to the adult MOET scheme. Because the maximum additional selection response that can be obtained from these genes was approached in an adult scheme, the absolute additional gain is not expected to be very great.

Gibson et al. (1990) and Pedersen (1991) examined selection for κ-casein genetic variants through deterministic simulation of a progeny testing scheme. Gibson et al. (1990) considered the situations where κ-casein genotypes only affected protein yield and where κ-casein genotypes had an additional effect on cheese yield, independent of its effect on protein yield. When the κ-casein genotype only had an effect on protein yield there was no advantage in genotyping sires. The results of Gibson et al. (1990) and those of the present study disagree because Gibson et al. (1990) examined a progeny testing scheme and this study examined

a MOET scheme. Selection for milk protein genotypes in the progeny testing scheme did not increase selection intensity, and information on 50 half sib daughters was available at the time of selection and thus accuracy of selection was not substantially improved. Where an additional effect of κ-casein genotypes on cheese yield was considered, typing of animals was beneficial in a progeny testing scheme, particulary when a large fraction of the milk was manufactured into cheese. As in the present study, Gibson et al. (1990) showed that inaccurate estimates of gene effects substantially reduced genetic responses. Pedersen (1991) considered the effects of one round of combined selection for κ-casein genotypes and milk yield. It was assumed that κ-casein genotypes had an effect on the value of milk and not on milk yield. Pedersen (1991) concluded that if the additional value of κ-casein BB milk was 70 kg of milk or more, selection for increased κ-casein frequency was profitable.

Figure 1 showed that even if κ-casein genotypes were not included in selection, the κ-casein B frequency increased. This seems to be contradictory with the observed κ-casein B frequency in the Dutch Black and White dairy cattle population (Bovenhuis and Van Arendonk, 1991). However, the increase observed in the simulation study reflects the increase of κ-casein B in combination with ß-casein A^1, A^2 or A^3. As a result of the negative associations between ß-casein B and yield traits, the frequency of κ-casein B in combination with ß-casein B is expected to decrease. Overall, the κ-casein B frequency is expected to show a slight increase. However, past weighing factors for INET were different from weighing factors used in the present study. From 1973 to 1979 weighing factors for INET were 0, 1 and 1 for milk, fat and protein yields. These weighing factors result in a slightly negative association of the κ-casein B allele with INET. Consequently, selection for INET will have resulted in a slight decrease of the κ-casein B frequency. From 1979 to 1989 weighing factors were -0.182, 6.28 and 7.92 for milk, fat and protein yields, respectively. Selection for this INET would also have caused a slight decrease of the κ-casein B frequency.

It was assumed in this study that the cheese to whey price ratio was 1 to 172. This ratio might be subject to fluctuations, and how the value of ß-lactoglobulin genotypes changes as a function of the cheese to whey price ratio was investigated. Figure 8 shows the value of the ß-lactoglobulin genotypes as a deviation from the value of the ß-lactoglobulin AB genotype. The cheese to whey price ratio was varied from 0 to 180 where values below one reflect a hypothetical situation where the value of whey is higher than the value of cheese. Further, a maximum value was calculated, i.e. a situation where the value of whey is zero. The value of ß-lactoglobulin genotypes is only sensitive to changes

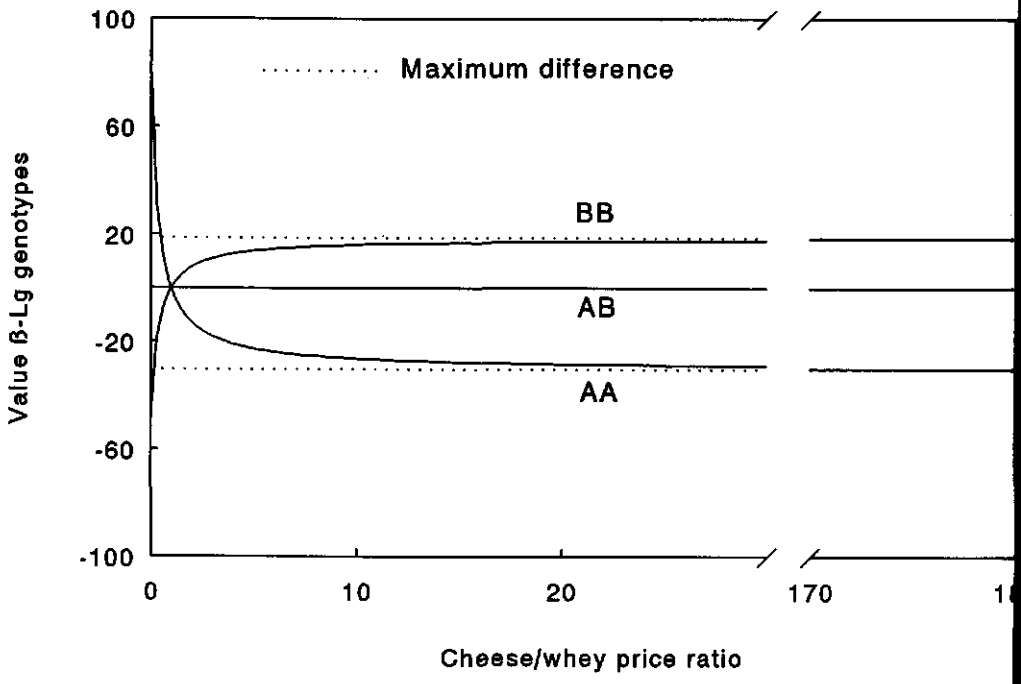


Figure 8. The value of ß-lactoglobulin genotypes as a function of the cheese to whey price ratio.

in the cheese to whey price ratio at values close to one, and the value of the genotypes stays rather constant if the cheese to whey price ratio exceeds a value of 20. The results of the present study are insensitive to a change in the cheese to whey price ratio over a broad range of values.

REFERENCES

Aleandri, R., L. G. Buttazzoni, J. C. Schneider, A. Caroli, and R. Davoli. 1990. The effects of milk protein polymorphisms on milk components and cheese-producing ability. J. Dairy Sci. 73:241.

Boer, I. J. M. de, and J. A. M. Van Arendonk. 1992. A stochastic model to study additive and dominance clonal responses in closed dairy cattle nucleus breeding schemes. In preparation.

Bovenhuis, H. and J. A. M. Van Arendonk. 1991. Estimation of milk protein gene frequencies in crossbred cattle by maximum likelihood. J. Dairy Sci. 74:2728.

Bovenhuis, H., J. A. M. Van Arendonk and S. Korver. 1992. Associations between

milk protein polymorphisms and milk production traits. Accepted for publication in J. Dairy Sci.

Bovenhuis, H. and J.I. Weller. 1992. Mapping and analysis of dairy cattle quantitative trait loci by maximum likelihood methodology using milk protein genes as genetic markers. Submitted for publication.

Dommerholt, J. and J.B.M. Wilmink. 1986. Optimal selection responses under varying milk prices and margins for milk production. Livest. Prod. Sci. 14:109.

Gibson, J. P., G.B. Jansen and P. Rozzi. 1990. The use of κ-casein genotypes in dairy cattle breeding. Proc. 4rd World Congr. Genet. Appl. Livest. Prod., Edinburgh. XIV:163.

Gonyon, D. S., R. E. Mather, H. C. Hines, G. F. W. Haenlein, C. W. Arave, and S. N. Gaunt. 1987. Associations of bovine blood and milk polymorphisms with lactation traits: Holsteins. J. Dairy Sci. 70:2585.

Graml, R., J. Buchberger, H. Klostermeyer, and F. Pirchner. 1985. Pleiotrope wirkungen von ß-lactoglobulin- und casein genotypen auf milchfett- und milchproteinmengen des Bayerischen Fleckviehs und Braunviehs. Z. Tierz. Zuechtgsbiol. 103:33.

Graml, R., J. Buchberger, H. Klostermeyer, and F. Pirchner. 1986. Pleiotrope wirkungen von ß-lactoglobulin- und casein genotypen auf milchinhaltsstoffe des Bayerischen Fleckviehs und Braunviehs. Z. Tierz. Zuechtgsbiol. 102:355.

Grosclaude, F. 1988. Le polymorphisme génétique des principales lactoprotéines bovines. Inst. Nat. Rech. Agron. Prod. Anim. 1:5.

Haenlein, G. F. W., D. S. Gonyon, R. E. Mather, and H. C. Hines. 1987. Associations of bovine blood and milk polymorphisms with lactation traits: Guernseys. J. Dairy Sci. 70:2599.

Produktschap voor zuivel. Statistisch jaaroverzicht 1990.

Medrano, J. V. 1990. Application of the polymerase chain reaction procedure for genetic evaluation in cattle. Proc. 4th World Congr. Genet. Appl. Livest. Prod., Edinburgh, Scotland. XIII:71.

Ng-Kwai-Hang, K. F., J. F. Hayes, J. E. Moxley, and H. G. Monardes. 1984. Association of genetic variants of casein and milk serum proteins with milk, fat, and protein production by dairy cattle. J. Dairy Sci. 67:835.

Pedersen, J. 1991. Selection to increase frequency of kappa-casein variant B in dairy cattle. J. Anim. Breed. Genet.

Schaeffer, L. R. and B. W. Kennedy. 1986. Computing solutions to mixed model equations. Proc. 3rd World Congr. Genet. Appl. Livest. Prod., Lincoln, Nebraska. XII:382.

Smith, C. and S. P. Simpson. 1986. The use of genetic polymorphisms in livestock improvement. J. Anim. Breed. and Genet. 103: 205.

Tier, B. Computing inbreeding coeeficients quickly. Gen. Sel. Evol., 22:419.

Van den Berg, G., J. T. M. Escher, P. J. de Koning and H. Bovenhuis. 1992. Genetic polymorphism of κ-casein and ß-lactoglobulin in relation to milk composition and processing properties. Neth. Milk Dairy J. 46: 145.

Van der Werf, J. H. J. and W. De Boer. 1989. Estimation of genetic parameters in a crossbred population of black and white dairy cattle. J. Dairy Sci. 72:2615.

SUMMARY

SUMMARY

Cheese production is of great importance for Dutch dairy industry, and milk protein genotypes affect cheese manufacturing properties. Consequently, the question can be raised whether, in addition to traditional selection, selection should be for milk protein genotypes to improve the quality of milk for cheese production. In addition to the effect on cheese manufacturing properties, the economic value of milk protein genotypes also depends on their associations with milk production traits. Furthermore, mapping and the analysis of genes affecting milk production traits has a potential use in animal breeding.

Chapter 2 provides the reader with a brief overview of what is known about milk proteins and milk protein genetic variants.

In chapter 3 a rapid method for the phenotyping of milk protein variants has been described. The method is based on the separation of milk protein variants by isoelectric focusing using PhastSystem. The method is suitable for phenotyping a large number of samples due to its short separation time and its high capacity.

In chapter 4 a maximum likelihood method is presented to estimate the fraction of animals misclassified and breed effects for milk protein gene frequencies based on crossbred data. Results of a simulation study indicate that the method proposed provides estimates of gene frequencies which agree closely with the true values. Estimates of gene frequencies in the Dutch Black and White and the Dutch Red and White crossbred populations, based on data on 10151 and 580 animals, respectively, were presented. Dutch Friesian and Holstein Friesian breeds differ for ß-casein and ß-lactoglobulin gene frequencies. Estimates for fractions misclassified were zero for α_{s1}-casein, 0.09 for ß-casein and ß-lactoglobulin and 0.12 for κ-casein. Differences between the Dutch Red and White and Red Holstein Friesian breeds were small. Estimates for fractions misclassified were high but have high approximated standard errors. Compared to the Black and White breeds the Red and Whites had a high κ-casein B gene frequency.

The aim of chapter 5 was to estimate associations between milk protein genotypes and milk production traits. For this purpose, data of 6803 first lactation cows were used. To obtain an exact test of associated hypotheses and unbiased estimates of genotype effects for a population containing related animals, an animal model was used. Milk protein genotype effects were estimated using a model in which each milk protein gene was analyzed separately (single-

gene analysis) and a model in which all milk protein genes were analyzed simultaneously (multigene analysis). The results of the two models indicate that some of the effects ascribed to certain milk protein genes in the single-gene analysis are not effects of the milk protein gene itself but of linked genes. From the results of the present study and from literature, it was concluded that the κ-casein gene or (a) very closely linked gene(s) affects protein percentage, whereas the ß-lactoglobulin gene or (a) very closely linked gene(s) affects fat percentage. Further, significant effects of ß-casein genotypes on milk yield, fat percentage, and protein yield were found, whereas ß-lactoglobulin genotypes had significant effects on milk yield and protein yield. For these effects, it is less clear whether they are due to effects of milk protein genes themselves or to effects of linked genes.

In chapter 6 maximum likelihood methodology was used to estimate effects of both a marker gene and a linked quantitative trait locus (QTL) on quantitative traits in segregating populations. Two alleles were assumed for the QTL. In addition to the genotype effects on the mean of the quantitative trait, recombination frequency between the loci, frequency of the QTL alleles, and the within marker and QTL standard deviation were also estimated. Thus six parameters were estimated in addition to the marker genotype means. The statistical model was tested on simulated data, and used to estimate direct and linked effects of the milk protein genes, ß-lactoglobulin, κ-casein, and ß-casein, on milk, fat, and protein yield and fat and protein percent in the Dutch dairy cattle population. The simulation study showed that the model provides estimates that agree with the simulated values. High empirical standard errors were found for the recombination fraction and the QTL genotype effects. In field data a significant direct effect of ß-lactoglobulin on fat percent was found. κ-Casein had a significant direct effect on protein percent. ß-Casein had significant direct effects on fat and protein percent. Linked QTLs with effects on fat percent were found for all three loci. Since the ß-casein and κ-casein genes are linked, it is likely that the same QTL was detected for those two markers.

The aim of chapter 7 was to quantify the potential effects of selection for κ-casein and ß-lactoglobulin genotypes by using stochastic simulation of a closed adult MOET nucleus breeding scheme. The results show that selection for κ-casein or ß-lactoglobulin genotypes can increase selection response in a MOET nucleus breeding scheme. For κ-casein, the total gain was 9 to 22 units of INET depending on whether κ-casein genotype effects needed to be estimated from the simulated data or κ-casein genotype effects were known without error. When selection was for κ-casein genotypes, the annual genetic progress was increased

by 2.4 to 4.8% in the first seven generations. The total additional response obtained by including ß-lactoglobulin genotypes was 16 units of INET. In the first seven generations the annual genetic progress was increased by 3.9%. For ß-lactoglobulin, genotype effects were not estimated and were assumed to be known without error. If estimates of ß-lactoglobulin genotype effects are inaccurate, the additional genetic response is expected to be lower. Additional gain for both milk protein genes was primarily due to an increased selection intensity on the male side. For a progeny testing scheme, it would be expected that increased genetic gain could be obtained if including milk protein genotypes results in an increased selection intensity.

Main conclusions

- Dutch Friesian and Holstein Friesian breeds differ for ß-casein and ß-lactoglobulin gene frequencies.
- Compared to the Black and White breeds the Red and Whites have a higher κ-casein B gene frequency.
- The κ-casein gene or (a) very closely linked gene(s) affects protein percentage.
- The ß-lactoglobulin gene or (a) very closely linked gene(s) affects fat percentage.
- Linked QTL with effects on fat percent were found for κ-casein, ß-casein and ß-lactoglobulin.
- Selection for κ-casein genotypes has the potential to increase cumulative selection response 9 to 22 units of INET.
- Selection for ß-lactoglobulin genotypes has the potential to increase cumulative selection response 16 units of INET.

SAMENVATTING

SAMENVATTING

De produktie van kaas is van groot belang voor de Nederlandse zuivelindustrie. Het is bekend dat melkeiwitgenotypen effect hebben op de kaasverwerkingseigenschappen van melk. Deze informatie in ogenschouw nemende ligt het voor de hand om de vraag te stellen of naast de op dit moment aangelegde selectiecriteria, aanvullend geselecteerd zou moeten worden op melkeiwitgenotypen. Hiermee zou de kwaliteit van de melk met betrekking tot de kaasproduktie verbeterd kunnen worden. Naast het reeds genoemde effect van melkeiwitgenotypen op kaasverwerkingseigenschappen wordt de economische waarde van melkeiwitgenotypen ook bepaald door de effecten op melkproduktiekenmerken. Een andere toepassing van melkeiwitgenotypen ligt in het lokaliseren en analyseren van genen die een effect hebben op melkproduktiekenmerken.

Hoofdstuk 2 geeft de lezer een beknopt overzicht van de huidige kennis over melkeiwitten en genetische varianten van melkeiwitten.

In hoofdstuk 3 is een snelle methode beschreven voor de typering van melkeiwitvarianten. De methode berust op de scheiding van melkeiwitvarianten door middel van isoelectrische focusering gebruik makend van het PhastSystem. De methode is vanwege de korte scheidingstijd en de hoge capaciteit geschikt voor de typering van een groot aantal monsters.

In hoofdstuk 4 is een maximum likelihood methode beschreven waarmee schattingen verkregen kunnen worden van de fractie fout getypeerde dieren. Op grond van gegevens van een kruisingspopulatie geeft de methode tevens schattingen van raseffecten op melkeiwitgenfrequenties. De resultaten van een simulatiestudie wijzen uit dat de methode schattingen van genfrequenties geeft die goed overeenkomen met de werkelijke waarden. De genfrequentieschattingen in de Nederlandse zwartbonte en roodbonte kruisingspopulaties zijn gebaseerd op respectievelijk 10151 en 580 dieren. Het Fries Hollandse en Holstein Friesian ras verschillen wat betreft genfrequenties voor ß-caseïne en ß-lactoglobuline. De schattingen voor de fracties fout getypeerde dieren waren nul voor α_{s1}-caseïne, 0.09 voor ß-caseïne en ß-lactoglobuline en 0.12 voor κ-caseïne. De verschillen tussen het MRIJ en Red-Holstein Friesian ras waren klein. De schattingen voor de fracties fout getypeerde dieren waren groot. De benaderde schattingsfouten voor deze fracties waren echter ook groot. De frequentie van κ-caseïne B is hoger voor de roodbonten dan voor de zwartbonten.

Het doel van de in hoofdstuk 5 beschreven studie was het schatten van de relaties tussen melkeiwitgenotypen en melkproduktiekenmerken. Hiervoor

werden gegevens van 6803 vaarzen gebruikt. Om in een populatie met gerelateerde dieren op correcte wijze de aanwezigheid van effecten van melkeiwitgenotypen te kunnen toetsen en om zuivere schattingen te verkrijgen van genotype effecten is er gebruik gemaakt van het diermodel. Effecten van melkeiwitgenotypen werden geschat gebruik makend van een model waarin elk melkeiwitgen afzonderlijk werd geanalyseerd (enkelvoudige gen analyse), en een model waarin alle melkeiwitgenen tegelijkertijd werden geanalyseerd (meervoudige gen analyse). De resultaten van beide modellen geven aan dat sommige effecten die in de enkelvoudige gen analyse toegeschreven werden aan bepaalde melkeiwitgenen, niet het effect van het melkeiwitgen zelf zijn maar van gekoppelde genen. Op grond van de resultaten van de huidige studie en van resultaten uit de literatuur werd geconcludeerd dat κ-caseïne, of een zeer nauw gekoppeld gen, effect heeft op het eiwitpercentage terwijl ß-lactoglobuline, of een zeer nauw gekoppeld gen, effect heeft op het vetpercentage. Verder zijn in het onderzoek significante effecten van ß-caseïne genotypen op de melkproduktie, het vetpercentage en de eiwitproduktie gevonden terwijl ß-lactoglobuline genotypen significante effecten hadden op de melk- en eiwitproduktie. Voor deze effecten is het minder duidelijk of ze de werking betreffen van het melkeiwitgen zelf danwel het gevolg zijn van een effect van een gen gekoppeld aan het melkeiwitgen.

In hoofdstuk 6 is de maximum likelihood methodologie gebruikt om tegelijkertijd effecten van een merker en een gekoppeld kwantitatief gen op kwantitatieve kenmerken te schatten. Er werd een kwantitatief gen met twee allelen verondersteld. Naast de effecten van de merkergenotypen op het gemiddelde van het kwantitatieve kenmerk werden het overkruisingspercentage tussen de beide loci, de allelfrequenties van het kwantitatieve gen en de standaarddeviatie binnen de merker en het kwantitatieve gen, geschat. Dit betekent dat naast de gemiddelden van de merkergenotypen nog zes parameters werden geschat. De statistische methode werd getest op gesimuleerde gegevens en gebruikt om directe en gekoppelde effecten van de ß-lactoglobuline, κ-caseïne en ß-caseïne melkeiwitgenen op de melk-, vet- en eiwitproduktie en het vet- en eiwitpercentage te schatten in de Nederlandse melkveepopulatie. De simulatiestudie gaf aan dat het model schattingen geeft die overeenkomen met de gesimuleerde waarden. Wel werden hoge empirische schattingsfouten waargenomen voor het overkruissingspercentage en voor de effecten van het kwantitatieve gen. In de praktijkdata werd een significant direct effect van ß-lactoglobuline op het vetpercentage gevonden. κ-Caseïne had een significant direct effect op het eiwitpercentage. ß-Caseïne had siginificante directe effecten

op de percentages vet en eiwit. Effecten van gekoppelde kwantitatieve genen op het vetpercentage werden gevonden voor ß-caseïne, κ-caseïne en ß-lactoglobuline. Omdat ook ß-caseïne en κ-caseïne gekoppeld zijn is het waarschijnlijk dat voor deze beide merkers hetzelfde gekoppelde kwantitatieve gen is gevonden.

Het doel van het onderzoek beschreven in hoofdstuk 7 was het kwantificeren van de mogelijke effecten van selectie op κ-caseïne en ß-lactoglobuline genotypen. Hiertoe is gebruik gemaakt van de stochastische simulatie van een gesloten kernfokprogramma met benutting van multiple ovulatie en embryo transplantatie. Dit fokprogramma zal verder aangeduid worden als een MOET fokprogramma. De resultaten wijzen uit dat selectie op κ-caseïne of ß-lactoglobuline genotypen de selectierespons in een MOET fokprogramma kan verhogen. Voor κ-caseïne was de extra cumulatieve respons 9 tot 22 eenheden INET, afhankelijk van de vraag of κ-caseïne genotype effecten geschat werden in de gesimuleerde gegevens of dat κ-caseïne genotype effecten zonder fout bekend waren. In de eerste zeven generaties was de jaarlijkse genetische vooruitgang 2.4 tot 4.8% hoger wanneer aanvullend geselecteerd werd op κ-caseïne genotypen. De extra cumulatieve respons die verkregen werd door ß-lactoglobuline genotypen bij de selectie te betrekken was 16 eenheden INET. In de eerste zeven generaties was de jaarlijkse genetische vooruitgang 3.9% hoger. Voor ß-lactoglobuline werden genotype effecten niet geschat maar bekend verondersteld. Wanneer de schattingen van ß-lactoglobuline effecten onnauwkeurig zijn dan is de verwachting dat de additionele selectie respons lager zal uitvallen. De extra winst die behaald werd door melkeiwitgenotypen te betrekken in de selectie kon voor het belangrijkste deel worden verklaard uit een verhoogde selectieintensiteit in de stieren. Voor het proef-wacht-fokstieren systeem is de verwachting dat het betrekken van melkeiwitgenotypen in de selectie ook daar kan zorgen voor een verhoogde genetische vooruitgang wanneer selectie op melkeiwitgenotypen resulteert in een hogere selectieintensiteit.

Belangrijkste conclusies

- Het Fries-Hollandse en het Holstein-Friesian ras verschillen wat betreft de ß-caseïne en ß-lactoglobuline genfrequenties.
- In vergelijking tot zwartbonten, is de κ-caseïne B frequentie hoger in roodbonten.
- Het κ-caseïne gen, of een erg nauw gekoppeld gen(en), beïnvloedt het eiwitpercentage.
- Het ß-lactoglobuline gen, of een erg nauw gekoppeld gen(en), beïnvloedt het vetpercentage.
- Gekoppelde kwantitatieve genen met effecten op het vetpercentage zijn gevonden voor κ-caseïne, ß-caseïne en ß-lactoglobuline.
- Selectie op de B-variant van κ-caseïne kan in totaal 9-22 gulden INET extra selectie respons opleveren.
- Selectie op de B-variant van ß-lactoglobuline kan in totaal 16 gulden INET extra selectie respons opleveren.

Curriculum vitae

Hendrik Bovenhuis werd op 22 december 1963 geboren in Rouveen (gemeente Staphorst). Na de plaatselijke gereformeerde lagere school te hebben doorlopen bezocht hij de Prof. dr. S. Greydanus scholengemeenschap te Zwolle. In 1982 werd hier het VWO diploma behaald. In datzelfde jaar werd begonnen met de studie Zoötechniek aan de toenmalige Landbouwhogeschool. Henk interesseerde zich met name voor de veevoeding en de veefokkerij. In 1987 werd aan de Landbouwuniversiteit de studie Zoötechniek met als oriëntatie veefokkerij, met lof afgesloten. Begin 1988 begon hij als AIO (assistent in opleiding) met een promotieonderzoek waarvan het thans voor u liggende proefschrift het resultaat is. Sinds 15 mei 1992 is hij aangesteld als tijdelijk medewerker bij de vakgroep veefokkerij.