Genetic and genomic studies on milk production and composition, and longevity in New Zealand dairy goats

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Animal Science

at Massey University, Manawatu,

New Zealand

Megan Rachel Scholtens

2020

Abstract

The New Zealand dairy goat industry is important for producing and exporting high-quality specialised dairy products aimed at niche markets. Efforts to increase the quantity and composition of goat milk will improve profits for farmers and deliver significant economic benefits to New Zealand. However, no formal program exists for the genetic improvement of dairy goats. Therefore, the general aim of this thesis was to perform genetic and genomic studies that contribute to the design of the breeding program for New Zealand dairy goats. The first studies estimated variance components and genetic parameters of total lactation yields of milk, fat and protein, somatic cell score and longevity. The main findings suggest sufficient variation and favourable genetic correlations between these traits, supporting their inclusion into a selection index that predicts profit per animal. A random regression test-day model was then used to predict lactation curves of milk, fat, protein and somatic cell score. Using this model for genetic evaluation will enable the dairy goat industry to move from total yields into the prediction of lactation curves, enabling more accurate predictions and the opportunity of selecting for extended lactations. The first genome-wide association study of dairy goats in New Zealand was conducted using 3,732 animals genotyped with the Caprine 50K SNP chip. A highly significant region on chromosome 19 was associated with yields of milk, fat and protein, and somatic cell score, and a region on chromosome 29 was associated with somatic cell score. A prototype single-step BayesC model was developed to predict genomic breeding values and demonstrated that including genomic information into the evaluation can increase the accuracy of predictions compared to the traditional methods based on pedigrees alone, which is currently implemented in the New Zealand dairy goat industry. This thesis demonstrates that a single-step prediction model that uses genomic information would put the New Zealand dairy goat industry in a very good position to implement a genomic selection scheme. Further studies are required to define clearer breeding objectives and to systematically design a breeding program for the genetic improvement of New Zealand dairy goats.

Acknowledgements

If you had asked the 21-year old me - an undergraduate with below average grades and absolutely no plans on sticking around - the typical “where do you see yourself in ten years?” question, I’m not sure what I would’ve said but it definitely wouldn’t have been “still in Palmy and completing a PhD thesis”. However, during my last year of my Bachelors, I was lucky to have the opportunity to undertake a Master’s with Professor Nicolas Lopez-Villalobos and Dr Sam Peterson, completing a prototype genetic evaluation of a dairy sheep flock. This sparked a passion for applying my effort to something that made a difference in the real-world, I decided I wanted to be a geneticist.

To my fantastic supervisors: Professor Nicolas Lopez-Villalobos, Professor Dorian Garrick and Professor Hugh Blair, I thank you for your unwavering support and encouragement throughout my studies at Massey University, I thoroughly enjoyed every meeting together and I absolutely wouldn’t have got this far without you.

To my chief supervisor, Professor Lopez-Villalobos, thank you for being a wonderful supervisor and mentor, for believing in my potential and for always having time to sit down and discuss any issues. Professor Dorian Garrick, thank you for your guidance, constructive criticism and patience, all have helped to build my understanding of the complex aspects of animal breeding. Professor Hugh Blair, thank you for your encouragement and for your quick feedback continuous advice. I cannot express, fully, the gratitude towards these supervisors, for every discussion, whether it be academic or general chat, was an absolute pleasure. You have all had a hand in shaping the scientist I have become.

I would also like to acknowledge the friendly environment provided by all personnel of IVABS and SAE, especially Miss Debbie Hill for your support during my time in IVABS, and Sharon Wright and Tara Penketh for your support during my time in SAE.

A few other people that I would like to thank are Professor Russell Snell and Dr Klaus Lehnert for their support and guidance during the genomic aspects of this thesis. My demonstrating team, Ying Jin, Seer Ikurior, Sandeep Karna and Komkiew Pinpimai for making Fridays an

enjoyable break from our PhDs. Also, to Dr Sam Peterson, to whom I must thank for my writing and presentation skills as well as his ongoing support.

To my previous office mates, Adrienne French, Dr Bex Lucas, Kat Littlewood and Hilary Webb, the first two years with an office full of witty and knowledgeable ladies was the best start to this journey and I am grateful for the time we had together. To my current officemates and especially Pavithra Ariyarathne, I am glad we got to share the ups and downs together and I wish you the best of luck in completing your thesis.

Finally, this has not been a journey I have taken alone. A big thanks to all my old and new friends for enjoying the journey with me. It has been great to have your support and encouragement to keep me going. I would especially like to thank my girls Cindy Inverarity and Seulyi So. Cindy Inverarity, thank you for being a great friend, your support and advice was instrumental in keeping me sane during the tough times. Seulyi So, I am so glad you stayed in Palmy, I couldn't have gotten through this without the weekly support. I'm sure all the cheese and wine really did help reduce the stress levels!

To Mum, thank you for always being there. It wasn't always an easy journey, but it was great knowing that I could call at any time and come home to visit, provided I booked in first! I would also like to thank my family, co-family and flat-families, your endless support means the world to me and I couldn't have done it without you all!

Funding acknowledgements

This thesis was part of a project funded by the Ministry of Business, Innovation and Employment (MBIE) grant number UOAX1502. This project titled "Accelerating sustainable productivity gains for high value export: Dairy Goat infant formula" was cofounded by MBIE and the Dairy Goat Cooperative (NZ) Ltd (Hamilton, New Zealand) and was a joint venture between the University of Auckland (Auckland, New Zealand) and Massey University (Palmerston North, New Zealand). Megan Scholtens was supported by a MBIE Research Studentship.

The financial support of the MU Doctoral Scholarship, Massey University Animal Science Postgraduate Fund and the New Zealand Society of Animal Production are acknowledged.

Contents

List of tables

List of figures

List of abbreviations

AI	Artificial insemination
AL	Actual longevity
BLUP	Best linear unbiased prediction
BV	Breeding value
dBayesC	GWAS model fitting diplotype and SNPs simultaneously using BayesC
DGC	Dairy Goat Cooperative (NZ) Ltd
DIM	Days in milk
EBV	Estimated breeding value
ECMY	Energy-corrected milk yield
FL	Functional longevity
FY	Fat yield
GBLUP	Genomic best linear unbiased prediction
GBV	Genomic estimated breeding value
GP	Genomic prediction
GWAS	Genome-wide association study
h^2	Heritability
hBayesC	GWAS model fitting haplotype and SNPs simultaneously using BayesC
LD	Linkage disequilibrium
MCMC	Markov chain Monte Carlo
MME	Mixed-model equations
MY	Milk yield
PBLUP	Pedigree-based best linear unbiased prediction
PY	Protein yield
QTL	Quantitative trait loci
RR-BLUP	Ridge-regression best linear unbiased prediction
RRM	Random regression models
sBayesC	GWAS model fitting all SNPs simultaneously using BayesC
SCS	Somatic cell score
sGWAS	Single-SNP genome-wide association study
SNP	Single nucleotide polymorphism
ssBC	Single-step BayesC
ssBR	Single-step Bayesian regression
ssGBLUP	Single-step genomic best linear unbiased prediction

Chapter 1
General Introduction

New Zealand has a small, well established goat milk industry that produces high value dairy products for export. Goat milk is becoming an increasingly common alternative for people with intolerances or allergies to cow milk (Bevilacqua et al., 2001; Lara-Villoslada et al., 2006). Compared to cow milk, goat milk is more readily digestible, more alkaline, lower in lactose, casein and protein, and with faster protein digestion, it is suitable for people suffering from eczema, asthma and stomach ulcers (Jandal, 1996; Haenlein, 2004). Unlike cow milk that is consumed as liquid milk, goat milk is used in the production of niche dairy foods and sold as high-quality dairy products (e.g. cheese and infant formula). Worldwide production of goat milk has more than doubled in the last 50 years and, if this trend is maintained, it is expected to increase by approximately 9.7 Mt (+53%) by 2030 (Pulina et al., 2018). This growing interest for non-bovine milk provides an opportunity for New Zealand to expand its goat milk sector and to continue producing high value exports.

The New Zealand dairy goat population is estimated at 66,100 dairy goats distributed in 92 farms (Scholtens et al., 2017). Most of the dairy goat farmers are organised as the Dairy Goat Cooperative (NZ) Ltd (DGC) which is located in the Waikato region. The DGC is the leading international manufacturer of goat milk based nutritional powders for infants and young children (Stafford and Prosser, 2016) and processes 80% of the milk from the New Zealand dairy goat population.

Although genetic evaluation of milk traits has been implemented since 1997 in New Zealand (Singireddy, 1997), no formal breeding program exists, and as a result, the national genetic improvement of dairy goats is stagnant. However, it is vital for the industry to implement a structured breeding program in order to increase the quantity and composition of goat milk produced in New Zealand. Improving dairy goat genes is achieved by having a well-designed and implemented breeding program which enables an increase in the average genetic merit of each successive generation of replacement does. Therefore, a co-ordinated breeding program is required to ensure farmers have access to animals of superior genetic merit for a defined breeding objective, such as profit per animal. Traditional genetic evaluation uses performance and pedigree records to estimate the genetic merit (breeding values) of individuals for traits of interest. These breeding values are the estimation of the sum of the

additive genetic effects for an individual that affect the trait of interest (Falconer and Mackay, 1996). Breeding values reflect the potential of an animal as a parent and are, therefore, widely used to rank animals and select candidates as the parents of the next generation. The most widely used method of genetic evaluation is the best linear unbiased prediction procedure (Henderson, 1975), which uses phenotypic and pedigree information to produce estimated breeding values (EBVs).

In order to remain competitive on the global goat milk stage, the DGC wants to improve genetic progress within the New Zealand dairy goat industry. This requires a breeding program that comprises a number of steps (Harris et al., 1984): 1) definition of the breeding goal (e.g. profit per doe), 2) definition of the breeding objective in which animal traits related to the breeding goal are defined and their economic values are estimated, 3) definition of a selection criteria, which generally is a selection index combining EBVs for traits defined in the breeding objective and other genetically related traits with their relative economic weights, 4) definition of a selection scheme in which superior animals for the breeding objective are identified based on the selection index, 5) definition of a dissemination system in which genes from superior animals are spread into the population, and 6) perform an economic analysis to evaluate the industry profitability and the cost of running the breeding program.

Currently, the genetic evaluation of New Zealand dairy goats produces EBVs for total lactation yields of milk, fat and protein, and average somatic cell score (Singireddy, 1997; Apodaca-Sarabia et al., 2009). This relies on a two-step process based on a first step of combining test-day records to phenotypically predict total lactation yields. In Chapter 3 a multi-trait repeatability model was used to estimate genetic parameters for these total lactation yields of milk, fat and protein and somatic cell score. Using a multi-trait repeatability model enables the estimation of genetic and phenotypic correlations between each of the traits which is essential for developing a selection index. Chapter 4 explored the potential of implementing a random regression test-day model that would provide more accurate estimates for each individual and selection programs could be devised to exploit the genetic variation throughout the lactation period.

Placing too much emphasis on production, whilst neglecting other traits may result in undesirable consequences on the health and fertility of animals, which could decrease longevity (Oltenacu and Broom, 2010). Longevity is an important trait for increasing the overall economic efficiency of a dairy goat farm and should be considered in the current genetic evaluation system. In Chapter 5 the heritability of longevity was estimated to explore the possibility of including this trait into the evaluation and subsequent selection index.

In Chapter 6 genome-wide association studies (GWAS) were performed to identify genetic markers associated with quantitative trait loci (QTL) that underlie the phenotypic expression of specific traits (Goddard and Hayes, 2012). The possibility of applying a single-step genomic evaluation of dairy goats in New Zealand was investigated in Chapter 7. Genomic prediction is a method that predicts genomic breeding values (GBVs) using information from genetic markers located across the entire genome in an attempt to capture all QTL influencing the variation in a trait (Hayes et al., 2009). Previously, genomic prediction was limited to genotyped animals, however, the introduction of a single-step process (Legarra et al., 2009) enables the prediction of GBVs for all animals in the population. Implementing single-step genomic evaluation in a breeding program for the New Zealand dairy goat industry would provide an opportunity to rapidly increase the quantity and composition of goat milk produced in the New Zealand.

The main aim of this thesis was to do genetic and genomic studies of economically important traits and explore the possibility of applying a single-step genomic evaluation of dairy goats in New Zealand. The focus of this work was to contribute in the design of the breeding program that will ensure the dairy goat industry increases the quantity and composition of goat milk delivering significant economic benefits to New Zealand. To achieve this, the main objectives of this thesis were to:

- estimate genetic parameters and variance components of total lactation yields of milk, fat and protein and somatic cell score of New Zealand dairy goats.
- estimate genetic parameters of daily yields of milk, fat and protein and somatic cell score throughout the lactation of New Zealand dairy goats.

- estimate heritability of longevity of New Zealand dairy goats.
- identify and evaluate any significant regions on the goat genome that influence yields of milk, fat and protein and somatic cell score in New Zealand dairy goats, and
- design a prototype prediction equation to estimate genomic breeding values of New Zealand dairy goats and to quantify the advantages of including genomic information compared to pedigree-based breeding values.

This thesis will advance the knowledge necessary for the design of a breeding program using genomic selection for the New Zealand dairy goat industry.

Chapter 2
Review of literature

2.1 Dairy goats worldwide

Originating from a handful of ancestral wild goat breeds in the Middle East, goats have descended and evolved into hundreds of different breeds around the world (Haenlein, 2001) with the current population estimated to be about 203 million goats (FAO, 2018). The popularity of goats and their products can be attributed to their ability to survive and reproduce under harsh environments and perform well under restricted nutrition (Escareño et al., 2013). In addition to their resilience, goats are advantageous over other production ruminants for their early maturity, shorter gestation period, higher prolificacy, longer lactation (up to 300 days of milking compared to 250 for sheep) and the ability to adapt to a broad range of environments.

The dairy goat industry is subject to competition from cattle, sheep, buffalo and camel milk products (Dubeuf et al., 2004). Cattle (83.1%), and buffalos (13.1%) are the most important milk producers in terms of world production, while goat milk represents only 1.9% (FAO, 2018). There is a growing demand for non-bovine milks. Worldwide goat milk production has more than doubled during the last 50 years and, if this trend is maintained, it is expected to increase by approximately 9.7 Mt (+53%) by 2030 (Figure 2.1.) (Pulina et al., 2018). Dairy goat products are primarily made for dietetic milk or cheese markets, but the profitability and competitive advantage of these products depends on their relative price and production systems (seasonality, herd size, goat productivity and milk characteristics) (Dubeuf et al., 2004). Historically, goats were farmed for home-consumption or sold within villages (Dubeuf et al., 2004). This is still the case in developing countries (predominantly in Asia and Africa), however, this has changed for a number of countries in Latin America and Europe (especially France, Italy, Spain and Greece), with the development of a specific dairy goat sector where goat milk is sold or transformed into cheese or candies (Escareño et al., 2013). The establishment of national professional organizations, technical centres, breeding and selection organizations, along with the steady growth of export markets for goat cheese has resulted in 90% of goat milk produced in France being sold as cheese (Dubeuf et al., 2004).

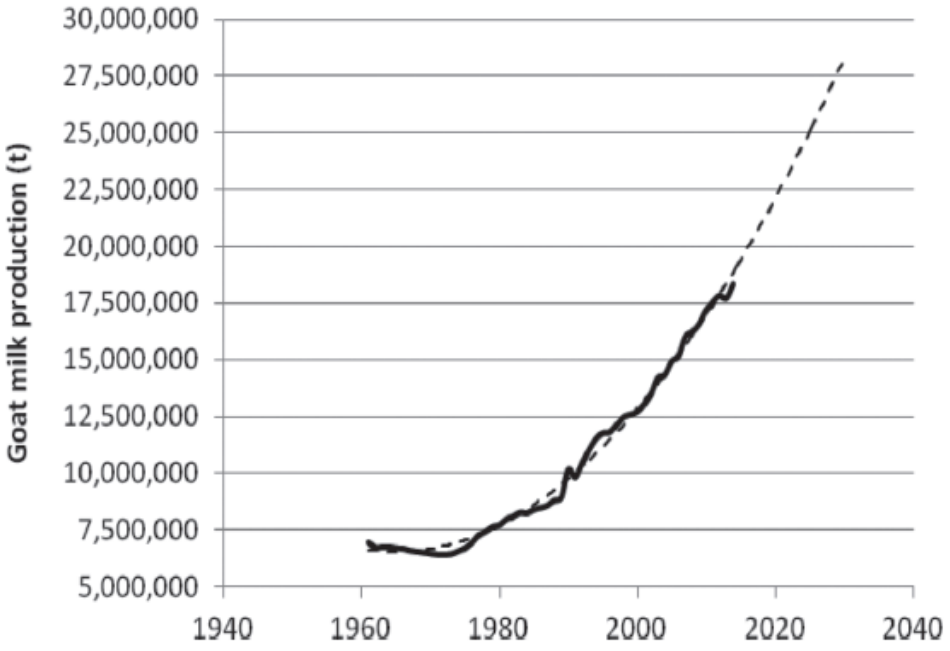

Figure 2.1. World goat milk production trends from 1960 to 2016 (solid line) and forecast to 2030 by using time-series model (dashed line) (adapted from Pulina et al., 2018).

In 2018 the worldwide production of goat milk was 18 million t (Table 2.1). Approximately 52.7% of this was produced in Asia, with remarkable amounts in the Indian subcontinent (i.e. Bangladesh, India and Pakistan), followed by Africa (25.7%), where Sudan and South Sudan are the largest contributors (FAO, 2018). Europe and Americas contribute 16.6% and 4.9%, respectively, to world goat milk production, while Oceania makes a negligible contribution to global production (<0.01%). The average milk yield of dairy goats worldwide ranges dramatically from 35 to 291 L/doe, with some countries such as Ukraine and France reaching average yields of 500 L/doe and 715 L/doe, respectively. In 2018, Asia contributed the greatest number of dairy goats and milk production of all continents, but produces on average 77 L/milk/doe/year, whereas Africa had the second largest dairy goat population and produces on average 54 L/milk/doe/year. This demonstrates how dairy goat productivity varies greatly between countries and continents. Goat farms in Asia, Africa and Latin America, are generally extensively managed grazing on communal land which tends to be overstocked, degraded and barely providing adequate nutrition (Escareño et al.,

2013). Some parts of the Middle East, such as Israel, have a well-organized intensive dairy goat sector with zero-grazing (Dubeuf et al., 2004), while in Vietnam goats are managed in either intensive (goats live in cages and are hand-fed), semi-intensive (goats graze during day and housed at night with supplements provided) or extensive systems (goats graze without supplementation).

America and Europe had similar population statistics contributing approximately 4% each to the total world goat population. However, in general, European countries are far more specialised in milk production than developing countries, contributing 16.6% compared to America which produced 4.9% of total goat milk produced worldwide in 2018. This could be attributed to the fact that in the last 20 years dairy goat operations in Europe (especially in countries such as France, Greece and Spain (de Rancourt et al., 2006)) have tried to eliminate seasonality and improve milk production by using breeds with high production potential and intensifying their goat systems (Castel et al., 2011). European goats are farmed under two types of systems: traditional (grazing in spring and autumn, housed indoors in winter with vertical transhumance in the summer) or intensive/semi-intensive (housed indoors under controlled feeding of hay and concentrates) (Nicoloso et al., 2015).

In America, dairy goats are milked to produce fluid or powdered milk (Dubeuf et al., 2004), while Canada and California have very active dairy goat sectors producing new goat cheeses, cosmetic products and candy from goat milk (Haenlein, 2000; Haenlein, 2001).

Brazil has a combination of intensive and semi-intensive goat operations. Southeast Brazil has predominantly intensive production systems where goats are confined and fed concentrates, while in Northeast Brazil, goats graze on native forests during the rainy season, then confined indoors for the remainder of the year (Lôbo et al., 2017). The dairy sector in other parts of the world is less organised or part of dual purpose (milk and meat) management, with the majority of milk being sold locally, or consumed at home (Dubeuf et al., 2004).

Table 2.1. Global dairy goat populations and average annual production per goat in 2018 (FAO, 2018).

Continent	Milking goats		Milk production		Milk production per goat
	Number (Million)	Contribution (%)	(Mt/year)	Contribution (%)	(L/year)
Africa	80	39.6	3.93	25.7	54
Americas	8	4.0	0.75	4.9	66
Asia	106	52.1	8.04	52.7	77
Europe	9	4.3	2.54	16.6	291
Oceania	<0.01	<0.01	<0.01	0	35
World	203	100	15.26	100	81

2.2 Dairy goats in New Zealand

Goat farming in New Zealand offers a reasonable economic option for low impact agricultural diversification when needed to meet environmental compliance conditions. The local dairy goat industry shows strength in its high-value products and has secure domestic and international markets. The dairy goat population is estimated at 66,100 goats distributed in 92 farms (Scholtens et al., 2017), however, the exact population is not known as there is no census undertaken for dairy goats in New Zealand (Stafford and Prosser, 2016). The New Zealand herd is predominantly of the Saanen breed (85%), but also including British Alpine, Toggenburg and Anglo-Nubian breeds. Most dairy goats are intensively managed, with approximately 72% of the dairy goat population located in the Waikato region (Table 2.2) with the remaining 28% being distributed throughout the rest of New Zealand (Orr et al., 2010). Dairy farms which supply goat milk processing plants have an average herd size of 750 milking does, while the smaller farms, which make their own cheese or supply local cheese makers, tend to have approximately 50 goats.

Table 2.2. Regional distribution of dairy goats in New Zealand in 2016.

Region	Herds (Number)	Goats (Number)	Proportion of population (%)
Auckland	3	3,732	6.0
Bay of Plenty	1	9	0.01
Hawkes Bay	1	1,000	1.5
Manawatu	5	6,000	9.1
Nelson	1	60	0.1
Northland	4	2,555	3.9
Otago	1	30	0.1
Taranaki	8	5,154	7.8
Waikato	67	47,485	71.8
Wellington	1	75	0.1
Total	92	66,100	100.0

Two types of farming systems are practiced in the New Zealand goat industry. Goats are housed indoors with a cut-and-carry feeding system, or in an outdoor system where they live and graze in paddocks (Robertson et al., 2015). The bulk of dairy goat farms are managed intensively (Morris et al., 1997), housed in open-sided barns and their food is brought to them two to three times a day. The goats are fed fresh pasture or crops, which are grown and harvested on-farm and cut and carried to the side of the barn (Solis-Ramirez et al., 2011). In an outdoor system, the goats live and graze in paddocks with supplements provided, if needed. Outdoor systems have animal health challenges due to internal parasitism that can be avoided in cut and carry systems.

Lactation length for goats farmed in indoor systems range from 190-324 days in milk (six-eleven months) (Robertson et al., 2015; Stafford and Prosser, 2016), with daily milk production averaging 2.7 L/doe/day with 3.5 L/doe/day at peak lactation (Stafford and Prosser, 2016). The industry average has been 625 milking does per farm and 86 kg milk solids/doe/year (Robertson et al., 2015; Stafford and Prosser, 2016). Yearling does produce

about 75 kg milk solids/doe/year while two and three-year old does can average 100 kg milk solids/doe/year. These averages are based on the greater proportion of indoor farms, which tend to have increased numbers and greater production per animal (Robertson et al., 2015).

Dairy goat farms which supply the Dairy Goat Cooperative (NZ) Ltd (DGC), on average, undertake herd-testing four times during the season. These herd-test records for daily milk yield (litres), concentrations of fat, protein and lactose, and somatic cell count are managed and stored in the Livestock Improvement Corporation database, together with individual animal information.

Definition of a breeding objective and selection indexes with the estimation of economic values for milk, fat, protein, lactose and somatic cell score (SCS) and longevity were proposed by Solis-Ramirez et al. (2014). The first genetic evaluation of New Zealand dairy goats was performed in 1997 for the estimation of breeding values (EBVs) for lactation yields for milk, fat and protein (Singireddy, 1997) using a univariate repeatability animal model. Currently, EBVs for lactation yields of milk, fat and protein and average SCS during lactation, are obtained using a multivariate repeatability animal model (Lopez-Villalobos, personal communication). Somatic cell score is calculated as SCS=Log2(somatic cell counts/1000). An economic breeding index combines EBVs for protein, fat and SCS with economic values for these traits.

A test-day model was proposed for the estimation of breeding values for somatic cell count in 2009 (Apodaca-Sarabia et al., 2009). In addition to the national evaluation for farmers supplying DGC, within herd evaluations have been published by Morris et al. (2006) for milk, fat and protein and Wheeler et al. (2013) for stayability.

Despite the efforts of proposing a selection index and implementing genetic evaluation, a well-structured selection scheme does not exist. The majority of farmers select bucks and does as parents for the new generation, based on female phenotypic records. There is limited use of artificial insemination (AI) to promote the use of superior bucks across herds. Nevertheless, DGC is progressing towards the use of AI and the possibility of a sire referencing scheme, but the exchange of animals is currently limited due to the risk of

spreading Caprine arthritis encephalitis that is present in some herds. In 2016, a team of bucks were selected from herds diagnosed as disease-free based on dam production and the economic breeding index of the herds where they were born. Bucks were genotyped for the α_{S1}-casein polymorphism and were selected if they had the FF genotypes which are associated with low levels of α_{S1}-casein concentration (Huitema, 2012). However, results from the animal evaluation have not been implemented as part of a broader selection scheme. There are no published reports on phenotypic or genetic trends.

2.3 Systematic design of a breeding program for dairy goats

Animal breeding is a tool involving the knowledge of genetic, phenotypic, economic and farm management factors to select the most suitable animals for the production system (Harris et al., 1984). Genetic improvement is the result of selecting genetically superior animals to be the parents of the next generation (Garrick and Fernando, 2014). In practice, this involves many challenges but can be summarized into seven steps as shown by Lopez-Villalobos and Garrick (2005) in dairy cattle which follows the systematic approach to the design of animal breeding programs as proposed by Harris et al. (1984).

Step 1 – Breeding goal

The breeding goal states the desired direction of improvement from the breeding program (Groen, 2000). Common breeding goals defined in the agricultural industry include, profit per animal, profit per hectare or, in regards to efficiency, profit per dry matter consumed (Lopez-Villalobos and Garrick, 2005). The breeding goal can vary for each species, breed, system and country. For example, the breeding goal of dairy goats in Norway is to increase milk solids produced per goat and year (Dagnachew et al., 2011). Therefore, this industry is focused on improving the quantity and quality of milk produced per doe. Likewise, in Europe, the majority of goat milk is used for cheese, but it is also commonly consumed as whole milk and yoghurt (Pulina et al., 2018), therefore the goal is more focused on protein and fat content rather than increasing milk yield (Tabbaa and Al-Atiyat, 2009). In

comparison, in Brazil, goat milk is sold as whole milk, therefore milk volume is more important than milk composition (Lopes et al., 2012). Once the breeding goal is defined, the next step is to decide on the animal traits that influence the goal.

Step 2 – Breeding objective

The breeding objective is a mathematical equation representing important traits which affect overall farm profit and aims to achieve the breeding goal (Newman et al., 1992; Charfeddine, 2000). The breeding objective can be described in two steps (Harris et al., 1984). First, the animal traits which influence the breeding goal are identified and second, the relative weight of each trait is quantified (Lopez-Villalobos and Garrick, 2005).

Step 3 – Selection criteria

Once the breeding goal and important traits have been defined the next step is to define the selection criteria. Selection criteria are the traits that can be measured in the animals at a young age and are genetically correlated with traits in the breeding objective (Lush, 1937). Once traits are identified and economic weights established, selection index theory (Hazel, 1943) can be used to derive a selection index, which predicts the breeding goal as accurately as possible. A selection index is a mathematical formula that amalgamates adjusted phenotypes or estimated breeding values (EBVs) for several traits and incorporates a relative economic weight. The selection index is a predictor of the aggregate economic value of an animal. The formulation was first described by Hazel and Lush (1942) and in their context requires knowledge of population parameters, namely heritabilities and phenotypic and genetic correlations. Estimates of these genetic parameters should be specific for the population under consideration (van der Werf and de Boer, 1989). This index allows the ranking of animals in the population, so that the best animals can be selected for replacements to achieve the breeding goal.

Genetic improvement programs for dairy goats typically began with an evaluation focused on improving milk yield, before expanding to an index that include yields of fat and protein,

and type traits. This has been the case for most selection indexes worldwide. Brazilian selection focuses on milk yield (Lopes et al., 2013), while the United States of America have progressed to select for milk, fat and protein yield (Analla et al., 1996; Wiggans and Hubbard, 2001), and Norway and France have progressed further by including health, reproduction, conformation and production traits (Aziz, 2010; Lopes et al., 2013). A good example of a selection index is the Index Combine Caprine which has been implemented in the French Alpine and Saanen breeds focused on yields of milk, fat and protein, percentages of fat and protein and udder traits (shape, attachment and teat placement) (Aziz, 2010).

Step 4 – Selection schemes

The design of a selection scheme includes deciding which and how many animals will be selected as parents for the next generation. Rendel and Robertson (1950) provided a general framework to systematically design selection schemes based on four pathways of selection. These pathways are selection of dams to breed female replacements, selection of dams to breed male replacements, selection of sires to breed female replacements and selection of sires to breed male replacements. They illustrated these four pathways of selection in the case of the design of a progeny test for dairy cattle (Robertson and Rendel, 1950). They applied the breeder equation (Equation 2.1) for each pathway to demonstrate the expected rate of genetic gain. The breeder's equation was proposed by Lush (1937) and can be represented in the following way:

$$\Delta G = \frac{\bar{\imath} r_{TI} \sigma_g}{L} \qquad (2.1)$$

where ΔG = genetic gain over time, $\bar{\imath}$ = standardised selection differential (selection intensity), r_{TI} = selection accuracy, σ_g = genetic standard deviation and L = generation interval. Increasing the rate of genetic gain can be achieved by changing any of the factors

of this equation, either increasing the selection intensity, accuracy or genetic variation, or decreasing the generation interval.

The dairy goat industry in France has applied the principles of four pathways of selection in a progeny test scheme since the 1980's (Clément et al., 2002). Each year the top 200 bucks from the selection base entered an individual performance test station. The bucks went into a 30-day period where conformation, growth, and sanitary conformation controls were carried out. Then, 120 males continued to be tested for sperm production and sexual behaviour. From this individual performance station, the top 80 bucks continued and were evaluated based on their on-farm progenies. The genetic evaluation of each buck was based on ~200 artificial insemination and performance records of 80 daughters, on average. Following progeny testing, 30 to 40 of the top bucks were retained each year as elite sires and used for AI (France Génétique Elevage, 2020).

The progeny test of a buck is completed when the buck is 4-6 years old (Carillier et al., 2013). This is an expensive and long process to estimate the breeding value of the bucks to allow the selection of the potential sires to produce male and female replacements. By delaying selection decisions more information accumulates, and accuracy of EBVs increases, but this leads to a long generation interval and a reduced rate of genetic gain (Meuwissen, 2003). Conversely, the information from the pedigree could be available immediately, which could assist with the selection of replacement does or bucks at an earlier age. However, the early selection would be based on fewer records, thus, reducing the accuracy of selection and consequently the rate of genetic gain.

The genomic information used in genomic selection schemes can improve the accuracy of selection as well as enabling identification of animals with desirable genotypes at a younger age, both increasing the rate of genetic gain. A genomic selection scheme requires genotypes of top sires which are included in the previous progeny testing scheme and their sons. Marker effects are estimated using the genotypes and progeny records. Then, all possible selection candidates to be a dam to breed male replacements are genotyped (the top dams) and genomic breeding values (GBVs) are estimated (Schaeffer, 2006). This genomic selection scheme can be used to rank young animals, allowing bucks to be selected

earlier in life and allowing earlier identification of replacement candidates (Meuwissen et al., 2013). In addition, the accuracy of the GBVs are sufficient for selection over several generations without repeated phenotyping, which reduces the cost and generation intervals (Habier et al., 2007).

Genomic selection has become a widely adopted method for ranking candidates for selection in dairy cattle (Hayes et al., 2009; VanRaden et al., 2009; Boichard et al., 2012; García-Ruiz et al., 2016), dairy sheep (Duchemin et al., 2012; Baloche et al., 2014) and dairy goats (Carillier et at., 2013; Carillier et al., 2014; Mucha et al., 2015).

Step 5 – Dissemination system

Designing an appropriate system for the transfer of genes from the genetically superior individuals into the commercial population is largely determined by the size of the commercial population, and the cost and efficacy of the biotechnologies available (e.g., AI, multiple ovulation embryo transfer, trans vaginal recovery and in vitro production, and sexed semen) (Harris et al., 1984; Lopez-Villalobos and Garrick, 2005). Most livestock industries have a structure that comprises at least two tiers – nucleus herds at the top and commercial production herds at the bottom (Figure 2.2). A nucleus herd consists of the animals in the population with the superior genes. These superior animals are generally owned by breeders or breeding companies and provide the next generation of sires to breed sires and sires to breed dams. These nucleus herds are recorded for a large number of traits and are the basis of genetic evaluations. From these herds, the superior genes are disseminated to the remaining population in the commercial herds, typically using AI. This is the case in the French dairy goat industry where the official milk recording and selective mating occurs in the nucleus herds. Then, the breeders associations manage the AI centers to raise the young bucks from weaning to the age at reproduction, and to organize their progeny test. The gene flow from the nucleus herds to the commercial herds is achieved through AI and natural mating males by sons of AI sires (Larroque et al., 2014). The relative size of the nucleus tier, its composition and the manner in which it interacts with the

commercial tier strongly influence the genetic improvement and overall profitability of the industry.

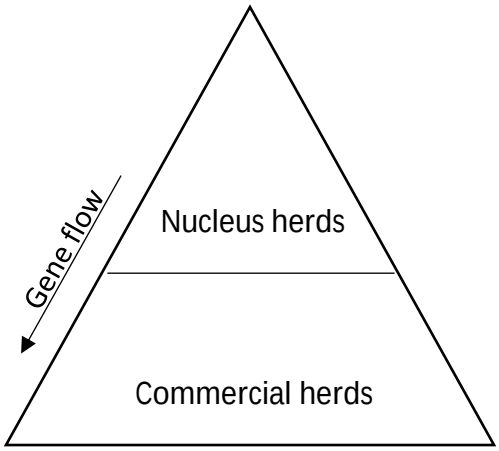

Figure 2.2. A pyramid livestock industry structure in which genetic improvements flow from the top to the base.

Step 6 – Mating plan

A mating plan is required to maximise the long-term genetic gains obtained in a breeding program. Designing a mating plan includes the decision of what mating strategy to use, whether it's crossbreeding, inbreeding, assortative mating or random mating (Harris et al., 1984). Within each of these strategies, decisions must be made on the mating ratio of female to male, and the number of breeding seasons the selected individuals will be used for (Lopez-Villalobos and Garrick, 2005).

Crossbreeding is a common mating plan implemented by farmers to exploit heterosis effects and has been successfully implemented in the New Zealand dairy cattle industry (Lopez-Villalobos and Garrick, 2005). Crossbreeding is also common practice in dairy goat industries where local breeds are crossed with specialised dairy breeds to "upgrade" local breeds to produce greater milk yields (Escareño et al., 2013).

Assortative mating is another common method for maximising long-term genetic gains through the use of AI. Artificial insemination is the best dissemination of superior genetics as it enables farmers to choose which bulls to breed to which cows which enables the chance of achieving genetic gains for specific traits within a herd. In addition, AI reduces the need for natural mating, thereby reducing the risk of spreading disease which is an issue in dairy goats. Admittedly, the use of AI will increase the selection intensity which can lead to increased rates of inbreeding (Granleese et al., 2015). However, genomic selection is expected to increase genetic gain of traits of interest, without increasing the level of inbreeding (Daetwyler et al., 2007; Dekkers et al., 2007).

Dairy goat breeding programs in Canada mainly use AI in the nucleus herds for assortative mating (Brito et al., 2013). While in France the gene flow from the nucleus to the commercial herds is based partly on AI males and partly on natural mating males by sons of AI sires (Carillier et al., 2013).

Step 7 – Economic analysis

The final stage of designing an effective genetic improvement program is the economic analysis of the proposed program. The economic analysis evaluates the effectiveness of the breeding program. This is a complex step which requires whole-system modelling including the cost of the evaluation system, selection scheme and dissemination system (Harris et al., 1984).

Designing an appropriate breeding program is essential to ensuring genetic improvement occurs in the right direction for the dairy goat industry. The structure and design of a breeding program is consistent worldwide however, the breeding goals and selection criteria will vary among different dairy goat systems. The genetic parameters and EBVs required for the genetic evaluation are population dependent and their derivation are explained in further detail.

2.4 Estimation of genetic parameters

Knowledge of genetic parameters is required for planning efficient animal breeding programs as the potential genetic improvement largely depends on the heritability of the traits as well as the genetic relationship (correlation) between the traits. Heritability, commonly referred to in quantitative genetics as narrow heritability, measures the proportion of variation of a trait which is due to the genetic variation between individuals in that population (Hayes and Goddard, 2014). Heritability provides valuable information as to whether the trait can be improved by selection or management practices, or both. Traits may also be contingent on one another, with either positive or negative correlations and genetic and phenotypic correlations that are either strong or weak. In some cases, indicator traits may be used to exploit such correlations, if they are more readily available than a trait of interest (Hazel, 1943). Repeatability is the correlation between multiple records on the same individual from the population (Hazel, 1943). This information can be used when constructing a selection index or to estimate the individual's productive ability for future records.

Estimation of breeding values requires knowledge of the variances and covariances of genetic effects. Genetic and environmental factors significantly influence milk yield and quality in ruminants (Selvaggi and Dario, 2015), and genetic gain from selection will be enhanced if these environmental factors are accounted for in the estimation of genetic merit. Knowledge of variance components for production traits will enable the design of an effective genetic evaluation strategy, allowing the selection of animals with superior overall genetic merit, optimizing direct and correlated selection responses for traits of economic importance (Barillet, 2007). Such procedures are routinely performed on dairy goats in South Africa (Muller, 2005), France (Boichard et al., 1989; Bélichon et al., 1998), Spain (Analla et al., 1996) and the United Kingdom (McLaren et al., 2016).

Many studies have published estimates of genetic parameters of milk production traits in dairy goat populations. Heritability and repeatability estimates for milk yield (MY), fat yield (FY), protein yield (PY) and somatic cell score (SCS) range from 0.10-0.45, 0.19-0.40, 0.04-0.38 and 0.09-0.25, respectively (summarised in Table 2.3). Similarly, genetic relationships

between these milk traits have also been published and are summarised in Table 2.4. The differing estimates of genetic parameters of milk production traits in goats can be due to the breed and population, structure of the data, management conditions, estimation errors associated with sample size, and the estimation methodology used (Moioli et al., 2007).

Table 2.3. Published heritability and repeatability values for milking traits in dairy goats.

Trait	Heritability	Average
Milk yield	0.10[a], 0.17[b], 0.21[c], 0.22[d,e1,f], 0.23[g], 0.24[e2,h], 0.29[i1], 0.30[j1], 0.31[i2], 0.32[k2], 0.34[j2,k1], 0.35[l,m,n], 0.37[o], 0.45[a]	0.28
Fat yield	0.19[b], 0.20[h], 0.25[f], 0.32[i2,j1], 0.35[j2,m], 0.37[k1,o], 0.39[i1], 0.40[k2]	0.32
Protein yield	0.04[h], 0.17[b], 0.23[f], 0.31[i1,j1], 0.34[j2,k2], 0.36[i2,k1], 0.37[m], 0.38[o]	0.29
Somatic cell score[1]	0.09[e], 0.12[p], 0.15[e], 0.20[j1], 0.21[q], 0.25[j2,p]	0.18
	Repeatability	
Milk yield	0.26[g], 0.37[h], 0.51[m], 0.64[e], 0.66[e]	0.42
Fat yield	0.22[h], 0.42[b], 0.49[m]	0.38
Protein yield	0.42[b], 0.52[m], 0.57[h]	0.50
Somatic cell score	0.46[p], 0.58[e], 0.59[e,p]	0.56

[1]Somatic cell score = average Log_2(somatic cell count).

[a] Mucha et al. (2014) (UK crossbred).
[b] Torres-Vázquez et al. (2009) (Saanen).
[c] Delfino et al. (2011) (Maltese).
[d] Valencia et al. (2007) (Saanen).
[e] Maroteau et al. (2014) ([1]Alpine and [2]Saanen).
[f] Selvaggi and Dario (2015) (Jonica).
[g] Morris et al. (1997) (Saanen).
[h] Rabasco et al. (1993) (Verata).
[i] Boichard et al. (1989) ([1]Alpine and [2]Saanen).
[j] Rupp et al. (2011) ([1]Alpine and [2]Saanen).
[k] Bélichon et al. (1998) ([1]Alpine and [2]Saanen).
[l] Morris et al. (2006) (Saanen).
[m] García-Peniche et al. (2012) (US goats).
[n] Valencia-Posadas et al. (2017) (Mixed).
[o] Castañeda-Bustos et al. (2014) (US goats).
[q] Apodaca-Sarabia et al. (2009) (Mixed).
[r] Bagnicka et al. (2016) (Polish White Improved and Polish Fawn Improved).

Table 2.4. Published phenotypic (above diagonal) and genetic correlations (below diagonal) between milking traits for dairy goats.

Trait	Milk yield	Fat yield	Protein yield	Somatic cell score
Milk yield		0.85[a]	0.95[a]	0.59[d]
		0.85-0.86[b]	0.93-0.95[b]	
		0.87-0.89[c]	0.94-0.95[c]	
Fat yield	0.72[a]		0.88[a,b]	
	0.77[b]		0.88-0.91[c]	
	0.80-0.84[c]			
Protein yield	0.87[a]	0.80[a]		
	0.89-0.92[b]	0.83-0.86[b]		
	0.90-0.93[c]	0.82-0.91[c]		
Somatic cell score	0.00-0.12[e]	-0.13--0.02[e]	-0.04-0.06[e]	

[a] Torres-Vázquez et al. (2009) (Saanen).
[b] Bélichon et al. (1998) (Alpine and Saanen).
[c] Boichard et al. (1989) (Alpine and Saanen).
[d] Bagnicka et al. (2016) (Polish White Improved and Polish Fawn Improved).
[e] Rupp et al. (2011) (Alpine and Saanen).

These genetic parameters are population specific and therefore must be estimated for the population of interest (van der Werf and de Boer, 1989). The only genetic parameters that have been published for milk production traits of dairy goats in New Zealand were for MY (Morris et al., 1997; Morris et al., 2006), SCS (Apodaca-Sarabia et al., 2009) and stayability (Wheeler et al., 2013). Although genetic parameters have been estimated for dairy goats worldwide, traits are influenced by genes and the environment, thus, using estimates from literature does not provide accurate estimates of genetic and environmental variation in New Zealand dairy goats. More research to obtain genetic parameters for milking traits of New Zealand dairy goats is required to develop an efficient breeding program for dairy goats in New Zealand. These genetic parameters will aid in calculating breeding values to be used in the construction of a selection index.

2.5 Genome-wide association study

With the declining cost of genotyping technologies and advances in computing capabilities, genomic studies are becoming increasingly popular in livestock breeding. In early 2007, the development of next-generation sequencing enabled de novo sequencing of the goat genome (Dong et al., 2013) which led to the development of the Illumina Goat SNP50 BeadChip, a high-density SNP chip containing 53,347 SNPs (Tosser-Klopp et al., 2014). A SNP stands for single nucleotide polymorphism which exhibits two or more nucleotide variants at a single base. These SNPs occur naturally across the genome and are a type of genetic marker commonly used in genomic studies. The advances in complex statistical models along with the availability of the Caprine 50K SNP chip provides multiple opportunities for the inclusion of genome-wide information into the genetic improvement program of the dairy goat industry through markers association studies and genomic prediction (GP).

A genome-wide association study (GWAS) is the analysis of genetic associations between genetic markers and specific traits (Goddard and Hayes, 2012). More specifically, these genetic markers are analysed for variation across the DNA sequence of the individual's genome (McCarthy et al., 2008). Identifying genetic markers associated with economically important traits provides the opportunity to increase the rate of genetic gain using genomic or marker-assisted selection. Animals with desirable genotypes can be identified and selected at a young age which can reduce the generation interval (Schaeffer, 2006) and increase rates of genetic gain. Genome-wide association studies have been performed in many livestock species, including dairy cattle (Mai et al., 2010; Pryce et al., 2010; Meredith et al., 2012), sheep (Zhao et al., 2011) and pigs (Sato et al., 2016; Le et al., 2017; Meng et al., 2017). Since release of the Caprine 50K SNP chip, associations of quantitative trait loci (QTL) in goats have been published for polledness (Kijas et al., 2013), milking speed (Palhière et al., 2014), wattles (Reber et al., 2015), coat colour (Becker et al., 2015; Martin et al., 2016a), supernumerary teats (Martin et al., 2016b), milk production and type traits (Maroteau et al., 2013; Martin et al., 2017; Mucha et al., 2018a). To date, there are no published papers reporting GWAS for dairy goats in New Zealand.

2.5.1 Fitting individual markers

One of the simplest GWAS is testing the association of a single marker and is referred to as a single-SNP GWAS (sGWAS). The sGWAS is based on a linear regression test of the fixed covariate effect of a single marker, which treats each SNP as if it has an additive effect. This model can be adjusted for the structure of the population by fitting principal components computed from the genomic relationship matrix (Price et al., 2006) as fixed effects. One problem with this model is that because each SNP is analysed separately, thousands of tests are performed, which incurs a multiple testing problem. To account for multiple testing the significance level must be stringent and can be calculated using the Bonferroni correction. However, setting a significance threshold combined with the testing of so many marker effects means that the markers likely to exceed the threshold are those with favourable error terms and in turn have overestimated effects.

A GWAS uses linkage disequilibrium (LD) based methods to identify the associations between the genetic markers and phenotypic expression of traits of interest. Linkage disequilibrium is the non-random association of alleles at two or more loci and is influenced by factors such as population history, effective population size, relatedness between individuals and the pattern of geographic subdivision (Slatkin, 2008). Analysing individual SNPs relies on the LD between each marker and the QTL. Therefore, the power of a sGWAS may suffer if the individual SNP is in low LD with the causal mutation and the LD contained in flanking markers is ignored (Fernando and Garrick, 2013). On the other hand, it is also possible to overestimate the effects of the individual SNPs. For example, several SNPs could all be in LD with the same QTL and therefore each SNP could either explain a part of the QTL effect, or each SNP could be explaining the same part of the QTL effect, which would lead to false positives (Martin et al., 2016b).

Overall, this single-SNP approach is capable of detecting a signal (causal mutation), however, most economically important traits are complex and controlled by several major loci with small effects. In this case, a more precise method of estimating the number of QTL is by calculating the variance explained by the effects of the SNPs in a specified chromosomal region (Habier et al., 2011).

2.5.2 Fitting all markers simultaneously

Bayesian multiple-regression models can be used to simultaneously fit thousands of SNPs as random effects to determine the proportion of variance explained by the markers (Fernando and Garrick, 2013). Fitting all SNPs simultaneously in the model takes into account the LD between neighboring SNPs which limits the false positive discoveries (Fernando and Garrick, 2013). Also, it is expected that SNPs near each other will be highly correlated, therefore, analysing markers within a genomic window would expect to capture most of the variability at a nearby trait locus (Fernando and Garrick, 2013). Thus, instead of using P-values to determine significant associations as with the single-SNP approach, these Bayesian methods make inferences of associations based on the variance explained by each genomic window (Misztal et al., 2020). The windows that explain the greatest variance can be used to identify the most informative genomic regions, facilitating the discovery of associated markers and possible causal mutations (Fernando and Garrick, 2013). By sliding the window over the chromosome and observing peaks that are greater than those obtained for the single SNPs, the number of actual QTLs may be inferred more accurately (Habier et al., 2011).

In addition to sGWAS or fitting all markers simultaneously, the SNPs can be combined into a haplotype block. A haplotype block is a cluster of SNPs that tend to be inherited together. Therefore, clustering SNPs into a haplotype block combines information of adjacent SNPs into composite multi-locus haplotype alleles which may be more informative than individual SNPs and may also capture the regional LD information, which is arguably more robust and powerful (Pritchard et al., 2000; Akey et al., 2001). Such haplotypes can be included in the GWAS analysis to further investigate the true associations obtained from the SNP analyses.

The discovery of thousands of SNPs and their application in GWAS has facilitated the identification and localisation of regions that control quantitative traits in dairy cattle (Georges et al., 1995; Jiang et al., 2010; Mai et al., 2010; Pryce et al., 2010; Meredith et al., 2012) and pigs (Sato et al., 2016; Le et al., 2017; Meng et al., 2017). A few studies have analysed SNPs and their associations with milking traits in dairy goats. For example, Martin et al. (2017) attempted a GWAS for dairy goats and revealed that two mutations (R251L and

R396W) of the Diacylglycerol O-Acyltransferase 1 (DGAT1) gene were responsible for decreased milk fat content. But, due to the small sample sizes (4,563 and 1,941 goats, respectively), the associations with candidate genes should be treated as an indication and further research is required before validation. Thus, there is still very little known about the loci controlling milk traits in goats.

2.6 Estimation of breeding values

2.6.1 Pedigree and phenotypic based breeding values

In animal breeding, the best linear unbiased prediction (BLUP) is the main method for predicting the genetic merit of individuals in a population (Henderson, 1950; Henderson, 1963). This method uses phenotypic records of the individual and its relatives (Garrick and Fernando, 2014) to derive unbiased estimates of the linear functions of the fixed effects, and the random animal effect included in the mixed model. Pedigree information is used to estimate the average genetic relationships among the individuals based on the probability that genes are identical by descent (Wright, 1922), i.e. half- siblings born to unrelated non-inbred parents are expected to share 0.5 of their alleles, and these probabilities are the basis for generating the average genetic relationship matrix (A) between close and distant relatives in the pedigree. A traditional single-trait analysis BLUP model is;

$$y = \mathbf{X}b + Za + e \qquad (2.2)$$

where

y is the vector of all observations,

X is the design matrix relating fixed effects in b to y,

b is the vector containing fixed effects,

Z is the design matrix relating genetic effects in a to y,

a is the vector containing the animal additive genetic effects,

e is the vector of residual effects.

It is assumed that the expectations (E) of the variables are: E(y) = Xb, E(a) = 0, E(e) = 0. It is assumed that residual effects, which includes random environmental and non-additive genetic effects, are independently distributed with variance σ^2_e; therefore, var(e) = $I\sigma^2_e$ = R; var(a) = $A\sigma^2_a$ = G and cov(a,e) = 0, where I is an identity matrix of order n (the number of records), A is the numerator relationship matrix from the pedigree and σ^2_a is the additive genetic variance. Since cov(a,e) = 0, then:

$$var(y) = V = ZGZ' + R \qquad (2.3)$$

The implications from the assumptions in the animal model described above include:

(i) All genetic values are from the same distribution and have common genetic variance, in the absence of inbreeding.
(ii) All residual effects have the same variance and are independent.
(iii) Random effects u and e are assumed to have zero covariance, equivalent to assuming no genotype-environment interaction.

In 1963, Henderson published a theory which combined the selection index theory (Hazel, 1943) with the least squares method, to find the best linear unbiased estimators of **β**, and to use these estimators, **β**° in predicting u satisfying the above criteria. Henderson's mixed-model equations (MME) correspond to a very general matrix model in which u can comprise of several random factors. Given the assumptions explained in the definition of the model (Equation 2.2), the MME reduce to (Henderson, 1975):

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z+\alpha A^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^\circ \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \qquad (2.4)$$

with

$$\alpha = \frac{\sigma^2_e}{\sigma^2_a}$$

where

A^{-1} is the inverse of the relationship matrix between individuals in u.

β° is the vector of solutions for fixed effects,

û is the vector of animal solutions, which are the EBVs.

This BLUP methodology has been applied for genetic evaluation of dairy goats in France (Boichard et al., 1989; Ducrocq, 1992; Clément et al., 2002) and Norway (Andonov et al., 2007). These MME can also be extended to multi-trait analysis and more recently genomic predictions (Meuwissen et al., 2001).

2.6.2 Prediction of genomic breeding values

Genomic prediction uses information from genetic markers covering the whole genome to predict GBVs for individuals without phenotypic records (Meuwissen et al., 2001). Using high-density arrays, it is assumed that all the QTL that contribute to trait variation are in high LD with at least one marker or haplotype (Meuwissen et al., 2001). The GBVs are calculated as the sum of the effects of markers or marker haplotypes across the entire genome, thereby potentially capturing all of the additive genetic variance of the trait (Hayes et al., 2009). Genomic prediction is now widely practiced across commercial livestock species such as dairy goats (Carillier et at., 2013; Carillier et al., 2014; Mucha et al., 2015), dairy cattle (Hayes et al., 2009; VanRaden et al., 2009; Boichard et al., 2012; García-Ruiz et al., 2016), dairy sheep (Duchemin et al., 2012; Baloche et al., 2014), meat sheep (Banks et al., 2009; Brito et al., 2017a), beef cattle (Weber et al., 2012; Guo et al., 2017; Zhu et al., 2019), pigs (Christensen et al., 2012; Knol et al., 2016), and poultry (Wolc et al., 2015). Although GP has been implemented in New Zealand dairy cattle (Spelman et al., 2013), sheep (Dodds et al., 2014; Nilforooshan, 2020) and trees (Suotama et al., 2019), this technology has not been applied in the New Zealand dairy goat industry.

2.6.2.1 Multi-step genomic prediction

The basic principle of GP includes a set of individuals that have phenotypic records and genotypic information (referred to as the training or reference population), that are used to construct a model that predicts the GBV of individuals for which only genetic information is available (validation population). Genomic prediction is a very active area of research where new algorithms, software and methods are constantly being developed. Within this context the following section will explain the general mixed model and briefly introduce a few key ideas of the most common genomic prediction models currently used in livestock breeding such as the ridge-regression BLUP (RR-BLUP), fixed regression least squares (FR-LS), genomic BLUP (GBLUP) and Bayesian approaches (BayesA, B, C and Cπ). The general equation used for the prediction of total genetic merit using genomic information is as follows (Meuwissen et al., 2001):

$$y = \mathbf{X}b + Ms + e \quad (2.5)$$

where

y is the vector of phenotypes,

X is the design matrix relating fixed effects in b to y,

b is the vector containing fixed effects,

M is the matrix of centred marker covariates observed on genotyped animals,

s is the vector containing additive marker effects,

e is the vector of residual effects.

This estimates of marker effects are obtained from the following MME (Henderson, 1984):

$$\begin{bmatrix} X'X & X'M \\ M'X & M'M+\lambda I \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{s} \end{bmatrix} = \begin{bmatrix} X'y \\ M'y \end{bmatrix} \quad (2.6)$$

where

$\hat{b}$ is the vector of solutions for fixed effects,

$\hat{s}$ is the vector of marker effects,

λ is the proportion of the residual variance that is explained by the SNPs and is calculated as $\sigma^2_e/\sigma^2_{SNP}$ where σ^2_e = residual variance and σ^2_{SNP} = marker variance,

I is an identity matrix of size equal to the number of markers.

Leading to:

$$\hat{u} = M\hat{s} \quad (2.7)$$

therefore

$\hat{u}$ is a vector of animal solutions, which are the estimated GBVs.

The main differences between the GP models are the number of SNPs included in the model and the assumptions of the distribution of SNP effects.

In RR-BLUP, all markers are included in the model and are assumed to contribute equal variance. This approach was proposed by Meuwissen et al. (2001), which includes a penalty parameter (λ) that shrinks marker effects uniformly towards zero and was calculated as σ^2_g/k where σ^2_g is the total genetic variance and k is the number of markers. However, Habier et al. (2007) demonstrated that this calculation of λ is statistically equivalent to BLUP using the average genetic relationship matrix (A), and suggested λ to be calculated as:

$$\lambda = \sigma^2_g / 2\sum_k p_k(1-p_k) \quad (2.8)$$

where σ^2_g is the additive genetic variance, and p_k is the allele frequency at marker locus k. Biologically, we would expect some SNPs to be in close LD to a gene and therefore have an effect on the trait of interest, while other SNPs may not be in LD and therefore have no effect on the trait. However, despite the underlying assumption that all SNPs contribute equal variance, the RR-BLUP approach performs well when the traits are controlled by many loci with small effects (Lorenz et al., 2011).

Meuwissen et al. (2001) also proposed a fixed regression least squares (FR-LS) method that predicts GBVs using information from all markers located across the entire genome in an attempt to capture all of the QTL influencing the variation in a trait (Hayes et al., 2009). Similar to RR-BLUP, for the FR-LS model it is assumed that the effects of all SNPs are normally distributed, and all SNPs have equal variance (λ = constant) (Meuwissen et al., 2001; VanRaden, 2008). The difference between RR-BLUP and FR-LS is that in FR-LS, λ is the proportion of the residual variance that is explained by the SNPs and is calculated as $\sigma^2_e/\sigma^2_{SNP}$, as in Equation 2.6. This ratio dictates the extent of shrinkage in the prediction of marker effects (Garrick et al., 2014).

An equivalent model to FR-LS is genomic BLUP (GBLUP), which fits a genomic relationship matrix (VanRaden, 2008) in place of the traditional average genetic relationship matrix (A) in the BLUP model (Equation 2.4) (Garrick, 2007; VanRaden, 2007). This GBLUP approach is preferred over FR-LS as regressions are on genotypes rather than haplotypes, and at high marker numbers, haplotyping would increase computation time with minimal gain in accuracy (Calus et al., 2008).

Bayesian models were proposed for GP to overcome the limitation of homogeneous shrinkage of marker effects by partitioning the genetic variance among markers. In Bayesian models a fraction (π) of the SNPs have an effect whereas a fraction (1-π) have no effect on the trait (Fernando and Garrick, 2013). This fraction allows the model to perform marker specific shrinkage of estimates by specifying an appropriate prior density. This assumption that not all markers have an effect, agrees with the fact that some of the chromosome segments contain QTL with large effects, some with small effects, and some have no QTL. The assumed prior density of marker effects determines the extent and type of shrinkage induced and

whether the model will induce variable selection and shrinkage or shrinkage only (de los Campos et al., 2013). Bayesian models are known as variable selection models as they only include markers in the prediction model if they are estimated to have an effect. A brief overview of the methodology of Bayesian models is described:

In Bayesian models, a prior probability, based on previous knowledge (which can be vague), is assigned to the data. The prior distribution for the variance of marker i (V_{ai}) assumed as:

V_{ai}= 0 with probability p

$V_{ai} \sim \chi^{-2}(v,S)$ with probability (1-p)

where p depends on the marker mutation rate and $\chi^{-2}(v,S)$ represents the inverse-Chi squared distribution, with v degrees of freedom and scale parameter S. Both parameters v and S depend on the distribution of the mutational effects of the markers, which, in practice, require estimation (Meuwissen, 2003). Through the Bayes theorem, this prior is combined with the evidence arising from the data (the likelihood) to obtain a posterior distribution from which inferences are made. The overall application of this Bayes theorem becomes complex as its calculation involves multiple integrations. For this reason, Bayesian analyses are commonly carried out with the use of Markov chain Monte Carlo (MCMC) methods, in which inferences on the parameters are obtained from statistics of samples obtained empirically from the likelihood iterations.

In BayesA, it is assumed that all markers have an effect ($\pi = 0$) and are included in the model (Meuwissen et al., 2001). For BayesA the genetic variance is partitioned unequally among all markers for marker-specific variances (σ^2_{snp}) (λ = varies for each marker). A student-t distribution is used as a prior for the SNPs with effects, which allows some SNPs to have large effects on the trait (Meuwissen et al., 2001). In BayesB, to accommodate the assumption that many SNPs have a zero effect ($\pi > 0$), these markers are excluded from the model and the genetic variance is partitioned unequally among the remaining subset of markers (Habier et al., 2011). The fraction of markers assumed to have non-zero effects are drawn from distributions with marker-specific variance (λ = varies for each marker) (Meuwissen et al., 2001). Kizilkaya et al. (2010) developed a BayesB-like mixture model

called BayesC which assumes that the all markers that have an effect explain equal genetic variance (λ = constant). However, BayesC is similar to BayesB in that they both assume that a proportion of the markers (π > 0) contribute to the genetic variance. As only a fraction of the markers are assumed to have an effect, the BayesB and BayesC models use a mixture of two priors, one with a point of mass at zero and the other that can either be a scaled-t distribution in BayesB (Meuwissen et al., 2001) or a Gaussian distribution in BayesC (Habier et al., 2011). Previous studies have indicated that Bayesian models that differ in their prior assumptions tend to produce different inferences about individual marker effects and GBVs; although, in cross-validation studies they often have similar predictive performance (Gianola, 2013). The main differences between these prediction models are shown in Table 2.5.

Table 2.5. Key parameters considered for each of the genomic prediction models.

Parameter	RR-BLUP	FR-LS	GBLUP	BayesA	BayesB	BayesC
Number of markers included in the model	All	All	All	All	1-π	1-π
Marker variance (λ)	Constant	Constant	Constant	Variable	Variable	Constant
π	NA	NA	NA	NA	Known	Known

Based on the underlying assumptions of these prediction models, when a trait is controlled by a few QTL with moderate-to-large effect, the Bayesian variable selection approaches are expected to perform better than the RR-BLUP or GBLUP models. This was demonstrated by Wang et al. (2019) in Chinese Simmental beef cattle, reporting that when the traits are influenced by fewer genes but of large effect, it seems that Bayesian methods have a small advantage over linear models such as GBLUP, whereas, GBLUP may outperform BayesB for a trait with many loci with small effects (Wang et al., 2019). This demonstrates that although these prediction methods have been successfully implemented in livestock species, these methods may perform differently for different traits (Hayes et al., 2009).

It must be noted that these GP models can only be applied to animals with genotypes, and in a practical scenario only a subset of the population, generally the top sires, are genotyped. However, this genomic information can still be applied to the population using a "multi-step" approach (Meuwissen et al., 2016): 1) pseudo-phenotypes for genotyped animals are calculated using information (phenotypes) on its ungenotyped relatives, 2) GP is performed on genotyped animals using the pseudo- phenotypes and their genotypes, 3) the pedigree-based EBVs and GBVs are combined into a total EBV (VanRaden, 2008). The pseudo-phenotypes used in this multi-step approach could be de-regressed breeding values (Garrick et al., 2009), which require preliminary evaluation of the performance of the buck's progeny, referred to as a progeny-test, or daughter-yield deviations (VanRaden and Wiggans, 1991). Although this method does provide greater accuracies than pedigree-based EBVs (Harris and Johnson, 2010), handling the data in multiple steps is clearly suboptimal.

2.6.2.2 Single-step genomic prediction

Legarra et al. (2009) proposed a method that uses phenotypes, genotypes and pedigree information to predict GBVs for both genotyped and non-genotyped individuals in one single-step, and is referred to as single-step GBLUP (ssGBLUP). Implementing ssGBLUP in the prediction of the genetic merit generally results in higher accuracy due to the utilisation of all available data (Silva et al., 2016). Moreover, predicting all genotyped and non-genotyped individuals simultaneously reduces prediction bias (Vitezica et al., 2011; Christensen et al., 2012). This method incorporates all individuals in the evaluation by combining the average genetic relationship matrix (A) with the genomic relationship matrix (G), into a modified relationship matrix (H) (Legarra et al., 2009):

$$H = \begin{pmatrix} A_{nn} + A_{ng}A_{gg}^{-1}(G - A_{gg})A_{gg}^{-1}A_{gn} & A_{ng}A_{gg}^{-1}G \\ GA_{gg}^{-1}A_{gn} & G \end{pmatrix} \quad (2.8)$$

where

A_{nn} is a sub-matrix of the A for non-genotyped animals,

A_{gg} is a sub-matrix for genotyped animals,

A_{ng} (or A_{gn}) are sub-matrices that describe the pedigree-based relationship between non-genotyped and genotyped animals.

This single-step approach uses Henderson's MME and the H to yield unbiased predictions under multivariate normality, even in populations that are undergoing selection and non-random mating. This single-step procedure increases both power and precision by taking advantage of phenotypes from related and unrelated animals. Despite these advantages, ssGBLUP requires computation of the G or its inverse, which can be computationally demanding when many animals are genotyped.

Several alternative single-step approaches have also been proposed in an attempt to reduce computations with many genotyped animals (Legarra and Ducrocq, 2012; Fernando et al., 2014; Liu et al., 2014; Taskinen et al., 2017). These approaches include equations where G is not inverted, the SNP effects are estimated for genotyped animals and a polygenci effect is fit for non-genotyped animals, or, SNP effects are estimated for all animals using imputed genotypes. For example, Fernando et al. (2014) proposed a class of single-step Bayesian regression methods that does not require the computation of the G or its inverse. Instead, this single-step Bayesian approach imputes marker covariates for non-genotyped animals based on their genotyped relatives and a genetic imputation error effect to accommodate the difference between true and imputed genotypes (Fernando et al., 2014). Later, Fernando et al. (2016) also proposed another single-step approach called a single-step hybrid model that utilises Bayesian regression analyses but requires considerably less computing effort.

In an attempt to minimise prediction bias of GBVs an extra polygenic term can be included in the prediction model to account for the additive genetic variance not explained by the markers (Goddard et al., 2007; Christensen and Lund, 2010; Liu et al., 2011). Including this term in the model tends to increase prediction accuracies if the prediction uses low marker

density panels, however, if high marker density panels are used, or the markers already explain most of the genetic variance, inclusion of an extra polygenic effect will hardly increase prediction accuracies (Calus and Veerkamp, 2007). Only a few published studies have applied the single-step method with an extra polygenic effect included in the model. For example, in a ssGBLUP evaluation of dairy goats in the UK, Mucha et al. (2015) included an extra polygenic effect of 10% of the additive genetic variance to avoid high variance and to minimise prediction bias of GBV.

The ssGBLUP approach has been explored in many simulation studies (Kang et al., 2017; Bradford et al., 2019) and successfully applied to different species, including dairy goats (Desire et al., 2017; Mucha et al., 2018b), dairy cattle (Aguilar et al., 2010; Harris et al., 2012; Liu et al., 2014; Koivula et al., 2015), beef cattle (Moore et al., 2018), sheep (Swan et al., 2012; McMillan and Swan, 2017; Brown et al., 2018), broilers (Chen et al., 2011) and pigs (Christensen et al., 2012). The single-step Bayesian regression has been implemented for beef cattle (Lee et al., 2017; Golden et al., 2018). In a simulation study comparing the prediction accuracy of single-step BayesA, single-step BayesB and ssGBLUP with various numbers of QTL (5, 50, and 500), Zhou et al. (2018) reported that single-step BayesA and single-step BayesB models were advantageous over ssGBLUP when there were fewer QTL affecting the trait. Concluding that single-step BayesA was the most robust and efficient model across all QTL scenarios. In addition, the authors noted that accuracies of ssGBLUP did not change significantly as the number of QTL changed, however single-step BayesA and single-step BayesB accuracies significantly decreased as the number of QTL increased. These results suggest single-step Bayesian models are more sensitive to the number of QTL affecting the trait while ssGBLUP is more robust model to handle scenarios with different number of QTL. Overall, these results are similar to the multi-step predictions, in which the Bayesian approaches are adventageous when there are fewer QTL but with medium to large effect.

2.7 Factors affecting accuracy of prediction

Simulation and empirical studies in animal breeding programs indicate that the accuracy of GP is influenced by several parameters including the heritability of the trait (Viana et al., 2017), the number of animals in the reference population (VanRaden et al., 2009; Daetwlyer et al., 2012), the relationship between animals in the reference population and the target animals to be predicted (Meuwissen et al., 2001; Solberg et al., 2008; Clark et al., 2012), the extent of LD between the SNPs and QTL (Meuwissen et al., 2001), the distribution of the QTL effects (Meuwissen et al., 2001; Goddard, 2009; Hayes et al., 2009) and of course the prediction method used (Calus, 2010).

2.7.1 Heritability

As trait heritability is the proportion of the total variation due to the genetic variance, it is not surprising that more heritable traits have greater prediction accuracies (Zhang et al., 2019). However, compared to phenotypic selection, the efficiency of genomic selection increases as the heritability of the trait decreases (Viana et al., 2017). This is primarily due to the fact that the genomic data provides more information to predict the breeding values (Bouquet and Juga, 2012). This can be seen from the results of using genomic selection in US dairy cattle, where the rate of genetic gain per year increased by 50-100% for high heritability traits such as milk yield, but increased by 300-400% for low heritability traits such as daughter pregnancy rate (García-Ruiz et al., 2016).

2.7.2 Size of reference population

The size of the reference population is an important factor influencing GP as this information is the basis of the predictions. The more animals in the reference population will provide more data available to estimate marker effects which in turn will increase prediction accuracies (Meuwissen et al., 2001; VanRaden et al., 2009). Meuwissen et al. (2001) showed that training populations consisting of 500, 1,000 and 2,200 animals obtained prediction accuracies of 0.58, 0.66, and 0.73 using the traditional BLUP approach,

and 0.71, 0.79, and 0.85 using a BayesB approach, respectively. In addition to population size, the relationship between animals in the reference and validation populations also has a significant effect on prediction accuracies (de los Campos et al., 2013). Close relationships between the two populations results in the greatest GBV accuracies (Habier et al., 2007; Habier et al., 2010; Clark et al., 2012; Garrick et al., 2012; Habier et al., 2013; Kang et al., 2017). Using a simulation study Zhou et al. (2018) investigated the influence of relationships between the training and validation populations on the accuracy of GBVs using various single-step prediction models. Zhang et al. (2018) reported that prediction accuracies of ssGBLUP, single-step BayesA and single-step BayesB models all decreased as the distance of the validation population increased. Thus, the contribution of genetic relationships to the prediction of GBVs is different in each generation and accuracies will decrease over generations (Habier et al., 2010; Kang et al., 2017). For example, the contribution to the prediction accuracy of the parents of individuals in the training population can be high, but the information from genetic relationships is halved each generation for the following generations. In addition, the contribution to prediction accuracy from the relationships between the two populations is likely correlated to the accuracy obtained from the extent of LD between the markers and QTLs, as the level of LD increases when subgroups are closely related (Daetwyler et al., 2012). However, the accuracy due to the extent of LD tends to be more persistent across generations and breeds than the accuracy due to relationships, which makes LD of particular importance in GP (Meuwissen et al., 2001; Habier et al., 2007; de Roos et al., 2009).

2.7.3 Level of linkage disequilibrium

Genomic prediction is based on the idea that all QTL that contribute to trait variation will be in LD with at least one marker and therefore captured in the prediction model (Meuwissen et al., 2001). The extent of LD is quantified as the correlation between two loci (r^2) which generally increases as the density of markers increases (Garrick et al., 2012). The greater level of LD between markers is related to more accurate GBVs while an r^2 value greater than 0.20 is suggested to be enough for genomic selection (Meuwissen et al., 2001;

Calus et al., 2008). The success of GP depends largely on the existence of LD across the population of interest (Carillier at al., 2013; Baloche et al., 2014). Therefore, when all individuals come from the same population, the LD between genetic markers and QTL persists from the training to the validation population. Marker densities also influence prediction accuracies, with lower densities resulting in poorer predictions (Solberg et al., 2008). Increasing marker density also provides greater chance that a QTL will be in LD with at least one marker, which will increase the prediction accuracy, until prediction accuracy reaches a plateau and does not increase further as marker density increases (Lee et al., 2017). For example, in New Zealand dairy cattle, Spelman et al. (2014) reported minimal improvement in GPs when moving from the 50K SNP panels (Harris et al., 2011) to using a 777K marker panel. Also, the increased marker density can increase the risk of slow convergence or even no convergence of MCMC iterations in Bayesian methods, which may result in low prediction accuracy (Zhang et al., 2019).

2.7.4 Prediction model

Genomic prediction models can use single markers, haplotypes of markers, or using an identical by descent approach (Meuwissen et al., 2001; Goddard and Hayes, 2007; Calus et al., 2008). Methods such as ssGBLUP use all markers and information from the genetic relationship matrix (Meuwissen et al., 2001), while Bayesian models can use a subset of markers (Habier et al., 2011). Also, the single-step approaches enable the use of pedigree and genotypic information from all animals in the population, usually resulting in greater accuracies than GBLUP, due to the utilisation of all available data (Carillier et al., 2014; Silva et al., 2016). An important factor when considering the appropriate model to implement is the genetic architecture (number and position of QTL, magnitude of QTL effects) of the traits of interest. For example, the standard GBLUP approaches assume that all SNPs follow the same distribution and contribute the same level of variance, thus all SNPs are assigned the same weight in the model (Legarra et al., 2009; Stranden and Garrick, 2009; Christensen and Lund, 2010; Wang et al., 2012; Zhang et al., 2016). Meanwhile Bayesian methods are able to consider that SNPs explain different proportions of the genetic variance (Habier et

al., 2011). Allocating more variance to a subset of SNPs allows these Bayesian methods to take into account the presence of QTL or major gene effects (Teissier et al., 2018). If prior information is known or assumed about the genetic architecture of the traits of interest then these can be used to modify the distribution of SNP effects (Teissier et al., 2018), or use a model which the underlying assumptions match. For example, the α_{s1} casein gene is known to have a significant effect on protein content in French dairy goats (Teissier et al., 2018). Teissier et al. (2018) demonstrated that prediction accuracies of a weighted-ssGBLUP model, in which weights for SNP variances are used when forming the genomic relationship matrix (Legarra and Ducrocq, 2012), was more accurate than the regular unweighted ssGBLUP model.

2.8 Summary of literature review

There is tremendous opportunity to improve the production of milk produced in the New Zealand dairy goat industry. This review of the literature has identified that there is an animal evaluation system in the industry, but there is no selection scheme to select and disperse genes from superior animals into the commercial population. Also, there is lack of reports of the estimation of genetic parameters for New Zealand dairy goats, restricting the estimation of breeding values for economically important traits. Additionally, there has been no genomic research despite the potential benefit of genomic information available to the New Zealand dairy goat industry.

Chapter 3

Genetic parameters for total lactation yields of milk, fat, protein, and somatic cell score in New Zealand dairy goats

This Chapter has been published in part elsewhere. It has been reformatted and presented here with permission:

Scholtens MR, Lopez-Villalobos N, Garrick DJ, Blair HT, Lehnert K, Snell RG. 2019. Genetic parameters for total lactation yields of milk, fat, protein, and somatic cell score in New Zealand dairy goats. Animal Science Journal DOI:10.1111/asj.13310.

Abstract

The aim of this study was to estimate genetic parameters for lactation yields of milk (MY), fat (FY), protein (PY) and somatic cell score (SCS) of New Zealand dairy goats. The analysis used 64,604 lactation records from 23,583 does, kidding between 2004 and 2017, distributed in 21 herds and representing 915 bucks. Estimates of genetic and residual (co) variances, heritabilities and repeatabilities were obtained using a multiple-trait repeatability animal model. The model included the fixed effects of contemporary group (does kidding in the same herd and year), age of the doe (in years) and as covariates, kidding day, proportion of Alpine, Nubian, Toggenburg, and "unknown" breeds (Saanen was used as the base breed), and heterosis. Random effects included additive animal genetic and doe permanent environmental effects. Estimates of heritabilities were 0.25 for MY, 0.24 for FY, 0.24 for PY, and 0.21 for SCS. The phenotypic correlations between MY, FY and PY ranged from 0.90 to 0.96, and the genetic correlations from 0.81 to 0.93. These results indicate lactation yield traits exhibit useful heritable variation and that multiple trait selection for these traits could improve milk revenue produced from successive generations of New Zealand dairy goats.

Introduction

Dairy goat farming in New Zealand is a profitable industry able to access niche markets for high value products. The New Zealand Dairy Goat Cooperative's members collectively manage 80% of the nation's dairy goat population (Scholtens et al., 2017). The cooperative is the leading international manufacturer of goat milk nutritional powders for infants and young children (Stafford and Prosser, 2016). The industry is dependent on goat milk with high total milk solids and low bacterial count for the manufacture of high-quality products. Selecting animals based on a selection index utilises estimates of genetic merit encompassing economically relevant traits. A breeding program would enable the index selection of genetically superior animals which would rapidly improve the quantity and

quality of milk produced from New Zealand dairy goats, and results in improved profits for farmers.

Modern breeding programs include routine genetic evaluation to enable selection of the best parents to produce the next generation of animals. Genetic and environmental factors significantly influence milk yield and quality in ruminants (Selvaggi and Dario, 2015), and genetic gain from selection will be enhanced if these environmental factors are accounted for in the estimation of genetic merit. Knowledge of variance components for production traits will enable the design of an effective genetic evaluation strategy, allowing the selection of animals with superior overall genetic merit, optimising direct and correlated selection responses for traits of economic importance (Barillet, 2007). Estimates of genetic parameters for economically relevant traits have been reported for dairy goats in South Africa (Muller, 2005), France (Boichard et al., 1989; Bélichon et al., 1998), Spain (Analla et al., 1996), the United Kingdom (McLaren et al., 2016) and New Zealand (Morris et al., 1997; Morris et al., 2006). The published values are summarised in Table 2.3 including heritability and repeatability statistics. The genetic and phenotypic correlations are summarised in Table 2.4.

Genetic parameters previously reported for dairy goats in New Zealand for yields of milk, fat plus protein (Morris et al., 1997; Morris et al., 2006) have been estimated based on data from a single Saanen goat herd with bi-variate repeatability models. The aim of this study was to use a dataset from a much larger multi-farm dairy goat population and estimate genetic parameters for MY, FY, PY and SCS by fitting a multiple-trait repeatability animal model.

Methods

Data

Pedigree information and lactation records for MY, FY, PY and SCS from Alpine, Nubian, Saanen, Toggenburg and crossbred dairy goats kidding between 2004 and 2017 were obtained from the herd-test database maintained by Livestock Improvement Corporation.

The original data set comprised 182,386 lactation records from 87,176 does distributed in 77 herds across the North Island of New Zealand. The pedigree included 1,076 sires and 23,949 dams. Herds were excluded when less than 80% of the dams were recorded. The final data set contained 64,604 lactation records from 23,583 does distributed in 21 herds. The does were offspring of 915 sires and 12,108 dams with pedigrees that spanned up to 9 generations. The 778 sires had progeny in only one herd, and 137 sires had progeny in two herds or more. Breed composition of each doe was calculated from pedigree proportions of Alpine, Nubian, Saanen, Toggenburg and "unknown" breeds. There was some crossbreeding but very few first-cross or purebred animals of Alpine, Nubian and Toggenburg breeds. Therefore, the proportion of these breeds were summed into a single combined breed group called ANT. Structure of the dataset is provided in more detail in Table 3.1. All herds have Saanen and it is used as a base breed for crossbred. There was a total of 136 herd-year contemporary groups, mostly consisting of more than one breed or cross.

Table 3.1. Number of animals and lactation records, and average breed composition of the goat population classified by proportion of Saanen.

Proportion of Saanen (%)	N[1]	Breed					Lactation records	NSR[2]
		Saanen	Alpine	Nubian	Toggenburg	Unknown		
>87.5	4,754	0.993	0.000	0.000	0.001	0.006	11,971	2,890
> 75 - ≤ 87.5	1,280	0.853	0.001	0.002	0.026	0.119	4,318	1,280
> 0.50 - ≤ 0.75	3,417	0.684	0.001	0.003	0.052	0.260	10,801	3,390
> 0.25 - ≤ 0.50	6,377	0.449	0.002	0.003	0.035	0.511	17,907	1,887
> 0.125 - ≤ 0.25	3,671	0.231	0.001	0.004	0.025	0.739	9,528	694
> 0 - ≤ 0.125	2,806	0.102	0.001	0.001	0.013	0.883	7,051	401
0	1,278	0.000	0.029	0.032	0.256	0.683	3,028	687

[1]N = number of animals, [2]NSR = this is the number of animals with sire recorded in the pedigree.

Statistical analysis

The test interval method as described by Sargent et al. (1968), which has a high accuracy (0.96-0.97) for estimating lactation yields from test-day records (Norman et al., 1999), was used by Livestock Improvement Corporation to calculate MY, FY and PY for either the actual realised lactation length or up to 305 days in milk (DIM) for those lactations with more than 305 DIM.

Current practice for the dairy goat industry is to start milking very soon after kidding (after the colostral phase is over) and start supplying The Dairy Goat Cooperative (NZ) Ltd (DGC) after the first eight milkings after kidding. Milking is twice daily and herd-testing occurs three to four times each season. Average SCS over the lactation was calculated as the mean Log_2(somatic cell count) from each herd-test.

Descriptive statistics of MY, FY, PY and SCS were obtained using the MEAN procedure of Statistical Analysis System version 9.4 (SAS Institute Inc., Cary, NC, USA). Normality was tested using the UNIVARIATE procedure of SAS.

The genetic parameters were estimated using a multiple-trait repeatability animal model represented as:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_n \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} + \begin{bmatrix} Z_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Z_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} W_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_n \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix} + \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} \quad (3.1)$$

where for traits i =1,2,..n:

y_i = vector of observations on all the measured animals,

b_i = vector of fixed effects for trait i,

a_i = vector of random (additive genetic) effects,

p_i = vector of permanent environmental effects,

e_i = vector of random residual effects,

X_i = incidence matrix for the fixed effects,

Z_i = incidence matrix relating observations to animals,

W_i = incidence matrix for the permanent environmental effects.

Fixed class effects included in b_i were contemporary group (defined as does kidding in the same herd and year), and doe age in years. Fixed covariables included day of kidding, proportion of ANT, and "unknown" breeds (Saanen was used as the base breed), and coefficient of general heterosis. General heterosis was calculated as $1 - \sum_{j=1}^{f} \mathbf{p}_j^2$ where p_j is the proportion of each of the f breeds (Gregory and Cundiff, 1980; Lamberson et al., 1993). General heterosis was calculated because the number of crossbred animals for each two-breed combination was inadequate for fitting specific pair-wise heterosis values. General heterosis assumes that first-cross heterosis is the same for all breed combinations. Random effects included in the model were additive genetic and permanent environmental animal effects. The distributional properties of the elements in the model, with expectation (E) and variance-covariance structures (Var) were as follows:

$$E\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_n \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \tag{3.2}$$

and

$$Var\begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} A\sigma^2_{a_1} & \cdots & A\sigma_{a_{1n}} \\ \vdots & \ddots & \vdots \\ A\sigma_{a_{n1}} & \cdots & A\sigma^2_{a_n} \end{bmatrix} \tag{3.3}$$

and

$$Var\begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix} = \begin{bmatrix} I_1\sigma^2_{p_1} & \cdots & I_1\sigma_{p_{1n}} \\ \vdots & \ddots & \vdots \\ I_1\sigma_{p_{1n}} & \cdots & I_1\sigma^2_{p_n} \end{bmatrix} \tag{3.4}$$

and

$$\text{Var}\begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix} = \begin{bmatrix} I_2\sigma^2_{e_1} & \cdots & I_2\sigma_{e_{1n}} \\ \vdots & \ddots & \vdots \\ I_2\sigma_{e_{1n}} & \cdots & I_2\sigma^2_{e_n} \end{bmatrix} \quad (3.5)$$

where

A = numerator relationship matrix among animals,

$\sigma^2_{a_i}, \sigma^2_{a_j}, \sigma_{a_{ij}}$ = animal (co)variance components for traits i and j,

$\sigma^2_{p_i}, \sigma^2_{p_j}, \sigma_{p_{ij}}$ = permanent environmental (co)variance components for traits i and j,

$\sigma^2_{e_i}, \sigma^2_{e_j}, \sigma_{e_{ij}}$ = residual (co)variance components for traits i and j,

I_1 = identity matrix for the permanent effects (order equal to the number of does with records),

I_2 = identity matrix for the residuals (of order equal to the number of records).

Estimates of variance and covariance components along with heritability and repeatability for lactation yields of MY, FY, PY and SCS were obtained using the Restricted Maximum Likelihood procedure in ASReml version 3 (Gilmour et al., 2009). Some animals had missing values of SCS. When this occurred, ASReml uses the genetic covariances between the traits to estimate breeding values for the missing values (Gilmour et al., 2009).

Results

Descriptive statistics are shown in Table 3.2. Mean FY was slightly greater than mean PY. The coefficients of variation for lactation length and milk production traits were high, reflecting the large phenotypic variation in all the traits.

Table 3.2. Descriptive statistics of milking traits of New Zealand dairy goats[1] kidding between 2004 and 2017.

Trait	N	Mean	SD[2]	Min	Max	CV[3]
Lactation length (days)	64604	226.4	89.6	60.0	700.0	40
Yields (up to 305 days)						
Milk (kg)	64604	727.9	331.1	30.0	2,262.7	45
Fat (kg)	64604	24.4	11.2	1.0	80.0	46
Protein (kg)	64604	22.2	10.0	1.0	69.0	45
SCS[4] (units)	61567	9.3	1.4	3.3	14.8	15

[1]N = number of records, [2]SD = raw standard deviation across herds, [3]CV = coefficient of variation, [4]SCS = calculated as average $\log_2$(somatic cell count).

Estimated effects of age, breed and general heterosis on milking traits are presented in Table 3.3 with the age group effects of nine-year-old does constrained to zero and the breed group effects of Saanen does constrained to zero. Three-year-old does produced the highest MY, FY and PY. Yearling does and then two- and three-year-old does tended to have the lowest SCS, which increased beyond three-year-olds with the age of the doe. Does of unknown breed produced the highest milk yield, while does in the ANT breed-group produced the lowest. For example, three-year-old does produced 281.3 kg milk more than the average milk production of the animals that were nine years or older. Similarly, purebred animals that were either Alpine, Nubian or Toggenburg breed produced on average, 32.1 kg less milk than purebred Saanen does.

Table 3.3. Effect (±standard error) of doe age relative to nine-year-old does, breed relative to Saanen, and heterosis for milking traits in New Zealand dairy goats kidding between 2004 and 2017.

Effect	Trait							
	Milk yield (kg)	SE	Fat yield (kg)	SE	Protein yield (kg)	SE	Somatic cell score[1]	SE
Age (years)								
1	20.47[h]	7.74	2.02[g]	0.27	1.41[h]	0.23	-1.79[h]	0.05
2	226.90[d]	7.64	9.01[c]	0.27	7.59[d]	0.23	-1.62[g]	0.05
3	281.30[a]	7.56	10.47[a]	0.26	9.23[a]	0.23	-1.30[f]	0.05
4	265.40[b]	7.51	9.70[b]	0.26	8.68[b]	0.22	-1.04[e]	0.05
5	231.90[c]	7.50	8.51[c]	0.26	7.64[c]	0.22	-0.81[d]	0.05
6	188.10[e]	7.58	6.95[d]	0.27	6.28[e]	0.23	-0.67[c]	0.05
7	142.20[f]	7.84	5.29[e]	0.27	4.88[f]	0.23	-0.42[b]	0.05
8	87.16[g]	8.45	3.49[f]	0.30	3.18[g]	0.25	-0.22[a]	0.05
Breed								
ANT[2]	-32.07[b]	8.52	-0.13	0.29	-0.63[b]	0.25	0.02	0.05
Unknown	44.32[a]	6.42	1.78	0.22	1.50[a]	0.19	-0.23	0.04
Heterosis	7.54*	4.15	0.17	0.14	0.26*	0.12	0.04*	0.02

[1]Somatic cell score = calculated as $\log_2$(somatic cell count), [2]ANT = breed group including Alpine, Nubian and Toggenburg animals, [a,b,c,d,e,f,g,h,I,j] Means with different superscript, within effect, are significantly different (P-value<0.05), * Significantly different to zero (P-value<0.05).

The variances (additive, permanent environment and residual), heritabilities and repeatabilities estimated with the multiple-trait animal model for MY, FY, PY, and SCS are presented in Table 3.4. Heritabilities were similar for the four traits studied as were estimates of repeatability values, falling within the ranges of 0.21-0.25 for heritability and 0.39-0.48 for repeatability.

Table 3.4. Estimates of additive, permanent environment and residual variances, heritability and repeatability and their corresponding standard errors (SE), for milking traits in New Zealand dairy goats kidding between 2004 and 2017.

Trait	Additive variance	Permanent environment variance	Residual variance	Heritability	Repeatability
Milk yield	12,215.70	7,886.44	28,784.20	0.25	0.41
	(526.77)	(425.60)	(198.20)	(0.01)	(0.00)
Fat yield	14.01	9.10	35.40	0.24	0.40
	(0.62)	(0.50)	(0.24)	(0.01)	(0.00)
Protein yield	10.19	6.16	25.84	0.24	0.39
	(0.45)	(0.36)	(0.18)	(0.01)	(0.00)
Somatic cell score	0.32	0.42	0.79	0.21	0.48
	(0.02)	(0.02)	(0.01)	(0.01)	(0.00)

Table 3.5 shows the genetic and phenotypic correlations between milk production traits and SCS. In general, correlations were high and positive for milk production traits while all phenotypic correlations with SCS were low and negative. The correlations between MY, FY and PY ranged between 0.90-0.96 for phenotypic effects and 0.81-0.93 for genetic effects.

Table 3.5. Estimates of genetic (below diagonal) and phenotypic (above diagonal) correlations and standard errors, among milking traits in New Zealand dairy goats kidding between 2004 and 2017.

Trait	Milk yield	Fat yield	Protein yield	Somatic cell score
Milk yield		0.90 (0.00)	0.96 (0.00)	-0.12 (0.01)
Fat yield	0.81 (0.01)		0.92 (0.00)	-0.16 (0.01)
Protein yield	0.93 (0.00)	0.93 (0.00)		-0.09 (0.01)
Somatic cell score	0.10 (0.04)	-0.01 (0.04)	0.10 (0.04)	

Discussion

The average production yields in the data set presented in this analysis are greater than those obtained from an earlier dataset representing British, Nubian, Saanen, Toggenburg and crossbred dairy goats in New Zealand (Singireddy et al., 1997). They are also greater than those reported in Alpine (456-648 kg MY, 14.7-22.7 kg FY and 12.5-19.9 kg PY in 231-250 DIM) and Saanen goats (512-676 kg MY, 15.7-21.8 kg FY and 13.6-19.9 kg PY in 240-250 DIM) in France (Boichard et al., 1989; Bélichon et al., 1998) or Maltese (532.3 kg MY in 230 DIM) and Jonica goats (281 kg MY, 10.5 kg FY and 9.8 kg PY in 240 DIM) in Italy (Delfino et al., 2011; Selvaggi and Dario, 2015). However, the New Zealand production figures were lower than those reported by Valencia et al. (2007) in Saanen goats in Mexico (800 kg MY in 285 DIM), and for US dairy goats (García-Peniche et al., 2012; Castañeda-Bustos et al., 2014) for which reported values ranged from 1026-1,043 kg MY, 37.1-38 kg FY and 30.5-32.0 kg PY from 305 day lactations. These differences in mean performance from the present study compared with those obtained in other studies may be attributable to a number of environmental factors. For example, MY was reported to be significantly influenced by the age of the goat, season and year of kidding in Saanen goats in Mexico (Valencia et al., 2002), Black Bengal goats in Bangladesh (Mahal et al., 2014), Alpine and Saanen breeds in Brazil (Brito et al., 2011) and Damascus goats in Cyprus (Mavrogenis et al., 1984; Mavrogenis et al., 1989). Also, as in the current study, season of kidding, parity and herd-year was reported to have a significant effect (P-value<0.01) on FY and PY in Saanen goats in Mexico (Torres-Vázquez et al., 2009). The difference between mean milk production values have been attributed to the different breeds and seasonality between regions of the world (Montaldo et al., 2010), climate and nutritional quality of food (Selvaggi and Dario, 2015) as well as other management factors (Castañeda-Bustos et al., 2014).

Average SCS values reported in other studies were calculated using a range of logarithmic scales, including $\log_2$, $\log_{10}$ or natural logarithms. In the current study, SCS was calculated using $\log_2$ transformation with an average SCS of 9.3 ± 1.4. This value is within the range of somatic cell counts previously reported for dairy goat populations worldwide (Bergonier et al., 2003; Paape et al., 2007; Apodaca-Sarabia et al., 2009; Rupp et al., 2011; Maroteau et

al., 2014), suggesting the 'health' of does in New Zealand is similar to the rest of the world. It is known that SCS are greater in dairy goats relative to dairy cattle and sheep (Rupp et al., 2011). It has been proposed that the higher measured SCS in goats may be due to anomalous measurements because of the presence of larger anucleated cytoplasmic particles in goat milk, and may be misinterpreted through the assay system as leukocyte concentration which of course lacks any pathological significance (Dulin et al., 1982; Rota et al., 1993; Paape et al., 2001). However, this cannot be the reason for the high SCS in this study as SCC was measured using FOSS technology (Fossmatic) which utilises a DNA stain to detect somatic cells, thus, anucleated particles are not included in the count.

Age had a significant effect (P-value<0.0001) on yield traits and SCS. Does in their first year and does which were eight- and nine-years-of-age had the lowest MY, FY and PY compared to does of three- and four-years-of-age which produced the greatest yields (Table 3.3). These differences were similar to those reported by Singireddy et al. (1997) and Morris et al. (1997) who also reported on the milk production of dairy goats in New Zealand. They showed that four-year-old dairy goats produced 56% (Singireddy et al., 1997), and 94% (Morris et al., 1997) more milk yield than yearlings. Like lactation yields, average SCS varied by age group. The lowest SCS was in yearling goats and increased as does aged. This is consistent with previous reports of dairy goats in New Zealand (Apodaca-Sarabia et al., 2009) and Poland (Bagnicka et al., 2016), suggesting that the health status of the mammary gland is best in primiparous goats (Barrón-Bravo et al., 2013). The health of the udder worsens in older animals due to changes in mammary physiology with succeeding lactations, resulting in increased susceptibility to infection (Rota et al., 1993; Anniss and McDougall, 2000).

Breed had a significant effect (P-value<0.0001) on milking traits, with does of unknown breed producing the greatest MY, FY and PY. As the majority of dairy goats in New Zealand are Saanen (Scholtens et al., 2017), and considering the effect of breed on milk production, it is likely that these high producing "unknown" animals are of Saanen origin. Despite relatively large breed effects on MY, there were no significant differences between SCS for ANT and Saanen does. This is consistent with the values reported by Apodaca-Sarabia et al.

(2009), but contrasts those of Paape et al. (2007) who observed that Saanen animals have significantly lower SCS than Alpine or Toggenburg does.

In this study all heterosis effects were positive, however, due to the limited number of animals across breeds, only general heterosis was calculated. General heterosis had significant effects on milk production traits, indicating that first cross does will produce 7.54 kg more milk compared to the average of the population. Two other studies on New Zealand dairy goats also calculated general heterosis and reported heterosis effects of +52 kg milk for Saanen, Nubian, British, Toggenburg and crossbred goats (Singireddy et al., 1997) and +0.072 for SCS in mixed-breed dairy goats (Apodaca-Sarabia et al., 2009), both of greater magnitude than the heterosis effects found in this study. This could suggest the New Zealand dairy goat population has become more inbred since these studies were undertaken and the genetic diversity has either been lost or become more introgressed within the population.

Heritability estimates calculated in the current study for MY, FY, PY and SCS were within the range of those published values from different goat populations (0.10-0.45, 0.19-0.40, 0.04-0.38 and 0.09-0.25, respectively) (summarised in Table 2.3). The heritability values estimated here suggest that there is genetic variation underlying these traits in the study population and the consistency of the magnitude of heritabilities of this study with that of other studies gives confidence in the values estimated.

Estimates of repeatability for the traits studied were generally twice the magnitude of heritability estimates. Repeatabilities of MY, FY, PY and SCS in this study were close to the average values reported for dairy goats in the other studies (Table 2.3). Of course the estimates of heritability and repeatability will differ due to the breed and population, structure of the data, management conditions, estimation errors, association with sample size, and estimation methodology used (Moioli et al., 2007).

The phenotypic correlations between all three milk traits were greater than those previously reported (Table 2.4). The sole value reported for the phenotypic correlation of SCS with milk production traits was a strong positive correlation between SCS and MY (0.59;

Bagnicka et al., 2016), much greater than the value of -0.12 obtained in the current study. This difference may be attributable to estimation errors because of the size of the dataset (4,417 records) compared with the current study (64,591 records).

All genetic correlations estimated for milk traits in this study were positive (Table 3.4). The genetic correlations between MY and FY and between MY and PY were similar to the range of values obtained from Alpine and Saanen goats in France (Table 2.4). Until recently, information of genetic correlations with SCS in dairy goats was limited. The genetic correlations between SCS and MY and between SCS and FY were all within the range of values estimated in Alpine and Saanen dairy goats in France (Table 2.4).

The high and positive genetic correlations between MY and FY and PY traits suggest that selection for MY alone should also result in an increase of both FY and PY. The slight positive correlations between SCS and MY and PY indicate that the quality of milk could be decreased by an increase in SCS if selection is based on high-yielding animals, however these correlations are relatively low which suggests that there would be minimal changes in the SCS based on a selection index targeting antagonistic traits. However, selection tools such as a selection index can be used to constrain changes in SCS while still allowing genetic improvement of milk production.

Overall, the genetic parameters estimated for MY, FY, PY and SCS in mixed-breed dairy goats in New Zealand are consistent with the values reported by others. Results from this study provide the first estimates of heritability for FY and PY and update estimates for MY and SCS for dairy goats in New Zealand, using a larger, structured data set fitted to a multiple-trait animal repeatability model. Once there are sufficient numbers of purebred animals in the population it is recommended that genetic parameters are estimated for each breed, to gain further understanding of genetic parameters of New Zealand dairy goats.

Using a multiple-trait animal model, this study produced variance and covariance components and genetic parameters required for genetic evaluation of lactation yields of milk, fat and protein, and SCS for dairy goats in New Zealand. Genetic evaluation will produce estimated breeding values that can be combined with corresponding economic

values and used in a selection index. Farmers can use the selection index to rank animals to be selected as parents for the next generation.

Genetic evaluation using a multiple-trait model allows breeding values to be estimated for all animals, even if an animal did not have phenotypic records for a trait. Therefore, if a selection index is constructed, all animals would be included in the evaluation, albeit with varying reliabilities. The heritabilities and genetic correlations suggest that milk income per animal can be improved in this mixed-breed population through selection for MY, FY and PY, while ameliorating clinical and subclinical mastitis by including a breeding value for SCS in an economic index (Bagnicka et al., 2016).

The results from this study suggest that there is adequate genetic variation for MY, FY, PY and SCS of New Zealand dairy goats to allow genetic change.

Conclusions

Positive genetic correlations between lactation yields suggest favourable correlated responses for selection on any combination of MY, FY and PY. Considerable variation exists within mixed-breed dairy goats farmed in New Zealand, and goat milk production can be increased through selection for these traits. These estimates of heritability, repeatability and correlations can be used for estimating breeding values for these traits and used in a selection index to enable the selection of animals with superior genetic merit to improve the quantity and quality of milk produced from successive generations of New Zealand dairy goats.

Chapter 4

Estimates of genetic parameters for lactation curves for milk, fat, protein and somatic cell score in New Zealand dairy goats

This Chapter has been published in part elsewhere. It has been reformatted and presented here with permission:

Scholtens MR, Lopez-Villalobos N, Garrick DJ, Blair HT, Lehnert K, Snell RG. 2019. Estimates of genetic parameters for lactation curves for milk, fat, protein and somatic cell score in New Zealand dairy goats. New Zealand Journal of Animal Science and Production 79: 177-182.

Abstract

The aim of this study was to estimate genetic parameters of daily yields of milk (MY), fat (FY), protein (PY) and somatic cell score (SCS) of New Zealand dairy goats. The analysis used 113,895 herd-test records from 14,187 does, kidding between 2010 and 2016, distributed in 11 herds and representing 377 sires. Estimates of genetic and residual (co)variances, heritabilities (h^2) and repeatabilities were obtained using a random regression test-day animal model. The model included the fixed effects of contemporary group (herd-test-day), age of the doe (in years) and as covariates, deviation from median kidding date for a given herd and year, proportion of genes from Alpine, Nubian, Toggenburg and "unknown and other" breeds (Saanen was used as a base breed), heterosis, and days in milk (DIM) modelled as a third-order orthogonal polynomial. Random effects included additive animal genetic and doe-lactation permanent environment effects modelled using third-order orthogonal polynomials. Estimates of h^2 and repeatabilities at different stages of lactation ranged from 0.13 to 0.35 and 0.39 to 0.81, respectively. These results provide an opportunity to estimate breeding values at any and every day of lactation, and to change the shape of the lactation curve by selection.

Introduction

The production efficiency of the dairy goat industry needs to improve to continue expanding and remain competitive with other dairy industries. Production efficiency can be improved by increasing milk production or maintaining persistent yields across the lactation period. Genetic improvement is the most attractive approach to promote permanent gains in livestock species. Milk solids (fat + protein + lactose yields) is among traits with highest economic relevance for the industry and maximising milk solid production is a key goal of dairy farmers. Characterising variation in milk-production traits at specific days during lactation is important for understanding the genetic associations between traits at different stages of lactation and would provide parameters required for estimating breeding values for milk traits at different stages of lactation.

Milk production during the lactation is an example of a longitudinal trait, characterised by repeated measures in the same individual over time. Random regression models (RRM) have become commonly adopted for the genetic analysis of longitudinal data and are currently used for the national genetic evaluation of production traits of dairy cattle in Australia, Belgium, Canada, Czech Republic, Denmark, Finland, Germany, Ireland, The Netherlands, New Zealand and Sweden (Interbull, 2017). Under RRM, records on the sample day are considered directly in the analysis, therefore, RRM can account more precisely for environmental factors that could affect animals differently during lactation (Schaeffer and Dekkers, 1994). In one class of RRM, a fixed curve for the population is calculated and individual curves are fitted as deviations from the population curve. Any RRM provides the opportunity to use a function to model any or all of the fixed and random curves.

The average shape of a lactation curve can inform the farmer of the predicted level of production over the lactation period, whereas the shape of lactation curves for individual animals can provide insight into the health status of the animal during the lactation process (i.e., health of the mammary gland, energy supply/deficit) and the environmental effects affecting its milk production (Hossein-Zadeh, 2016). Information about milk production levels and the characteristics of the lactation curves allow evaluation of the production performance and subsequent implementation of improvement strategies, i.e., feeding, breeding and economic management. Knowledge of heritabilities (h^2) at each test-day and covariances among test days across the lactation period would provide the opportunity to estimate breeding values at any and every point in time during lactation.

The objective of this study was to estimate genetic parameters for daily yields of MY, FY, PY and SCS across the lactation period of New Zealand dairy goats using a random regression test-day model.

Materials and methods

The original data set was provided by Livestock Improvement Corporation and included 304,648 herd-test records across seven lactations from 48,113 does kidding up to parity 14

between 2010 and 2016 and from 55 herds across the North Island of New Zealand. The pedigree of these lactating does included 422 sires and 11,803 dams. Breed composition of each doe was calculated from pedigree proportions of Alpine, Nubian, Saanen, Toggenburg and "unknown and other" breeds. There were a total of 797 herd-test-day contemporary groups. Somatic cell score was calculated as $SCS=\log_2$(somatic cell count) at each herd-test.

Herds comprising does with <80% of dams known or with contemporary groups with <10 herd-tested does were removed (44 herds and 604 herd-test-day groups). After cleaning, the data set contained 113,895 herd-test records across seven lactations, representing 14 parities and included 14,187 does distributed in 11 herds. The does were the offspring of 377 sires and 8,043 dams and the known pedigree spanned up to nine generations. Many of the does were crossbred and there were very few purebred animals of Alpine (1), Nubian (3) or Toggenburg (40) breeds and, therefore, the covariates reflecting the proportion of these breeds were summed in a single combined-breed group hereafter called ANT. The remaining does included 1,358 Saanen, 12,766 crossbred and 19 animals of unknown and other breeds. There was a total of 229 herd-test-day contemporary groups, mostly consisting of more than one breed or cross. The structure of the dataset is provided in more detail in Table 4.1.

Table 4.1. Summary characteristics of dataset comprising test-day yields of dairy goats in New Zealand.

Breed	Herds	Animals	Herd-test records	Animals with known sires
ANT[1]	3	44	184	1
Saanen	10	1,358	8,757	74
Crossbred	11	12,766	104,808	368
Unknown + other	6	19	136	3
Total	-	14,187	113,895	446

[1]ANT = combined breed group including Alpine, Nubian and Toggenburg animals.

Estimates of variance and covariance components for test-day yields of MY, FY, PY and SCS were obtained using the restricted maximum likelihood procedure in ASReml version 3 (Gilmour et al., 2009) to fit a single-trait random regression test-day model. The model included the fixed effects of contemporary group (herd-test-day) and age of the doe (in years). Covariates in the model included deviation from median kidding date, proportion of genes from ANT and "unknown and other" breeds (Saanen was used as a base breed) and general heterosis. Days in milk (DIM) was modelled as a third-order orthogonal polynomial. Random effects included additive animal genetic and doe-lactation permanent environment, both modelled using third-order orthogonal polynomials.

General heterosis was calculated as $1-\sum_{j=1}^{f} p_j^2$ where p_j is the proportion of each of the f breeds (Gregory and Cundiff, 1980; Lamberson et al., 1993). General heterosis assumes that first-cross heterosis is the same for all breed combinations. Residual variances were assumed to be heterogenous across six stages of lactation, with different residual variances for 0 to 50, 51 to 100, 101 to 150, 151 to 200, 201 to 250 and 251 to 270 DIM. The genetic (co)variances across all DIM were estimated as:

$$\delta = \phi' K \phi$$

in which δ is the variance or (co)variance matrix for the traits, ϕ is the matrix of orthogonal polynomials for each DIM, and K is the matrix of the additive genetic (co)variance matrix of random orthogonal polynomial coefficients.

Initially, six RRMs were tested in order to identify the one which best fitted the production records. Models ranged in the orders of the orthogonal polynomial used to describe the additive animal genetic and doe-lactation permanent environment effects. The model with best fit had polynomials of 3rd order for the fixed curve of the population and for both the additive animal genetic and doe-lactation permanent environment effects.

Results

Descriptive statistics for MY, FY, PY and SCS are shown in Table 4.2. Primiparous does produced the lowest yields compared to does in later parities. Parity had a significant (P-value<0.001) effect on test-day yields of MY, FY, PY and SCS over that lactation period. The coefficients of variation on test-day yields were high, reflecting considerable phenotypic variation in the shape of the lactation curve.

Table 4.2. Descriptive statistics of milking traits of New Zealand dairy goats kidding between 2010 and 2016.

		Milk yield (kg/day)			Fat yield (g/day)			Protein yield (g/day)			Somatic cell score (units/day)		
	N	Mean	SD[1]	CV[2]	Mean	SD[1]	CV[2]	Mean	SD[1]	CV[2]	Mean	SD[1]	CV[2]
Population	113,895	3.4	1.3	37	114.4	44.6	39	106.5	37.2	35	9.1	1.7	19
Parity													
1	31,728	2.6	0.9	35	90.1	32.8	36	84.1	27.8	33	8.7	1.8	21
2	29,305	3.6	1.1	32	122.3	42.5	35	112.8	34.1	30	8.9	1.7	19
3	19,329	3.9	1.3	33	130.5	46.4	36	121.4	37.6	31	9.2	1.7	18

[1]SD = standard deviation across herds, [2]CV = coefficient of variation (%).

Orthogonal polynomials of order 3 were used to estimate lactation curves for daily yields of MY, FY, PY and SCS for does of different parity (Figure 4.1). The shape of lactation curves were the same for each parity for each trait. Primiparous does had the lowest yields for all traits while third-parity does had the highest yields. Peak yields were around day 95 for MY, day 1 for FY, days 12 to 106 for PY and day 270 for SCS.

Figure 4.2 graphically illustrates the additive genetic (σ^2_a), permanent environment (σ^2_{pe}), residual (σ^2_e) and phenotypic (σ^2_p) variances estimated using a single-trait RRM for test-day yields of MY, FY, PY and SCS during the lactation period. The σ^2_{pe}, and σ^2_p were greatest at the beginning of lactation for all traits while the trajectories of σ^2_a and σ^2_e varied among traits.

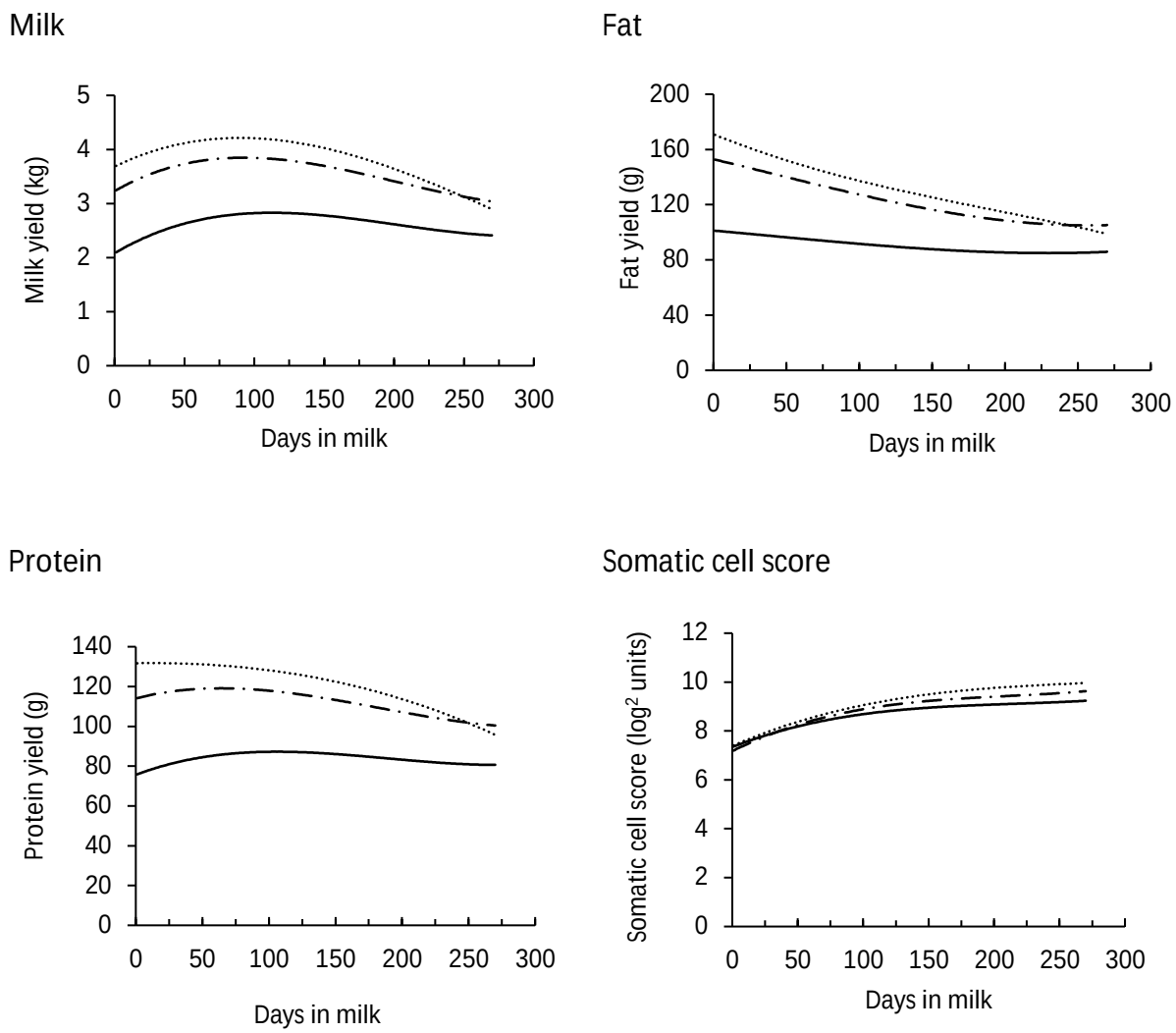

Figure 4.1. Lactation curves of daily yields of milk, fat and protein and somatic cell score during the 270-day lactation for does in parity 1 (–), 2 (–•–) and 3 (•••) modelled with orthogonal polynomials of order 3.

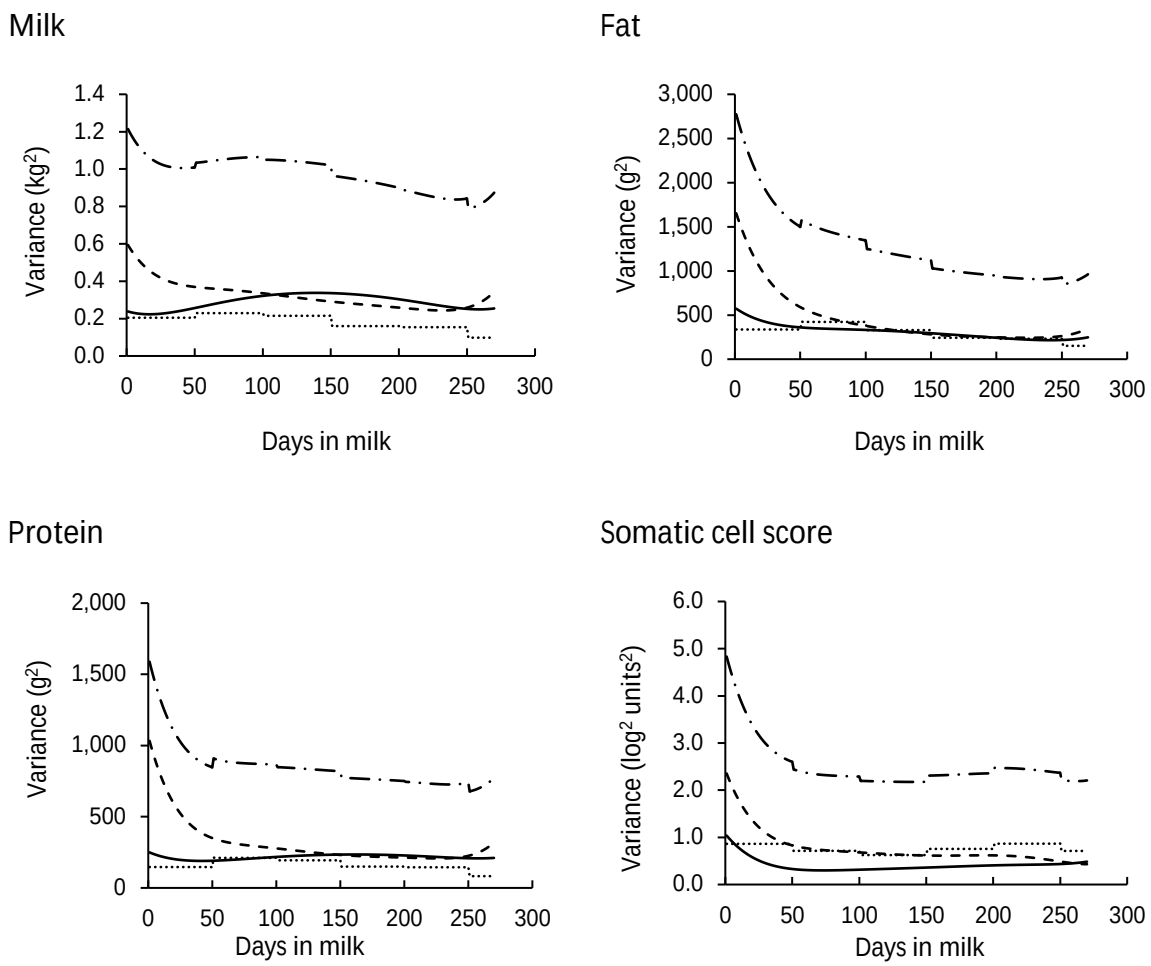

Figure 4.2. Estimates of additive genetic (–), permanent environment (---), residual (•••) and phenotypic variance (–•–) of test-day yields of milk, fat, protein and somatic cell score during the 270-day lactation in New Zealand dairy goats.

Estimated h^2 by day of lactation for MY, FY, PY and SCS are in Figure 4.3. The h^2 ranged from 0.20 to 0.35 for MY, 0.21 to 0.28 for FY, 0.16 to 0.31 for PY and 0.13 to 0.22 for SCS. Average repeatability estimates were 0.63, 0.57, 0.61 and 0.45 for MY, FY, PY and SCS, respectively.

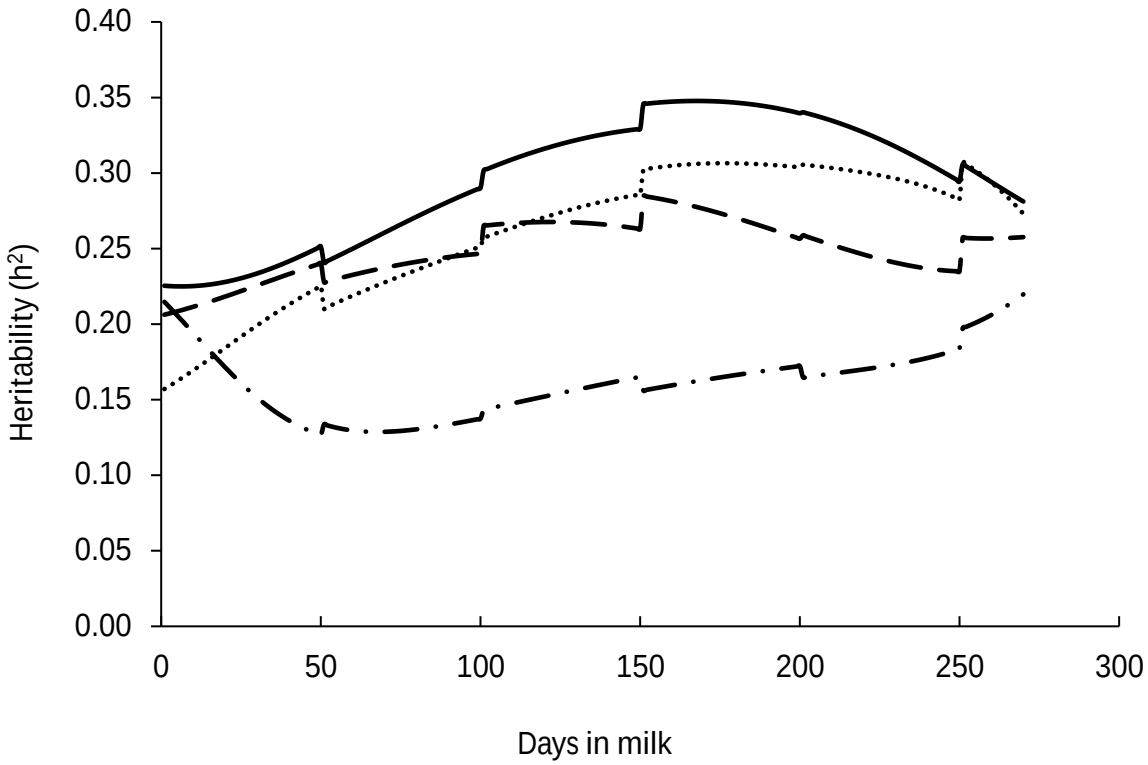

Figure 4.3. Estimates of heritability for test-day yields of milk (–), fat (---), protein (•••) and somatic cell score (–•–) during the 270-day lactation in New Zealand dairy goats.

Discussion

The mean and standard deviation for daily MY, FY and PY were greater than those reported in literature: MY in crossbred goats in Brazil (2.61±0.71 kg/day; Lobo et al., 2017) and Murciano-Granadina goats in Spain (1.93±1.01 and 2.18±1.01; Menéndez-Buxadera et al., 2010), and FY and PY in Murciano-Granadina goats (0.098 to 0.109±0.05 g/day for FY and 0.068 to 0.072±0.03 g/day for PY; Menéndez-Buxadera et al., 2010). The lower MY could be due to climatic effects, especially for semi-arid climates in Brazil, which are known to negatively affect milk production through reduced feed intake and increased maintenance requirements needed for homeothermy (sweating, panting) (Salama et al., 2014). In addition, the lower FY and PY in the Spanish study could be due to parity effect, as they only analysed does in first and second-parity, which are known to have lower milk production compared to older does (Amin et al., 2017; Lobo et al., 2017). The values obtained in this

study for SCS were similar to 9.35±1.69 units/day, previously reported in goats in New Zealand (Apodaca-Sarabia et al., 2009).

A typical lactation curve of MY for French dairy goats has been reported to comprise of a rapid increase at the start of lactation, peak around 50 DIM, then stable for 50 days and a gradual decrease through to the end of lactation (Arnal et al., 2018). A similar pattern was also observed in the current study, but the peak was not until 90 to 100 DIM. For primiparous does, the peak yield is typically earlier and the lactation curve is generally flatter than multiparous does (Fernandez et al., 2002). The early peak was not observed for primiparous does in this study, however, the curve was flatter than for does in second and third parity. The lactation curves estimated for does of first, second and third parity followed a similar trend for each of the traits. Does in third parity produced the greatest MY, FY and PY throughout the lactation, until 250 DIM. Overall, second- and third-parity does produced significantly greater daily yields than primiparous does. This is in agreement with findings of Amin (2017) and Lobo et al. (2017) who reported that yields increased as the age of doe increased until about fourth parity, and then decreased. This parity effect is common for milk-production animals and is suggested to be due to younger does having lower body weight, body condition score and body reserves than older does, and consequently, the body and mammary gland of young animals are still developing during the first lactation. It is also common that the level of SCS is lowest at the beginning of lactation and increases during the lactation period (Lobo et al., 2017) and that SCS is lowest in primiparous animals and increases as the number of lactations increases (Amin, 2017; Lobo et al., 2017).

Orthogonal polynomials of third order were applied for modelling the average production curve of the population (fixed effect) and for modelling production curves of the random effects (additive animal genetic and doe-lactation permanent environment). Although models with more parameters are more accurate (Brito et al., 2017b), this generates greater computational demand and can lead to problems of convergence and estimation, especially when using a RRM for genetic evaluations on large datasets. Therefore, it is important to use less-parameterised models without losing quality of fit. Considering heterogeneity of

residual variances rather than homogenous residual variance can improve the modelling of the random effects (Brito et al., 2017b). Therefore, if performing genetic evaluation on a large data set, we recommend using a RRM with orthogonal polynomials of order 3 for fixed and random effects and assume heterogenous residual variances.

For MY, the highest value for σ_a^2 was observed just after peak production (124 DIM). This trajectory was similar to that seen in dairy goats in Brazil using a multiple-trait analysis, but their peak lactation was around 60-90 days (Irano et al., 2015) and 40 days (Brito et al., 2017b). In contrast, Silva et al. (2013) found higher values of σ_a^2 at the end of the lactation curve using a single-trait animal model for random regression.

The σ_a^2 declined from the beginning of the lactation for FY, PY and SCS until days 40-50, then for SCS the variance increased for the rest of the lactation, for PY this increased for 100 days before slightly decreasing and for FY the variance plateaued for 100 days before slightly decreasing along the remaining days in milk. Despite there being no literature on dairy goats, this decrease in genetic variability during the lactation period is commonly observed in dairy cattle (Biassus et al., 2011).

In this study, σ_{pe}^2 followed the same pattern for MY, FY and PY, decreasing as the number of lactation days increased until 250 DIM before slightly increasing for the last 20 DIM. These inflated variances at the start of the lactation period were also reported by Thepparat et al. (2015) using a single-trait RRM, and Oliveira et al. (2016) using a multiple-trait RRM with different functions to describe each trait. In contrast, Brito et al. (2017b) reported that σ_{pe}^2 and σ_p^2 were lower in early lactation and increased at the end of lactation compared with the other lactation stages. They proposed that this was expected as there was a greater reduction in the number of records at the end of lactation, however, this reduction in records also occurred in this study. Moreover, more variability during the beginning and end of lactation would be expected, as non-genetic factors, such as management, tend to influence milk production more expressively during this period. The σ_p^2 variance generally decreased as the number of lactation days increased, but also followed a similar trend to

that of residual variance, which peaked during the 2nd and 3rd stage of lactation (days 51-100 and 101-150) for MY, FY and PY.

The h^2 estimates were greatest from the middle to late lactation. This is expected and is in agreement with reports by Sarmento et al. (2008) and Oliveira et al. (2016), but are in contrast to the report by Menéndez-Buxadera et al. (2010) who reported h^2 values decreased throughout the lactation period and Silva et al. (2013) and Brito et al. (2017b) who found that the h^2 values increased in the final third of the lactation. The lower estimates for MY, FY and PY during early lactation and late lactation could be due to the greater influence of environmental effects at these stages, while production in mid-lactation is more influenced by the genetic and permanent effects. For example, most does are in an advanced stage of pregnancy during late lactation, which can explain the decreased h^2 at that latter stage.

Despite the differing patterns of h^2 in the literature, the h^2 estimates obtained in this study were similar, with the range of values published for daily MY of 0.12 to 0.66 (Andonov et al., 2007; Sarmento et al., 2008; Zumbach et al., 2008; Menéndez-Buxadera et al., 2010; Irano et al., 2015; Thepparat et al., 2015; Oliveira et al., 2016; Brito et al., 2017b), daily FY of 0.12 to 0.25 (Menéndez-Buxadera et al., 2010), daily PY of 0.10 to 0.18 (Menéndez-Buxadera et al., 2010) and daily SCS of 0.12 to 0.25 (Apodaca et al., 2009). Our results suggest that there is enough genetic variability to make genetic progress for test-day MY, FY, PY and SCS in dairy goats in New Zealand and it would be possible to modify the shape of the curve by selective breeding. In order to maximise milk production, the estimation of breeding values during early lactation would enable identification and selection of animals with low genetic merit for culling and of high genetic merit for breeding decisions. However, the lower σ_a^2 and h^2 near the beginning and end of the lactation would result in lower genetic response when selecting for increased yield in just the first or last part of lactation. Instead, the point with the greatest heritability would be the most applicable stage for practicing selection. The estimate of the h^2 to select for increased persistency between 150 and 250 days in milk in this goat population, was between 0.23 and 0.35 for MY, FY and PY.

At farm level, the biological interpretation of the parameters estimated in this study can contribute to the improvement of goat milk production throughout the lactation period. The estimation and interpretation of test-day yields and h^2 suggest that selection can help to raise milk production and persistency in dairy goat herds in New Zealand. This will be of great relevance to the implementation of genetic evaluations in dairy goats.

Conclusion

Parity has a significant effect on milk production in New Zealand dairy goats. Most lactation curves of does in second and third parity had similar shapes, while does in first parity tended to have lower values throughout the lactation. The results showed that the amount of variation changes during the lactation and the lactation curves for production varies with parity. The use of a RRM for genetic evaluation of dairy goats may allow for selection to alter the shape of the lactation curve. Estimates of h^2 obtained throughout the lactation were moderate, indicating there is enough genetic variability to make genetic progress for test-day yields of MY, FY, PY and SCS in dairy goats.

Chapter 5

Heritability of longevity in New Zealand dairy goats

This Chapter has been published in part elsewhere. It has been reformatted and presented here with permission:

Scholtens MR, Lopez-Villalobos N, Garrick DJ, Blair HT. 2018. Heritability of longevity in New Zealand dairy goats. New Zealand Journal of Animal Science and Production 78: 11-15.

Abstract

Longevity, defined for a doe as the age when it leaves the milking herd, is a trait of economic importance in dairy goat farming. Actual longevity (AL) is defined as the number of days from birth to when the animal leaves the herd, whereas functional longevity (FL) is defined as AL adjusted for first lactation energy-corrected milk yield (ECMY). This study reports the heritability (h^2) for AL and FL in New Zealand dairy goats. Records of longevity from 12,108 does born between 1993 and 2011 were analysed with a model that included the fixed effects of herd-year (does born in the same herd and year) and covariates for the proportion of Alpine, Nubian, Toggenburg and heterosis, and the random effect of animal. The model for FL was the same as AL but included ECMY as a covariate. Average AL was 1,891 (SD=832) days. Estimates of h^2 were 0.07 for AL and FL. The estimated regression coefficient for ECMY in AL was 0.56 days/kg. There were significant differences in longevity among herds, indicating that management and feeding are important factors affecting longevity. Further research is required to estimate genetic correlations with economically important traits.

Introduction

Increased longevity of multi-parous animals such as dairy goats, enables an older age structure and consequently greater milk production by the herd, it also reduces replacement costs (Serradilla et al., 1997; Castañeda-Bustos et al., 2017). Until recently, improving production of milk, fat and protein per doe have been the main breeding objective traits in genetic improvement programs of dairy goats (Desire et al., 2017; Valencia-Posadas et al., 2017). Placing too much emphasis on production, whilst neglecting other traits, may result in unexpected and undesirable consequences on the health and fertility of animals, which decrease longevity (Oltenacu and Broom, 2010). Either direct or indirect evaluation of longevity, if used in selection, will increase the overall economic efficiency of the dairy goat industry.

In dairy ruminants, AL takes into account all reasons the animal was removed from the herd, while FL takes into account all reasons except milk productivity. Adjusting for milk production results in a longevity value that reflects the animal's ability to avoid

involuntary culling due to health and reproductive challenges (Castañeda-Bustos et al., 2014).

Longevity, or similar traits, such as stayability or survival, are now typically included in the breeding objective for most dairy cattle breeding programs (Miglior et al., 2005), but longevity is not yet included in dairy goat breeding programs (Castañeda-Bustos et al., 2014; Castañeda-Bustos et al., 2017; Valencia-Posadas et al., 2017; Palhière et al., 2018). Only three papers report longevity of dairy goats in New Zealand (Wheeler et al., 2013; Wheeler et al., 2014; Gautam et al., 2017). The studies by Wheeler et al. (2013; 2014), provide estimates of stayability and AL in a single herd. The study by Gautam et al. (2017) is a retrospective study that analysed risk factors for the animals leaving the herd. None of these provided estimates of heritability for FL. The objective of this study was to estimate the heritability of AL and FL of dairy goats in New Zealand.

Materials and methods

Data

The dataset used in this study was obtained from Livestock Improvement Corporation and comprised 112,009 dairy goats of Alpine, Nubian, Saanen, and Toggenburg breeds as well as crossbred animals, born between 1973 and 2016. The goats were from 164 herds located throughout North Island, descending from 26,720 dams and 1,284 sires.

Individual birth dates of animals were not available for all animals; some farmers allocated a single birth date for a group of animals. Contemporary groups that had these single birth dates for a group of animals were removed from the analysis. Pedigree information was incomplete; of the does, only 18% had known sires and 44% had known dams. Therefore, only records from those few better-recorded farms were used to calculate longevity. The farms whose data were used were those that had more than 70% of does with known sires and with more than 15 does born in a specific year.

Animals were removed from the analysis if they were born before 1993 or after 2011. These dates were chosen because an exploratory analysis showed that contemporary groups (animals born in the same herd and year) typically comprised more than 15

animals. Those does born after 2011 would not yet have had an uncensored opportunity to express survival.

The number of days from birth to when the animal left the herd defined AL whereas FL was defined as longevity adjusted by the ECMY in the first lactation, calculated as, ECMY = 0.327×MY + 12.86×FY + 7.65×PY, where MY, FY and PY were estimated first lactation yields of milk, fat and protein, respectively (Flores et al., 2009).

Statistical analysis

Descriptive statistics were obtained using the MEAN procedure of Statistical Analysis System version 9.4 (SAS Institute Inc., Cary, NC, USA). The Kaplan-Meier survival curve was obtained using the LIFETEST procedure of SAS which is a nonparametric maximum likelihood estimate of the survivor function (Kaplan and Meier 1958).

Contemporary groups were defined as comprising does born in the same herd and year. The original dataset included 112,009 animals in 2,058 groups and 12,108 animals in 123 groups after editing.

Analyses of AL and FL were performed using ASReml software (Gilmour et al., 2009) fitting a mixed linear animal model. The model included the fixed effects of herd-year (contemporary group), and covariates for proportion of Alpine, Nubian, Toggenburg and general heterosis, and a random animal effect. General heterosis, instead of specific two-breed heterosis, was calculated because the number of crossbred animals was insufficient to determine heterosis for each of the breed combinations. General heterosis assumes that first-cross heterosis is the same for all breed combinations (Olfati et al., 2011).

Analysis for FL was the same as AL, but including ECMY as a covariate. Phenotypic variance was the sum of animal and residual variances, genetic variance was the animal variance, and h^2 was calculated as the proportion of genetic variance with respect to the phenotypic variance. To explore the level of variation caused by the herd-year effect, the statistical model was run a second time, but with herd-year considered as a random effect rather than fixed effect.

Results

The main studies on different measures of longevity of dairy goats are presented in Table 5.1. The numbers of goats in these studies ranged from 4,910 to 1,137,793 animals and included the same breeds as in this study. The final dataset included 12,108 does from eight herds, representing descendants of 407 sires.

Table 5.1. Longevity traits studied in dairy goats around the world.

Country	Trait	Study
New Zealand	Survivability	Wheeler et al. (2013)
	Longevity	Wheeler et al. (2014)
	Risk factors associated with the length of productive life	Gautam et al. (2017)
Mexico	Stayability	Pérez-Razo et al. (2004)
	Productive life	Torrero (2010)
France	Productive life	Palhière et al. (2018)
United States	Productive life	Valencia-Posadas et al. (2010)
	Functional productive life	Castañeda-Bustos et al. (2014)
	Functional stayability	Castañeda-Bustos et al. (2017)
		Valencia-Posadas et al. (2017)

Descriptive statistics for first lactation MY, FY, PY, ECMY and AL are in Table 5.2. There was a 6 kg difference between first-lactation milk yield and ECMY for the same lactation and a reduction in the maximum milk yield from 1,538 to 1,510 kg, respectively. Average longevity of does born between 1993 and 2011 was 1,891±832 days.

Table 5.2. Descriptive statistics for first-lactation milk production and longevity of New Zealand dairy goats, born between 1993 and 2011.

Trait	N	Mean	SD[1]	Min	Max	CV[2]
First lactation yields (kg)						
Milk	12,108	502.9	223.2	30.0	1,538.1	44
Fat	12,108	16.9	7.6	1.1	54.3	45
Protein	12,108	15.1	6.5	1.0	43.0	43
Energy-corrected milk	12,108	496.9	216.7	31.4	1,510.0	44
Actual longevity (days)	12,108	1,891	832	400	6,551	44

[1]SD = standard deviation, [2]CV = coefficient of variation (%).

Figure 5.1 shows the longevity of dairy goats born in the eight herds. There was large variation in mean longevity among herds, from 1,638 to 2,088 days. When herd-year was included as a random effect, it was found that herd-year effect explained 35% of the total variation.

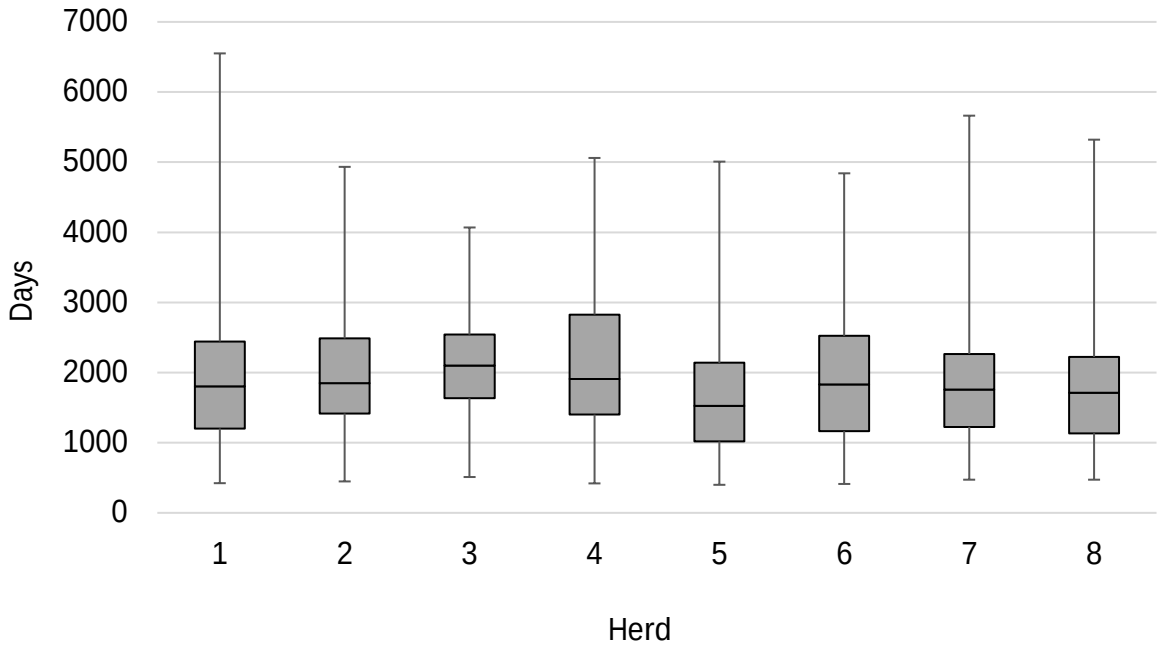

Figure 5.1. Boxplot of the longevity of dairy goats born between 1993 and 2011, in eight herds throughout the North Island of New Zealand. Number of does in each herd were; herd 1=2,721, herd 2=1,835, herd 3=611, herd 4=1,610, herd 5=1,327, herd 6=1,208, herd 7=2,117 and herd 8=679 does.

Figure 5.2 shows a Kaplan-Meier survival curve that included the longevity of 12,108 does. After 1,000 days, 85% of the goats remained in the herd and by 2,000 days, only 40% of the animals remained in the herd.

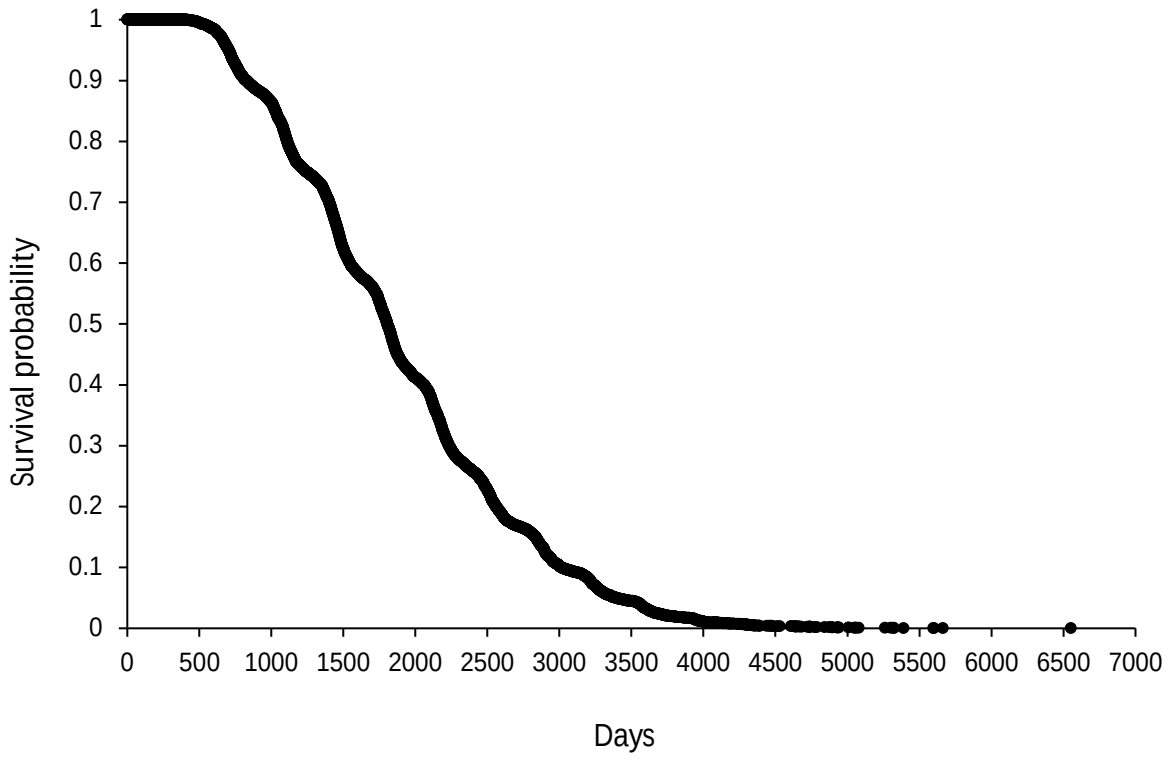

Figure 5.2. Kaplan-Meier survival curve for longevity of 12,108 New Zealand dairy goats born between 1993 and 2011.

The estimates of breed and heterosis effects were not significantly different from zero. The estimate of the regression coefficient of AL on ECMY was 0.56 days/kg ECMY (P-value<0.0001). Estimates of variance components and h^2 for AL and FL are shown in Table 5.3. The estimates of h^2 for AL and FL were the same at 0.07. The phenotypic and genetic coefficients of variation were 40 and 11% for AL and 40 and 10% for FL, respectively.

Table 5.3. Estimates of heritability and additive and phenotypic variances for actual and functional longevity of New Zealand dairy goats born between 1993 and 2011, obtained using a single-trait analysis.

	Actual longevity			Functional longevity		
	Estimate	SE[1]	CV[2]	Estimate	SE	CV
Genetic variance	39,547	4.85	11	38,482	4.76	10
Residual variance	529,157	52.61	39	523,033	52.54	38
Total variance	568,700	7,388	40	561,520	7,294	40
Heritability	0.07	0.01		0.07	0.01	

[1]SE = standard error, [2]CV = coefficient of variation (%).

Discussion

Longevity is an economically important trait in production animals, with genetic parameters being published for dairy cattle, sheep, pigs and rabbits, but there are only a few reports of the genetics of longevity in dairy goats. The main studies found in the literature included the same breeds as in this study in addition to the La Mancha (Castañeda-Bustos et al., 2014; Castañeda-Bustos et al., 2017; Valencia-Posadas et al., 2017) and Granadina (Pérez-Razo et al., 2004) breeds.

Descriptive statistics suggest that the average longevity of does in New Zealand was 1,891±832 days, which is longer than the 1,644 days previously reported (Gautam et al., 2017). Both studies used the same dataset but with different editing criteria. The dataset in this study accounted for pedigree information that limited the number of observations, whereas, Gautam et al. (2017) did not account for pedigree, so the dataset included more herds (38 herds).

Comparison with other studies was difficult because of differences in the definition of survival. Palhière et al. (2018) reported a decline in the length of productive life in Saanen and Alpine goats born in France from 1991 to 2011. The length of productive life of animals born in 1991 was 1,150 and 1,175 days for Saanen and Alpine, respectively, whereas the length of productive life of animals born in 2011 was 800 and 850 days for the respective breeds. These values of longevity are lower than the values found in this study, assuming that first kidding for those studies was at 365 days of age.

The proportion of does surviving as they become older shows noticeable dips every 200-300 days. A similar pattern was observed by Gautam et al. (2012), who modelled the instantaneous removal hazard (expressed as a probability of removal per day) as a function of age. Their results showed a reoccurring pattern which has crests representing the dry period within a lactation cycle (low risk of being culled), followed by a large dip (high risk of being culled). Therefore, the dips in (Figure 5.2) represent the end of each lactation period, before the animals are dried off, as this is typically the time when a farmer will cull undesirable animals, for example, those that are not pregnant or have low production.

The regression of AL on first-lactation ECMY was significant indicating that longevity was increased by 0.56 days for each extra kilogram of ECMY, indicating that does with higher production lasted longer. However, the estimate of h^2 for FL was the same as the estimate of h^2 for AL. This result agrees with Castañeda-Bustos et al. (2014), who reported similar h^2 estimates for productive and functional productive life. Nevertheless, knowledge about genetic parameters for FL are still important, as genetic improvement of longevity would be more efficient when the effects of voluntary culling can be taken into account (Castañeda-Bustos et al., 2014).

Despite low h^2 estimates for AL and FL, the genetic coefficients of variation for AL and FL (11% and 10%, respectively) suggest there is genetic variation of longevity in this dairy goat population. These results agree with those of Valencia-Posadas et al. (2017) who reported that there was sufficient additive genetic variation to justify the inclusion of functional stayability at 24 and 36 months of age, into a breeding program.

Estimates of h^2 of longevity vary in different species with low estimates for sows, cows and sheep (0.02-0.08 VanRaden and Klaaskate, 1993; Serenius and Stalder, 2004; El-Saied et al., 2005) and larger estimates for rabbits (0.15 Piles et al., 2006). Comparison of the h^2 estimates from this study with other estimates from other studies of goats, warrants caution as different definitions of longevity and statistical models have been used. Estimates of h^2 for length of productive life in French dairy goats (Palhière et al., 2018), FL of US dairy goats (Valencia-Posadas et al., 2017) and stayability of New Zealand dairy goat were low (0.07 to 0.09). Whereas, estimates of h^2 for length of productive life at 72 months old of US dairy goats (Castañeda-Bustos et al., 2014; Castañeda-Bustos et al., 2017) were higher (0.14 to 0.17). Overall, the h^2 estimates in this study were within the range of published values for longevity of dairy goats.

This study analysed the effect of does kidding in the same herd and year, breed, heterosis and the individual animal effect on longevity. In addition to these, the effects of birth month and dam age on survival of progeny, were also investigated. Including month of birth as a covariate was attempted, but birth date of animals was not accurately recorded in all farms, therefore, this factor could not be included in the model. Age of dam was calculated using pedigree data but the dataset containing the birth date of dams was incomplete. With many missing records, including this variable

in the analysis would have required that a significant proportion of data would have been filtered out and excluded from the analysis. However, using an incomplete dataset, results showed no significant effect of dam age on longevity.

Results from this study suggest that if selection for longevity is included in a selection index, there is adequate genetic variation for longevity of New Zealand dairy goats to allow genetic improvement for this trait. Solis-Ramirez et al. (2018) estimated the economic value for longevity of $0.04 per day, enabling the straightforward inclusion of this trait into a selection index. However, further work is required, especially in quantifying the genetic and phenotypic correlations with other traits, to enable the inclusion of longevity in the current genetic evaluation system.

Chapter 6

Genome-wide association studies of lactation yields of milk, fat, protein and somatic cell score in New Zealand dairy goats

This Chapter has been published in part elsewhere. It has been reformatted and presented here with permission:

Scholtens MR, Jiang A, Smith A, Littlejohn M, Lehnert K, Snell RG, Lopez-Villalobos N, Garrick DJ, Blair HT. 2020. Genome-wide association studies of lactation yields of milk, fat, protein and somatic cell score in New Zealand dairy goats. Journal of Animal Science and Biotechnology DOI:10.1186/s40104-020-00453-2.

Abstract

Identifying associations between genetic markers and traits of economic importance will provide practical benefits for the dairy goat industry, enabling genomic prediction of the breeding value of individuals, and facilitating discovery of the underlying genes and mutations. Genome-wide association studies (GWAS) were implemented to detect genetic regions that are significantly associated with effects on lactation yields of milk (MY), fat (FY), protein (PY) and somatic cell score (SCS) in New Zealand dairy goats. A total of 4,840 goats were genotyped with the Caprine 50K SNP chip (Illumina Inc., San Diego, CA, USA). After quality filtering, 3,732 animals and 41,989 SNPs were analysed assuming an additive linear model. Four GWAS models were performed, a single-SNP additive linear model and three multi-SNP BayesC models. For the single-SNP GWAS, SNPs were fitted individually as fixed covariates, while the BayesC models fit all SNPs simultaneously as random effects. A cluster of significant SNPs were used to define a haplotype block whose alleles were fitted as covariates in a Bayesian model. The corresponding diplotypes of the haplotype block were then fit as class variables in another Bayesian model. Across all four traits, a total of 43 genome-wide significant SNPs were detected from the SNP GWAS. At a genome-wide significance level, the single-SNP analysis identified a cluster of variants on chromosome 19 associated with MY, FY, PY, and another cluster on chromosome 29 associated with SCS. Significant SNPs mapped in introns of candidate genes (45%), in intergenic regions (36%), were 0-5 Kb upstream or downstream of the closest gene (14%) or were synonymous substitutions (5%). The most significant genomic window was located on chromosome 19 explaining up to 9.6 % of the phenotypic variation for MY, 8.1% for FY, 9.1% for PY and 1% for SCS. The quantitative trait loci for yield traits on chromosome 19 confirms reported findings in other dairy goat populations. There is benefit to be gained from using these results for genomic selection to improve milk production in New Zealand dairy goats.

Introduction

The majority of dairy goats in New Zealand are housed and their milk is primarily used to manufacture powdered nutritional products for sale in international markets. There are estimated to be 92 farms in New Zealand milking 66,100 dairy goats. Current estimates indicate that 85% of the dairy goats belong to the Saanen breed, while Toggenburg, British Alpine, and Nubian type crosses comprise the remaining 15%. The Dairy Goat Cooperative (NZ) Ltd (DGC) is the main processor of goat milk in New Zealand, and accounts for 80% of the dairy goat production. Farms that supply DGC, and undertake herd testing, participate in an annual genetic evaluation for MY, FY and PY and for SCS. Breeding values for these traits were estimated for each animal from a multi-trait repeatability animal model using available pedigree (Lopez-Villalobos and Garrick, 2001).

Genome-wide association studies identify associations between genetic markers and phenotypic expression of traits of interest. Genetic markers are analyzed for variation across the DNA sequence of the individual's genome (McCarthy et al., 2008). A GWAS allows the statistical evaluation or association of polymorphic loci with phenotypic variance to be quantified in a given population and can provide the genetic architecture of the complex traits which can be useful in medicine, agriculture and evolution (Goddard et al., 2016). One type of genetic marker commonly used in GWAS is characterised by single-nucleotide polymorphisms (SNPs), which exhibit two or more nucleotide variants at a single base. Genome-wide association studies have been performed in many livestock species, including dairy cattle (Mai et al., 2010; Pryce et al., 2010; Meredith et al., 2012), sheep (Zhao et al., 2011) and pigs (Sato et al., 2016; Le et al., 2017; Meng et al., 2017). Since release of the Illumina Caprine 50K BeadChip (Illumina Inc., San Diego, CA, USA), association of quantitative trait loci (QTL) in goats have been published for polledness (Kijas et al., 2013), milking speed (Palhière et al., 2014), wattles (Reber et al., 2015), coat colour (Becker et al., 2015; Martin et al., 2016a), supernumerary teats (Martin et al., 2016b), milk production and type traits (Maroteau et al., 2013; Martin et al., 2017).

Although the simplest and perhaps the most popular GWAS test for associations is between a single marker and a quantitative trait, the power of this method may suffer because a

single SNP may have only low linkage disequilibrium (LD) with the causal mutation and the LD contained jointly in flanking markers is ignored. An alternative method is to fit SNPs simultaneously using Bayesian methods, which take into account the LD between neighboring SNPs, limiting the false positive discoveries (Fernando and Garrick, 2013). Also, the SNP sliding window approach of the multi marker methods can be used to identify the most informative genomic regions, facilitating the discovery of associated markers and possible causal mutations. In addition, SNPs can be combined into a haplotype block. Clustering SNPs into a haplotype block combines information of adjacent SNPs into composite multi-locus haplotype alleles which may be more informative than individual SNPs and may also capture the regional LD information, which is arguably more robust and powerful (Pritchard et al., 2000; Akey et al., 2001).

Knowledge of genetic markers associated with milk production traits provides an opportunity to increase the rate of genetic gain using genomic or marker-assisted selection. Animals of above-average genetic merit can be identified at an early age and with a higher selection accuracy than conventional approaches, creating options for implementing selection schemes that reduce generation intervals (Schaeffer, 2006) and increase rates of genetic gain.

To date, a few GWAS have been conducted for milking traits of dairy goats. Studies that identified SNPs associated with milk production in dairy goats were performed by Martin et al. (2017; 2018), Palhière et al. (2018) and Mucha et al. (2018a). There are no published papers reporting GWAS for dairy goats in New Zealand. The objective of this study was to identify SNPs and genomic regions significantly associated with milk production traits in New Zealand dairy goats using the Caprine 50K SNP chip.

Materials and methods

Data

Phenotypic and pedigree records were provided by DGC from a dataset maintained by Livestock Improvement Corporation that included estimates of 305-day lactation records

for MY, FY, PY and SCS. The test interval method (Sargent, 1968), was used by Livestock Improvement Corporation to calculate MY, FY and PY for the actual realised lactation length, or up to 305 days in milk (DIM) for those lactations with more than 305 DIM. The dataset included 106,289 animals and 236,858 lactation records. The breed composition of the goats included Alpine (592), Nubian (374), Saanen (63,370), Toggenburg (1,741) and crossbred (34,054) animals, located in the Waikato region of New Zealand. Animals were considered crossbred unless the proportion of the major breed was >0.85. Breed composition was "unknown" for some goats (4,941). The pedigree contained 105,072 individuals spanning 5 generations, representing 1,322 sires and 27,180 dams. The records from a farm were included in the analysis if the farm supplied milk to DGC, performed herd-testing during 2017 or 2018, and contributed records for genetic evaluation. Phenotypes for the GWAS were pre-corrected for non-genetic factors using the GLM procedure of Statistical Analysis System version 9.4 (SAS Institute Inc., Cary, NC, USA) that produced residuals after fitting the fixed effects of herd-year and parity. The significance of association between the SNP effect or haplotype effect and the phenotype adjusted for herd-year and parity, as represented by the residual, was calculated at each SNP position.

Genotyping

Skin samples from 3,894 animals distributed in 21 herds were collected for SNP genotyping with the Illumina Caprine 50K BeadChip (Illumina Inc., San Diego, CA, USA). For three of the herds, only does in their first or second parity were sampled (14% of genotyped animals). Does of all parities were sampled in the remaining 18 herds (86% of genotyped animals). The recorded ancestors of the sampled animals were born between 2003 and 2015 and included 154 sires and 2,024 dams. Genotyped animals were of Saanen (1,436), crossbred (1,669), or unknown (789) breeds. A total of 51,462 SNPs were obtained.

The SNP & Variation Suite v8 (SVS) (Golden Helix, Inc., Bozeman, MT, USA) software was used for quality control, principal component analysis and two of the GWAS. Quality control was performed to remove genotypes from unreliable SNPs or animals. Records were

removed for individuals with >2% missing genotypes across all SNPs (call rate <98% which excluded 162 animals), SNPs with >1% missing genotypes across all individuals (call rate <99%), that deviated significantly from Hardy-Weinberg equilibrium threshold of P-value>10^{-6} or had minor allele frequency <1%. After these quality control edits, 3,732 animals and 41,989 SNPs remained for association analysis and the average distance between SNPs was 58.2 Kb and the average r^2 between two neighbouring SNPs was 0.15.

Genome-wide association study

A single-SNP GWAS was performed in SVS to identify SNPs significantly associated with the milk traits. The single-SNP GWAS (sGWAS) is based on a linear regression test of the fixed covariate effect of a single marker, which treats each SNP as if it had an additive effect. Population structure was estimated by principal component analysis in SVS using the method described by Price et al. (2006). The genomic relationship matrix was used to compute the principal components. The top 50 principal components captured 47% of the variation and were subsequently included as fixed effects in the sGWAS method. To correct for multiple testing, a Bonferroni correction of $\alpha = 0.05$ was applied to the genome-wide significance threshold (Significance threshold = α/number of SNP). The SNP effects were declared significant at a genome-wide level of P-value = 1.1×10^{-6} (0.05/41,989). Quantile-quantile plots were examined to determine the validity of the P-value for the sGWAS.

A BayesC GWAS was implemented in GenSel Software (Fernando and Garrick, 2013) fitting all SNPs simultaneously (sBayesC) to determine the proportion of variance explained by the SNPs. The algorithm uses Markov chain Monte Carlo (MCMC) methods to calculate samples from the posterior distributions of marker effects and variances, and inferences were made using the posterior means. The chains include 20,000 iterations after a burn-in of 1,000 iterations. For this model the priors for the genetic and residual variances were based on posterior means in a previous analysis (Scholtens et al., 2019). It was assumed that 99.8% of the SNPs have no effect on the trait. The genome was partitioned into 1 Mb windows and the multi-locus contribution to genetic variance of the combined effects of SNPs within

every one of these intervals were simultaneously estimated by sBayesC (Fernando and Garrick, 2013). The 1 Mb windows that explained >1% of genetic variance were considered to be associated with the traits.

The seven most significant SNPs clustered on chromosome 19 were combined into a haplotype block to further investigate true associations from the SNP analyses. The BayesC method was implemented a second time but with the alleles in the haplotype block included as fixed covariates while the remaining panel SNPs were fitted simultaneously as random effects (hBayesC). Thus, covariates for haplotype allele dosage were fitted instead of the dosage of alleles at each individual SNP in the QTL region. An expectation-maximisation algorithm was used to estimate haplotype allele frequencies and haplotype alleles with an expectation-maximisation probability >= 50% were included in the analysis (10 of the 28 haplotype alleles).

To test for non-additive effects of the haplotype alleles, a BayesC model was re-run again in GenSel, but fitting diplotypes (pairs of haplotypes) (dBayesC). Diplotypes were defined as class effects, but were only constructed for the two most common haplotypes. The effects of these diplotypes and the remaining eight haplotypes were fitted as fixed with the remaining SNPs simultaneously fitted as random effects.

Effects of haplotypes and diplotypes on the production traits were obtained using the GLM procedure of SAS. The model fitted for each trait and each haplotype, was $Y_i = b_0 + x_i b + e_i$ where Y_i is the residual phenotype of animal i, b_0 is the intercept, x_i is a row-vector indicating which haplotype and how many copies of the haplotype are carried by the animal; b is the effect of the haplotype and e_i is a residual effect. For the diplotype analysis, diplotype was treated as a class effect based on the number of copies of the two most common haplotypes.

Ensembl (Zerbino et al., 2018) was used to search for genes closest to the most significant SNPs. Gene annotation was retrieved if the SNP was located on an intron, lying 0-5 Kb upstream or downstream from gene boundaries, or, if the SNP was located in intergenic regions, the SNPs were assigned to the closest gene.

Results

Descriptive statistics for raw lactation yields of first and second parity genotyped does are presented in Table 6.1.

Table 6.1. Descriptive statistics of milking traits of genotyped New Zealand dairy goats in their first and second parity (N=7,284).

Trait	Mean	SD[1]	Min	Max	CV[2]
Lactation length (days)	272.3	117.0	60.0	696.0	43
Lactation yields					
Milk (kg)	804.5	290.4	58.4	2005.8	36
Fat (kg)	26.7	10.3	2.0	76.5	39
Protein (kg)	25.0	8.9	2.4	63.0	36
SCS[3]	8.6	1.3	3.5	13.7	15

[1]SD = standard deviation across herds, [2]CV = coefficient of variation (%), [3]SCS = somatic cell count calculated as $\log_2$ (somatic cell count).

Figure 6.1 shows the Manhattan plot for the sGWAS for lactation yields of MY, FY and PY and average SCS. A total of 43 genome-wide significant SNPs were detected across all four traits. A highly significant region (19:24836694-19:28953102) was identified on chromosome 19 for all four traits. In this region, 26 SNPs are associated with MY, 24 SNPs associated with FY and PY and 11 SNPs associated with SCS. Another significant region was identified on chromosome 29 (29:24850418-29:25328810) with 11 SNPs associated with SCS. The two top SNPs associated with MY, FY and PY were detected on chromosome 19 (19:26610610 and 19:26662281) with significance levels of $-\log_{10}$(P-value) = 22.51 and 21.67 for MY, 19.14 and 19.60 for FY, and 19.93 and 19.31 for PY. These two SNPs were also the top SNPs on chromosome 19 associated with SCS ($-\log_{10}$(P-value) = 8.22 and 7.93, respectively). Results obtained from the sGWAS model showed that the top SNP (19:26610610) explained 4.4% of the total variance for MY and 3.4% for FY and PY.

The Quantile-Quantile plot (QQ-plot) in Figure 6.2 shows the observed and expected P-values (expressed as $-\log_{10}$(P-value)) of the sGWAS for MY, FY, PY and SCS. The dashed line represents the distribution of the SNPs under the null hypothesis that there is no association of SNPs with the trait of interest. The strong deviation of the observed from the expected P-values for all eight QQ-plots indicate that there were more SNPs significantly associated with all of the four traits than would be expected by chance.

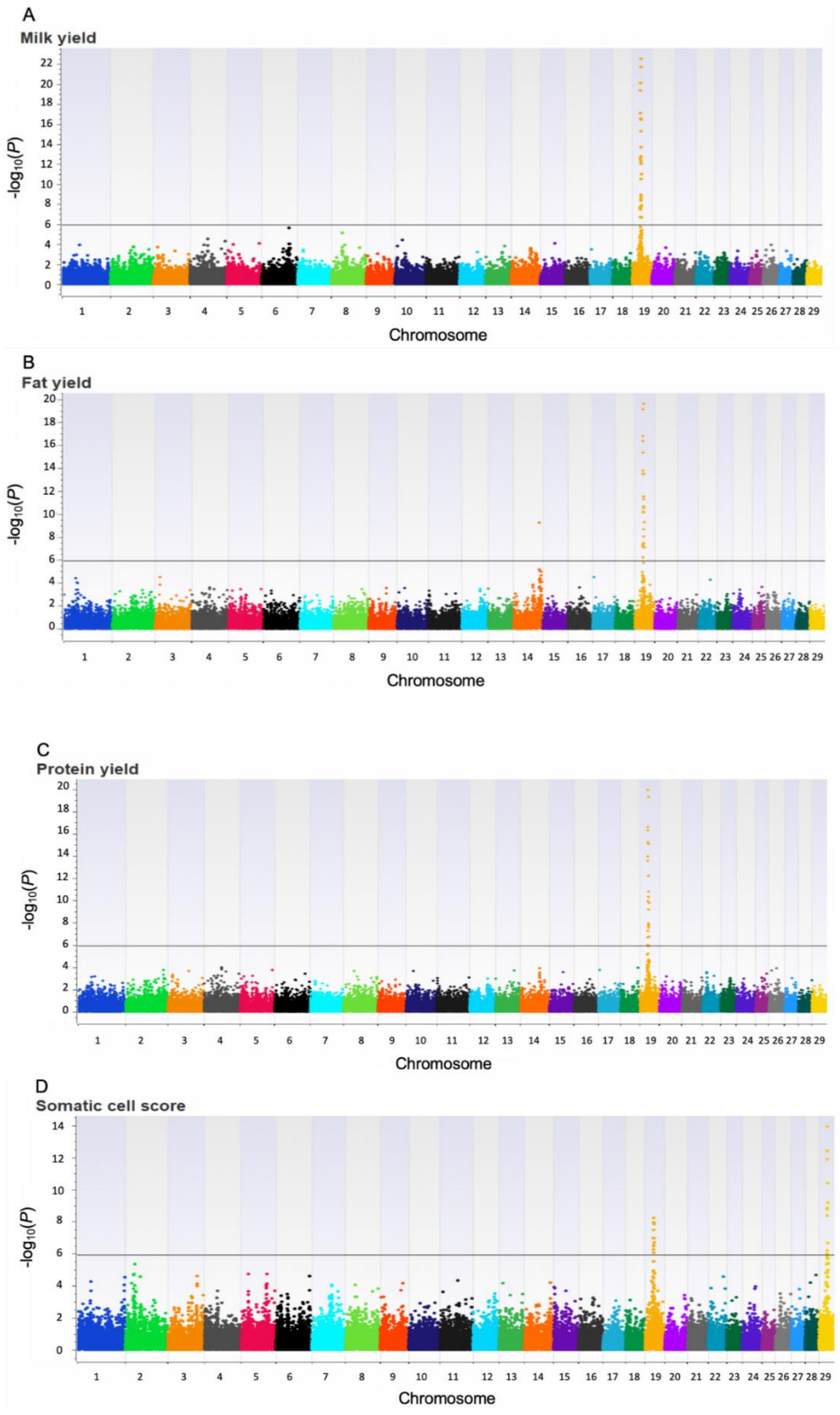
A
Milk yield
B
Fat yield
C
Protein yield
D
Somatic cell score
-log10(P)
Chromosome

Figure 6.1. Manhattan plot of sGWAS for lactation yields of milk (A), fat (B) and protein (C) and average somatic cell score (D), using the Illumina Caprine 50K BeadChip (Illumina Inc., San Diego, CA, USA) in 3,732 New Zealand dairy goats. The P-values (-$\log_{10}$ (P-value)) for each SNP are shown on the y-axis and chromosomes 1-29 are shown on the x-axis. The horizontal line indicates the Bonferroni-corrected genome-wide threshold at P-value 0.05.

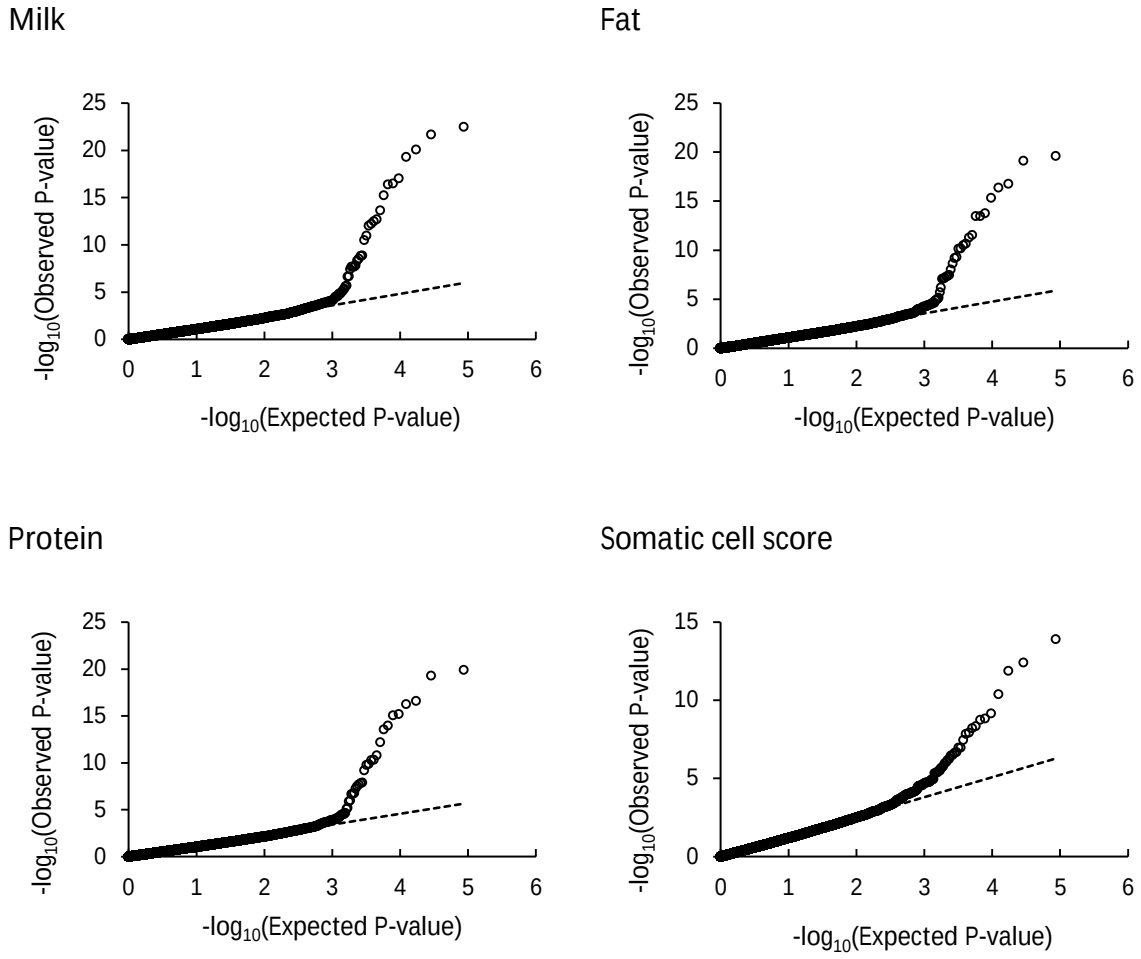

Figure 6.2. Quantile-quantile plots observed and expected P-values (expressed as –$\log_{10}$ (P-value)) of the sGWAS for yields of milk, fat, protein and somatic cell score in New Zealand dairy goats.

In this study the sBayesC model partitioned the genome into 2,520 1 Mb SNP windows with an average of 17 SNPs per window. The windows were sorted based on the proportion of genetic variance each window captured. The genomic region with the highest proportion of explained genetic variance for MY, FY and PY was on chromosome 19 (19:26029220-19:26956209). The combined effect of the 15 SNPs within this window was estimated to explain 9.62% of the genomic variance for MY, 8.09% for FY, 9.09% for PY and 0.94% for SCS. The probability that this window explained more than the average genetic variance expected under an infinitesimal model of inheritance was 1.00 for MY, FY and PY and 0.98 for SCS. Other windows of interest included one on chromosome 6 (6:86050148-6:86990478) explaining 1% of the genomic variance of MY, chromosome 14 (14:81032694-14:81952406) explaining 2% of the genomic variance of FY and a window on chromosome 29 (29:25025234-29:25972909) explaining 3% of genomic variance of SCS.

Table 6.2 shows the variances obtained from the Bayesian analyses in GenSel. The proportion of phenotypic variance explained by all SNPs fitted in the sBayesC model was 18% for MY, 16% for FY, 14% for PY and 20% for SCS. The genetic variances were reduced for MY, FY and PY, when the haplotypes or diplotypes in the QTL region were fitted as fixed effects (hBayesC and dBayesC models), the reduction representing the genetic variance explained by the haplotypes and diplotypes. When the haplotypes were fitted into the hBayesC model, the remaining SNPs explained 12% of the total variance for MY, 11% for FY, 9% for PY and 20% for SCS.

When the BayesC model was adjusted for the effects of the haplotypes or diplotypes (hBayesC or dBayesC, respectively), the SNPs that showed the highest model frequency were located on chromosomes 6 and 8. This suggested that all of the informative SNPs located on chromosome 19 were accurately included in the haplotype block.

Table 6.2. Summary of variances estimated from BayesC GWAS for lactation yields of milk, fat and protein and average somatic cell score, using the Illumina Caprine 50K BeadChip (Illumina Inc., San Diego, CA, USA) in 3,732 New Zealand dairy goats.

Trait	Model[1]	Genetic variance	Phenotypic variance	Phenotypic variance explained by SNPs (%)
Milk yield				
	sBayesC	12925	73865	18
	hBayesC	8739	71128	12
	dBayesC	8484	70893	12
Fat yield				
	sBayesC	13.15	83.27	16
	hBayesC	9.23	80.82	11
	dBayesC	9.06	80.53	11
Protein yield				
	sBayesC	8.61	63.91	14
	hBayesC	5.48	62.06	9
	dBayesC	5.38	61.89	9
Somatic cell score				
	sBayesC	0.28	1.45	20
	hBayesC	0.28	1.45	20
	dBayesC	0.28	1.45	20

[1]Models = sBayesC= BayesC model fitting all SNPs simultaneously, hBayesC = BayesC model fitting 10 haplotype alleles as fixed effect and remaining SNPs as random effects simultaneously, dBayesC = BayesC model fitting diplotypes of h1 and h2 as well as the 8 remaining haplotypes, and the remaining SNPs as random effects simultaneously.

The population frequency of the haplotype alleles and their diplotypes are presented in Table 6.3. The commonest haplotypes, h1 and h2, had estimated frequencies of 49% and 17%, respectively. Diplotypes were derived based on the occurrence of h1 and h2, of which, 79% of the population is estimated to have at least one copy of h1 and 34% of the population is estimated to have at least one copy of h2.

The estimated effect of haplotypes and diplotypes on milk traits are reported in Table 6.4. The most frequent haplotype h1 has the greatest positive effect on MY, FY and PY, while haplotype h7 had the greatest negative effect on MY, FY and PY. The diplotype with the

greatest effect on yields includes two copies of h1 (h1-h1), of which 29% of the population is estimated to carry. The diplotype with the largest negative effect on MY and PY comprised of one copy of h2 (h2-h0), which is estimated to represent 11% of the population.

Table 6.3. Estimated population frequency of the 10 most frequent haplotypes, and diplotypes within the most significant region on chromosomes 19 associated with milk production in New Zealand dairy goats.

Haplotype number	Haplotype	Frequency
h1	TCTTCTG	49%
h2	CTCCTGA	17%
h3	CTCCTTG	11%
h4	CCCCTTG	5%
h5	TCCCTTG	4%
h6	CCTCTTG	4%
h7	CCCCTGA	2%
h8	CCTCTGA	2%
h9	CCTTCGA	1%
h10	TTCCTTG	1%

Diplotype number	Frequency
h1-h0	30%
h1-h1	29%
h1-h2	20%
h2-h0	11%
h2-h2	3%
h0-h0	9%

Diplotype numbers with a h0 refers to the occurrence of any haplotype other than h1 and h2.

Table 6.4. Effects (± standard error) of haplotypes and diplotypes located within the most significant region on chromosome 19 on milk traits in New Zealand dairy goats.

	Milk yield (kg)	Fat yield (kg)	Protein yield (kg)	SCS[1] (units)
Haplotypes				
h1	73.6 (3.1)[a]	2.15 (0.11)[a]	1.91 (0.09)[a]	0.16 (0.02)[a]
h2	-57.0 (4.1)[a]	-1.87 (0.14)[a]	-1.54 (0.12)[a]	-0.08 (0.02)[a]
h3	-28.7 (5.6)[a]	-0.59 (0.19)[b]	-0.68 (0.16)[a]	-0.20 (0.03)[a]
h4	-59.1 (6.9)[a]	-1.62 (0.24)[a]	-1.64 (0.20)[a]	-0.12 (0.04)[a]
h5	-70.0 (7.9)[a]	-2.24 (0.27)[a]	-1.89 (0.23)[a]	-0.10 (0.04)[b]
h6	-5.7 (7.7)	-0.10 (0.26)	-0.07 (0.22)	-0.06 (0.04)
h7	-107.6 (20.9)[a]	-3.59 (0.72)[a]	-2.58 (0.61)[b]	-0.07 (0.11)
h8	-43.4 (18.9)[b]	-2.24 (0.65)[a]	-1.92 (0.55)[b]	0.01 (0.10)
h9	7.8 (13.1)	1.49 (0.45)[a]	1.04 (0.38)[b]	-0.02 (0.07)
h10	-44.7 (21.7)[b]	-2.15 (0.75)[b]	-1.05 (0.64)	0.16 (0.12)
Diplotypes				
h1-h0	53.4 (4.1)[a]	1.88 (0.14)[a]	1.61 (0.12)[a]	-0.18 (0.02)[a]
h1-h1	103.6 (3.8)[a]	3.10 (0.13)[a]	2.75 (0.11)[a]	0.05 (0.02)[b]
h1-h2	42.4 (4.6)[a]	1.17 (0.16)[a]	1.21 (0.14)[a]	-0.11 (0.02)[a]
h2-h0	-58.0 (6.9)[a]	-1.49 (0.24)[a]	-1.44 (0.20)[a]	-0.17 (0.04)[a]
h2-h2	-45.9 (13.5)[a]	-1.83 (0.47)[a]	-1.33 (0.40)[a]	-0.41 (0.07)[a]
h0-h0	-37.3 (8.4)[a]	-1.08 (0.29)[a]	-0.93 (0.25)[a]	-0.32 (0.04)[a]

[1] SCS = $\log^2$(somatic cell count), [a]P-value<0.001, [b]P-value<0.05, Diplotype numbers with a h0 refers to the occurrence of any haplotype other than h1 and h2.

Table 6.5 shows the 43 genome-wide significantly associated SNPs with the milk production traits and positional candidate genes (i.e. annotated genes that are nearest to each marker). Half of significant SNPs on chromosome 19 are mapped to introns, 27% to intergenic regions, 7% introducing synonymous substitutions and the remaining 17% located upstream or downstream to the closest genes. The two top SNPs from the sGWAS, are located on chromosome 19 at 26610610 bp, introducing a synonymous substitution in the RNASEK gene (ENSCHIG00000009505) and at 26662281 bp, located within the intron of the ASGR2 gene (ENSCHIG00000003690). Both SNPs were significantly associated with all four

milk traits. Other SNPs included in the haplotype block were SNP 19:26724454, located within the intron of the DLG4 gene (ENSCHIG00000009974), and 19:26780952, located downstream of the ELP5 gene (ENSCHIG00000010521), that were also significantly associated with MY, FY and PY. The functional annotation of SNP (19:27854624) resulted in a synonymous substitution in MYH10 (ENSCHIG00000018616) and SNP (19: 28079607) located within an intergenic region but the closest gene being 166 Kb from the MYH10 gene. Both SNPs were significantly associated with SCS.

Of the 11 SNPs on chromosome 29 significantly associated with SCS, four were mapped to introns and seven were in intergenic regions. The most significant SNP on chromosome 29 (29: 25328810) is located in an intergenic region and is 60 Kb from the closest gene, PTPN5 (ENSCHIG00000008345). Another significant SNP (29: 25366901), is also near the same gene (22 Kb). In addition, two SNPs (29: 25649038 and 29: 27144973) significantly associated with SCS were located within introns of the LDHC gene (ENSCHIG00000013476) and OR8B4 (ENSCHIG00000012776), respectively. The two remaining SNPs on chromosome 29 (26:25175690 and 29:25206548), were located within introns are of the ZDHH13 gene (ENSCHIG00000024992).

Table 6.5. Genes linked to the 43 genome-wide significant SNPs for yields of milk, fat, protein and somatic cell score in New Zealand dairy goats.

Chr[1]	Position	Trait[2]	-log[10](P)	Annotation	Gene name	Gene description
14	81658443	FY	9.21	Upstream	ZNF16	Zinc finger protein 16
19	24836694	SCS	6.97	Intron	MYBBP1A	MYB binding protein 1a
19	25087981	MY, PY	6.7.2-7.7	Intron	KIAA0753	KIAA0753 ortholog
19	25413768	MY, FY, PY	7.2-7.7	Intergenic	WSCD1	WSC domain containing 1
19	25782297	MY	6.7	Intergenic	NLRP1	NLR family pyrin domain containing 1
19	25823025	MY, FY, PY	9.9-12.5	Intergenic	NLRP1	NLR family pyrin domain containing 1
19	26072328	MY, FY, PY	16.3-19.3	Intergenic	RABEP1	Rabaptin, RAB GTPase binding effector protein 1
19	26115456	MY, FY, PY	6.0-7.4	Intergenic	ZNF232	Zinc finger protein 232
19	26148755	MY, FY, PY	16.6-20.1	Downstream	ZFP3	Zinc finger protein
19	26192128	MY, FY, PY	15.2-17.1	Downstream	KIF1C	KIF1C Kinesin family member 1C
19	26420506	MY, FY, PY	13.5-15.2	Intron	ZMYND15	Zinc finger MYND-type containing 15
19	26542254	MY, FY, PY	6.7.3-7.7	Downstream	none	Arachidonate 12-lipoxygenase, epidermal-type
19	26578775	MY, FY, PY, SCS	6.5-16.4	Intergenic	none	Arachidonate 12-lipoxygenase, epidermal-type
19	26610610	MY, FY, PY, SCS	8.2-22.5	Synonymous	RNASEK	Ribonuclease K
19	26662281	MY, FY, PY, SCS	7.9-21.7	Intron	ASGR2	Asialoglycoprotein receptor 2
19	26724454	MY, FY, PY	7.9-8.9	Intron	DLG4	Discs large MAGUK scaffold protein 4
19	26780952	MY, FY, PY	7.5-8.6	Downstream	ELP5	Elongator acetyltransferase complex subunit 5
19	27360768	MY, FY, PY	7.5-8.4	Intron	CNTROB	Centrobin, centriole duplication and spindle assembly protein

19	27401023	MY, FY, PY, SCS	6.3-13.6	Intron	GUCY2D	Guanylate cyclase 2D, retinal
19	27480793	MY, FY, PY, SCS	6.5-12.7	Intron	ALOXE3	Arachidonate lipoxygenase 3
19	27529983	MY, FY, PY, SCS	6.1-12.2	Intron	none	Vesicle associated membrane protein 2
19	27558520	MY, FY, PY	10.3-12.0	Intron	TMEM107	Transmembrane protein 107
19	27605322	MY, FY, PY	9.2-10.5	Intron	CTC1	CST telomere replication complex component 1
19	27744036	SCS	7.5	Upstream	NDEL1	NudE neurodevelopment protein 1 like 1
19	27854624	SCS	7.0	Synonymous	MYH10	Myosin heavy chain 10
19	28038645	MY, FY, PY, SCS	6.7-16.5	Intron	PIK3R6	Phosphoinositide-3-kinase regulatory subunit 6
19	28079607	SCS	7.8	Intergenic	MYH10	Myosin heavy chain 10
19	28202268	MY, FY, PY	7.1-8.9	Intergenic	NTN1	Netrin 1
19	28578424	MY	6.7	Intron	STX8	Syntaxin 8
19	28730193	MY, FY, PY	9.3-11.0	Intron	GLP2R	Glucagon like peptide 2 receptor
19	28953102	MY, FY, PY	6.8-7.8	Intron	none	Growth arrest specific 7
29	24850418	SCS	6.2	Intergenic	NAV2	Neuron navigator 2
29	25175690	SCS	11.9	Intron	ZDHHC13	Zinc finger DHHC-type containing 13
29	25206548	SCS	12.4	Intron	ZDHHC13	Zinc finger DHHC-type containing 13
29	25328810	SCS	14	Intergenic	PTPN5	Protein tyrosine phosphatase, non-receptor type 5
29	25366901	SCS	8.3	Intergenic	PTPN5	Protein tyrosine phosphatase, non-receptor type 5
29	25649038	SCS	8.8	Intron	LDHC	L-lactate dehydrogenase C chain
29	26381310	SCS	6.0	Intergenic	OR10D3	Putative olfactory receptor 10D3

29	26502551	SCS	8.8	Intergenic	none	Olfactory receptor 145-like
29	27144973	SCS	10.4	Intron	OR8B4	Olfactory receptor family 8 subfamily B member 4
29	27407592	SCS	6.7	Intergenic	PANX3	Pannexin 3
29	27967983	SCS	9.2	Intergenic	PKNOX2	PBX/knotted 1 homeobox 2

[1]Chromosome, [2]MY = milk yield, FY = fat yield, PY = protein yield, SCS = somatic cell score.

Discussion

Genome wide association studies have been used to identify associations between genetic markers and candidate genes for traits of economic importance. This study evaluated the associations of 41,989 SNPs with MY, FY, PY and SCS from 3,732 New Zealand dairy goats.

The sGWAS identified 43 SNPs significantly associated with MY, FY, PY and SCS in this population. A cluster of highly significant SNPs were identified on chromosome 19 for all four traits and on chromosome 29 for SCS. The two strongest signals were identified at SNP 19:26610610 and 19:26662281. These two SNPs were in high LD ($r^2 = 0.94$) and it is highly probable that these SNPs were in high LD with a QTL or causal variant that had a very significant effect on MY, FY and PY in this dairy goat population.

Quantile-quantile plots (Figure 6.2) of the observed and expected P-values of the sGWAS for each trait indicated that a large proportion of the observed P-values were clearly more significant than expected under the null hypothesis. This suggested there were some true associations between SNPs and genes controlling these traits.

The main advantage of the sGWAS is the ease of significance testing. However, single-SNP analysis relies on LD between the marker and QTL, therefore the results do not provide information about the location of the causal variant, instead they correspond to the location of the marker. Also, fitting SNPs individually may result in the same signal picked up in multiple single SNP tests, thus overestimating the number of actual QTLs detected. And finally, although a significant signal is identified, if a trait is controlled by many QTLs, which

is the case for most quantitative traits, the single-locus tests may prove inaccurate compared with methods where grouped (haplotypes) or all SNPs are jointly considered. For these reasons, an additional analysis was performed fitting all SNPs simultaneously in a BayesC GWAS.

The BayesC GWAS that fits all SNPs simultaneously, can improve the accuracy of detecting QTLs (Wolc et al., 2012) and the 1 Mb window variances provide greater insight for identifying the genomic region of the casual variant (Fernando and Garrick, 2013) and estimates the proportion of variance explained by the SNPs.

In the Bayesian analysis, the percentage of genetic variance explained by 1 Mb genomic windows are used to make inference about the proportion of variance explained by a QTL and whether the QTL bleeds over multiple windows. The genomic window that explained the greatest level of genetic variance (8-9%) for MY, FY and PY included 15 SNPs and ranged from 26420506 to 26780952 bp on chromosome 19. Two of the SNPs located in this window were also the most frequent SNPs included in the model (suggesting they are informative SNPs that contribute to the model) and were the top SNPs identified in the sGWAS to be associated with MY, FY and PY.

Combing these results from the sGWAS and Bayesian analyses provides strong evidence that those SNPs with the highest model frequency within the genomic window on chromosome 19 with the largest effect, are likely to be in LD with the causal variant.

To learn more about this potential QTL on chromosome 19 the seven most significant SNPs identified in the sGWAS were combined into a haplotype block and the Bayesian analysis was re-run by adjusting for the SNPs in the haplotype block. Fitting covariates for haplotype alleles rather than the SNP alleles provides higher LD between causal mutations and haplotype alleles as the multi-locus haplotype takes into account not only the LD information from the SNPs within the haplotype but also other important polymorphisms within the QTL cluster region. In addition, the use of haplotypes can provide information regarding the genetic determinants that cannot be captured by the biallelic markers. For example, when a SNP is fit in the model there is no guarantee that its alleles are in high LD

with the QTL allele, whereas in the haplotype, provided there are enough SNPs to represent them, at least one haplotype must contain the favourable QTL allele and at least one must include the unfavourable allele. When the haplotypes were fitted into the hBayesC model as a fixed effect, there were no other signals on chromosome 19 of large effect, indicating that the majority of the QTL was indeed captured within the genomic region of that haplotype block. Also, the genetic variance from the hBayesC model was lower than the sBayesC, indicating that the seven SNPs located in the haplotype are capturing the variation that exists in that genomic region.

The haplotype effects on milk production in this dairy goat population were estimated for the 10 haplotypes. Haplotypes h1 and h9 had the greatest positive effect on the milk traits. Animals that carry one copy of h1 or h9 are estimated to produce +73.6 and +7.8 kg milk, +2.2 and +1.5 kg fat and +1.9 and +1.0 kg more protein, respectively, per lactation, compared with the average of the population. Both h1 and h9 are the only haplotypes that contain the T allele at the loci 19:26610610, which had the strongest signal on the milk traits as well as the C allele at the loci 19:2666281, which had the second strongest signal on the milk traits. This suggests that an animal carrying the T and C alleles at the corresponding loci will have the greatest yields per lactation compared to the average of the population. The positive effect of these loci on milk production traits should be used in combination with performance and pedigree information to generate more accurate breeding values. When selecting animals for breeding replacements, genotyped males carrying the desirable alleles can be used for mating to females to produce replacements that carry the desirable alleles and thus the potential to be high yielding animals. In addition, the h7 haplotype had the greatest negative effects on the milk traits, therefore, farmers could identify animals with this haplotype and avoid breeding or as a selection method for culling purposes.

When haplotypes are fitted in the model as dosage covariates we assume the haplotypes have an additive effect, which may not be true. Therefore, to test whether the effect of the haplotype block was truly additive, we fit diplotypes (pairs of haplotypes) into the model. Fitting diplotypes allows the estimation of the effect of the heterozygote without assuming

it is intermediate between the opposite homozygotes, which can determine whether that haplotype allele is additive, or dominant or over-dominant etc.

The diplotypes included in the trend regression were derived from the two most frequent haplotypes in the population, h1 and h2. The predominant diplotype (29%) in the study population had two copies of h1. Animals that do not carry either h1 or h2 had an average effect of -37.3 kg milk per lactation, relative to the population average. If an animal has only one copy of h2 then they will have -58.0 kg, which is 20.7 kg less than animals with neither h1 nor h2. If an animal carries one copy of h1, they will have +53.4 kg milk, producing 90.7 kg more than an animal that carries neither h1 nor h2. If an animal carries two copies of h1 then it will have +103.6 kg milk than the population, producing an extra 50.2 kg milk more than an animal with one copy of h1. These results follow a similar trend for FY and PY and suggest that h1 has a positive effect on milk traits and can lead to increased productive value of dairy goats in New Zealand.

Several studies have identified QTL significantly associated with milk production traits in goats (Table 6.6). Results from our study confirmed the presence of a QTL reported on chromosome 19 for MY, FY, PY and SCS and on chromosome 29 for SCS. In addition, several novel regions were identified, including a QTL for FY on chromosome 14 and genetic regions associated with MY, FY and PY on chromosome 23 and SCS on chromosome 5.

Table 6.6. Reported QTL associated with milk production traits in dairy goats.

Trait	Chromosome	Reference
Milk yield	6, 8, 14, 19 and 21	Roldán et al. (2008), Maroteau et al. (2013), Martin et al. (2017), Mucha et al. (2018a)
Fat yield	2, 14 and 19	Maroteau et al. (2013), Martin et al. (2017)
Protein yield	19 and 20	Maroteau et al. (2013), Martin et al. (2017)
Fat content	6, 7, 14, 20, 21 and 25	Roldán et al. (2008), Maroteau et al. (2013), Martin et al. (2017)
Protein content	1, 3, 5, 6, 11, 20, 21, 28	Roldán et al. (2008), Maroteau et al. (2013), Martin et al. (2017)
Fatty acid	1, 7, 8, 11, 14 and 29	Maroteau et al. (2013)
SCS	19, 29	Maroteau et al. (2013), Martin et al. (2018)
Morphology traits	29	Maroteau et al. (2013), Martin et al. (2018)

The QTL on chromosome 19 that was strongly associated with all four traits, was reported in the French Saanen dairy goat population (Palhière et al., 2014) and a mixed breed population (Mucha et al., 2018a). In addition to milk traits in dairy goats, this highly significant region was also associated with type traits (Martin et al., 2018), udder floor position (Palhière et al., 2014), functional longevity (Palhière et al., 2018) and semen production (Oget et al., 2018), suggesting a pleiotropic QTL effect. Further investigation into this genomic region (chromosome 19, 25-29 Mb) revealed that the SNPs significantly associated with MY in the current study were different to the SNPs identified by Mucha et al. (2018a) in their mixed breed goat population. This could be because both studies analysed mixed breed populations, thereby having different levels of linkage disequilibrium (de Roos et al., 2008), thus, the loci on the SNP have different levels of linkage disequilibrium with the unknown causal. With that said, although the individual SNPs differed in statistical significance between the goat populations, this highly significant region identified in both studies suggests the segregation of a common gene that has a major effect on milk production in dairy goats.

In this study, the most significantly associated SNP (19:26610610) was located on chromosome 19 introducing a synonymous substitution in the RNASEK gene

(ENSCHIG00000009505). RNASEK is a transmembrane protein ubiquitously expressed and highly conserved across mammals. RNASEK localises to the cell surface and endosomal pathway and closely associates with the vacuolar ATPase (V-ATPase) proton pump. RNASEK is required for endocytosis that prevents the replication of multiple pathogenic viruses such as rhinovirus, influenza A and dengue (Perreira et al., 2015). This most significant SNP was strongly associated with all four milk traits, but no previous studies have reported this SNP or any association with this gene in goats. However, this SNP is in strong LD ($r^2 = 0.94$) with SNP 19:26662281, which was also strongly associated with all four milk traits. This SNP (19:26662281), is located within the intron of the ASGR2 gene (ENSCHIG00000003690) and is in the same region where Mucha et al. (2018a) reported a locus (19:26150581) that is strongly association with udder depth of mixed breed dairy goats. The ASGR2 gene encodes a subunit of the asialoglycoprotein receptor involved with the glycoprotein metabolic process, lipid homeostasis and the regulation of protein stability. Therefore, the possibility of the ASGR2 gene's involvement with the milk production traits is supported by its activity in lipid homeostasis and protein stability.

Another signal strongly associated with all four milk traits is SNP (19:27480793) which is located within the intron of the ALOXE3 gene. A SNP (19:26972244) in the same gene region was reported by Mucha et al. (2018a) to be associated with udder depth of mixed breed dairy goats. This gene is part of the lipoxygenase family of enzymes and is involved in the metabolic pathway during formation of the epidermal barrier (Krieg et al., 2013). As this process includes the activity in cell differentiation, cell proliferation and fat metabolism, it is possible that this gene is involved with udder conformation (Mucha et al., 2018a), and subsequently milk production.

Another association which was reported by Mucha et al. (2018a) was for SNP (19:26066457), which is located near the ALOX12 gene (GOAT_ENSP00000251535) and has a significant effect on MY in dairy goats (Mucha et al., 2018a). However, this SNP and chromosome region were not significantly associated with milk production traits in this current dairy goat population.

In the current study, the SNP (19:26192128) was significantly associated with MY, FY and PY and is located downstream from the KIF1C gene (ENSCHIG00000000772). This gene is involved in the movement of molecules from the Golgi back to the endoplasmic reticulum. This SNP was also reported at the genome-wide significance level, to be associated with functional longevity in Saanen dairy goats (Palhière et al., 2018). In the same population, Martin et al. (2017) also reported the same genomic region to be associated with milk production. This is not surprising as multiple studies have published a positive genetic correlation between milk production and longevity in dairy goats (Castañeda-Bustos et al., 2014; Wheeler et al., 2014).

Two significant SNPs were mapped within and close to the MYH10 gene (ENSCHIG00000018616) which is involved in mitotic cytokinesis. The SNP (19: 27854624) causes a synonymous substitution and the 19: 28079607 SNP is located 166 Kb from the gene. Both SNPs were significantly associated with SCS and not with the other milk traits.

Other genes on chromosome 19 associated with milk production traits in dairy goats include the GH1 gene located in the 47 Mb region (Lan et al., 2007; Dettori et al., 2013) and the STAT5A gene located in the 42 Mb region (An et al., 2012). However, in our study there were no associations for milk traits detected in these regions.

We identified a peak of significant SNPs on chromosome 29 associated with SCS. It is evident there is a QTL located on this chromosome for SCS as we detected strong signals for 11 SNPs from the sGWAS. However, further investigation using Bayesian methods would provide more information about the genomic region of the QTL and the level of variance it explains. Previous studies have reported a chromosome-wide significant SNP on chromosome 29 associated with MY (Mucha et al., 2018a) and fatty acid composition (Maroteau et al., 2013) in French dairy goats and associated with gastrointestinal nematode resistance in dairy goats in Zimbabwe (Zvinorova, 2017).

Two of the top SNPs associated with SCS (29: 25175690 and 29:25206548) are within introns of the ZDHHC13 gene (ENSCHIG00000024992), which is associated with signal transducer activity and palmitoyltransferase activity. Palmitoyltransferase is important for the positive

regulation of I-kappaB kinase/NF-kappaB signalling, which is an inflammatory signalling pathway. This gives credibility to the SNP being associated with SCS in this study.

Other genomic regions that may be involved in SCS include the LDHC gene, which is involved in carbohydrate metabolic processes such as the chemical reactions and pathways resulting in the formation of ATP, a universally important coenzyme and enzyme regulator, and the OR8B4 gene, which changes the activity or state of a cell in response to a chemical stimulus by chemoreceptors i.e. smell perception.

Only one genome-wide significant signal was detected on chromosome 14 (14:81658443) with FY. Although associated with FY, the genome-wide significant SNP was not located in the immediate region of the DGAT1 gene, a gene known to have a major effect on milk fat content in goats (Martin et al., 2017) and cattle (Grisart et al., 2002). Instead, the SNP was located upstream of the ZNF16 gene (ENSCHIG00000020215). Although not studied in goats, this gene promotes cell proliferation and inhibits cell apoptosis in humans (Li et al., 2011).

Despite only a few papers reporting GWAS studies in dairy goats, candidate genes related to milk traits have been widely studied. Some polymorphisms associated with milk production in goats include the LALBA gene (chromosome 5) which is linked to milk yield, lactose content and milk coagulation properties (Dettori et al., 2015a), the MTHFR gene (chromosome 6) involved in milk protein synthesis (An et al., 2015), the β-lactoglobulin gene (chromosome 11) (Dettori et al., 2015b; El Hanafy et al., 2015) associated with milk yield and daily fat and protein yield, the TLR2 gene (chromosome 17) which is important in the recognition of the innate immune system of mastitis causing bacteria (Ruiz-Rodriguez et al., 2017) and the PRLR gene (chromosome 20) (Hou et al., 2013) and the STAT5A gene (chromosome 19) (An et al., 2012), that are associated with milk yield. But none of the significant SNPs in the current study were located in the regions of these genes.

Although numerous studies have provided evidence of polymorphisms within specific genes influencing milk production, there are limited studies using GWAS methodologies to identify QTL for milk production traits in dairy goats. All of the previous GWA studies identified SNPs

that were of varying significance levels for different breeds (Martin et al., 2017; Martin et al., 2018; Mucha et al., 2018a; Palhière et al., 2018). In our study, the goats were of mixed breeds, representative of the New Zealand dairy goat population.

Results from the GWAS strongly show a QTL located on chromosome 19 and the trend regression analysis suggest this is biallelic with h1 containing the desirable allele. This was detected when analysing the effects of the haplotypes and confirmed by the estimated diplotype effects. All diplotypes containing h1 resulted in positive effects on milk traits, while every diplotype that contained h2 had negative effects on the milk traits, compared with the average of the population. The fact that animals carrying one copy of both h1 and h2 still had positive effects on the milk traits shows the greater magnitude of the positive effect of h1 over the negative effect of h2.

The results from this study provide evidence that there is a likely QTL strongly associated with milk traits in this population. It is possible that the QTL has an additive effect and is biallelic. In addition, it is concluded that this QTL has a pleiotropic effect as it has been identified in other goat populations and associated with a range of traits besides milk production traits.

Although the study population was small, the significant regions identified were also reported in other studies, which gives confidence in the results. Nevertheless, the results require validation. If the new results are consistent with the current results, the identified markers could be used for marker-assisted-selection. This will enable the prediction of genetic and phenotypic value of individuals. For example, to predict the future phenotypes of offspring so that those with the best breeding values can be selected as parents of the next generation (Goddard et al., 2016). At the same time, the information on the genomic regions found in this study, can be used to facilitate the identification of candidate genes for these milk traits. Doing so would enable a greater understanding of the biology underlying the response from genomic selection, and managing possible consequences of selecting for mutations with undesirable pleiotropic effects (Goddard and Hayes, 2012). Ultimately, these results provide an opportunity for adopting genomic selection within the New Zealand dairy goat population. Implementing genomic selection will increase the

accuracy of predicted genetic and phenotypic values and reduce the generation interval, leading to increased rates of genetic improvement within this dairy goat population.

Conclusion

The study identified one region strongly associated with milk production traits in New Zealand dairy goats. The highly significant region identified on chromosome 19 was also reported in French dairy goat populations and suggests a major pleiotropic QTL for milk production traits in dairy goats. The significant SNPs will increase the accuracy of predicted genetic and phenotypic values of individuals to allow for genomic selection. The results demonstrated in this study require validation using a larger dataset before implementing genomic selection within the New Zealand dairy goat population.

Chapter 7

Advantage of including genomic information to predict breeding values for lactation yields of milk, fat, and protein or somatic cell score in a New Zealand dairy goat herd

Abstract

Selection on genomic breeding values (GBVs) is now readily available for ranking candidates in improvement schemes. Our objective was to quantify benefits from including genomic information in single trait estimation of breeding values (BVs) for a New Zealand mixed breed dairy goat herd. The dataset comprised phenotypic and pedigree records of 839 does. The phenotypes comprised estimates of 305-day lactation yields of milk, fat and protein and average somatic cell score from the 2016 season. Only 388 goats were genotyped with a Caprine 50K SNP chip and 41,981 SNPs passed quality control. Pedigree-based best linear unbiased prediction (PBLUP) was used to obtain across-breed breeding values (EBVs) whereas a single-step BayesC model (ssBC) was used to estimate across-breed GBVs. The average prediction accuracies ranged from 0.20 to 0.22 for EBVs and 0.34 to 0.43 for GBVs. Accuracies of GBVs were up to 103% greater than EBVs. Breed effects were more reliably estimated in the ssBC model compared to the PBLUP model. The greatest benefit of genomic prediction was for individuals with no pedigree or phenotypic records. Including genomic information improved the prediction accuracy of BVs compared to the PBLUP method currently implemented in the New Zealand dairy goat population.

Introduction

The purpose of selection is to improve the performance of a population. Selection based on GBVs can improve the rate of genetic gain compared to using only performance and pedigree records and has become a widely adopted method for ranking candidates for selection in animal breeding schemes (Cole and VanRaden, 2018). The benefits of genomic prediction (GP) are greatest when traits of interest are difficult to measure, expensive to record, measured late in an animal's life, or the traits have low heritability (e.g. disease resistance, feed efficiency, slaughter traits, survivability and fertility). Nevertheless, GP can still be beneficial for easy-to-measure heritable traits like milk production traits as GP can be applied to young animals allowing earlier identification of replacement candidates, thereby reducing replacement costs and also shortening the generation interval, which may

increase the rate of genetic improvement provided the accuracy of selection is not greatly reduced. Genomic prediction can increase the accuracy of GBVs, especially if no records are available on the selection candidates.

Until recently, the standard method of estimating breeding values was to use PBLUP that uses phenotypic records of the individual and its relatives (Garrick and Fernando, 2014). That method uses pedigree information to estimate the average genetic relationships among the individuals based on the probability that genes are identical by descent (Wright, 1922), i.e. half- siblings born to unrelated non-inbred parents are expected to share 0.5 of their alleles, and these probabilities are the basis for generating the average genetic relationship matrix (A) between close and distant relatives in the pedigree. Meuwissen et al. (2001) proposed a genomic best linear unbiased prediction (GBLUP) method, that predicted GBV using information from genetic markers located across the entire genome in an attempt to capture all quantitative trait loci (QTL) influencing the variation in a trait (Hayes et al., 2009). Including information from all markers, i.e. GBLUP, can provide greater accuracy for estimating breeding values compared to PBLUP. The GBLUP method uses single-nucleotide polymorphisms (SNPs) to identify alleles identical in state that can be shared through common ancestors (not necessarily recorded in the pedigree) to generate a genomic relationship matrix (G).

Legarra et al. (2009) suggested a single-step genomic BLUP (ssGBLUP) approach using phenotypes, genotypes and pedigree information to predict GBVs for both genotyped and non-genotyped individuals, simultaneously. The method combines pedigree information from the A and genomic information from the G into a modified genetic relationship matrix (H). This single-step approach uses Henderson's mixed model equations (MME) and the H to yield unbiased predictions under multivariate normality, even in populations that are undergoing selection and non-random mating. A single-step procedure increases both power and precision by taking advantage of phenotypes from related and unrelated animals. Despite these advantages, ssGBLUP requires computation of the G or its inverse which can be computationally demanding when many animals are genotyped. Fernando et al. (2014) proposed a class of single-step Bayesian regression (ssBR) methods that does not

require the computation of the G or its inverse. Instead, this ssBR approach imputes marker covariates for non-genotyped animals based on their genotyped relatives and a genetic imputation error effect to accommodate the difference between true and imputed genotypes (Fernando et al., 2014). Another difference between ssGBLUP and ssBR methods are the assumptions of the distribution of SNP effects and the number of SNPs included in the model. For ssGBLUP it is typically assumed that the effects of all SNP are normally distributed, and all SNPs have the same variance (Meuwissen et al., 2001; VanRaden, 2008). Meanwhile, the Bayesian methods incorporate prior information into the model that assumes a fraction (π) of the SNPs have an effect whereas a fraction (1-π) have no effect on the trait. BayesA and BayesB use a student-t distribution as a prior for the SNPs with effects, which allows some SNPs to have large effects on the trait (Meuwissen et al., 2001) while BayesC assumes SNP effects are normally distributed and have the same variance (Habier et al., 2011). Based on these assumptions, if there are known QTL with large effect on traits within the population, and many SNPs that are unlikely to be causal, then it would be more appropriate to fit a mixture model where some of the SNPs are assumed to have zero effect.

Estimating breeding values involves a so-called training population that has phenotypes and genotypes. The prediction model uses this "training data" in ssBR to predict the influence of genetic markers by regression of the observed phenotypes on marker genotypes. Then, the marker effects are summed across all loci to get the GBVs of individuals in another dataset that don't have observed phenotypic records and are referred to as the validation population. This method performs best when all individuals come from the same population and therefore the linkage disequilibrium (LD) between genetic markers and QTL persists from the training to the validation population. This LD is the non-random association of alleles at two or more loci and is influenced by population history and the pattern of geographic subdivision (Slatkin, 2008).

The success of GP depends largely on the existence of LD across the population of interest. This level of LD persists across larger distances of the genome when the effective population size is smaller and therefore persists more within breeds than across breeds and as a result,

GP using 50K SNP markers is generally more successful in purebred populations (Moghaddar et al., 2014).

Dairy Goat Cooperative (NZ) Ltd (DGC) processes 80% of goat milk in New Zealand. Farms that supply DGC, and undertake herd testing, participate in an annual genetic evaluation for lactation yields of milk (MY), fat (FY) and protein (PY) and for average somatic cell score (SCS). Breeding values for these traits are estimated for each animal from an across-breed multi-trait repeatability animal model using available pedigree information (Lopez-Villalobos and Garrick, 2001). In total, there are believed to be 92 farms in New Zealand milking an estimated 66,100 dairy goats. Current estimates indicate that 85% of the dairy goats are of the Saanen breed, while Toggenburg, British Alpine, and Nubian type crosses comprise the remaining 15%. Therefore, although efficacy of GP within-breed is promising, it is necessary to develop and evaluate across-breed predictors in order for these genomic tools to be applied to the New Zealand dairy goat industry. Breed covariates are included in the evaluation in order to account for the differences in expected value of the breeding values for animals of different breeds or cross.

Only 4,840 of the animals that comprise the New Zealand dairy goat population have been genotyped. This means that the reference population is relatively small, which will limit the accuracy of GP using GBLUP or Bayesian methods (Goddard, 2009). The aim of this chapter was to quantify the benefit from the inclusion of genomic information in the estimation of breeding values for a single New Zealand dairy goat herd.

Materials and methods

Data

Phenotypic and pedigree records were provided for a single dairy goat herd by the DGC from a herd-test database maintained by Livestock Improvement Corporation. The original dataset comprised lactation records from the 2016 season for 883 dairy goats. The phenotypic records were estimates of 305-day lactation yield records for MY, FY, PY and SCS. The test interval method (Sargent, 1968), was used by Livestock Improvement

Corporation to calculate MY, FY and PY for either the realised lactation length, or up to 305 days in milk (DIM) for those lactations with more than 305 DIM. Average SCS over the lactation was calculated as the mean Log2(somatic cell count) from each herd-test. Lactation yields were removed if the lactation length was <105 days, MY <100 kg, FY or PY <3 kg, deviation from median kidding date was less than -90 or more than +90 days. The final dataset contained 839 animals that were offspring of 46 sires and 589 dams. Contemporary group was defined as the group of does of the same lactation number (1, 2, 3 4 and ≥5). Breed composition of each animal was calculated from pedigree proportions of Alpine, Nubian, Saanen, Toggenburg, "other" and "unknown" breeds. There was some crossbreeding but very few first-cross or purebred animals of Alpine, Nubian, Toggenburg and "other" breeds. The breed composition of animals in this herd consisted of 21 purebred Saanen and 818 animals with mixed breed composition. For this analysis the breeds were described as proportion of Saanen, or the sum of all other breeds referred to as ANTO (Alpine, Nubian, Toggenburg, other breeds and unknown breed).

Genotyping

Skin samples were collected for SNP genotyping with the Illumina Caprine 50K BeadChip (Illumina Inc., San Diego, CA, USA) in 2016. Of the 51,462 SNPs obtained, a total of 41,981 SNPs per animal remained after quality control. Quality control of genotypic data was performed using the SNP & Variation Suite v8 (SVS) (Golden Helix, Inc., Bozeman, MT, USA) software. Individuals were discarded if they had a call rate <95% or if they didn't have phenotypic records. SNPs were discarded if they had a call rate <90%, MAF <1% or deviated significantly from Hardy-Weinberg equilibrium based on a threshold of P-value<10^{-6}. The majority of genotypes were from does in their second parity (246 genotyped animals), while the remaining genotyped animals were in parity one (19 animals), three (90 animals), four (30 animals) or older than fourth parity (3 animals). The 388 genotyped animals were of Saanen (14) or unknown (374) breeds.

Statistical evaluation

In this analysis, a PBLUP model was used as the base model to estimate across-breed EBVs. A single-step BayesC model (ssBC) was used to estimate across-breed GBVs. Phenotypic and pedigree records from 839 animals were included in both models and genotypes of the 388 animals were also included in the ssBC model.

Pedigree-based BLUP evaluation

The PBLUP was performed using a single-trait animal model to predict EBV using pedigree and phenotypic records for all animals in the pedigree. The PBLUP model was performed using ASReml 3.0 software package (Gilmour et al., 2009) with the following model:

$$y = Xb + ZDd + Za + e \quad (7.1)$$

where y is the vector of phenotypes comprising at most one lactation record for MY, FY, PY or SCS,

b is the vector of fixed effects,

d is the vector of effects of ANTO and unknown breeds,

a is the vector of additive genetic effects (random effects of animal),

e is the vector of random residual effects (residual errors not accounted for by the fixed and random effects),

X and Z are design matrices relating the fixed and additive genetic effects, respectively,

D is a matrix with a row for each animal in the pedigree and columns for the proportion of ANTO and the proportion of unknown breed (Saanen was used as the base breed to constrain the regression coefficient for the Saanen breed effect to zero).

Fixed effects included in b were contemporary group and as covariates, day of kidding, days in milk and general heterosis. General heterosis was calculated as $1 - \sum_{j=1}^{f} d_j^2$ where d_j is the proportion of each of the f=3 breeds (Saanen, ANTO) (Gregory and Cundiff, 1980). The additive animal genetic effect was included as a random effect, and assumed to have a normal distribution with mean Dd and variance As^2_g, where A is the numerator relationship matrix from the pedigree and s^2_g is the within-breed additive genetic variance. Residuals were assumed to have a normal distribution with mean zero and variance Is^2_e, where I is an identity matrix of size equal to the number of animals with a lactation record, and s^2_e is the residual variance.

Estimated breeding values were calculated as:

$$\widehat{EBV} = D\hat{d} + \hat{a} \qquad (7.2)$$

where

$\widehat{EBV}$ is the vector of across-breed EBVs,

$\hat{d}$ is the solutions for the ANTO and unknown breed effects,

$\hat{a}$ is the vector of the solution for within-breed random animal effects.

Prediction accuracy for each model and trait were assessed using a validation process by splitting the herd into two subsets: the training set of the oldest 70% of the herd (587 animals) and the validation set with the youngest 30% (252 animals). The PBLUP model was used on the training set to produce pedigree-based EBVs (EBV) for the validation set.

The prediction accuracy was also assessed and summarised based on the different levels of pedigree information available in the evaluation. Using predicted EBVs of animals in the validation set (252 animals), the average prediction accuracies were calculated when the animal had: A) neither the sire nor dam were recorded (1 animal), B) the dam was recorded and had ≥1 lactation record (6 animals), C) the sire was recorded and had ≥5 progeny in the herd (155 animals), or D) both the sire and dam recorded (161 animals). To demonstrate

the impact on prediction accuracy when both the sire and dam are unknown, the PBLUP evaluation was re-run an additional time but by masking the pedigree records of animals that previously had records of both the sire and dam, resulting in 162 animals in the evaluations for scenario (A).

Single-step BayesC genomic evaluation

In this population there is a known QTL with a large effect on milk traits (Scholtens et al., 2020). Therefore, it is appropriate to fit a BayesC model which assumes a mixture of marker effects, with a point mass at zero with a probability of p and a univariate normal distribution with probability 1- p for all marker effects. This model was fitted using the JWAS Julia package (Cheng et al., 2016) fitting 50,000 Markov chain Monte Carlo (MCMC) iterations (including 1,000 burn in) and p was assumed known at 0.98. The model in matrix notation was:

$$\begin{bmatrix} y_n \\ y_g \end{bmatrix} = \begin{bmatrix} X_n \\ X_g \end{bmatrix} b + \begin{bmatrix} Z_n D_n \\ Z_g D_g \end{bmatrix} d + \begin{bmatrix} Z_n J_n \\ Z_g J_g \end{bmatrix} q + \begin{bmatrix} Z_n M_n \boldsymbol{\alpha} + \boldsymbol{\epsilon} \\ Z_g M_g \boldsymbol{\alpha} \end{bmatrix} + \begin{bmatrix} W_n & 0 \\ 0 & W_g \end{bmatrix} u + e \qquad (7.3)$$

where the vectors and matrices for non-genotyped animals are denoted with subscript n and those for the genotyped animals with a subscript g. Thus:

y_n and y_g are the vectors of phenotypes,

X_n and X_g are the incidence matrices for fixed effects,

b is a vector of fixed effects including the contemporary group (does kidding in the same parity) and as covariates, day of kidding, days in milk and general heterosis,

D_n and D_g are matrices with a row for each non-genotyped and genotyped animal in the pedigree and columns for the proportion of ANTO and the proportion of unknown breed,

d is a vector of effects of ANTO and unknown breeds,

J_n and J_g are matrices with a row for each non-genotyped and genotyped animal in the pedigree and columns for the J covariate for each breed group that were included as fixed effects to fit the difference between the genotyped founder and non-genotyped founder breeds. The matrix J_n is computed as $A_{ng}A_{gg}^{-1}J_g$, for breed f, where A_{ng} and A_{gg}^{-1} are submatrices of the numerator relationship matrix A, and J_g is the matrix of breed fractions identical to D_g except that it includes the vector of breed fractions for Saanen,

q is the vector containing the regression covariates for J, which account for the difference in breeding value between genotyped and non-genotyped animals of the same breed (Hsu et al., 2017),

Z_n and Z_g are incidence matrices that relate the breeding values of animals,

M_g is the matrix of centred marker covariates for genotyped animals,

$M_n = A_{ng}A_{gg}^{-1}M_g$, is the matrix of marker covariates for the non-genotyped animals that are imputed from genotyped relatives,

a is the vector of random marker effects,

$\boldsymbol{\epsilon}$ is the vector of genetic imputation error effects,

W_n and W_g are incidence matrices that relate the residual polygenic effects,

u is the vector of residual polygenic effects,

e is a vector of residuals.

The fixed effects are assumed to have flat priors. The prior for the marker effects depends on the marker variance, s^2_{ak}, and the prior probability p that SNP k has zero effect and follows a two-component mixture prior:

$$a_k|\pi, \sigma^2_{a_k} = \begin{cases} 0 & \text{with probability } \pi, \\ \sim N(0,\sigma^2_{a_k}) & \text{with probability } (1-\pi), \end{cases} \quad (7.4)$$

where $s^2_{ak} \sim v_a, S_a^2 \, X^2_{va}$. A previous study in this population reported that the markers captured 12% of the genetic variance (Scholtens et al., 2020). To recognise the markers did not explain the total genetic variance, a residual polygenic effect was included in the model accounting for 88% of the additive genetic variance. The vector of imputation residual deviations, $\epsilon \sim N[0,(A_{nn}-A_{ng}A_{gg}^{-1}A_{gn})(1-w)s^2_g]$ (Fernando et al., 2014), where A_{nn}, A_{ng}, A_{gg} and A_{gn} are submatrices of A, s^2_g is the total genetic variance with $(s^2_g|v_g,S_g^2) \sim v_g,S_g^2 \, X_{vg}^2$, and w is the ratio of residual polygenic to total genetic variance (0.88), $u \sim N(0, Aws^2_g)$ that are not captured by markers and e is $e_i|s^2_e \sim_{iid} N(0, s^2_e)$ with $(s^2_e|v_e, S_e^2) \sim v_e \, S_e^2 \, X_{ve}^2$.

Genomic breeding values were calculated as:

$$\widehat{GBV} = D\hat{d} + \begin{bmatrix} J_n \\ J_g \end{bmatrix} \hat{q} + \begin{bmatrix} \hat{M}_n \\ M_g \end{bmatrix} \hat{\boldsymbol{\alpha}} + \begin{bmatrix} Z_n \\ 0 \end{bmatrix} \hat{\boldsymbol{\epsilon}} \qquad (7.5)$$

where

$\widehat{GBV}$ are the across-breed GBV's,

$\hat{d}$ is the solutions for the ANTO and unknown breed effects,

$\hat{q}$ is a vector of regression coefficients for the J covariates for each breed group,

$\hat{\boldsymbol{\alpha}}$ is the vector of solutions for random marker effects,

$\hat{\boldsymbol{\epsilon}}$ is a vector of solutions for imputation residuals.

The BayesC mixture model used in the single-step analysis requires that unknowns to be estimated using MCMC techniques. Due to the limited number of observations in this evaluation (a single herd), variance components were estimated using the ssBC model and data from a larger dataset first (phenotypic records from 24,317 individuals and 41,981 markers on 2,681 individuals). That posterior residual variance and the heritability values

previously estimated for this population (Scholtens et al., 2019) for each of the traits were used to calculate the total genetic variance. The "known" variance components were then considered known in both the PBLUP and ssBC models (Table 7.1). All ssBC models were fitted for 50,000 MCMC iterations including a burn-in of 1,000 iterations using the JWAS package in Julia. Convergence of MCMC iterations were assessed using the coda package in RStudio based on the method of Geweke (1991).

Table 7.1. Prior across-breed variance components fitted in the PBLUP and ssBC models.

Trait	Polygenic variance	SNP variance	Residual variance	π	h^2
Milk yield	9,098.6	1,011.0	30,345.0	0.98	0.25
Fat yield	8.66	1.33	35.20	0.98	0.24
Protein yield	5.69	0.88	23.10	0.98	0.24
Somatic cell score	0.51	0.08	2.50	0.98	0.21

Reliabilities of PBLUP EBVs were calculated as $(1-(PEV/s_g^2))$ where PEV is the predicted error variance calculated by inverting the coefficient matrix of the MME (Henderson, 1975) and s_g^2 is the total genetic variance. For ssBC the PEV were computed from the Bayesian posterior variance of GBV samples. Prediction accuracies were calculated as the square root of the reliability.

The across-breed EBVs and GBVs were standardised to a base population mean of zero because PBLUP and ssBC were independent evaluations. The EBVs were standardised by subtracting each EBV by the mean EBV and each GBV was standardised by subtracting by the mean GBV, resulting in the population means of EBVs and GBVs being zero.

Results

Descriptive statistics are shown in Table 7.2. Mean FY and PY were both 31.8 kg. The coefficients of variation for the milking traits reflect the phenotypic variation in this herd.

Table 7.2. Descriptive statistics of milking traits of 839 dairy goats kidding in the 2016 season from a single New Zealand herd.

Trait	Mean	SD[1]	Min	Max	CV[2]
Lactation length (days)	288	23	185	305	8
Yields (up to 305 days)					
Milk yield (kg)	1,002.0	268.3	292.4	1811.3	27
Fat yield (kg)	31.8	8.9	8.2	67.1	28
Protein yield (kg)	31.8	8.4	8.9	58.5	26
SCS[3] (units)	9.5	1.2	6.5	12.6	12

[1]SD = raw standard deviation across the herd, [2]CV = coefficient of variation (%), [3]SCS = calculated as average $\log_2$(somatic cell count).

Average accuracies of GBVs obtained for MY, FY, PY and SCS were greater than the average accuracies of EBVs (Table 7.3). The greatest increase in accuracy was obtained for FY with +103% more accurate GBVs compared to EBVs.

Table 7.3. Accuracies (r) of EBV and GBV of milk traits for animals in the validation population using PBLUP[1] and ssBC[2] methods, N=100.

	EBV		GBV		Gain
Trait	r	SE	r	SE	
Milk yield	0.22	0.01	0.38	0.01	+73%
Fat yield	0.21	0.01	0.43	0.01	+103%
Protein yield	0.21	0.01	0.34	0.01	+64%
Somatic cell score	0.20	0.01	0.39	0.01	+95%

[1]PBLUP = pedigree-based best linear unbiased prediction model, [2]ssBC = single-step BayesC model, [3]SE = average standard error from animals in the validation population.

For all scenarios and all traits, the GBVs had greater accuracies than the EBVs (Figure 7.1). When the individual had no lactation records but had a sire and dam recorded in the pedigree (Scenario D), the average accuracies of EBV and GBV were 0.27 and 0.43 for MY, 0.26 and 0.47 for FY, 0.24 and 0.39 for PY and 0.24 and 0.43 for SCS, respectively. The greatest increase in accuracy between the two prediction models was if the animal had no

phenotypic or pedigree information (Scenario A), with GBV prediction accuracies of 0.39 for MY, 0.30 for FY and SCS and 0.33 for PY, compared to 0 for EBVs.

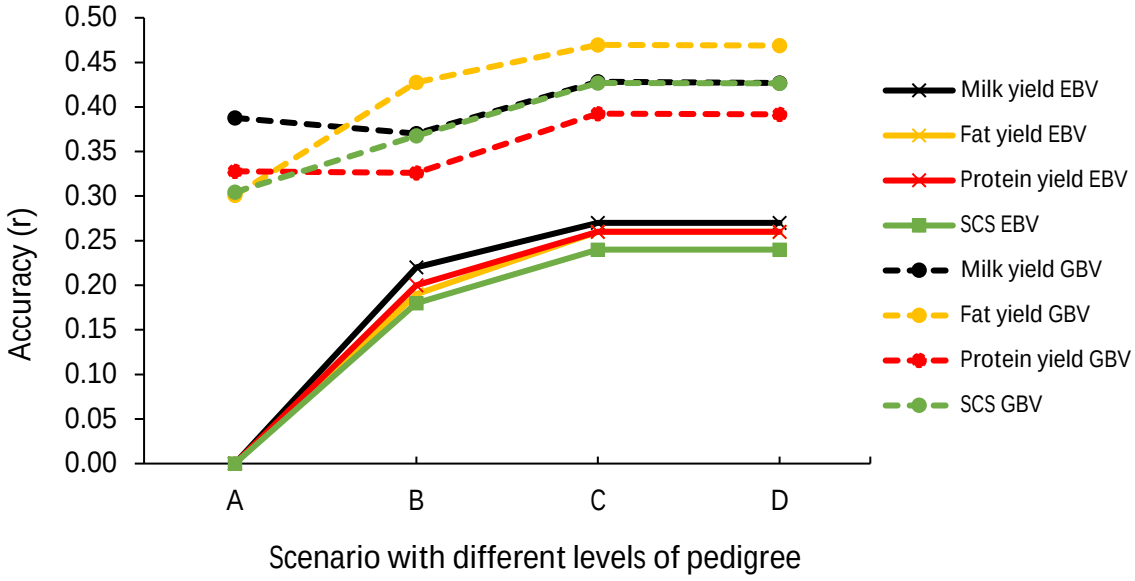

Figure 7.1. Average prediction accuracies of EBV and GBVs of milk traits obtained using PBLUP and ssBC models, respectively, for validation animals. Animals in scenarios B, C and D were obtained from the same evaluations and accuracies were summarised for animals when both the sire and dam was recorded (Scenario D, N=161 animals), the sire was recorded and had ≥5 progeny in the herd (Scenario C, N=155 animals), the dam was recorded and had ≥1 lactation record (Scenario B, N=6 animals). Animals in scenario A were obtained from a second evaluation where the pedigree records of animals that previously had records of both the sire and dam were masked (N=162 animals).

Figure 7.2 shows scatterplots of GBV against EBV of milk traits of the animals in the validation population that have both the sire and dam known after base adjustment. The correlations ranged from 0.603 to 0.978.

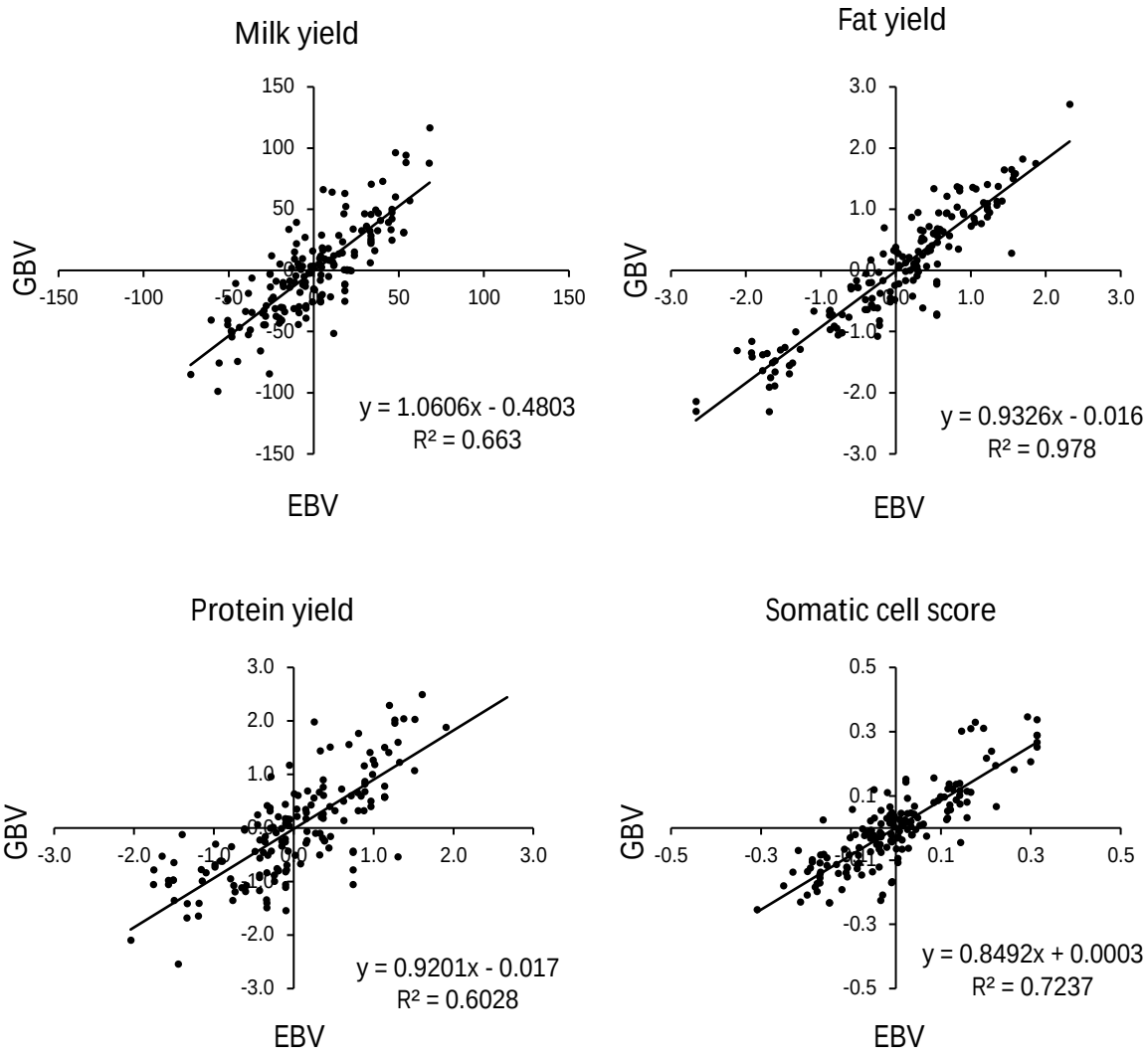

Figure 7.2. Scatterplot of standardised-GBV against standardised-EBV of milk yield, fat yield, protein yield and somatic cell score of animals in the validation set that have both the sire and dam recorded, N= 161.

Table 7.4 shows the breed coefficients obtained from the PBLUP model and the sum of the breed and J covariate coefficients obtained from the ssBC model for each breed group. The breed effects of the Saanen breed group are constrained to zero. Based on pedigree and phenotype records, the effect of either ANTO or unknown breed groups are lower than the Saanen breed, however, when genotypes are included in the model the group for the

animals of unknown breed is estimated to be better than that for the Saanen breed for MY and PY. The breed effects are much more reliably estimated in the ssBC model compared to the PBLUP model, as evident by comparing their SE.

Table 7.4. Estimated breed coefficients and standard errors (SE) of milk traits obtained from PBLUP[1] and the sum of breed and J covariate coefficients obtained from ssBC[2].

		Breed coefficient of traits							
Model	Breed	Milk yield (kg)	SE	Fat yield (kg)	SE	Protein yield (kg)	SE	SCS[3]	SE
PBLUP									
	Saanen	0		0		0		0	
	ANTO	-100.60	138.70	-0.52	4.69	-1.07	3.80	-0.02	1.22
	Unknown	-33.12	45.10	-1.18	1.52	-1.50	1.24	0.11	0.40
ssBC									
	Saanen	-112.54	2.52	-0.59	0.08	-2.11	0.07	0.47	0.02
	ANTO	-364.56	5.82	-5.28	0.19	-8.87	0.16	0.01	0.05
	Unknown	-64.04	2.34	-0.92	0.08	-1.42	0.07	0.46	0.02

[1]PBLUP = pedigree-based best linear unbiased prediction model, [2]ssBC = single-step BayesC model, [3]SCS = calculated as average $\log_2$(somatic cell count).

Breed and J covariate coefficients for each breed group obtained from the ssBC model are shown in Table 7.5. The broad range between breeds reflect the large differences between the breed groups, particularly the values of the J covariate for the ANTO breed group for MY. The SEs of ssBC estimates of breed or J covariates are much larger than those for the sum of the breed effects and J covariates shown in Table 7.4, indicating that the breed and J covariates in the ssBC model are confounded.

Table 7.5. Estimated breed and J covariate coefficients and standard errors (SE) of milk traits obtained from the ssBC[1] model.

	Breed and J coefficients of traits							
	Milk yield		Fat yield		Protein yield			
Breed	(kg)	SE	(kg)	SE	(kg)	SE	SCS[2]	SE
Saanen	0		0		0		0	
ANTO	94.04	247.57	4.99	8.54	7.03	6.73	1.03	2.25
Unknown	-164.89	83.04	-2.29	2.67	-4.53	2.37	0.58	0.72
	Breed specific J covariate coefficient of traits							
J_{Saanen}	-112.54	78.95	-0.59	2.56	-2.11	2.25	0.47	0.71
J_{ANTO}	-458.60	300.00	-10.28	10.23	-15.90	8.19	-1.02	2.76
$J_{Unknown}$	100.85	34.88	1.38	1.14	3.11	1.00	-0.12	0.31

[1]ssBC = single-step BayesC model, [2]SCS = calculated as average $\log_2$(somatic cell count).

Discussion

The aim of this study was to evaluate how the inclusion of genomic information would impact the accuracy of predicting breeding values for milk traits in New Zealand dairy goats. Across-breed breeding values were estimated for MY, FY, PY and SCS using the PBLUP model and a ssBC model. A training population was used to develop the prediction equation that was then used to predict EBVs and GBVs of the animals in the validation population. In addition to comparing prediction accuracies of EBVs and GBVs of animals in the validation population, the effect of the level of pedigree information available in the evaluations were also explored.

This study reported that a ssBC model using genotypes, pedigree and phenotypic records can obtain more accurate predictions of animal genetic merit compared to the PBLUP model currently used for the genetic evaluation of dairy goats in New Zealand. The average prediction accuracies ranged from 0.20 to 0.22 for EBVs and 0.34 to 0.43 for GBVs. The increase in accuracy is particularly valuable as these GBVs can be estimated for all individuals in the evaluation, even those without phenotypes or pedigree information. Whereas with the PBLUP evaluation those animals without phenotypic or pedigree records

would have otherwise been excluded. Another benefit of predicting GBVs is the early selection of young bucks that would be used for breeding without the need to wait for progeny testing or does producing their first lactation records.

The expected prediction accuracy of PBLUP EBV of MY of this population when both the sire and dam of an individual is known should be ~0.36 ($0.71*\sqrt{h^2}$) (Van Vleck, 1993). Likewise, the average prediction accuracy when the dam is known and has a lactation record is expected to be ~0.25 ($0.5*\sqrt{h^2}$). The low prediction accuracies obtained in this study from the PBLUP model suggest the current pedigree records provide limited information to the genetic evaluation of animals in this herd. Whereas the genomic information provides a significant amount of information to the evaluation as shown by the increased prediction accuracies. Despite this increase, these accuracies are much lower than those reported in other dairy goat populations. Multiple studies using ssGBLUP approaches have published prediction accuracies for milk traits in dairy goats including accuracies of 0.61 for MY in the UK (Mucha et al., 2015), 0.69 for MY in Spain (Molina et al., 2018), from 0.64 to 0.74 for MY, FY and PY (Carillier et al., 2014) and from 0.73 to 0.77 for protein content in France (Teissier et al., 2018). These studies (Carillier et al., 2014; Mucha et al., 2015; Molina et al., 2018; Teissier et al., 2018) reported that the genomic models only increased accuracies by +5% to +12% compared to PBLUP models, while in the current study, the accuracies increased by +64% to +103%. The minimal increase in accuracy in the UK, Spanish and French populations demonstrates that when the population has rich pedigree records, including genomic information is not as advantageous. Meanwhile, the significant increase in accuracy obtained in the current population demonstrates the substantial benefit that genomic information can have on the prediction of GBVs in this population. This is due to the ability of the ssBC model to capture additive genetic relationships between the individual and its relatives from their shared genotypes.

The greatest benefit of including genomic information in the evaluation is for animals that have no pedigree or phenotypic records. In a PBLUP genetic evaluation these animals would not be linked to the pedigree and excluded from the evaluation. Whereas, the ssBC model can capture the additive genetic relationships, enabling the evaluation to predict a GBV

based on the GBV of their relatives. For this reason, the average prediction accuracies of GBVs when neither the sire or dam were recorded, ranged from 0.30 to 0.39 for the four traits. It should be noted that these GBV accuracies obtained when neither the sire and dam are known, are also greater than the accuracies of EBVs when both the sire and dam of the individual is known (0.24 to 0.27), demonstrating the benefit of including genomic information into the evaluation.

When animals in the validation population have a dam in the reference population with at least one lactation record, this animal was able to be included in the evaluation, as the PBLUP model could then include information from the dam to estimate the genetic merit of the individual. In this scenario, prediction accuracies of EBVs range from 0.18 to 0.22. The accuracy of the GBVs of FY and SCS was slightly greater when the dam was recorded in the pedigree and had her own lactation record, compared to the animal having no pedigree records. However, the prediction accuracy was lower for prediction of MY and PY GBVs. These contradictory accuracies of GBVs could be due to the limited number of animals in the validation population that had dams with lactation records (6 animals), and therefore was not an accurate representation of the true effect of having this additional information in the reference population. On the other hand, these differing accuracies between traits could suggest that the inclusion of the dam and her lactation records does not add much to the accuracy of GPs. The latter coincides with other GP studies which also suggest that adding females to the reference population does not contribute a great deal to the accuracy of GPs (Cooper et al., 2014; Mucha et al., 2015).

When the animals have a sire recorded in the pedigree that has at least five progeny records, the average prediction accuracy of both models was greater than the accuracies obtained when the dam was recorded in the pedigree and had her own lactation record. This difference in accuracy suggests that most of the information is captured by the males present in the reference population, rather than the females. Although actual sires were not included in the reference population, the link between the sires and their progeny provides greater benefit to the predictions compared to the information provided by the dams with lactation records.

The prediction accuracy of EBVs and GBVs when animals have a sire recorded that has at least five progeny, was similar to corresponding accuracies when both the sire and dam was known. This suggests that animals in the validation population that have recorded sires linked to the reference population, provides as much information to the prediction of breeding values as the animals with both the sire and dam recorded.

Accuracy of predictions are important in livestock genetic improvement programmes as this gives confidence of a reliable estimate of the individuals true breeding values. If the accuracy is low, there is greater risk that the EBVs and GBVs are not reliable and there is more chance of selecting a genetically dud animal. The accuracy of GP largely depends the size of the reference population (Daetwlyer et al., 2012), the relationship between animals in the reference population and the target animals to be predicted (Clark et al., 2012), the LD between the SNPs and QTL and the distribution of the QTL effects (Hayes et al., 2009), the heritability of the trait and of course the prediction method used. All of which can be changed to improve accuracies.

Unlike the dairy cattle industry that has high accuracies due to a well-established recording system and large reference populations (Harris et al., 2008), the dairy goat population is significantly smaller and pedigree records are more often incomplete. The more information provided in the genetic evaluation, the greater the accuracy, therefore, as animals are included in the reference population these accuracies should increase. Due to Mendelian sampling, the maximum reliability from additional information from siblings is constrained to 0.25 for half-sibs and 0.5 for full-sibs. To achieve greater reliabilities like the dairy cattle industry, progeny testing and / or genomic information is required. Likewise, the lower accuracies obtained in this study for the across-breed EBVs and GBVs could be due to an insufficient number of genotyped and phenotyped animals in the reference population required to accurately represent all breeds in the validation population. Furthermore, this study used a medium-density SNP chip which limited the LD between SNP and QTL, and consequently limited the prediction accuracies. However, using a denser SNP chip or whole-genome-sequencing could increase the LD, providing greater accuracy of across-breed GPs. Additionally, reducing environmental or residual components will

increase the heritability of the trait, since the heritability is the proportion of phenotypic variation attributed to genetic variation. This could be achieved by adjusting for covariates that explain part of the environmental factors in the analysis. Lastly, the prediction model fitted in this study was a single-step BayesC approach, which is different to the ssGBLUP approaches generally used in GP studies of milk traits in dairy goats (Carillier et al., 2014; Mucha et al., 2015; Molina et al., 2018; Teissier et al., 2018). Currently, dairy goat populations are relatively small and therefore the ssGBLUP approach provides an efficient process for obtaining GBVs. In this population there is a large QTL known to have a significant effect on milk production in dairy goats. While it is true that ssGBLUP assumes a normal distribution of marker effects, this approach can accommodate different weightings for different loci (Teissier et al., 2018). However, Bayesian methods use a prior allowing for genes of moderate to large effect, therefore fitting a mixture model such as BayesC seemed appropriate. Even though the prediction accuracies using this ssBC model were lower than those obtained in other studies using the ssGBLUP approaches, this could be due to the limited information available in the evaluation, rather than the model. The prediction accuracies obtained in the current study were based on a single herd however, these accuracies are expected to increase as the size of the reference population increases.

Without knowing the actual genotype of an individual and when the animals are young and have not yet produced their own records, the EBVs are based on the average of their parents EBVs. This was the method of genetic evaluation in animal breeding before the introduction of genomic technologies. Genomic breeding values were predicted to demonstrate the effect of including genomic information into the genetic evaluation system. The slope of standardised-GBV against standardised-EBV is used as a measure of genomic inflation. The expected value is 1, indicating that the GPs are on a similar scale as the EBVs i.e. not inflated of deflated. In this study, the regression coefficients obtained for FY, PY and SCS were less than 1 (0.85 to 0.93), indicating GBVs of these traits are slightly inflated, and the regression coefficient for MY is 1.06, indicating slight deflation. Inaccurate prediction of GBVs could potentially lead to selection of the wrong candidates, for example, deflated GBVs would result in high producing animals being underestimated and the low

producing animals would be overestimated. Likewise, for inflated GBVs, high producing animals would be overestimated, and low producing animals would be underestimated. However, the regression coefficients obtained in this study give confidence that the standardised-GBVs are similar to the EBVs and given the improved accuracy of GBVs, reiterates the potential of including genomic information in the genetic evaluation of these traits.

Genetic evaluation of New Zealand dairy goats is performed annually and carried out using a PBLUP multi-trait animal model to produce EBVs of milk production traits using pedigree and phenotypic records. The dairy goat population in New Zealand consists of animals with various and often unknown breed compositions. The predominant breed is Saanen, but there are too few animals of other breeds to carry out effective within-breed evaluations. Thus, the industry currently uses an "across-breed" genetic evaluation enabling evaluation of all purebred and crossbred animals. Across-breed GBVs are calculated by including the breed effect. The breed effect can have a large influence on the accuracy of EBVs and GBVs in a multibreed genetic evaluation and will be important for ranking of animals. Although this data set is relatively small, the regression coefficients obtained for these breed groups illustrate the importance of breed effects. Likewise, J coefficients obtained for the difference between the genotyped founder and non-genotyped founder breeds ranged from -472.5 kg MY for ANTO to +101.67 kg MY for the "unknown" breed group. These coefficients are included in the across-breed GBVs for each animal where the value for non-genotyped animals will vary widely, depending on how closely related they are to the genotyped animals and their breed proportions of each breed group. Previously, GBVs were predicted based solely on their pedigree and genotypes, which led to overestimation as the genotyped animals were generally the most superior in the population. However, correcting for the difference between genotyped and non-genotyped founder breeds enables prediction of genetic merit in a population where selection is absent as the analysis is conditional on the data used for selection.

A single-step approach that includes both genotyped and non-genotyped animals would be recommended for the genetic evaluation of the New Zealand dairy goat population as the

inclusion of genomic information improved the accuracy of prediction of across-breed breeding values for all traits and for all scenarios. The accuracies obtained for the different scenarios demonstrate what farmers could expect with varying degrees of relationships to the reference animals. Prediction accuracies are important for farmers as this shows how reliable the estimation is and provides trust that the GBVs between selection candidates are consistent. Results from this study suggest GP is possible in the New Zealand dairy goat population however, this was based on a single-herd and requires further investigation before implementing for the entire population.

Conclusion

Including genomic information improved the prediction accuracy of breeding values compared to the pedigree-based BLUP method currently implemented in the New Zealand dairy goat industry. Prediction accuracies were slightly lower than other populations, but these accuracies are expected to increase as more animals enter the reference population. The use of a higher density SNP chip or whole-genome-sequencing would increase the extent of LD which would improve accuracies of across-breed GPs. Although this genomic evaluation was of a single New Zealand dairy goat herd, the inclusion of genomic information would enable prediction of GBV for all animals, even those without known pedigree or phenotypic records, which would benefit the New Zealand dairy goat population.

Chapter 8

Overall discussion and conclusion

8.1 Review of important findings

No formal breeding program exists in the New Zealand dairy goat industry, and as a result, the national genetic improvement of dairy goats is stagnant. Although only a small proportion of the New Zealand dairy goat population has been genotyped to date, the inclusion of genomic information into a single-step genomic evaluation will enable prediction of genomic breeding values (GBVs) of all genotyped individuals and their recorded relatives. This will allow ranking of selection candidates at a young age, and if these rankings are used for selection will provide real genetic progress. Therefore, implementing single-step genomic evaluation in a breeding program for the New Zealand dairy goat industry would provide an opportunity to increase quantity and composition of milk produced in the New Zealand dairy goat industry.

In order to achieve the rate of genetic gain offered by genomic selection there needs to be a well-defined breeding program. There is a logical process to the development of a breeding program and this thesis investigated a number of these aspects. This discussion covers the genetic parameters of traits of interest and their suitability of inclusion in the breeding objective. Genomic studies identified significant regions on the goat genome that influence the milk traits and a prototype single-step genomic evaluation model was developed. Important aspects required to successfully implement genomic evaluation in this population were highlighted including: re-defining the breeding objective, considering traits other than production, establishing a database and improving pedigree records, and managing the level of inbreeding, were discussed. Last of all, this information was consolidated into a summary for the Dairy Goat Cooperative (NZ) Ltd (DGC) to successfully implement genomic evaluation in the New Zealand dairy goat population.

8.1.1 Estimation of genetic parameters (milk traits)

The estimation of genetic parameters is an essential step to develop an effective breeding program (Harris et al., 1984). Genetic parameters such as heritability, repeatability, (co)variances, phenotypic and genetic correlations of traits are estimated to assess the

sources of variation and to evaluate relationships between traits of interest. These genetic parameters are specific to the population in which they are estimated as they can change due to selection, migration of genes from one population to another, or changing environmental conditions (van der Werf and de Boer, 1989). Estimates of the heritability and phenotypic variances for total lactation yields of milk (MY), fat (FY) and protein (PY) and somatic cell score (SCS) were reported in Chapter 3 and suggest that most traits are under moderate genetic control and show sufficient phenotypic variation to achieve reasonable genetic gains through selection. Favourable genetic correlations between these traits support the use of an economic index which appropriately weights the traits in order to maximise the economic response for the farmers.

8.1.2 Estimation of genetic parameters (lactation curves)

Currently the New Zealand genetic evaluation relies on a two-step process based on a first step of combining test-day records to phenotypically predict total lactation yields. However, a random regression test-day model that considers sample day records directly in the analysis can account more precisely for environmental factors that could affect animals differently during the lactation (Schaeffer and Dekkers, 1994). In addition, this test-day model can be used with incomplete lactation records (Freeman, 1998). This model was used to estimate genetic parameters of daily MY, FY, PY and SCS in New Zealand dairy goats. Results from this study are important to the farmer as the average shape of the lactation curve provides the predicted level of production over the lactation period, enabling the farmer to make informed management decisions such as feeding, breeding and economic management. Lactation curves obtained for individual animals provide farmers insight into the health status of the animal during the lactation process and the environmental effects affecting its milk production (Hossein-Zadeh, 2016). The genetic parameters of the lactation curves provide insight into the genetic associations between the traits at different stages of the lactation and enable the estimation of breeding values to select for improving milk traits over the whole lactation. Heritability estimates obtained for the lactation curves (Chapter 4) indicate sufficient genetic variability to make genetic progress for these traits. The results

from this analysis were similar to those obtained in dairy goat populations in Brazil, Spain, Norway, Germany, Thailand which have already adopted this test-day approach in their genetic evaluations (Andonov et al., 2007; Zumbach et al., 2008; Menéndez-Buxadera et al., 2010; Irano et al., 2015; Thepparat et al., 2015; Oliveira et al., 2016; Brito et al., 2017b). Moving to a random regression test-day model would provide more accurate estimates for each individual and selection programs could be devised to exploit the genetic variation of the lactation period.

8.1.2.1 Potential for extended lactations

Extending the lactation period in dairy goats would enable a continual supply of milk to producers without producing potentially unwanted kids. In addition, extending the lactation period would reduce the metabolic stress related to negative energy balance during early lactation (Knight, 1997). However, goats have a narrow seasonal breeding pattern, making it difficult to achieve year-round dairy production (Desire et al., 2017). In France, attempts to extend the lactation period in goats led to a decrease in milk yield across the lactation, and the high genetic merit females had fewer opportunities to contribute high-merit herd replacements (Desire et al., 2017). Conversely, in Holland and Spain, an extended lactation period was successfully adopted in dairy goats, enabling goats to be milked consistently for 2 to 7 years without significant losses in milk yield (Salama et al., 2005; Schuiling, 2007). In the dataset used in this thesis many animal lactations were identified as extended, suggesting some dairy goat farmers have already adopted the practice of extending the lactation period of their does. The random regression test-day model introduced in Chapter 4 could easily be extended to estimate genetic parameters and breeding values to enable genetic evaluation of these extended lactation traits (Portolano et al., 2001). The shape of the lactation curves obtained in Chapter 4 were explored to examine the potential for selecting for extended lactations in the New Zealand dairy goat population. Results showed that heritability estimates of daily yields were greatest between days 150 and 250, indicating that production during mid-lactation is more influenced by the genetic effects of the individual and less by environmental factors. With this in mind, genetic correlations

between different attributes of the lactation curve (peak yield, day at peak and persistency) and total lactation yields should be estimated to understand the relationship between these traits, as selecting for desired characteristics of the lactation curve could negatively influence yield traits (Ferris et al., 1985). Extending the lactation period could be a useful strategy for simplifying herd management and would mitigate production of surplus males and should be seriously considered for inclusion within the New Zealand breeding scheme.

8.1.3 Heritability of survival

Milk traits are currently included in the annual genetic evaluation as the DGC is focused on producing milk with high total milk solids and low bacterial count for the manufacture of high-quality products. However, placing too much emphasis on production, whilst neglecting other traits may result in undesirable consequences on the health and fertility of animals, which decreases longevity (Oltenacu and Broom, 2010). Longevity is an important trait for increasing the overall economic efficiency of a dairy goat farm as it results in an older age structure of the herd, leading to greater milk production within the herd and reduced replacement costs (Serradilla et al., 1997; Castañeda-Bustos et al., 2017). Therefore, longevity should be considered in the current genetic evaluation system. In Chapter 5 the estimated heritability of longevity in this population was reported to be low (0.07), but the coefficients of variation ranged from 43 to 45 indicating useful levels of phenotypic variation that could be exploited by its inclusion in a breeding program. Although longevity-type traits are not currently included in breeding objectives in dairy goats (Castañeda-Bustos et al., 2017), they are exploited in the breeding program of New Zealand dairy cattle and have a similar heritability value of 0.06 (DairyNZ, 2020a). An economic value for longevity in New Zealand already exists which would enable easy inclusion of this trait into a selection index (Solis-Ramirez et al., 2018), but further work is required to quantify the genetic and phenotypic correlations with other traits currently included in the index.

8.1.4 Genome-wide association studies

We performed the first genome-wide association study (GWAS) of dairy goats in New Zealand (Chapter 6) using single nucleotide polymorphism (SNP) genotypes obtained from 3,732 animals using the Caprine 50K SNP chip (Illumina Inc., San Diego, CA, USA). A genomic region on chromosome 19 was significantly associated with MY, FY, PY and SCS and a region on chromosome 29 was associated with SCS. It is possible the quantitative trait locus (QTL) on chromosome 19 has major pleiotropic effects in dairy goats as it was also significantly associated with type traits, udder morphology, functional longevity and semen production in the French dairy goat population (Palhiere et al., 2014, Martin et al., 2018; Oget et al., 2018; Palhiere et al., 2018). Although the haplotype frequencies obtained in the New Zealand population suggest that the major QTL on chromosome 19 is not yet fixed in the New Zealand dairy goat population, it exhibits undesirable pleiotropic effects on milk production and udder traits in French dairy goats (Martin et al., 2018). Genetic markers affecting gene function, in high linkage disequilibrium with genes, or known to be causal mutations can be fitted in the model as fixed effects to improve the accuracy of genomic predictions (Xu et al., 2020). Further research is recommended to distinguish whether these negative pleiotropic effects occur within the current population before implementing marker-assisted selection. If further analysis identifies favourable effects on important traits, then results from the GWAS could be used in selection programs for implementing marker-assisted-selection. Alternatively, genomic prediction captures all QTL across the genome and therefore, would account for both positive and negative effects, providing the greatest benefit over marker-assisted-selection (Xu et al., 2020).

8.1.5 Estimation of genomic breeding values

Currently, the genetic evaluation of New Zealand dairy goats relies on a multi-step process and high-quality pedigree records to estimate breeding values (EBVs). These multi-step prediction models require pseudo-phenotypes such as de-regressed breeding values (Garrick et al., 2009), which require preliminary evaluation of the performance of the buck's

progeny, referred to as a progeny-test. The New Zealand dairy goat industry have poor pedigree records and no progeny test scheme, therefore the reliabilities of bucks EBVs are low. The benefits of a single-step genomic evaluation will enable more accurate estimates of GBVs of bucks and the ability to accurately evaluate bucks at a very young age, instead of waiting for them to be evaluated through progeny testing of their daughters, will achieve faster genetic gains on sex-limited traits.

In Chapter 7, a prototype genomic prediction evaluation of a single-herd proved that across-breed genomic prediction could be implemented in the multi-breed dairy goat population in New Zealand. A single-step BayesC (ssBC) model that included phenotypic, pedigree and genotypic information from genotyped and non-genotyped animals was fitted to predict GBVs for a single dairy goat herd. Prediction accuracies of GBVs were significantly greater than prediction accuracies of EBVs obtained from a pedigree-based BLUP method which is currently implemented in the New Zealand dairy goat industry that uses phenotypes and pedigree records (Singireddy et al., 1997). Including genomic information substantially increased prediction accuracy within a single herd and is expected to provide even greater benefits for the rest of the population that has even more animals with missing pedigree records than the herd evaluated. The genomic prediction model used for this single herd can be applied to the wider New Zealand dairy goat population and with the increased animals in the training population, the prediction accuracies are expected to increase.

This chapter demonstrates the benefits of the single-step prediction model compared to the current evaluation, and adopting this model would put the New Zealand dairy goat industry in a very good position to implement genomic selection.

8.2 Important aspects to improve the New Zealand dairy goat industry

A few areas that the DGC needs to focus on in order to successfully implement genomic selection were identified in this thesis and discussed in the sections below. These include; re-defining a breeding objective, considering traits other than production, improving pedigree records and managing inbreeding.

8.2.1 Re-defining the breeding objective

Although this thesis focusses on improving the quantity and quality of goat milk in New Zealand, the potential rate of genetic gains possible through the use of genomic information also introduces great risk. Before implementing genomic prediction, great care must be taken to decide on the desired direction of genetic gain. The first step in a breeding program is to define the breeding goal which states the desired direction of improvement from the breeding program (Groen, 2000). However, the most important decision in the design of a breeding program, is the definition of the breeding objective (James, 1982). This requires identification of traits that influence the breeding goal and the relative importance of each trait (Garrick and Fernando, 2014). If too much emphasis is put on a trait, then genetic gains will be achieved at a rapid rate, but in the wrong direction. Therefore, the DGC needs to take care in re-defining a clear breeding objective. For example, the national breeding objective in the New Zealand dairy cattle industry is Breeding Worth (BW). The traits and relative emphasis of each trait on the breeding objective includes milk fat (24%), protein (17%), milk volume (13%), live weight (11%), fertility (13%), somatic cell score (6%), residual survival (9%) and body condition score (7%) (DairyNZ, 2020b). Thus, the BW index is currently putting the greatest emphasis on increasing milk fat while also ensuring genetic gain is not detrimentally increasing other traits. To achieve genetic improvement towards a more sustainable dairy goat industry, it is important to broaden the breeding objective to include traits other than production, similar to the dairy cattle industry. For dairy goat farming the traits currently perceived as being of primary importance are milk yield, body size, fertility, growth rate and disease tolerance (Bett et al., 2009), length of lactation, reproductive traits (Lopes et al., 2013) and udder morphology traits (Martin et al., 2018). Incorporating live weight and feed efficiency into a genetic improvement program could reduce feed costs per unit of output. In addition to the enormous potential for reducing costs to producers, genetic improvement of live weight and feed efficiency also have positive implications from an environmental sustainability standpoint. Therefore, live weight, longevity, fertility and feed efficiency are important traits that should be considered in the breeding objective of the New Zealand dairy goat industry.

Once the breeding objective is clearly defined, genetic parameters and the relative importance of each trait (economic values) can be estimated and a selection index that includes the traits and economic weights can be implemented.

8.2.2 Including traits other than production

The current genetic evaluation produces EBVs for MY, FY, PY and SCS. The DGC uses an index that weights each of these traits with respective economic values. With the majority of goat milk being sold as infant formula, the index is currently focused on increasing PY. However, placing too much emphasis on production whilst neglecting other traits will result in undesirable consequences on the health and fertility of animals. For example, in the French dairy goat population, MY has an antagonistic association with udder type traits (Manfredi et al., 2001) and the highly successful selection of milk production led to a deterioration in udder shape (Martin et al., 2018). Five udder type traits found to explain 80% of the genetic variability of udder and teat morphology (Clément et al., 2006) have since been included in the selection index to simultaneously improve milk production and udder shape (Martin et al., 2018). To avoid issues like this occurring in the New Zealand dairy goat population, the DGC is now moving towards the inclusion of other traits such as live weight, longevity and fertility into the selection index. The genetic parameters required for inclusion in the index were estimated for longevity in Chapter 5 but could not be obtained for live weight or fertility due to lack of records. Also, economic values have already been developed for longevity and live weight, but one is required for fertility. Other traits that should be considered and are currently included in other dairy goat indices around the world include feed efficiency (Desire et al., 2018), mammary health traits (Manfredi et al., 2001; Martin et al., 2018), disease tolerance (Berton et al., 2017) and fertility (Desire et al., 2017).

If the goal of the DGC is to improve the quality of milk to produce leading infant formulas, it would be beneficial to invest in understanding the composition of goat milk and the possibility of estimating genetic parameters for these elements and potentially begin some form of selection to produce animals that produce milk more suited to infant formula. Key

issues would be determining principal milk constituents and knowledge of fatty acid composition of the goat milk in New Zealand. The development of automated infrared instruments (mid-infrared and near-infrared spectroscopy) enables rapid analysis of milk components, which has been adopted into the routine analysis of samples in the dairy cattle industry (Tiplady et al., 2020). With the infrared spectra, milk components important for infant formula can be predicted, and just like other phenotypes, GBVs can be developed.

8.2.3 Establishing a database and improving pedigree records

Once the breeding objective is clearly defined, the DGC must set up a strong foundation for recording and managing records. This will require an organised database and implementing appropriate recording techniques to ensure that sufficient high-quality pedigree and phenotypic data are collected and managed. Currently, pedigree-records in the New Zealand dairy goat population are incomplete with some pedigree errors. Inconsistencies were observed when the A-matrix (based on pedigree records) was compared to the G-matrix (based on genomic information). Additionally, the parents recorded in the pedigree file were often not confirmed through the genotypes, indicating parent misidentification. These pedigree errors can affect the genetic evaluation by introducing biases in the estimation of genetic parameters and EBVs (Bradford et al., 2019). For example, 10% of pedigree errors can reduce genetic progress by up to 4% (Jiménez-Gamero et al., 2006). In such cases, genomic tools are advantageous in optimising breeding programs by verifying or assigning parentage (Talenti et al., 2018). Genomic information can be used to discover maternal parentage, specifically maternal grandsires and great-grandsires when the dam is not known, as well as assessing breed composition (Strucken et al., 2017), which is another inconsistency in the records of the New Zealand dairy goat population. A possible way to address the poor pedigree recording in this population is to use the genomic information provided by the genotyped animals. Although this is a small number of genotyped animals, information from genotypes will provide a lot more information about the relationships within the population which will increase the accuracy of predictions. Provided such analyses are limited to the genotyped animals, genotyping animals with poor pedigree

records would provide a more structured pedigree and substantially improve the accuracy of predictions.

High-quality pedigree and phenotype information is crucial for the success of a genetic evaluation system. Accurate phenotypes will reduce experimental errors and environmental effects, improving the estimation of heritabilities, and since heritability is a function of phenotypic variation, improve prediction accuracies. Accurately recorded data will result in more reliable genetic indexes, providing farmers with more accurate breeding values to make informed decisions when selecting candidates for breeding. Although recording phenotypes and pedigree is an extra task amongst an already busy schedule of work, increasing the number and quality of records will be a worthwhile investment in the long term (LIC, 2020).

In order to maximise the genetic improvement of the New Zealand dairy goat herd, increased data recording is essential. Three areas in which farmers can provide more data include the collection of pedigree records, fertility results and health events. Pedigree records are the most important piece of data that can be recorded on any animal as this is the link required in the genetic evaluation to obtain a breeding value based on performance records of the individual and its relatives. The full potential of genotyping will only be realised if it is paralleled by these high-quality phenotypes and a well-designed database.

8.2.4 Managing inbreeding

Due to the inadequate breeding structure in the dairy goat sector, inbreeding should be a serious concern for New Zealand dairy goat breeders. Traditional selection methods such as BLUP can lead to an increased rate of inbreeding per generation because the covariance between EBV of family members may be high, especially for young animals using EBVs derived from ancestral information (Clark et al., 2013). In addition, increasing the use of reproductive and genetic technologies such as (e.g., artificial insemination and multiple ovulation and embryo transfer) can increase rates of genetic gains but also increases selection intensity for females in breeding programs, which can significantly increase the

rates of inbreeding (Granleese et al., 2015). To ensure inbreeding is not unnecessarily increased, breeders should monitor inbreeding levels to limit the potential decrease in performance caused by inbreeding. Meuwissen (1997) introduced the Optimal Contribution Selection method that determines the optimal levels of the genetic contributions of selection candidates in an attempt to maximise the rate of genetic gain for a specified level of inbreeding. However, if genomic prediction is applied to the New Zealand dairy goat population, the rates of inbreeding per generation are expected to reduce, due to the GBVs explaining more of the mendelian sampling variation compared to EBVs (Daetwyler et al., 2007; Dekkers et al., 2007). Therefore, implementing genomic prediction in this population is expected to increase genetic gain of traits of interest, while maintaining the populations genetic diversity (Clark et al., 2013).

8.3 Summary for the New Zealand Dairy Goat Cooperative

The DGC needs to redefine the breeding objective, set up a genotyping program and establish a breeding structure to achieve the rapid genetic progress offered by genomic evaluation. Most importantly, the DGC should consider defining the breeding goal in terms of profit per kilogram of dry-matter intake. This goal will ensure the breeding program produces efficient animals and a sustainable dairy goat industry. The breeding objective should include traits that effect profitability such as lactation yields of milk, fat and protein, somatic cell score, live weight, longevity and fertility. Including traits other than production traits will ensure genetic gain does not negatively impact fitness traits such as health and fertility. The relative emphasis of each trait is crucial as this ensures the genetic gains occur at the desired level and in the desired direction. These relative weights depend on the priority of the DGC but should consider the long-term goal. With the breeding goal and objective clearly defined, the DGC must set up a strong foundation for storing, recording and managing the data. This step is paramount to the success of a breeding program and ensuring high-quality phenotypic data can be collected. The recording system should include the sire and dam, knowledge of the herd, contemporary group, sex, date of birth, age of the animal, and the genotypes. The DGC should put strong emphasis on the quality

of such information as inconsistencies in the database can introduce bias and low prediction accuracies. The DGC could establish an incentive to encourage farmers to collect accurate records and invest in genotyping, as these would benefit the interest of farmers and the DGC. For example, if the farm participates in genotyping, or, provides pedigree records with minimal errors, then they receive the GBVs for their herd, otherwise they will only receive the BLUP EBVs. The handling of the genetic information and database management is a vital aspect of a genetic improvement program. The DGC must decide on how the genetic information is managed and reported back to farmers. In addition, the DGC needs to establish a genotyping scheme and method of disseminating the superior genes to the rest of the population. The use of artificial insemination and optimum contributions selection would be a fast and safe (low risk of disease) method of disseminating superior genes while managing inbreeding.

8.4 General conclusion

The results of this thesis contribute in the design of the breeding program that will ensure the dairy goat industry increases the quantity and composition of goat milk produced in New Zealand. Genetic parameters and favourable genetic correlations between milk traits support the use of a selection index to predict the breeding objective. A random regression test-day model was developed, enabling more accurate estimates for each individual and the option of extending lactation traits. Longevity showed sufficient variation to warrant inclusion into the evaluation, but further work is required to quantify the genetic and phenotypic correlations with other traits currently included in the index. Genomic regions were identified on chromosome 19 for MY, FY, PY and SCS and on chromosome 29 for SCS, that can be exploited for a more desirable milk composition. A single-step genomic prediction model demonstrated that inclusion of genomic information into the evaluation can achieve significantly greater prediction accuracies than the traditional pedigree-based evaluation currently implemented in the New Zealand dairy goat industry. This single-step model will enable the evaluation of bucks at a younger age which can rapidly increase in the rate of genetic gain within the New Zealand dairy goat industry. However, the DGC should

re-define the breeding objective to ensure the genetic progress occurs in the right direction and should consider including traits other than production in the evaluation. To achieve the full potential of genomic evaluation, it is paramount that the DGC establishes a well-structured database and recording system to ensure high-quality pedigree and phenotypic records are maintained. Furthermore, the DGC should consider the risk of inbreeding when disseminating superior genetics.

Overall, the results presented in this thesis advance the knowledge required for the design of a breeding program using genomic information and provide a framework of statistical tools and steps required to implement genomic prediction in the New Zealand dairy goat industry. Including the single-step model in the breeding program will dramatically improve the quantity and composition of goat milk produced in New Zealand, enabling the DGC to remain competitive on the global stage.

Bibliography

Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. Journal of Dairy Science 93: 743-752.

Akey J, Jin L, Xiong M. 2001. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? European Journal of Human Genetics 9: 291-300.

Amin AA. 2017. Genetic and phenotypic correlation between somatic cell count and milk yield in Saudi dairy goats using random regression animal model. Journal of Animal and Plant Sciences 27: 1432-1439.

An X, Hou J, Zhao H, Zhu C, Yan Q, Song Y, Wang J, Cao B. 2012. Mutations in caprine DGAT1 and STAT5A genes were associated with milk production traits. Engineering 4: 30-34.

An X, Song Y, Hou J, Han P, Peng J, Zhang L, Wang J, Cao B. 2015. Mutations in the MTHFR gene and their associations with milk production traits in dairy goats. Small Ruminant Research 130: 76-80.

Analla M, Jiménez-Gamero I, Muñoz-Serrano A, Serradilla JM, Falagán A. 1996. Estimation of genetic parameters for milk yield and fat and protein contents of milk from Murciano-Granadina goats. Journal of Dairy Science 79: 1895-1898.

Andonov S, Ødegård J, Boman IA, Svendsen M, Holme IJ, Ådnøy T, Vukovic V, Klemetsdal G. 2007. Validation of test-day models for genetic evaluation of dairy goats in Norway. Journal of Dairy Science 90: 4863-4871.

Anniss F, McDougall S. 2000. Prevalence of bacterial infection and somatic cell count in early postpartum milking goats. Proceedings of the New Zealand Society of Animal Production 60: 284-286.

Apodaca-Sarabia CA, Lopez-Villalobos N, Blair HT, Prosser GG. 2009. Genetic parameters for somatic cell score in dairy goats estimated by random regression. Proceedings of the New Zealand Society of Animal Production 69: 206-209.

Arnal M, Robert-Granié C, Larroque H. 2018. Diversity of dairy goat lactation curves in France. Journal of Dairy Science 101: 11040-11051.

Aziz MA. 2010. Present status of the world goat populations and their productivity. Lohman Information 45: 42-52.

Bagnicka E, Łukaszewicz M, Adnoy T. 2016. Genetic parameters of somatic cell score and lactose content in goat's milk. Journal of Animal and Feed Sciences 25: 210-215.

Baloche G, Legarra A, Sallé G, Larroque H, Astruc J-M, Robert-Granié C, Barillet F. 2014. Assessment of accuracy of genomic prediction for French Lacaune dairy sheep. Journal of Dairy Science 97: 1107-1116.

Banks RG, van der Werf JHJ. 2009. Economic evaluation of whole genome selection, using meat sheep as a case study. Proceedings of the 18th Conference of the Association for the Advancement of Animal Breeding and Genetics, Adelaide, Australia 430-433.

Barillet F. 2007. Genetic improvement for dairy production in sheep and goats. Small Ruminant Research 70: 60-75.

Barrón-Bravo OG, Gutiérrez-Chávez AJ, Ángel-Sahagún CA, Montaldo HH, Shepard L, Valencia-Posadas M. 2013. Losses in milk yield, fat and protein contents according to different levels of somatic cell count in dairy goats. Small Ruminant Research 113: 421-431.

Becker D, Otto M, Ammann P, Keller I, Drögemüller C, Leeb T. 2015. The brown coat colour of Coppernecked goats is associated with a non-synonymous variant at the TYRP1 locus on chromosome 8. Animal Genetics 46: 50-54.

Bélichon S, Manfredi E, Piacère A. 1998. Genetic parameters of dairy traits in the Alpine and Saanen goat breeds. Genetics Selection Evolution 30: 529-534.
Bergonier D, De Crémoux R, Rupp R, Lagriffoul G, Berthelot X. 2003. Mastitis of dairy small ruminants. Veterinary research 34: 689-716.
Berton MP, Silva RMO, Peripolli E, Stafuzza NB, Martin JF, Álvarez MS, Gavinã BV, Toro MA, Banchero G, Oliveira PS. 2017. Genomic regions and pathways associated with gastrointestinal parasites resistance in Santa Inês breed adapted to tropical climate. Journal of Animal Science and Biotechnology DOI:10.1186/s40104-017-0190-4.
Bett RC, Kosgey IS, Kahi AK, Peters KJ. 2009. Analysis of production objectives and breeding practices of dairy goats in Kenya. Tropical Animal Health and Production 41: 307-320.
Bevilacqua C, Martin P, Candalh C, Fauquant J, Piot M, Roucayrol A, Pilla F, Heyman M. 2001. Goats' milk of defective α s1-casein genotype decreases intestinal and systemic sensitization to β-lactoglobulin in guinea pigs. Journal of Dairy Research 68: 217-227.
Biassus IO, Cobuci JA, Costa CN, Rorato PRN, Braccini Neto J, Cardoso LL. 2011. Genetic parameters for production traits in primiparous Holstein cows estimated by random regression models. Revista Brasileira de Zootecnia 40: 85-94.
Boichard D, Bouloc N, Ricordeau G, Piacere A, Barillet F. 1989. Genetic parameters for first lactation dairy traits in the Alpine and Saanen goat breeds. Genetics Selection Evolution 21: 205-215.
Boichard D, Guillaume F, Baur A, Croiseau P, Rossignol M-N, Boscher MY, Druet T, Genestout L, Colleau J-J, Journaux L. 2012. Genomic selection in French dairy cattle. Animal Production Science 52: 115-120.
Bouquet A, Juga J. 2012. Integrating genomic selection into dairy cattle breeding programmes: A review. Animal 7: 1-9.
Bradford HL, Masuda Y, Cole JB, Misztal I, VanRaden PM. 2019. Modeling pedigree accuracy and uncertain parentage in single-step genomic evaluations of simulated and US Holstein datasets. Journal of Dairy Science 102: 2308-2318.
Brito L, Schenkel F, Jafarikia M, Vermette S, Maignel L, Wyss S, Weaver K, Girouard S, Sullivan B. 2013. Goat herd improvement on productivity and health using genomics. DOI:10.13140/RG.2.1.1136.4723.
Brito LF, Silva FG, Melo ALP, Caetano GC, Torres RA, Rodrigues MT, Menezes GRO. 2011. Genetic and environmental factors that influence production and quality of milk of Alpine and Saanen goats. Genetics and Molecular Research 10: 3794-3802.
Brito LF, Kijas JW, Ventura RV, Sargolzaei M, Porto-Neto LR, Cánovas A, Feng Z, Jafarikia M, Schenkel FS. 2017a. Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. BMC genomics. DOI:10.1186/s12864-017-3610-0.
Brito LF, Silva FG, Oliveira HR, Souza NO, Caetano GC, Costa EV, Menezes GRO, Melo ALP, Rodrigues MT, Torres RA. 2017b. Modelling lactation curves of dairy goats by fitting random regression models using Legendre polynomials or B-splines. Canadian Journal of Animal Science 98: 73-83.
Brown DJ, Swan AA, Boerner V, Li L, Gurman PM, McMillan AJ, van der Werf JHJ, Chandler HR, Tier B, Banks RG. 2018. Single-step genetic evaluations in the Australian sheep industry. Proceedings of the World Congress on Genetics Applied to Livestock Production 11: 460.
Calus MPL, Veerkamp RF. 2007. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. Journal of Animal Breeding and Genetics 124: 362-368.
Calus MPL, de Roos APW, Veerkamp RF. 2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics 178: 553-561.

Calus MPL. 2010. Genomic breeding value prediction: methods and procedures. Animal 4: 157-164.

Carillier C, Larroque H, Palhière I, Clément V, Rupp R, Robert-Granié C. 2013. A first step toward genomic selection in the multi-breed French dairy goat population. Journal of Dairy Science 96: 7294-7305.

Carillier C, Larroque H, Robert-Granié C. 2014. Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population. Genetics Selection Evolution. DOI:10.1186/s12711-014-0067-3.

Castañeda-Bustos VJ, Montaldo HH, Torres-Hernández G, Pérez-Elizalde S, Valencia-Posadas M, Hernández-Mendo O, Shepard L. 2014. Estimation of genetic parameters for productive life, reproduction, and milk-production traits in US dairy goats. Journal of Dairy Science 97: 2462-2473.

Castañeda-Bustos VJ, Montaldo HH, Valencia-Posadas M, Shepard L, Pérez-Elizalde S, Hernández-Mendo O, Torres-Hernández G. 2017. Linear and nonlinear genetic relationships between type traits and productive life in US dairy goats. Journal of Dairy Science 100: 1232-1245.

Castel JM, Mena Y, Ruiz FA, Camúñez-Ruiz J, Sánchez-Rodríguez M. 2011. Changes occurring in dairy goat production systems in less favoured areas of Spain. Small Ruminant Research 96: 83-92.

Charfeddine N. 2000. Economic aspects of defining breeding objectives in selection programmes. Analysis and definition of the objectives in genetic improvement programmes in sheep and goats-an economic approach to increase their profitability. CIHEAM Series1-17.

Chen CY, Misztal I, Aguilar I, Legarra A, Muir WM. 2011. Effect of different genomic relationship matrices on accuracy and scale. Journal of Animal Science 89: 2673-2679.

Cheng H, Garrick DJ, Fernando RL. 2016. JWAS: Julia implementation of whole-genome analyses software using univariate and multivariate Bayesian mixed effects model. QTL. rocks. http://qtl.rocks/.

Christensen OF, Lund MS. 2010. Genomic prediction when some animals are not genotyped. Genetics Selection Evolution. DOI:10.1186/1297-9686-42-2.

Christensen OF, Madsen P, Nielsen B, Ostersen T, Su G. 2012. Single-step methods for genomic evaluation in pigs. Animal. DOI:10.1017/S1751731112000742.

Clark SA, Hickey JM, Daetwyler HD, van der Werf JHJ. 2012. The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Genetics Selection Evolution. DOI:10.1186/1297-9686-44-4.

Clark SA, Kinghorn BP, Hickey JM, van der Werf JHJ. 2013. The effect of genomic information on optimal contribution selection in livestock breeding programs. Genetics Selection Evolution. DOI:10.1186/1297-9686-45-44.

Clément V, Boichard D, Piacere A, Barbat A, Manfredi E. 2002. Genetic evaluation of French goats for dairy and type traits. Proceedings of the 7th World Congress on Genetics Applied to Livestock Production. Montpellier, France.

Clément V, Martin P, Barillet F. 2006. Elaboration d'un index synthétique caprin combinant les caractères laitiers et des caractères de morphologie mammaire. Renccontres Recherches Ruminants 13: 209-212.

Cole JB, VanRaden PM. 2018. Symposium review: Possibilities in an age of genomics: The future of selection indices. Journal of Dairy Science 101: 3686-3701.

Cooper TA, Wiggans GR, VanRaden PM. 2014. Including cow information in genomic prediction of Holstein dairy cattle in the US. Proceedings of the 10th World Congress of Genetics Applied to Livestock Production. Vancouver, Canada.

Daetwyler HD, Villanueva B, Bijma P, Woolliams JA. 2007. Inbreeding in genome-wide selection. Journal of Animal Breeding and Genetics 124: 369-376.

Daetwyler HD, Kemper KE, van der Werf JHJ, Hayes BJ. 2012. Components of the accuracy of genomic prediction in a multi-breed sheep population. Journal of Animal Science 90: 3375-3384.

Dagnachew BS, Thaller G, Lien S, Ådnøy T. 2011. Casein SNP in Norwegian goats: additive and dominance effects on milk composition and quality. Genetics Selection Evolution. DOI:10.1186/1297-9686-43-31.

DairyNZ. 2020a. Breeding values (BV) are an estimate of a cow or bull's genetic merit for a trait [Online]. Available: https://www.dairynz.co.nz/animal/animal-evaluation/interpreting-the-info/breeding-values/ [Accessed 5 August 2020].

DairyNZ. 2020b. Our national breeding objective (NBO) in New Zealand is to breed dairy cows that are able to efficiently convert feed into profit [Online]. Available: https://www.dairynz.co.nz/animal/animal-evaluation/interpreting-the-info/all-about-bw/ [Accessed 5 August 2020].

de los Campos G, Vazquez AI, Fernando R, Klimentidis YC, Sorensen D. 2013. Prediction of complex human traits using the genomic best linear unbiased predictor. PLOS Genetics. DOI:10.1371/journal.pgen.1003608.

de Rancourt M, Fois N, Lavín MP, Tchakérian E, Vallerand F. 2006. Mediterranean sheep and goats production: An uncertain future. Small Ruminant Research 62: 167-179.

de Roos APW, Hayes BJ, Spelman RJ, Goddard ME. 2008. Linkage disequilibrium and persistence of phase in Holstein Friesian, Jersey, and Angus cattle. Genetics 179: 1503-1512.

de Roos APW, Hayes BJ, Goddard ME. 2009. Reliability of genomic predictions across multiple populations. Genetics 183: 1545-1553.

Dekkers JCM. 2007. Prediction of response to marker-assisted and genomic selection using selection index theory. Journal of Animal Breeding and Genetics 124: 331-341.

Delfino RP, Selvaggi M, Celano GV, Dario C. 2011. Heritability estimates of lactation traits in Maltese goat. World Academy of Science, Engineering and Technology 78: 666-668.

Desire S, Mucha S, Coffey M, Mrode R, Broadbent J, Conington J. 2017. Pseudopregnancy and aseasonal breeding in dairy goats: genetic basis of fertility and impact on lifetime productivity. Animal. DOI:10.1017/S17517311170030561-8.

Desire S, Mucha S, Coffey M, Mrode RA, Broadbent J, Conington J. 2018. Deriving genomic breeding values for feed intake and body weight in dairy goats. Proceedings of the World Congress on Genetics Applied to Livestock Production 11.818.

Dettori ML, Rocchigiani AM, Luridiana S, Mura MC, Carcangiu V, Pazzola M, Vacca GM. 2013. Growth hormone gene variability and its effects on milk traits in primiparous Sarda goats. Journal of Dairy Research 80: 255-262.

Dettori ML, Pazzola M, Paschino P, Pira MG, Vacca GM. 2015a. Variability of the caprine whey protein genes and their association with milk yield, composition and renneting properties in the Sarda breed. 1. The LALBA gene. Journal of Dairy Research 82: 434-441.

Dettori ML, Pazzola M, Pira E, Puggioni O, Vacca GM. 2015b. Variability of the caprine whey protein genes and their association with milk yield, composition and renneting properties in the Sarda breed: 2. The BLG gene. Journal of Dairy Research 82: 442-448.

Dodds KG, Auvray B, Lee M, Newman SAN, McEwan JC. 2014. Genomic selection in New Zealand dual purpose sheep. Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, Vancouver, Canada.

Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, Tosser-Klopp G, Wang J, Yang S, Liang J. 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nature biotechnology 31: 135-141.

Dubeuf JP, Morand-Fehr P, Rubino R. 2004. Situation, changes and future of goat industry around the world. Small Ruminant Research 51: 165-173.

Duchemin SI, Colombani C, Legarra A, Baloche G, Larroque H, Astruc J-M, Barillet F, Robert-Granié C, Manfredi E. 2012. Genomic selection in the French Lacaune dairy sheep breed. Journal of Dairy Science 95: 2723-2733.

Ducrocq V. 1992. Solving animal model equations through an approximate incomplete Cholesky decomposition. Genetics Selection Evolution 24: 193-209.

Dulin AM, Paape MJ, Wergin WP. 1982. Differentiation and enumeration of somatic cells in goat milk. Journal of Food Protection 45: 435-439.

El-Saied UM, De la Fuente LF, Carriedo JA, San Primitivo F. 2005. Genetic and phenotypic parameter estimates of total and partial lifetime traits for dairy ewes. Journal of Dairy Science 88: 3265-3272.

El Hanafy AAM, Qureshi MI, Sabir J, Mutawakil M, Ahmed MMM, El Ashmaoui H, Ramadan HAMI, Abou-Alsoud M, Sadek MA. 2015. Nucleotide sequencing and DNA polymorphism studies of beta-lactoglobulin gene in native Saudi goat breeds in relation to milk yield. Czech Journal of Animal Science 60: 132-138.

Escareño L, Salinas-González H, Wurzinger M, Iñiguez L, Sölkner J, Meza-Herrera C. 2013. Dairy goat production systems. Tropical Animal Health and Production 45: 17-34.

Falconer DS, Mackay TFC. 1996. Introduction to quantitative genetics 4th edition. Harlow, UK: Longmans.

FAO. 2018. FAO statistical databases and data-sets. http://www.fao.org/faostat/en/#data.

Fernández C, Sánchez A, Garcés C. 2002. Modeling the lactation curve for test-day milk yield in Murciano-Granadina goats. Small Ruminant Research 46: 29-41.

Fernando RL, Garrick DJ. GenSel—User manual for a portfolio of genomic selection related analyses. Journal of Animal Breeding and Genetics, Iowa State University, Ames, Available at: http://bigs.ansci.iastate.edu/bigsgui/login.html (2008). [Accessed 26 June 2019].

Fernando RL, Garrick DJ. 2013. Bayesian methods applied to GWAS. In: Gondro C, van der Werf J, Hayes B (eds). Genome-wide association studies and genomic prediction.237-274. Springer.

Fernando RL, Dekkers JCM, Garrick DJ. 2014. A class of Bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. Genetics Selection Evolution. DOI:10.1186/1297-9686-46-50.

Fernando RL, Cheng H, Golden BL, Garrick DJ. 2016. Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. Genetics Selection Evolution. DOI:10.1186/s12711-016-0273-2.

Ferris TA, Mao IL, Anderson CR. 1985. Selecting for lactation curve and milk yield in dairy cattle. Journal of Dairy Science 68: 1438-1448.

Flores A, Mendoza G, Pinos-Rodriguez JM, Plata F, Vega S, Bárcena R. 2009. Effects of rumen-protected methionine on milk production of dairy goats. Italian Journal of Animal Science 8: 271-275.

Freeman AE. 1998. Dairy cattle breeding. Proceedings of the 6th World Congress on Genetics applied to livestock production 23: 293-294.

France Génétique Elevage. 2020. Effective breeding programmes - Dairy goats breeds selection [Online]. Available: http://en.france-genetique-elevage.org/Dairy-goats-breeds-selection/ [Accessed 26 August 2020].

García-Peniche TB, Montaldo HH, Valencia-Posadas M, Wiggans GR, Hubbard SM, Torres-Vázquez JA, Shepard L. 2012. Breed differences over time and heritability estimates for production and reproduction traits of dairy goats in the United States. Journal of Dairy Science 95: 2707-2717.

García-Ruiz A, Cole JB, VanRaden PM, Wiggans GR, Ruiz-López FJ, Van Tassell CP. 2016. Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. Proceedings of the National Academy of Sciences 113: 3995-4004.

Garrick DJ. 2007. Equivalent mixed model equations for genomic selection. Journal of Animal Science 85: 376-376.

Garrick DJ, Taylor JF, Fernando RL. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genetics Selection Evolution. DOI:10.1186/1297-9686-41-55.

Garrick DJ, Baumgard LH, Neibergs HL. 2012. Invited review: Genomic analysis of data from physiological studies. Journal of Dairy Science 95: 499-507.

Garrick DJ, Dekkers JCM, Fernando RL. 2014. The evolution of methodologies for genomic prediction. Livestock Science 166: 10-18.

Garrick DJ, Fernando RL. 2014. Genomic prediction and genome-wide association studies in beef and dairy cattle. In: The genetics of cattle. 474-501. CABI, Wallingford, UK.

Gautam M. 2012. Epidemiological study of removals in New Zealand dairy goats. Master of veterinary studies thesis, Massey University, Palmerston North, New Zealand.

Gautam M, Stevenson M, Lopez-Villalobos N, McLean V. 2017. Risk factors for culling, sales and deaths in New Zealand dairy goat herds, 2000 to 2009. Frontiers in Veterinary Science. DOI:10.3389/fvets.2017.00191.

Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT, Sargeant LS, Sorensen A, Steele MR, Zhao X. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 139: 907-920.

Geweke J. 1991. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Federal Reserve Bank of Minneapolis, Research Department Minneapolis.

Gianola D. 2013. Priors in whole-genome regression: the Bayesian alphabet returns. Genetics 194: 573-596.

Gilmour AR, Gogel BJ, Cullis BR, Thompson R, Butler D. 2009. ASReml user guide release 3.0. VSN International Ltd, Hemel Hempstead, UK.

Goddard ME. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. Genetica 136: 245-257.

Goddard ME, Hayes BJ. 2007. Genomic selection. Journal of Animal Breeding and Genetics 124: 323-330.

Goddard ME, Hayes BJ. 2012. Genome-wide association studies and linkage disequilibrium in cattle. In: Womack JE (ed). Bovine Genomics.192-210. Wiley-Blackwell.

Goddard ME, Kemper KE, MacLeod IM, Chamberlain AJ, Hayes BJ. 2016. Genetics of complex traits: prediction of phenotype, identification of causal polymorphisms and genetic architecture. Proceedings of the Royal Society B: Biological Sciences. DOI:10.1098/rspb.2016.0569.

Golden BL, Spangler ML, Snelling WM, Garrick DJ. 2018. Current single-step national beef cattle evaluation models used by the American Hereford Association and International Genetic Solutions, computational aspects, and implications of marker selection. Proceedings of the

Beef Improvement Federation 11th Genetic Prediction Workshop Refining Genomic Evaluation and Selection Indices 14-22.

Granleese T, Clark SA, Swan AA, van der Werf JHJ. 2015. Increased genetic gains in sheep, beef and dairy breeding programs from using female reproductive technologies combined with optimal contribution selection and genomic breeding values. Genetics Selection Evolution. DOI:10.1186/s12711-015-0151-3.

Gregory KE, Cundiff LVJ. 1980. Crossbreeding in beef cattle: evaluation of systems. Journal of Animal Science 51: 1224-1242.

Grisart B, Coppieters W, Farnir F, Karim L, Ford C, Berzi P, Cambisano N, Mni M, Reid S, Simon P. 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. Genome Research 12: 222-231.

Groen AF. 2000. Breeding goal definition. In: Galal S, Boyazoglu J, Hammond K (eds). Proceedings workshop on developing breeding strategies for lower input animal production environments 25-104. ICAR.

Guo P, Zhu B, Xu L, Niu H, Wang Z, Guan L, Liang Y, Ni H, Guo Y, Chen Y. 2017. Genomic prediction with parallel computing for slaughter traits in Chinese Simmental beef cattle using high-density genotypes. PLOS ONE. DOI:10.5061/dryad.4qc06.

Habier D, Fernando RL, Dekkers JCM. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177: 2389-2397.

Habier D, Tetens J, Seefried F-R, Lichtner P, Thaller G. 2010. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genetics Selection Evolution. DOI:10.1186/1297-9686-42-5.

Habier D, Fernando RL, Kizilkaya K, Garrick DJ. 2011. Extension of the Bayesian alphabet for genomic selection. BMC Bioinformatics. DOI:10.1186/1471-2105-12-186.

Habier D, Fernando RL, Garrick DJ. 2013. Genomic BLUP decoded: a look into the black box of genomic prediction. Genetics 194: 597-607.

Haenlein GFW. 2000. History of goat milk in America. Dairy Practices.

Haenlein GFW. 2001. Past, present, and future perspectives of small ruminant dairy research. Journal of Dairy Science 84: 2097-2115.

Haenlein GFW. 2004. Goat milk in human nutrition. Small Ruminant Research 51: 155-163.

Harris DL, Stewart TS, Arboleda CR. 1984. Animal breeding programs: A systematic approach to their design. Advances in Agricultural Technology 1-14.

Harris BL, Johnson DL, Spelman RJ, Sattler JD. 2008. Genomic selection in New Zealand and the implications for national genetic evaluation. Proceedings of the Interbull Meeting, Niagara Falls, Canada.

Harris BL, Johnson DL. 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. Journal of Dairy Science 93: 1243-1252.

Harris BL, Creagh FE, Winkelman AM, Johnson DL. 2011. Experiences with the Illumina high density bovine beadchip. INTERBULL bulletin.

Harris BL, Winkelman AM, Johnson DL. 2012. Large-scale single-step genomic evaluation for milk production traits. INTERBULL bulletin.

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. Journal of Dairy Science 92: 433-443.

Hayes BJ, Goddard ME. 2014. Prediction of breeding values using marker-derived relationship matrices. Journal of Animal Science 86: 2089-2092.

Hazel LN, Lush JL. 1942. The efficiency of three methods of selection. Journal of Heredity 33: 393-399.

Hazel LN. 1943. The genetic basis for constructing selection indexes. Genetics 28: 476-490.
Henderson CR. 1950. Estimation of genetic parameters. In: Biometrics 6: 186-187. International Biometric Society, Washington DC.
Henderson CR. 1963. Selection index and expected genetic advance. Statistical genetics and plant breeding 982: 141-163.
Henderson CR. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 423-447.
Henderson CR. 1984. Applications of linear models in animal breeding. University of Guelph, Guelph.
Hossein-Zadeh NG. 2016. Modelling lactation curve for fat to protein ratio in Holstein cows. Animal Science Papers and Reports 34: 233-246.
Hou JX, An XP, Song YX, Wang JG, Ma T, Han P, Fang F, Cao BY. 2013. Combined effects of four SNPs within goat PRLR gene on milk production traits. Gene 529: 276-281.
Hsu W, Garrick DJ, Fernando RL. 2017. The accuracy and bias of single-step genomic prediction for populations under selection. G3: Genes, Genomes, Genetics 7: 2685-2694.
Huitema NA. 2012. How nutritional and genetic factors affect New Zealand goat milk composition, Master of Science thesis, University of Waikato, Waikato, New Zealand.
Interbull. 2017. Description of national genetic evaluation systems. http://www.interbull.org/ib/geforms [Accessed 12 August 2020].
Irano N, Bignardi AB, Baldi F, Santana Jr ML, Teixeira IAMA, Tonhati H, Albuquerque LG. 2015. Multivariate analysis of test-day and total milk yield in goats. Genetics and Molecular Research 14: 13719-13727.
James JW. 1982. Economic aspects of developing breeding objectives: general considerations. Future developments in the genetic improvement of animals/edited by JSF Barker, Keith Hammond, AE McClintock.
Jandal JM. 1996. Comparative aspects of goat and sheep milk. Small Ruminant Research 22: 177-185.
Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y, Zhang Q. 2010. Genome wide association studies for milk production traits in Chinese Holstein population. PLOS ONE. DOI:10.1371/journal.pone.0013661.
Jiménez-Gamero I, Dorado G, Muñoz-Serrano A, Analla M, Alonso-Moraga A. 2006. DNA microsatellites to ascertain pedigree-recorded information in a selecting nucleus of Murciano-Granadina dairy goats. Small Ruminant Research 65: 266-273.
Kang H, Zhou L, Mrode R, Zhang Q, Liu JF. 2017. Incorporating the single-step strategy into a random regression model to enhance genomic prediction of longitudinal traits. Heredity 119: 459-467.
Kaplan EL, Meier P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association. 53: 457-481.

Kijas JW, Ortiz JS, McCulloch R, James A, Brice B, Swain B, Tosser-Klopp G. 2013. Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs. Animal Genetics 44: 325-335.
Kizilkaya K, Fernando RL, Garrick DJ. 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. Journal of Animal Science 88: 544-551.
Knight CH. 1997. Biological control of lactation length. Livestock Production Science 50: 1-3.
Knol EF, Nielsen B, Knap PW. 2016. Genomic selection in commercial pig breeding. Animal Frontiers 6: 15-22.

Koivula M, Strandén I, Pösö J, Aamand GP, Mäntysaari EA. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. Journal of Dairy Science 98: 2775-2784.

Krieg P, Rosenberger S, De Juanes S, Latzko S, Hou J, Dick A, Kloz U, van der Hoeven F, Hausser I, Esposito I. 2013. Aloxe3 knockout mice reveal a function of epidermal lipoxygenase-3 as hepoxilin synthase and its pivotal role in barrier formation. Journal of Investigative Dermatology 133: 172-180.

Lamberson B, Massey J, Whittier JC. 1993. Crossbreeding systems for small herds of beef cattle [Online]. University of Missouri Department of Animal Sciences Extension Factsheet. Available: https://extension2.missouri.edu/g2040 [Accessed 3 November 2018].

Lan XY, Chen H, Pan CY, Ming LJ, Lei CZ, Hua LS, Zhang CL, Hu SR. 2007. Polymorphism in growth hormone gene and its association with production traits in goats. Journal of Applied Animal Research 32: 55-60.

Lara-Villoslada F, Debras E, Nieto A, Concha A, Gálvez J, López-Huertas E, Boza J, Obled C, Xaus J. 2006. Oligosaccharides isolated from goat milk reduce intestinal inflammation in a rat model of dextran sodium sulfate-induced colitis. Clinical Nutrition 25: 477-488.

Larroque H, Barillet F, Baloche G, Astruc J, Buisson D, Shumbusho F, Clément V, Lagriffoul G. 2014. Toward genomic breeding programs in French dairy sheep and goats. Proceedings of the 10th World Congress on Genetics Applied to Livestock Production. Vancouver, Canada.

Le TH, Christensen OF, Nielsen B, Sahana G. 2017. Genome-wide association study for conformation traits in three Danish pig breeds. Genetics Selection Evolution. DOI:10.1186/s12711-017-0289-2.

Lee J, Cheng H, Garrick D, Golden B, Dekkers J, Park K, Lee D, Fernando R. 2017. Comparison of alternative approaches to single-trait genomic prediction using genotyped and non-genotyped Hanwoo beef cattle. Genetics Selection Evolution. DOI:10.1186/s12711-016-0279-9.

Legarra A, Aguilar I, Misztal I. 2009. A relationship matrix including full pedigree and genomic information. Journal of Dairy Science 92: 4656-4663.

Legarra A, Ducrocq V. 2012. Computational strategies for national integration of phenotypic, genomic, and pedigree data in a single-step best linear unbiased prediction. Journal of Dairy Science 95: 4629-4645.

Li X, Chen J, Deng M, Wang F, Du Z, Zhang J. 2011. Zinc finger protein HZF1 promotes K562 cell proliferation by interacting with and inhibiting INCA1. Molecular Medicine Reports 4: 1131-1137.

LIC. 2020. The value of genetic improvement [Online]. Available: https://www.lic.co.nz/news/value-genetic-improvement/ [Accessed 5 August 2020].

Liu Z, Seefried FR, Reinhardt F, Rensing S, Thaller G, Reents R. 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. Genetics Selection Evolution. DOI:10.1186/1297-9686-43-19.

Liu Z, Goddard ME, Reinhardt F, Reents R. 2014. A single-step genomic model with direct estimation of marker effects. Journal of Dairy Science 97: 5833-5850.

Lôbo AMBO, Lôbo RNB, Facó O, Souza V, Alves AAC, Costa AC, Albuquerque MAM. 2017. Characterization of milk production and composition of four exotic goat breeds in Brazil. Small Ruminant Research 153: 9-16.

Lopes FB, Borjas AR, da Silva MC, Facó O, Lôbo RN, Fiorvanti MCS, McManus C. 2012. Breeding goals and selection criteria for intensive and semi-intensive dairy goat system in Brazil. Small Ruminant Research 106: 110-117.

Lopes FB, Silva MC, Miyagi ES, Fioravanti MCS, Facó O, McManus C. 2013. Comparison of selection indexes for dairy goats in the tropics. Animal Sciences 35: 321-328.
Lopez-Villalobos N, Garrick DJ. 2001. Estimation of genetic parameters for lactation yields of milk, fat and protein of New Zealand dairy goats. Journal Animal Science 79: 70.
Lopez-Villalobos N, Garrick DJ. 2005. Methodology for the design and enhancement of genetic improvement programs illustrated in the context of the New Zealand dairy industry. Agrociencia 9: 553-568.
Lorenz AJ, Chao S, Asoro FG, Heffner EL, Hayashi T, Iwata H, Smith KP, Sorrells ME, Jannink J-L. 2011. Genomic selection in plant breeding: knowledge and prospects. In: Advances in agronomy. 110: 77-123. Elsevier.
Lush JL. 1937. Animal breeding plans. Iowa State College Press, Ames, Iowa.
Mahal Z, Khandoker MAMY, Haque MN. 2014. Effect of non genetic factors on productive traits of Black Bengal goats. Journal of the Bangladesh Agricultural University 11: 79-86.
Mai MD, Sahana G, Christiansen FB, Guldbrandtsen B. 2010. A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. Journal of Animal Science 88: 3522-3528.
Manfredi E, Piacere A, Lahaye P, Ducrocq V. 2001. Genetic parameters of type appraisal in Saanen and Alpine goats. Livestock Production Science 70: 183-189.
Maroteau C, Palhière I, Larroque H, Clément V, Tosser-Klopp G, Rupp R. 2013. QTL detection for traits of interest for the dairy goat industry. Proceedings of the 64th Annual Meeting of the European Federation of Animal Science. Nantes, France.
Maroteau C, Palhière I, Larroque H, Clément V, Ferrand M, Tosser-Klopp G, Rupp R. 2014. Genetic parameter estimation for major milk fatty acids in Alpine and Saanen primiparous goats. Journal of Dairy Science 97: 3142-3155.
Martin PM, Palhière I, Ricard A, Tosser-Klopp G, Rupp R. 2016a. Genome wide association study identifies new loci associated with undesired coat color phenotypes in Saanen goats. PLOS ONE. DOI:10.1371/journal.pone.0152426.
Martin PM, Palhière I, Tosser-Klopp G, Rupp R. 2016b. Heritability and genome-wide association mapping for supernumerary teats in French Alpine and Saanen dairy goats. Journal of Dairy Science 99: 8891-8900.
Martin PM, Palhière I, Maroteau C, Bardou P, Canale-Tabet K, Sarry J, Woloszyn F, Bertrand-Michel J, Racke I, Besir H. 2017. A genome scan for milk production traits in dairy goats reveals two new mutations in Dgat1 reducing milk fat content. Scientific Reports. DOI:10.1038/s41598-017-02052-0.
Martin PM, Palhière I, Maroteau C, Clément V, David I, Tosser-Klopp G, Rupp R. 2018. Genome-wide association mapping for type and mammary health traits in French dairy goats identifies a pleiotropic region on chromosome 19 in the Saanen breed. Journal of Dairy Science 101: 5214-5226.
Mavrogenis AP, Constantinou A, Louca A. 1984. Environmental and genetic causes of variation in production traits of Damascus goats 2. Goat productivity. Animal Production 38: 99-104.
Mavrogenis AP, Papachristoforou C, Lysandrides P, Roushias A. 1989. Environmental and genetic effects on udder characteristics and milk production in Damascus goats. Small Ruminant Research 2: 333-343.
McCarthy M, Abecasis G, Cardon L, Goldstein D, Little J, Ioannidis J, Hirschhorn J. 2008. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature Reviews Genetics 9: 356-369.

McLaren A, Mucha S, Mrode R, Coffey M, Conington J. 2016. Genetic parameters of linear conformation type traits and their relationship with milk yield throughout lactation in mixed-breed dairy goats. Journal of Dairy Science 99: 5516-5525.

McMillan AJ, Swan AA. 2017. Weighting of genomic and pedigree relationships in single step evaluation of carcass traits in Australian sheep. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 22: 557-560.

Menéndez-Buxadera A, Molina A, Arrebola F, Gil MJ, Serradilla JM. 2010. Random regression analysis of milk yield and milk composition in the first and second lactations of Murciano-Granadina goats. Journal of Dairy Science 93: 2718-2726.

Meng Q, Wang K, Liu X, Zhou H, Xu L, Wang Z, Fang M. 2017. Identification of growth trait related genes in a Yorkshire purebred pig population by genome-wide association studies. Asian-Australasian Journal of Animal Sciences 30: 462-469.

Meredith BK, Kearney FJ, Finlay EK, Bradley DG, Fahey AG, Berry DP, Lynn DJ. 2012. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genetics. DOI:10.1186/1471-2156-13-21.

Meuwissen THE. 1997. Maximizing the response of selection with a predefined rate of inbreeding. Journal of Animal Science 75: 934-940.

Meuwissen THE, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-1829.

Meuwissen THE. 2003. Genomic selection: the future of marker assisted selection and animal breeding. Marker Assisted selection: a fast track to increase genetic gain in plants and animal breeding 54-59.

Meuwissen THE, Hayes BJ, Goddard ME. 2013. Accelerating improvement of livestock with genomic selection. The Annual Review of Animal Biosciences 1: 221-237.

Meuwissen THE, Hayes BJ, Goddard ME. 2016. Genomic selection: A paradigm shift in animal breeding. Animal Frontiers 6: 6-14.

Miglior F, Muir BL, Van Doormaal BJ. 2005. Selection indices in Holstein cattle of various countries. Journal of Dairy Science 88: 1255-1263.

Misztal I, Lourenco D, Legarra A. 2020. Current status of genomic evaluation. Journal of Animal Science. DOI:10.1093/jas/skaa101.

Moghaddar N, Swan AA, van der Werf JHJ. 2014. Comparing genomic prediction accuracy from purebred, crossbred and combined purebred and crossbred reference populations in sheep. Genetics Selection Evolution 46: 58.

Moioli B, D'Andrea M, Pilla F. 2007. Candidate genes affecting sheep and goat milk quality. Small Ruminant Research 68: 179-192.

Molina A, Munoz E, Diaz C, Menendez-Buxadera A, Ramon M, Sanchez M, Carabano MJ, Serradilla JM. 2018. Goat genomic selection: Impact of the integration of genomic information in the genetic evaluations of the Spanish Florida goats. Small Ruminant Research 163: 72-75.

Montaldo HH, Valencia-Posadas M, Wiggans GR, Shepard L, Torres-Vázquez JA. 2010. Short communication: Genetic and environmental relationships between milk yield and kidding interval in dairy goats. Journal of Dairy Science 93: 370-372.

Moore K, Moran A, Mrode R, Coffey M. 2018. Using commercial data and genomics to improve female fertility and calf survival of Limousin beef cattle in the UK. Proceedings of the World Congress on Genetics Applied to Livestock Production Species–Bovine (Beef). Auckland, New Zealand.

Morris CA, Wheeler M, Hosking BC, Watson TG, Hurford AP, Foote BJ, Foote JF. 1997. Genetic parameters for milk yield and faecal nematode egg count in Saanen does. New Zealand Journal of Agricultural Research 40: 523-528.

Morris CA, Wheeler M, Lanuzel M. 2006. Genetic trend and parameter estimates for milk yield traits and kidding date in a Saanen goat herd in New Zealand. New Zealand Journal of Agricultural Research 49: 175-181.

Mucha S, Mrode R, Coffey M, Conington J. 2014. Estimation of genetic parameters for milk yield across lactations in mixed-breed dairy goats. Journal of Dairy Science 97: 2455-2461.

Mucha S, Mrode R, MacLaren-Lee I, Coffey M, Conington J. 2015. Estimation of genomic breeding values for milk yield in UK dairy goats. Journal of Dairy Science 98: 8201-8208.

Mucha S, Mrode R, Coffey M, Kizilaslan M, Desire S, Conington J. 2018a. Genome-wide association study of conformation and milk yield in mixed-breed dairy goats. Journal of Dairy Science. DOI:10.3168/jds.2017-12919.

Mucha S, Desire S, Geddes L, Mrode R, Coffey M, Conington J. 2018b. Implementation of genomic selection in UK dairy goats. Proceedings of the 10th World Congress on Genetics Applied to Livestock Production, Vancouver, Canada.

Muller CJC. 2005. Genetic parameter estimation and breeding plans for the South African dairy goat herd. Stellenbosch: University of Stellenbosch.

Newman S, Morris CA, Baker RL, Nicoll GB. 1992. Genetic improvement of beef cattle in New Zealand: breeding objectives. Livestock Production Science 32: 111-130.

Nicoloso L, Bomba L, Colli L, Negrini R, Milanesi M, Mazza R, Sechi T, Frattini S, Talenti A, Coizet B, Chessa S, Marletta D, D'Andrea M, Bordonaro S, Ptak G, Carta A, Pagnacco G, Valentini A, Pilla F, Ajmone-Marsan P, Crepaldi P. 2015. Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. Genetics Selection Evolution. DOI:10.1186/s12711-015-0140-6.

Nilforooshan MA. 2020. Application of single-step GBLUP in New Zealand Romney sheep. Animal Production Science 60: 1136-1144.

Norman HD, VanRaden PM, Wright JR, Clay JS. 1999. Comparison of test interval and best prediction methods for estimation of lactation yield from monthly, a.m.–p.m., and trimonthly testing. Journal of Dairy Science 82: 438-444.

Oget C, Clément V, Palhiere I, Tosser-Klopp G, Fabre S, Rupp R. 2018. Genome-wide study finds a QTL with pleotropic effects on semen and production traits in Saanen goats. Proceedings of the 69th Annual Meeting of the European Federation of Animal Science. Dubrovnik, Croatia.

Olfati JA, Samizadeh H, Rabiei B. 2011. Griffing's methods comparison for general and specific combining ability in cucumber. Journal of Biometrics and Biostatistics. DOI:10.4172/2155-6180.1000122.

Oliveira HR, Silva FF, Siqueira OHGBD, Souza NO, Junqueira VS, Resende MDV, Borquis RRA, Rodrigues MT. 2016. Combining different functions to describe milk, fat, and protein yield in goats using Bayesian multiple-trait random regression models. Journal of Animal Science 94: 1865-1874.

Oltenacu PA, Broom DM. 2010. The impact of genetic selection for increased milk yield on the welfare of dairy cows. Animal Welfare 19: 39-49.

Orr N, Back W, Gu J, Leegwater P, Govindarajan P, Conroy J, Ducro B, Van Arendonk JA, MacHugh DE, Ennis S, Hill EW, Brama PA. 2010. Genome-wide SNP association-based localization of a dwarfism gene in Friesian dwarf horses. Animal Genetics. DOI:10.1111/j.1365-2052.2010.02091.

Paape MJ, Poutrel B, Contreras A, Marco Juan C, Capuco AV. 2001. Milk somatic cells and lactation in small ruminants. Journal of Dairy Science 84: 237-244.

Paape MJ, Wiggans GR, Bannerman DD, Thomas DL, Sanders AH, Contreras A, Moroni P, Miller RH. 2007. Monitoring goat and sheep milk somatic cell counts. Small Ruminant Research 68: 114-125.

Palhière I, Larroque H, Clément V, Tosser-Klopp G, Rupp R. 2014. Genetic parameters and QTL detection for milking speed in dairy Alpine and Saanen goats. Proceedings of the 10th World Congress on Genetics Applied to Livestock Production. Vancouver, Canada.

Palhière I, Oget C, Rupp R. 2018. Functional longevity is heritable and controlled by a major gene in French dairy goats. Proceedings of the 11th World Congress on Genetics Applied to Livestock Production. Auckland, New Zealand.

Pérez-Razo M, Sánchez F, Torres-Hernández G, Becerril-Pérez C, Gallegos-Sánchez J, González-Cosío F, Meza-Herrera C. 2004. Risk factors associated with dairy goats stayability. Livestock Production Science 89: 139-146.

Perreira JM, Aker AM, Savidis G, Chin CR, McDougall WM, Portmann JM, Meraner P, Smith MC, Rahman M, Baker RE, Gauthier A, Franti M, Brass AL. 2015. RNASEK is a V-ATPase-associated factor required for endocytosis and the replication of rhinovirus, influenza A virus, and dengue virus. Cell Reports 12: 850-863.

Piles M, Garreau H, Rafel O, Larzul C, Ramon J, Ducrocq V. 2006. Survival analysis in two lines of rabbits selected for reproductive traits. Journal of Animal Science 84: 1658-1665.

Portolano B, Montalbano L, Militi W. 2001. Genetic and environmental sources of variation for milk yield traits in Barbaresca siciliana breed. Small Ruminant Research 41: 195-202.

Price A, Patterson N, Plenge R, Weinblatt M, Shadick N, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics 38: 904-909.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Pryce JE, Bolormaa S, Chamberlain AJ, Bowman PJ, Savin K, Goddard ME, Hayes BJ. 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. Journal of Dairy Science 93: 3331-3345.

Pulina G, Milán MJ, Lavín MP, Theodoridis A, Morin E, Capote J, Thomas DL, Francesconi AHD, Caja G. 2018. Invited review: Current production trends, farm structures, and economics of the dairy sheep and goat sectors. Journal of Dairy Science 101: 6715-6729.

Rabasco A, Serradilla JM, Padilla JA, Serrano A. 1993. Genetic and non-genetic sources of variation in yield and composition of milk in Verata goats. Small Ruminant Research 11: 151-161.

Reber I, Keller I, Becker D, Flury C, Welle M, Drögemüller C. 2015. Wattles in goats are associated with the FMN1/GREM1 region on chromosome 10. Animal Genetics 46: 316-320.

Rendel JM, Robertson A. 1950. Estimation of genetic gain in milk yield by selection in a closed herd of dairy cattle. Journal of Genetics 50: 1-8.

Robertson A, Rendel JM. 1950. The use of progeny testing with artificial insemination in dairy cattle. Journal of Genetics 50: 21-31.

Robertson K, Symes W, Garnham M. 2015. Carbon footprint of dairy goat milk production in New Zealand. Journal of Dairy Science 98: 4279-4293.

Roldán DL, Rabasa AE, Saldaño S, Holgado F, Poli MA, Cantet RJC. 2008. QTL detection for milk production traits in goats using a longitudinal model. Journal of Animal Breeding and Genetics 125: 187-193.

Rota AM, Gonzalo C, Rodriguez PL, Rojas AI, Martín L, Tovar JJ. 1993. Effects of stage of lactation and parity on somatic cell counts in milk of Verata goats and algebraic models of their lactation curves. Small Ruminant Research 12: 211-219.

Ruiz-Rodriguez CT, Brandt JR, Oliverio R, Ishida Y, Guedj N, Garrett EF, Kahila Bar-Gal G, Nikolaidis N, Cardoso FC, Roca AL. 2017. Polymorphisms of the toll-like receptor 2 of goats (Capra hircus) may be associated with somatic cell count in milk. Animal Biotechnology 28: 112-119.

Rupp R, Clément V, Piacere A, Robert-Granié C, Manfredi E. 2011. Genetic parameters for milk somatic cell score and relationship with production and udder type traits in dairy Alpine and Saanen primiparous goats. Journal of Dairy Science 94: 3629-3634.

Salama AAK, Caja G, Such X, Casals R, Albanell E. 2005. Effect of pregnancy and extended lactation on milk production in dairy goats milked once daily. Journal of Dairy Science 88: 3894-3904.

Salama AAK, Caja G, Hamzaoui S, Badaoui B, Castro-Costa A, Facanha DAE, Guilhermino MM, Bozzi R. 2014. Different levels of response to heat stress in dairy goats. Small Ruminant Research 121: 73-79.

Sargent FD, Lytton VH, Wall OG. 1968. Test interval method of calculating dairy herd improvement association records. Journal of Dairy Science 51: 170-179.

Sarmento JLR, Albuquerque LG, Torres RA, Rodrigues MT, Lopes PS, Reis Filho JC. 2008. Comparação de modelos de regressão aleatória para estimação de parâmetros genéticos em caprinos leiteiros. Revista Brasileira de Zootecnia 37: 1788-1796.

Sato S, Uemoto Y, Kikuchi T, Egawa S, Kohira K, Saito T, Sakuma H, Miyashita S, Arata S, Kojima T. 2016. SNP-and haplotype-based genome-wide association studies for growth, carcass, and meat quality traits in a Duroc multigenerational population. BMC Genetics. DOI:10.1186/s12863-016-0368-3.

Schaeffer LR, Dekkers JCM. 1994. Random regressions in animal models for test-day production in dairy cattle. Proceedings of the World Congress of Genetics Applied Livestock Production 18: 443-446.

Schaeffer LR. 2006. Strategy for applying genome-wide selection in dairy cattle. Journal of Animal Breeding and Genetics 123: 218-223.

Scholtens MR, Smith RMC, Lopez-Lozano SR, Lopez-Villalobos N, Burt D, Harper L, Tuohy M, Thomas DG, Carr AJ, Gray DI, Tozer P, Schreurs NM. 2017. Brief communication: The current state of the New Zealand goat industry. Proceedings of New Zealand Society of Animal Production 77: 164-168.

Scholtens MR, Lopez-Villalobos N, Garrick DJ, Blair HT, Lehnert K, Snell RG. 2019. Genetic parameters for total lactation yields of milk, fat, protein, and somatic cell score in New Zealand dairy goats. Animal Science Journal. DOI:10.1111/asj.13310.

Scholtens MR, Jiang A, Smith A, Littlejohn M, Lehnert K, Snell RG, Lopez-Villalobos N, Garrick DJ, Blair HT. 2020. Genome-wide association studies of lactation yields of milk, fat, protein and somatic cell score in New Zealand dairy goats. Journal of Animal Science and Biotechnology. DOI:10.1186/s40104-020-00453-2.

Schuiling HJ. 2007. Duurmelken bij geiten = Prolonged lactations in goats. Animal Sciences Group, Wageningen University. https://edepot.wur.nl/42847.

Selvaggi M, Dario C. 2015. Genetic analysis of milk production traits in Jonica goats. Small Ruminant Research 126: 9-12.

Serenius T, Stalder KJ. 2004. Genetics of length of productive life and lifetime prolificacy in the Finnish Landrace and Large White pig populations. Journal of Animal Science 82: 3111-3117.

Serradilla JM, Sanchez-Palma A, Micheo JM. 1997. Preliminary analysis of the possibility of including longevity as a breeding goal of Maguena goats. In: Gabina D, Bodin L (eds). Data collection and definition of objectives in sheep and goat breeding programmes: New

prospects. 33: 207-219. Options Mediterraneennes. Serie A: Seminaires Mediterraneens (CIHEAM).

Silva FG, Torres RA, Brito LF, Euclydes RF, Melo ALP, Souza NO, Ribeiro Jr JI, Rodrigues MT. 2013. Random regression models using Legendre orthogonal polynomials to evaluate the milk production of Alpine goats. Genetics and Molecular Research 12: 6502-6511.

Silva RMO, Fragomeni BO, Lourenco DAL, Magalhães AFB, Irano N, Carvalheiro R, Canesin RC, Mercadante MEZ, Boligon AA, Baldi FS, Misztal I, Albuquerque LG. 2016. Accuracies of genomic prediction of feed efficiency traits using different prediction and validation methods in an experimental Nelore cattle population. Journal of Animal Science 94: 3613-3623.

Singireddy SR, Lopez-Villalobos N, Garrick DJ. 1997. Across-breed genetic evaluation of New Zealand dairy goats. Proceedings of the New Zealand Society of Animal Production 57: 43-45.

Slatkin M. 2008. Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics 9: 477-485.

Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE. 2008. Genomic selection using different marker types and densities. Journal of Animal Science 86: 2447-2454.

Solis-Ramirez J, Lopez-Villalobos N, Blair HT. 2011. Dairy goat production systems in Waikato, New Zealand. Proceedings of the New Zealand Society of Animal Production 71: 86-91.

Solis-Ramirez J. 2014. A genetic improvement programme for New Zealand dairy goats. PhD in Animal Science, Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand.

Solis-Ramirez J, Lopez-Villalobos N, Blair HT. 2018. Estimation of economic values for longevity and somatic cell score in New Zealand dairy goats. Proceedings of the World Congress on Genetics Applied to Livestock Production 11: 367.

Spelman RJ, Hayes BJ, Berry DP. 2013. Use of molecular technologies for the advancement of animal breeding: genomic selection in dairy cattle populations in Australia, Ireland and New Zealand. Animal Production Science 53: 869-875.

Spelman RJ, Tiplady K, Harland C, Winkelman AM, Sherlock RG, Keehan MD, Littlejohn MD. 2014. Application of large-scale sequence datasets for the discovery of genomic variations of economic importance in dairy cattle. Proceedings of the 10th World Congress of Genetics Applied to Livestock Production. Vancouver, Canada.

Stafford K, Prosser C. 2016. Goat production. In: Stafford K (ed). Livestock production in New Zealand: the complete guide to dairy cattle, beef cattle, sheep, deer, goats, pigs and poultry. Massey University Press, New Zealand.

Strandén I, Garrick DJ. 2009. Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. Journal of Dairy Science 92: 2971-2975.

Strucken EM, Al-Mamun HA, Esquivelzeta-Rabell C, Gondro C, Mwai OA, Gibson JP. 2017. Genetic tests for estimating dairy breed proportion and parentage assignment in East African crossbred cattle. Genetics Selection Evolution 49: 67-67.

Suontama M, Klápště J, Telfer E, Graham N, Stovold T, Low C, McKinley R, Dungey H. 2019. Efficiency of genomic prediction across two Eucalyptus nitens seed orchards with different selection histories. Heredity 122: 370-379.

SVS. SNP & Variation Suite ™ [Software]. Bozeman, MT: Golden Helix, Inc. http://www.goldenhelix.com.

Swan AA, Johnston DJ, Brown DJ, Tier B, Graser H. 2012. Integration of genomic information into beef cattle and sheep genetic evaluations in Australia. Animal Production Science 52: 126-132.

Tabbaa MJ, Al-Atiyat R. 2009. Breeding objectives, selection criteria and factors influencing them for goat breeds in Jordan. Small Ruminant Research 84: 8-15.

Talenti A, Palhière I, Tortereau F, Pagnacco G, Stella A, Nicolazzi EL, Crepaldi P, Tosser-Klopp G. 2018. Functional SNP panel for parentage assessment and assignment in worldwide goat breeds. Genetics Selection Evolution. DOI:10.1186/s12711-018-0423-9.

Taskinen M, Mäntysaari EA, Strandén I. 2017. Single-step SNP-BLUP with on-the-fly imputed genotypes and residual polygenic effects. Genetics Selection Evolution. DOI:10.1186/s12711-017-0310-9.

Teissier M, Larroque H, Robert-Granie C. 2018. Weighted single-step genomic BLUP improves accuracy of genomic breeding values for protein content in French dairy goats: a quantitative trait influenced by a major gene. Genetics Selection Evolution. DOI:10.1186/s12711-018-0400-3.

Thepparat M, Boonkum W, Duangjinda M, Tumwasorn S, Nakavisut S, Thongchumroon T. 2015. Genetic evaluation using random regression models with different covariance functions for test-day milk yield in an admixture population of Thailand goats. Animal Science Journal 86: 655-660.

Tiplady KM, Lopdell TJ, Littlejohn MD, Garrick DJ. 2020. The evolving role of Fourier-transform mid-infrared spectroscopy in genetic improvement of dairy cattle. Journal of Animal Science and Biotechnology 11: 1-13.

Torrero GY. 2010. Estimación de covarianzas para características de longevidad y producción en cabras. Universidad de Guanajuato, Irapuato, Gto. México.

Torres-Vázquez JA, Valencia-Posadas M, Castillo-Juárez H, Montaldo HH. 2009. Genetic and phenotypic parameters of milk yield, milk composition and age at first kidding in Saanen goats from Mexico. Livestock Science 126: 147-153.

Tosser-Klopp G, Bardou P, Bouchez O, Cabau C, Crooijmans R, Dong Y, Donnadieu-Tonon C, Eggen A, Heuven HCM, Jamli S. 2014. Design and characterization of a 52K SNP chip for goats. PLOS ONE. DOI:10.1371/journal.pone.0086227.

Valencia-Posadas M, Torrero-Garza Y, Vicencio-Reyes CV, Shepard L, Montaldo HH. 2010. Phenotypic relationships between characteristics of conformation with the ability of permanence to the 36 months in Alpine goats. Acta Universitaria 20: 40-44.

Valencia-Posadas M, Torrero-Garza Y, Torres-Vázquez JA, Ángel-Sahagún CA, Gutiérrez-Chávez AJ, Shepard L, Montaldo HH. 2017. Genetic parameters for functional stayability to 24 and 36 months of age and first lactation milk yield in dairy goats. Small Ruminant Research 149: 209-213.

Valencia M, Doble J, Arbiza SI. 2002. Sources of environmental variation affecting lactation and pre-weaning growth characteristics in Saanen goats. Cuban Journal of Agricultural Science 36: 117-122.

Valencia M, Dobler J, Montaldo HH. 2007. Genetic and phenotypic parameters for lactation traits in a flock of Saanen goats in Mexico. Small Ruminant Research 68: 318-322.

van der Werf JHJ, de Boer W. 1989. Estimation of genetic parameters in a crossbred population of black and white dairy cattle. Journal of Dairy Science 72: 2615-2623.

Van Vleck LD. 1993. Selection index and introduction to mixed model methods. CRC Press, Boca Raton.

VanRaden PM, Wiggans GR. 1991. Derivation, calculation, and use of national animal model information. Journal of Dairy Science 74: 2737-2746.

VanRaden PM, Klaaskate EJH. 1993. Genetic evaluation of length of productive life including predicted longevity of live cows. Journal of Dairy Science 76: 2758-2764.

VanRaden PM. 2007. Genomic measures of relationship and inbreeding. INTERBULL bulletin 33-33.

VanRaden PM. 2008. Efficient methods to compute genomic predictions. Journal of Dairy Science 91: 4414-4423.

VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. Journal of Dairy Science 92: 16-24.

Viana JMS, Piepho H. 2017. Quantitative genetics theory for genomic selection and efficiency of genotypic value prediction in open-pollinated populations. Scientia Agricola 74: 41-50.

Vitezica Z, Aguilar I, Misztal I, Legarra A. 2011. Bias in genomic predictions for populations under selection. Genetics Research 93: 357-366.

Wang H, Misztal I, Aguilar I, Legarra A, Muir WM. 2012. Genome-wide association mapping including phenotypes from relatives without genotypes. Genetics Research 94: 73-83.

Wang X, Miao J, Chang T, Xia J, An B, Li Y, Xu L, Zhang L, Gao X, Li J. 2019. Evaluation of GBLUP, BayesB and elastic net for genomic prediction in Chinese Simmental beef cattle. PLOS ONE. DOI:10.5061/dryad.4qc06.

Weber KL, Thallman RM, Keele JW, Snelling WM, Bennett GL, Smith TPL, McDaneld TG, Allan MF, Van Eenennaam AL, Kuehn LA. 2012. Accuracy of genomic breeding values in multibreed beef cattle populations derived from deregressed breeding values and phenotypes. Journal of Animal Science 90: 4177-4190.

Wheeler M, Foote BJ, Foote JF. 2013. Brief communication: genetic parameters of stayability measures in a dairy goat herd. Proceedings of the New Zealand Society of Animal Production 73: 165-167.

Wheeler M, Foote BJ, Foote JF. 2014. Brief Communication: relationships between longevity and herd-test traits in a dairy goat herd. Proceedings of the New Zealand Society of Animal Production 74: 49-51.

Wiggans GR, Hubbard SM. 2001. Genetic evaluation of yield and type traits of dairy goats in the United States. Journal of Dairy Science 84: 69-73.

Wolc A, Arango J, Settar P, Fulton JE, O'sullivan NP, Preisinger R, Habier D, Fernando R, Garrick DJ, Hill WG. 2012. Genome-wide association analysis and genetic architecture of egg weight and egg uniformity in layer chickens. Animal Genetics 43: 87-96.

Wolc A, Zhao HH, Arango J, Settar P, Fulton JE, O'sullivan NP, Preisinger R, Stricker C, Habier D, Fernando RL. 2015. Response and inbreeding from a genomic selection experiment in layer chickens. Genetics Selection Evolution. DOI:10.1186/s12711-015-0133-5.

Wright S. 1922. Coefficients of inbreeding and relationship. The American Naturalist 56: 330-338.

Xu Y, Liu X, Fu J, Wang H, Wang J, Huang C, Prasanna BM, Olsen MS, Wang G, Zhang A. 2020. Enhancing genetic gain through genomic selection: from livestock to plants. Plant Communications. DOI:10.1016/j.xplc.2019.100005.

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, Billis K, Cummins C, Gall A, Girón CG, Gil L, Gordon L, Haggerty L, Haskell E, Hourlier T, Izuogu OG, Janacek SH, Juettemann T, To JK, Laird MR, Lavidas I, Liu Z, Loveland JE, Maurel T, McLaren W, Moore B, Mudge J, Murphy DN, Newman V, Nuhn M, Ogeh D, Ong CK, Parker A, Patricio M, Riat HS, Schuilenburg H, Sheppard D, Sparrow H, Taylor K, Thormann A, Vullo A, Walts B, Zadissa A, Frankish A, Hunt SE, Kostadima M, Langridge N, Martin FJ, Muffato M, Perry E, Ruffier M, Staines DM, Trevanion SJ, Aken BL, Cunningham F, Yates A, Flicek P. 2018. Ensembl 2018. Nucleic Acids Research 46: 754-761.

Zhang H, Yin L, Wang M, Yuan X, Liu X. 2019. Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. Frontiers in Genetics. DOI:10.3389/fgene.2019.00189.

Zhang X, Lourenco D, Aguilar I, Legarra A, Misztal I. 2016. Weighting strategies for single-step genomic BLUP: An iterative approach for accurate calculation of GEBV and GWAS. Frontiers in Genetics. DOI:10.3389/fgene.2016.00151.

Zhao X, Dittmer KE, Blair HT, Thompson KG, Rothschild MF, Garrick DJ. 2011. A novel nonsense mutation in the DMP1 gene identified by a genome-wide association study is responsible for inherited rickets in Corriedale sheep. PLOS ONE. DOI:10.1371/journal.pone.0021739.

Zhou L, Mrode R, Zhang S, Zhang Q, Li B, Liu J. 2018. Factors affecting GEBV accuracy with single-step Bayesian models. Heredity. DOI:10.1038/s41437-017-0010-9.

Zhu B, Guo P, Wang Z, Zhang W, Chen Y, Zhang L, Gao H, Wang Z, Gao X, Xu L. 2019. Accuracies of genomic prediction for twenty economically important traits in Chinese Simmental beef cattle. Animal Genetics 50: 634-643.

Zumbach B, Tsuruta S, Misztal I, Peters KJ. 2008. Use of a test day model for dairy goat milk yield across lactations in Germany. Journal of Animal Breeding and Genetics 125: 160-167.

Zvinorova PI. 2017. A genome-wide association study on mechanisms underlying genetic resistance to gastrointestinal parasites in goats, Zimbabwe PhD thesis, Stellenbosch University.

Curriculum vitae

Megan Scholtens was born on the 19th October 1993 in Whakatane, Bay of Plenty, New Zealand. She completed her primary education in 2006 at Waiotahe Valley School, Bay of Plenty, New Zealand. She finished her secondary school in 2011 at Opotiki College, Bay of Plenty, New Zealand. In 2014 she was awarded Massey Equine Student of the Year and in 2015 she graduated with a Bachelors of Agri Science from Massey University, New Zealand. In 2016, she obtained a Masters Degree with First Class Honours from the Department of Animal Science, Massey University, New Zealand with the thesis "Genetic evaluation of milk traits, live weight, somatic cell score, and litter size at birth, and development of a selection index for dairy sheep", under the supervision of Prof. Nicolas Lopez-Villalobos and Dr Sam Peterson. In 2017 she was awarded a full scholarship by MBIE to realise her Ph. D studies. She started her Ph. D program in the Department of Animal Science, Massey University, New Zealand. In 2018 she was awarded the Young Member Finalist Award from the New Zealand Society of Animal Production, New Zealand. During her PhD she attended and presented at multiple conferences including the World Congress for Genetics Applied to Livestock Production in New Zealand, 2018 and the European Association of Animal Production in Belgium, 2019. She is going to start a Genetics and Breeding job at the Cawthron Institute, Nelson, New Zealand working with salmon, mussels and oysters.

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Megan Scholtens
Name/title of Primary Supervisor:	Prof. Nicolas Lopez-Villalobos
In which chapter is the manuscript /published work:	3

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

Scholtens MR, Lopez-Villalobos N, Garrick DJ, Blair HT, Lehnert K, Snell RG. 2019. Genetic parameters for total lactation yields of milk, fat, protein, and somatic cell score in New Zealand dairy goats. Animal Science Journal 10.1111/asj.13310

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:

- The percentage of the manuscript/published work that was contributed by the candidate:

- Describe the contribution that the candidate has made to the manuscript/published work:

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	Megan Scholtens Digitally signed by Megan Scholtens Date: 2020.08.26 14:16:46 +12'00'
Date:	26-Aug-2020
Primary Supervisor's Signature:	Nicolas Lopez-Villalobos Digitally signed by Nicolas Lopez-Villalobos DN: cn=Nicolas Lopez-Villalobos, c=NZ, o=Massey University, ou=Institute of Veterinary, Animal and Biomedical Sciences, email=N.Lopez-Villalobos@massey.ac.nz Date: 2020.08.26 15:15:43 +12'00'
Date:	26-Aug-2020

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Megan Scholtens
Name/title of Primary Supervisor:	Prof. Nicolas Lopez-Villalobos
In which chapter is the manuscript /published work:	4

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

Scholtens MR, Lopez-Villalobos N, Garrick DJ, Blair HT, Lehnert K, Snell RG. 2019. Estimates of genetic parameters for lactation curves for milk, fat, protein and somatic cell score in New Zealand dairy goats. New Zealand Journal of Animal Science and Production 79: 177-182.

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:

- The percentage of the manuscript/published work that was contributed by the candidate:

- Describe the contribution that the candidate has made to the manuscript/published work:

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	Megan Scholtens Digitally signed by Megan Scholtens Date: 2020.08.26 14:25:27 +12'00'
Date:	26-Aug-2020
Primary Supervisor's Signature:	Nicolas Lopez-Villalobos Digitally signed by Nicolas Lopez-Villalobos DN: cn=Nicolas Lopez-Villalobos, c=NZ, o=Massey University, ou=Institute of Veterinary, Animal and Biomedical Sciences, email=N.Lopez-Villalobos@massey.ac.nz Date: 2020.08.26 15:16:52 +12'00'
Date:	26-Aug-2020

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Megan Scholtens
Name/title of Primary Supervisor:	Prof. Nicolas Lopez-Villalobos
In which chapter is the manuscript /published work:	5

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

Scholtens MR, Lopez-Villalobos N, Garrick DJ, Blair HT. 2018. Heritability of longevity in New Zealand dairy goats. New Zealand Journal of Animal Science and Production 78: 11-15.

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:

- The percentage of the manuscript/published work that was contributed by the candidate:

- Describe the contribution that the candidate has made to the manuscript/published work:

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	Megan Scholtens Digitally signed by Megan Scholtens Date: 2020.08.26 14:28:08 +12'00'
Date:	26-Aug-2020
Primary Supervisor's Signature:	Nicolas Lopez-Villalobos Digitally signed by Nicolas Lopez-Villalobos DN: cn=Nicolas Lopez-Villalobos, c=NZ, o=Massey University, ou=Institute of Veterinary, Animal and Biomedical Sciences, email=N.Lopez-Villalobos@massey.ac.nz Date: 2020.08.26 15:17:38 +12'00'
Date:	26-Aug-2020

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Megan Scholtens
Name/title of Primary Supervisor:	Prof. Nicolas Lopez-Villalobos
In which chapter is the manuscript /published work:	6

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

Scholtens MR, Jiang A, Smith A, Littlejohn M, Lehnert K, Snell RG, Lopez-Villalobos N, Garrick DJ, Blair HT. 2020. Genome-wide association studies of lactation yields of milk, fat, protein and somatic cell score in New Zealand dairy goats. Journal of Animal Science and Biotechnology 10.1186/s40104-020-00453-2

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:

- The percentage of the manuscript/published work that was contributed by the candidate:

- Describe the contribution that the candidate has made to the manuscript/published work:

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	Megan Scholtens Digitally signed by Megan Scholtens Date: 2020.08.26 14:32:31 +12'00'
Date:	26-Aug-2020
Primary Supervisor's Signature:	Nicolas Lopez-Villalobos Digitally signed by Nicolas Lopez-Villalobos DN: cn=Nicolas Lopez-Villalobos, c=NZ, o=Massey University, ou=Institute of Veterinary, Animal and Biomedical Sciences, email=N.Lopez-Villalobos@massey.ac.nz Date: 2020.08.26 15:18:05 +12'00'
Date:	26-Aug-2020

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.