Genomic inbreeding estimation and effective population size of four South African dairy breeds

by

Janka Limper

Submitted in partial fulfilment of the requirements for the degree

MSc Animal Breeding and Genetics

Faculty of Natural and Agricultural Science

Department of Wildlife Science

University of Pretoria

Pretoria

South Africa

December 2018

Declaration

I declare that the thesis/dissertation, which I hereby submit for the degree MSc (Agric) Animal Science at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at any other tertiary institution.

Signature……..

Date……..

Acknowledgements

Firstly, I would like to thank my parents for supporting me throughout my studies. Without their continuous love and support both emotionally and financially I would not have been able to achieve all I have in my life. Furthermore, I would like to thank my sister, Ruska Limper and my brother, Ludwig Limper for their support and belief in me.

To my dear husband, thank you for always believing in me and always pushing me a little further when I don't believe I can reach any higher. Thank you for your love and support in everything that I set to achieve and for always being my rock.

Thank you to Dr. Carina Visser and Prof Esté van Marle-Köster for the opportunity to have completed my MSc under your guidance and supervision. Thank you for your patience and giving me the extra push when I needed it most. Thank you for allowing me to follow my passion for animal breeding and genetics.

I would also like to thank my fellow students Simon Frederick Lashmar and Rulien Grobler for all their assistance throughout my studies.

Thank you to my current employer Mr Allan Pennel for allowing me to finish my MSc whilst working for him as an intern.

Thank you to the Dairy Genomics Project (DGP) for the financial assistance towards the funding of this research. The opinions and conclusions are those of the author and are not necessarily to be attributed to the TIA.

Lastly, but most importantly I would like to thank the Lord for giving me this opportunity to finish a degree that most people dream of doing. Thank You for blessing me with the knowledge, I needed to be able to finish my studies and thank You for an amazing support system that backed me all the way.

Genomic inbreeding estimation and effective population size of four South African dairy breeds

by

Janka Limper

Supervisor:	Dr. C. Visser
Co-supervisor:	Prof E. van Marle-Köster
Department:	Animal and Wildlife Sciences
Degree:	MSc (Agric) Animal Science

Abstract

In this study 1002 dairy cattle, representing four South African dairy breeds (Ayrshire, Holstein, Jersey and SA Dairy Swiss) were genotyped at the ARC-Biotechnology Platform (Onderstepoort, 0110) with the Infinium Bovine SNP 50-24 V3.0 Beadchip. Genotypes for these animals originated from the Dairy Genomics Program (DGP) as part of an ongoing study aimed at integrating genomic information into the selection of South African dairy cattle. Genotypic data for all registered, genotyped animals participating in Logix Milk Recording was received from SA Stud Book, representing the Ayrshire, Holstein and Jersey breeds. The SA Dairy Swiss had no genotypic data available, thus hair samples from 62 individual animals were collected from three registered breeders to represent this breed. Raw Illumina genotype files were received via a downstream link from the ARC. These files were converted into PLINK (Purcell *et al.*, 2007) files using SNP Convert v1.0 (Nicolazzi *et al.*, 2016) for further analysis. The aim of the study was to estimate genomic inbreeding and effective population sizes for these breeds. The average call rate obtained across the samples was 99%. The observed heterozygosity values obtained for the populations were 0.355, 0.359, 0.340 and 0.345 for the Ayrshire, Holstein, Jersey and SA Dairy Swiss, respectively. Linkage disequilibrium (LD) estimation revealed average r^2 values of 0.181 (Ayrshire), 0.311 (Holstein), 0.349 (Jersey) and 0.291 (SA Dairy Swiss). Two different inbreeding estimates were calculated individual inbreeding coefficients (F_{IS}) and runs of homozygosity (F_{ROH}) and correlations were estimated between the inbreeding estimates. Pedigree-based inbreeding (F_{PED}) estimates were received from SA Stud Book for the Ayrshire, Holstein and Jersey and compared to the genomic inbreeding estimates. The mean individual inbreeding coefficient (F_{IS}) was -0.039 (Ayrshire), -0.007 (Holstein), -0.010 (Jersey) and -0.019 (SA Dairy Swiss), which indicates effective on farm management against inbreeding in the populations in this study. $F_{ROH > 16\ Mb}$ ranged from 0.227 to 0.255 (Ayrshire and Holstein). These relatively high F_{ROH} values indicate recent inbreeding in these populations. The strongest correlations were observed between F_{IS} and $F_{ROH>1}$ ranging from 0.454 to 0.686 (SA Dairy Swiss and Jersey) respectively, while lower correlations were found between F_{IS} and $F_{ROH>16}$, ranging from 0.071 to 0.377 (SA Dairy Swiss and Jersey) respectively. Very low correlations were found between F_{PED} and F_{ROH}, which may have been due to shallow pedigree depth. The highest correlation between F_{PED} and F_{ROH} (0.186) was observed for the Holstein at an ROH length of 4000kb. The N_e for the four populations included in the study has decreased to 117, 133, 120 and 112 for the Ayrshire, Holstein, Jersey and SA Dairy Swiss, respectively from approximately five generations ago. The four populations were separated into four separate clusters using principal component analysis (PCA). This corresponded with ADMIXTURE where the populations were also separated into the four respective populations. This indicates that the four populations are genetically distinct and were developed as separate breeds which is also consistent with the history of the four breeds. The high levels of genomic inbreeding could be explained by the increased use of artificial insemination in the populations studied. This is a concern as an increase in inbreeding leads to a reduction in the effective population size which was also evident in the populations included in the study.

Table of contents

List of Tables

List of Figures

Abbreviations

AI	Artificial Insemination
ARC	Agricultural Research Council
ARC-API	Agricultural Research Council – Animal Production Institute
BLAD	Bovine Leukocyte Adhesion Deficiency
BLUP	Best Linear Unbiased Prediction
BTA	Bos Taurus autosomes
BTP	Biotechnology Platform
CBD	Convention biological diversity
cM	centimorgan
CV	Cross Validation
CVM	Complex vertebral malformation
EBV	Estimated Breeding Value
EH	Expected heterozygosity
EM	Expectation Maximization
DAFF	Department of Agriculture Fisheries and Forestry
DGP	Dairy Genomics Program
DNA	Deoxyribonucleic Acid
DUMPS	Deficiency of Uridine Monophosphate Synthase
F_{GRM}	Genomic Relationship Matrix Inbreeding Coefficient
F_{IS}	Individual Inbreeding Coefficient
F_{PED}	Pedigree-based Inbreeding Coefficient
F_{ROH}	Inbreeding Coefficient based on ROH
F_{SNP}	SNP-based Inbreeding Coefficient
F_X	Inbreeding coefficient
ΔF	Rate of inbreeding per generation
FAO	Food and Agriculture Organisation
GEBV	Genomic Estimated Breeding Value
GRM	Genomic Relationship Matrix
GS	Genomic Selection
H_e	Expected Heterozygosity
IBD	Identical by descent

IBS	Identical by state
INTERGRIS	Integrated Registration and Genetic Information System
IFCN	International Farm Comparison Network
INTERBULL	International Bull Evaluation Service
kg	Kilogram
LD	Linkage disequilibrium
LDP	Linkage disequilibrium-based pruning
Logix	Livestock Operational and Genetic Information Exchange
MAF	Minor allele frequency
MOET	Multiple Ovulation and Embryo Transfer
mRNA	Messager RNA
N_e	Effective population size
OCS	Optimum Contribution Selection
PCA	Principal component analysis
QC	Quality Control
RNA	Ribonucleic Acid
ROH	Runs of Homozygosity
R_{SD}	Relationship between respective parents
SA	South Africa
SNP	Single Nucleotide Polymorphism
SOP	Standard Operating Procedures
TIA	Technology Innovation Agency
UMPS	Uridine monophosphate synthase
US	United States

Chapter 1

1.1. Introduction

Domestication and artificial selection led to the development of individual cattle breeds that differ in production and adaptability characteristics, such as their meat or milk yield and disease and internal and external parasite resistance as well as phenotypic diversity (Diamond, 2002; Flori *et al.*, 2009). Genetic diversity is essential for improvement and environmental adaptation of livestock (FAO, 2007; Woolliams and Oldenbroek, 2017). Without genetic variation breeders cannot develop new breeds that will be able to adapt to changing conditions (climate change), disease threats and changing market conditions (Bijma and Wooliams, 2002; FAO, 2010; Woolliams and Oldenbroek, 2017). Brotherstone and Goddard (2005) viewed the importance of genetic diversity from an economic point stating that large genetic variation both within and between breeds may result in more profitable cattle. Traditionally, dairy breed development was focused on an increase in milk and fat production along with the selection for body conformation traits. (Brotherstone and Goddard 2005; Sewalem *et al.*, 2006). In dairy cattle the economically important production traits (milk yield, butterfat and protein percentage) are only observed in female animals and the recording of these traits has been implemented since the 1900 for management and genetic improvement purposes world-wide (Brotherstone and Goddard 2005).

During the 1800s breed societies were established with closed herdbooks which resulted in the elimination of gene flow between populations that were perceived as different breeds (Brotherstone and Goddard 2005). In some cases, initial breed development occurred from relatively small effective population sizes (N_e) such as beef breeds (Hereford and Shorthorn) (MacHugh *et al.*, 1998). The development of reproductive technologies such as artificial insemination (AI) in the 1950s, together with the use of frozen semen has allowed breeders to use genetically superior bulls across the world (Fleming *et al.*, 2018). Best linear unbiased prediction (BLUP) was developed in the 1970s by Henderson (1973) as a method to estimate the genetic merit of bulls and cows that enabled the selection of genetically superior animals to contribute to the next generation. The development of these technologies led to an increase in the world-wide trade of semen resulting in the need for accurate comparisons between animals in different countries (Brotherstone and Goddard, 2005). Thus, INTERBULL was developed in 1983 with the purpose of providing standardized international genetic evaluations for all dairy

cattle (Brotherstone and Goddard, 2005). This enabled breeders to make world-wide selection decisions based on these genetic evaluations. The increased use of AI and the availability of international genetic evaluations available to farmers across the world accompanied with the drive for increased milk production resulted in a decrease in genetic diversity in the global dairy cattle population (Brotherstone and Goddard, 2005). This led research into selection methods which allowed for the selection of cattle for improved genetic gain whilst at the same time maintaining genetic variation.

The development and implementation of AI made it possible for a small number of sires to provide semen for more than 60 000 services in 1968, which led to 7 344 420 females to be artificially bred during 1970 in the US (Fleming *et al.*, 2018). Weigel (2001) reported that due to the extensive use of AI, a few elite bulls may have as many as 250 000 milking daughters worldwide. This use of a relatively small number of sires to contribute to the next generation led to a reduction in the genetic pool and resulted in an increased relatedness between cattle across the world (Weigel, 2001; Fleming *et al.*, 2018).

The Holstein breed showed the largest reduction in genetic variation as a result of modern breeding programs (Brotherstone and Goddard, 2005; Zenger *et al.*, 2007). Traditional breeding objectives focused on income from the sale of milk (Brotherstone and Goddard, 2005; Sewalem *et al.*, 2006). The ability of Holstein cows to produce more milk than other dairy breeds resulted in strong selection pressure on these animals for even higher milk production and saw this breed imported from the USA to other countries across the world (Brotherstone and Goddard, 2005; Zenger *et al.*, 2007). In the past decades breeding objectives has changed in order to include more economically important traits (Miglior *et al.*, 2005) such as longevity, live weight, calving interval and somatic cell score (Banga *et al.*, 2014). Fertility traits have also recently been included into selection programs due to the widely reported antagonistic effect between the selection for increased milk yield and fertility (Haile-Mariam *et al.*, 2003; Makgahlela *et al.*, 2007). Banga *et al.* (2014) stresses the importance of including liveweight in selection programs for dairy cattle as an increase in liveweight results in a decrease in profit due to increased feed costs.

Before the development of dense single nucleotide polymorphism (SNP) arrays, inbreeding and genetic variability were assessed for a population from its recorded pedigree information (Van Raden *et al.*, 2011; Rodriguez-Ramilo *et al.*, 2015). Estimates of pedigree-based inbreeding (F_{ped}) can be unreliable due to a lack of pedigree depth and pedigree errors (Cassell *et al.*, 2003; Bjelland *et al.*, 2013; Pryce *et al.*, 2014). Inbreeding estimates are expected to be

more accurate when they are based on genomic estimates using genome-wide DNA markers, such as microsatellite markers (Lenstra *et al.*, 2012; Cole *et al.*, 2013) and SNPs (Bjelland *et al.*, 2013; Rodriguez-Ramilo *et al.*, 2015). These molecular markers serve as tools in animal identification, parentage verification and the assessment of genetic distance within and between breeds (Lenstra *et al.*, 2012; Cole *et al.*, 2013; Sabir *et al.*, 2014). More recently SNPs have become available and are able to provide information on the selection of dairy cattle (Van Raden *et al.*, 2009; Seidel, 2010). These markers are expensive to develop but with the routine genotyping of cattle for genomic selection, as well as the ability to accurately estimate parentage and inbreeding the cost of these markers has decreased significantly (Hayes *et al.*, 2009; Seidel, 2010).

With the use of genomic information to confirm parentage, it has been shown that the frequency of misidentifying sires in United States dairy cattle can be as high as 13.9% (Wiggans *et al.*, 2012), which could mainly be attributed to pedigree errors. Runs of homozygosity (ROH) has in the past been used to infer population history in human and cattle populations and to examine the effect of deleterious homozygotes due to inbreeding. This has led to the proposition for the use of ROH as a way of estimating genomic inbreeding in cattle (Ferenčaković *et al.*, 2011; Bjelland *et al.*, 2013; Purfield *et al.*, 2017). F_{ROH} is expected to be more accurate than traditional pedigree-based inbreeding coefficients and can be used to distinguish between past versus recent inbreeding (Bjelland *et al.*, 2013; Purfield *et al.*, 2017). ROH may also give information associated with production and disease traits within a population and may also give new insights into selection signatures (Purfield *et al.*, 2012).

1.2 Aim and objectives

This study forms part of a project funded by the Technology Innovation Agency via the Dairy Genomics Program (DGP) that was established in 2016, with the overall aim to integrate genomic information in selection of SA dairy cattle. In this program genotypes have been generated using the Infinium Bovine SNP 50-24 V3.0 Beadchip for Holstein, Jersey, Ayrshire and SA Dairy Swiss cattle. The DGP is a three-year project, coordinated by the University of Pretoria and the Milk Producers Organization (MPO) with representation from the different breed societies, ARC-API, SA Stud Book and AI companies.

Information on effective population size (N_e), genomic diversity and inbreeding is important for maintaining genetic diversity in these breeds subjected to selection. Genomic inbreeding and N_e have not previously been estimated for South African dairy cattle.

The aim of the study is to estimate genomic inbreeding and effective population sizes for four SA dairy breeds. This will be achieved by attaining the following objectives:

1. Estimate genetic diversity parameters for four SA dairy breeds.
2. Estimate F_{SNP} and F_{ROH} inbreeding levels for each breed and compare these with F_{PED} values.
3. Estimate the effective population sizes of each breed.
4. Perform a population structure analyses across breeds to evaluate population differentiation.

Chapter 2

Literature Review

2.1. Introduction

More than 800 modern cattle breeds can be found around the world (MacHugh *et al.*, 1997). These breeds are commonly subjected to artificial selection for increased production. Genetic variation is important to be maintained in any species as it ensures its survival (Zenger *et al.*, 2007) and is crucial for a response to selection (Cardellino and Boyazoglu, 2009). Genetic diversity allows animals to adapt to changing environments, which enable them to maintain constant growth and reproduction not only important for survival but to also reach production requirements (Zenger *et al.*, 2007; FAO, 2007). Intense selection may lead to inbreeding, especially when only a few proven sires are used for mating and may result in a reduction of animal performance.

Traditionally inbreeding and genetic variability were assessed for a population using its pedigree information (Van Raden *et al.*, 2011; Rodriguez-Ramilo *et al.*, 2015). Recently genomic based inbreeding estimates have become the method of choice to assess population inbreeding and effective population size (N_e). These estimates are expected to be more accurate when it is based on genomic estimates using genome-wide DNA markers, usually single nucleotide polymorphism (SNP) (Bjelland *et al.*, 2013; Rodriguez-Ramilo *et al.*, 2015). Runs of homozygosity (ROH) can also be used as a method of estimating inbreeding and is a useful measure of recent versus ancient inbreeding (Kirin *et al.*, 2010; Bjelland *et al.*, 2013). The aim of this section is to provide a brief overview of the SA dairy industry and review relevant literature with regard to genomic diversity and inbreeding.

2.2. Brief overview of South African dairy industry

The total world milk production in 2016 was 826 million tons of which 96% was contributed by cows and buffalos (IFCN, 2017). World milk production is expected to increase by 22% (178 million tons) by 2026 (FAO, 2017). It is anticipated that most of this increase in milk production will come from developing countries (FAO, 2017). South Africa produces approximately 2.8 million tons of milk per annum over a ten-year period (DAFF, 2017), representing up 0.5% of the total

world milk production (Lacto Data, 2018). Local milk production has increased by 26% from 2009 (2 587 000 t) to 2017 (3 253 000 t) (Lacto Data, 2018).

According to the Department of Agriculture Fisheries and Forestry (DAFF, 2018) there were approximately 13 million cattle in South Africa, of which 1.41 million were dairy cows in 2016/17. Per capita consumption of fresh milk has increased from 35.8kg/year in 2010/11 to 39.0 kg/year in 2016/17, indicating that the demand for fresh milk is growing. This demand has been accompanied by an 83% increase in milk production over the same period (DAFF, 2018). Table 2.1 shows the per capita consumption of fresh milk as compared to other livestock products in South Africa.

Table 2.1 Production and per capita consumption of animal protein products (Adapted from DAFF, 2018)

Product	Production* (1 000t)	Per capita consumption* (kg/year)
Meat		
White	1 676	38.76
Red	1 473	26.34
Fresh milk	2 207	39.0
Dairy Products#		
Butter	12	0.3
Cheese	38	0.8
Condensed milk and milk powder	340	7.3
Pigs (pork)	232.5	4.6
Sheep and Goats	177.4	3.3
Eggs	455	7.28

*2016/17

#No new data was found for dairy products, values here are for 2004/05

The average number of cows in milk per producer during 2017 was highest for the Eastern Cape (606), followed by KwaZulu-Natal (594) and Western Cape (268) (Lacto Data, 2018), while in Gauteng the herd average for cows in milk was 188 per producer. In Figure 2.1 the number of milk producers per province in the country for October 2017, is shown, with the Eastern Cape having the highest number of producers and Limpopo the lowest. This is due to the suitable climate of the coastal regions for milk production with mild temperatures and good quality pastures which reduces the need for high cost total mixed ration (TMR) systems (DAFF, 2017; Lacto Data, 2018).

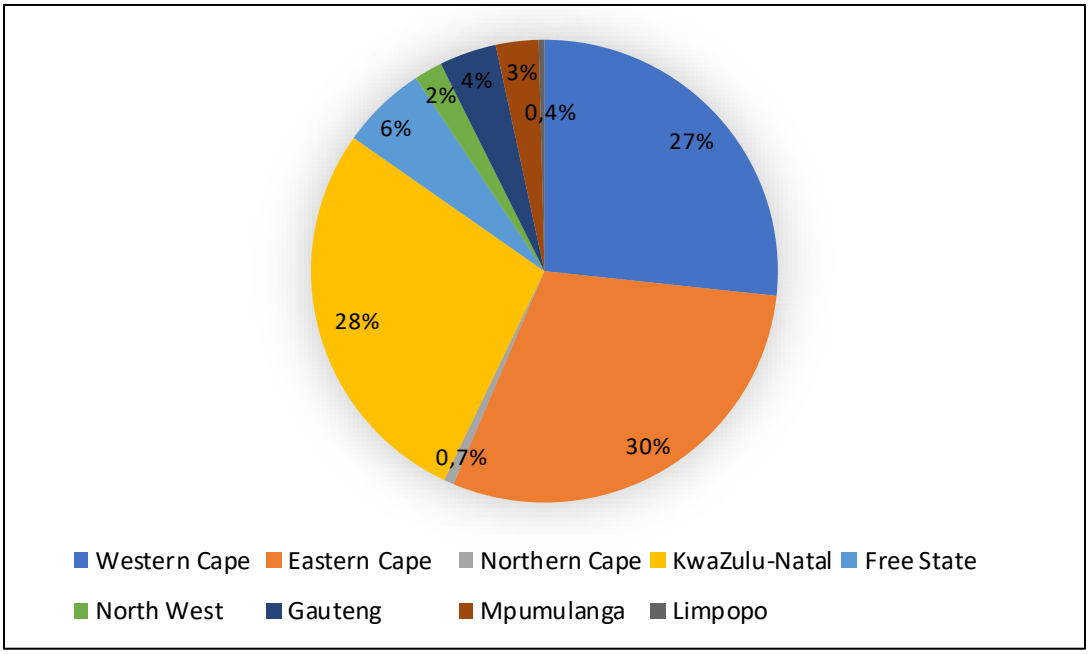

Figure 2.1 Number of South African milk producers per province according to statistics for October 2017 (Adapted from Lacto Data, 2018)

2.2.1 Main South African dairy breeds

The most numerous dairy cattle breeds in South Africa are the Holstein, Jersey, Guernsey and Ayrshire (DAFF, 2017). The Holstein breed account for 57% of cows participating in Milk Recording in South Africa, followed by the Jersey (38%) and Ayrshire (4%) (Mostert, 2007). In recent years cows in Milk Recording have decreased to less than 20% of all cows in milk (SA Stud Book, 2016).

South African Holstein

The modern Holstein descended from the Friesian breed, which originated in North-Western Europe (South African Livestock Breeding, 2004). The first Friesian black pieds were imported into South Africa in 1854, followed by importation of the Canadian Holstein in 1963 after which the breed was established (South African Livestock Breeding, 2004). Although the first Holstein-Friesland were registered in 1906 at SA Stud Book, the Breeders' Society was only founded a few years later (1912) (Duvenhage, 2017). The Breeders' Society was first known as The Friesland Cattle Breeder's Association of South Africa but has since changed its name to the Holstein-Friesland Cattle Breeders' Association of South Africa (Duvenhage, 2017). Currently there are 71 296 (2017, B. Mostert, Pers. Comm, bernice@studbook.co.za; 2018, C. Banga, Pers. Comm, cuthbert@arc.agric.za) Holstein cows participating in Milk Recording (Table 2.2). In Figure 2.2 a SA Holstein cow, as representative of the breed is shown.

Figure 2.2 A typical SA Holstein cow as presented by the Holstein-Friesland Cattle Breeders' Association of South Africa (*www.saholstein.co.za*)

South African Jersey

The Jersey breed was developed on the Island of Jersey from a breed that migrated from North Africa to France (Nel, 1968). On the Island of Jersey, the population was closed for nearly 250 years (South African Livestock Breeding, 2004). In 1881 the first Jerseys were imported to South Africa (Nel, 1968). The South African Jersey Cattle Breeders' Society was established in 1920 in Pietermaritzburg with the main objective to encourage Jersey breeding in South Africa and to maintain the purity of the breed (Nel, 1968). In the 1990s, 19% of cattle registered under the Milk Recording Scheme were Jersey cattle (www.jerseysa.co.za) and this number doubled by 2002. Currently there are 51 102 registered and commercial Jersey cows participating in Milk Recording (2017, B. Mostert, Pers. Comm, bernice@studbook.co.za; 2018, C. Banga, Pers.

Comm, cuthbert@arc.agric.za) (Table 2.2). Figure 2.3 shows a SA Jersey cow, as representative of the breed according to the Breed Society.

Figure 2.3 A typical SA Jersey as presented by the South African Jersey Cattle Breeders' Society (*www.jerseysa.co.za*)

South African Ayrshire

The Ayrshire breed originated in the Cunningham district of Ayrshire, Scotland where this breed was known as Cunningham cattle (McCreath, 1913). The Ayrshire breed received its name in 1821 when it was for the first time called the 'true Ayrshire breed of milk cattle' (South African Livestock Breeding, 2004). Two bulls and eight cows were imported to South Africa between 1890 and 1893 (www.ayrshire.co.za). The Ayrshire Cattle Breeders' Society of South Africa was established in April 1916, with the objective to control the registration of the breed and to maintain the purity of the breed. Currently there are 4 233 Ayrshire cows participating in Milk Recording (2017, B. Mostert, Pers. Comm, bernice@studbook.co.za; 2018, C. Banga, Pers. Comm, cuthbert@arc.agric.za) (Table 2.2) Figure 2.4 shows a SA Ayrshire, as representative of the breed according to the Breed Society.

Figure 2.4 A typical SA Ayrshire as presented by the Ayrshire Cattle Breeders' Society of South Africa (*www.ayrshire.co.za*)

South African Dairy Swiss

The first Brown cattle (Braunvieh) were imported from the USA into South Africa at the beginning of the previous century as a dual-purpose breed (www.dairyswiss.co.za). The SA Dairy Swiss was developed from the Brown Swiss in 1974 and was recognized by the Department of Agriculture in 1995 as a separate breed in terms of the Livestock Improvement Act (www.dairyswiss.co.za). The SA Dairy Swiss Cattle Breeders' Society currently consists of 10 members (www.dairyswiss.co.za). There are currently 567 registered dairy cows participating in Milk Recording as shown in Table 2.2 (2017, B. Mostert, Pers. Comm, bernice@studbook.co.za; 2018, C. Banga, Pers. Comm, cuthbert@arc.agric.za). In Figure 2.5 a SA Dairy Swiss cow, as representative of the breed according to the Breed Society.

Figure 2.5 A typical SA Dairy Swiss as presented by the South African Dairy Swiss Cattle Breeders' Society (*www.dairyswiss.co.za*)

2.2.2 Production statistics of South African dairy breeds

Data presented in Table 2.2 was obtained as personal communications from SA Stud Book (2017, B. Mostert, Pers. Comm, bernice@studbook.co.za) and the ARC (2018, C. Banga, Pers. Comm, cuthbert@arc.agric.za), respectively. The ARC does not currently have any SA Dairy Swiss animals in Milk Recording, thus production averages for the SA Dairy Swiss in Table 2.2 is representative of animals registered at SA Stud Book. SA Stud Book's Milk Recording system is called the Logix Milk Recording System (SA Stud Book, 2016), whereas the ARC makes use of INTERGRIS (Van Graan, 2016).

From Table 2.2 it is clear that the Holstein is numerically the largest breed, making up approximately 56% of the dairy cattle participating in Milk Recording, while the SA Dairy Swiss makes up only 0.45%

Table 2.2 Production averages of active dairy cows participating in Milk Recording (SA Stud Book and ARC)

Breed	Production parameter	305-day Lactation averages (kg)	Production averages per day (kg)	Total number of active cows
Holstein	Milk yield	8 636 – 8 950	28 - 29	
	Butter Fat	332 - 339	1.09 – 1.11	71 296
	Protein	278 - 285	0.91 – 0.93	
Jersey	Milk yield	5 363 - 5 762	17 - 19	
	Butter Fat	264 - 273	0.8 - 0.9	51 102
	Protein	205 - 215	0.6 - 0.7	
Ayrshire	Milk yield	5 874 – 8 153	19 - 26	
	Butter Fat	249 - 317	0.8 – 1.0	4 233
	Protein	266 - 196	0.6 – 0.9	
SA Dairy	Milk yield	8 612.7	28.24	
Swiss	Butter Fat	357.72	1.17	567
	Protein	302.32	0.99	

2017, B. Mostert, Pers. Comm, bernice@studbook.co.za

2018, C. Banga, Pers. Comm, cuthbert@arc.agric.za

2.3. Maintaining genetic diversity and effective population size

In order to ensure sustainable agricultural production to meet present and future human food needs, the conservation and sustainable utilization of farm animal genetic diversity are global obligations under the Convention on Biological Diversity (CBD) (FAO 2007). Domestic animal genetic diversity can be defined as genetic differences within or between breeds of species used for food production (Cardellino and Boyazoglu, 2009; Woolliams and Oldenbroek, 2017). Intense selection pressure together with the overuse of a few sires have led to a loss of genetic diversity and the survival of dairy breeds across the world (Bjelland *et al.*, 2013; Rodriguez-Ramilo *et al.*, 2015). Artificial selection focuses on a specific part of the genome that increases fitness or improves production, thus increasing the frequency of the favorable alleles, resulting in the fixation and thus reduced variation at that region of the genome (O'Brien *et al.*, 2014; Makina, 2015). Genetic diversity is essential for natural selection where breeds need to adapt to different environmental stressors and for artificial selection for improved production performance (Woolliams and Oldenbroek, 2017).

Traditionally genetic diversity within and between breeds were estimated by studying the phenotypic diversity between individuals within a breed and comparing this to individuals within the same species that were representative of the breed (Sabir *et al.*, 2014; Woolliams and Oldenbroek, 2017). More recently with the introduction of DNA-analysis, it has become possible to assess genetic diversity at a DNA level (Woolliams and Oldenbroek, 2017). Before the discovery of DNA analysis, it was assumed that two offspring would have half the chance to inherit the same allele from their parents, thus sharing 50% of the DNA inherited from the parents (Leutenegger *et al.*, 2003; Woolliams and Oldenbroek, 2017). With the development of DNA-markers, such as microsatellite markers and SNPs, this has been shown not to be true. This can be attributed to the Mendelian Sampling effect (Pryce and Daetwyler, 2012; Clark *et al.*, 2013), but also the mis-identification of parents (Wiggans *et al.*, 2012).

Microsatellite markers are several copies of short tandem repeats (1-5bp long) that are evenly distributed throughout the eukaryotic genome within coding and non-coding regions (Sabir *et al.*, 2014). These markers may be used for parentage verification, the assessment of genetic distance within and between breeds, the identification of disease carrier animals, as well as the mapping of genes and marker-assisted selection (Lenstra *et al.*, 2012; Cole *et al.*, 2013). Although microsatellite markers proved to be powerful research tools, they require labor intensive development which is linked to high costs (Sabir *et al.*, 2014). This has led to research into the idea of inferring microsatellite markers developed for one species (*Bos taurus*) to another species (buffalo; *Bubalus bubalis*) (Moore *et al.*, 1991; Navani *et al.*, 2002). Microsatellite markers may be subject to several disadvantages namely, high costs associated with the development of these markers as mentioned previously, misclassification of heterozygotes as homozygotes, underestimation of genetic divergence, it does not provide information on functional traits and covers only a small portion of the genome (Erhardt and Weimann, 2007; Yang *et al.*, 2013; Fernăndez and Bennewitz, 2017).

SNPs are bi-allelic, single-locus markers that are located abundantly throughout the *Bos taurus* genome (Seidel, 2010; Sabir *et al.*, 2014). SNPs can be found approximately every 700bp in the *Bos taurus* genome (Bovine HapMap Consortium, 2009) amounting to approximately 4 million SNPs across the genome (Seidel, 2010). Small samples of SNPs are identified using an approach known as the SNP chip, which is made up of a small piece of plastic with thousands of small dots that are able to bind DNA (Seidel, 2010). These dots each correspond to specific SNPs and depending on whether the animal inherited the SNP from neither, one or both its parents; zero, one or two copies of the SNP will be present on the genome of a given animal (Seidel,

2010). The Illumina 50K SNP chip is currently the most widely used SNP chip used in cattle (Illumina Inc., San Diego, CA, USA) and although some of the SNPs within this chip are redundant and contain ambiguous information, they are still able to provide useful information with regards to selection in dairy cattle populations (Van Raden *et al.*, 2009; Seidel, 2010; Sabir *et al.*, 2014). Currently Illumina has a Bovine HD Beadchip available that features 777 962 SNPs that are relatively evenly spaced throughout the genome (Illumina Inc., San Diego, CA, USA). Hayes *et al.* (2009) demonstrated that incorporating the use of SNPs into selection programs can double the response of selection as a result of the identification of genetically superior animals at a young age. The use of SNPs for selection also comes with a few limitations, in that it requires good phenotypic records from a large number of animals within a given population (Van Tassel *et al.*, 2008; Hayes *et al.*, 2009). Another drawback with the use of SNPs is that a separate system has to be set up for each population in order to improve the accuracy of SNP chips as SNPs match up to different alleles within different breeds (Hayes *et al.*, 2009).

2.3.1 Measures of genetic diversity

Currently microsatellite markers are the most widely explored method for the estimation of genetic diversity in livestock species (Lenstra *et al.*, 2012; Sabir *et al.*, 2014). Genetic diversity can also be measured by the allelic diversity (Fernǎndez and Bennewitz, 2017), although this parameter is used mainly in conservation programs (Toro *et al.*, 2009). Allelic diversity refers to the number of different alleles present at one or more loci in a chromosome and is thus used as a measure of genetic diversity within a population (Toro *et al.*, 2009). Allelic diversity is an uncomplicated measure of genetic diversity as a high number of alleles implies variation (Fernǎndez and Bennewitz, 2017). The most common measure of within breed genetic diversity is the expected heterozygosity (H_E). H_E is the probability that two alleles chosen at random from the population, are different (Nei, 1973; Melka and Schenkel, 2012) and can be calculated for a specific locus as follows (Fernǎndez and Bennewitz, 2017):

$$EH = 1 - \sum p^2$$

where: EH = Expected heterozygosity

p = allele frequency

H_E measures the ability of a given population to respond to either natural or artificial selection within a short period of time (Fernǎndez and Bennewitz, 2017). The maintenance of

genetic diversity is essential in mitigating unpredictable changes in socio-economic needs and environmental changes (Hoffmann, 2010).

The effective population size (N_e) is an important parameter that can be used to assess the rate of inbreeding and the loss of genetic diversity within a population (Groeneveld *et al.*, 2010) and has in the past been estimated using recorded pedigree information (Falconer and Mackay, 1996). This method of estimating N_e is not reliable as it depends on the completeness and correctness of pedigree information (Barbato *et al.*, 2015). N_e from pedigree information can be calculated as follows (Falconer and Mackay, 1996):

$$Ne = \frac{1}{2\Delta F_L}$$

N_e can be used to assess population performance based on genetic variation and inbreeding over long periods of time, (Fernández *et al.*, 2005) and can be defined as the number of individuals in a given population as reflected by the rate of inbreeding for the population (Wright 1969). The census size of a population is not an accurate reflection of its N_e (Rodriguez-Ramilo *et al.*, 2015; Jiménez-Mena and Bataillon, 2016). This could be attributed to the fact that the idealized population may deviate as a result of unequal sex ratios, variations in family size, successive generations may have unequal numbers as well as the overlapping of generations (Falconer and Mackay, 1996). Although the Holstein is a numerically large breed, dominating dairy production worldwide (Zenger *et al.*, 2007), it has experienced a reduction in genetic diversity due to a high selection intensity and the increased use of AI. The N_e for the Italian Holstein expected to be approximately 69 (Mastrangelo *et al.*, 2016) and 101 for the Spanish Holstein (Rodriguez-Ramilo *et al.*, 2015). Falconer and Mackay (1996) proposed that in order to maintain a constant effective population size, an equal number of male and female animals should be kept.

The low effective population size of most livestock breeds (Leroy *et al.*, 2013) can mainly be attributed to the intense selection of only a few proven animals contributing to the next generation. An N_e size of at least 50 should be maintained to avoid inbreeding depression in the short term (Frankham *et al.*, 2002), whilst an N_e of 500 should be maintained to sustain reproductive fitness and to prevent a reduction in genetic diversity in the long term (Holt *et al.*, 2005; Meuwissen, 2009). Table 2.3 contains the N_e size of four South African dairy breeds, as well as N_e sizes for their global counterparts.

Table 2.3 Summary of effective population size (N_e) reported for Holstein, Jersey, Ayrshire and SA Dairy Swiss across countries

Breed	Effective Population Size	Reference	Country
Holstein	69.61	Mastrangelo *et al.*, 2016	Italy
	101	Rodriguez-Ramilo *et al.*, 2015	Spain
	98.7	Marras *et al.*, 2015	Italy
	114	Stachowicz *et al.*, 2011	Canada
	137*	Maiwashe *et al.*, 2006	South Africa
	39*	Weigel, 2001	USA
Jersey	108*	Maiwashe *et al.*, 2006	South Africa
	30*	Weigel, 2001	USA
Ayrshire	148*	Maiwashe *et al.*, 2006	South Africa
	161*	Weigel, 2001	USA
SA Dairy	90.7	Marras *et al.*, 2015	Italy
Swiss	132*	De Ponte-Bouwer *et al.*, 2013	South Africa
	61*	Weigel, 2001	USA

*Estimated from recorded pedigree data

The N_e size could be calculated from linkage disequilibrium (LD), however it is not feasible as demographic processes like genetic drift can cause LD signatures (Khatkar *et al.*, 2008). If this occurs alleles at different loci becomes associated independent of their proximity in the genome (Barbato *et al.*, 2015). This is supported by Gomez-Romano *et al.* (2013) who found that a low N_e is associated with high levels of LD. Barbato *et al.* (2015) developed a new software tool (SNeP tool) to estimate N_e across generations making use of SNP data. The SNeP tool is able to correct for sample size, phasing and recombination rate. It should be noted that inbreeding across the genome is heterogenous, leading to some regions undergoing inbreeding at a faster rate than other regions (Jiménez-Mena and Bataillon, 2016; Howard *et al.*, 2017).

Maintaining genetic diversity can be achieved by maximizing the effective population size (N_e) or minimizing the rate of inbreeding (De Cara *et al.*, 2011; Gómez-Romano *et al.*, 2016). Genetic diversity loss can thus be avoided by regulating the level of inbreeding in livestock populations. A trade-off between the maintenance of uniformity and the minimization of unfavorable effects due to inbreeding occurs in dairy cattle populations.

2.4. Inbreeding in dairy cattle

In a finite population the prevention of inbreeding is impossible as there is an exponential increase in the number of ancestors in each generation (Howard *et al.*, 2017). Finite populations are also subjected to a random fluctuation of gene frequencies or genetic drift during gamete formation which will eventually lead to a loss of alleles during the sampling process (Falconer and Mackay, 1996). Intense selection may however also lead to inbreeding, especially when only a few proven sires are used for mating and may result in inbreeding depression (Weigel, 2001; Du Toit *et al.*, 2012).

Inbreeding occurs when two individuals that share at least one common ancestor are mated together (Weigel, 2001; Fleming *et al.*, 2018). Inbreeding results in a decrease in the additive genetic variation as well as a decreased response to selection for the traits under selection, as well as other traits (Kristensen and Sorensen, 2005). Thus, inbreeding results in changes in the distribution of genetic variance which in return leads to allelic fixation. Inbreeding within a population may increase due to several reasons. These include intense directional selection over a number of generations (Robertson, 2007), the intense use of AI with a small number of superior sires (Nicholas and Smith, 1983), as well as the use of BLUP in conjunction with truncation selection (Verrier *et al.*, 1993), which led to the extensive use of related individuals as parents.

If inbreeding increases it will eventually result in a reduction in animal performance and thus inbreeding depression (Weigel, 2001; Du Toit *et al.*, 2012). Inbreeding has the greatest effect on fitness traits such as survival, reproduction and disease resistance (Kristensen and Sorensen, 2005; Howard *et al.*, 2017; Fleming *et al.*, 2018) but may also affect economically important traits such as fertility traits, age at first calving (Smith *et al.*, 1998) and calving interval (Pryce *et al.*, 2014), as well as milk production (Thompson *et al.*, 2000a, b; González-Recio *et al.*, 2007) and longevity (Thompson *et al.*, 2000a, b; Sewalem *et al.*, 2006).

Inbreeding depression assuming no epistasis can be calculated as follow (Crow and Kimura, 1970):

$$2F\sum_{i=1}^{n} p_i q_i d_i$$

where: F = inbreeding coefficient

p_i and q_i = allele frequencies

d_i = dominance deviation at the i^{th} locus

Inbreeding depression can be explained by two main hypotheses, namely the partial dominance hypothesis or the over-dominance hypothesis. With the partial dominance hypothesis, it is assumed that inbreeding depression occurs as a result of the expression of deleterious recessive alleles in a homozygous individual. Thus, with an increase in inbreeding within a population, the frequency of deleterious recessive homozygotes will increase resulting in the expression of deleterious homozygotes that were hidden in the heterozygote (Kristensen and Sorensen, 2005; Howard *et al.*, 2017). The over-dominance hypothesis displays heterozygote advantage. With an increase in the level of inbreeding the level of heterozygote genotypes will be reduced and the superior heterozygote genotypes will become less frequent (Kristensen and Sorensen, 2005; Howard *et al.*, 2017). The long-term effects of these two hypotheses are different in the sense that with over-dominance selection would favor a heterozygote state at multiple loci whereas with partial dominance selection would remove unfavorable alleles that were generated as a result of mutations, within a population (Kristensen and Sorensen, 2005). Most of the inbreeding depression that is observed in loci occurs as a result of partial dominance (Charlesworth and Charlesworth, 1987; Kristensen and Sorensen, 2005). Another explanation for inbreeding depression that will not be discussed here is inbreeding depression due to epistasis between dominance effects across loci. This results in a decreased frequency of favorable gene combinations among heterozygous genotypes (Jain and Allard, 1966; Howard *et al.*, 2017).

Studies indicate that inbreeding is increasing in dairy cattle populations across the world (González-Recio *et al.*, 2007; Bjelland *et al.*, 2013; Pryce *et al.*, 2014; Marras *et al.*, 2015). The overuse of high impact sires and linebreeding resulted in high levels of relatedness with the local dairy cattle populations, which resulted in increased levels of inbreeding. In 1998 Smith *et al.* reported that a 1% increase in inbreeding would result in a loss of 37 kg milk, 1.2 kg fat and 1.2 kg protein, per lactation in Holstein cattle. These findings were confirmed by Croquet *et al.* (2006 and 2007), who reported a 15.42 kg reduction in 305d-milk, 0.64 kg reduction in fat yield and 0.59 kg reduction in protein yield for every 1% increase in inbreeding in Walloon Holstein cows. Thus, an increase in inbreeding results in a reduced lifetime performance of dairy cattle, which in turn leads to economic losses (Smith *et al.*, 1998; Thompson *et al.*, 2000a, b; Brotherstone and Goddard, 2005). More recent research estimated that a 1% increase in inbreeding would result in a 0.7-day increase in the calving interval and a 0.3% decrease in the survival to second lactation in Irish Holstein-Friesland cows (Mc Parland *et al.*, 2007). This was supported by Bjelland *et al.*, (2013), who also found that an increase in inbreeding would be associated with a reduction in lifetime milk yield and reproductive ability in US Holstein cows.

As a result of an increase in the number of homozygous loci due to inbreeding, several recessive mutations have been observed in dairy cattle in addition to decreased performance and profitability. The three most important inherited disorders of Holstein cattle are bovine leukocyte adhesion deficiency (BLAD) (Kehrli *et al.*, 1992), complex vertebral malformation (CVM) (Agerholm *et al.*, 2001) and a deficiency in uridine monophosphate synthase (DUMPS) (Shanks and Robinson 1990; Kaminski *et al.*, 2005). In Table 2.4 a summary of the causative mutations for each of the common disorders are given. BLAD is a rare, autosomal recessive disorder that is prevalent in young animals and is characterized by recurrent pneumonia, enteritis, delayed wound healing and death at an early age (Kehrli *et al.*, 1992; Nagahata, 2004). This disorder is associated with two-point mutations of which the second is a silent mutation in the CD18 gene (Kehrli *et al.*, 1992; Nagahata, 2004).

A second inherited disorder found in Holstein cattle, is DUMPS (Shanks and Robinson, 1990). DUMPS is an autosomal recessive disease caused by a deficiency of the uridine 5' monophosphate synthetase (UMPS) enzyme (Robinson *et al.*, 1983). DUMPS result in the loss of homozygous affected embryos at day 40 of pregnancy (Shanks and Robinson, 1990) and reduced enzyme activity in the milk and urine of lactating cows heterozygous for DUMPS (Shanks and Greiner, 1992). The disorder manifests itself as a result of a point mutation resulting in a premature stop codon in the bovine UMPS messager RNA (mRNA). Carriers of DUMPS were all descended from a Holstein bull, Skokie Sensation Ned, born in 1957 (Kaminski *et al.*, 2005).

CVM is a third inherited disorder found in Holstein cattle, which mainly results in the abortion of homozygous affected fetuses before day 260 of gestation, or stillbirths (carry two copies of the mutant allele) (Hemati *et al.*, 2015). Visible deformities such as a short neck, curved legs and abnormal ribs, can be seen in the dead calve (Nagahata *et al.*, 1987). In rare cases calves may be born alive, these calves will show visible malformations of the carpal and tarsal joints, low birth weights, cardiac abnormalities and will usually die within a few days of birth (Agerholm, 2007). CVM is the cause of a point mutation at chromosome three which encodes a uridine 5-diphosphate-N-acetyl-glucosamine transporter resulting in the amino acid valine being substituted by phenylalanine at position 180 (Thomsen *et al.*, 2006). The above-mentioned inherited disorders spread world-wide in the global Holstein population due to the extensive use of AI bulls that were carriers for these diseases (Schwenger *et al.*, 1994; Nagahata *et al.*, 2002; Agerholm, 2007).

Table 2.4 Summary of common genetic disorders in dairy cattle due to inbreeding as reported in literature

Disorder	Mutation	Reference
BLAD	A → G at nucleotide 383	Kehrli *et al.*, 1992; Nagahata, 2004
	C → T at nucleotide 775	Hemati *et al.*, 2015
DUMPS	C → T at codon 405	Schwenger *et al.*, 1994
CVM	G → T at nucleotide 559	Thomsen *et al.*, 2006

Mandatory DNA analysis is performed for these diseases before a bull can be selected as a sire for the next generation (Schütz *et al.*, 2008). In Germany the number of BLAD carriers has decreased from approximately 10% in 2003 to 1.6% in 2007 (Schütz *et al.*, 2008). For the same period, the number of CVM carriers has slightly decreased from approximately 16% in 2003 to 14.6% in 2007 (Schütz *et al.*, 2008). This corresponds to Nagahata *et al.* (2002) who reported that 13.2% of the AI sires used in Germany during 2001 were CVM carriers. All of the above-mentioned diseases result in major economic losses due to the loss of the fetus and longer calving intervals.

2.5 Estimation of inbreeding

Due to recent advances in reproductive technologies and genomic selection as a way to identify and select animals with superior genotypes to contribute to the next generation, it has become almost impossible to find dairy animals without genetic relationships to certain superior individuals (Weigel, 2001). The level of inbreeding can be measured by the inbreeding coefficient (F_X) which can be defined as the probability that two alleles inherited by an individual from its parents are identical by descent (IBD) (Bourdon, 2014). Two alleles are IBD when the parents of an individual transmit two identical alleles which they inherited from a common ancestor to their offspring (Bourdon, 2014).

Raymond Pearl made the first attempts to quantify inbreeding. During the period 1913 to 1917, he published a series of papers in an effort to quantify inbreeding based on pedigree information (Curik *et al.*, 2014). However, it was Wright (1922) whom a few years later developed the inbreeding coefficient. Since its development, the inbreeding coefficient has been mainly estimated from pedigree information (F_{PED}). According to Malécot and Blaringhem (1948) as cited by Curik *et al.* (2014), if there is no selection or mutations occurring, it is assumed that all loci are

segregating in the same hereditary pattern and is therefore expected to have a similar inbreeding coefficient (F_{PED}).

F_X can be calculated as follows (Bourdon, 2014):

$$\sum_{CA=1}^{k} \left(\frac{1}{2}\right)^{n_1+n_2+1} (1 + F_{CA})$$

where, CA = Individual *X*'s sire and dam's common ancestor

k = number of common ancestors in individual *X*'s pedigree

n_1 = number of generations that separate the sire of individual *X* from the common ancestor

n_2 = number of generations that separate the dam of individual *X* from the common ancestor

F_{CA} = Common ancestor's inbreeding coefficient

The rate of inbreeding per generation can be calculated as follows (Falconer and Mackay, 1996):

$$\Delta F = \frac{F_t - F_{t-1}}{1 - F_{t-1}}$$

where F_t and F_{t-1} represent the average inbreeding of the offspring and their parents, respectively.

Figure 2.6 represents the common matings that can be expected, in the form of arrow diagrams. From the figure the inbreeding coefficient (F_x) and Wright's coefficient of relationship between the respective parents (R_{SD}) can be seen.

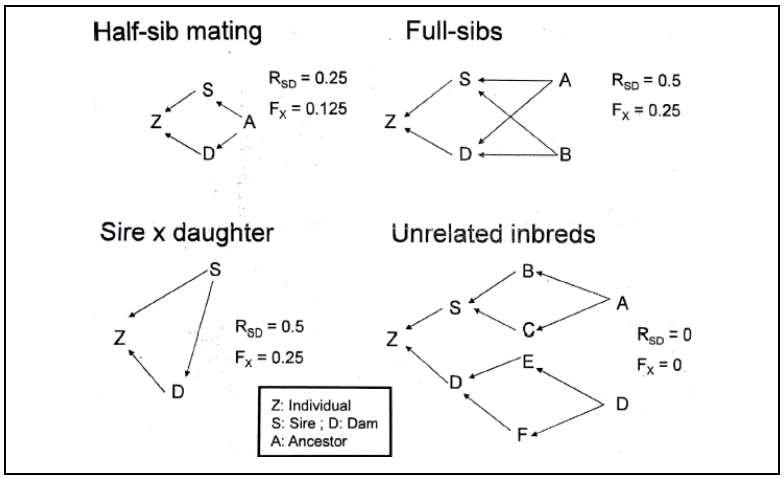

Figure 2.6 Illustration of common matings and their associated inbreeding coefficients (Bourdon, 2014)

With full-sib mating the inbreeding coefficient for the offspring will be 0.25, this is the same as the inbreeding coefficient that can be expected when a bull is mated with its daughter. With an increase in R_{SD} the inbreeding coefficient of the offspring will also increase.

2.5.1 Pedigree-based inbreeding estimates in dairy cattle

According to Weigel (2001), inbreeding is unrelated to the size of the population. The rates of inbreeding in numerically small breeds (Dairy Swiss) are not very different from inbreeding rates in numerically large breeds (Holstein). However, the impact of inbreeding on numerically small breeds could be more detrimental compared to the numerically large breeds (Maiwashe *et al.*, 2006; Stachowicz *et al.*, 2011). This is because the loss of genetic diversity in a numerically small population may be higher compared to the numerically large populations due to the small number of animals (Maiwashe *et al.*, 2006; Stachowicz *et al.*, 2011). Levels of inbreeding reported in literature are shown in Table 2.5.

Table 2.5 A summary of previously reported pedigree based inbreeding coefficients for different dairy breeds

Breed	**Inbreeding coefficient**	**Reference**	**Country**
Holstein	0.075	Kelleher *et al.*, 2017	Ireland
	0.056	Van Raden *et al.*, 2011	
	0.060	Maiwashe *et al.*, 2006	South Africa
	0.032	Sewalem *et al.*, 2006	Canadian
	0.026	Wiggans *et al.*, 1995	United States
Jersey	0.173	Kelleher *et al.*, 2017	Ireland
	0.061	Van Raden *et al.*, 2011	
	0.070	Maiwashe *et al.*, 2006	South Africa
	0.036	Sewalem *et al.*, 2006	Canadian
	0.035	Wiggans *et al.*, 1995	United States
Ayrshire	0.050	Maiwashe *et al.*, 2006	South Africa
	0.040	Sewalem *et al.*, 2006	Canadian
	0.047	Wiggans *et al.*, 1995	United States
Brown Swiss	0.080	De Ponte-Bouwer *et al.*, 2013	South Africa
	0.049	Van Raden *et al.*, 2011	
	0.030	Wiggans *et al.*, 1995	United States

In a study done by Maiwashe *et al.* (2006) it was reported that the breed with the highest annual inbreeding rate was the Jersey breed (0.07%), this was followed by the Holstein breed (0.06%) and the Ayrshire having an annual inbreeding rate of 0.05%. Although inbreeding rates for SA dairy breeds are lower than for their global counterparts (Table 2.5), intense inbreeding may still occur occasionally and if not managed may become a cause for concern. Nicholas (1989) reported that the critical level of inbreeding per year is estimated at approximately 0.5%. If this level is exceeded within breed genetic variation will be adversely affected and will result in a reduction in the effective population size (N_e) (Weigel, 2001).

The International Bull Evaluation Service (INTERBULL) makes estimated breeding values (EBVs) available for use on an international scale (Weigel, 2001; Kearney *et al.*, 2004), which enables breeders to make decisions on importing semen from foreign sires. Due to the intense pressure on breeders for improved production several South African semen companies are importing semen from high ranking foreign sires and this has led to an 'overuse' of a small number of elite sires (Maiwashe *et al.*, 2006; Rodriguez-Ramilo *et al.*, 2015) and thus led to a decrease in the male N_e (Fleming *et al.*, 2018). Maiwashe *et al.* (2006) reported that in the period between 2000 and 2003, 25 to 36% of progeny from the four major South African dairy breeds (Ayrshire, Holstein, Jersey and Guernsey) were sired by foreign sires. This is made possible by the importation of semen mainly from the USA, Australia, Austria, Canada and Germany (Maiwashe *et al.*, 2006). For Canadian cattle, Stachowicz *et al.* (2011) reported that 10 Canadian Holstein (2000 to 2008) and 10 Canadian Jersey (2000 to 2007) ancestors of high genetic merit contributed to 62% and 60% respectively of the gene pool. This corresponds to estimates of Weigel (2001), who reported that globally 50% of Holstein offspring originated from the 10 most popular sires.

The pedigree-based inbreeding coefficient (F_{PED}) for the most inbred individual animals estimated in 2006 for the four respective South African cattle breeds is shown in Table 2.6. These animals were born in 1985 for the Holstein, 1991 for the Jersey and 1988 for the Ayrshire (Maiwashe *et al.*, 2006) and 2000 for the SA Dairy Swiss (De Ponte-Bouwer *et al.*, 2013).

Pedigree depth for South African dairy cattle increased during the period 1975 to 1985 due to an increase in pedigree recording (Maiwashe *et al.*, 2006) The number of parents known increased to 70% for the Guernsey and 90% for the Jersey for the same period. Holstein had the most parents known, and Ayrshire was intermediate (Maiwashe *et al.*, 2006). Even with this increase in pedigree recording and with accurate pedigree data, the avoidance of inbreeding is still difficult (Bjelland *et al.*, 2013). Inbreeding estimates based on pedigree information has

several disadvantages that may lead to under-estimation of the inbreeding coefficient. Firstly, pedigree errors may occur as a result of misidentification and incorrect or incomplete pedigree recording (Cassell *et al.*, 2003; Bjelland *et al.*, 2013; Pryce *et al.*, 2014).

Table 2.6 Pedigree-based inbreeding coefficients reported for the most inbred individuals in SA Dairy cattle

Breed	Inbreeding coefficient	Reference
Holstein	41.2	Maiwashe *et al.*, 2006
Jersey	42.2	Maiwashe *et al.*, 2006
Ayrshire	37.5	Maiwashe *et al.*, 2006
SA Dairy Swiss	39.2	De Ponte-Bouwer *et al.*, 2013

*Base population

F_{PED} is the expected proportion of the genome that is IBD, does not take the stochastic nature of inheritance into account and estimates inbreeding using recorded pedigree data (Ferenčaković *et al.*, 2011; Curik *et al.*, 2014). During meiosis, grandchildren inherit varying proportions of DNA from their grandparents. Thus, the offspring of first cousins are expected to have an inbreeding coefficient of 0.0625 and a standard deviation of 0.0243 (Carothers *et al.*, 2006). This variance increases with each meiosis event, making it possible for the offspring of third cousin matings to be more autozygous than offspring of second cousin matings (McQuillan *et al.*, 2008). Due to this, F_{PED} is only an approximate estimate of the individual level of autozygosity. Another disadvantage of F_{PED} is that the proportion of an individual's genome that is IBD is estimated relative to that of a poorly characterized founder generation (McQuillan *et al.*, 2008; Ferenčaković *et al.*, 2011; Fernández and Bennewitz, 2017). The reference population is used as the founder generation and founders or individuals not represented in the pedigree are assumed to be unrelated (Curik *et al.*, 2014; Fernández and Bennewitz, 2017). This is inaccurate as individuals in historical populations were often related (McQuillan *et al.*, 2008). Thus, F_{PED} does not capture ancient relatedness between individuals and therefore, may lead to under estimations of autozygosity. A third drawback of using F_{PED} is that it does not account for the potential bias that may be introduced during selection and assumes that the genome was not subjected to any selection (Curik *et al.*, 2014). Due to all the above-mentioned drawbacks with the use of pedigree-based inbreeding estimates it could be assumed that genomic inbreeding estimates should be more accurate in estimating inbreeding (Bjelland *et al.*, 2013; Rodriguez-Ramilo *et al.*, 2015).

2.5.2 Genomic estimates of inbreeding

Genomic estimation of inbreeding is expected to be more accurate compared to pedigree-based estimates (Bjelland *et al.*, 2013; Rodriguez-Ramilo *et al.*, 2015). This can be attributed to genomic estimation of inbreeding reflecting the realized proportion of homozygous loci for an animal, whereas pedigree estimates provide only expected values (Gómez-Romano *et al.*, 2016). In other words, the genomic calculation is based upon the percentage of the genome that is homozygous for an animal (Rodriguez-Ramilo *et al.*, 2015). Although genomic inbreeding estimation is expected to be more accurate compared to pedigree-based inbreeding, there are certain limitations.

Genomic inbreeding is calculated from a SNP-derived genomic relation matrix (GRM) (Van Raden, 2008; Bjelland *et al.*, 2013) and is the realized proportion of the genome shared by two individuals (Goddard *et al.*, 2011). GRM is used to measure inbreeding by estimating the actual allele sharing as opposed to pedigree-based estimates relationships that only estimates fractions of the allele that are expected to be identical by descent (Van Raden *et al.*, 2011; Pryce *et al.*, 2014). An advantage of using GRM as a method to estimate inbreeding is that it is able to predict inbreeding more accurately than pedigree-based inbreeding even when only a few records of genotyped information is available (Pryce *et al.*, 2012). An example of this can be the estimation of inbreeding between full sibs or half sibs. Where pedigree relationships between full sibs or half sibs may be equal, the genomic relationship between them may vary (Pryce *et al.*, 2012). GRM estimates genomic inbreeding by examining identical by state (IBS) information marker by marker (Van Raden, 2008). One limitation of using GRM is that it does not distinguish between alleles that are IBD and IBS.

The GRM (G) can be calculated as follows (Van Raden *et al.*, 2011):

$$G = \frac{zz'}{2\sum p\,(1-p)'} - 1$$

where: Z = the matrix containing the subtraction of a base population allele frequency from the given marker values, contains the values 0 – $2p$ for homozygotes, 1 - $2p$ for heterozygotes and 2 – $2p$ for opposite homozygotes

p = allele frequency

The Z matrix contains values of +1 or -1 for homozygotes and 0 for heterozygotes, which makes F_{GRM} a measure of homozygosity which has been transformed to follow a distribution that

is similar to that of F_{PED}. The values on the diagonal element of G denote the relationship of the animal to itself or its genomic inbreeding coefficient (Bjelland *et al.*, 2013).

In a study by Van Raden *et al.* (2011) genomic relationship matrix inbreeding coefficient (F_{GRM}) values with F_{PED} values were compared and correlations of 0.59 for the Holsteins, 0.68 for the Jerseys and 0.61 for the Brown Swiss were reported. This was supported by Hayes and Goddard (2010) who reported correlations of 0.69 between F_{PED} and F_{GRM} in Australian Angus bulls. Table 2.7 contains SNP derived inbreeding coefficients for the four respective breeds.

Table 2.7 Summary of inbreeding coefficients based on the genomic relationship matrix reported in literature

Breed	F_{GRM} / F_{IS} / F_{SNP}	Reference	Country
Holstein	0.023	Zhang *et al.*, 2015a	Danish
	0.101	Marras *et al.*, 2015	Italy
	0.164*	Van Raden *et al.*, 2011	
	0.328#	Van Raden *et al.*, 2011	
Dairy Swiss / Brown	0.196	Marras *et al.*, 2015	Italy
Swiss	0.070*	Van Raden *et al.*, 2011	
	0.341#	Van Raden *et al.*, 2011	
Jersey	-0.062	Zhang *et al.*, 2015a	Danish
	0.081*	Van Raden *et al.*, 2011	
	0.433#	Van Raden *et al.*, 2011	

*Base population

#0.5 Allele frequency

In a study done by Van Raden *et al.* (2011) F_{GRM} estimates based on the base population was lower than estimates where F_{GRM} was adjusted to account for alleles shared (this is equivalent to using an allele frequency of 0.5) (Table 2.7). Allele frequencies had an effect on the diagonal element of the GRM in the base population, which may result in elevated inbreeding coefficients (Van Raden *et al.*, 2011; Bjelland *et al.*, 2013; Pryce *et al.*, 2014). As a possible solution to overcome this problem GRM could be calculated with allele frequencies fixed at 0.5 (Van Raden *et al.*, 2011; Bjelland *et al.*, 2013). Using allele frequencies of 0.5 results in higher correlations between F_{PED} and F_{GRM} (Van Raden *et al.*, 2011; Bjelland *et al.*, 2013).

The use of the genomic relationship matrix as a method to control inbreeding may result in larger reduction in the frequency of homozygous minor alleles than pedigree-based methods,

and thus increasing the heterozygosity of SNPs not under selection (Pryce *et al.*, 2012). Pryce *et al.* (2012) reported that inbreeding can be reduced almost twice as much when using the genomic relationship matrix compared to pedigree-based inbreeding at the same rate of genetic gain. Thus, the use of GRM may be an effective method to reduce inbreeding within a population with a minimal effect on the rate of genetic gain for the given population. Both pedigree-based inbreeding and genomic inbreeding estimates may have disadvantages in that pedigree data may have pedigree errors or lack pedigree depth (Pryce *et al.*, 2014), whereas genomic data from SNP data may be subject to errors due to incorrect sample identification (Pryce *et al.*, 2012).

An alternative method to estimate genomic inbreeding or alternatively assess genetic diversity (McQuillan *et al.*, 2008; Keller *et al.*, 2011; Fleming *et al.*, 2018) has recently been proposed, which involves genomic runs of homozygosity (ROH). ROH is able to distinguish between markers that are IBD and identical by state (IBS) (Pryce *et al.*, 2012). F_{ROH} strongly correlates with homozygosity, making it a more preferable inbreeding measure than F_{PED} and other genomic inbreeding methods (Keller *et al.*, 2011).

The F_{ROH} can then be calculated as follows (McQuillan *et al.*, 2008):

$$F_{ROH} = \frac{\sum F_{ROH}}{\sum L_{AUTO}}$$

where: $\sum F_{ROH}$ = the length of ROH in one individual

$\sum L_{AUTO}$ = the length of the genome covered by SNPs

The software package PLINK has been shown to be useful for estimation of ROH (Purcell *et al.*, 2007). PLINK uses a sliding window of 50 SNPs, in one SNP intervals, across the genome to estimate homozygosity (Purcell *et al.*, 2007). In PLINK a few parameters can be changed to account for genotyping error, i.e. the sliding window, number of missing genotypes allowed, the number of heterozygous genotypes allowed, the number of SNPs constituting a ROH, the minimum length of an ROH and the number of overlapping windows (Purcell *et al.*, 2007). The decision of which parameters to use is important as these parameters influence the estimation of an ROH (Fernández and Bennewitz, 2017). A few authors have highlighted the importance of setting these parameters and reported that the minimum number of SNPs that is used to define an ROH should be defined according to the available SNP density as this will influence autozygosity estimates (Bjelland *et al.*, 2013; Ferencăković *et al.*, 2013b; Signer-Hasler *et al.*, 2017). The criteria used to detect ROH differs between studies especially for the minimum length used to define an ROH and the minimum number of SNPs allowed, thus making it difficult to

compare results from different studies (Mastrangelo *et al.*, 2016; Signer-Hasler *et al.*, 2017). Few F_{ROH} studies has been performed in dairy cattle, but the strong correlation between F_{PED} and F_{ROH} may see this approach being strongly incorporated into animal breeding.

ROH lengths follow an exponential distribution with mean (Fisher, 1954):

$$= \frac{100}{2g}\ x\ cM$$

where: g = the number of generations since the last common ancestor

Short ROH segments (~ 1 Mb) indicate a common ancestor many generations ago (Kirin *et al.*, 2010; Howrigan *et al.*, 2011; Purfield *et al.*, 2012). Long ROH segments (~ 10 Mb) indicate a more recent common ancestor and thus an indication of more recent inbreeding (Kirin *et al.*, 2010; Howrigan *et al.*, 2011; Purfield *et al.*, 2012). If we assume that 1 cM equals 1 Mb, then an ROH length of 25, 10 and 2.5 Mb, may for example indicate inbreeding due to a common ancestor 2, 5 and 20 generations ago, respectively (Howard *et al.*, 2017). Inbreeding due to a recent common ancestor will have a larger effect on animal performance than a distant common ancestor, as selection has not had time to remove unfavorable alleles from the population (Holt *et al.*, 2005). In studies done by Purfield *et al.* (2012) and Ferenčaković *et al.* (2013a) it was found that ROH segments shorter than 4 Mb are less likely to be due to IBD compared to longer segments and that ROH length influences the inbreeding estimates.

A few studies have focused on ROH as a way to estimate inbreeding in cattle (Sölkner *et al.*, 2010; Ferenčaković *et al.*, 2011; Purfield *et al.*, 2012). All of these studies confirmed that ROH-based inbreeding estimates are more accurate than pedigree-based estimates. Rodriguez-Ramilo *et al.* (2015) reported a correlation of 0.57 between F_{ROH} and $F_{PED.}$ This corresponds to results from Ferenčaković *et al.* (2013a) who reported correlations ranging from 0.50 to 0.67 for the Dairy Swiss. Correlations of 0.71 to 0.75 was reported between F_{ROH} and F_{PED} for $F_{ROH > 10000\ kb}$ and F_{ROH} respectively (Purfield *et al.*, 2012). Marras *et al.* (2015) reported correlations of 0.588 and 0.561 between F_{PED} and F_{ROH} for the Italian Brown and Holstein, respectively, at an ROH length of 16 Mb. When longer ROH segments ($F_{ROH > 16Mb}$) are considered, the correlation between F_{PED} and F_{ROH} are more closely related because longer ROH segments indicate recent inbreeding (Marras *et al.*, 2015). A difference between F_{PED} and F_{ROH} may be attributed to the fact that ROH captures both recent and ancient inbreeding, whereas F_{PED} only estimates inbreeding based on recorded pedigree which may only extend a few generations back (Marras *et al.*, 2015). This is supported by McQuillan *et al.* (2008) for humans and by Van Raden *et al.* (2011) and

Ferenčaković *et al.* (2013a) for cattle. In this regard methods utilizing GRM to estimate inbreeding gives a more uniform estimate of inbreeding (Marras *et al.*, 2015). Marras *et al.* (2015) also showed that the correlation between F_{GRM} and F_{ROH} is lower than the correlation between F_{PED} and F_{ROH}, this was also reported by Keller *et al.* (2011). The correlation between F_{GRM} and F_{ROH} reported by Marras *et al.* (2015) was 0.079 and 0.390 for the Italian Brown and Holstein, respectively, at an ROH length of 16 Mb. A study done by Pryce *et al.* (2012) reported a low correlation between the GRM and ROH. This could be due to the fact that ROH is able to account for more subtle differences in the genome than GRM due to the shared runs of homozygosity (Pryce *et al.*, 2012). Therefore, individuals that may have very similar GRM values may have different ROH values due to different genetic backgrounds. Table 2.8 shows previous F_{ROH} estimated inbreeding coefficients for four dairy cattle breeds from different countries.

It is important to note that the frequency of ROH across the genome is not uniform across all populations, thus some regions of the genome may have a higher or lower ROH frequency compared to other regions (Purfield *et al.*, 2012). Certain ROH are common in all breeds and can give information about the history of the given population (Purfield *et al.,* 2012). ROH can also be used to infer population history in the absence of conclusive population information (McQuillan *et al.*, 2008). Intense selection pressure in combination with small effective population sizes in dairy cattle populations, has led to an increase in the frequency of homozygous segments across the genome (Kim *et al.*, 2015). Studies have found that ROH are enriched with deleterious variants within both the human and cattle genome, but that the length at which the highest frequency of deleterious mutations occur is longer in humans than in cattle (Zhang *et al.*, 2015b). Deleterious variants within long ROH segments mainly arise due to rare IBD haplotypes, found in the ROH, combining at low frequencies (Howard *et al.*, 2017). These low frequency variants are more likely to be deleterious than more commonly found variants (Howard *et al.*, 2017). Pryce *et al.* (2014) reported that 0.66 and 0.72% of the genome for Holstein and Jersey respectively, was homozygous and that Holstein cattle had shorter ROH segments than Jersey cattle. This could indicate that Jersey cattle are more inbred than Holstein, but could also be due to ascertainment bias, as many of the SNPs used to discover the genotyping panel were discovered using Holstein cattle (Matukumalli *et al.*, 2009). Increases in genomic inbreeding can be associated with a reduction in lactation performance in Holstein cattle (Bjelland *et al.*, 2013). This is supported by Pryce *et al.* (2014), who found that shorter ROH segments had a less unfavorable effect on lactation than longer ROH segments.

Table 2.8 F_{ROH} inbreeding coefficients reported in literature for dairy cattle at different ROH lengths (Mb) from different countries

Breed	Inbreeding coefficient	Reference	ROH length (Mb)	Country
Holstein	0.058	Signer-Hasler *et al.*, 2017		Switzerland
	0.077	Rodriguez-Ramilo *et al.*, 2015	4	Spain
	0.066	Zhang *et al.*, 2015a		Denmark
	0.073	Marras *et al.*, 2015	16	Italy
	0.026	Marras *et al.*, 2015	4	Italy
	0.042	Mastrangelo *et al.*, 2016		
Dairy Swiss / Brown Swiss	0.091	Signer-Hasler *et al.*, 2017		Switzerland
	0.09	Gomez-Romano *et al.*, 2016	4	Austria
	0.097	Marras *et al.*, 2015	4	Italy
	0.034	Marras *et al.*, 2015	16	Italy
	0.103	Ferenčaković *et al.*, 2013b	4	
	0.039	Ferenčaković *et al.*, 2013b	16	
Jersey	0.070	Zhang *et al.*, 2015a		Denmark

2.6. Other Applications of Genomic Information

In the following section a short discussion will follow focusing on other applications of genomic information with regards to the dairy industry. Details with regards to selection signatures and genomic selection is beyond the scope of this study but a short overview of the impacts and possible advantages will be discussed in this section.

2.6.1 Selection Signatures

Artificial selection for increased milk and meat production and disease resistance has occurred since the domestication of cattle (Zhao *et al.*, 2015; Randhawa *et al.*, 2016). This together with recent inbreeding and decreased population sizes in dairy breeds worldwide has resulted in changes in the cattle genome (Zhao *et al.*, 2015; Makina, 2015). Positive selection pressure for a specific trait of interest alters the frequency of the favorable alleles over time and thus increases the frequency of the favorable alleles within the genome, this is known as selective sweeps or selection signatures (Purfield *et al.*, 2017; Zhao *et al.*, 2015). With the development of the genomic era and new computational biology tools it is now possible to characterize the impacts of selection on the genome (Nielsen, 2005). Previous studies found selective sweeps on a number of chromosomes, such as BTA-6 (*ABCG2*, Casein cluster) (Cohen-Zinder *et al.*, 2005), BTA-14 (*DGAT1*) and BTA-20 (*GHR*) (Cohen-Zinder *et al.*, 2005; Lillehammer *et al.*, 2008) in multiple dairy breeds.

Zhang *et al.* (2015b) reported that ROH patterns across the genome are not randomly distributed but are shared among individuals most likely as a result of selection events. Thus, ROH can be used as a method to explore selection signatures. The uncovering of selection signatures is important as it may lead to a better understanding of the mechanisms underlying economically important traits under both natural and artificial selection (Akey *et al.*, 2002). Table 2.9 contains some of the candidate genes that has been identified in dairy cattle to date, as well as the trait that is associated with.

Table 2.9 Candidate genes and their associated traits identified in dairy cattle

Trait	Candidate Gene	Reference
Coat color	*MC1R, KIT*	Randhawa *et al.*, 2016
Stature	*NCAPG, LCORL, PLAG1*	Randhawa *et al.*, 2016
Reproduction	*MGAT1, FGF1*	Randhawa *et al.*, 2016
Milk production	*ABCG2, DGAT1, GHR,*	Randhawa *et al.*, 2016
	AGTRAP	Stella *et al.*, 2010
Feed Efficiency	*R3HDMI, ZRANB3*	Gautier and Naves, 2011;
		Qanbari *et al.*, 2011
Disease resistance	*HBEGF*	Stella *et al.*, 2010

The *PLAG1* gene that is associated with stature, has also been found to be associated with fertility traits at the region of BTA-14 (Randhawa *et al.*, 2016). The *ABCG2* gene which is involved in milk yield has also been found to be involved as a lactation regulator in dairy cattle (Sheehy *et al.*, 2009). The *MGAT1* and *FGF1* genes are strongly implicated with reproduction traits due to their association in fertilization as well as embryonic development and growth (Qanbari *et al.*, 2011).

2.6.2 Genomic Selection

Dairy cattle breeding has undergone a significant change since genomic information became available (Hayes *et al.*, 2009) as a tool to predict breeding values in a cost-effective manner (Meuwissen *et al.*, 2001). Genomic selection (GS) has revolutionized the dairy industry and has led to an increase in the number of animals, male and female, that undergo routine genotyping as potential parents for the next generation of offspring. The demonstration by Meuwissen *et al.* (2001) that accurate selection decisions can be made from the use of dense DNA-marker data without phenotypic data has led to the implementation of genomic selection in modern dairy breeding programs world-wide (Hayes *et al.*, 2009). Genomic selection programs make use of a prediction equation from a reference population of individuals that has genotype and phenotype data available, to predict breeding values in populations without phenotypic data (Meuwissen *et al.*, 2001; Hayes and Goddard, 2010).

Genomic selection can be used to double the rate of genetic gain per generation (Hayes *et al.*, 2009; De Roos *et al.*, 2011). This is supported by other researchers who reported that GS may lead to a 50% increase in genetic gain (Pryce *et al.,* 2010). This increase in genetic gain will be greatest in traits that are expressed late in life and for sex limited traits e.g. female reproduction (Hayes and Goddard, 2010) and all milk-associated traits. Schaeffer (2006) reported that increases in genetic gain is due to selection based on GEBV, as it is more accurate than selection decisions based on parent averages (Clark *et al.*, 2013). This is supported by Van Raden (2008), who reported that GS is more reliable for young animals than parent averages (Hayes *et al.*, 2009; Hayes and Goddard, 2010; Pryce and Daetwyler, 2012). Accurate selection decisions can therefore be made for young animals without records (Hayes *et al.*, 2009; Pryce *et al.*, 2010), resulting in young bulls to be selected for breeding as soon as they are physiologically able to reproduce, leading to a shorter generation interval (Hayes *et al.*, 2009; Hayes and Goddard, 2010).

Results from studies with simulated data suggest that inbreeding per generation should be reduced when selection decisions are based on GEBV (Hayes and Goddard, 2010; Clark *et al.*, 2013). This is because full sibs will receive different EBVs due to Mendelian sampling (Pryce and Daetwyler, 2012; Clark *et al.*, 2013). With traditional parent averages, full sibs receive the same mid-parent value, thus increasing the selection of relatives (Pryce and Daetwyler, 2012). Selection based on GEBV should thus increase genetic gain, especially for lowly heritable traits, whilst maintaining genetic variation (Hayes and Goddard, 2010). However, a reduction in the generation interval can also lead to an increase in inbreeding per generation due to strong selection on a small number of young animals with high GEBV (Pryce *et al.*, 2010; De Roos *et al.*, 2011; Fleming *et al.*, 2018).

Although genomic selection has led to an increase in the number of bulls sampled for testing, the number of bulls used as parents for the next generation has remained relatively constant. Fifty percent of the young bulls used in AI programs are still being sired by the same number of bulls before the implementation of genomic selection (Miglior and Beavers, 2014).

2.7 Conclusion

A general increase in inbreeding can be seen in the world-wide dairy cattle populations. This is a concern as an increase in inbreeding also leads to a reduction in genetic variation. This increase in inbreeding has been the result of strong selection pressure on dairy cattle for increased milk production. Genomic inbreeding estimations together with the implementation of selection programs can help to reduce inbreeding.

Chapter 3

Material and Methods

3.1 Introduction

In this study genotypic data from four South African dairy breeds were analyzed. The genotypes for the breeds originate from the Dairy Genomics Program (DGP). Consent for use of the data as well as hair sample collection was received from the associated breed societies (South African Holstein Friesland, Jersey South Africa, Ayrshire Cattle Breeders' Society of South Africa and South African Dairy Swiss). Approval for hair sample collection from live animals (Dairy Swiss) as well as the use of external data obtained from the DGP was received from the Animal Ethics committee of the Faculty of Natural and Agricultural Sciences at the University of Pretoria (EC170627-135). In this study genomic inbreeding and effective population sizes for four South African dairy breeds were estimated.

3.2 Animal resources and data collection

The South African Holstein, Jersey and Ayrshire breeds are part of routine genotyping performed by the DGP. BTP DNA-extraction and genotyping for these animals were done using the Infinium Bovine SNP 50-24 V3.0 Beadchip (Illumina, Inc. San Diego, CA 92122 USA). Genotypes from the current active populations were received from the Logix SA Stud Book database to represent these breeds. These genotypes were received from breeders across South Africa (KwaZulu Natal, Limpopo, North-West Province, Gauteng, West-, East- and South Cape).

Sixty-two individual animals were sampled for the SA Dairy Swiss from three registered breeders. These breeders are located in Gauteng, Mpumulanga and North-West Province. Sixty tail hairs were collected per animal in a paper envelope according to the standard operating procedures (SOP) of the DGP. Samples were submitted to SA Stud Book, where they were verified, logged and sent to the ARC-Biotechnology Platform (Onderstepoort, 0110) for BTP DNA-extraction and genotyping using the Infinium Bovine SNP 50-24 V3.0 Beadchip (Illumina, Inc. San Diego, CA 92122 USA). In total 1 002 animals with genotypic information was included in the study, as shown is Table 3.1.

Table 3.1 Summary of the available genotypes representing the four dairy populations included in the study

Breed	Number of animals
Ayrshire	112
SA Dairy Swiss	62
Holstein	412
Jersey	416
Total	**1002**

Genotypic data was downloaded from a downstream link received by the ARC and put into SNP Convert v1.0 (Nicolazzi *et al.*, 2016) in order to convert the Illumina Genotype files into PLINK (Purcell *et al.*, 2007) files. The Illumina ROW format option of SNP Convert (Nicolazzi *et al.*, 2016) was used and the final report and SnpMap for the respective breeds were used to generate the PLINK (Purcell *et al.*, 2007) .ped and .map files.

3.3 Statistical Analysis

3.3.1. Quality Control (QC)

All statistical analyses performed in the current study, with the exception of ADMIXTURE and PCA, was first done on the individual populations, after which the four data sets were merged, and statistical analysis performed on the merged data set. Animal and marker-based quality control (QC) was performed using PLINK v1.90 (Purcell *et al.*, 2007) software. All non-informative SNPs and individual animals were removed using the following criteria. Animals were removed based on high rates of missing genotypes and deviating from the average heterozygosity rate by more than three standard deviations. PLINK's --missing command was then used to estimate the missing SNPs per individual outputted in an .imiss file. Heterozygosity data was generated using the --het command in PLINK, from which the heterozygosity rate per individual was calculated. Microsoft Excel (2016) was used to plot the missing genotypes against the heterozygosity rates per individual. This was done in order to establish the most appropriate threshold values in order to retain the maximum number of samples for further analysis.

SNPs with an animal call rate below 90%, a marker call rate of 5% (< 95%) and minor allele frequency (MAF <0.02) were removed from further analysis. Only the 29 autosomal chromosomes were included in the analysis. Hardy-Weinberg was not included in the QC in order

to have as many markers included in the analysis as possible and to allow for the estimation of runs of homozygosity.

3.3.2 Linkage based pruning

Linkage disequilibrium-based (LD) pruning was done to remove bias due to markers that are closely linked. This enables analyses to be performed on a subset of SNPs with the maximum number of independent markers (Davis *et al.*, 2011). LD pruning was performed using PLINK's --indep-pairwise (50 5 0.8) command. PLINK uses the Expectation Maximization (EM) algorithm, proposed by Excoffier and Slatkin (1995) in order to calculate r^2 and D' values between SNPs. LD pruning was done only for the estimation of H_o, H_E, r^2, and admixture. ROH was calculated before LD pruning in order to accurately estimate the amount of homozygosity and thus inbreeding estimates based on ROH.

3.3.3 Summary statistics

Individual inbreeding and heterozygosity were calculated using the --het command in PLINK before and after LD pruning. The following formulas were used to estimate H_o and H_E:

$$H_O = \frac{(N(NM) - O(Hom))}{N(NM)}$$

$$H_E = \frac{(N(NM) - E(Hom))}{N(NM)}$$

where N(NM) represents the number of non-missing genotypes, O(Hom) represents the observed number of heterozygotes and E(Hom) represents the expected number of heterozygotes.

Minor allele frequency (MAF) values were calculated for each breed separately before LD pruning using PLINK's --freq command. The minor allele frequency (MAF) distribution was calculated using PLINK's --maf command, at different MAF intervals (0.02; 0.05; 0.1; 0.2; 0.3; 0.4; 0.5) in order to indicate the most appropriate MAF value. Microsoft Excel (2016) was used to graph the number of SNPs available for further analysis at each minor allele frequency threshold (0.02; 0.05; 0.1; 0.2; 0.3; 0.4; 0.5).

LD, measured as r^2, was estimated for each individual breed using PLINK's --r2 command. r^2 estimates were calculated for all autosomal SNPs that passed quality control before and after LD pruning.

3.3.4 Merging of data sets

The four data sets were merged using --merge command in PLINK, and QC was performed again on the one data set. Quality control as well as LD pruning was performed according to the threshold parameters described above and binary (FAM., BED. And BIM) files were obtained for further analysis. After QC there were 994 animals remaining (six animals were removed from the Holstein and two animals from the Jersey population) and 43 976 SNPs. After LD pruning 39 999 SNPs remained for analysis. The minor allele frequency (MAF) distribution was calculated at the same intervals as for the individual data sets.

3.3.5 Runs of Homozygosity (ROH)

ROH were calculated for each respective population, after which it was calculated for the merged data set. ROH was estimated using PLINK v1.90 (Purcell *et al.*, 2007) software. PLINK uses a sliding window of 50 SNPs, in one SNP intervals, across the genome in order to estimate homozygosity. ROH was calculated according to the parameters used by Purfield *et al.* (2012). Using the --homozyg-window-het 1, command no more than one possible heterozygous genotype was allowed and no more than two missing genotypes were allowed per window using the --homozyg-window-missing 2, command. The minimum SNP density was set to 1 SNP every 120 kb with the –homozyg-density 100 command with no restriction placed on the minimum number of SNPs in a ROH. The maximum gap length allowed between two consecutive SNPs was no more 1000 kb, which is the default value used by PLINK. ROH was also calculated at different lengths according to different length categories used by Ferenčaković *et al.* (2013a). These ROH lengths were defined by using the --homozyg-kb command and are as follows: ROH1 = 1000 kb, ROH2 = 2000 kb, ROH4 = 4000 kb, ROH8 = 8000 kb, ROH16 = 16 000 kb. The mean sum of ROH was estimated for each respective population and presented on a graph using Microsoft Excel (2016). F_{ROH} was estimated using the following formula:

$$F_{ROH} = \frac{\sum F_{ROH}}{\sum L_{AUTO}}$$

where: $\sum F_{ROH}$ = the length of ROH in one individual

$\sum L_{AUTO}$ = the length of the genome covered by SNPs

3.3.6 Genetic Relatedness

PLINK v1.90 (Purcell *et al.*, 2007) was used to estimate eigenvectors and eigenvalues for the principal component analysis (PCA). The --pca 4 command was used to obtain the eigenvalues and eigenvectors for the first four principal components. The .eigenvectors file was then converted to a .evec file with a plink2evec command so that the file could be exported to Genesis version 0.2.3. (Buchmann and Hazelhurst, 2014). Genesis was used to generate PCA plots.

3.3.7 Population structure analysis

ADMIXTURE version 1.23 (Alexander *et al.*, 2009) was used in order to determine the most appropriate K-value. Since the data set consisted of four breeds, K-values of 1-5 were used. ADMIXTURE was run using the --cv command. ADMIXTURE determines the K-value by following a cross-validation (CV) procedure. For each K-value, a .P file with the allele frequencies as well as a .Q file with the ancestry fractions is obtained. After CV errors were estimated for each K-value, the lowest K-value or the K-value at which the inflection occurs, when visualized on a graph, indicates the optimal K-value to be used for population structure analysis.

Genesis version 0.2.3. (Buchmann and Hazelhurst, 2014) was used to construct population structure plots for the appropriate K-value. Genesis requires a .Q file, generated by ADMIXTURE, as well as a PLINK .fam file for the construction of the bar plot.

3.3.8 Effective Population Size (N_E)

The effective population sizes (N_E) were estimated for each of the four respective populations using the SNeP v1.1 tool (Barbato *et al.*, 2015). N_E was calculated using a MAF of 0.02 and Microsoft Excel (2016) was used to construct a line graph to visualize the effective population sizes of the four breeds.

Chapter 4

Results

4.1. Introduction

This chapter presents the results obtained from the analyses described in Chapter three. Genotypic data representative of four South African dairy breeds were used to calculate genomic inbreeding estimates and effective population size for each of the respective populations.

4.1.1 Quality Control

Animal-based quality control was first performed in order to remove individuals with high levels of missing genotypes and heterozygosity rates deviating with more than three degrees of freedom from the average heterozygosity rate. Figures 4.1a to 4.1d shows the heterozygosity rate plotted against the proportion of missing genotypes for each of the four populations respectively.

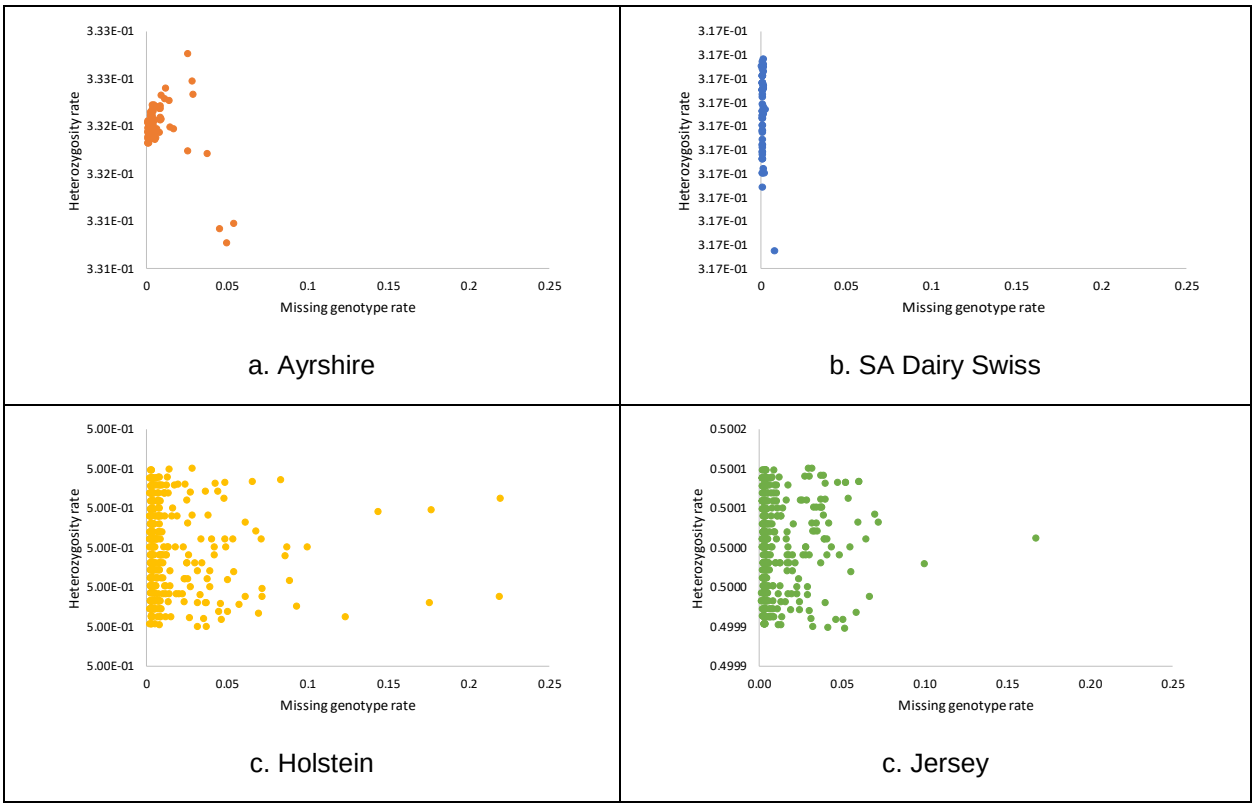

Figure 4.1 Heterozygosity rate and proportion of missing genotypes for the four respective populations

All Ayrshire and SA Dairy Swiss samples fell well below the 10% missing genotype rate, respectively (Figure 4.1). The Holstein and Jersey data sets had six and two samples, respectively that exceeded the 10% missing genotype call rate. Given the above figures, an animal call rate threshold was set at 90%. The results following animal-based quality control were summarized in Table 4.1.

Table 4.1 Number of individuals removed and average call rates after sample-based quality control

Breed	Number of individuals removed (Sample call rate < 90%)	Average call rate
Ayrshire	0	0.944
SA Dairy Swiss	0	0.999
Holstein	6	0.988
Jersey	2	0.989

After animal-based quality control the six Holstein and two Jersey animals failing QC were removed from further analysis. Table 4.2 contains the number removed and remaining after marker-based quality control.

Table 4.2 Number of SNPs removed and remaining after marker-based quality control

Breed	SNP call rate (<95%)	SNP MAF (<8%)	Total SNPs removed	SNPs available for analysis
Ayrshire	1241	8691	9932	41 487
SA Dairy Swiss	206	8755	8961	42 259
Holstein	2944	6808	9752	41 667
Jersey	3208	11 327	14 535	36 884

As shown in Table 4.2 the Jersey data set had the most SNPs removed due to the SNP call rate. The data set with the highest number of SNPs left for further analysis was the SA Dairy Swiss.

After individual QC of the four respective breeds, the data sets were merged, and animal and marker-based QC was performed on the merged data. One-thousand-and-two cattle and 51 419 SNPs were available for quality control after the merging of the data sets. Figure 4.2 shows the choice for the most appropriate animal call rate.

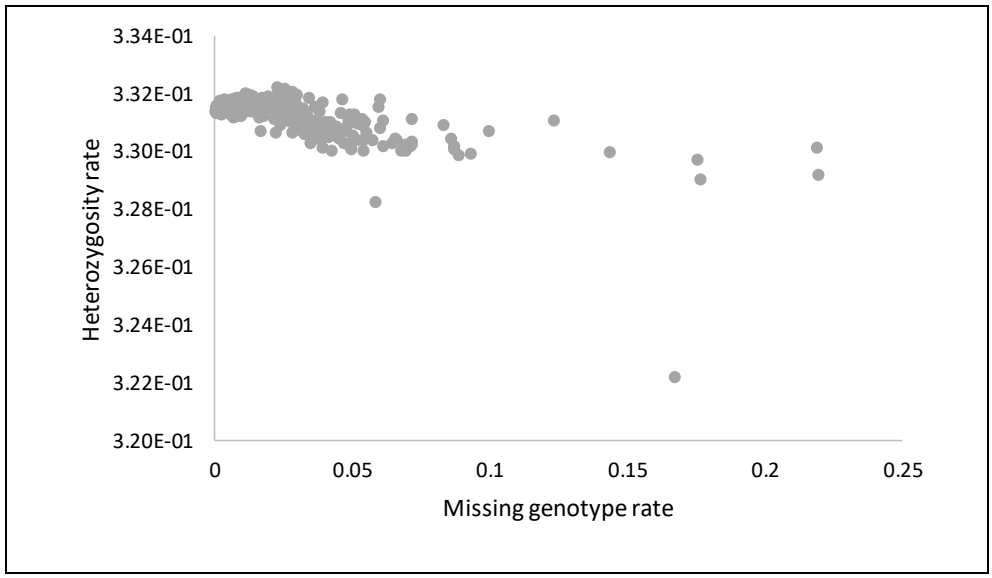

Figure 4.2 Heterozygosity rate and proportion of missing genotypes for the merged populations

From Figure 4.2 an animal call rate threshold was set at 90% and eight individuals were removed from further analysis for the merged data set. After marker-based QC 2 476 and 4 967 SNPs were removed at a marker call rate of 95% and a MAF of 8%, respectively.

Minor allele frequencies (MAF) were calculated for all four dairy populations respectively, before quality control. This was done in order to observe the distribution of the complete set of SNPs within the different MAF intervals for each respective population. From this an average MAF across all four populations were obtained and the uninformative SNPs removed from further analysis. Figure 4.3 illustrates the minor allele frequency ranging from 0.02-0.5 for the four different populations included in the study.

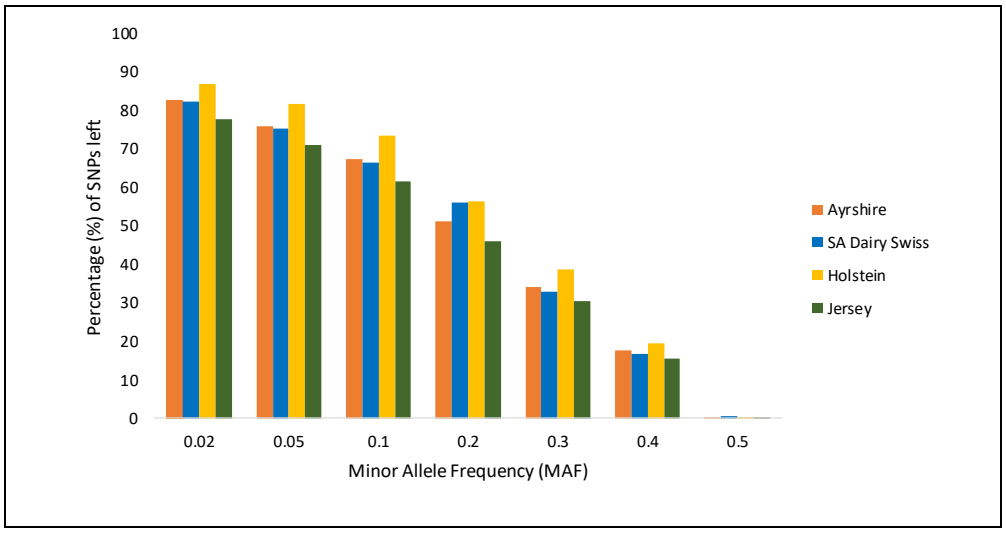

Figure 4.3 The percentage of SNPs remaining at different minor allele frequency (MAF)

The Holstein had on average a higher proportion of SNPs across all MAF ranges compared to the other three breeds. Table 4.3 contains the number of polymorphic loci, for the four respective populations, obtained in this study.

Table 4.3 Number of polymorphic loci obtained for the four dairy populations

Breed	Sample size	Number of polymorphic loci
Ayrshire	112	42 591
SA Dairy Swiss	62	42 433
Holstein	412	44 741
Jersey	416	39 997
Merged	994	44 680

The number of polymorphic loci for the four populations included in the study was very similar, with the exception of the Jersey which had the lowest number of polymorphic loci across the four populations. After marker and sample-based quality control there were 944 animals and 43 976 SNPs left for further analysis. An approximate set of 50 000 SNPs is required for PCA (Anderson *et al.*, 2010) and therefore linkage disequilibrium (LD) based pruning was only performed for the observed (H_O) and the expected (H_E) heterozygosity, as well as the r^2 linkage disequilibrium measure, in order to maintain the maximum number of SNPs for further analysis.

4.2 Population parameters

4.2.1 Observed and expected heterozygosity

Table 4.4 summarizes the average and median MAF values at a minor allele frequency of at least 0.02, as well as H_O and H_E before and after LD pruning.

Table 4.4 Summary statistics before and after LD pruning for the four respective dairy populations

Breeds	Average MAF	Median MAF	H_o*	H_o**	H_e*	H_e**
Ayrshire	0.253	0.250	0.355	0.349	0.342	0.341
SA Dairy Swiss	0.253	0.250	0.345	0.365	0.338	0.358
Holstein	0.271	0.277	0.359	0.357	0.357	0.355
Jersey	0.252	0.251	0.340	0.338	0.336	0.334
Merged	0.253	0.250	0.318	0.317	0.355	0.353

*Before Linkage Disequilibrium Pruning

**After Linkage Disequilibrium Pruning

Table 4.4 indicates that the Holstein had the highest (27%) average MAF before LD pruning. With the exception of the Holstein the other three populations had similar MAF values.

A limited loss of heterozygosity (H_E<H_O) can be seen across all four breeds. Due to the limited difference between H_O, H_E and F_{IS} before and after linkage disequilibrium (LD) pruning it was decided that all further analysis, with the exception of the linkage disequilibrium (r^2) measures, will be performed without LD pruning of the four respective data sets.

4.2.2 Linkage Disequilibrium (LD)

Table 4.5 contains the r^2 estimates for the four respective dairy breeds included in the study before and after linkage disequilibrium-based pruning.

Table 4.5 Linkage disequilibrium (r^2) measures before and after LD pruning obtained for the studied populations

Breed	r^2*	r^2**
Ayrshire	0.181	0.181
SA Dairy Swiss	0.291	0.177
Holstein	0.311	0.182
Jersey	0.350	0.187

*Before Linkage Disequilibrium Pruning

**After Linkage Disequilibrium Pruning

Linkage disequilibrium (r^2) measures for the four respective populations showed a large difference before and after LD pruning. The Jersey breed had the highest r^2 followed by the Holstein, while the Ayrshire showed the lowest r^2 value.

4.2.3 Inbreeding and effective population size (N_e) estimates

Table 4.6 shows the average inbreeding coefficients for the four respective populations as well as the inbreeding coefficients for the most and least inbred individuals for each population.

Table 4.6 Average inbreeding coefficients for the respective populations and most and least inbred individuals

Breeds	**F_{IS}**	**F_{IS} Most Inbred Individual**	**F_{IS} Least Inbred Individual**
Ayrshire	-0.039	-0.153	-0.0003
SA Dairy Swiss	-0.019	0.191	-0.004
Holstein	-0.007	0.164	-0.107
Jersey	-0.010	0.128	-0.219

As shown in Table 4.6 the SA Dairy Swiss had the highest inbred individual, followed by the Holstein. The Ayrshire had lowest inbred individual as well as the lowest average inbreeding coefficient across all four populations.

F_{ROH} was estimated across five different length categories in order to distinguish the degree of recent versus past inbreeding. F_{ROH} was calculated at 1000kb, 2000kb, 4000kb, 8000kb and 16 000kb, respectively. ROH was calculated before LD pruning, in order to ensure that all homozygous segments are accounted for. Longer ROH segments indicate more recent inbreeding, whereas shorter segments indicate past inbreeding. The mean sum of ROH, within each ROH length category, was calculated by summing the number ROH per animal, in each ROH length category and then averaging this per breed. Figure 4.4 contains the mean sum of ROH lengths per population.

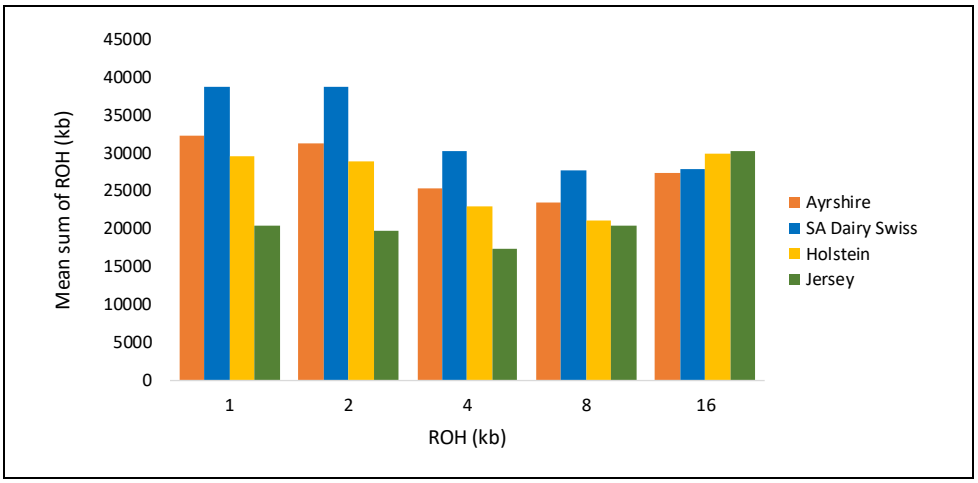

Figure 4.4 Mean ROH lengths per animal per population

A clear difference within the different ROH length categories can be observed between the populations (Figure 4.4). A general decrease in ROH was observed across all four populations between the ROH length categories of 1 000 kb to 8 000 kb, then followed by an increase in ROH segment coverage at an ROH length of 16 000kb. The four most homozygous animals in the study had on average 2 365.41 Mb of their genome classified as ROH.

Table 4.7 contains the genomic inbreeding coefficients (F_{ROH}) estimated at different ROH length categories as well as pedigree inbreeding coefficients (F_{PED}) for the four individual breeds. Pedigree inbreeding coefficients were received from SA Stud Book (2017, B. Mostert, Pers. Comm, bernice@studbook.co.za) for the Ayrshire, Holstein and Jersey. SA Stud Book does not currently have pedigree inbreeding coefficients available for the SA Dairy Swiss, thus F_{PED} estimates reported by De Ponte-Bouwer *et al.* (2013) were used in order to estimate correlations between the three different measures of inbreeding.

Table 4.7 F_{ROH} distribution and F_{PED} for the respective populations

Breeds	**F_{ROH1}**	**F_{ROH2}**	**F_{ROH4}**	**F_{ROH8}**	**F_{ROH16}**	**F_{PED}**
Ayrshire	0.064	0.068	0.095	0.144	0.227	0.049
SA Dairy Swiss	0.073	0.076	0.103	0.152	0.231	0.080
Holstein	0.065	0.068	0.097	0.154	0.252	0.022
Jersey	0.056	0.062	0.093	0.159	0.247	0.038

F_{ROH1} = 1Mb, F_{ROH2} = 2Mb, F_{ROH4} = 4Mb, F_{ROH8} = 8Mb, F_{ROH16} = 16Mb

From the genomic inbreeding estimates a general increase in the level of inbreeding was observed over the last few generations (Table 4.7). The Holstein had the highest (0.252) amount of recent (F_{ROH16}) inbreeding, while the Ayrshire had the lowest (0.227).

Table 4.8 contains the genomic (F_{ROH}) and pedigree (F_{PED}) inbreeding coefficients for the most inbred individuals of the four respective populations.

Table 4.8 F_{ROH} and F_{PED} for the most inbred individuals for the studied populations

Breeds	**F_{ROH}**	**F_{PED}**
	Most inbred individual	**Most inbred individual**
Ayrshire	0.851	0.990
SA Dairy Swiss	0.836	0.392
Holstein	0.866	0.417
Jersey	0.952	0.430

The population with the most inbred individual was the Jersey with an individual inbreeding coefficient of 0.95 (Table 4.8), while the Ayrshire had the most inbred individual when pedigree inbreeding estimates were used. Table 4.9 (a-d) contains the correlations between the respective inbreeding coefficients and the four populations included in the study.

Table 4.9a Correlation coefficients between the different inbreeding coefficients for the Ayrshire

	F_{IS}	$F_{ROH>1}$	$F_{ROH>2}$	$F_{ROH>4}$	$F_{ROH>8}$	$F_{ROH>16}$	F_{PED}
F_{IS}	1						
$F_{ROH>1}$	0.654	1					
$F_{ROH>2}$	0.640	0.996	1				
$F_{ROH>4}$	0.626	0.987	0.990	1			
$F_{ROH>8}$	0.407	0.716	0.724	0.738	1		
$F_{ROH>16}$	0.262	0.540	0.543	0.560	0.689	1	
F_{PED}	0.479	0.010	0.010	-0.003	-0.056	0.010	1

From Table 4.9a moderate to high correlation can be observed between F_{IS} and relatively short F_{ROH} lengths, ranging from 0.407 to 0.654. A lower correlation can be seen between F_{IS} and $F_{ROH>16Mb}$. Low correlations can be observed between F_{PED} and F_{ROH}.

Table 4.9b Correlation coefficients between the different inbreeding coefficients for the SA Dairy Swiss

	F_{IS}	$F_{ROH>1}$	$F_{ROH>2}$	$F_{ROH>4}$	$F_{ROH>8}$	$F_{ROH>16}$
F_{IS}	1					
$F_{ROH>1}$	0.454	1				
$F_{ROH>2}$	0.461	0.999	1			
$F_{ROH>4}$	0.493	0.989	0.990	1		
$F_{ROH>8}$	0.251	0.817	0.813	0.807	1	
$F_{ROH>16}$	0.071	0.523	0.513	0.490	0.611	1

The SA Dairy Swiss has the lowest correlation between F_{IS} and $F_{ROH>8}$ ranging from 0.071 to 0.454 amongst the four populations included in the study indicating low to moderate correlations. No pedigree inbreeding information was available for the SA Dairy Swiss; thus, correlations could not be estimated between F_{PED} and the other measures of inbreeding.

Table 4.9c Correlation coefficients between the different inbreeding coefficients for the Holstein

	F_{IS}	$F_{ROH>1}$	$F_{ROH>2}$	$F_{ROH>4}$	$F_{ROH>8}$	$F_{ROH>16}$	F_{PED}
F_{IS}	1						
$F_{ROH>1}$	0.594	1					
$F_{ROH>2}$	0.591	0.997	1				
$F_{ROH>4}$	0.596	0.988	0.990	1			
$F_{ROH>8}$	0.529	0.858	0.858	0.862	1		
$F_{ROH>16}$	0.355	0.538	0.543	0.554	0.621	1	
F_{PED}	0.256	0.180	0.169	0.186	0.168	0.053	1

Correlations between F_{IS} and F_{ROH} ranged from 0.355 to 0.594. Lower correlation was observed between F_{IS} and $F_{ROH>8}$ for the Holstein compared to the Ayrshire. Correlations between F_{PED} and F_{ROH} ranged from 0.053 to 0.180 for the Holstein population, with the highest correlation observed between F_{PED} and $F_{ROH>4}$.

Table 4.9d Correlation coefficients between the different inbreeding coefficients for the Jersey

	F_{IS}	$F_{ROH>1}$	$F_{ROH>2}$	$F_{ROH>4}$	$F_{ROH>8}$	$F_{ROH>16}$	F_{PED}
F_{IS}	1						
$F_{ROH>1}$	0.686	1					
$F_{ROH>2}$	0.683	0.998	1				
$F_{ROH>4}$	0.677	0.988	0.991	1			
$F_{ROH>8}$	0.602	0.904	0.905	0.911	1		
$F_{ROH>16}$	0.377	0.522	0.520	0.529	0.593	1	
F_{PED}	0.081	0.116	0.113	0.117	0.075	-0.039	1

From Table 4.9d the Jersey breed has the highest correlation between F_{IS} and $F_{ROH>8}$ compared to the other three populations. The correlation between F_{IS} and $F_{ROH>16}$ is the lowest of all correlations between F_{IS} and F_{ROH}, ranging from 0.071 in the SA Dairy Swiss to 0.377 in the Jersey. Figure 4.5 (a-d) shows the correlations between F_{IS} and F_{ROH} for the four dairy populations at an ROH length of 1000 kb. The Ayrshire had the lowest correlations between F_{PED} and $F_{ROH,}$ while the Holstein had the highest. The highest correlation between F_{PED} and F_{ROH} can be seen at an ROH length of 4000kb for the Holstein population.

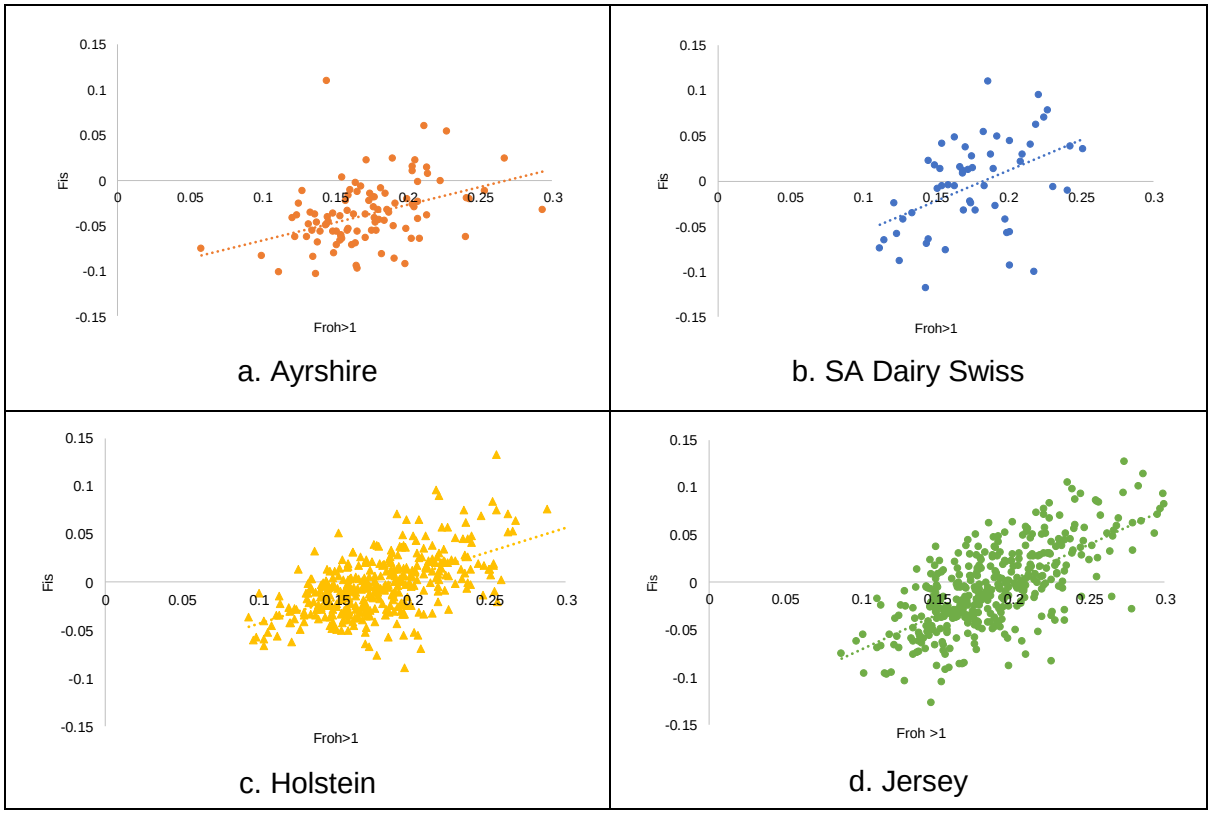

Figure 4.5 Scatterplots of F_{IS} vs F_{ROH} at an ROH length of 1000 kb for the populations

From Figure 4.5 the Jersey had the highest correlation between F_{IS} and $F_{ROH>1}$ among the four breeds. From Figure 4.5b the low to moderate correlations between F_{IS} and F_{ROH} for the SA Dairy Swiss can clearly be visualized. Figure 4.6 (a-c) below depicts the correlation between F_{PED} and $F_{ROH>1}$ for the Ayrshire, Holstein and Jersey at an ROH length of 1000 kb.

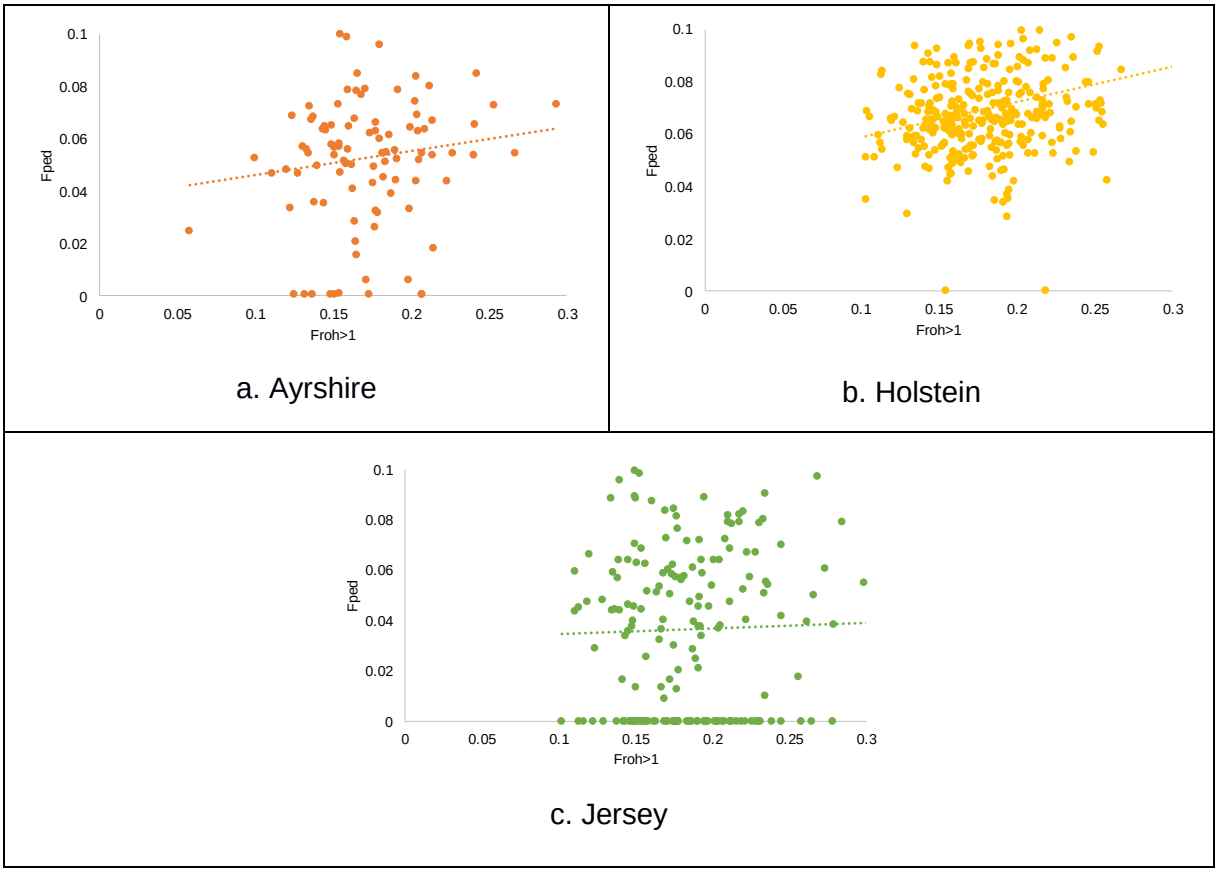

Figure 4.6 Scatterplots of F_{PED} vs F_{ROH} at an ROH length of 1000 kb for the populations

From the above figures the low correlation between F_{PED} and F_{ROH} compared to the high correlations between F_{IS} and F_{ROH} is clear. Correlation between F_{PED} and F_{ROH} range from -0.056 (F_{PED} vs $F_{ROH>8}$) for the Ayrshire to 0.186 (F_{PED} vs $F_{ROH>4}$) for the Holstein.

Figure 4.7 represents the effective population size (N_e) for the four respective dairy populations. SNeP was used to calculate the N_e for each of the four populations. N_e was plotted for all populations from 900 to approximately 13 generations ago.

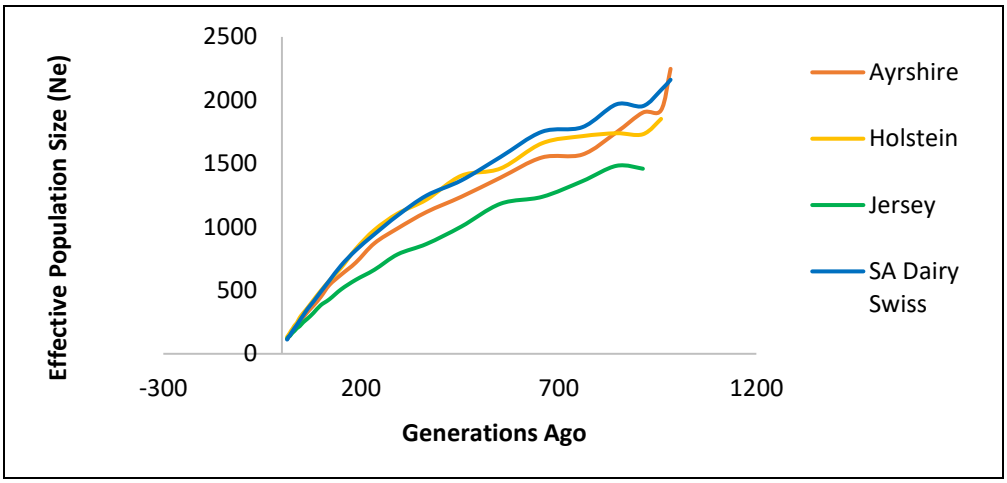

Figure 4.7 Trends in historic effective population size (Ne) for the four dairy populations

A general decrease in N_e can be seen for all four populations, indicating a loss of genetic diversity. The SA Dairy Swiss showed the largest decrease in Ne from 1903 animals 913 generations ago, to 112 animals 13 generations ago. Currently the SA Dairy Swiss has the smallest (112) effective population size of the four populations, followed by the Ayrshire with an N_e of 117.

4.2.4 Principal component analysis (PCA)

The genetic relatedness between the individual animals of the four different populations were investigated by principal component analysis (PCA). The first (PCA 1) and second (PCA 2) was plotted against each other in Figure 4.8a and the first (PAC1) and third (PCA3) principal components in Figure 4.8b.

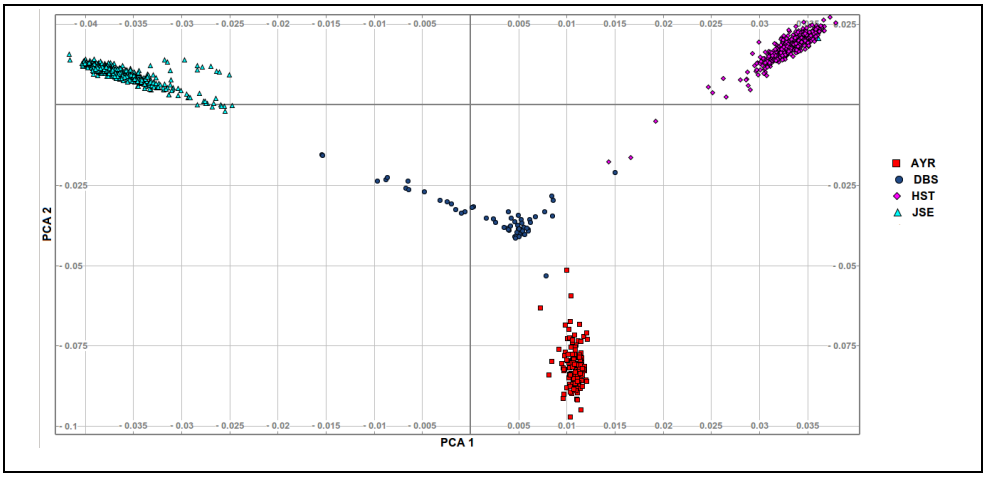

Figure 4.8a. The genetic relationships among the four populations as seen when plotting the first and second principal components (PCA1 and PCA2)

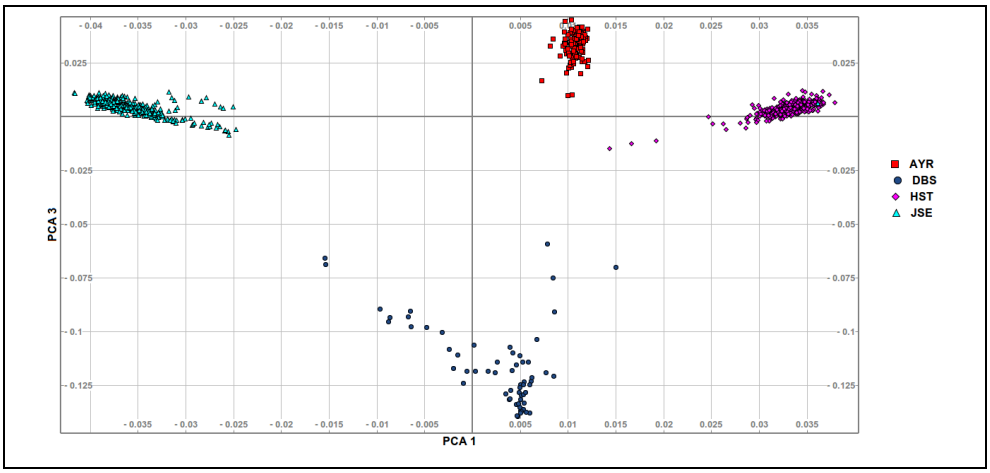

Figure 4.8b. The genetic relationships among the four populations as seen when plotting the first and third principal components (PCA1 and PCA3)

As can be seen in Figure 4.8a the individual animals from the four different populations clustered together within the respective populations. A few outliers were visible within all four populations with the SA Dairy Swiss being the most spread out of the four populations. The Ayrshire formed the tightest cluster of the four respective breeds. Figure 4.8b shows all the dairy breeds maintaining tight clusters, except for the SA Dairy Swiss, which was more diverse.

4.2.5 Population structure analysis

ADMIXTURE was used in order to investigate the population structure based on the shared ancestral SNP genotypes. Cross-validation scores for K-values of 1-5 were plotted in order to select the most appropriate K-value for population structure analysis.

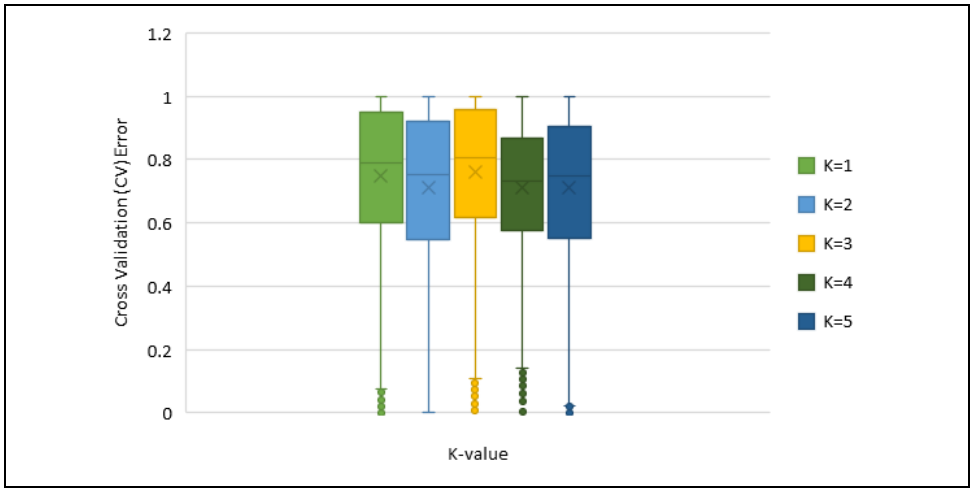

Figure 4.9. A cross validation plot, indicating the choice of the most appropriate K-value (after LDP)

From Figure 4.9 a K-value of 4 was chosen and used to construct a population structure plot, as can be seen in Figure 4.10. The K-value with the least variation around the mean CV-point was chosen.

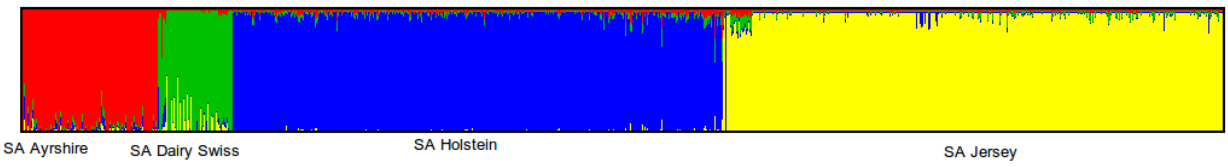

Figure 4.10. Population structure plot (K=4) of the populations

From Figure 4.10 it is clear that the four dairy breeds have their own distinct ancestral backgrounds, and this is in agreement with the results from the PCA plot in Figure 4.8a and b. The SA Dairy Swiss shows admixture with the Jersey.

Chapter 5

Discussion

Genomic inbreeding and effective population size (N_e) estimates have not previously been estimated for South African dairy cattle. These parameters reveal important information regarding the genetic diversity of breeds under selection. The current study was thus performed to estimate genomic inbreeding and effective population sizes for four South African dairy breeds as part of a project funded by the Dairy Genomics Program (DGP) with the aim of incorporating genomic information into the selection of South African dairy cattle.

5.1 Quality Control

Before the analysis of SNP data, quality control (QC) must be performed in order to remove potential errors that might have occurred during sampling and genotype calling to avoid biased results (Anderson *et al.*, 2010). It is important to note that the removal of markers impacts the final results of a study and it is thus important to ensure that only substandard markers are removed (Anderson *et al.*, 2010).

At a MAF of at least 0.02 in the current study the number of polymorphic SNPs ranged from 39 997 (Jersey) to 44 741 (Holstein). The number of polymorphic SNPs for the South African Holstein in the current study is higher than that reported by Qwabe *et al.* (2013). A similar trend was observed in the study done by Matukumalli *et al.* (2009) who observed 42 849 polymorphic SNPs in the European Holstein. The differences in the number of polymorphic SNPs between the current study and the study done by Qwabe *et al.* (2013) may be due to the different MAF thresholds used for defining polymorphic loci as well as a difference between the versions of the Bovine 50K SNP Beadchip.

Differences in the number of polymorphic loci between populations within this study, may be due to ascertainment bias resulting from the choice of SNPs during the development of the bovine 50K SNP chip (Nielsen, 2004; O'Brien *et al.*, 2014). The successful application of the Bovine 50K SNP chip depends mainly on their degree of polymorphisms in the different cattle breeds (Fan *et al.*, 2010). Ascertainment bias occurs when a relatively small sample was used for SNP discovery and the panel is not wholly representative of all the breeds (Nielsen, 2004; Matukumalli *et al.*, 2009). This results in a deficiency of SNP loci with rare alleles (Nielsen, 2004). During the SNP discovery for development of the bovine 50K SNP chip, different *Bos taurus*

breeds were included, including the Holstein, Angus, Charolais, Gelbvieh, Herefored, Limousin, Red Angus and Simmental (Van Tassell *et al.*, 2008). Although the Ayrshire, SA Dairy Swiss and Jersey breeds were not included in the initial SNP discovery, high utility was expected for these breeds as they are also classified as *Bos taurus* cattle and share similar characteristics with the discovery breeds. Thus, a large proportion of the polymorphic SNPs discovered during the initial SNP discovery is expected to exist in the Brown Swiss, Ayrshire and Jersey.

Heterozygosity rate is an important parameter to analyze during quality control as excessive heterozygous genotypes due to possible DNA sample contamination or a shortage of heterozygous genotypes due to possible inbreeding may lead to biased results (Anderson *et al.*, 2010). The average observed and expected heterozygosity rates across populations were 0.318 and 0.355, respectively. This corresponds to average heterozygosity estimates of 0.30 (Matukumalli *et al.*, 2009), 0.30 (Gautier *et al.*, 2010), 0.381 (Kelleher *et al.*, 2017) and 0.380 (Melka and Schenkel, 2012) across different cattle breeds.

The Holstein and Jersey had average H_E rates of 0.356 and 0.336 respectively. Both these heterozygosity values were lower compared to heterozygosity values of 0.372 and 0.390 obtained by Kelleher *et al.* (2017) for the European Holstein-Friesland and Jersey, respectively. Melka and Schenkel (2012) reported H_E values of 0.31 and 0.26 for the North American Holstein and Jersey, respectively. This corresponds to H_E values reported by Engelsma *et al.* (2012) on 90 Holstein heifers. The cattle were divided into two groups and their H_E estimated at high and low EBV. H_E for the high EBV group was lower (0.303) than that of the low EBV group (0.312) indicating a loss of genetic diversity in these cows. (Engelsma *et al.*, 2012). The H_E value reported by Makina *et al.* (2014) for the South African Holstein was 0.310, which corresponds to the current H_E estimates reported in the current study. Chagunda *et al.* (2018) also reported similar observed and expected heterozygosity values for the Rwanda Holstein and Jersey to the current study.

Melka and Schenkel (2012) estimated that the average H_O rates for the North American Holstein has been declining. They reported that H_O has declined from 0.361, 4 generations ago to 0.353 one generation ago. This is similar to the H_O of 0.359 for the current study. The H_E rates for the Ayrshire and SA Dairy Swiss was 0.342 and 0.338, respectively, which was higher than the H_E (0.27) reported by Melka and Schenkel (2012) for the Dairy Swiss. The higher H_O for the Holstein (0.359) compared to that of the other three breeds (Ayrshire, SA Dairy Swiss and Jersey) is consistent with the higher effective population size of the Holstein (Stachowicz *et al.*, 2011). This is consistent with the findings of Melka and Schenkel (2012) who also reported a higher

observed heterozygosity for the European Holstein than for the Brown Swiss and Jersey. The high H_E reported for the four populations indicates that these populations have a high degree of genetic variability, but the higher H_O than H_E reported for the four populations is a cause for concern as this indicates that inbreeding has occurred in these populations.

The average MAF for the current study ranged from 0.252 (Jersey) to 0.271 (Holstein). This is in agreement with the high degree of variation reported by Matukumalli *et al.* (2009) between breeds. Qwabe *et al.* (2013) reported MAF values of 0.22 for the South African Holstein during the evaluation of the bovine SNP50 bead chip for the use in South African cattle populations. This was lower than the average MAF value obtained in the current study and may be attributable to the small number (40) of samples included in the study done by Qwabe *et al.* (2013) compared to 412 samples in the current study. Another possible explanation for the difference in MAF values between the two studies may be due to the different minor allele frequency parameters (0.05 and 0.02, respectively) chosen during quality control. In a study done by Chagunda *et al.* (2018) on dairy cattle in smallholder farming operations, average MAF values of 0.29 and 0.23 are reported for the Rwanda Holstein and Jersey, respectively.

The higher MAF value for the Holstein could be indicative of a high level of polymorphism for this breed, while the lower MAF values obtained for the other three breeds could indicate that a higher proportion of alleles may be fixed within these breeds. This thus indicates a higher level of heterozygosity in the Holstein than the other three breeds.

The high linkage disequilibrium (r^2) values before LD pruning for the SA Dairy Swiss, Holstein and Jersey, could be as a result of strong selection pressure on these populations. When selection focuses on a specific trait of interest (milk yield) this part of the genome is preferentially kept in a population. The frequency of the favorable allele will increase, as well as that of the neutral loci that surround this region that are in LD with it. This will thus drive the frequency of the given haplotypes in the region towards fixation (Biswas and Akey, 2006; Stephan *et al.*, 2006). The r^2 value reported for the Holstein in the current study (0.311) is higher than that reported for the Nordic (0.18) (Su *et al.*, 2012) and German Holstein (0.29) (Qanbari *et al.*, 2010), but is much lower than r^2 value of > 0.80 reported by Kim and Kirkpatrick (2009) for the North American Holstein. It should be noted that different sample sizes, LD measures, marker densities, as well as recent and historical population demographics makes it difficult to compare LD levels between different studies (Pritchard and Przeworski, 2001).

Cattle populations have longer LD regions compared to humans due to the recent, strong selection on these breeds (Bovine HapMap Consortium, 2009). The fact that more SNPs were removed for the Holstein and Jersey than for the Ayrshire following LD pruning, may imply that the other three breeds' genomes contained larger regions of high LD, which indicates a higher selection pressure on these breeds than for the Ayrshire. In the case of the SA Dairy Swiss a larger number of SNPs may have been removed due to the small number of animals used in the current study.

5.2 Inbreeding and effective population size

Pedigree based inbreeding has been used in animal breeding to estimate inbreeding coefficients for more than 50 years. One disadvantage of pedigree-based inbreeding estimation is the lack of pedigree depth and the misidentification of pedigrees which may result in unreliable inbreeding estimates (Cassell *et al.*, 2003; Bjelland *et al.*, 2013; Pryce *et al.*, 2014). A second disadvantage of F_{PED} is that the proportion of an individual's genome that is IBD is estimated relative to that of a poorly characterized founder generation (McQuillan *et al.*, 2008; Ferenčaković *et al.*, 2011; Fernández and Bennewitz, 2017). The reference population is used as the founder generation and founders or individuals not represented in the pedigree, are assumed to be unrelated (Curik *et al.*, 2014; Fernández and Bennewitz, 2017). This is inaccurate as individuals in historical populations were often related (McQuillan *et al.*, 2008). With the utilization of ROH as a way to examine population history in humans (McQuillan *et al.*, 2008; Kirin *et al.*, 2010), ROH has become a common method of estimating genomic inbreeding in animal breeding. In this study the average inbreeding coefficients (F_{IS}) for all four populations were negative. This could be indicative of effective on farm management against inbreeding as well as the use of both local and international bulls. The most inbred individual was in the Holstein population with an inbreeding coefficient of 0.164. This is followed by the Jersey (0.128), SA Dairy Swiss (0.110) and Ayrshire (0.109). These values are lower than the pedigree inbreeding coefficients (F_{PED}) estimated by SA Stud Book. Lower F_{PED} values compared to F_{IS} could be explained by the fact that F_{IS} measures the actual allele sharing as opposed to F_{PED} that estimates fractions of the allele that are expected to be identical by descent (Van Raden *et al.*, 2011; Pryce *et al.*, 2014).

It is crucial to set appropriate parameters to detect an ROH as the minimum number of SNPs that is used to define an ROH should be done according to the available SNP density. The density of the SNP panel to detect ROH is an important factor which strongly affects the autozygosity estimates (Bjelland *et al.*, 2013; Ferenčaković *et al.*, 2013b; Signer-Hasler *et al.*,

2017). In the current study the number of ROH in the different length categories were distributed approximately evenly across the genome. This was supported by the study done by Signer-Hasler *et al.* (2017) on nine Swiss dairy cattle, which reported that most ROH segments were observed in the length category of five to ten Mb. This was in contrast with other studies who found that longer ROH segments were found far less frequent than shorter segments in human (Kirin *et al.*, 2010) ROH studies and in cattle populations (Ferencăković *et al.*, 2013a). Mastrangelo *et al.* (2016) also reported higher frequencies of ROH at longer length categories than shorter lengths. Signer-Hasler *et al.* (2017) reported that the minimum number of SNPs used to define a ROH will also have an influence on the number of ROH segments found in the different length categories.

The higher number of longer ROH segments in the current study could be due to the increased level of recent inbreeding (Signer-Hasler *et al.*, 2017) in South African dairy cattle. The high number of long ROH segments may also be due to the fact that one heterozygous SNP were allowed within the ROH. Ferencăković *et al.* (2013b) suggested different levels of heterozygous calls for the different lengths of ROH. According to a study done by Mastrangelo *et al.* (2016) on Italian dairy cattle, allowing one, two and three heterozygous genotypes showed no significant effects on the levels of inbreeding. Due to genotyping errors in SNP chip data, it may be reasonable to allow some heterozygous genotypes. This may be especially so for long segments of ROH that are more frequent in cattle populations (Ferencăković *et al.*, 2013b) than they are in humans (Kirin *et al.*, 2010). For longer ROH segments (> 5000 to 6000 SNPs) it is acceptable to allow some heterozygous genotypes as long as there is a limit on the number allowed (Ferencăković *et al.*, 2013b). Ferencăković *et al.* (2013b) also showed that inaccurate ROH calls may be defined when allowing certain minimum numbers of heterozygous genotypes, especially at the end of an ROH segment. In a study done by Marras *et al.* (2015), the number of longer ROH segments increased dramatically when heterozygous genotypes were allowed within the ROH.

F_{ROH} estimates for the four South African dairy populations indicate an increase in inbreeding across the last ten generations. The Jersey had the lowest level of genomic inbreeding (0.056) at an ROH of 1 Mb, while the SA Dairy Swiss had the highest (0.073). At an ROH length of 16 Mb the Ayrshire had the lowest (0.227) genomic inbreeding coefficient while the Holstein had the highest (0.252). Mastrangelo *et al.* (2016) reported that an ROH segments of ~ 10 Mb is indicative of inbreeding up to five generations ago. Ferencăković *et al.* (2013a) reported F_{ROH} values of 0.039 at an ROH length of 16 Mb this is much lower than the F_{ROH} values estimated in the current study. F_{ROH} estimates for the Dairy Swiss in the current at an ROH length of 4 Mb

correlated with the F_{ROH} (0.103) value reported by Ferencăković *et al.* (2013a) but was higher (0.097) than that reported by Marras *et al.* (2015) for the Italian Brown Swiss.

Two different estimates of genomic inbreeding were calculated for each of the four populations included in the study, namely F_{IS} and F_{ROH}. The F_{PED} inbreeding coefficients were lower than the F_{ROH} inbreeding estimates. This difference may be due to F_{PED} only capturing inbreeding based on recorded pedigree that may only extend few generations back, whereas ROH is able to capture both ancient and recent inbreeding. In the current study the strongest correlation between the two inbreeding estimates were found between F_{IS} and $F_{ROH>1}$. A moderate to high correlation was found between F_{IS} and F_{ROH} up to 8 000 kb ranging from 0.407 to 0.998. Several other studies done on dairy cattle also reported moderate to high correlations between F_{GRM} and F_{ROH} (Marras *et al.*, 2015; Mastrangelo *et al.*, 2016). Bjelland *et al.* (2013) reported moderate to high correlations (0.810) in the US Holstein, and a moderate correlation of 0.620 was reported by Pryce *et al.* (2014) for the Australian Holstein and Jersey. Lower correlations were observed between F_{IS} and $F_{ROH>16}$ ranging from 0.262 to 0.377. These moderate to high correlations between F_{IS} and F_{GRM} suggests that ROH is an accurate estimator of the IBD genomic proportion.

Very low correlations between F_{PED} and F_{ROH} were observed across the four populations in the current study. The low correlations between F_{PED} and F_{ROH} for the Ayrshire may indicate incomplete pedigree data for this population. This is supported by Peripolli *et al.* (2018) who also reported low correlations between F_{PED} and F_{ROH} relating to shallow pedigree depth. The higher correlation between F_{PED} and F_{ROH} found in the Holstein and the Jersey may suggest that the correlation between these parameters are dependent on pedigree depth. The correlation between these two parameters has been shown to increase with an increase in pedigree depth (Ferencăković *et al.*, 2011; Purfield *et al.*, 2012; Marras *et al.*, 2015). Ferencăković *et al.* (2011) reported correlations between F_{PED} and F_{ROH} ranging from 0.61 to 0.67. This was in agreement with Purfield *et al.* (2012) and Gurgul *et al.* (2016), both reported higher correlations between F_{PED} and F_{ROH} with an increase in pedigree depth. Low pedigree completeness is commonly accepted as one of the limitations in F_{PED} calculations (Cassell *et al.*, 2003; Bjelland *et al.*, 2013; Pryce *et al.*, 2014). F_{PED} was not able to recover ancient relatedness indicating that F_{PED} is not an accurate measure of inbreeding. Low to moderate correlations between F_{PED} and F_{IS} in the current study correlate with results obtained by Pryce *et al.* (2014) (Australian Holstein and Jersey); Marras *et al.* (2015) (Italian Brown Swiss and Holstein) and Zhang *et al.* (2015a) (Danish Holstein and Jersey).

The F_{ROH} value for the Holstein breed in the current study at an ROH length of 4 Mb was 0.097, which is higher than the F_{ROH} value (0.049) reported by Mastrangelo *et al.* (2016) for the Italian Holstein at the same length. Marras *et al.* (2015) reported an F_{ROH} value of 0.073 for the Italian Holstein which is higher than that reported by Mastrangelo *et al.* (2016) but lower than the F_{ROH} estimated in the current study. This could be due to the fact that no heterozygous SNPs was allowed in the study done by Marras *et al.* (2015). F_{ROH} values reported by Signer-Hasler *et al.* (2017) for the Swiss, Brown Swiss and Holstein, was lower than estimates reported in the current study and the study done by Marras *et al.* (2015). Ferencăković *et al.* (2013a) reported that the 50K SNP panel may overestimate the number of ROH segments between 1 to 4 Mb long and reveals an abundance of small segments, which suggests that it is not sensitive enough for the accurate estimation of small segments. Criteria used to detect ROH differs between studies especially for the minimum length used to define an ROH and the minimum number of SNPs allowed, thus making it difficult to compare results from different studies (Mastrangelo *et al.*, 2016; Signer-Hasler *et al.*, 2017).

F_{ROH} has become the preferred method of estimating inbreeding in dairy cattle as it is able to accurately predict the amount of autozygosity within the genome as well as being able to give information on past inbreeding (Howrigan *et al.*, 2011; Keller *et al.*, 2011). F_{ROH} also has the added advantage of being able to estimate autozygosity in animals without any pedigree information (Keller *et al.*, 2011).

In the current study a decrease in N_e is visible, with the SA Dairy Swiss showing the largest decrease in N_e to an N_e of 112 animals approximately 15 generations ago. The SA Dairy Swiss was also the population with the lowest current effective population size of the four breeds included in the study. This was confirmed in a study done by Marras *et al.* (2015) who reported an effective population size of 90.7 for the Italian Brown Swiss up to five generations ago, which is significantly lower than the 237.6 animals, fifty generations ago. The authors also reported N_e values for the Italian Holstein of 98.7 and 284.3 animals five and fifty generations ago, respectively. These N_e estimates are consistent with the N_e values reported for the South African Holstein in the current study at 45 generations ago, but significantly lower than N_e estimates for the SA Dairy Swiss in the same time period. The effective population size (N_e) is an important parameter that can be used to assess inbreeding rates and thus genetic diversity within populations (Groeneveld *et al.*, 2010). The census size of a population is not a true reflection of its N_e, thus it is an important parameter to measure in order to ensure enough genetic diversity

exists within a breed in order to improve the breed. The N_e sizes for all four breeds included in this study have decreased. This decrease in N_e is accompanied with a loss of genetic diversity.

The effective population sizes for the South African dairy cattle populations included in this study were all lower than pedigree-based N_e reported by Maiwashe *et al.* (2006) and De Ponte-Bouwer *et al.* (2013) for the same breeds, with the exception of the Jersey, who had a slightly higher N_e of 120 in the current study compared to 108 by Maiwashe *et al.* (2006). The different N_e estimates reported by Maiwashe *et al.* (2006) could be due to the fact that these authors estimated N_e based on parentage information. The N_e for South African Holstein is higher than for their global counterparts. Rodriguez-Ramilo *et al.* (2015) reported an effective population size of 101 for the Spanish Holstein, while an N_e of 114 was reported by Stachowicz *et al.* (2011) for the Canadian Holstein.

The strong selection pressure practiced on dairy breeds globally, over the past few decades, has resulted in high rates of genetic gain along with increases in inbreeding (Rodríguez-Ramilo *et al.*, 2015) and thus decreases in the effective population size.

5.2.1 Genetic relatedness and population structure

Principal component analysis (PCA) is the most common method to distinguish between different ancestries through the identification of genetically related samples (Anderson *et al.*, 2010) and has been used in various studies to investigate genetic diversity in cattle (Gautier *et al.*, 2010; Lewis *et al.*, 2011; Harris *et al.*, 2014; Edea *et al.* 2015; Kelleher *et al.*, 2017). Both the PCA and ADMIXTURE results in the current study were in agreement and separated the four dairy breeds into four non-overlapping clusters. This could be an indication of unique genes shared within each breed.

A study done by Blott *et al.* (1998) observed that the European Ayrshire and Friesland breeds were grouped into the same cluster, which could indicate that these two breeds share a common ancestor. In the current study the close clustering between these breeds were not observed, in fact the Ayrshire breed formed the tightest within-breed cluster of the four breeds.

The distinct separation between the Jersey and the other three breeds included in this study when plotting the first and second, and the first and third principal components against each other, supports the isolated development of this breed on the Jersey Island in the English Channel off the coast of France (Nel, 1968; MacHugh *et al.*, 1997). During the isolation of the Jersey cattle,

strict rules on importing and breeding practices were implemented in order to keep the breed pure (MacHugh *et al.*, 1997). This was confirmed by Kelleher *et al.* (2017) who found a distinct separation between the breeds originating from the British Isle from the rest of the breeds included in the study. Kantanen *et al.* (2000) also found that the Danish Jersey had no close genetic relationship with other breeds.

As seen with the plotting of the first and second principal components against each other, there was a slight overlap between the SA Dairy Swiss and the Holstein breed, which was removed when plotting the first and third principal components. This could be due to the fact that some breeders have in the past cross bred SA Dairy Swiss cattle with Holstein cattle in order to assess the economic value of cross-breeding with the Holstein (SA Dairy Swiss Journal). The SA Dairy Swiss was the most spread out when plotting both PCA1 and PCA2 against each other, as well as at PCA1 and PCA3 which could be indicative of cross-breeding between the SA Dairy Swiss and other breeds. From the PCA and ADMIXTURE results it is clear that the four populations included in the current study had distinct ancestries which could mainly be attributed to the diverse origin and development of these breeds.

Chapter 6

Conclusion and recommendations

6.1 Conclusion

In this study 1002 dairy cattle, representing four South African dairy breeds, were genotyped with the Infinium Bovine SNP 50-24 V3.0 Beadchip. The genotypes used in the current study originated from the Dairy Genomics Program (DGP). The DGP is a three-year project that is aimed at integrating genomic information into the selection of South African dairy cattle. Funding for the project was received from the DGP as well as the Technology Innovation Agency (TIA) through the DGP.

The higher observed than expected heterozygosity rates for the four dairy populations were comparable with heterozygosity values obtained for the same breeds in other studies. The number of ROH segments found at each length category was higher at a ROH length of 16 Mb than an ROH length of 4 Mb. $F_{ROH.> 16\ Mb}$ estimates were higher than those reported in other studies, but indicate that recent inbreeding has occurred in the four South African dairy populations. This implies that breeders should consider the risk of inbreeding when selecting for improved genetic merit. Individual inbreeding coefficients indicate that breeders are already applying strategies to reduce inbreeding in these breeds. $F_{ROH > 4\ Mb}$ values were more closely related with F_{PED} values than $F_{ROH > 16\ Mb}$, which is in contrast to other studies that found higher correlations between longer F_{ROH} segments than shorter segments. The effective population size for all four populations has decreased over the last few generations, which is in agreement with other studies. A PCA analysis showed that the four populations formed distinct clusters from one another. PCA results were consistent with the history of each of the four populations.

The results obtained in the study indicated that F_{ROH} is the most accurate measure of inbreeding when compared to F_{PED} and F_{IS}. The general increase in inbreeding and the subsequent decrease in effective population size emphasizes the importance of implementing effective breeding programs in order to maintain genetic diversity within populations. Genomic inbreeding derived from ROH is a useful method of estimating inbreeding. ROH is also able to give insight on pedigree-based inbreeding when pedigree data is incomplete or unavailable. Analyses with ROH at different length categories allows for the estimation of past versus recent inbreeding.

6.2 Recommendations

From the above study it is clear that inbreeding in the four South African dairy breeds have been increasing in the past few years. Breeders should use commercial mating programs that include all available generations' pedigree information from both the sire and dam-lines' side to effectively manage and avoid inbreeding as far as possible. ROH as a measure of inbreeding and genetic diversity should be incorporated into these methods in order to minimize inbreeding. According to the knowledge of the author no other genomic studies to estimate the inbreeding and effective population size has been performed on South African dairy cattle. Further studies have to be done in order to identify the optimal parameters for detecting a ROH in a wide range of livestock species.

References

Agerholm, J.S., 2007. Inherited disorders in Danish cattle. APMIS. 115, Suppl s122, 1-76.

Agerholm, J.S., Bendixen, C., Andersen, O. & Arnbjerg, J., 2001. Complex vertebral malformation in Holstein calves. *J. of Vet. Diagn. Invest.* 13, 4, 283-289.

Akey, J.M., Zhang, G., Zhang, K., Jin, L. & Shriver, M.D., 2002. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* 12, 12, 1805-1814.

Alexander, D.H., Novembre, J. & Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655-1664.

Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P. & Zondervan, K.T., 2010. Data quality control in genetic case-control association studies. *Nature Protocols* 5, 9, 1564-1573.

Banga, C.B., Neser, F.W.C. & Garrick, D.J., 2014. Breeding objectives for Holstein cattle in South Africa. *S. Afr J. Anim. Sci.* 44, 3, 199-214.

Barbato, M., Orozco-terWengel, P., Tapio, M. & Bruford, M.W., 2015. SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Front. Genet.* 6.

Bijma, P. & Woolliams, J.A., 2002. Design of sustainable breeding programs in developed countries. In Proceedings of the 7th World Congress on Genetics Applied to Livestock Production. Montpellier, 19-23 August. Vol. 30. pp. 133-133.

Biswas, S. & Akey, J.M., 2006. Genomic insights into positive selection. *Trends Genet. 22*, 8, 437-446.

Bjelland, D.W., Weigel, K.A., Vukasinovic, N. & Nkrumah, J.D., 2013. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *J. Dairy Sci.* 96, 7, 4697-4706.

Blott, S.C., Williams, J.L. & Haley, C.S., 1998. Genetic relationships among European cattle breeds. *Anim. Genet.* 29, 4, 273-282.

Bourdon, R.M., 2014. Mating Strategies Based on Pedigree Relationship: Inbreeding and Outbreeding. In: Understanding Animal Breeding (2nd ed.). Publisher, Pearson. pp. 345, 346, 347.

Bovine HapMap Consortium, 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science.* 324, 5926, 528-532.

Brotherstone, S. & Goddard, M., 2005. Artificial selection and maintenance of genetic variance in the global dairy cow population. In: Philosophical Transactions of the Royal Society of London B. *Biolog. Sci.* 360, 1459, 1479-1488.

Buchmann, R. & Hazelhurst, S., 2014. Genesis Manual, University of the Witwatersrand, Johannesburg. http://www.bioinf.wits.ac.za/software/genesis/Genesis.pdf

Cardellino, R.A. & Boyazoglu J., 2009. Research opportunities in the field of animal genetic resources. *Livest. Sci.* 120, 166-173.

Carothers, A.D., Rudan, I., Kolcic, I., Polasek, O., Hayward, C., Wright, A.F., Campbell, H., Teague, P., Hastie, N.D. & Weber, J.L., 2006. Estimating human inbreeding coefficients: comparison of genealogical and marker heterozygosity approaches. *Ann. Hum. Genet.* 70, 5, 666-676.

Cassell, B.G., Adamec, V. & Pearson, R.E., 2003. Effect of incomplete pedigrees on estimates of inbreeding and inbreeding depression for days to first service and summit milk yield in Holsteins and Jerseys. *J. Dairy Sci.* 86, 9, 2967-2976.

Chagunda, M., Dusingizimana, T., Kamana, O., Cheruiyot, E., Mujibi, F.D. & Mwai, O., 2018. Use of high density single nucleotide polymorphism (SNP) arrays to assess genetic diversity and population structure of dairy cattle in smallholder dairy systems: the case of Girinka Programme in Rwanda. *FR Genet.* 9, p.438.

Charlesworth, D. & Charlesworth, B., 1987. Inbreeding depression and its evolutionary consequences. *Annu. Rev. Ecol. Syst.* 18, 1, 237-268.

Clark, S.A., Kinghorn, B.P., Hickey, J.M. & van der Werf, J.H., 2013. The effect of genomic information on optimal contribution selection in livestock breeding programs. *Genet. Sel. Evol.* 45, 1, p.44.

Cohen-Zinder, M., Seroussi, E., Larkin, D.M., Loor, J.J., Everts-van der Wind, A., Lee, J.H., Drackley, J.K., Band, M.R., Hernandez, A.G., Shani, M. & Lewin, H.A., 2005. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Res.* 15, 7, 936-944.

Cole, J. B., Lewis, R. M., Maltecca, C., Newman, S., Olson, K. M. & Tait Jr., R. G., 2013. Identifying relationships among markers, genes, and phenotypes. *J. Anim. Sci.* 91, 521-522.

Croquet, C., Mayeres, P., Gillon, A., Vanderick, S. & Gengler, N., 2006. Inbreeding depression for global and partial economic indexes, production, type, and functional traits. *J. Dairy Sci.* 89, 6, 2257-2267.

Croquet, C., Mayeres, P., Gillon, A., Hammami, H., Soyeurt, H., Vanderick, S. & Gengler, N., 2007. Linear and curvilinear effects of inbreeding on production traits for Walloon Holstein cows. *J. Dairy Sci.* 90, 1, 465-471.

Crow, J.F. and Kimura, M., 1970. An introduction to population genetics theory. Harper and Row, New York.

Curik, I., Ferenčaković, M. & Sölkner, J., 2014. Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livest. Sci.* 166, pp.26-34.

DAFF, 2017. A Profile of the South African Dairy Market Value Chain. Department of Agriculture Forestry and Fisheries, Pretoria, South Africa.

DAFF, 2018. Abstract of Agricultural Statistics. Department of Agriculture, Forestry and Fisheries, Pretoria, South Africa.

Davis, N.A., Pandey, A. & McKinney, B.A., 2011. Real-world comparison of CPU and GPU implementations of SNPrank: a network analysis tool for GWAS. Bioinformatics 27, 284-285.

De Cara, M.A.R., Fernández, J., Toro, M.A. & Villanueva, B., 2011. Using genome-wide information to minimize the loss of diversity in conservation programmes. *J. Anim. Breed. Genet.* 128, 6, 456-464.

De Ponte-Bouwer, P., Visser, C. & Mostert, B.E., 2013. Analysis of inbreeding of the South African Dairy Swiss breed. *S. Afr. J. Anim. Sci.* 43, 1, 38-43.

De Roos, A.P.W., Schrooten, C., Veerkamp, R.F. & Van Arendonk, J.A.M., 2011. Effects of genomic selection on genetic improvement, inbreeding, and merit of young versus proven bulls. *J. Dairy Sci.* 94, 3, 1559-1567.

Diamond, J., 2002. Evolution, consequences and future of plant and animal domestication. *Nature.* 418, 6898, 700.

Du Toit, J., Van Wyk, J.B. & Maiwashe, A., 2012. Assessment of inbreeding depression for functional herd life in the South African Jersey breed based on level and rate of inbreeding. *S. Afr. J. of Anim. Sci.* 42, 1, 55-62.

Duvenhage, H., 2017. How well do you know your Holstein history? In: The Dairy Mail (February). Volume 24, No 2. http://www.agriconnect.co.za/tdm-digital-magazines/ accessed 8 July 2018.

Edea, Z., Dadi, H., Kim, S.W., Dessie, T., Lee, T., Kim, H., Kim, J.J. & Kim, K.S., 2013. Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. *Front. Genet.* 4, p.35.

Engelsma, K.A., Veerkamp, R.F., Calus, M.P.L., Bijma, P. & Windig, J.J., 2012. Pedigree-and marker-based methods in the estimation of genetic diversity in small groups of Holstein cattle. *J. of Anim. Breeding Genet.* 129, 3, 195-205.

Erhardt, G. & Weimann, C., 2007. Use of molecular markers for evaluation of genetic diversity and in animal production. *Arch. Latinoam. Prod. Anim.* 15, 63-66.

Excoffier, L. & Slatkin, M., 1995. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol. Biol. Evol. 12, 5, 921-927.

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

Fan, B., Du, Z.Q., Gorbach, D.M. & Rothschild, M.F., 2010. Development and application of high-density SNP arrays in genomic studies of domestic animals. Asian Austral. *J. Anim. Sci.* 23, 7, 833-847.

FAO, 2007. The State of the World's Animal Genetic Resources for Food and Agriculture, Rome, Italy.

FAO, 2010. http://www.fao.org/unfao/procurement/statistics-from-2010-2015/statistics-2010/en/ accessed 8 July 2018.

FAO, 2017. Dairy and Dairy Products. In: OECD-FAO Agricultural Outlook 2017-2026, OECD Publishing, Paris.

Ferenčaković, M., Hamzić, E., Gredler, B., Curik, I. & Sölkner, J., 2011. Runs of homozygosity reveal genome-wide autozygosity in the Austrian Fleckvieh cattle. *Agriculturae Conspectus Scientificus* (ACS). 76, 4, 325-329.

Ferenčaković, M., Hamzić, E., Gredler, B., Solberg, T.R., Klemetsdal, G., Curik, I. & Sölkner, J., 2013a. Estimates of autozygosity derived from runs of homozygosity: Empirical evidence from selected cattle populations. *J Anim. Breed. Genet.* 130, 286–293.

Ferenčaković, M., Sölkner, J., & Curik, I., 2013b. Estimating autozygosity from high-throughput information: Effect of SNP density and genotyping errors. *Genet. Sel. Evol.* 45:42.

Fernández, J. & Bennewitz, J. ed., 2017. Chapter 2 “Defining genetic diversity based on genomic tools”. In: Genomic management of animal genetic diversity. Publisher, Wageningen Academic. pp. 55-62.

Fernández, J., Villanueva, B., Pong-Wong, R. & Toro, M.Á., 2005. Efficiency of the use of pedigree and molecular marker information in conservation programs. *Genet.* 170, 3, 1313-1321.

Fisher, R.A., 1954. A fuller theory of “junctions” in inbreeding. *Heredity.* 8, 2, p.187.

Fleming, A., Abdalla, E.A., Maltecca, C. & Baes, C.F., 2018. Invited review: Reproductive and genomic technologies to optimize breeding strategies for genetic progress in dairy cattle. *Arch. Anim. Breed.* 61, 1, 43-57.

Flori, L., Fritz, S., Jaffrézic, F., Boussaha, M., Gut, I., Heath, S., Foulley, J.L. & Gautier, M., 2009. The genome response to artificial selection: a case study in dairy cattle. *Plos One.* 4, 8, pp. e6595.

Frankham R., Ballou J.D. & Briscoe D.A., 2002. Introduction to Conservation Biology. Camb. Univ. Press.

Gautier, M. & Naves, M., 2011. Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Mol. Ecol.* 20, 15, 3128-3143.

Gautier, M., Laloë, D. & Moazami-Goudarzi, K., 2010. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PloS One.* 5, 9, pp.e13038.

Goddard, M.E., Hayes, B.J. and Meuwissen, T. H. E., 2011. Using the genomic relationship matrix to predict the accuracy of genomic selection. *J. Anim. Breed. Genet.* 128, 6, 409-421.

Gómez-Romano, F., Villanueva, B., De Cara, M.Á.R. & Fernández, J., 2013. Maintaining genetic diversity using molecular coancestry: The effect of marker density and effective population size. *Genet. Sel. Evol.* 45, 1, 38.

Gómez-Romano, F., Villanueva, B., Sölkner, J., Cara, M.A.R., Mészáros, G., Pérez O'Brien, A.M. & Fernández, J., 2016. The use of coancestry based on shared segments for maintaining genetic diversity. *J. Anim. Breed. Genet.* 133, 5, 357-365.

González-Recio, O., De Maturana, E.L. & Gutiérrez, J.P., 2007. Inbreeding depression on female fertility and calving ease in Spanish dairy cattle. *J. Dairy Sci.* 90, 12, 5744-5752.

Groeneveld, L.F., Lenstra, J.A., Eding, H., Toro, M.A., Scherf, B., Pilling, D., Negrini, R., Finlay, E.K., Jianlin, H., Groeneveld, E. & Weigend, S., 2010. Genetic diversity in farm animals – a review. *Anim. Genet.* 41, pp. 6-31.

Gurgul, A., Szmatoła, T., Topolski, P., Jasielczuk, I., Żukowski, K. & Bugno-Poniewierska, M., 2016. The use of runs of homozygosity for estimation of recent inbreeding in Holstein cattle. *J. Appl. Genet.* 57, 4, 527-530.

Haile-Mariam, M., Morton, J.M. & Goddard, M.E., 2003. Estimates of genetic parameters for fertility traits of Australian Holstein-Friesian cattle. *Anim. Sci.* 76, 35-42.

Harris, B.L., Winkelman, A.M. & Johnson, D.E., 2014. Across-breed genomic prediction in dairy cattle. In: Proceedings of the 10th World Congress of Genetics Applied to Livestock Production. Livest. Impr. Corp., 17-22, Hamilton, New Zealand.

Hayes, B., & Goddard, M., 2010. Genome-wide association and genomic selection in animal breeding. *Genome*. 53, 11, 876-883.

Hayes, B.J., Bowman, P.J., Chamberlain, A.J. & Goddard, M.E., 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92, 2, 433-443.

Hemati, B., Gharaie-Fathabad, S., Fazeli, M.H., Namvar, Z. & Ranji, M., 2015. Investigation of bovine leukocyte adhesion deficiency (BLAD) and complex vertebral malformation (CVM) in a population of Iranian Holstein cows. Iran. *J. Appl. Anim. Sci.* 5, 1, 69-72.

Henderson, C.R., 1973. Sire evaluation and genetic trends. *J Anim. Sci.* pp.10-41.

Hoffmann, I., 2010. Climate change and characterization, breeding and conservation of animal genetic resources. *Anim. Genet.* 41, 32-46.

Holt, M., Meuwissen, T. & Vangen, O., 2005. The effect of fast created inbreeding on litter size and body weights in mice. *Gen. Sel. Evol.* 37, 5, 523-537.

Howard, J.T., Pryce, J.E., Baes, C. & Maltecca, C., 2017. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. *J. Dairy Sci.* 100, 8, 6009-6024.

Howrigan, D.P., Simonson, M.A. & Keller, M.C., 2011. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *BMC Genomics* 12, 1, 460.

Illumina Inc., 2012. BovineHD genotyping beadchip: more than 777 000 SNPs that deliver the densist coverage available for the bovine genome. Available at URL: www.illumina.com/Documents/products/datasheets/datasheet_bovineHD.pdf.

International Farm Comparison Network, Dairy Report, 2017. https://ifcndairy.org/wp-content/uploads/2017/05/Dairy_Report_2017_extraction.pdf accessed 8 July 2018.

Jain, S.K. & Allard, R.W., 1966. The effects of linkage, epistasis, and inbreeding on population changes under selection. *Genet.* 53, 4, p.633.

Jiménez-Mena, B. & Bataillon, T., 2016. Heterogeneity in effective population size and its implications in conservation genetics and animal breeding. *Conserv. Genet.* Res. 8, 1, 35-41.

Kaminski, S., Grzybowski, G., Prusak, B. & Rusc, A., 2005. No incidence of DUMPS carriers in Polish dairy cattle. *J. Appl. Genet.* 46, 4, 395-397.

Kantanen, J., Olsaker, I., Holm, L.E., Lien, S., Vilkki, J., Brusgaard, K., Eythorsdottir, E., Danell, B. & Adalsteinsson, S., 2000. Genetic diversity and population structure of 20 North European cattle breeds. *J. Hered.* 91, 6, 446-457.

Kearney, J.F., Wall, E., Villanueva, B. & Coffey, M.P., 2004. Inbreeding trends and application of optimized selection in the UK Holstein population. *J. Dairy Sci.* 87, 10, 3503-3509.

Kehrli Jr, M.E., Ackermann, M.R., Shuster, D.E., Van Der Maaten, M.J., Schmalstieg, F.C., Anderson, D.C. & Hughes, B.J., 1992. Bovine leukocyte adhesion deficiency. Beta 2 integrin deficiency in young Holstein cattle. AM *J. Pathol.* 140, 6, 1489.

Keller, M.C., Visscher, P.M. & Goddard, M.E., 2011. Quantification of inbreeding due to distant ancestors and its detection using dense SNP data. *Genetics.* pp.111.

Kelleher, M.M., Berry, D.P., Kearney, J.F., McParland, S., Buckley, F. & Purfield, D.C., 2017. Inference of population structure of purebred dairy and beef cattle using high-density genotype data. *Anim.* 11, 1, 15-23.

Khatkar, M.S., Nicholas, F.W., Collins, A.R., Zenger, K.R., Cavanagh, J.A., Barris, W., Schnabel, R.D., Taylor, J.F. & Raadsma, H.W., 2008. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics.* 9, 1, 187.

Kim, E.S. & Kirkpatrick, B.W., 2009. Linkage disequilibrium in the North American Holstein population. *Animal Genetics*. 40, 3, 279-288.

Kim, E.S., Sonstegard, T.S., Van Tassell, C.P., Wiggans, G. & Rothschild, M.F., 2015. The relationship between runs of homozygosity and inbreeding in Jersey cattle under selection. *PLoS One* 10, 7, pp. e0129967.

Kirin, M., McQuillan, R., Franklin, C.S., Campbell, H., McKeigue, P.M. & Wilson, J.F., 2010. Genomic runs of homozygosity record population history and consanguinity. *PloS One.* 5,11, p. e13996.

Kristensen, T.N. & Sørensen, A.C., 2005. Inbreeding–lessons from animal breeding, evolutionary biology and conservation genetics. *Anim. Sci.* 80, 2, 121-133.

Lacto Data Vol21 No1 May 2018. https://milksa.co.za/sites/default/files/LAC005%20May%202018.pdf accessed 3 July 2018.

Lenstra, J.A., Groeneveld, L.F., Eding, H., Kantanen, J., Williams, J.L., Taberlet, P., Nicolazzi, E.L., Sölkner, J., Simianer, H., Ciani, E. & Garcia, J.F., 2012. Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. *Anim. Genet.* 43, 5, 483-502.

Leroy, G., Mary-Huard, T., Verrier, E., Danvy, S., Charvolin, E. & Danchin-Burge, C., 2013. Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle and horse. *Genet. Sel. Evol.* 45, 1, 1.

Leutenegger, A.L., Prum, B., Génin, E., Verny, C., Lemainque, A., Clerget-Darpoux, F. & Thompson, E.A., 2003. Estimation of the inbreeding coefficient through use of genomic data. *AM J. Hum. Genet.* 73, 3, 516-523.

Lewis, J., Abas, Z., Dadousis, C., Lykidis, D., Paschou, P. & Drineas, P., 2011. Tracing cattle breeds with principal components analysis ancestry informative SNPs. *PloS One.* 6, 4, pp.e18007.

Lillehammer, M., Goddard, M., Nilsen, H., Sehested, E., Olsen, H.G., Lien, S. & Meuwissen, T., 2008. QTL by environment interaction for milk yield traits on Bos Taurus Autosome 6. *Genet.* 179, 1539-1546.

MacHugh, D.E., Shriver, M.D., Loftus, R.T., Cunningham, P., & Bradley, D.G. 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and Zebu cattle (Bos taurus and Bos indicus). *Genet.* 146, 1071-1086.

MacHugh, D.E., Loftus, R.T., Cunningham, P. & Bradley, D.G., 1998. Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Anim. Genet.* 29, 5, 333-340.

Maiwashe, A., Nephawe, K.A., Van der Westhuizen, R.R., Mostert, B.E. & Theron, H.E., 2006. Rate of inbreeding and effective population size in four major South African dairy cattle breeds. S. Afr. J. Anim. Sci. 36, 1, 50-57.

Makgahlela, M.L., Banga, C.B., Norris, D., Dzama, K. & Ng'ambi, J.W., 2007. Genetic correlations between female fertility and production traits in South African Holstein cattle. *S. Afr. J. Anim. Sci.* 37, 3, 180-188.

Makina, S. O. 2015., A Genome Wide Scan for Signatures of Selection in Six South African Cattle Breeds. Ph.D. (Agric) Diss., University of Pretoria, South Africa.

Makina, S.O., Muchadeyi, F.C., van Marle-Köster, E., MacNeil, M.D. & Maiwashe, A., 2014. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. *Front. Genet.* 5, pp. 333.

Malécot, G. & Blaringhem, L.F., 1948. The mathematics of heredity. Masson, Paris. pp 1-64 (cited by Curik *et al.*, 2014).

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., Williams, J.L. & Macciotta, N.P., 2015. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim. Genet.* 46, 2, 110-121.

Mastrangelo, S., Tolone, M., Di Gerlando, R., Fontanesi, L., Sardina, M.T. & Portolano, B., 2016. Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. *Animal.* 10, 5, 746-754.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'connell, J., Moore, S.S., Smith, T.P., Sonstegard, T.S. & Van Tassell, C.P., 2009. Development and characterization of a high density SNP genotyping assay for cattle. *PloS One* 4, 4, p.e5350.

McCreath, J.F., 1913. Ayrshire Cattle. *Agric. J. of the Union of SA.* 5, 4, pp.529.

Mc Parland, S., Kearney, J.F., Rath, M. & Berry, D.P., 2007. Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. *J. Dairy Sci.* 90, 9, 4411-4419.

McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R., Franklin, C.S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A. & MacLeod, A.K., 2008. Runs of homozygosity in European populations. *AM. J. Hum. Genet.* 83, 3, 359-372.

Melka, M.G. & Schenkel, F.S., 2012. Analysis of genetic diversity in Brown Swiss, Jersey and Holstein populations using genome-wide single nucleotide polymorphism markers. *BMC Res.* Notes 5, 1, 161.

Meuwissen, T., 2009. Genetic management of small populations: A review. *Acta Agr. Scand A.* 59, 2, 71-79.

Meuwissen, T.H.E., Hayes, B.J., & Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819-1829.

Miglior, F., & Beavers. L., 2014. Genetic Diversity and Inbreeding: Before and after Genomics. http://www.progressivedairy.com/dairy-basics/ai-and-breeding/12422-geneticdiversity-and-inbreeding-before-and-after-genomics. Accessed 8 July 2018.

Miglior F., Muir, B.L. & Van Doormaal, B.J. 2005. Selection Indices in Holstein cattle of various countries. *J. Dairy Sci.* 88, 1255–63.

Moore, S.S., Sargeant, L.L., King, T.J., Mattick, J.S., Georges, M. & Hetzel, D.J.S., 1991. The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics.* 10, 3, 654-660.

Mostert, B.E., 2007. The suitability of test-day models for genetic evaluation of dairy cattle in South Africa. Ph.D. (Agric) Diss, University of Pretoria, South Africa.

Nagahata, H., 2004. Bovine leukocyte adhesion deficiency (BLAD): A Review. *J. of Vet Med. Sci.* 66, 12, 1475-1482.

Nagahata, H., Noda, H., Takahashi, K., Kurosawa, T. & Sonoda, M., 1987. Bovine granulocytopathy syndrome: Neutrophil dysfunction in Holstein Friesian calves. *Transbound. Emerg. Dis.* 34, 1-10, 445-451.

Nagahata, H., Oota, H., Nitanai, A., Oikawa, S., Higuchi, H., Nakade, T., Kurosawa, T., Morita, M. & Ogawa, H., 2002. Complex vertebral malformation in a stillborn Holstein calf in Japan. *J. Vet. Med. Sci.* 64, 12, 1107-1112.

Navani, N., Jain, P.K., Gupta, S., Sisodia, B.S. & Kumar, S., 2002. A set of cattle microsatellite DNA markers for genome analysis of riverine buffalo (Bubalus bubalis). *Anim. Genet.* 33, 2, 149-154.

Nei, M., 1973. Analysis of gene diversity in subdivided populations. *Proceedings Nat. Academy Sci.* USA 70, 3321-3323.

Nel, G.D., 1968. Historical Origin of Jerseys (Extract from "Jerseys of S.A." https://www.jerseysa.co.za/Breed-History.htm accessed 8 July 2018.

Nicholas, F. W., 1989. Incorporation of new reproductive technology in genetic improvement programmes. *Evol. Anim. Breed.* pp.203-209 (cited by Nomura *et al.*, 2001).

Nicholas, F.W. & Smith, C., 1983. Increased rates of genetic change in dairy cattle by embryo transfer and splitting. *Anim. Sci.* 36, 3, 341-353.

Nielsen, R., 2004. Population genetic analysis of ascertained SNP data. Hum. *Genom.* 1, 3, 218.

Nielsen, R., 2005. Molecular signatures of natural selection. *Annu. Rev. Genet.* 39, 197-218.

Nicolazzi, E.L., Marras, G. & Stella, A., 2016. SNPConvert: SNP array standardization and integration in livestock species. *Microarrays.* 5, 2, 17.

O'Brien, A.M.P., Utsunomiya, Y.T., Mészáros, G., Bickhart, D.M., Liu, G.E., Van Tassell, C.P., Sonstegard, T.S., Da Silva, M.V., Garcia, J.F. & Sölkner, J., 2014. Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. *Genet. Sel. Evol.* 46, 1, 19.

Peripolli, E., Stafuzza, N.B., Munari, D.P., Lima, A.L.F., Irgang, R., Machado, M.A., do Carmo Panetto, J.C., Ventura, R.V., Baldi, F. & da Silva, M.V.G.B., 2018. Assessment of runs of homozygosity islands and estimates of genomic inbreeding in Gyr (Bos indicus) dairy cattle. BMC Genomics. 19, 1, 34.

Pritchard, J.K. & Przeworski, M., 2001. Linkage disequilibrium in humans: models and data. *AM. J. Hum. Genet.* 69, 1, 1-14.

Pryce, J.E. & Daetwyler, H.D., 2012. Designing dairy cattle breeding schemes under genomic selection: a review of international research. *Anim. Prod. Sci.* 52, 3, 107-114.

Pryce, J.E., Goddard, M.E., Raadsma, H.W. & Hayes, B.J., 2010. Deterministic models of breeding scheme designs that incorporate genomic selection. *J. Dairy Sci.* 93, 11, 5455-5466.

Pryce, J.E., Hayes, B.J. & Goddard, M.E., 2012. Novel strategies to minimize progeny inbreeding while maximizing genetic gain using genomic information. *J. Dairy Sci.* 95, 1, 377-388.

Pryce, J.E., Haile-Mariam, M., Goddard, M.E. & Hayes, B.J., 2014. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genet. Sel. Evol.* 46, 1, 71.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., De Bakker, P.I., Daly, M.J. & Sham, P.C., 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *AM J. Hum. Genet.* 81, 3, 559-575.

Purfield, D.C., Berry, D.P., McParland, S. & Bradley, D.G., 2012. Runs of homozygosity and population history in cattle. *BMC Genet.* 13, 1, p.70.

Purfield, D.C., McParland, S., Wall, E. & Berry, D.P., 2017. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS One.* 12, 5, p.e0176780.

Qwabe, S.O., Maiwashe, A. & Muchadeyi, F.C., 2013. Evaluation of the BovineSNP50 genotyping array in four South African cattle populations. *S. Afr. J. Anim. Sci.* 43, 1, 64-67.

Qanbari, S., Pimentel, E.C.G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A.R. & Simianer, H., 2010. The pattern of linkage disequilibrium in German Holstein cattle. *Animal Genetics* 41, 4, 346-356.

Qanbari, S., Gianola, D., Hayes, B., Schenkel, F., Miller, S., Moore, S., Thaller, G. & Simianer, H., 2011. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics.* 12, 1, 318.

Randhawa, I.A., Khatkar, M.S., Thomson, P.C. & Raadsma, H.W., 2016. A meta-assembly of selection signatures in cattle. *PLoS One.* 11, 4, pp. e0153013.

Robertson, A., 2007. Inbreeding in artificial selection programmes. Gen. Res. 89, (5-6), 275-280.

Robinson, J.L., Drabik, M.R., Dombrowski, D.B. & Clark, J.H., 1983. Consequences of UMP synthase deficiency in cattle. In: *Proceedings of the National Academy of Sciences.* 80, 2, 321-323.

Rodríguez-Ramilo, S.T., Fernández, J., Toro, M.A., Hernández, D. & Villanueva, B., 2015. Genome-wide estimates of coancestry, inbreeding and effective population size in the Spanish Holstein population. *PloS One.* 10, 4, pp. e0124157.

SA Stud Book, 2016. Logix Milk Annual Report 2015/16. http://www.sastudbook.co.za/images/photos/Melk_Jaarverslag2015-16_FIN.pdf accessed 6 December 2018.

Sabir, J., Mutwakel, M., El-Hanafy, A., Al-Hejin, A., Abdel Sadek, M., Abou-Alsoud, M., Qureshi, M., Saini, K. & Ahmed, M., 2014. Applying molecular tools for improving livestock performance: from DNA markers to next generation sequencing technologies. *J. Food, Agric. Environment* 12, 2, 541-553.

Schaeffer, L.R., 2006. Strategy for applying genome-wide selection in dairy cattle. *J. Anim. Breed.* 123, 4, 218-223.

Schwenger, B., Tammen, I., & Aurich, C., 1994. Detection of homozygous recessive genotype for deficiency of uridine monophosphate synthase by DNA typing among bovine embryos produced in vitro. *J. Reprod. Fertil.* 100, 511–14.

Schütz, E., Scharfenstein, M. & Brenig, B., 2008. Implication of complex vertebral malformation and bovine leukocyte adhesion deficiency DNA-based testing on disease frequency in the Holstein population. *J. Dairy Sci.* 91, 12, 4854-4859.

Seidel, G.E., 2010. Brief introduction to whole-genome selection in cattle using single nucleotide polymorphisms. *Reprod. Fert. Develop.* 22, 138-144.

Sewalem, A., Kistemaker, G.J., Miglior, F. & Van Doormaal, B.J., 2006. Analysis of inbreeding and its relationship with functional longevity in Canadian dairy cattle. *J. Dairy Sci.* 89, 6, 2210-2216.

Shanks, R.D. & Greiner, M.M., 1992. Relationship Between Genetic Merit of Holstein Bulls and Deficiency of Uridine-5'-Monophosphate Synthase1. *J. Dairy Sci.* 75, 7, 2023-2029.

Shanks, R.D. & Robinson, J.L., 1990. Deficiency of uridine monophosphate synthase among Holstein cattle. *The Cornell Veterinarian.* 80, 2, 119-122.

Sheehy, P.A., Riley, L.G., Raadsma, H.W., Williamson, P. & Wynn, P.C., 2009. A functional genomics approach to evaluate candidate genes located in a QTL interval for milk production traits on BTA6. *Anim. Genet.* 40, 4, 492-498.

Signer-Hasler, H., Burren, A., Neuditschko, M., Frischknecht, M., Garrick, D., Stricker, C., Gredler, B., Bapst, B. & Flury, C., 2017. Population structure and genomic inbreeding in nine Swiss dairy cattle populations. *Genet. Sel. Evol.* 49, 1, 83.

Smith, L.A., Cassell, B.G. & Pearson, R.E., 1998. The effects of inbreeding on the lifetime performance of dairy cattle. *J. Dairy Sci.* 81, 10, 2729-2737.

Sölkner, J., Ferenčaković, M., Gredler, B. & Curik, I., 2010. Genomic metrics of individual autozygosity, applied to a cattle population. In: 61st Annual Meeting of the European Association of Animal Production.

South African Livestock Breeding, 2004. Dairy Breeds. In: South African Livestock Breeding. Publisher, Picasso Headline. pp 58-65.

Stachowicz, K., Sargolzaei, M., Miglior, F. & Schenkel, F.S., 2011. Rates of inbreeding and genetic diversity in Canadian Holstein and Jersey cattle. *J. Dairy Sci.* 94, 10, 5160-5175.

Stella, A., Ajmone-Marsan, P., Lazzari, B. & Boettcher, P., 2010. Identification of selection signatures in cattle breeds selected for dairy production. *Genetics.* 185, 1451-1461.

Stephan, W., Song, Y.S. & Langley, C.H., 2006. The hitchhiking effect on linkage disequilibrium between linked neutral sites. *Genetics.* 172, 2647-2663.

Su, G., Brøndum, R.F., Ma, P., Guldbrandtsen, B., Aamand, G.P. & Lund, M.S., 2012. Comparison of genomic predictions using medium-density (~ 54,000) and high-density (~ 777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. *J. Dairy Sci.* 95, 8, 4657-4665.

Thomsen, B., Horn, P., Panitz, F., Bendixen, E., Petersen, A.H., Holm, L.E., Nielsen, V.H., Agerholm, J.S., Arnbjerg, J. & Bendixen, C., 2006. A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. *Genome Res.* 16, 1, 97-105.

Thompson, J.R., Everett, R.W. & Hammerschmidt, N.L., 2000a. Effects of Inbreeding on Production and Survival in Holsteins. *J. Dairy Sci.* 83, 8, 1856-1864.

Thompson, J.R., Everett, R.W. & Wolfe, C.W., 2000b. Effects of Inbreeding on Production and Survival in Jerseys. *J. Dairy Sci.* 83, 9, 2131-2138.

Toro, M. A., Fernandez, J. & Caballero, A., 2009. Molecular characterization of breeds and its use in conservation. *Livest. Sci.* 120, 174-194.

Van Graan, A.C., 2016. Annual Milk Bulletin. Agricultural Research Council, Pretoria, South Africa.

Van Raden, P.M., 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 11, 4414-4423.

Van Raden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F. & Schenkel, F.S., 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92, 1, 16-24.

Van Raden, P.M., Olson, K.M., Wiggans, G.R., Cole, J.B. & Tooker, M.E., 2011. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. *J. Dairy Sci.* 94, 11, 5673-5682.

Van Tassell, C.P., Smith, T.P., Matukumalli, L.K., Taylor, J.F., Schnabel, R.D., Lawley, C.T., Haudenschild, C.D., Moore, S.S., Warren, W.C. & Sonstegard, T.S., 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods.* 5, 3, 247.

Verrier, E., Colleau, J.J. & Foulley, J.L., 1993. Long-term effects of selection based on the animal model BLUP in a finite population. *Theor. Appl. Genet.* 87, 4, 446-454.

Weigel, K.A., 2001. Controlling inbreeding in modern breeding programs. *J. Dairy Sci.* 84, E177-E184.

Wiggans, G.R., Van Raden, P.M. & Zuurbier, J., 1995. Calculation and use of inbreeding coefficients for genetic evaluation of United States dairy cattle. *J. Dairy Sci.* 78, 7, 1584-1590.

Wiggans, G.R., Cooper, T.A., Van Raden, P.M., Olson, K.M. & Tooker, M.E., 2012. Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation1. *J. Dairy Sci.* 95, 3, 1552-1558.

Woolliams, J.A. & Oldenbroek, J.K. ed., 2017. Chapter 1 “Genetic diversity issues in animal populations in the genomic era”. In: Genomic management of animal genetic diversity. Publisher, Wageningen Academic. pp13-44.

Wright, S., 1922. Coefficients of inbreeding and relationship. *AM Nat.* 56, 645, 330-338.

Wright, S., 1969. The theory of gene frequencies. Vol. 2. In: Evolution and the genetics of populations (Cited by Keller & Waller, 2002).

Yang, W., Kang, X., Yang, Q., Lin, Y. & Fang, M., 2013. Review on the development of genotyping methods for assessing farm animal diversity. *J. Anim. Sci. Biotech.* 4, 1, 2.

Zenger, K.R., Khatkar, M.S., Cavanagh, J.A.L., Hawken, R.J. & Raadsma, H.W., 2007. Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into Australian and global population variability, including impact of selection. *Anim. Genet.* 38, 1, 7-14.

Zhang, Q., Calus, M.P., Guldbrandtsen, B., Lund, M.S. & Sahana, G., 2015a. Estimation of inbreeding using pedigree, 50k SNP chip genotypes and full sequence data in three cattle breeds. *BMC Genet.* 16 1, 88.

Zhang, Q., Guldbrandtsen, B., Bosse, M., Lund, M.S. & Sahana, G., 2015b. Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC Genomics* 16, 1, 542.

Zhao, F., McParland, S., Kearney, F., Du, L. & Berry, D.P., 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. *Genet. Sel. Evol.* 47, 1, 49.