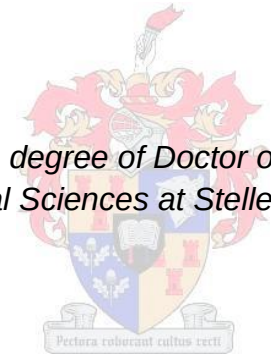
Identification of SNPs associated with robustness and greater reproductive success in the South African Merino sheep using SNP chip technology

by

Lise Sandenbergh



Dissertation presented for the degree of Doctor of Philosophy in Genetics in the Faculty of Natural Sciences at Stellenbosch University

Promotor: Prof. Schalk Cloete,

Co-promotors: Prof. Rouvay Roodt-Wilding, Dr. Aletta van der Merwe

March 2015

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

This dissertation includes original papers published in or submitted to peerreviewed journals or books and unpublished publications. The development and writing of the papers (published and unpublished) were the principal responsibility of myself and, for each of the cases where this is not the case, a declaration is included in the dissertation indicating the nature and extent of the contributions of co-authors.

March 2015

Abstract

Reproduction and robustness traits are integral in ensuring sustainable, efficient and profitable sheep farming. Increases in genetic gain of reproduction and robustness traits are however, hampered by low heritability coupled with the difficulty in quantification of these traits for traditional selective breeding strategies. The aim of the current study was therefore to identify genomic regions underlying variation in reproduction traits and elucidate quantitative trait loci (QTL) and/or genes associated with reproductive traits. The Elsenburg Merino flock has been divergently selected for the ability to raise multiple offspring and has resulted in a High and a Low line that differ markedly with regard to reproductive output and other robustness traits. The flock thus served as an ideal platform to identify genomic regions subject to selection for reproductive traits. To pinpoint genomic regions subject to selection, a whole-genome genotyping platform, the OvineSNP50 chip, was selected to determine the genotype of more than 50 000 SNPs spread evenly across the *ovine* genome. The utility of the OvineSNP50 chip was determined for the Elsenburg Merino flock as well as additional South African Merino samples and three other important South African sheep breeds, the Blackheaded Dorper, South African Mutton Merino (SAMM) and the Namaqua Afrikaner. Although genotyping analysis of the Elsenburg Merino flock indicated some signs of poor genotype quality, the overall utility of the genotype data were successfully demonstrated for the South African Merino and the other two commercial breeds, the Dorper and SAMM. Genotyping results of the Namaqua Afrikaner and possibly other indigenous African breeds may be influenced by SNP ascertainment bias due to the limited number of indigenous African breeds used during SNP discovery. Analysis of pedigree, phenotypic records and SNP genotype data of the Elsenburg Merino cohort used in the current study, confirmed that the lines are phenotypically as well as genetically distinct. Numerous putative genomic regions subject to selection were identified by either an F_{ST} outlier approach or a genomic scan for regions of homozygosity (ROH) in the High and Low lines. Although annotated genes with putative roles in reproduction were identified, the exact mechanism of involvement with variation in reproduction traits could not be determined for all regions and genes. Putative ROH overlapped with QTL for several reproduction, milk, production and parasite resistance traits, and sheds some light on the possible function of these regions. The overlap between QTL for production and parasite

resistance with putative ROH may indicate that several, seemingly unrelated traits add to the net-reproduction and may have been indirectly selected in the Elsenburg Merino flock. A SNP genotyping panel based solely on reproduction traits may therefore be ineffective to capture the variation in all traits influencing reproduction and robustness traits. A holistic selection strategy taking several important traits, such as robustness, reproduction and production into account may as such be a more effective strategy to breed animals with the ability to produce and reproduce more efficiently and thereby ensure profitable and sustainable sheep farming in South Africa.

Opsomming

Reproduksie- en gehardheids-eienskappe is noodsaaklik om volhoubare, doeltreffende en winsgewende skaapboerdery te verseker. ‘n Toename in genetiese vordering in reproduksie- en gehardheids-eienskappe word egter bemoeilik deur lae oorerflikhede tesame met die probleme in kwantifisering van hierdie eienskappe vir tradisionele selektiewe diereteelt strategieë. Die doel van die huidige studie was dus om gebiede in die genoom onderliggend tot variasie in reproduksie-eienskappe te identifiseer en die rol van verwante kwantitatiewe eienskap loki (KEL) en/of gene met reproduktiewe eienskappe te bepaal. Die Elsenburg Merinokudde is uiteenlopend geselekteer vir die vermoë om meerlinge groot te maak en het gelei tot 'n Hoë en 'n Lae lyn wat merkbaar verskil ten opsigte van reproduksie-uitsette en ander gehardheids-eienskappe. Die kudde het dus gedien as 'n ideale platform om genomiese areas onderhewig aan seleksie vir reproduksie-eienskappe te identifiseer. Om vas te stel waar genomiese areas onderhewig aan seleksie gevind kan word, is ‘n heel-genoom genotiperingsplatform, die OvineSNP50 skyfie, gekies om die genotipes van meer as 50 000 enkel nukleotied polimorfismes (ENPs) eweredig versprei oor die skaap genoom, te bepaal. Die nut van die OvineSNP50 skyfie is bepaal vir die Elsenburg Merinokudde sowel as addisionele Suid-Afrikaanse Merinos en drie ander belangrike Suid-Afrikaanse skaaprasse, die Swartkop Dorper, Suid-Afrikaanse Vleismerino (SAVM) en die Namakwa Afrikaner. Hoewel genotipe resultate van die Elsenburg Merino kudde sommige tekens van swak genotipe gehalte getoon het, kon die algehele nut van die genotipering resultate vir die Suid-Afrikaanse Merino en die ander twee kommersiële rasse, die Dorper en SAVM, bevestig word. Genotipering resultate van die Namakwa Afrikaner en moontlik ook ander inheemse Afrika rasse kan deur ENP vasstellingspartydigheid beïnvloed word as gevolg van die beperkte aantal inheemse Afrika rasse gebruik tydens ENP ontdekking. Ontleding van stamboom inligting, fenotipe rekords en ENP genotipe data van die Elsenburg Merino-kohort gebruik in die huidige studie, het bevestig dat die lyne fenotipies asook geneties verskil. Talle vermeende genomiese areas onderhewig aan seleksie is geïdentifiseer deur 'n F_{ST} uitskieter benadering of deur ‘n genomiese skandering vir gebiede van homogositeit (GVH) in die Hoë en Lae lyne. Hoewel geannoteerde gene met potensiële rolle in reproduksie geïdentifiseer is, kan die presiese meganisme van betrokkenheid by variasie in reproduksie-eienskappe nie bevestig word vir al die

gebiede en gene nie. Vermeende GVH oorvleuel met KEL vir 'n paar reproduksie-, melk-, produksie- en parasietweerstand-eienskappe, en werp daarom lig op die moontlike funksie van hierdie gebiede. Die oorvleueling tussen KEL vir produksie en parasietweerstand met vermeende GVH kan daarop dui dat 'n hele paar, skynbaar onverwante, eienskappe bydrae tot net-reproduksie, wat indirek geselekteer mag wees in die Elsenburg Merino-kudde. ‘n ENP genotiperingspaneel uitsluitlik gebaseer op reproduksie-eienskappe mag daarom onvoldoende wees om die variasie in alle eienskappe wat betrekking het op reproduksie- en gehardheids-eienskappe, in te sluit. ‘n Holistiese seleksie strategie wat verskeie belangrike eienskappe, soos gehardheid, reproduksie en produksie in ag neem, mag ‘n meer effektiewe strategie wees om diere te teel met die vermoë om in 'n meer doeltreffende manier te produseer en reproduseer en om daardeur winsgewende en volhoubare skaapboerdery in Suid-Afrika te verseker.

List of publications and presentations

Publications:

Sandenbergh L, Roodt-Wilding R, Van der Merwe AE, Cloete SWP (2013) Analysis of a South African Merino flock divergently selected for reproductive potential. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 20: 98-102

Cloete SWP, Olivier JJ, Sandenbergh L, Snyman MA (2014) The adaption of the South Africa sheep industry to new trends in animal breeding and genetics: A review. South African Journal of Animal Science 44: 307-321

In review:

Sandenbergh L, Cloete SWP, Roodt-Wilding R, Snyman G, Van der Merwe AE. Evaluation of the OvineSNP50 genotyping array for use in four South African sheep breeds. Submitted to the South African Journal of Animal Sciences.

Presentations:

Sandenbergh L, Cloete SWP, Kruger ACM. The 44th South African Society of Animal Sciences Congress, July 2011, Stellenbosch, South Africa. Oral presentation: Variation in reproduction traits contained in a divergently selected South African Merino flock.

Sandenbergh L, Cloete SWP, Roodt-Wilding R, Van der Merwe AE. The 4th International Congress on Quantitative Genetics, June 2012, Edinburgh, Scotland. Poster presentation: Chromosome-specific analyses of SNP chip data from a South African Merino flock divergently selected for reproductive potential.

Sandenbergh L, Cloete SWP, Roodt-Wilding R, Van der Merwe AE. The 22nd South African Genetics Society Conference, September 2012, Stellenbosch, South Africa. Oral presentation: Chromosome-specific analyses of SNP chip data from a South African Merino flock selected for divergent reproductive potential.

Sandenbergh L, Cloete SWP, Roodt-Wilding R, Van der Merwe AE. The 20th Conference of the Association for the Advancement of Animal Breeding and Genetics, October 2013, Napier, New Zealand. Oral presentation: Analysis of a South African Merino flock divergently selected for reproductive potential.

Cloete SWP, Cloete JJE, Hough D, Naidoo P, Olivier JJ, Sandenbergh L, Scholtz AJ, Van Wyk JB. The Cape Wools 9th World Merino Conference, May 2014, Stellenbosch, South Africa. Oral presentation: 'The high line' - continuous selection for the best.

Sandenbergh L, Cloete SWP, Roodt-Wilding R, Snyman G, Van der Merwe AE. The 47th South African Society of Animal Sciences Congress, July 2014, Pretoria, South Africa. Oral presentation: Evaluation of the OvineSNP50 genotyping array for use in the South African sheep industry.

Acknowledgements

Several institutions and individuals have contributed to this PhD project through funding, resources, academic expertise and experience, and technical skills. I would like to thank all contributors to this study:

- Project funding was obtained by Prof. Schalk Cloete from the Technology and Human Resources for Industry Programme (THRIP), Cape Wools and the Western Cape Agricultural Research Trust.
- Research facilities and infrastructure was provided by Stellenbosch University and the Western Cape Department of Agriculture.
- Ms. Annelie Kruger was responsible for data collection and management of the Elsenburg Merino flock. She was assisted by Messrs. Zonwabele Stentyi, Davey Marang and Sakhekile Stentyi.
- Mr. Ryan Turner and Dr. Roelof Hugo assisted with blood collection at Nortier Research Farm. Blood samples from Cradock and Grootfontein Research farms were contributed by Dr. Gretha Snyman on behalf of the Biobank at Grootfontein Agricultural Development Institute.
- My sincerest gratitude to my three supervisors, Prof. Schalk Cloete, Prof. Rouvay Roodt-Wilding and Dr. Aletta Van der Merwe for academic guidance and support.
- Ms. Gerty Mostert and Mr. Alwyn Benson administrated all financial aspects of the project.
- Dr. Ansie Scholtz assisted with editing of the thesis and her attention to detail and willingness to help has contributed immensely in the completion of this thesis.
- Dr. Buks Olivier imparted practical knowledge of the South African sheep industry and afforded me the time to complete this project.

The following people deserve a special thank you:

- My parents.
- Jessica Vervalle for her input in this project as lab manager, friend and confidant.
- …and Matt, who has only been encouraging and supportive.

Table of Contents

List of Figures

List of Tables

List of abbreviations

ACTH: Adrenocorticotropic hormone

AGIS: Agricultural Geo-Referenced Information System

BAC: Bacterial artificial chromosome

BLUP: Best linear unbiased prediction

BMP15: Bone morphogenetic protein 15

BMPR-1B: Bone morphogenetic protein receptor 1B

bp: basepair

CNV: Copy number variation

CRH: Corticotropin releasing hormone

CSIRO: Commonwealth Scientific and Industrial Research Organisation

CV: Coefficient of variation

CYCL2: Cylicin basic protein of sperm head cytoskeleton 2

CYP17: P450 17-hydroxylase/17,20-lyase

DAD-IS: Domestic Animal Diversity Information System

DAFF: Department of Agriculture, Forestry and Fisheries

DNA: Deoxyribonucleic acid

SD: Standard deviation

SE: Standard error

EBV: Estimated breeding value

F1: First generation

FAO: Food and Agriculture Organisation of the United Nations

F_{IS}: Inbreeding coefficient

F_{ST}: Fixation index

GDF9: Growth differentiation factor 9 (GDF9)

GDP: Gross domestic product

He (obs): Observed heterozygosity

HL: High line

HPAA: Hypothalamo-pituitary-adrenal axis

ISGC: International sheep genomics consortium

kb: kilobasepair

LD: Linkage disequilibrium

LL: Low line

MAF: Minor allele frequency

Mb: megabasepair

MCMC: Markov chain Monte Carlo

n: sample number

NA: Namaqua Afrikaner

NCBI: The National Centre for Biotechnology Information

NE: Nebraska

Ne: Effective population size

NLB: Number of lambs born per ewe per joining

NLW: Number of lambs weaned per ewe per joining

PCR: Polymerase chain reaction

QTL: Quantitative trait loci

QTLdb: Quantitative trait loci database

RH: Radiation hybrid

ROH: Regions of homozygosity

RRS: Reduced representational sequencing

SAMM: South African Mutton Merino

SNP: Single nucleotide polymorphism

TGFB: Transforming growth factor beta

TWW: Total weight of lamb weaned per ewe per joining

USA: United States of America

YAC: Yeast artificial chromosome

ZAR: South African Rand

Chapter 1:

Introduction and literature study

1.1. Ovine taxonomy:

Sheep form part of the order Ungulata that contains all hoofed animals. Due to their hoof morphology sheep are further classified into the Artiodactyla, with other even toed Ungulata such as pigs, hippopotami, camels, deer and antelope. Sheep are ruminants with unbranched, hollow and continuously growing horns and are therefore placed in the family Bovidae with species such as bison, gazelle, impala and domestic cattle. The subfamily of Caprinae is reserved for sheep- and goat-like species and is further divided into 10 genera; *Ovis* referring to sheep (Table 1.1). Although sheep and goats are quite similar, several distinctions exist and goats are therefore classed into a separate genus, *Capra*. The distinction between sheep and goat is made with regard to the position of scent glands, horn formation and the absence of beards and knee calluses in sheep (Franklin 1997, Fedosenko & Blank 2005, Parrini *et al.* 2009).

Table 1.1: Scientific classification of sheep, *Ovis aries L*.

Classification	Scientific Name
Kingdom	Animalia
Phylum	Cordata
Class	Mammalia
Order	Ungulata
Suborder	Artiodactyla
Section	Pecora
Family	Bovidae
Subfamily	Caprinae
Genus	*Ovis*
Species	*Ovis aries*

The species partitioning between members of the genus *Ovis* is controversial, given that most of the species can interbreed and have overlapping distributions. Two distinct geographical groups are nonetheless acknowledged: the Eurasian and Asian-American groups. Seven wild *Ovis* species have been proposed in the past (Maijala

1997) based on differences in body size, coat colour, horn morphology, chromosome number and geographic distribution (Fedosenko & Blank 2005, Bunch *et al*. 2006). The Eurasian species includes the argali (*Ovis ammon*), urial (*O. vignei*), Asian mouflon (*O. orientalis*) and European mouflon (*O. musimon*), whereas the bighorn (*O. canadensis*), thinhorn (*O. dalli*) and snow sheep (*O. nivicola*) represent the Asian-American species (Table 1.2) (Maijala 1997).

Table 1.2: The chromosome number and geographical distribution of the proposed extant wild sheep species.

Scientific name	Common name	Chromosome number	Distribution
Ovis vignei	Urial	58	North-east Iran, Afghanistan and north-west India
Ovis ammon	Argali	56	Mountains of central Asia
Ovis orientalis	Asian mouflon	54	Western Asia
Ovis musimon	European mouflon	54	Europe
Ovis canadensis	Bighorn	54	Western North America and Mexico
Ovis dalli	Thinhorn	54	Alaska, western Canada and North America
Ovis nivicola	Snow sheep	52	North-east Asia and Siberia

Adapted from Maijala (1997) and Rezaei *et al.* (2010)

Rezaei *et al*. (2010) combined nuclear and mitochondrial sequence fragments of wild sheep species to infer molecular phylogenies using Maximum parsimony, Bayesian inferences, Maximum likelihood and Neighbour-joining methods. European mouflon samples clustered within the Asian mouflon clade, while samples from other species resulted in monophyletic clades. These results suggest that the European mouflon represents a subspecies of the Asian mouflon and not a separate species as previously thought. Rezaei *et al.* (2010) furthermore proposed that wild sheep underwent successive speciation events concurrent with a migration from their Eurasiatic origin toward central Asia and North America. The Eurasiatic origin of sheep and a consequent migration towards North America through Asia has also been substantiated by fossil records and karyotyping studies (Bunch *et al*. 2006).

1.2. Domestication of sheep:

Archaeozoological evidence, such as an increased frequency of *ovine* remains in archaeological sites, a general change in the body and horn size and the occurrence of sub-adult remains (Maijala 1997, Zohary *et al*. 1998, Pedrosa *et al*. 2005), indicates that sheep and goats were domesticated soon after the domestication of the dog (*Canis lupus familiaris*) more than 14 000 years ago (Vilà *et al.* 1997, Leonard *et al*. 2002). During the Neolithic period (9 000 to 5 000 BC), as sheep and goats were being domesticated, human settlements became more sedentary and increasingly dependent on livestock farming and cereal cultivation (Maijala 1997, Zohary *et al*. 1998, Pedrosa *et al*. 2005).

In the past, the urial, mouflon, argali and hybrids of these species have all been considered as the ancestor of domestic sheep (*O. aries*) (Hiendleder *et al*. 2002). Currently, the mouflon (*O. orientalis*) is thought to be the wild ancestor of the domestic sheep and wild mouflon still roam the area considered to be the centre of domestication – the Near East (South-west Asia). Research is ongoing to determine the location of *ovine* domestication, the progenitor species and the number of domestication events (Zohary *et al*. 1998, Pedrosa *et al*. 2005, Meadows *et al*. 2007). However, the increased level of *ovine* nucleotide diversity in the proximity of the Near East does suggest that this area may have been the site of domestication (Meadows *et al*. 2007).

Currently five mitochondrial lineages have been identified in domestic sheep (Meadows *et al.* 2007). These multiple mitochondrial lineages serve as evidence for the inclusion of separate lines into the gene pool of domestic sheep and are suggestive of multiple domestication events. Individuals from progenitor species were therefore integrated in the sheep gene pool at several instances during domestication. This trend is also evident in other domestic species, such as cattle, goats and pigs (Meadows *et al*. 2007, Naderi *et al*. 2008, Amaral *et al.* 2011, McTavish *et al*. 2013, Decker *et al*. 2014).

During domestication, humans provided shelter, better nutrition and protection from natural predators to sheep. Humans also facilitated selection on behavioural traits that enabled the co-habitation of humans and sheep. Husbandry practices such as culling

young male animals also affected the social structure and morphology of sheep. The direct (intentional) and indirect selection pressure applied to sheep during the domestication process has consequently resulted in domestic sheep being more docile, less agile, shorter limbed, smaller, less camouflaged and possessing varying horn formation in comparison with their wild ancestors (Maijala 1997, Zohary *et al.* 1998).

1.3. Sheep breeds:

Ovis aries has spread from its putative centre of domestication, South-west Asia, across the globe and at present sheep have a global distribution and are found on every continent except Antarctica (Cottle 2010). Initially sheep were mainly kept for meat, milk and skins, while secondary products, such as wool, were only utilised since approximately 6000 BC (Zohary *et al*. 1998, Chessa *et al*. 2009, Cottle 2010, Gutiérrez-Gil *et al*. 2014). Selection for meat, milk or wool under varying environmental conditions has resulted in a spectrum of modern breeds specialised for a single products or a combination of products (Kijas *et al*. 2012). Domestic sheep are divided into more than 2 400 breeds and comprise 17% of all known livestock breeds (DAD-IS 2014). In some instances the difference between animals from the same breed are more pronounced than the differences between separate breeds (Maijala 1997). This is the result of adaptation to the local environment, differences in selection strategies or outbreeding within breeds. Prefixes or a completely novel name is used to distinguish strains that have diverged from the breed at large, as is the case with the South African Merino, South African Mutton Merino, Dohne Merino and Australian Merino (Cottle 2010).

1.3.1 Merino:

The Merino sheep is a fine wool sheep with its origin in Spain. Archaeological evidence indicates that Merino-like animals inhabited the south of Spain since pre-historic times (Cottle 2010). Scrolls dating back to the Roman occupation of Iberia, about 2 000 years ago, make reference to sheep with superior wool quality. The Romans selected these sheep for a white coat colour and outcrossed them to other breeds, mainly African sheep. Merino-like sheep spread throughout the Spanish peninsula through the migration of people and their livestock to seasonal grazing, but the spread of the breed was curbed by a ban on the export of sheep from Spain (Cottle 2010). During

this time the Merino gained protection and was officially named by a newly formed group of Spanish sheep breeders. The ban on export was only lifted during the 1700's and led to a spread of the Merino to nearby European countries. The Merino reached other continents in the late 1700's through export to Australia, North America, Argentina and South Africa (Maijala 1997, Diez-Tascón *et al*. 2000, Cottle 2010).

The Merino has formed the foundation stock for other breeds in several countries (Diez-Tascón *et al*. 2000, Cottle 2010) and its influence on the genetic structure of sheep breeds is evident in the extent of haplotype sharing between the Merino and other breeds. Haplotype sharing between the Merino and European breeds is particularly common (Kijas *et al*. 2012). The 45 or more strains of Merino or Merino-based breeds have a global distribution and differ in traits such as wool production, reproduction and the presence or absence of horns. Although the Merino is mainly a wool breed, several strains have been bred for meat production (Maijala 1997).

1.3.2 Merino in South Africa:

In the late 1700's the king of Spain gave the Dutch government six Merino sheep; two rams and four ewes (Cottle 2010). These sheep did not thrive in the Netherlands and were moved to the Dutch Cape Colony. The Dutch settlers at the Cape Colony considered the Merino as an oddity rather than a profitable breed and the number of Merinos only increased considerably after the Colony was surrendered to Britain in the early 1800's. The increase in Merino numbers resulted in a concurrent increase in the quantity of wool exported to England as well as a steady increase in the carcass and wool quality as uniformity was achieved in the flocks maintained in the Cape and further inland (McKee 1913). The introduction of other Merino strains, most notably those from Australia, occurred during the last 200 years and today the Merino or Merino-like sheep comprise more than 50% of the national sheep flock in South Africa (Cloete & Olivier 2010).

The South African Merino is the primary fine wool producing breed in South Africa and is also utilised for meat production. Other Merino-based breeds are specifically bred for the dual production of wool and meat and include breeds such as the South African Mutton Merino and Dohne-Merino. The South African Merino excels at its intended purpose and the volume of Merino wool sold at auctions is more than double that of

all other breeds combined. In 2011/2012 South Africa produced approximately 30 100 tonnes of Merino wool in comparison to 13 500 tonnes of wool from other breeds. The price of Merino wool also exceeds the price of wool from other breeds by more than ZAR15/kg (ZAR43.58/kg in comparison to ZAR28.46/kg for the other breeds in 2010/11, DAFF 2013).

1.3.3. Other sheep breeds in South Africa:

1.3.3.1. Dorper:

Several factors including the economic decline after the First World War, a drop in the wool price and the production of surplus meat, led to an increased interest in the export of sheep meat to Britain in the 1930's. Lambs with a fast growth rate that could produce good quality carcasses under veld conditions, were in demand. This led to crossbreeding of indigenous sheep with British mutton breeds to establish a local breed that would meet the export market needs (Milne 2000). The newly developed breed needed to replace the local fat-tailed sheep, which were unacceptable to the British markets at the time. These sheep also needed to be adaptable to the winter rainfall regions of the Cape as well as survive the low rainfall and harsh conditions in the Karoo. Subsequently, the South African Department of Agriculture experimented with crossbreeding the Blackheaded Persian and Dorset Horn. The Blackheaded Persian originated in Somalia and Saudi-Arabia, but is considered an indigenous South African breed due to its protracted history in this country (Soma *et al.* 2012). The breed is well-suited to harsh environmental conditions and exhibits superior mothering ability and was therefore used as dams in crossbreeding trials. The Dorset Horn is a British mutton breed and was selected as the paternal line for crossbreeding trials as the breed exhibited a longer breeding season than other British breeds and has superior carcass qualities. The resulting breed, the Dorper, is a 50-50 composite of the Dorset Horn and Blackheaded Persian. The Dorper is a meat producing breed that is easily maintained and is highly adaptable to challenging environmental conditions. For these reasons the breed has gained wide-spread popularity and is currently the major meat producing breed in South Africa. The Dorper is the second most abundant breed after the Merino in the country (Cloete & Olivier 2010).

1.3.3.2. Namaqua Afrikaner:

The Namaqua Afrikaner is a hardy, fat-tailed sheep indigenous to South Africa and is primarily maintained under extensive conditions in smallholding farming systems (Qwabe *et al*. 2013). The history of the breed dates back to the migration of the Khoikhoi tribes and their sheep into South Africa around 400 AD (Snyman *et al*. 1993, Deacon & Deacon 1999, Farm Animal Conservation Trust 2001). Under challenging environmental conditions the Namaqua Afrikaner exhibits acceptable levels of production, and reproduction rates that are comparable to those of commercial South African breeds such as the Dorper, Afrino and Merino (Snyman *et al.* 1993, Schoeman *et al*. 2010). The Namaqua Afrikaner stores fat reserves in its tail which results in poor fat distribution throughout the carcass. The aforementioned fat deposition characteristics of a Namaqua Afrikaner carcass has made the breed unfavourable for commercial production in the past (Milne 2000, Schoeman *et al.* 2010). Namaqua Afrikaner lambs have been shown to be inferior to Dorper and South African Mutton Merino (SAMM) lambs for carcass meat yield, expressed as a percentage, in a study of Burger *et al*. (2013). Namaqua Afrikaner lambs have also been shown to exhibit a higher percentage of bone in all retail cuts compared to the commercial breeds, as well as a lower carcass weight (Burger *et al.* 2013). These characteristics have been an extenuating factor in the decline of the breed and have resulted in the breed being at risk of extinction (Qwabe *et al.* 2013). The South African Department of Agriculture intervened in 1966 and has been maintaining Namaqua Afrikaner flocks at several research farms to ensure the survival of the breed. The breed is however still endangered with only 100 to 1 000 breeding ewes and 6 to 20 breeding rams remaining (FAO 2000, Snyman *et al.* 2013, Qwabe *et al.* 2013).

1.3.3.3. South African Mutton Merino (SAMM):

The first German Merinos, or German Merinofleischschaf, were imported to South Africa by the Elsenburg Agricultural College in 1932 (Schoeman *et al*. 2010). The German Merino was bred to produce wool and meat and was introduced to South Africa to improve meat production of the local Merino. Owing to the influence of the German Merino and continued selective breeding for a locally adapted dual-purpose sheep breed, the SAMM was recognised as a separate breed in 1971. The breed is the major dual-purpose breed in South Africa and has been exported widely and has gained popularity, especially in Australia (Cloete & Olivier 2010, Cottle 2010). The

SAMM is known for producing a heavier slaughter lamb without excessive fat in the carcass (Cloete *et al*. 2007a), while also producing good quality white wool. These characteristics have made the breed popular in feedlots where emphasis is placed on feed conversion efficiency and carcass quality (Cottle 2010).

1.4. Resource flocks:

Several sheep resource flocks have been established and are housed across a range of environmental conditions in South Africa. These flocks are maintained by agricultural research and training institutes and used for research, conservation of endangered breeds and demonstration purposes. Research on these flocks has been cardinal in improving the scientific knowledge of sheep breeding practices in South Africa (Schoeman *et al.* 2010). Most of these flocks are maintained either in structured breeding experiments or as studs that reflect breed-specific breeding practices (Schoeman *et al.* 2010). Flocks maintained on four research farms in the Western and Eastern Cape provinces formed part of the current study and include the Elsenburg Merino flock; the Dorper, SAMM and Namaqua Afrikaner flocks at Nortier Research Farm and the Merino studs at Cradock and Grootfontein (Figure 1.1). In sections 1.4.1 to 1.4.4 the background of each of the aforementioned research flocks is summarised.

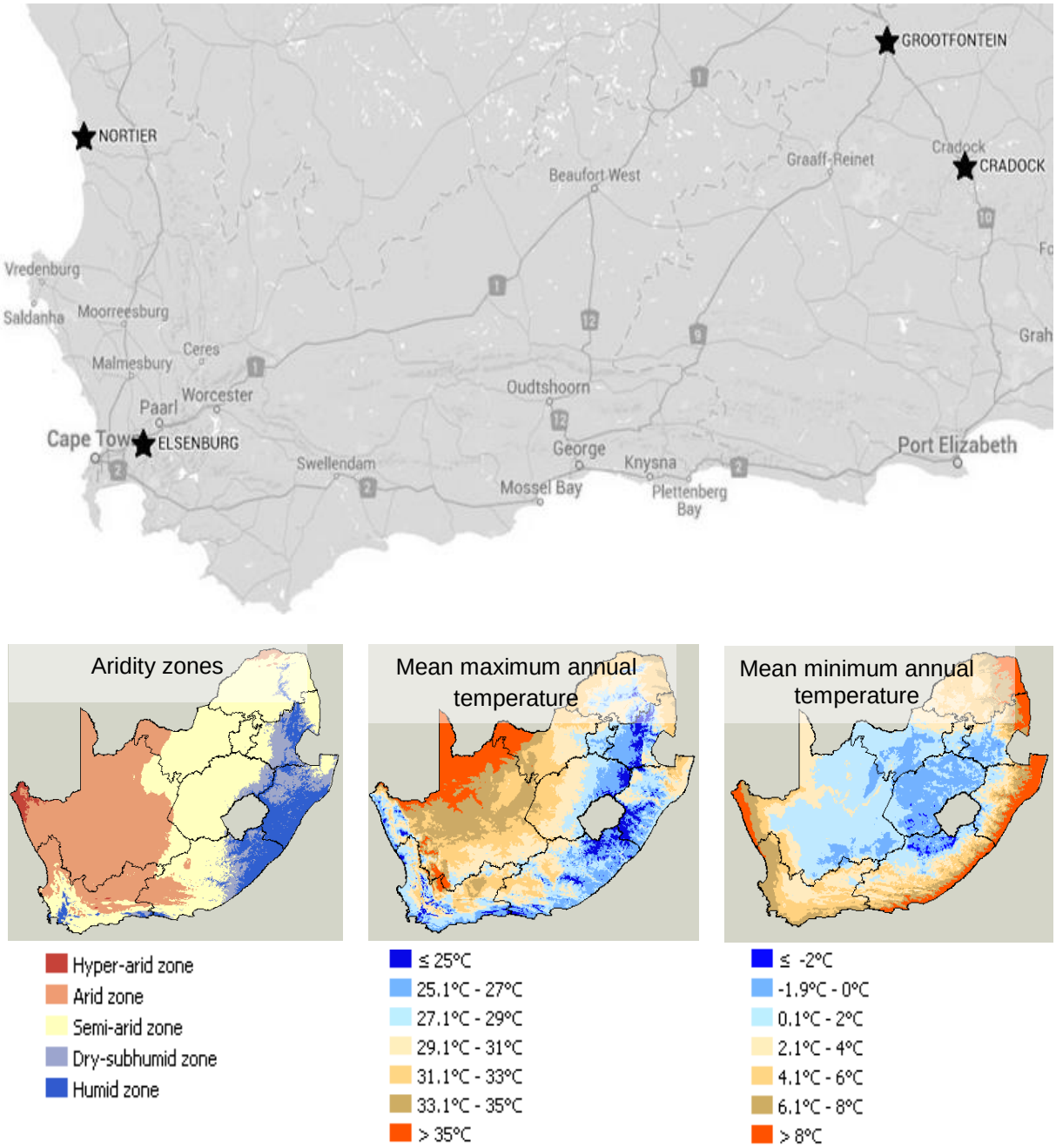


Figure 1.1: The location of four research farms housing the sheep flocks relevant to the current study (top) and the prevailing aridity and temperature zones of South Africa (bottom).
Map created using Google (www.google.com/maps) with information generated by AfriGIS (Pty) Ltd (2014) (www.afrigis.co.za). Aridity and temperature map information adapted from AGIS (2007).

1.4.1. Elsenburg Merino flock:

The Elsenburg Research Farm is situated near Stellenbosch in the Boland region of the Western Cape Province of South Africa. The Elsenburg Merino flock was established from animals that formed part of a former selection trial at the Tygerhoek Research farm near Riviersonderend in the Overberg district of the Western Cape. Three lines were maintained at Tygerhoek; a control line and two wool selection lines. The selection lines were bred for an increase in clean fleece weight and a larger secondary to primary follicle ratio (S:P line), respectively (Heydenrych *et al*. 1984, Schoeman *et al.* 2010). A total of 240 ewes, divided according to their reproductive output, were selected from the Tygerhoek S:P line and formed the base of the Elsenburg Merino selection lines. Rams were selected from the clean fleece selection line or the control line. The experimental design of Elsenburg flock centres on divergent selection for reproduction, with most of the selection emphasis having been placed on number of lambs weaned per mating (Cloete *et al.* 2004, 2009). Since the first lambs were born in 1987, selection pressure has been applied for approximately nine generations during the ongoing trial. Divergent selection for the ability to rear multiple offspring has been applied by selecting breeding stock on maternal ranking values and since 2003, by evaluating their best linear unbiased prediction (BLUP) estimated breeding value (EBV) for the number of lambs weaned per ewe per joining (NLW). This has resulted in a 1.5% per year increase in the number of lambs weaned per ewe and a 1.8% increase in the total weight of lamb weaned per joining in the line selected for improved lamb rearing ability, further on referred to as the ‘High line’. Conversely, a 0.8 to 1.0% decline in the number of lambs weaned and a 1.0 to 1.2% decline for total weight of lamb weaned per joining was noted in the ‘Low line’ line selected for a decrease in lamb rearing ability (Cloete *et al*. 2004, 2007b). The divergence in these traits also led to divergence in the weaning weight, yearling weight and lamb survival (Cloete *et al.* 2003, 2005a, 2009) as well as other traits related to robustness, such as the animal’s reaction to humans in an arena test (Cloete *et al.* 2010) and susceptibility to breech blowfly strike (Scholtz *et al*. 2010). Furthermore, behavioural differences that influence lamb survival have been noted between the two lines (Cloete *et al.* 2003, 2005b). Although the lines differ significantly with regard to several reproductive traits, wool quantity (clean fleece weight) and quality (fibre diameter) was unaffected by selection for reproduction (Cloete *et al*. 2005a). At least one polymorphism in the *cytochrome P450 17α-hydroxylase/17,20-lyase* (*CYP17*)

gene has been linked to variation in robustness traits related to stress response. Heterozygous individuals (WT1/WT2) from the Low line performed poorly in stress tests, while homozygous (WT1/WT1) individuals from both lines as well as heterozygous individuals (WT1/WT2) from the High line coped better in stressful situations (Van der Walt *et al*. 2009, Hough *et al*. 2010).

1.4.2. Dorper, Namaqua Afrikaner and SAMM flocks at Nortier Research Farm:

The Nortier Research Farm is situated near Lambert's Bay on the west coast of the Western Cape Province of South Africa. Several purebred sheep breeds are maintained on the farm and include the Dorper, Namaqua Afrikaner and SAMM. The Namaqua Afrikaner flock has been bred from individuals from the Namaqua Afrikaner conservation flock at Klerefontein in the Karoo (see Schoeman *et al.* 2010 for more information on this flock) and unrelated individuals have occasionally been introduced from other conservation flocks. Commercial rams are introduced regularly to the Dorper and SAMM flocks and therefore these flocks closely resemble the gene pool of the national sheep flock. Although the Namaqua Afrikaner, Dorper and SAMM animals are kept as a single ewe flock, the Namaqua Afrikaner is exempt from the commercially driven breeding strategies applied to the other two breeds. The Namaqua Afrikaner flock is maintained mostly for conservation purposes and the preservation of genetic diversity within the breed is of greater importance than an improvement in commercial production traits. This resource flock has been used to compare the available breeds for carcass traits (Burger *et al*. 2013), flock-isolation behaviour (Cloete *et al*. 2013a), tick counts and udder health (Cloete *et al.* 2013b). Namaqua Afrikaner sheep exhibited poorer carcass composition, were more likely to bleat when isolated from their contemporaries, had better udder health and lower body tick counts than the commercial ewes maintained alongside.

1.4.3. Grootfontein Merino flock:

The Grootfontein Merino stud is situated near Middelburg in the Eastern Cape, an area that forms part of the Karoo. The history of the stud can be traced as far back as 1912 when two rams and 63 ewes were imported from New South Wales in Australia for breeding purposes. Until 1958 ewes from local Merino stud breeders as well as imported Australian rams were introduced to the flock. In 1958 the Grootfontein Merino stud was registered with the Merino Breeders' Society and from this point on new

genetic material was rarely introduced. In 1973 the stud was reduced to 300 ewes of which 20 to 25% was replaced annually (Olivier 1989). During 1968 to 1985 the stud's main selection aim was conformation, but this shifted towards wool and reproduction traits as methods for accurately estimating breeding values became available. Prior to 1986, selection based on conformation led to gradual increases in the annual breeding values for live weight and clean fleece weight, and an unwanted increase in fibre diameter (Olivier *et al*. 1995, Schoeman *et al.* 2010). After selection objectives were amended to maintain clean fleece weight while increasing live weight and reduce fibre diameter, gains were observed in the desired direction. The former study was important as it demonstrated the value of using objective selection criteria to the broader Merino industry. From 1999 onwards, selection was applied to a subset of the Grootfontein flock to reduce the fibre diameter, while another subset was selected for an increased live weight of lambs born and clean fleece weight. The stud has been utilised for estimating genetic parameters, such as heritability, covariance and correlation of many wool and reproduction traits through the years and has greatly contributed to industry related research (Schoeman *et al.* 2010).

1.4.4. Cradock Merino flock:

The Cradock Merino flock, formerly known as the Halesowen Fine wool stud, is situated near Cradock in the north-eastern Karoo of the Eastern Cape. A need for fine wool sheep and research pertaining to fine wool in the South African context led to the establishment of the stud in 1988 (Schoeman *et al.* 2010). The aim was to establish a genetic fine wool resource that would supply rams to industry to capitalise on the price premium for finer wool at that stage. A further aim was to enable research on fine wool under intensive pasture and under more limiting extensive conditions in South Africa. Initially, South African fine wool ewes were bred with Australian fine wool rams for two consecutive years and thereafter selection was applied to increase live weight of lambs born and clean fleece weight. After 1995 selection was applied to reduce fibre diameter and increase live weight of lambs born while maintaining clean fleece weight (Olivier 2014). Genetic trends for three phases of selection in the flock history were also reported by the latter author.

1.5. The role of sheep in global agriculture:

The world's human population has doubled since the late 1970's and now amounts to around 7 billion. If current population growth rates continue, the population will grow to 9 billion by 2050. To meet the projected short-term nutritional demands of the world's population, global agriculture production will have to increase by 60% from its 2005 to 2007 levels (FAO 2013). Not only is agriculture essential in meeting the demands created by population expansion, but it also provides a means by which millions of people make a living. Approximately 2.5 billion people are directly dependent on agricultural exploits for their livelihood. Agriculture is a building block in economic growth and also safeguards developing countries in particular against global economic and financial crises (FAO 2013). Investment and innovation in the growth of the agriculture sector is therefore important for the development and growth of economies, particularly those of developing countries. Agricultural innovation, stemming from cutting-edge agricultural research, further ensures global competitiveness.

The global consumption of livestock products has increased steadily and will continue to increase based on current estimates. In Asia, where the greatest population increase has taken place, a 3 to 5% annual increase in the consumption of meat or dairy products occurred between 2000 and 2010. The global response has been an increase in livestock production with an associated increase in the production of animal feeds by the cropping and fishery sectors. Livestock farming occupies the largest section of agricultural land with crops and pastures comprising 30% of all available land on earth. Due to the widespread and far-reaching effects of agricultural activities, natural resources and the environment are increasingly threatened. The expansion of livestock industries have resulted in increased deforestation, overgrazing, effluent production and the production of greenhouse gasses (FAO 2013). Improvements in the efficiency of livestock production to minimise the impact on the environment is therefore essential to keep providing food for the global population.

Globally, cattle, buffalo, pigs, sheep, goats and poultry species are the main meat producing livestock animals. Of the 300 million tonnes of meat produced worldwide in 2010, pork comprised the largest quantity by weight. Beef, buffalo, sheep and goat

meat contributed the least to the global total. However, in Africa pigs contributed the least to the meat produced, while sheep and goat meat comprises a considerable larger percentage in comparison to the global outlook. In South Africa, beef and poultry are the main meat producers while sheep and goat meat only contribute 6% to the total (Figure 1.2). Asia produced more than half the sheep and goat meat in 2010, with Europe and the combination of Australia and New Zealand producing most of the remainder. This is not surprising considering that Asia houses nearly half of the total of 1 billion sheep in the world (Cottle 2010). South Africa is a relatively small producer and only contributes 1.3% of the sheep meat produced globally (Figure 1.3) (FAO 2013).

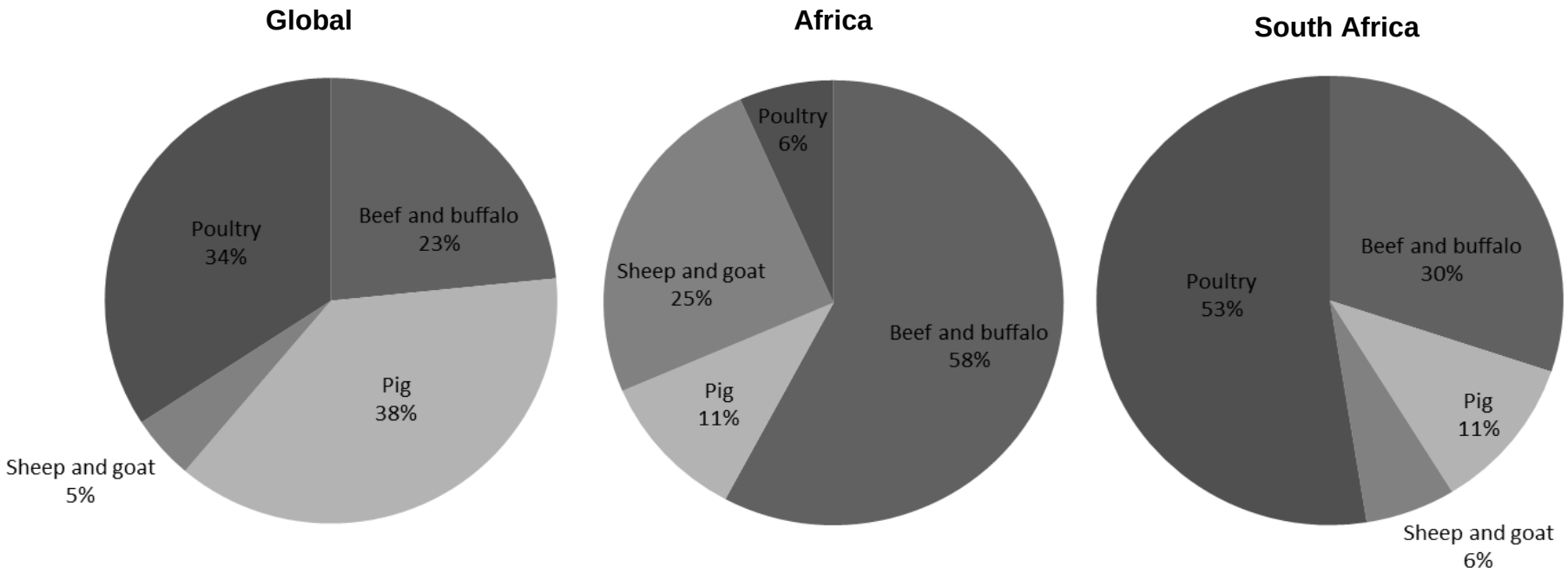


Figure 1.2: Meat production on a global, regional (Africa) and national (South Africa) scale according to farmed livestock species.
Adapted from FAO 2013.

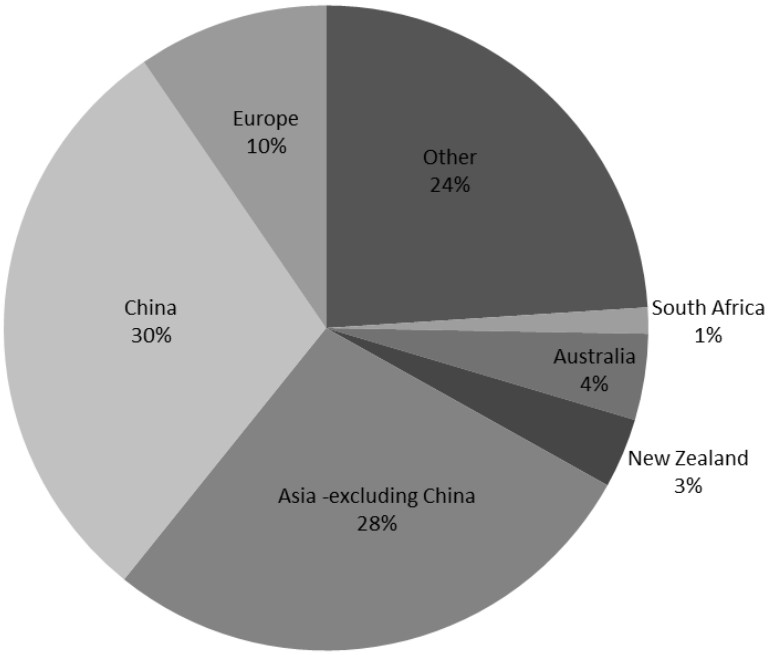


Figure 1.3: The countries producing sheep and goat meat worldwide.
Adapted from FAO 2013.

Wool plays a major role in the economics of the global sheep industry. Unfortunately the wool price has varied considerably in recent history and has resulted in uncertainty in projections based on the wool price. The profitability of sheep meat in relation to wool has increased and in some instances surpassed that of wool (Cloete & Olivier 2010). This has been a contributing factor in the 38% decrease in the amount of wool produced globally between 1990 and 2007 (Cottle 2010). In the last decade China, Australia and New Zealand have been responsible for more than 50% of the global wool production. South Africa is in the top 20 wool producing countries and contributes about 2.3% to the global wool production per weight (Cottle 2010). South African wool is mostly used in high value apparel wool products while several of the other larger wool producing countries produce mainly less valuable carpet wool (Cloete & Olivier 2010).

Even though Asia and specifically China is the largest producer of sheep meat and wool, Australia and New Zealand remain the main exporters of lamb, mutton and wool owing to their greater production of wool and meat per sheep as well as their small domestic market (Cottle 2010).

1.5.1. Sheep industry in South Africa:

South Africa is considered a developing economy where agriculture, forestry, fisheries and hunting contributes 2.4% to the annual national GDP (DAFF 2013). A large proportion of South Africa is considered arid to semi-arid (Figure 1.1). Only 12.4% of South African land mass is considered arable and much of the western and central districts are suitable only for extensive livestock production (AGIS 2007, DAFF 2013). Extensive livestock production therefore plays an important role in utilising land mass that would otherwise be unsuitable for intensive or semi-intensive plant production (Cloete & Olivier 2010). Livestock production contributes 48% of the gross agricultural production in South Africa (DAFF 2013). Sheep and goat farming is practiced throughout South Africa, but plays an especially important role in the arid to semi-arid regions, such as the Karoo, where the scope for other agricultural activities are limited. Sheep farming is also practiced in areas where intensive pasture-cropping and horticulture is practised as sheep are able to utilise crop residues and by-products from these industries. In these regions sheep also serve as a fall-back to assure the livelihood of farmers in the event of crop-failure or decreases in crop prices (Cloete & Olivier 2010).

South Africa's sheep population amounts to approximately 22 million (DAFF 2013) and encompasses roughly 2% of the global sheep population (FAO 2013). Wool and meat of sheep and goats account for about 8% of the gross value of livestock products and about 4% of the total agricultural value in South Africa. Wool is among the top ten South African export products and R2 billion worth of wool was exported in 2011 (DAFF 2013). South Africa produced 176 000 tonnes of sheep meat in 2010 (FAO 2013) and approximately 15 000 tonnes were imported additionally (DAFF 2013).

Considering the aforementioned importance of sheep faming to the South African economy and the benefit to farming communities, the improvement and expansion of the industry is of great socio-economic importance. The demand for sheep meat is greater than the local supply and as a result South Africa remains a net importer of sheep meat. An opportunity therefore exists for increasing the production of mutton and lamb in South Africa.

1.6. Genomics in agriculture:

Numerous molecular and bioinformatics tools were developed during the sequencing and completion of the human genome in the 2000's. These technologies have easily been transferred to other mammals, such as livestock, and have driven the advancement of genomic research in these species (Hu *et al*. 2009, Fan *et al*. 2010). Since the genome sequence of the red junglefowl became available in 2004, livestock genomic resources have increased and the genome sequences of many livestock species, including pigs, cattle, sheep, horses and rabbits are now available (International Chicken Genome Sequencing Consortium 2004, Zimin *et al.* 2009, Fan *et al*. 2010, Bai *et al*. 2012, Jiang *et al.* 2014). Genomic technologies, including whole-genome sequencing, next-generation sequencing and high-throughput genotyping, have many potential applications in livestock research and therefore the implementation of these technologies is expected to increase (Fan *et al*. 2010, Bai *et al*. 2012).

Whole-genome sequencing strategies have enabled the identification of large numbers of genetic markers, including many single nucleotide polymorphisms (SNPs). Millions of SNPs have been identified in livestock species such as chicken (~2.8 million SNPs), cattle (~2.2 million SNPs) and horse (~1.1 million SNPs) (International Chicken Polymorphism Map Consortium 2004, The Bovine HapMap Consortium 2009, Wade *et al*. 2009). The use of SNPs in population genetic and animal breeding studies has become widespread as these markers are distributed widely throughout the genome and are easy to evaluate and interpret (Brumfield *et al*. 2003, Morin *et al*. 2004, Kijas *et al.* 2009, Fan *et al.* 2010). A panel of polymorphic SNPs with sufficient coverage of the genome have numerous applications, including determining levels of linkage disequilibrium (LD), species' evolution, domestication and breed formation studies (The Bovine HapMap Consortium 2009, Fan *et al*. 2010, Kijas *et al*. 2012, Decker *et al*. 2014). A further application of SNPs include pinpointing genomic regions underlying complex livestock traits and use in genomic selection strategies aimed at increasing genetic gains (Hayes *et al*. 2009, Daetwyler *et al*. 2010, Fan *et al*. 2010, Snelling *et al*. 2010, Bolormaa *et al*. 2011).

Microarray SNP genotyping platforms offer an alternative to whole-genome sequencing by enabling genotyping of thousands of SNPs throughout the genome in

a timesaving and cost-effective manner, thereby enabling the use of large sample cohorts necessary for a variety of livestock applications (Brumfield *et al*. 2003, Morin *et al*. 2004, Fan *et al*. 2010). These SNP genotyping arrays are increasingly utilised in the field of livestock research and is changing traditional animal breeding and genetic studies (Hu *et al*. 2009, Fan *et al*. 2010). Illumina and Affymetrix are the main manufacturers of whole-genome SNP genotyping platforms for livestock species (Fan *et al*. 2010). Currently, Illumina manufactures most of the commercial livestock SNP chips owing to the affordability and flexibility of the Illumina Infinium II genotyping platform (Perkel 2008, Fan *et al*. 2010). The number of SNPs tested by species-specific SNP genotyping platforms currently range from 96 (cattle parentage chip) to more than 700 000 ('high-density' cattle chip) (Bai *et al*. 2012, Mullen *et al.* 2013).

Most of the economically important livestock traits are quantitative traits that are influenced by many loci spread throughout the genome (Andersson & Georges 2004, Goddard & Hayes 2009). Pinpointing these loci is challenging and require large sample cohorts together with sufficient numbers of marker loci spread throughout the genome (Goddard & Hayes 2009, Hu *et al*. 2009, Zhang *et al*. 2012). The aforementioned livestock genomic resources together with newly developed analysis strategies may prove successful and indeed have already been relatively successful in identifying regions underlying quantitative trait loci (QTL) in livestock species (Fan *et al*. 2010, Zhang *et al*. 2012). Association studies employing whole-genome SNP data have identified candidate genes underlying production traits such as feed conversion (Barendse *et al*. 2007), milk quality (Schopen *et al.* 2011), carcass quality (Bolormaa *et al*. 2011), fertility (Demars *et al*. 2013), body weight (Snelling *et al*. 2010), and disease phenotypes (Zhao *et al*. 2011) in cattle, pigs and sheep. Whole-genome SNP data have also been incorporated into population genetic approaches to estimate population divergence (MacEachern *et al.* 2009, The Bovine Hapmap Consortium 2009), breed formation (The Bovine Hapmap Consortium 2009, Kijas *et al*. 2012), whole-genome LD patterns (McKay *et al*. 2007, Kijas *et al*. 2014), and signatures of selection (Stella *et al*. 2010, Moioli *et al*. 2013, Phua *et al*. 2014) of livestock species. Recently, copy number variation, parentage assignment, traceability and genomic selection strategies relying on whole-genome SNP data have also been investigated (Goddard & Hayes 2009, Hayes *et al*. 2009, Daetwyler *et al*. 2010, Fan *et al*. 2010, Weller *et al*. 2010, Liu *et al*. 2013, Rowe *et al.* 2013, Heaton *et al*. 2014).

1.6.1. Ovine genomics:

Since its inception in 2002, the International Sheep Genomics Consortium (ISGC) has been responsible for the development of most of the sheep genomic resources, including the publication and online availability of the *ovine* genome. The draft *ovine* genome became publically available in 2010 (International Sheep Genomics Consortium *et al*. 2010) and an updated version, Oar v3.1, followed in 2013 (http://www.ncbi.nlm.nih.gov/genome/genomes/83). Oar v3.1 was produced from two Texel sheep, a ram and a ewe, and has an assembled length of 2.61 Gb with approximately 99% of the sequence anchored onto the 26 autosomal chromosomes and the X-chromosome (Jiang *et al.* 2014). Approximately 0.2% of each of the respective genomes were heterozygous SNP loci and of these, 25% were heterozygous in both individuals.

Other sheep genomic resources include microsatellite, parentage and barcoding panels, 1 536 and 50 000 SNP genotyping panels (see Section 1.6.3. for a description of the microarray genotyping platform), linkage maps, bacterial artificial chromosome (BAC) and yeast artificial chromosome (YAC) libraries, radiation hybrid (RH) panels and physical genome maps (Maddox & Cockett 2007, Kijas *et al.* 2009, International Sheep Genomics Consortium *et al*. 2010). The National Centre for Biotechnology Information (NCBI) maintains a sheep genome resource page (http://www.ncbi.nlm.nih.gov/genome?term=ovis%20aries) where NCBI resources such as Blast, Mapview and UniGene as well as links to other online resources can be accessed. The website of the ISGC (www.sheephapmap.org) and the Animal, Food and Health Sciences livestock genomics website of CSIRO (http://www.livestockgenomics.csiro.au/sheep) contains useful links to the virtual sheep genome and other online resources such as Ensemble (http://www.ensembl.org/Ovis_aries/Info/Index/) and scientific publications. Recent advances in sheep genomics have been the development of a 163 SNP parentage panel as well as a 600 000 SNP genotyping platform (Heaton *et al*. 2014, Kijas *et al*. 2014). The parentage panel produced by Heaton *et al*. (2014) may prove useful in accurately determining pedigrees as well as ensuring the traceability of sheep and sheep products, while the 600 000 SNP chip will increase the coverage of the *ovine* genome.

The aforementioned genomic resources have been used to study genome-wide LD patterns (Miller *et al*. 2011, Kijas *et al.* 2012), identify QTL through the use of traditional QTL mapping methods (Van der Werf *et al*. 2007), genome-wide association (Zhang *et al*. 2012) and selection sweep (Moioli *et al.* 2013, Gutiérrez-Gil *et al.* 2014, Phua *et al*. 2014). The Sheep QTL database (http://www.animalgenome.org/cgi-bin/QTLdb/OA/index, Hu *et al*. 2013) contains curated sheep QTL and association data from 90 scientific publications. Presently, a total of 789 QTL for 219 different traits have been deposited in the database. Most QTL are involved with meat characteristics (32%), health (21%), milk (20%) or production (16%) traits, while traits involved with reproduction (5%), wool (4%) and the animal's external appearance (3%) are not as well represented.

Copy number variations (CNV) have also been identified throughout the genome using OvineSNP50 genotype data (Liu *et al*. 2013). The potential to impute genotypes from low density SNP genotype data to a high density SNP genotype panel has been proven and may enable the use of lower density SNP genotyping panels; thereby lower genotyping costs (Hayes *et al*. 2012).

1.6.2. Illumina genotyping microarrays:

Illumina produces SNP genotyping microarrays with probes for standard species-specific panels as well as custom designed panels. Most of the species-specific genotyping is based on the Infinium BeadChip assay which enables unlimited multiplexing and whole-genome genotyping without the need for polymerase chain reaction (PCR) (Gunderson *et al*. 2005). The single tube array-based genotyping system is considerably more time- and cost-effective than traditional PCR genotyping or earlier attempts at array-based genotyping (Morin *et al*. 2004, Gunderson *et al*. 2005, Steemers *et al*. 2006). This microarray technology is centred on 3-micron silica beads that are evenly spaced every 5.7 microns within microwells on fiber optic bundles or planar silica slides. The beads are covered with 50-mer oligonucleotide probes that act as a hybridisation sequences for specific marker loci. Marker allele detection is a two-step process where probes are selectively bound to marker loci, followed by either allele-specific primer extension (Infinium I) or single-base extension (Infinium II) of a labelled nucleotide that facilitates a colour reaction (Fan *et al.* 2010, Illumina 2014). Detection of the labelled nucleotide is enabled by dual-colour

fluorescent staining. Imaging technology (HiScan/iScan) is used to identify the colour and intensity of signals at individual loci. Red or green colour signals indicate a homozygote marker locus, while red/green signals designate heterozygotes. The protocol involves an initial amplification of sample DNA followed by fragmentation and then hybridisation to the microarray. Once sample DNA has hybridised to loci-specific oligonucleotides, enzymatic allele-specific primer extension facilitates a colour reaction. The colour reaction is detected by imaging technology which enables the calling of genotypes (Gunderson *et al.* 2005, Steemers *et al.* 2006).

1.6.3. OvineSNP50 BeadChip:

Illumina manufactures several livestock genotyping arrays, including the OvineSNP50 BeadChip that became commercially available in 2009. The chip was developed through collaboration between Illumina and the ISGC comprising researchers from several countries including New Zealand, Australia, the United States of America (USA), China, Spain, Austria, France, Finland, Germany, Italy, Kenya, Iran and Israel (International Sheep Genomics Consortium *et al.* 2010). The SNPs included in the panel were identified in three separate discovery attempts; BAC end Sanger sequencing, Roche 454 whole-genome shotgun sequencing and reduced representational sequencing (RRS). More than 70% of the SNPs included on the panel were identified in Roche 454 sequencing data of six individuals from six different breeds. This sequencing data was also used in creating the draft *ovine* genome sequence. BAC end Sanger sequencing of nine individuals, each from a different breed, yielded approximately 1% of the SNPs included in the chip. The remaining SNPs included in the chip design were identified through the use of RRS data of 60 individuals from 15 different breeds. Seven breeds and 23 individuals from African and Asian origin, and 19 breeds and 47 individuals from American and European breeds were used for SNP discovery (Kijas *et al*. 2012). Therefore, a bias towards the better represented American and European breeds may be observed in SNP genotyping results. The chip has nonetheless been validated in more than 75 international sheep breeds and has been successfully used for a wide range of applications (Zhao *et al*. 2011, Kijas *et al*. 2012, Zhang *et al*. 2012, Kemper *et al.* 2012, Demars *et al*. 2013, Heaton *et al.* 2013, Våge *et al.* 2013, McRae *et al*. 2014, Phua *et al*. 2014).

1.7. Climate change and adaptability:

The effect climate change will have on agriculture is uncertain at this stage due to a lack of data pertaining to the phenomenon as well as the difficulty in quantifying the myriad of potential indirect effects. An increase in temperature may be a direct effect of climate change and could influence animal health, production and reproduction (Nardone *et al*. 2010). Some possible indirect effects include water scarcity, soil infertility, reductions in grain yields and quality as well as changes in the frequency and distribution of pest species. Pasture-based extensive livestock farming will be affected more than intensive systems owing to the increased risk of solar radiation and high temperatures on animal health and the potential increase in the frequency of drought (Nardone *et al.* 2010). In Africa and Asia, where extensive pasture-based farming is prevalent, a 25% loss in animal production is predicted (Sequin 2008). Adaptability is essential in dealing with the potential challenges of climate change while ensuring the sustainability of livestock production (Nardone *et al.* 2010). The ability of an animal to adapt to changing and/or challenging environmental conditions while maintaining a high level of production is referred to as robustness (Knapp 2005, Star *et al*. 2008). The robustness of livestock animals influences not only production, but also the health and welfare of individual animals. Animals that can adapt and thrive despite external stressors will be less susceptible to health disorders or behavioural issues associated with climate change and may have a greater production potential (Star *et al*. 2008). The overall robustness of livestock can therefore be evaluated by examining certain health, reproduction and longevity traits (Klopčič & Kuipers 2009, Strandberg 2009).

1.8. Reproduction traits:

In animal production systems, reproductive traits are especially important due to the significance of these traits in production efficiency and their role in the general robustness of an animal (Klopčič & Kuipers 2009, Strandberg 2009). Reproductive traits are essential for efficient and profitable sheep production, even more so than growth traits (Purvis & Hillard 1997, Olivier 1999). Increases in the number of lambs born and raised per ewe increases the number of lambs available for slaughter as well as the number of animals available for selection and therefore results in an increase in genetic gain (Olivier 1999, De Graaf 2010). Reproduction traits are also vital in the maintenance of small flocks or flocks maintained under adverse conditions, as is often

found in resource poor areas (Purvis & Hillard 1997). The number of ewes that lamb after being mated (fertility); the number of lambs born per ewe lambed (litter size or fecundity) and the survival rate of lambs underwrite reproductive efficiency (De Graaf 2010). These traits are components of the composite trait, net-reproduction rate, defined as number of lambs weaned or weight of lamb weaned per lambing opportunity (Olivier 1999, Snowder & Fogarty 2009). Several additional factors also contribute to a ewe's reproductive characteristics including age of puberty, fertility, seasonality of reproduction, ovulation rate, embryonic mortality, mothering ability and lamb survival. Male measures of reproductive ability lie with traits such as testicular size, spermatogenesis and libido. In the current study, the focus will be on female reproductive traits as these traits have been shown to have a greater impact on the economic viability of sheep farming than male reproductive characteristics (Olivier 1999).

Ovulation rate is the foundation of a ewe's reproductive potential as it determines fertility, maximum conception rate and therefore also litter size. Ovulation is a complex trait that is influenced by genetic as well as environmental factors and there is great variation in this trait among sheep breeds (Montgomery *et al*. 2001). Mutations in three major genes, the *bone morphogenetic protein receptor 1B* (*BMPR-1B*), *bone morphogenetic protein 15* (*BMP15*) and growth *differentiation factor 9* (*GDF9*), have been linked to variation in ovulation rate (Notter 2008). Mutations in these three genes results in major changes to the ovulation rate of female animals carrying the mutation, however, several highly proliferative lines do not carry mutations in these genes (Davis *et al*. 2006, Juengel *et al*. 2012). As ovulation rate is a complex trait, many (QTL) may be responsible for variation in this trait (Andersson & Georges 2004). Several QTL, each contributing to a small degree of the variation in litter size, may underlie reproductive performance of highly proliferative lines that do not carry mutations in the aforementioned genes. It is therefore not surprising that putative QTL of reproductive traits are more challenging to elucidate and require denser marker panels and larger sample cohorts in comparison to the identification of a single causative mutation (Andersson & Georges 2004, Goddard & Hayes 2009).

Reproduction traits are notoriously difficult to quantify as specialised equipment and trained staff are necessary to measure traits such as conception rate and embryo

survival. Reproduction traits are also sex-limited and only measurable after participation in reproductive opportunities (Notter 2012). A further confounding factor is that most reproductive traits show low heritability and environmental factors contribute greatly to variation in these traits (Table 1.3). Selective breeding programmes rely on accurate trait recording for the selection of genetically superior breeding stock. All the aforementioned factors contribute to the difficulty in quantifying reproduction traits at the phenotypic and genetic level and consequently result in relatively slow genetic gains of these traits.

Table 1.3: The weighted mean heritability for reproduction traits in sheep as reported in literature.

Trait	Number of studies	Heritability±SE
Lambs weaned/ewe joined	11	0.07±0.01
Lambs born/ewe joined	19	0.10±0.01
Weight weaned/ewe joined	7	0.13±0.03
Lambs weaned/ewe lambing	8	0.05±0.01
Lambs born/ewe lambing	49	0.13±0.01
Lambs born alive/ewe lambing	2	0.10±0.05
Weight weaned/ewe lambing	11	0.11±0.02
Ewes lambing/ewe joined (fertility)	18	0.08±0.01
Ewe rearing ability	7	0.06±0.02
Lamb survival	16	0.03±0.01
Embryo survival	4	0.01±0.01

SE: Standard error. Adapted from Safari *et al.* (2005)

In comparison with quantitative trait loci (QTL) reported for production traits, such as meat and health traits, very few QTL relating to reproduction traits have been reported for sheep (Zhang *et al.* 2012, Hu *et al*. 2013). The QTL relating to aspects of reproduction have been identified through QTL interval-mapping studies (Mateescu & Thonney 2010) as well as association analyses of SNPs located in candidate genes and reproduction traits (Table 1.4).

Table 1.4: QTL affecting reproduction in sheep reported in literature.

Chromosome	Number of QTL	Trait (number of QTL)	Reference
1	2	Aseasonal reproduction (2)	Mateescu & Thonney 2010
3	7	Aseasonal reproduction (2), Stillbirths (4), Total lambs born (1)	Mateescu & Thonney 2010, Darlay *et al.* 2011, Chu *et al*. 2012
5	6	Ovulation rate (1), Total lambs born (5)	Silva *et al*. 2011, Chu *et al.* 2004, 2011a, Javanmard *et al*. 2011
6	1	Total lambs born (1)	Chu *et al.* 2011b
11	1	Total lambs born (1)	Javanmard *et al.* 2011
12	1	Aseasonal reproduction (1)	Mateescu & Thonney 2010
16	1	Aseasonal reproduction (1)	Mateescu & Thonney 2010
17	4	Aseasonal reproduction (1)	Mateescu & Thonney 2010
19	2	Aseasonal reproduction (1)	Mateescu & Thonney 2010
20	2	Aseasonal reproduction (1)	Mateescu & Thonney 2010
24	1	Aseasonal reproduction (1)	Mateescu & Thonney 2010
26	2	Onset of fertility (1), Interval between first and second lambing (1)	Mateescu & Thonney 2010
X	1	Total lambs born (1)	Shabir *et al.* 2013

Taken from www.animalgenome.org/cgi-bin/QTLdb/OA/index (Hu *et al.* 2013)

1.9. Aims and objectives:

The projected increase in the global food demand and the potential threat of climate change necessitates the implementation of efficient and sustainable agricultural production. The South African sheep industry plays an important role in food security, utilisation of land mass that is unsuitable for intensive crop production and also in the livelihood of agricultural communities occupying areas of low cropping potential. The market demand for an increase in South African sheep meat production is evident in the annual net import of approximately 15 000 tonnes of sheep meat. In order to increase the profitability, global competitiveness, yield and efficiency of the sheep

industry, the reproductive potential of the South African national sheep flock needs to be improved. Due to the challenges involved with quantifying reproduction traits as well as the substantial impact from environmental factors, progress towards improving these traits has been relatively slow. An approach utilising advanced genotyping techniques may provide the opportunity to enhance the genetic gains in reproduction of South African sheep.

The aim of the current study was to assess the utility of a whole-genome SNP genotyping strategy to identify genomic regions associated with greater reproductive success and other robustness traits by exploiting the divergent selection applied to the Elsenburg Merino flock. As divergent selection has been applied to this flock for several generations, it served as an ideal starting point to identify SNPs associated with genomic regions under selection as well as determine putative QTL and genes linked to reproduction traits.

The Elsenburg Merino flock is characterised in Chapter 2 and the selection history and subsequent change in mean reproduction traits are examined. Genotyping results of four South African sheep breeds are assessed to determine the utility of the OvineSNP50 genotyping array for South African sheep (Chapter 3). Genotype data from the Elsenburg Merino flock is used to identify individual SNP loci and genomic regions subject to selection by means of F_{ST} outlier methods (Chapter 4) and by identifying overlapping homozygous region characteristic of selection sweep (Chapter 5). The main findings, possible implications, potential limitations of the study and future research are summarised in the final chapter (Chapter 6).

1.10. References:

AGIS (2007) Agricultural Geo-Referenced Information System. Accessed May 2014, from: www.agis.agric.za

Amaral AJ, Ferretti L, Megens H-J, Crooijmans RPMA, Nie H, Ramos-Onsins SE, Perez-Enciso M, Schook LB, Groenen MAM (2011) Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. Public Library of Science ONE 6: e14782

Andersson L, Georges M (2004) Domestic-animal genomics: Deciphering the genetics of complex traits. Nature Reviews Genetics 5: 202-212

Bai Y, Sartor M, Cavalcoli J (2012) Current status and future perspectives for sequencing livestock genomes. Journal of Animal Science and Biotechnology 3: 8

Barendse W, Reverter A, Bunch RJ, Harrison BE, Barris W, Thomas MB (2007) A validated whole-genome association study of efficient food conversion in cattle. Genetics 176: 1893-1905

Bolormaa S, Neto LR, Zhang YD, Bunch RJ, Harrison BE, Goddard ME, Barendse W (2011) A genome-wide association study of meat and carcass traits in Australian cattle. Journal of Animal Science 89: 2297-2309

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. Trends in Ecology and Evolution 18: 249-256

Bunch TD, Wu C, Zhang Y-P, Wang S (2006) Phylogenetic analysis of snow sheep (*Ovis nivicola*) and closely related taxa. Journal of Heredity 97: 21-30

Burger A, Hoffman LC, Cloete JJE, Muller M, Cloete SWP (2013) Carcass composition of Namaqua Afrikaner, Dorper and SA Mutton Merino ram lambs reared under extensive conditions. South African Journal of Animal Science 43: S27-S32

Chessa B, Pereira F, Arnaud F, Amorim A, Goyache F, Mainland I, Kao RR, Pemberton JM, Beraldi D, Stear MJ, Alberti A, Pittau M, Iannuzzi L, Banabazi MH, Kazwala RR, Zhang YP, Arranz JJ, Ali BA, Wang Z, Uzun M, Dione MM, Olsaker I, Holm LE, Saarma U, Ahmad S, Marzanov N, Eythorsdottir E, Holland MJ, Ajmone-Marsan P, Bruford MW, Kantanen J, Spencer, Palmarini M (2009) Revealing the history of sheep domestication using retrovirus integrations. Science 324: 532-536

Chu MX, Li BX, Wang JY, Ye SC, Fang L (2004) Association between PCR-SSCP of *growth differentiation factor 9* gene and high prolificacy in Small Tail Han sheep. Animal Biotechnology 15: 111-120

Chu MX, Yang J, Feng T, Cao GL, Fang L, Di R, Huang DW, Tang QQ, Ma YH, Li K, Li N (2011a) *GDF9* as a candidate gene for prolificacy of Small Tail Han sheep. Molecular Biology Reports 38: 5199-5204

Chu MX, Jia L, Zhang Y, Jin M, Chen H, Fang L, Di R, Cao G, Feng T, Tang Q, Ma Y, Li K (2011b) Polymorphisms of coding region of *BMPR-IB* gene and their relationship with litter size in sheep. Molecular Biology Reports 38: 4071-4076

Chu MX, Guo XH, Feng CJ, Li Y, Huang DW, Feng T, Cao GL, Fang L, Di R, Tang QQ, Ma YH, Li K (2012) Polymorphism of 5' regulatory region of ovine *FSHR* gene and its association with litter size in Small Tail Han sheep. Molecular Biology Reports 39: 3721-3725

Cloete JJE, Cloete SWP, Olivier JJ, Hoffman LC (2007a) Terminal crossbreeding of Dorper ewes to Ile de France, Merino Landsheep and SA Mutton Merino sires: Ewe production and lamb performance. Small Ruminant Research 69: 28-35

Cloete JJE, Cloete SWP, Hoffman LC (2010) Behaviour of Merinos divergently selected for multiple rearing ability in response to external stimuli. Proceedings of the World Congress on Genetics Applied to Livestock Production, August 2010, Leipzig, Germany

Cloete JJE, Cloete SWP, Scholtz AJ, Hoffman LC (2013a) Behaviour response of Namaqua Afrikaner, Dorper and SA Mutton Merino lambs towards humans. South African Journal of Animal Science 43: S116-S120

Cloete JJE, Cloete SWP, Scholtz AJ, Matthee S (2013b) The effect of breed, ewe age and season on tick counts of indigenous and commercial sheep in South Africa. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 20: 187-190

Cloete SWP, Van Wyk JB, Scholtz AJ, Gilmour AR (2003) Genetic and environmental parameters for lambing behaviour in Merino lines divergently selected for ewe maternal rearing ability. Proceedings of the New Zealand Society of Animal Production 63: 169-172

Cloete SWP, Gilmour AR, Olivier JJ, Van Wyk JB (2004) Genetic and phenotypic trends and parameters in reproduction, greasy fleece weight and live weight in Merino lines divergently selected for multiple rearing ability. Australian Journal of Experimental Agriculture 44: 745-754

Cloete SWP, Olivier JJ, Van Wyk JB, Schoeman SJ, Erasmus GJ (2005a) Genetic parameters and trends for hogget traits in Merino lines divergently selected for multiple rearing ability. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 16: 24-27

Cloete SWP, Scholtz AJ, Cloete JJE, Van Wyk JB (2005b) The ability of Merino ewes and lambs to reunite after separation, as affected by divergent selection for ewe multiple rearing capacity. Australian Journal of Experimental Agriculture 45: 1131-1137

Cloete SWP, Olivier JJ, Olivier WJ (2007b) Genetic change in South African Merino resource flocks. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 17: 320-323

Cloete SWP, Misztal I, Olivier JJ (2009) Genetic parameters and trends for lamb survival and birth weight in a Merino flock divergently selected for multiple rearing ability. Journal of Animal Science 87: 2196-2208

Cloete SWP, Olivier JJ (2010) South African sheep and wool industry. In: Cottle DJ (ed.) The international sheep and wool handbook, pg. 95-112. Nottingham University press, Nottingham, England

Cottle DJ (2010) World sheep and wool production. In: Cottle DJ (ed.) The international sheep and wool handbook, pg. 1-48. Nottingham University press, Nottingham, England

DAD-IS (Domestic Animal Diversity Information System) (2014) Food and Agriculture Organization of the United Nations. Accessed May 2014, from: http://dad.fao.org/

Daetwyler HD, Hickey JM, Henshall JM, Dominik S, Gredler B, Van der Werf JHJ, Hayes BJ (2010) Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. Animal Production Science 50: 1004-1010

DAFF (Department of Agriculture, Forestry and Fisheries) (2013) Abstract of agricultural statistics. Compiled by directorate statistics and economic analysis. Published in the Republic of South Africa by Department of Agriculture, Forestry and Fisheries. Accessed May 2014, from: http://www.nda.agric.za/docs/statsinfo/Abstact2013.pdf

Darlay RJ, McCarthy AJ, Illot NE, Smith JE, Shaw MA (2011) Novel polymorphisms in ovine immune response genes and their association with abortion. Animal Genetics 42: 535-543

Davis GH, Balakrishnan L, Ross IK, Wilson T, Galloway SM, Lumsden BM, Hanrahan JP, Mullen M, Mao XZ, Wang GL, Zhao ZS, Zeng YQ, Robinson JJ, Mavrogenis AP, Papachristoforou C, Peter C, Baumung R, Cardyn P, Boujenane I, Cockett NE, Eythorsdottir R, Arranz JJ, Notter DR (2006) Investigation of the Booroola (FecB) and Inverdale (FexXI) mutations in 21 prolific breeds and strains of sheep sampled in 13 countries. Animal Reproduction Science 92: 87-96

De Graaf SP (2010) Reproduction. In: Cottle DJ (ed.) The international sheep and wool handbook, pg. 189-222. Nottingham University press, Nottingham. England

Deacon HJ, Deacon J (1999) Human beginnings in South Africa: uncovering the secrets of the Stone Age, pg. 181. Rowman Altamira, Walnut Creek, California, USA

Decker JE, McKay SD, Rolf MM, Kim J, Molina Alcalá A, Sonstegard TS, Hanotte O, Götherström A, Seabury CM, Praharani L, Babar ME, Correia de Almeida Regitano L, Yildiz MA, Heaton MP, Liu W-S, Lei C-Z, Reecy JM, Saif-Ur-Rehman M, Schnabel RD, Taylor JF (2014) Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. Public Library of Science Genetics 10: e1004254

Demars J, Fabre S, Sarry J, Rossetti R, Gilbert H, Persani L, Tosser-Klopp G, Mulsant P, Nowak Z, Drobik W, Martyniuk E, Bodin L (2013) Genome-wide association studies identify two novel *BMP15* mutations responsible for an atypical hyperprolificacy phenotype in sheep. Public Library of Science Genetics 9: e1003482

Diez-Tascón C, Littlejohn RP, Almeida PAR, Crawford AM (2000) Genetic variation within the Merino sheep breed: Analysis of closely related populations using microsatellites. Animal Genetics 31: 243-251

Fan B, Du Z-Q, Gorbach DM, Rothchild MF (2010) Development and application of high-density SNP arrays in genomic studies of domestic animals. Asian-Australian Journal of Animal Sciences 23: 833-847

FAO (Food and Agriculture Organisation of the United Nations) (2000) World watch list for domestic animal diversity, 3rd edn, Scherf B (ed.), Rome

FAO (Food and Agriculture Organisation of the United Nations) (2013) FAO Statistical yearbook: World food and agriculture. ISBN 978-92-5-107396-4

Farm Animal Conservation Trust (2001) Ramsay K, Harris L, Kotze A (eds.) Landrace breeds: South Africa's indigenous and locally developed farm animals. ISBN 0-620-25493-9

Fedosenko A, Blank D (2005) *Ovis ammon*. Mammalian Species 773: 1-15

Franklin IR (1997) Systematics and phylogeny. In: Piper L, Ruvinsky A (eds.) The genetics of sheep, pg. 1-12. CAB International, Oxon, UK

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Reviews Genetics 10: 381-391

Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS (2005) A genome-wide scalable SNP genotyping assay using microarray technology. Nature Genetics 37: 549-554

Gutiérrez-Gil B, Arranz JJ, Pong-Wong R, García-Gámez E, Kijas J, Wiener P (2014) Application of selection mapping to identify genomic regions associated with dairy production in sheep. Public Library of Science ONE: e94623

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

Hayes BJ, Bowman PJ, Daetwyler HD, Kijas JW, Van der Werf JHJ (2012) Accuracy of genotype imputation in sheep breeds. Animal Genetics 43: 72-80

Heaton MP, Kalbfleisch TS, Petrik DT, Simpson B, Kijas JW, Clawson ML, Chitko-McKown CG, Harhay GP, Leymaster KA, the International Sheep Genomics Consortium (2013) Genetic testing for *TMEM154* mutations associated with lentivirus susceptibility in sheep. Public Library of Science ONE 8: e55490

Heaton MP, Leymaster KA, Kalbfleisch TS, Kijas JW, Clarke SM, McEwan J, Maddox JF, Basnayake V, Petrik DT, Simpson B, Smith TPL, Chitko-McKown CG and the International Sheep Genomics Consortium (2014) SNPs for parentage testing and traceability in globally diverse breeds of sheep. Public Library of Science ONE: e94851

Heydenrych HJ, Du Plessis JJ, Cloete SWP (1984) Increasing the wool production of Merino sheep in the South Western districts of South Africa by direct and indirect selection. In: Hofmeyr JH, Meyer EHH (eds.) Proceedings of the 2nd World Congress on Sheep and Beef Cattle Breeding 16–19 April, Pretoria. South African Stud Book and Livestock Improvement Association, Bloemfontein, pg. 399-412

Hiendleder S, Kaupe B, Wassmuth R, Janke A (2002) Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. Proceedings of the Royal Society: Biological Sciences 269: 893-904

Hough D, Storbeck K, Cloete SWP, Swart P (2010) Improving stress management: A study of the *CYP17* genotype in *Ovis aries* and *Capra hircus*. Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, August 2010, Leipzig, Germany

Hu X, Gao Y, Feng C, Liu Q, Wang X, Du Z, Wang Q, Li N (2009) Advanced technologies for genomic analysis in farm animals and its application for QTL mapping. Genetica 136: 371-386

Hu Z-L, Park CA, Wu X-L, Reecy JM (2013). Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Research 41: D871-D879

Illumina (2014) Infinium bead micrarry. Acceseed May 2014, from: www.illumina.com/technology/beadarray-technology/infinium-hd-assay.ilmn

International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432: 695-716

International Chicken Polymorphism Map Consortium (2004) A Genetic variation map for chicken with 2.8 million single nucleotide polymorphism. Nature 432: 717-722

International Sheep Genomics Consortium, Archibald AL, Cockett NE, Dalrymple BP, Faraut T, Kijas JW, Maddox JF, McEwan JC, Hutton Oddy V, Raadsma HW, Wade C, Wang J, Wang W, Xun X (2010) The sheep genome reference sequence: a work in progress. Animal Genetics 41: 449-453

Javanmard A, Azadzadeh N, Esmailizadeh AK (2011) Mutations in *bone morphogenetic protein 15* and *growth differentiation factor 9* genes are associated with increased litter size in fat-tailed sheep breeds. Veterinary Research Communications 35: 157-167

Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, Wu C, Muzny DM, Li Y, Zhang W, Stanton J-A, Brauning R, Barris WC, Hourlier T, Aken BL, Searle SMJ, Adelson DL, Bian C, Cam GR, Chen Y, Cheng S, DeSilva U, Dixen K, Dong Y, Fan G, Franklin IR, Fu S, Fuentis-Utrilla P, Guan R, Highland MA, Holder ME, Huang G, Ingham AB, Jhangiani SN, Kalra D, Kovar CL, Lee SL, Liu, W, Liu X, Lu C, Lv T, Mathew T, McWilliam S, Menzies M, Pan S, Robelin D, Servin B, Townley D, Wang W, Wei B, White SN, Yang X, Ye C, Yue Y, Zeng P, Zhou Q, Hansen JB, Kristiansen K, Gibbs RA, Flicek P, Warkup CC, Jones HE, Oddy VH, Nicholas FW, McEwan JC, Kijas JW, Wang Y, Worley KC, Archibald AL, Cockett N, Xu X, Wang W, Dalrymple BP (2014) The sheep genome illuminates biology of the rumen and lipid metabolism. Science 344: 1168-1173

Juengel JL, Connell AR, French MC, Proctor LE, Wheeler R, Farquhar PA, Dodds KG, Galloway SM, Johnstone PD, Davis GH (2012) Identification of a line of sheep carrying a putative autosomal gene increasing ovulation rate in sheep that does not appear to interact with mutations in the transforming growth factor beta superfamily. Biology of Reproduction 85: 113-120

Kemper KE, Daetwyler HD, Visscher PM, Goddard ME (2012) Comparing linkage and association analyses in sheep points to a better way of doing GWAS. Genetics Research 94: 191-203

Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, Tang D, McEwan J, Cockett N, Oddy HV, Nicholas FW, Raadsma H, for the International Sheep Genomics Consortium (2009) A genome-wide survey of SNP variation reveals the genetic structure of sheep breeds. Public Library of Science ONE 4: e4668

Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J, Dalrymple B, other members of the International Sheep Genomics Consortium (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. Public Library of Science Biology 10: e1001258

Kijas JW, Porto-Neto L, Dominik S, Reverter A, Bunch R, McCulloch R, Hayes BJ, Brauning R, McEwan J, the International Sheep Genomics Consortium (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. Animal Genetics 45: 754-757

Klopčič M, Kuipers A (2009) Perception of robustness traits in breeding goals for dairy cattle in a new EU country. In: Klopčič M, Reents R, Philipsson J, Kuipers A (eds.) Breeding for robustness. EAAP publication 126: 249-262

Knapp PW (2005) Breeding robust pigs. Australian Journal of Experimental Agriculture 45: 763-773

Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillén S, Vilà C (2002) Ancient DNA evidence for Old World origin of New World dogs. Science 298:1613-1616

Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Zhang S, Zhou X, Wei C, Zhao F, Du L (2013) Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. BioMed Central Genomics 14: 229

MacEachern S, Hayes B, McEwan J, Goddard M (2009) An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bos taurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in domestic cattle. BioMed Central Genomics 10: 181

Maddox JF, Cockett NE (2007) An update on sheep and goat linkage maps and other genomic resources. Small Ruminant Research 70: 4-20

Maijala K (1997) Systematics and phylogeny. In: Piper L, Ruvinsky A (eds.) The genetics of sheep, pg. 13-49. CAB International, Oxon, UK

Mateescu RG, Thonney ML (2010) Genetic mapping of quantitative trait loci for aseasonal reproduction in sheep. Animal Genetics 41: 454-459

McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, Crews D, Dias Neto E, Gill CA, Gao C, Mannen H, Stothard P, Wang Z, Van Tassell CP, Williams JL, Taylor JF, Moore SS (2007) Whole genome linkage disequilibrium maps in cattle. BioMed Central Genetics 8: 74

McKee WM (1913) South African sheep and wool. William Clowes & Sons Ltd., London

McRae KM, McEwan JC, Dodds KG Gemmell NJ (2014) Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. BioMed Central Genomics 15: 637

McTavish EJ, Decker JE, Schnabel RD, Taylor JF, Hillis DH (2013) New World cattle show ancestry from multiple independent domestication events. Proceedings of the National Academy of Sciences 110: E1398-E1406

Meadows JRS, Cemal I, Karaca O, Gootwine E, Kijas JW (2007) Five ovine mitochondrial lineages identified from sheep breeds of the Near East. Genetics 175: 1371-1379

Miller JM, Poissant J, Kijas JW, Coltman DW, the International Sheep Genomics Consortium (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. Molecular Ecology Resources 11: 314-322

Milne C (2000) The history of the Dorper sheep. Small Ruminant Research 36 99-102

Moioli B, Scatà MC, Steri R, Napolitano F, Catillo G (2013) Signatures of selection identify loci associated with milk yield in sheep. BioMed Central Genetics 14: 76

Montgomery GW, Galloway SM, Davis GH, McNatty KP (2001) Genes controlling ovulation rate in sheep. Reproduction 121: 843-852

Morin PA, Luikart G, Wayne RK, SNP workshop group (2004) SNPs in ecology, evolution and conservation. Trends in Ecology and Evolution 19: 208-216

Mullen MP, McClure MC, Kearney JF, Waters SM, Weld R, Flynn P, Creevey CJ, Cromie AR, Berry DP (2013) Development of a custom SNP chip for dairy and beef cattle breeding, parentage and research. Interbull Bulletin 47: 58-66

Naderi S, Rezaei H-R, Pompanon F, Blum MGB, Negrini R, Naghash H-R, Balkız Ö, Mashkour M, Gaggiotti OE, Ajmone-Marsan P, Kence A, Vigne J-D, Taberlet P (2008) The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. Proceedings of the National Academy of Sciences 105: 17659-17664

Nardone A, Ronchi B, Lacetera N, Ranieri MS, Bernabucci U (2010) Effects of climate changes on animal production and sustainability of livestock systems. Livestock Science 130: 57-69

Notter DR (2008) Genetic aspects of reproduction in sheep. Reproduction of Domestic Animals 43: 122-128

Notter DR (2012) Genetic improvement of reproductive efficiency of sheep and goats. Animal Reproduction Science 130: 147-151

Olivier JJ (1989) Genetic and environmental trends in the Grootfontein Merino stud. PhD dissertation, University of the Free State, South Africa

Olivier JJ, Erasmus GJ, Van Wyk JB, Konstantinov KV (1995) Response to selection on BLUP of breeding values in the Grootfontein Merino stud. South African Journal of Animal Science 25: 13-15

Olivier JJ (1999) The South African Merino performance testing scheme. In: Rising to the challenge – Breeding for the 21st century customer. Beef Industry and CRC for Premium Quality Wool Industry Symposia. Supplement to the Proceedings of the Association for the Advancement of Animal Breeding and Genetics 13: 119-124

Olivier WJ (2014) The evaluation of a South African fine wool genetic resource flock. PhD dissertation, Stellenbosch University, South Africa

Parrini F, Cain JW, Krausman PR (2009) *Capra ibex* (Artiodactyla: Bovidae). Mammalian Species 830: 1-12

Pedrosa S, Uzun M, Arranz J-J, Gutiérrez-Gil B, San Primitivo F, Bayón Y (2005) Evidence of three maternal lineages in near eastern sheep supporting multiple domestication events. Proceedings of the Royal Society: Biological Sciences 272: 2211-2217

Perkel J (2008) SNP genotyping: six technologies that keyed a revolution. Nature Methods 5: 447-453

Phua SH, Brauning R, Baird HJ, Dodds KG (2014) Identifying chromosomal selection-sweep regions in facial eczema selection-line animals using an ovine 50K-SNP array. Animal Genetics 45: 240-247

Purvis IW, Hillard M (1997) Biology and genetics of reproduction. In: Piper L, Ruvinsky A (eds.) The genetics of sheep, pg. 375-394. CAB International, Oxon, UK

Qwabe SO, Van Marle-Köster E, Visser C (2013) Genetic diversity and population structure of the endangered Namaqua Afrikaner sheep. Tropical Animal Health and Production 45: 511-516

Rezaei HR, Naderi S, Chintauan-Marquier IC, Taberlet P, Virk AT, Naghash HR, Rioux D, Kaboli M, Pompanon F (2010) Evolution and taxonomy of the wild species of the genus *Ovis* (Mammalia, Artiodactyla, Bovidae). Molecular Phylogenetics and Evolution 54: 315-326

Rowe JB; Gill S, Banks RG; Van der Werf JHJ (2013) Genomics for the Australian sheep industry: from design to delivery. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 20: 14-17

Safari E, Fogarty NM, Gilmour AR (2005) A review of genetic parameter estimates for wool, growth, meat and reproduction traits in sheep. Livestock Production Science 92: 271-289

Schoeman SJ, Cloete SWP, Olivier JJ (2010) Returns on investment in sheep and goat breeding in South Africa. Livestock Science 130: 70-82

Scholtz AJ, Cloete SWP, Van Wyk JB, Kruger ACM, Van der Linde TCD (2010) Influence of divergent selection for reproduction on the occurrence of breech strike in mature Merino ewes. Animal Production Science 50: 203-209

Schopen GC, Visker MH, Koks PD, Mullaart E, Van Arendonk JA, Bovenhuis H (2011) Whole-genome association study for milk protein composition in dairy cattle. Journal of Dairy Science 94: 3148-3158

Seguin B (2008) The consequences of global warming for agriculture and food production. In: Rowlinson P, Steele M, Nefzaoui A (eds.). Proceedings of the International Conference on Livestock and Global Climate Change: 9-11

Shabir M, Ganai TA, Misra SS, Shah R, Ahmad T (2013) Polymorphism study of *growth differentiation factor 9B (GDF9B)* gene and its association with reproductive traits in sheep. Gene 515: 432-438

Silva BD, Castro EA, Souza CJ, Paiva SR, Sartori R, Franco MM, Azevedo HC, Silva TA, Vieira AM, Neves JP, Melo EO (2011) A new polymorphism in the *Growth and Differentiation Factor 9 (GDF9)* gene is associated with increased ovulation rate and prolificacy in homozygous sheep. Animal Genetics 42: 89-92

Snelling WM, Allan MF, Keele JW, Kuehn LA, McDaneld T, Smith TPL, Sonstegard TS, Thallman RM, Bennett GL (2010) Genome-wide association study of growth in crossbred beef cattle. Journal of Animal Science 88: 837-848

Snowder GD, Fogarty NM (2009) Composite trait selection to improve reproduction and ewe productivity: A review. Animal Production Science 49: 9-16

Snyman MA, Olivier JJ, Cloete JAN (1993) Productive and reproductive performance of Namaqua Afrikaner sheep. Karoo Agric 5: 21-24

Snyman MA, Van Marle-Köster E, Qwabe SO, Visser C (2013) Genetic and phenotypic profile of three South African Namaqua Afrikaner sheep flocks. Grootfontein Agric 13 (1). Accessed May 2014, from: http://gadi.agric.za/Agric/Vol13No1_2013/namaqua.php

Soma P, Kotze A, Grobler JP, Van Wyk JB (2012) South African sheep breeds: population genetic structure and conservation implications. Small Ruminant Research 103: 112-119

Star L, Ellen ED, Uitdehaag K, Brom WA (2008) A plea to implement robustness into a breeding goal: Poultry as an example. Journal of Agriculture and Environmental Ethics 21: 109-125

Steemers FJ, Chang W, Lee G, Barker DL, Shen R, Gunderson KL (2006) Whole-genome genotyping with the single-base extension assay. Nature Methods 3: 31-33

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

Strandberg E (2009) The role of environmental sensitivity and plasticity in breeding for robustness: Lessons from evolutionary genetics. In: Klopčič M, Reents R, Philipsson J, Kuipers A (eds.) Breeding for robustness. EAAP publication 126: 17-34

The Bovine HapMap Consortium (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science: 324: 528-532

Våge DI, Husdal M, Kent MP, Klemetsdal G, Boman IA (2013) A missense mutation in *growth differentiation factor 9 (GDF9)* is strongly associated with litter size in sheep. BioMed Central Genetics 14: 1

Van der Walt D, Cloete SWP, Storbeck K, Swart P (2009) The role of cytochrome P450 17-alpha-hydroxylase/17,20 lyase (CYP17) in the stress coping ability in a divergently selected Merino sheep population. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 18: 100-103

Van der Werf JHJ, Marshall K, Lee S (2007) Methods and experimental designs for detection of QTL in sheep and goats. Small Ruminant Research 70: 21-31

Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK (1997) Multiple and ancient origins of the domestic dog. Science 276: 1687-1689

Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear TL, Adelson DL, Bailey E, Bellone RR, Blöcker H, Distl O, Edgar RC, Garber M, Leeb T, Mauceli E, MacLeod JN, Penedo MC, Raison JM, Sharpe T, Vogel J, Andersson L, Antczak DF, Biagi T, Binns MM, Chowdhary BP, Coleman SJ, Della Valle G, Fryc S, Guérin G, Hasegawa T, Hill EW, Jurka J, Kiialainen A, Lindgren G, Liu J, Magnani E, Mickelson JR, Murray J, Nergadze SG, Onofrio R, Pedroni S, Piras MF, Raudsepp T, Rocchi M, Røed KH, Ryder OA, Searle S, Skow L, Swinburne JE, Syvänen AC, Tozaki T, Valberg SJ, Vaudin M, White JR, Zody MC; Broad Institute Genome Sequencing Platform; Broad Institute Whole Genome Assembly Team, Lander ES, Lindblad-Toh K (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 326: 865-867

Weller JI, Glick G, Zeron Y, Seroussi E, Ron M (2010) Paternity validation and estimation of genotyping error rate for the BovineSNP50 BeadChip. Animal Genetics 41: 551-553

Zhang H, Wang A, Wang S, Li H (2012) Progress of genome wide association study in domestic animals. Journal of Animal Science and Biotechnology 3: 26-36

Zhao X, Dittmer KE, Blair HT, Thompson KG, Rothschild MF, Garrick D (2011) A novel nonsense mutation in the *DMP1* gene identified by a genome-wide association study is responsible for inherited rickets in Corriedale sheep. Public Library of Science ONE 6: e21739

Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, Marcais G, Roberts M, Subramanian P, Yorke JA, Salzberg SL (2009) A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biology 10: R42

Zohary D, Tchernov E, Kolska Horwitz L (1998) The role of unconscious selection in the domestication of sheep and goats. Journal of Zoology 245: 129-135

Chapter 2:

Variation in reproduction traits of a divergently selected South African Merino flock

2.1. Introduction:

Quantitative traits typically exhibit a continuous distribution of measureable phenotypic values and may be influenced by environmental and genetic factors (Andersson 2001). In theory, quantitative traits are controlled by an infinite number of additive genes, each with an infinitesimal effect. However, studies have revealed that the effect of individual quantitative trait loci (QTL) is not necessarily equal and that some loci with a larger effect can be identified for a specific trait (Andersson & Georges 2004). Livestock provide a valuable resource for elucidating the QTL underlying complex traits due to the high level of phenotypic variation within and between breeds as well as the availability of additional information, such as breeding values and pedigrees (Andersson 2001, Dekkers & Hospital 2002, Andersson & Georges 2004). The presence of phenotypic variation not only facilitates the elucidation of QTL, but also forms the basis of any effective selective breeding programme. The presence of phenotypic variation is indicative of the inherent genetic variation that could be utilised to breed towards a specified breeding goal (Dekkers & Hospital 2002). By utilising trait and pedigree information as well as knowledge of the additive nature of most quantitative traits, the mean of the trait of interest can be shifted by employing a selective breeding programme. The success of such a breeding programme depends on the accuracy of the estimated breeding values for selection candidates, the selection intensity and the genetic variation contained in the population (Dekkers & Hospital 2002). Phenotypic variation, if controlled by genetic factors, is therefore an indicator of the potential of a population to undergo a change in mean trait values. Variation not only plays a role in the selection potential of a population, but also contributes to the adaptability of a population to changing environmental conditions. Populations with a higher degree of genetic variation are considered better equipped to adapt to changing conditions than less diverse populations (Notter *et al*. 1999, Hoffmann & Sgrò 2011).

The Elsenburg Merino flock has been divergently selected since the 1980's for the ability to rear multiple offspring. The base flock originated from a former selection experiment for wool quality at Tygerhoek Research Farm; the three wool selection lines being crossbred to form the base flock for the Elsenburg flock (Heydenrych *et al.* 1984, Schoeman *et al.* 2010). Initially, male and female breeding stock were selected according to maternal ranking values based on maternal phenotypes for number of lambs weaned per ewe joined. Breeding values derived from a single-trait repeatability model using best linear unbiased prediction (BLUP) estimates for number of lambs weaned per ewe per joining (NLW) were used to guide selection decisions from 2001 onwards. The selection processes used were detailed by Cloete *et al*. (2004, 2009) and Scholtz *et al*. (2010).

Divergent selection for multiple lamb rearing ability resulted in a 1.5% per year increase in the number of lambs weaned per ewe and a 1.8% increase in the total weight of lamb weaned per joining in the line selected for improved lamb rearing ability (High line). Conversely, a 0.8 to 1.0% decline in the number of lambs weaned and a 1.0 to 1.2% decline for total weight of lamb weaned per joining was noted in the line selected for a decrease in lamb rearing ability (Low line) (Cloete *et al*. 2004, 2007). The divergence in these traits also led to divergence in the weaning weight, lamb weight at one year and lamb survival (Cloete *et al*. 2003, 2005a, 2009). Furthermore, behavioural differences that influence lamb survival, have been recorded between the two lines (Cloete *et al.* 2003, 2005b). Wool quantity (clean fleece weight) and -quality (fibre diameter) of yearling hoggets were unaffected by selection for reproduction (Cloete *et al*. 2005a). In 2003, crossbreeding efforts involving High line and Low line individuals were initiated and the first crossbred individuals were able to reproduce during 2005. Initially, reciprocal crosses of High line and Low line individuals were carried out to produce crossbreds which were subsequently backcrossed to either High line or Low line sires. The result was crossbred individuals originating from either a cross between a High line dam and a Low line sire or High line sire and Low line dam; and backcross individuals with female parents from the crossbred lines and sires from either the High line or Low line.

By applying selection pressure to phenotypic characteristics such NLW, a decrease in the variation of these characteristics is expected. In accordance with the

aforementioned decrease in phenotypic variation, a decrease in genetic variation may also occur. It can therefore be assumed that the loci-specific degree of heterozygosity within each of the Elsenburg Merino lines will be lower than expected for an unselected base population. A general increase in homozygosity of certain loci may therefore have taken place due to ongoing selection (Johansson *et al*. 2010). It is assumed that individuals from the two lines are likely to be homozygous for disparate alleles at several loci and a cross of the two lines would result in all offspring being heterozygotes for all these loci and therefore exhibiting similar phenotypes and less variation around the population mean. Crossbreeding between lines is expected to result in individuals that exhibit less phenotypic variance than would be expected for an unselected base population. In theory, backcrossing the crossbreds to either of the foundation lines would result in offspring with a higher degree of phenotypic variation as this would result in heterozygous and homozygous offspring. In reality, most selected livestock lines are not homozygous for all loci and the aforementioned model of selection is complicated by the fact that most production traits are complex traits with several loci contributing to the trait variation (Dekkers & Hospital 2002). A further complication may be the contribution of loci under balancing selection that add to the favourable trait of interest.

By characterising the change in reproduction traits of the Elsenburg Merino flock, valuable information relating to the complexity of the reproduction trait inheritance as well as the future strategies needed to characterise QTL underlying these traits, may be gained. The aim of the current study was to characterise the change in the mean and variation of three reproduction traits during the selection experiment in the Elsenburg Merino flock. A further aim was to characterise the difference in mean and variance of reproduction traits between the lines, crosses and backcrosses in an attempt to gain more knowledge regarding the genetic architecture of putative QTL underlying the trait divergence and variance in this flock.

2.2. Materials and Methods:

At the onset of the selection experiment in the 1980's, breeding stock were selected according to maternal ranking values based on maternal phenotypes for number of lambs weaned per ewe joined. Progeny of ewes that reared more than one lamb per reproductive opportunity were preferred as replacements in the High line. Replacements for the Low line were mainly descended from ewes that reared one lamb per mating. From 20018, the maternal ranking values were augmented by EBVs for number of lambs weaned per ewe joined. The EBVs were derived from a single-trait repeatability model, as described by Cloete *et al.* (2004). During the course of the selection experiment progeny of ewes that reared one lamb per mating were occasionally accepted into either line. In recent years, all High line replacements have favourable EBVs and are the progeny of ewes that have reared more than one lamb per mating. Due to the unavailability of selection candidates in the Low line, the continued selection of progeny from ewes that reared one lamb per mating has taken place. Rams were selected on three or more maternal reproductive opportunities. Female replacements were selected from less than three maternal reproductive opportunities due to the greater demand for female replacements. Ewes remained in the breeding flock for at least five reproductive opportunities, unless teeth or udder defects necessitated culling. At the onset of the experiment, approximately 120 breeding ewes represented each line. The H line gradually increased to between 130 and 140 breeding ewes, whereas the number of breeding ewes in the Low line decreased to between 40 and 80. Between 1998 and 2002, dams that participated in five lambing opportunities were screened into both the High and Low lines from the selection lines maintained at Tygerhoek Research farm in an attempt to augment the number of ewes in the Low line. Ewes that reared more than seven lambs in five reproductive opportunities were screened into the High line, while ewes that reared between one and three lambs in five reproductive opportunities were introduced into the Low line. Between eight and eighteen ewes were introduced per year and 74 were introduced in total. Until the mid-1990's, five breeding rams were used per line and these rams were replaced every year. From 1992, one to two rams were retained for breeding in the subsequent year. The number of breeding rams in the High line ranged between four and six, while between two and four rams were used in the Low line to ensure that the ratio of male to female breeding stock remained constant within each line. The average generation interval was approximately 3.2 years; 2.3 years for rams

and 4.1 years for ewes. Reciprocal crossbreeding of some members of the High and Low lines took place and the first crossbreds were born in 2003. Crossbred individuals were backcrossed to either the High or Low lines to form High line and Low line backcross groups.

The reproductive performance of breeding ewes in the Elsenburg Merino flock has been recorded since the inception of the selection experiment. Breeding values were estimated from these records using a linear mixed model maximum likelihood method, referred to as a best linear unbiased prediction (BLUP) (Gilmour *et al.* 1995). The fixed effects included year of birth, sex, dam age, and birth type (single or multiple). The number of lambs born per ewe per joining (NLB); number of lambs weaned per ewe per joining (NLW); the total weight of lamb weaned per ewe per joining (TWW), and estimated breeding values (EBVs) (as described in Cloete *et al.* 2004) were considered in the current study. Ewes from the divergent lines (High line and Low line), crosses and backcrosses that were present for at least two lambing opportunities were included. To investigate the change in the values for NLB, NLW and TWW and the concomitant EBVs, raw trait value means, standard deviations and standard errors were calculated for individuals born from 1985 to 2009 in the High and Low lines. The annual phenotypic means and EBVs were regressed on birth year to investigate linear trends. The regression model was forced through the origin for all mean EBVs.

From a total of 1458 Elsenburg Merino ewes contributing reproductive data, a subset of individuals were identified to examine the variation in the lines; crossbreds and backcrosses. As the crossbreds became available in 2003, only individuals born after 2002 were included for all groups. Individuals included in this subset had between two and five lambing opportunities. The number of individuals in the subset amounted to 412, of which 50% were High line individuals, 10% Low line, 20% crossbreds and 20% backcrosses to either the High or Low lines. The crossbreds and backcrosses were produced by reciprocal crosses and could thus be divided into breeding groups according to their parental joining. An alphabetic identifier was arbitrarily assigned to the different groups and used in the results presented below to identify the specific groups. The theoretical genetic resemblance in relation to the High line was assigned to each group depending on the line or group the grandsire, granddam, sire and dam belonged to (Table 2.1). Raw mean trait values, standard deviations and standard

errors were calculated for NLB, NLW and TWW as well as the concomitant EBVs. Variance coefficients were also calculated for the means of NLB, NLW and TWW to compare difference in trait variance between the two lines and different breeding groups. A pairwise F-test, followed by a pairwise t-test, was performed to test for homogenous variances and determine whether significant differences occurred between the means of the breeding groups. The degree of inbreeding, as determined by pedigree-based methods (ΔF) (Quaas 1976), was reported as an average value per group.

Table 2.1: The number of individuals included in the data subset to compare the breeding groups (A to H) within the lines, crosses and backcrosses.

Description	Reciprocal crosses	n	Group	Genetic resemblance to the High line (%)
High line (HL)	-	206	A	100
Low line (LL)	-	43	B	0
F1 (LL x HL)	LL sire x HL dam	43	C	50
	LL dam x HL sire	39	D	50
Backcross to Low line (F1 x LL)	LL sire x Group C	19	E	25
	LL sire x Group D	19	F	25
Backcross to High line (F1 x HL)	HL sire x Group C	20	G	75
	HL sire X Group D	23	H	75

LL: Low line; HL: High line; n: Number of individuals

2.3. Results:

The linear regression results indicated an increase in the yearly mean for NLB, NLW and TWW in the High line, while a general decrease was evident in the Low line from 1985 to 2009 (Figure 2.1). A yearly increase of approximately 0.009 in NLB, 0.011 in NLW and 0.239 in TWW was predicted by fitting a linear regression model to the mean yearly trait values of the High line (Table 2.2). The linear regression results for the means of NLB, NLW and TWW exhibited a yearly decrease of 0.009, 0.001 and 0.136, respectively, for the Low line. The R^2 value, indicating the variation in the dependant variable explained by the linear regression model, for NLB, NLW and TWW per year were very low for both lines. This indicated that a large proportion of the variation in the data set is not explained by the linear regression model, especially in the Low line where all values were <0.09 (Table 2.2). The standard error value, a measure of the standard deviation of the mean, increased between 1985 and 2009 for the Low line. A higher degree of variation is therefore evident in the mean values for Low line individuals in 2009 than was present originally in 1985. In the High line, the variation across different years remained relatively constant with a slight decrease since 2008 (see Appendix Table A2.1 for the values of the mean, standard error and sample number). The relatively high degree of variation in the trait means of the Low line could be the result of low sample numbers in recent years. A large sample size results in a more accurate representation of the group mean and thereby also a decrease in the sample deviation from the mean (provided the mean remains constant). The number of Low line individuals has decreased from a mean (SD) of 19.4 (5.62) individuals per year in the period 1985 to 1994, to a mean of 6.2 (3.01) individuals per year from 2000 to 2009.

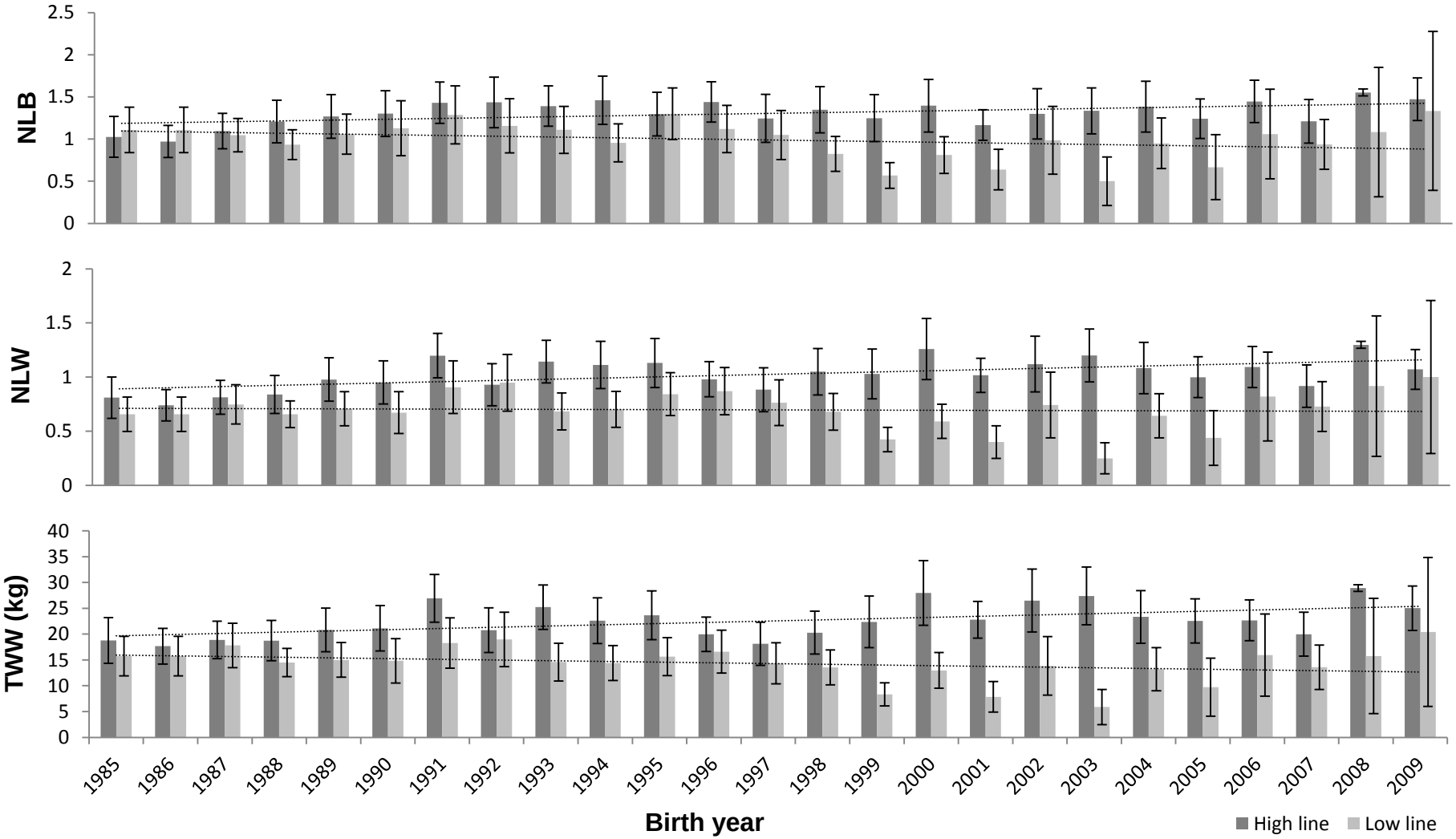


Figure 2.1: Mean number of lambs born (NLB), number of lambs weaned (NLW) and total weaning weight of lambs per joining (TWW) values for ewes from the High and Low line of the Elsenburg Merino flock. Linear regression lines have been added. Error bars indicate the standard error of the mean.

Table 2.2: A linear regression model of mean number of lambs born (NLB), number of lambs weaned (NLW) and total weaning weight of lambs per joining (TWW) per year for the High and Low lines.

Trait	Line	Regression equation	R^2-value
NLB	HL	y= 0.0099x + 1.779	0.2588
	LL	y= -0.0089x + 1.104	0.0899
NLW	HL	y= 0.0111x + 0.8812	0.3211
	LL	y= -0.0012x + 0.7137	0.0025
TWW	HL	y= 0.2393x + 19.406	0.2904
	LL	y= -0.1356x + 16.057	0.0862

HL: High line; LL: Low line

A general yearly increase in the EBVs for NLB, NLW and TWW for the High line and a decrease in those for the Low line were observed (Figure 2.2, Appendix Table A2.2). BLUP EBVs are calculated by comparing an individual's own trait value with that of the population mean using a mixed linear model to account for environmental effects such as year of birth. It was therefore not surprising that most of the mean values for the Low line were less than zero owing to the fact that these values were below that of the larger population. The mean trait values for the High line were, in contrast, greater than zero and therefore above the population average. The value of the standard error increased in both lines from 1985 to 2009. As mentioned before, the increase in the variation around the mean could be the result of a decrease in the yearly sample numbers. The sample number of the High line has not been reduced to the same extent as that of the Low line; nonetheless there is a reduction in sample number from the first 10 sampling years (1985 to1994) to the last 10 years (2000 to 2009). The mean yearly sample size (SD) for the High line for the first 10 years of the study was 46.20 (7.55), while for the last 10 years this value decreased by 24.02% to 35.10 (8.75). The variation around the mean was relatively high in years with very low sample numbers, such as 2004 and 2006. The regression models applied to the EBVs of NLB, NLW and TWW could account for a much larger degree of the variation in the data in comparison to those of the raw means for NLB, NLW and TWW (Table 2.2 and 2.3). The R^2 values for the High line were still considerably larger than those of the Low line. A yearly increase of 0.011, 0.010 and 0.273 in the EBVs of NLB, NLW and TWW, respectively, was predicted by the linear regression model of the yearly means

of the High line (Table 2.3). The linear regression results for the means of the EBVs for NLB, NLW and TWW exhibited a yearly decrease of approximately 0.008, 0.007 and 0.168, respectively, for the Low line.

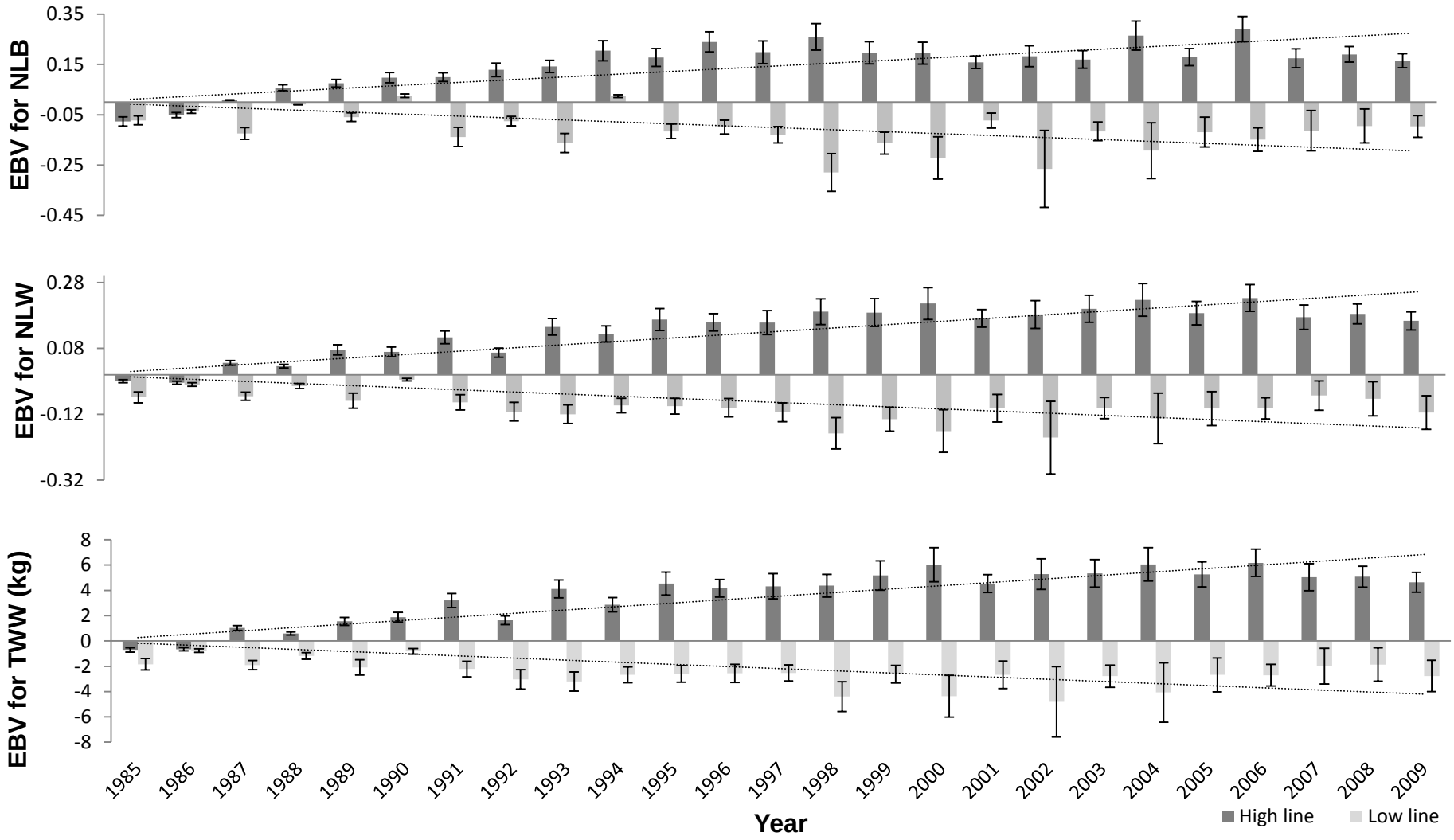


Figure 2.2: Mean estimated breeding values (EBVs) for number of lambs born (NLB), number of lambs weaned (NLW) and total weaning weight of lambs per joining (TWW) for ewes from the Elsenburg Merino flock. Linear regression lines have been added. Error bars indicate the standard error of the mean.

Table 2.3: A linear regression model of the estimated breeding values (EBVs) for mean number of lambs born (NLB), number of lambs weaned (NLW) and total weaning weight of lambs per joining (TWW) per year for the High and Low lines.

Trait	Line	Regression equation	R^2-value
EBV	HL	y = 0.0109x	0.5408
NLB	LL	y = -0.0077x	0.0229
EBV	HL	y = 0.0101x	0.7158
NLW	LL	y = -0.0065x	0.3460
EBV	HL	y = 0.2733x	0.7585
TWW	LL	y = -0.1683x	0.5390

HL: High line; LL: Low line

Employing the subset data, the phenotypic mean NLB, NLW and TWW values were calculated for the lines, crossbreds and backcrosses (Figure 2.3). The High line exhibited the largest mean values for all three traits, while the Low line exhibited the smallest. The crossbreds had intermediate values in relation to the High line and Low line for all three traits. It was therefore expected that the mean values for the backcross groups would also be intermediate to that of the foundation lines. However, the mean trait values for backcrosses to the High line were less than those of the crossbreds for NLW and TWW. For NLB, the backcross group's mean value was still less than the average of the foundation groups. Backcrosses to the Low line displayed larger mean trait values for all three traits than would be expected considering the average of the foundation groups' means. The High line exhibited the least sample variation around the mean, while the Low line exhibited the highest degree of variation. The crosses and backcrosses had intermediate measures of variation around the mean for all three traits (Appendix Table A2.3).

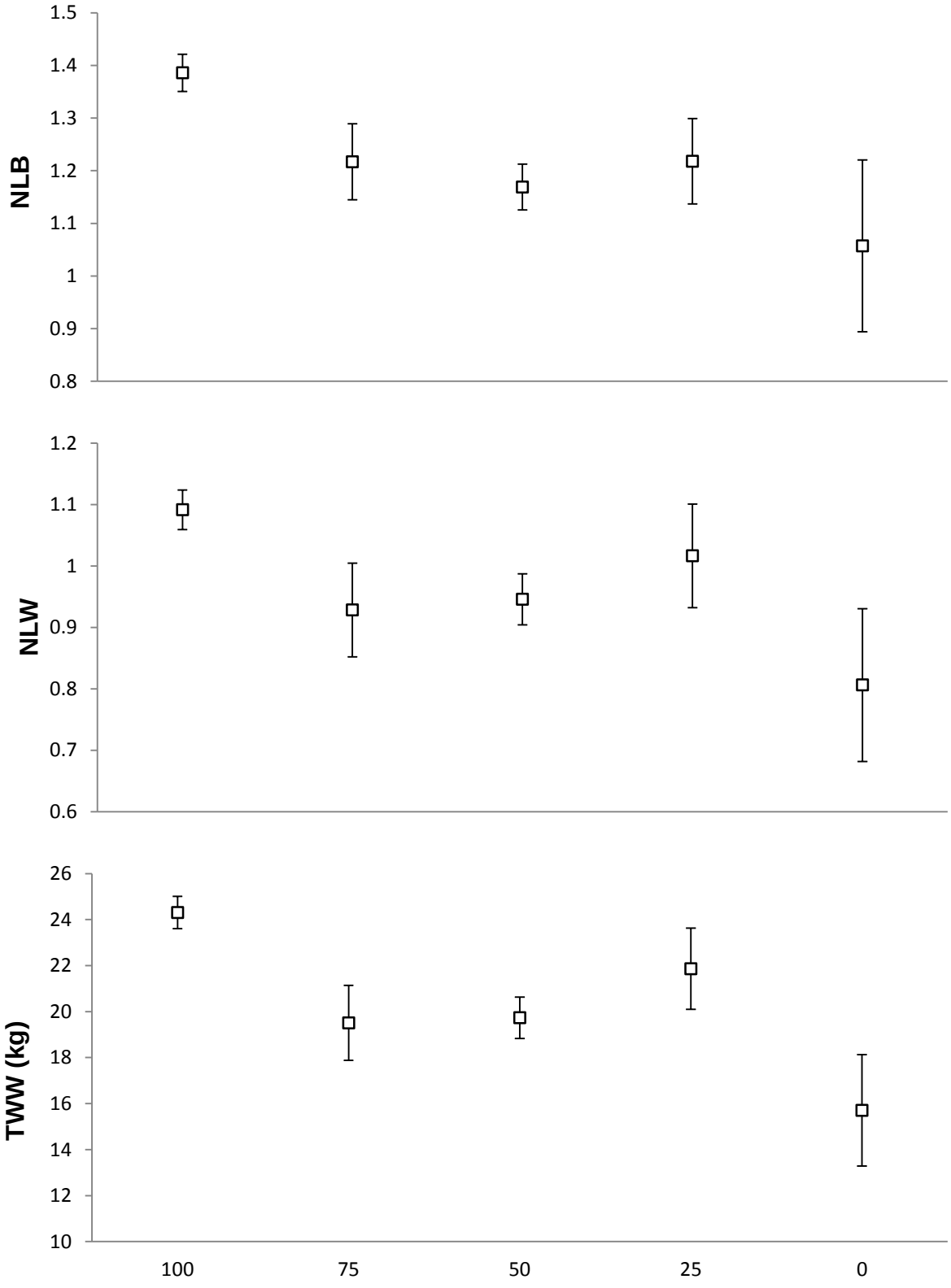


Figure 2.3: Mean number of lambs born (NLB), number of lambs weaned (NLW) and total weaning weight of lambs per joining (TWW) values for individuals in the data subset in relation to their theoretical genetic resemblance to the High line. Error bars indicate the standard error of the mean.

By calculating the mean trait values for all the breeding groups, the source of variation within the groups and the change in the mean trait values could be further investigated (Figure 2.4). The backcrosses to the High line are represented by Group G and H; where Group G comprised the individuals resulting from a cross between a High line ram and a crossbred ewe; the crossbreed ewe being the result of a cross between a Low line sire and a High line ewe. Group H was also the result of a High line ram crossed with a crossbred ewe, but this crossbred ewe was the result of a High line sire and a Low line dam (Table 2.1). The sample size (n=20 for Group G; n=23 for Group H) and the variation around the mean was comparable between the two groups, yet Group G's mean values for all three traits were larger than that of Group H. The two groups within the backcrosses to the Low line also exhibited differences in the mean trait values. Group E represented a backcross of a Low line sire with a crossbred dam originating from a cross between a Low line sire and High line dam. Group F was also a backcross to the Low line, but had a Low line grand-dam and High line grand-sire. Group E exhibited larger mean values for NLB and NLW than group F, while Group F had a larger mean value for TWW (Figure 2.4). This is surprising considering that the TWW is calculated for the combined weight of all lambs weaned and because Group E weaned more lambs on average than Group F, it would be expected that Group E would exhibit a larger TWW value than Group E as well. Group E and F have the smallest sample size of all the breeding groups (n=19 for Group E; n=19 for Group F) and some of the largest standard error values of the mean (only second to the Low line). Pairwise t-tests indicated significant ($p<0.05$) differences in the mean of the High line and Low line; High line and the crossbreds; and the High line and the backcrosses to the High line for all three traits ($p<0.05$). A significant ($p<0.05$) difference in the mean also existed between the Low line and backcrosses to the Low line for NLW and TWW. For TWW, a significant difference in the mean between the Low line and crossbreds were observed. There were no significant differences in the mean trait values for the three traits tested within the crossbreds (Groups C and D) and backcrosses (Groups E and F; and Groups G and H).

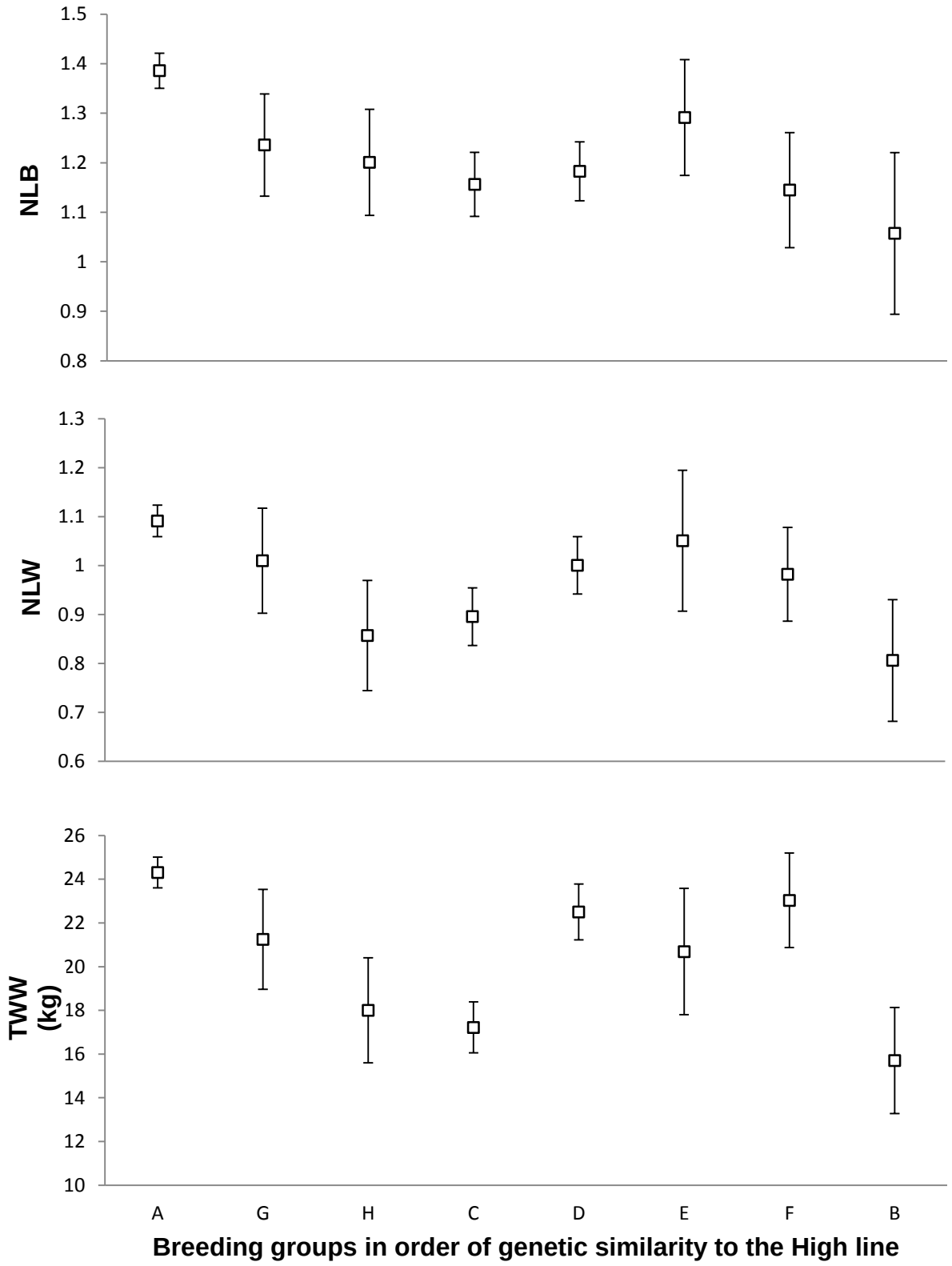


Figure 2.4: Mean number of lambs born (NLB), number of lambs weaned (NLW) and total weaning weight of lambs per joining (TWW) values for individuals in the data subset divided into the breeding groups. Error bars indicate the standard error of the mean. Group A and B represent the High and Low line, respectively; C and D are the crossbreds; G and H are the backcrosses to the High line; and E and F are the backcrosses to the Low line.

The coefficient of variation (CV), a measure of the dispersion of sample values around the mean, was calculated for the lines, crossbreds and backcrosses (Figure 2.5) and for the different breeding groups (Figure 2.6). As discussed earlier, the two divergent lines were expected to exhibit the least phenotypic variation within lines followed by the crossbreds with relatively little variation. The High line's CV values were between 37 and 43% and relatively small relative to the other groups. However, the Low line had a relatively high CV value of approximately 50% for all three traits. The crossbreds group's CV values were relatively low for all three traits, while the backcross groups exhibited relatively large CV values. Group H and Group E exhibited the largest CV values of between 58 and 65% for NLW and TWW (Figure 2.6). The CV values for all the breeding groups for NLW and TWW followed the same trend with similar values being recorded across the individual groups. Generally the CV for NLB was less than those for NLW or TWW and the trend in this trait differed slightly from the other two traits. Group H and E exhibited the largest CV values for NLW and TWW, while Group F and the Low line exhibited the largest CV values for NLB in comparison with the other groups.

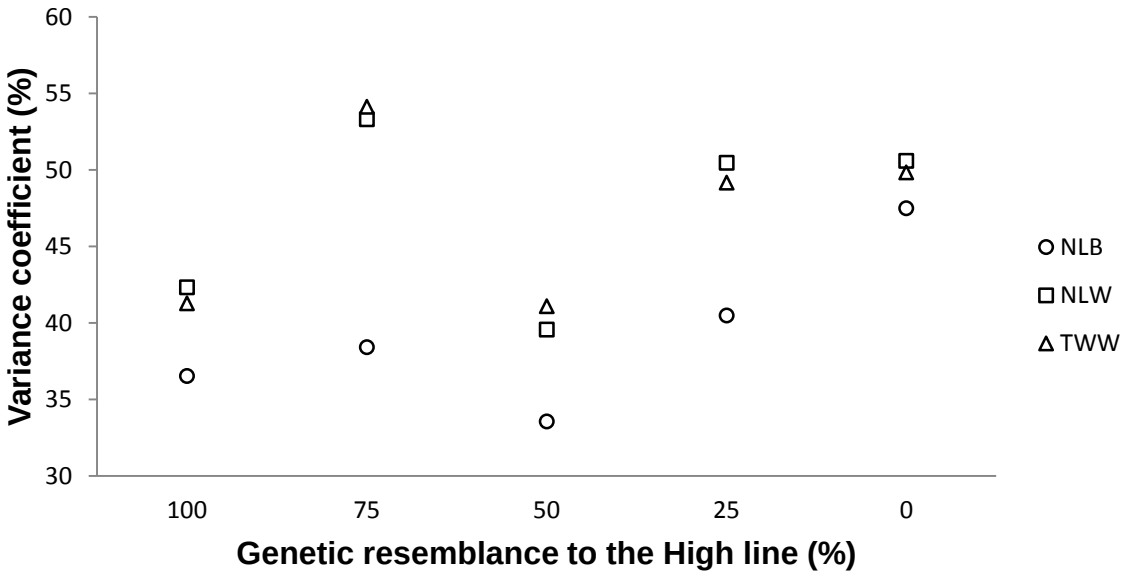


Figure 2.5: Variance coefficient of the mean number of lambs born (NLB), number of lambs weaned (NLW) and total weaning weight of lambs per joining (TWW) values for individuals in the data subset in relation to their theoretical genetic similarity to the High line.

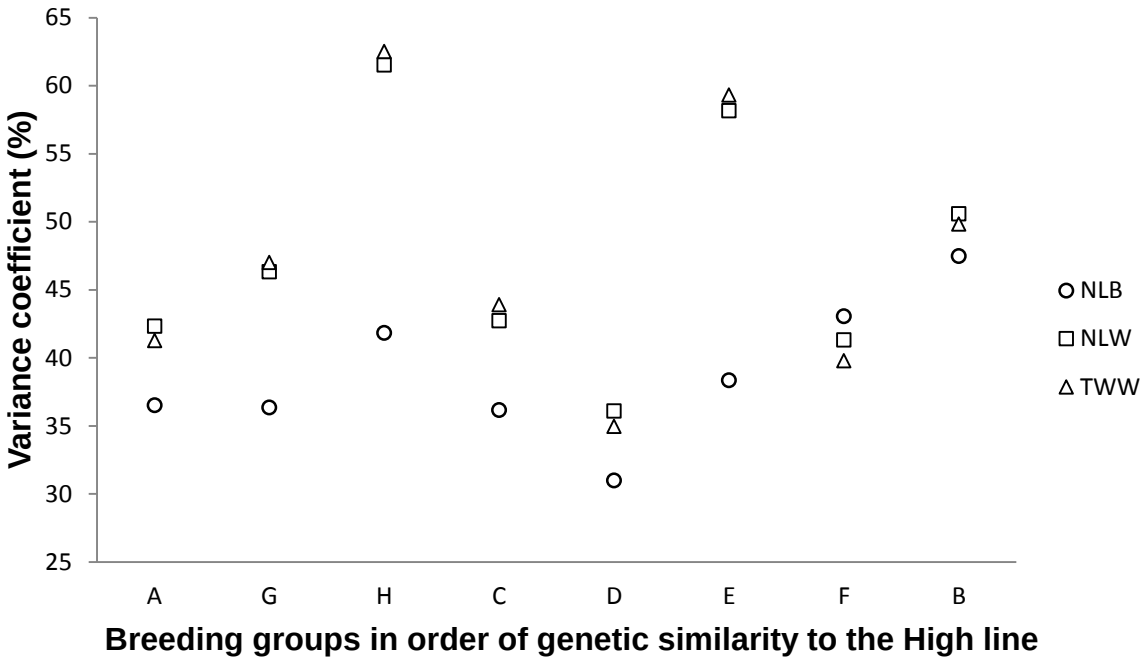


Figure 2.6: Variance coefficient of the mean number of lambs born (NLB), number of lambs weaned (NLW) and total weaning weight of lambs per joining (TWW) values for individuals in the data subset according to their breeding groups. Group A and B represent the High and Low line respectively; C and D are crossbreds; G and H are the backcrosses to the High line; and E and F are the backcrosses to the Low line.

The High line exhibited the largest mean EBVs for all three traits, while the Low line exhibited the lowest values (Figure 2.7). The other breeding groups exhibited intermediate values depending on their genetic background (Figure 2.8). Generally a linear relationship between the genetic similarity to the High line and breeding values for all the traits were observed. The backcross group's EBVs for all three traits were intermediate to their foundation groups and did not surpass or fall below that of the foundation groups as was the case when comparing the phenotypic mean for NLB, NLW and TWW. The two reciprocal cross groups of the backcrosses also did not differ for the EBVs to the same extent as for the NLB, NLW and TWW values. The High line, Low line and backcrosses to the High line exhibited the highest degree of variation, with relatively minor sample variation around the mean evident for all the other groups.

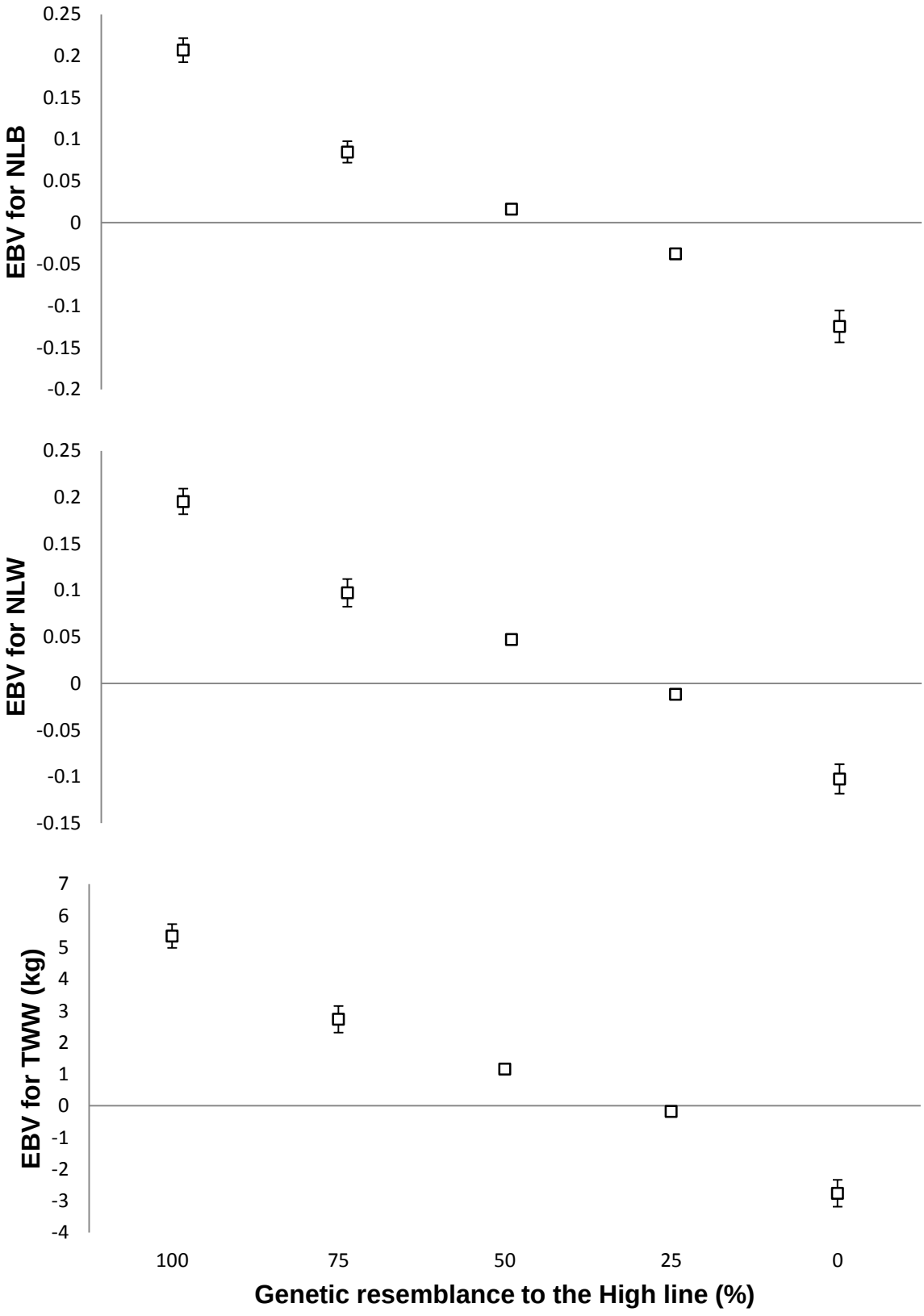


Figure 2.7: Mean EBVs for number of lambs born (NLB), number of lambs weaned (NLW) and total weaning weight of lambs per joining (TWW) values for individuals in the data subset data in relation to their theoretical genetic resemblance to the High line. Error bars indicate the standard error of the mean.

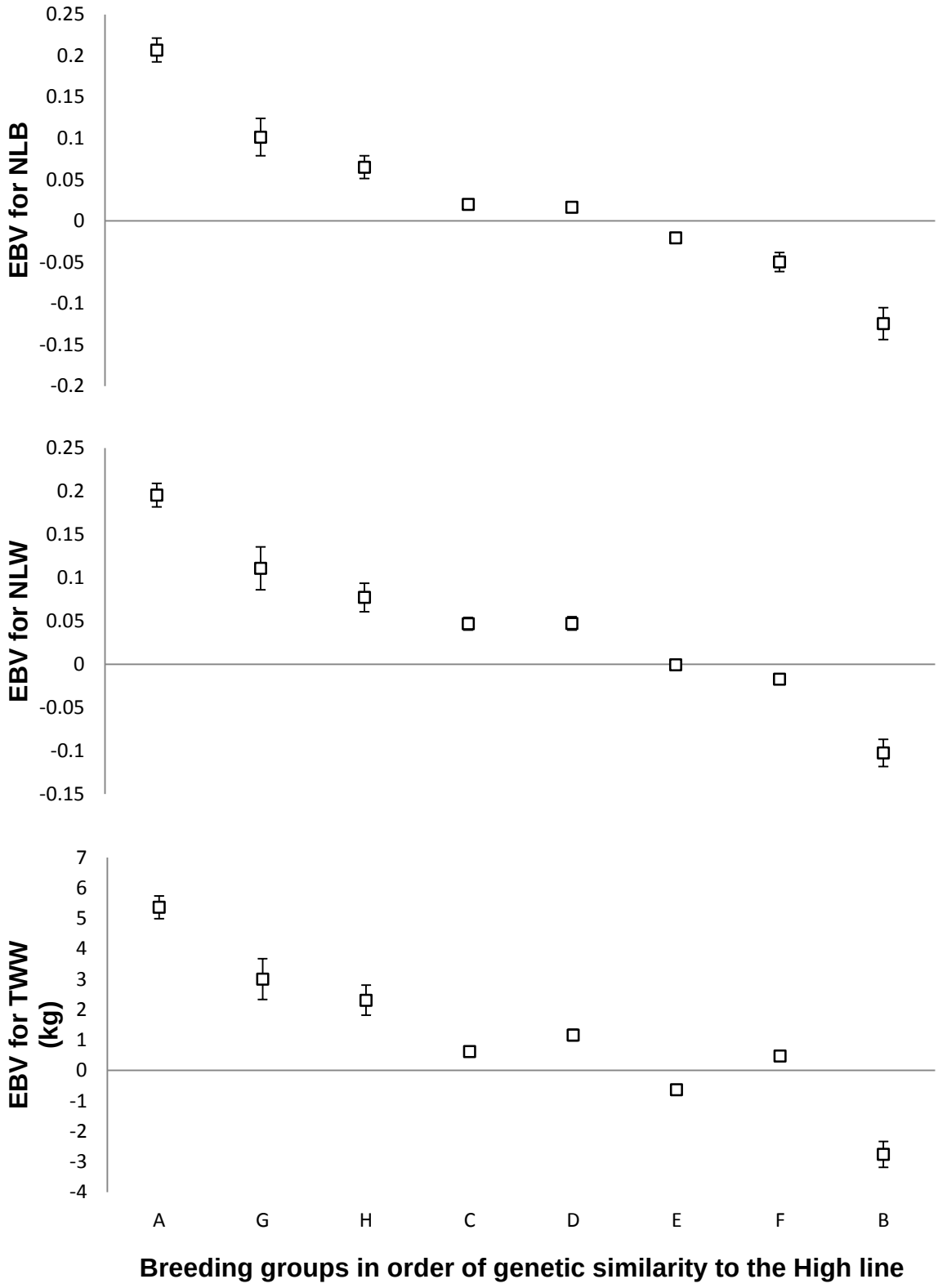


Figure 2.8: Mean number of lambs born (NLB), number of lambs weaned (NLW) and total weaning weight of lambs per joining (TWW)values for individuals in the subset data separated into the different breeding groups. Error bars indicate the standard error of the mean. Group A and B represent the High and Low line respectively; C and D are the F1 crossbreds; G and H are the backcrosses to the High line; and E and F are the backcrosses to the Low line.

Due to the limited number of breeding ewes and rams contributing to the divergent lines in the Elsenburg Merino flock, the extent of shared ancestry, referred to as inbreeding, has steadily increased over the duration of the experiment. The inbreeding was assumed to be zero at the onset of the trial. Pedigree information was used to determine the degree of inbreeding per individual (Quaas 1976). A mean inbreeding coefficient (Miglior 2000) of <4% was found to be present in all the groups, except the crossbreds, where no inbreeding was present (Figure 2.9). The crossbreds are the result of crossing two divergent lines and therefore did not share any ancestors. The High and Low lines exhibited the highest degree of inbreeding, 3.8% and 3.0%, respectively. The backcrosses had inbreeding values of between 2.5% and 1.8% which were intermediate to that of their foundation groups. The inbreeding in all groups was <6.25% and can therefore be described as passive inbreeding as a result of a limited population size, as defined by Miglior (2000).

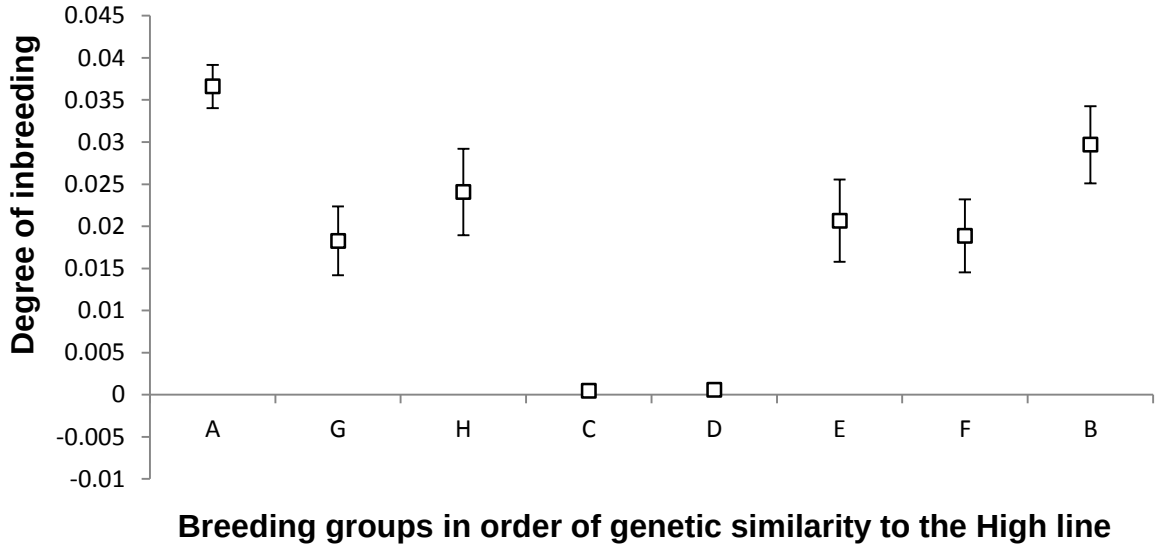


Figure 2.9: The mean inbreeding coefficient in the lines, crosses and backcrosses for individuals in the data subset separated into the breeding groups. Error bars indicate the standard error of the mean. Group A and B represent the High and Low line respectively; C and D are the F1 crossbreds; G and H are the backcrosses to the High line; and E and F are the backcrosses to the Low line.

2.4. Discussion:

A general yearly increase in mean NLB, NLW and TWW values was evident for High line individuals with a concurrent decrease for Low line individuals born between 1985 and 2009. The same trend was observed for the mean EBVs for NLB, NLW and TWW across the same timeframe. The linear regression models explained a larger degree of the variation of the mean yearly EBVs of NLB, NLW and TWW in comparison to the degree of variation explained by the linear regression models fitted to the raw NLB, NLW and TWW values. Due to a high degree of variation in the mean yearly NLW, NLB and TWW values only a weak relationship was evident between the linear regression model and the variation in the raw trait values. A BLUP EBV compensates for phenotypic variation introduced through environmental factors by including fixed and random effects in the mixed model prediction equations (Gilmour *et al*. 1995). The BLUP EBV also incorporates pedigree information and therefore the performance of an individual's relatives will also influence their BLUP EBV to an extent (Henderson & Quaas 1976). The end-result is an EBV relaying an accurate value for the genetic potential of an individual by compensating for environmental variation and taking the performance of relatives into account (Cameron 1997). BLUP EBVs are therefore less prone to reflect spurious environmentally introduced variation than phenotypic values. Production traits and especially reproduction traits with a relatively low heritability are influenced by changing environmental factors and are therefore subject to a high degree of variation particularly across different breeding opportunities or years (Safari *et al*. 2005). In the current study, less within line variation in mean yearly trait values was evident from the EBVs than for the phenotypic values as would be expected taking the aforementioned features of BLUP EBVs into account. Both the mean phenotypic yearly trait data and the mean yearly EBVs indicated a general increase in the degree of phenotypic variation since the onset of the trial. The initial low degree of phenotypic variation within each line was expected as only a limited number of animals, comprising the extremes of the unselected base flock, were selected to launch the trial.

When considering the mean yearly values for NLB, NLW and TWW, an initial increase in these values was observed for the Low line. It is only after 1997 that a noticeable decrease in these values occurred. Although the mean yearly values of the Low line indicated an increase over the initial trial period, these mean yearly values were

typically still less than those of the High line. A general change in the environment that the lines were maintained in may have contributed to the increase in phenotypic variation over the period of the trial. Indeed, the flock was translocated in 1993 from Tygerhoek Research Farm near Riviersonderend to Elsenburg Research Farm near Stellenbosch.

A change in the yearly sample size could also influence the degree of phenotypic variation observed. A large sample size results in a more accurate representation of the group mean and therefore a decrease in the sample deviation from the mean. The yearly sample sizes decreased by 24.02% in the High line and 68.04% in the Low line from the first 10 years of the trial (1985 to 1994) to the last ten years included in the current study (2000 to 2009). It was therefore not surprising that the degree of variation increased in both lines and that the Low line exhibited a much larger degree of variation than the High line.

The selection goal of the Low line and the concomitant selection pressure needs to be taken into account when considering the phenotypic variation in this line. The breeding goal of the Low line is to decrease the number of lambs weaned per ewe per joining. This will ultimately result in the extinction of the line if sufficient selection pressure can be applied and therefore increasingly fewer offspring are born each year to replace older stock. As a consequence of a decrease in the number of offspring weaned per year, fewer breeding animals are available for selection and therefore selection pressure has decreased. The High line's breeding goal is for an increase in the number of lambs weaned per ewe per year and therefore this line does not face the same potential extinction risks as the Low line. The capacity to apply selection pressure and the associated response in the High line is much greater than the Low line. The artificial selection pressure applied to the Low line is contrary to the natural selection for survival of this line and these two opposing forces as well as the decrease in sample size, resulted in the presence of a greater variation in this line in comparison to the High line. A similar asymmetrical response to selection has also been reported in other divergently selected sheep and chicken lines. Johansson *et al.* (2010) and Phua *et al.* (2014) reported that the line selected for a decrease in a trait essential for welfare or survival, such as body weight or disease resistance, exhibited greater

phenotypic variation and less response to selection than the line selected for an improvement in these traits.

The crossbred group exhibited intermediate values for NLB, NLW and TWW, and the associated EBVs in comparison to the values of the High and Low lines. The backcrosses to the Low line tended toward mean trait values that exceeded that of their foundation lines, while the backcrosses to the High line's mean trait values fell short of the expected intermediate values of their foundation lines. Significant differences in the trait means could be observed between the High line and the Low line, and between the two aforementioned lines and the respective backcrosses to those lines for most traits. These results confirm that the two lines differ significantly with regard to reproductive performance and that the difference in reproductive performance between the lines and backcrosses were also significant. This is a surprising result considering that the crossbreed group exhibited trait values that were intermediate to that of the High and Low lines. In view of the fact that selection pressure was continually applied to the High and Low lines, but not to the crosses and backcrosses, different time-points during the progression in selection may have been compared in the current study. However, attempts at correcting for the ongoing selection in the High and Low lines by comparing the trait mean of a subset representing a specific time-point during selection, did not change the relationship of the mean trait values of the backcrosses with those of the respective parental groups (data not shown).

The unexpected difference in trait means between the lines and the backcrosses to those lines may be the result of epistatic interaction between loci from different genetic backgrounds. Epistatic interactions between loci can result in an enhanced or reduced phenotype in comparison to additive loci and is described as either positive (synergistic), or negative (antagonistic) epistasis (Whitlock & Bourguet 2000, Bonhoeffer *et al*. 2004). The backcrosses to the Low line exhibit mean trait values that were significantly higher than would be expected and therefore a unique and favourable interaction may exist between the genetic background inherited from the dam and the sire. The backcrosses to the low line (Groups E and F) resulted from the combination of a Low line paternal line and a hybrid maternal line. Group E, possessing a maternal grandsire from the Low line, exhibited larger values for NLB

and NLW in comparison with Group F. This suggests that the combination of a Low line sire and a High line ewe line may result in a certain epistatic synergy that causes an enhanced reproductive phenotype. In contrast, the backcrosses to the High line (Groups G and H) underperformed in comparison to their foundation lines. Group G exhibited larger mean trait values than that of Group H, although no significant difference in mean trait values was evident between the two groups. Nonetheless, difference in trait means may indicate some synergy in the parental lines of the two groups. Although Group G and H share a paternal High line sire, Group G's maternal grandsire is from the Low line (Group C) and therefore a Low line sire is also present in this lineage. This also suggests that the combination of a Low line paternal line and a hybrid maternal line may facilitate an epistatic interaction that results in an enhanced reproductive phenotype. Phua *et al.* (2014) also suggest a synergetic interaction between the genetic background of outcrossed Romney selection lines (Phua *et al*. 2009). The study by Phua *et al.* (2014) failed to identify all QTL previously identified by a study conducted on outcrossed half-sib selection lines (Phua *et al.* 2009) and therefore suggests that a synergistic interaction may underlie the initial identification of certain QTL. Although the effects of epistatic interaction may explain the differences in trait means between the breeding groups in the current study, the limited sample numbers may influence inferences made from this data and therefore further studies utilising more samples together with genomic information (Eronen *et al*. 2010) would be necessary to verify the occurrence of epistatic interactions.

Heterosis was also considered as a cause for the mean trait values of the backcrosses to the Low line exceeding that of their foundation lines. Heterosis usually occurs in hybrid offspring resulting from a cross between unrelated and somewhat inbred lines as inbreeding depression is countered and the combination of dissimilar alleles at certain loci result in an offspring that performs better than either parental lines (Dekkers & Hospital 2002). Therefore, heterosis would be expected in the crossbred group rather than the backcrosses. The fact that heterosis is not observed for NLB, NLW or TWW of the crossbreds may be the consequence of comparatively low levels of inbreeding in the Elsenburg Merino flock.

2.5. Conclusions:

The Elsenburg Merino flock has been divergently selected for the ability to rear multiple offspring and has resulted in a High and Low line that has become increasingly divergent for NLB, NLW and TWW during the selection experiment. A small sample size in the Low line may have resulted in an apparent increase in phenotypic variation, however, other selection studies have also reported greater phenotypic variation in the line in which essential survival traits are selected against. The High line is expected to reach a plateau in reproductive performance while a continued reduction in reproductive performance in the Low line will proceed slowly and will eventually result in the extinction of the line. Evaluation of the crossbreds of the lines and backcrosses to the lines indicated the reproductive performance of the crossbreds to be intermediate to that of the two lines while the backcrosses exhibited mean trait values that either exceeded or fell below that of the parental lines. Surprisingly, the backcrosses to the Low line exhibited a superior reproductive performance, while the backcrosses to the High line exhibited lower reproductive performance in comparison to the respective parental lines. An epistatic interaction between the genetic background of the parental lines may underlie these results. Future studies relying on genomic information may be able to pinpoint the genomic regions responsible for variation in reproduction traits between the two lines. The current study confirms the high level of genetic variation contained in the base population from which the Elsenburg Merino lines have been bred and confirms the value of these lines as an animal resource for genetic research of South African Merino sheep.

2.6. References:

Andersson L (2001) Genetic dissection of phenotypic diversity in farm animals. Nature 2: 130-138

Andersson L, Georges M (2004) Domestic-animals genomics: Deciphering the genetics of complex traits. Nature 5: 202-212

Bonhoeffer S, Chappey C, Parkin NT, Whitcomb JM, Petropoulos CJ (2004) Evidence for positive epistasis in HIV-1. Science 306: 1547-1550

Cameron ND (1997). Selection indices and prediction of genetic merit in animal breeding. CAB International, Wallingford, UK

Cloete SWP, Van Wyk JB, Scholtz AJ, Gilmour AR (2003) Genetic and environmental parameters for lambing behaviour in Merino lines divergently selected for ewe maternal rearing ability. Proceedings of the New Zealand Society of Animal Production 63: 169-172

Cloete SWP, Gilmour AR, Olivier JJ, Van Wyk JB (2004) Genetic and phenotypic trends and parameters in reproduction, greasy fleece weight and live weight in Merino lines divergently selected for multiple rearing ability. Australian Journal of Experimental Agriculture 44: 745-754

Cloete SWP, Olivier JJ, Van Wyk JB, Schoeman SJ, Erasmus GJ (2005a) Genetic parameters and trends for hogget traits in Merino lines divergently selected for multiple rearing ability. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 16: 24-27

Cloete SWP, Scholtz AJ, Cloete JJE, Van Wyk JB (2005b) The ability of Merino ewes and lambs to reunite after separation, as affected by divergent selection for ewe multiple rearing capacity. Australian Journal of Experimental Agriculture 45: 1131-1137

Cloete SWP, Olivier JJ, Olivier WJ (2007) Genetic change in South African Merino resource flocks. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 17: 320-323

Cloete SWP, Misztal I, Olivier JJ (2009) Genetic parameters and trends for lamb survival and birth weight in a Merino flock divergently selected for multiple rearing ability. Journal of Animal Science 87: 2196-2208

Dekkers JCM, Hospital F (2002) The use of molecular genetics in the improvement of agricultural populations. Nature 3: 22-32

Eronen V-P, Lindén RO, Lindroos A, Kanerva M, Aittokallio T (2010) Genome-wide scoring of positive and negative epistasis through decomposition of quantitative genetic interaction fitness matrices. Public Library of Science ONE 5: e11611

Gilmour AR, Thompson R, Cullis BR (1995) Average information REML, an efficient algorithm for variance parameter estimation in linear mixed models. Biometrics 51: 1440-1450

Henderson CR, Quaas RL (1976) Multiple trait evaluation using relatives' records. Journal of Dairy Science 43: 1188-1197

Heydenrych HJ, Du Plessis JJ, Cloete SWP (1984) Increasing the wool production of Merino sheep in the South Western districts of South Africa by direct and indirect selection. In: Hofmeyr JH, Meyer EHH (eds.), Proceedings of the second world congress of sheep and beef cattle breeding, 16-19 April, Pretoria. South African Stud Book and Livestock Improvement Association, Bloemfontein pg. 399-412

Hoffmann AA, Sgrò CM (2011) Climate change and evolutionary adaptation. Nature 470: 479-485

Johansson AM, Pettersson ME, Siegel PB, Carlborg Ö (2010) Genome-wide effects of long-term divergent selection. Public Library of Science Genetics 6: e1001188

Miglior F (2000) Impact of inbreeding - managing a declining Holstein gene pool. Proceedings from the World Holstein Friesian Federation Conference 10: 108-113. Sydney, Australia

Notter DR (1999) The importance of genetic diversity in livestock populations of the future. Journal of Animal Science 77: 61-69

Phua SH, Dodds KG, Morris CA, Henry HM, Beattie AE, Garmonsway HG, Towers NR, Crawford AM (2009) A genome-screen experiment to detect quantitative trait loci affecting resistance to facial eczema disease in sheep. Animal Genetics 40: 73-79

Phua SH, Brauning R, Baird HJ, Dodds KG (2014) Identifying chromosomal selection-sweep regions in facial eczema selection-line animals using an ovine 50K-SNP array. Animal Genetics 45: 240-247

Quaas RL (1976) Computation of the diagonal elements and inverse of a large numerator relationship matrix. Biometrics 32: 949-953

Safari E, Fogarty NM, Gilmour AR (2005) A review of genetic parameter estimates for wool, growth, meat and reproduction traits in sheep. Livestock Production Science 92: 271-289

Schoeman SJ, Cloete SWP, Olivier JJ (2010) Returns on investment in sheep and goat breeding in South Africa. Livestock Science 130: 70-82

Scholtz AJ, Cloete SWP, Van Wyk JB, Kruger ACM, Van der Linde TCD (2010) Influence of divergent selection for reproduction on the occurrence of breech strike in mature Merino ewes. Animal Production Science 50: 203-209

Whitlock MC, Bourguet D (2000) Factors affecting the genetic load in *Drosophila*: Synergistic epistasis and correlations among fitness components. Evolution 54: 1654-1660

Chapter 3:

Evaluation of the OvineSNP50 genotyping array in four South African sheep breeds

3.1 Introduction:

The OvineSNP50 BeadChip was developed by Illumina in collaboration with the International Sheep Genomics Consortium and became commercially available in 2009. The microarray-based genotyping system is designed to determine the genotype of approximately 54 000 single nucleotide polymorphisms (SNPs) spaced evenly across the *ovine* genome. The mean genomic distance between SNPs included in the chip is approximately 51kb with a median distance equal to 42kb (Illumina 2010a, International Sheep Genomics Consortium *et al.* 2010, Kijas *et al*. 2012). Approximately 500 SNPs discovered through BAC end Sanger sequencing of nine animals, each representing a different breed, are incorporated in the 50K chip. Roche 454 whole-genome shotgun sequencing of six individuals of six different breeds provided approximately 33 000 additional SNPs included in the chip. A final 15 000 SNPs, identified by reduced representational sequencing (RRS), were also incorporated into the SNP chip. The aforementioned RRS was carried out on 60 individuals from 15 different breeds (Illumina 2010a, Kijas *et al*. 2012). The SNP chip has been validated in more than 75 international sheep breeds and used in a wide range of applications that include the elucidation of quantitative trait loci (QTL) and genomic selection strategies (Daetwyler *et al.* 2010, Miller *et al.* 2011, Kijas *et al*. 2012, Demars *et al.* 2013, Liu *et al.* 2013, Våge *et al*. 2013, Gutiérrez-Gil *et al.* 2014, Baloche *et al.* 2014, Phua *et al.* 2014).

Single nucleotide polymorphism (SNP) panels developed from a limited number of individuals or breeds are not necessarily representative of the polymorphisms and frequency distribution of alleles contained in the genome of all members of the species; this being referred to as ascertainment bias (Nielsen & Signorovitch 2003, Morin *et al.* 2004). Ascertainment bias may influence parameters estimated from genotyping results and can therefore bias the conclusions drawn from such data (Kijas *et al.* 2009, Albrechtsen *et al.* 2010). Ascertainment bias can also influence assumptions regarding a test population as well as the association between specific

traits and molecular markers in that population (Heslot *et al.* 2013). It is therefore vital to recognise the effect of possible ascertainment bias of a marker panel on inferences regarding allele frequency distribution, linkage disequilibrium (LD), population genetic structure, association studies and selection strategies (Clark *et al.* 2005, Heslot *et al.* 2013).

The limited number of individuals and breeds used in the development of the OvineSNP50 BeadChip may have resulted in ascertainment bias and consequently will influence the use of the chip in breeds underrepresented during the initial SNP discovery stage. This may also have an effect on certain applications that require a representative SNP panel (Kijas *et al*. 2009). By genotyping more than 75 international sheep breeds with the OvineSNP50 chip, Kijas *et al*. (2012) confirmed that the number of polymorphic loci as well as the distribution of the minor allele frequencies (MAFs) differed between sheep breeds. Most SNPs exhibited MAFs>0.3 and loci with rare alleles or equal distribution of both SNP alleles (MAF≥0.4) were in the minority for the European and American breeds tested. This is in contrast with the results obtained for the African and Asian breeds where the groups exhibited more SNP loci with low MAF, fewer loci with a MAF≥0.3 and more loci with an equal distribution of major and minor alleles in comparison to the European and American breeds (Kijas *et al*. 2012). Seven breeds and 23 individuals from African and Asian breeds were used in SNP discovery for the *ovine* SNP chip, while 19 breeds and 47 individuals from American and European breeds were used. The differences in the number of polymorphic loci and SNP allele frequencies reported by Kijas *et al*. (2012) may therefore be a consequence of ascertainment bias that resulted from the use of a limited number of individuals and breeds during SNP discovery as well as the use of reductionist SNP discovery methods such as RRS (Clark *et al*. 2005).

Although marker ascertainment bias is of concern for the interpretation of genotyping results, the genotype call remains the basis upon which all further analyses are built and therefore it is crucial to ensure accurate calls (Pare 2010). Genetic markers and individual samples of inferior genotype call quality can be identified by calculating confidence statistics for genotype calls. Indicators of the genotype call quality of Illumina Infinium genotyping platform include the GenTrain score, which is a reflection of the cluster quality. The former refers to the similarity of the normalised intensity data

within and between different genotype calls of individual SNP loci. The GenCall score in turn indicates the reliability of a call by considering the position of an individual sample within a cluster (Fan *et al.* 2004, Illumina 2005). Using GenomeStudio Software (Illumina) raw intensity data is normalised to account for variation within each run before being clustered according to the normalised intensity value. From the normalised data, calls are made depending on the GenTrain score and the GenCall score. The GenCall score (a value between 0 and 1) is used in the no-call cut-off for samples, where a value <0.25 is considered a no-call for the OvineSNP50 genotyping platform. The GenCall score is calculated by considering information of the clustering of the samples, with samples that are furthest from the centre of a cluster having a value closer to zero. The GenCall scores are ranked for all loci and the 10^{th} and 50^{th} percentiles are usually used as quality control measures (Fan *et al.* 2004, Illumina 2005). The overall call rate of samples, the call rate per locus and MAF are additional quality control measures that indicate the DNA quality of individual samples and the genotyping success of a particular locus (Pare 2010, Kijas *et al.* 2012).

The *ovine* SNP chip may be a useful genotyping tool for South African sheep research and for selection strategies related to commercial farming. However, the success of whole-genome SNP studies is reliant on sufficient numbers of polymorphic SNPs spaced throughout the genome of the test population (Kijas *et al.* 2009). It is therefore necessary to determine whether South African sheep breeds are polymorphic for sufficient numbers of SNPs included in the OvineSNP50 chip. The potential ascertainment bias of markers included in the OvineSNP50 chip and the potentially unique SNP profile of relatively rare African sheep breeds and subtypes further supports evaluating the use of the *ovine* chip in South African sheep breeds. The aim of the current study was to determine the utility of the OvineSNP50 chip for four South African sheep breeds; the Blackheaded Dorper, Namaqua Afrikaner, South African Merino and the South African Mutton Merino (SAMM) by determining the number of loci meeting quality control measures, the number of polymorphic loci and MAF distribution. The Merino, SAMM and Dorper were chosen for the study as they represent the most important wool, dual-purpose and meat breeds in the country (Cloete & Olivier 2010). The Namaqua Afrikaner was included as it represents the indigenous fat-tailed sheep of South Africa (Qwabe *et al*. 2013).

3.2. Materials and Methods:

3.2.1. Samples:

Two separate sets of genotyping were performed; the first set comprised 112 samples of South African Merino sheep from an experimental flock maintained at Elsenburg (see Chapters 1 and 2 for full details on the flock). The second genotyping set included 20 Dorper, 20 Namaqua Afrikaner, 20 SAMM and a 100 additional South African Merino sheep samples. The Dorper, Namaqua Afrikaner and SAMM samples were from a resource flock on the west coast of the Western Cape Province of South Africa at the Nortier Research Farm near Lambert's Bay. The additional South African Merino samples were obtained from the resource flocks maintained at Cradock and Grootfontein in the Eastern Cape Province and 50 samples from each location were included in the current study (see Chapter 1 for a full description of the flocks).

Samples from the Elsenburg Merino flock were selected for SNP genotyping by considering the available pedigree and best linear unbiased prediction (BLUP) estimated breeding values (EBV) based on between two and five own reproduction events in ewes. EBVs were calculated using a linear mixed model maximum likelihood method implemented in ASREML (Gilmour *et al.* 1995). Individuals that were least related to other selected individuals, had the least inbreeding (calculated using the method described by Quaas (1976)) and represented the extremes of the EBV range for number of lambs weaned (NLW) and total weight of lamb weaned (TWW), were selected. The latter selection criteria was included to ensure that methods aimed at identifying differences in allele frequencies and genomic signatures of selection could be implemented in subsequent studies (see Chapters 4 and 5). Due to the relatively small number of rams used in the breeding program, all rams born between 2002 and 2010 that contributed to the next generation were included and amounted to 27 samples. The 85 Elsenburg Merino ewes included in the study were also born between 2002 and 2010 (see Table 3.1 for the sample structure). The pedigree structure and relatedness were visualised in Pedigraph (Garbe & Da 2008) and some closely related individuals were excluded in this manner. Samples for the second genotyping set were selected based on records for NLW and TWW from at least two to five reproductive opportunities.

Table 3.1: The total number of samples in the first genotyping set grouped according to sex and selection line.

	Ewes	Rams	Total
High line	45	19	64
Low line	40	8	48
Total	85	27	112

3.2.2. Sample collection and genotyping:

Blood samples were collected by a trained veterinarian or veterinary technician from the jugular vein using a Vacuette blood collection system with 16 gauge hypodermic needles and EDTA blood tubes (Greiner bio-one, Germany). Ethical clearance was obtained from Stellenbosch University and from the Western Cape Department of Agriculture's Ethics Committee for Research on Animals (DECRA reference number R11/35). The first set of blood samples, from the Elsenburg Merinos, were stored at -20°C for a prolonged period and underwent at least one freeze-thaw cycle during this time. The samples in the second genotyping set were collected since 2007 and stored between -20°C and -80°C and only thawed for sample preparation for the current study. Once thawed, samples (from both batches) were applied to bloodcards and forwarded to GeneSeek Inc. (Lincoln, NE, USA) for genotyping using the OvineSNP50 BeadChip.

3.2.3. Genotyping quality assessment:

GenomeStudio Software version 1.0 (Genotyping Module, Illumina) was utilised to call genotypes from SNP intensity data and to examine the clustering of individual loci as well as to ensure the stringency of quality control parameters. SNPs with poor performance were identified by filtering for loci with low quality control measures such as those for cluster separation and call frequency. Loci exhibiting very low MAF were also investigated as were loci with low clustering separation scores on the X-chromosome in male individuals. Males should be homozygous for loci on the X-chromosome due to a single copy of the X-chromosome being present. Loci on the X-chromosome in males that appeared to be heterozygous were treated with suspicion and manually reclustered if they did not correlate with the standard clustering or

excluded if their quality measures did not improve. Following the initial evaluation of the genotype quality control, loci that passed the following quality control measures were included for further analyses: >0.25 GenCall score; >0.5 GenTrain score; >0.01 MAF; >0.85 call rate and a sample call rate >0.85. These quality control cut-offs were chosen to ensure data quality and are similar to those suggested by the manufacturer's protocol and those used in other published studies (Illumina 2010b, Miller *et al*. 2011, Zhao *et al*. 2011, Kijas *et al*. 2012, Therkildsen *et al*. 2013, Våge *et al*. 2013). Genotype data that met the necessary quality control criteria were used to determine the number of polymorphic loci and the MAF of loci for each sampling group.

3.3. Results:

3.3.1. First genotyping set:

By ranking ewes according to their EBVs for NLW and TWW, both expressed per ewe mated, the ewes with the highest and lowest estimated breeding values, respectively, could be identified out of approximately 1400 ewes available in the Elsenburg Merino database. Due to the divergent selective breeding of this flock (described in Chapters 1 and 2), the animals with the highest EBVs belonged to the High line, while the lowest values were found in animals belonging to the Low line. It should also be emphasised that individuals with the highest degree of inbreeding according to pedigree records were excluded as well as animals that were highly related to other individuals in the group. The mean (SD) for the High line ewes for EBVs for NLW, TWW and inbreeding amounted to 0.32 (0.08); 8.56 (1.46) and 0.03 (0.01), respectively. In the Low line ewes the mean (SD) of the EBVs for NLW, TWW and inbreeding were -0.12 (0.07); -3.00 (1.80) and 0.03 (0.02), respectively.

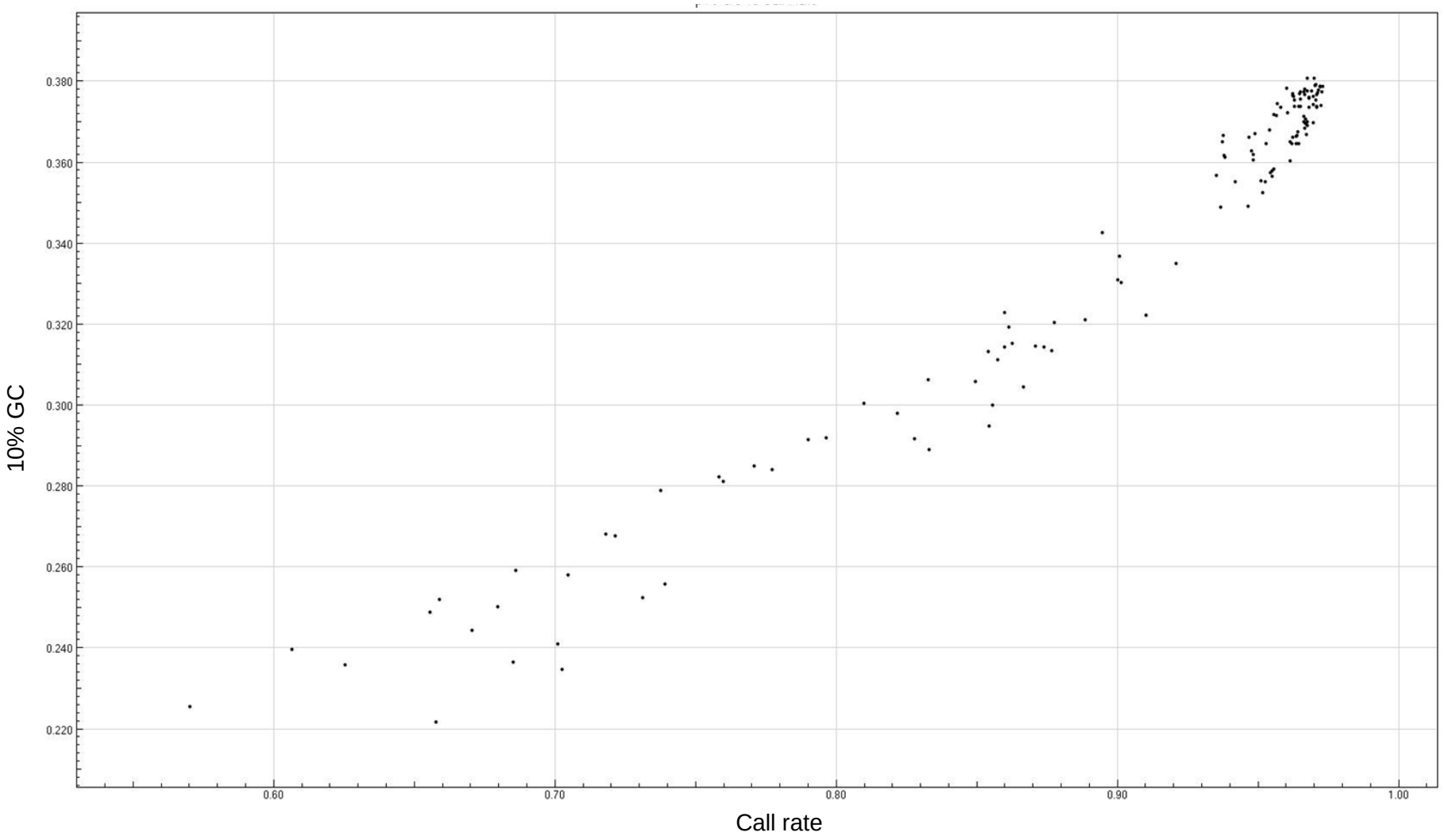


Figure 3.1: A scatter-plot depicting the 10% GenCall score plotted against the call rate of individual samples for the first genotyping set. Samples that cluster at the top right of the graph performed well, while samples that have a low call rate and low call reliability are at the bottom left of the diagonal distribution.

Initial investigation of the genotype call quality yielded evidence of several samples with poor performance. By plotting the 10% GenCall score against the call rate of individual samples, samples with low GenCall scores as well as low call rates could be identified (on the lower left in Figure 3.1). These samples were considered for reprocessing or exclusion (Illumina 2010b).

Examples of SNP genotype graphs used for additional manual quality control are provided in Figure 3.2. Successfully genotyped loci can be seen at Figure 3.2A and Figure 3.2B. In contrast, loci in Figure 3.2C and Figure 3.2D were excluded from analyses due to ambiguous clustering. The locus in Figure 3.2B is on the X-chromosome and therefore male individuals (cyan dots) are only present within the two homozygous clusters. Black dots represent samples that could not be called and can be seen in Figure 3.2C and Figure 3.2D.

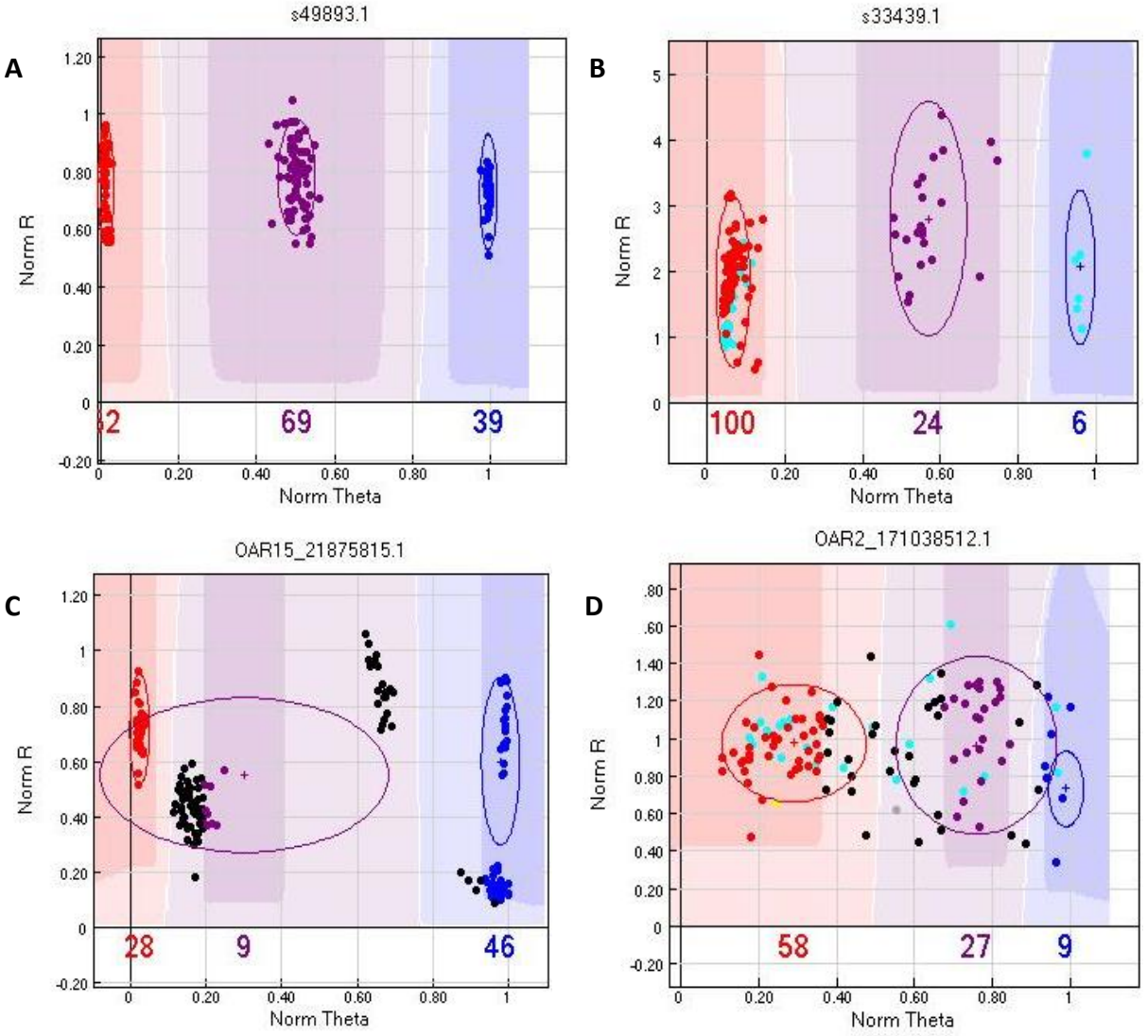


Figure 3.2: SNP genotype graphs with the normalised signal intensity on the y-axis plotted against the normalised angle deviation from pure signals on the x-axis for individual loci. Shaded areas (pink, purple and blue) correspond to the three possible genotypes (homozygous for A, heterozygous (AB) and homozygous for B) and are defined by the GenCall score.

A total of 112 individual samples from the Elsenburg Merino flock were genotyped, of which 91 samples met quality control measures (Table 3.2). These remaining 91 samples had a mean call rate of 95%. A total of 23 781 SNP loci from a possible 54 241 met the quality control measures and exhibited a mean and median MAF of 0.24.

Table 3.2: The number of samples in the first genotyping set that met quality control measures grouped according to sex and selection line.

	Ewes	Rams	Total
High line	38	16	54
Low line	30	7	37
Total	68	23	91

3.3.2. Second genotyping set:

A considerable improvement in the overall sample performance was evident for the second genotyping set. A plot of the 10% GenCall score against the call rate demonstrated that most samples had a call rate close to a 100% (Figure 3.3). All the samples from the second genotyping set met quality control measures and had an average call rate of 98%. A total of 49 517 SNPs, 91% of SNPs on the chip, met the quality control measures and had a MAF of >0.01. The Namaqua Afrikaner, SAMM, Dorper, Grootfontein Merino and Cradock Merino were polymorphic for 71%, 82%, 83%, 84% and 87% of the SNPs, respectively. To compare the current study results with those of the animal resource of the Sheep HapMap project (www.sheephapmap.org) presented in the manufacturer's guidelines (Illumina 2010a), the MAF minimum was raised to >0.05. Implementing the aforementioned MAF cut-off resulted in a total of 47 942 SNPs, or 88% of the SNPs on the chip, meeting quality control measures (Table 3.3). The number of polymorphic loci ranged from 34 448 in the Namaqua Afrikaner to 44 936 in the Grootfontein Merino.

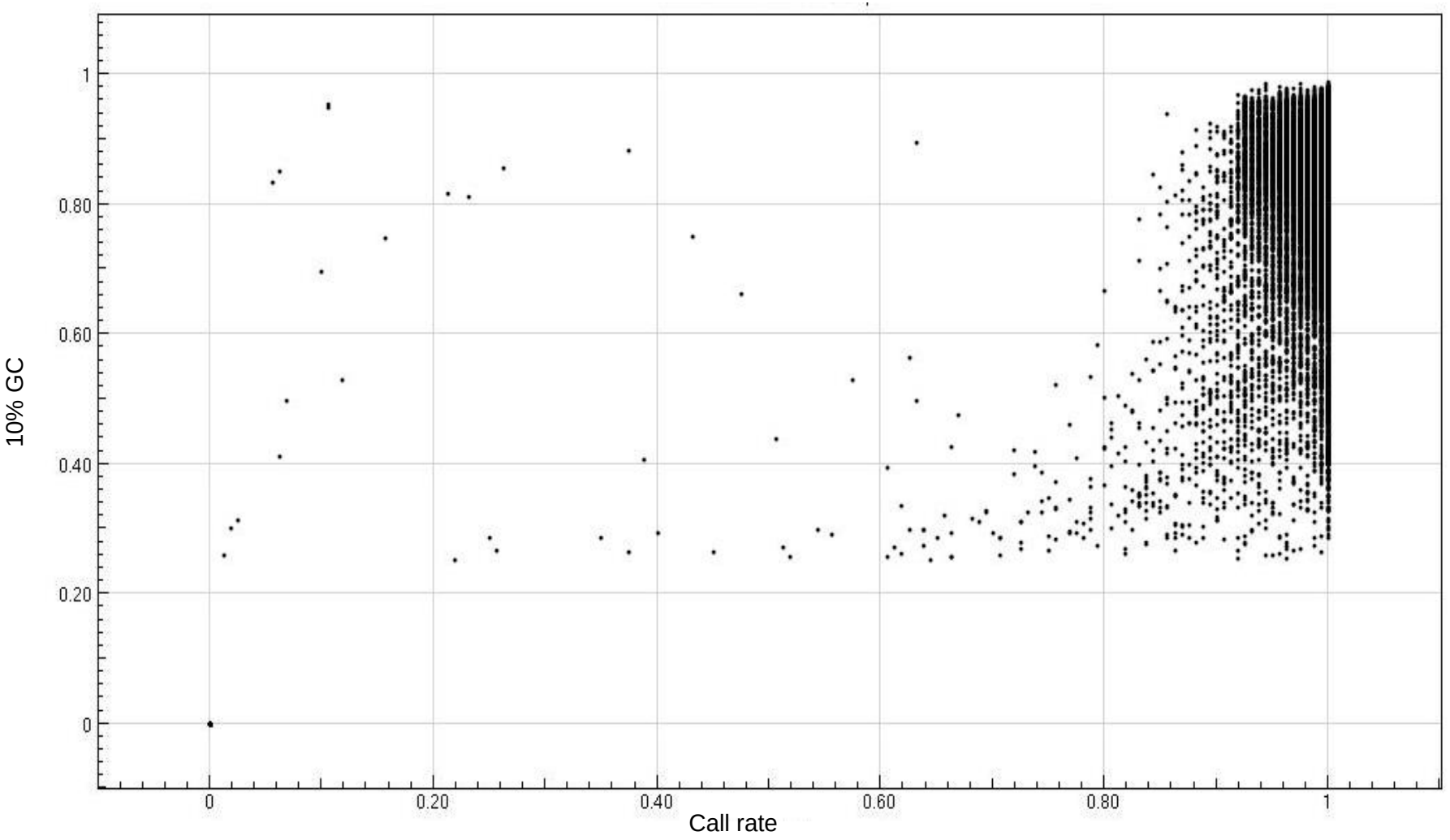


Figure 3.3: A scatter-plot depicting the 10% GenCall score plotted against the call rate of individual samples for the second set of genotyping. Most samples performed well and can be seen in the top right corner of the graph.

Table 3.3: The number of polymorphic loci and minor allele frequencies of the respective sample groups in the second genotyping set.

Breed	n	Polymorphic loci (MAF>0.05)	Polymorphic loci (%)	Mean MAF	Median MAF
Namaqua Afrikaner	20	34 448	64	0.19	0.18
Dorper	20	41 352	76	0.23	0.23
SAMM	19	41 948	77	0.23	0.24
Cradock Merino	50	44 595	82	0.27	0.27
Grootfontein Merino	50	44 936	83	0.26	0.27
Total	160	47 942	88	0.27	0.29

n: Number of samples

A total of 98% of loci were polymorphic (MAF>0) and 24% of loci exhibited a MAF of between 0.4 and 0.5 across all breeds in the second genotyping set (Figure 3.4). The percentage of SNP loci with a MAF equal to zero ranged between 3% and 23% depending on the breed surveyed in the second genotyping set. The Namaqua Afrikaner exhibited the largest number of loci with a MAF equal to zero (23%), while the Merino exhibited considerably less (3%). A difference of >12% could be observed between the number of non-polymorphic loci of the Namaqua Afrikaner and the three other breeds. Most SNP loci exhibited a MAF of between 0.4 and 0.5 in the Merino and Dorper breeds, while the most SNP loci exhibited a MAF of between 0.3 and 0.4 in the SAMM. A fairly equal distribution of the number of loci across the MAF range ($0<MAF\leq0.5$) was observed for the Namaqua Afrikaner.

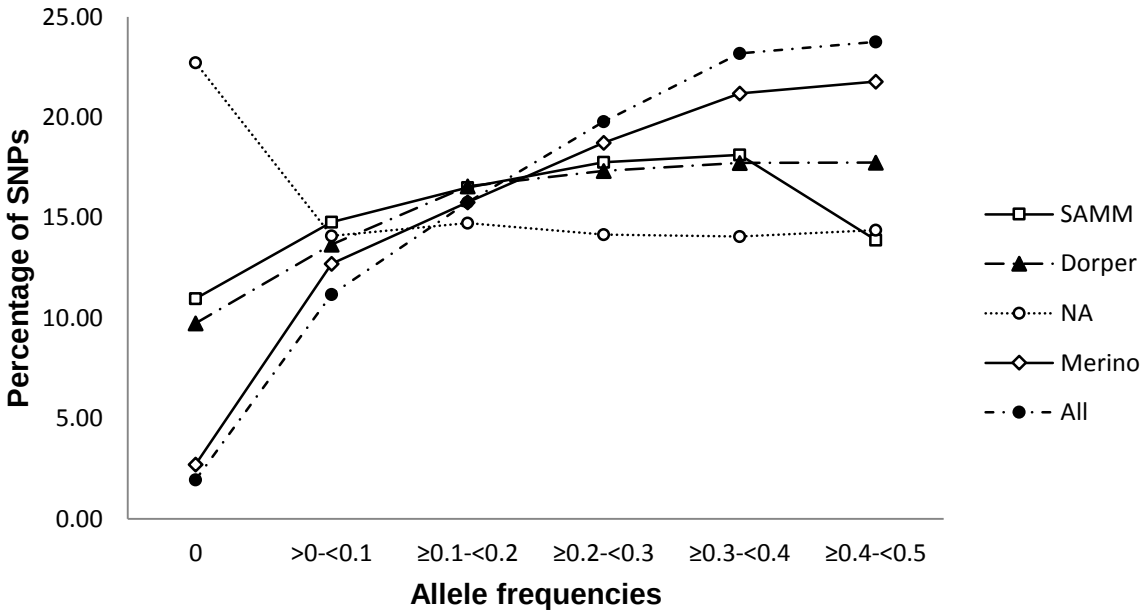


Figure 3.4: Frequency distribution of the SNP alleles for the four breeds tested in the second genotyping set as well as a combined sample.
Merino: Cradock and Grootfontein Merino samples; SAMM: South African Mutton Merino; NA Namaqua Afrikaner; All: Combined sample.

3.4. Discussion:

The current study evaluated OvineSNP50 BeadChip genotyping results of South African sheep breeds by examining the number of polymorphic markers, MAF distribution and potential SNP ascertainment bias across two separate genotyping sets. Samples from the Blackheaded Dorper, Namaqua Afrikaner, South African Merino and SAMM breeds were included to represent the most important commercial and indigenous breeds in South Africa. The first genotyping set comprised of samples from the High and Low lines of the Elsenburg Merino flock, while the second genotyping set comprised of Merino, Dorper, SAMM and Namaqua Afrikaner samples from the Grootfontein, Cradock and Nortier Research Farms.

A call rate cut-off of 98% is frequently used for microarray-based genotype results (Demars *et al*. 2013, Kijas *et al*. 2012, Liu *et al*. 2013, Våge *et al*. 2013) however, samples and loci with >85% call rate were included in the current study. As will be discussed below, poor DNA quality may have impacted genotype results from the first genotyping set and therefore a less stringent call rate cut-off value needed to be employed. Nonetheless, the value used in the current study was still more stringent

than some call rate cut-off values reported in literature. Therkildsen *et al.* (2013), for example, genotyped samples suspected of having poor DNA quality with the GoldenGate (Illumina) array-based assay and only excluded samples with a call rate <50%. Zhao *et al.* (2011) reports a call rate cut-off value of >80% for *ovine* samples genotyped with the OvineSNP50 chip. Other quality control measures, such as the GenTrain score and GenCall score, used in the current study, were similar to or exceeds the values suggested by the manufacturer's protocol (Illumina, 2010b) and are consistent with the values used in similar *ovine* studies (Miller *et al.* 2011, Zhao *et al*. 2011, Kijas *et al.* 2012, Våge *et al.* 2013).

Implementing genotype quality control measures resulted in 66% of the SNP loci and 19% of the samples being excluded from the first genotyping set, while only 9% of the loci and none of the samples in the second genotyping set were excluded. A marked difference in the number of loci and samples meeting quality control measures were evident across the two genotyping sets. The Merino results obtained from the second genotyping set were similar to those of other studies on Merino sheep where >80% of SNP loci were polymorphic and met quality control measures (Kijas *et al*. 2009, Kijas *et al.* 2012). Poor DNA quality may be responsible for the low quality of calls and the subsequent exclusion of samples and loci from the first genotyping set. The samples used in the first genotyping set were collected from 2002, stored for an extended period of time and were thawed at least once during this time. This process of freeze-thawing may have resulted in degradation of the DNA contained within the blood samples. All samples in the second genotyping set remained frozen at between -20˚C and -80˚C since collection in 2007. These samples all met quality control measures and exhibited a mean call rate of 98%. A total of 91% of loci could be included for further analyses. Studies utilising the OvineSNP50 chip to genotype commercial breeds reported more than 98% and 90% for sample call rate and SNP loci call rate, respectively (Shariflou *et al*. 2012, Våge *et al.* 2013, Phua *et al.* 2014). The results from the second genotyping set therefore compared favourably to those recently reported in literature.

While the DNA quality of samples in the first genotyping set were poor, the samples and calls with acceptable quality control measures are reliable and the samples and calls remaining after quality control could be used in further analysis. The exclusion of

SNP loci with poor data quality resulted in an arbitrary loss of loci throughout the genome and therefore it is unlikely that additional bias would have been incorporated into the data in this way. As a result of the poor genotype call quality the quantity of data obtained from the first genotyping set was considerably reduced in comparison with the second genotyping set and other *ovine* studies using the 50K SNP chip. The whole-genome marker coverage obtained by the remaining markers is nonetheless of great value and has been obtained in a more time and cost-effective manner than traditional PCR-based SNP genotyping (Meadows *et al*. 2008, Fan *et al*. 2010). Magee *et al*. (2010) reported that the inclusion of a whole-genome amplification step may be beneficial when using forensic, ancient and archival sheep DNA samples for whole-genome SNP genotyping and in future this approach may be considered when genotyping samples with poor DNA quality. Additional studies based on the current data set may consider increasing the quality control cut-off values for individual samples and thereby exclude more samples. The exclusion of samples with a lower call rate may result in an overall increase in the SNP call rates and call quality parameters and thereby increase the number of SNP loci that can be included in downstream analyses. However, the overall aim of the study needs to be considered before excluding individual samples for the potential gain in SNP loci numbers.

Although Namaqua Afrikaner samples from five individuals were included in SNP discovery for the *ovine* SNP array (Kijas *et al.* 2012), this breed exhibited the lowest number of polymorphic loci in the second genotyping set and also had the lowest overall MAF. Kijas *et al.* (2012) reported 68% polymorphic loci for Namaqua Afrikaner samples used in their study when excluding loci with MAF≤0.01. When employing the same MAF cut-off, a total of 71% SNPs were found to be polymorphic for the Namaqua Afrikaner in the current study. Information provided by the manufacturer's protocol regarding the Sheep HapMap resource flock reports that Namaqua Afrikaner samples had 62% polymorphic loci, with a mean MAF of 0.15 and a median MAF of 0.18 when excluding loci with a MAF≤0.05 (Illumina 2010a). These values are comparable to those generated in the current study where 64% of the loci were polymorphic, and a mean MAF of 0.19 and a median MAF of 0.18 are reported. The percentage of polymorphic loci and MAF values presented here are slightly higher than reported in the aforementioned studies and can be attributed to a difference in the allele frequencies between the sample cohorts of the respective studies. The Namaqua

Afrikaner samples forming part of the animal resource of the Sheep HapMap project used by Illumina (2010a) and the study by Kijas *et al.* (2012) were obtained from animals originally imported from Africa and maintained as a closed flock in Western Australia (Personal communication, Kijas JW, April 2014). It can therefore be assumed that these 12 to 17 Namaqua Afrikaner individuals were not necessarily representative of the MAF distribution of all Namaqua Afrikaner individuals and thus differ with regard to the number of polymorphic loci and MAF from the Nortier Namaqua Afrikaner samples used in the current study.

The South African Merino and another Merino-based breed, the SAMM, exhibited the largest number of polymorphic markers and had the highest overall MAF. This is not surprising as the Merino was included in all three sequencing approaches used for SNP discovery for the 50K chip (Kijas *et al*. 2012). Accordingly, the Merino and other closely related breeds, such as the Australian Merino and Merinolandschaf, have been shown to be polymorphic for >80% of the SNP loci included on the chip (Illumina 2010a, Kijas *et al*. 2012). Of the three Merino groups (Cradock, Grootfontein and SAMM) included in the second genotyping set, the SAMM had the lowest number of polymorphic loci and MAF; the values for the SAMM being fairly similar to that of the Dorper. Although the SAMM has been derived from the German Merino, which is also a Merino type, it has been extensively selected for hardiness and adaptability to the local climate and this may have resulted in a change in allele frequencies and divergence from other Merino breeds.

The lower number of polymorphic loci and lower MAF in the Dorper (in relation of the results of the Merino) may be related to the fact that this breed was not well represented during SNP discovery. The Dorper was originally developed from the Blackheaded Persian and Dorset Horn breeds (Milne 2000). The Poll Dorset, a breed closely related to the Dorset Horn, was included in all three SNP discovery methods. However, Dorset Horn samples have been shown to have fewer polymorphic loci and a lower MAF than the Poll Dorset (Illumina 2010a, Kijas *et al.* 2012), thereby indicating a difference in the SNP profiles of these two breeds. It should also be considered that the Dorper, Namaqua Afrikaner and SAMM samples included in the current study may not necessarily be representative of the SNP frequencies of these breeds throughout South Africa and the rest of the world. The number of animals included in the current

study was limited to 20 per breed and therefore may not necessarily be an exhaustive resource of the genetic stock present globally. Also, these samples were sourced from animals maintained as resource flocks and therefore may not be representative of all commercial flocks. In contrast, the Cradock and Grootfontein Merino flocks have routine exchange of genetic stock with the national Merino flock and should therefore be a relatively good representation of the polymorphisms and allele frequencies present in the national flock. The number of samples from these two flocks is greater than those from the other breeds, thereby further increasing the possibility of thorough representation of the commercial South African Merino.

The MAFs of the Dorper, Merino and SAMM exhibited a distribution pattern fairly similar to those seen for most European and American breeds (Kijas *et al.* 2012). The majority of SNP loci exhibited a MAF>0.3 for the three aforementioned breeds and loci with rare alleles were in the minority. In the case of the Namaqua Afrikaner, however, SNP loci with rare alleles were more frequent than in the other three breeds and in fact, the Namaqua Afrikaner samples exhibited an equal distribution of loci across the MAF range ($0<MAF\leq0.5$). The percentage of non-polymorphic loci in the Namaqua Afrikaner also exceeded all other breeds. A greater percentage of SNP loci with low MAFs, and less polymorphic loci have also been observed in BovineSNP50 (Illumina) genotype results of African and indicine cattle breeds (Matukumalli *et al.* 2009). The limited number of African and indicine individuals and breeds utilised during SNP discovery have been identified as underlying the aforementioned genotype results. Matukumalli *et al.* (2009) therefore reports the BovineSNP50 chip's utility in European and other commercial taurine breeds, but warns of reduced power in African and indicine breeds. Similarly, the OvineSNP50 chip will be of greater use in commercial and/or American and European sheep breeds than in African breeds.

Kijas *et al.* (2012) employed a linkage disequilibrium (LD) pruning technique aimed at compensating for SNP ascertainment bias, thereby allowing the unbiased interpretation of the OvineSNP50 genotype data (López Herráez *et al.* 2009). This was accomplished by excluding SNPs in LD within specified genomic windows from downstream analyses. This was repeated for the full set of SNPs included in the chip, SNP subsets partitioned according to their discovery method and for SNPs that were found to be polymorphic in non-domestic sheep breeds. Kijas *et al.* (2012) reported

that this approach did not significantly change the outcome of the OvineSNP50 data used for population genetic estimates, such as estimates of genetic distance (Reynold's distance) and principle component analysis. However, the ascertainment bias introduced by the SNPs included in the chip still has to be considered to interpret results effectively. The ascertainment bias contained in the *ovine* SNP chip could result in the misinterpretation of data relating to indigenous South African breeds (Kijas *et al*. 2009). However, this technology still provides immense potential for the use of genomic technology in the four South African breeds tested and provides the opportunity to gain a substantial quantity of genomic information in a relatively rapid and inexpensive manner, even though the number of polymorphic SNPs may be less in rare indigenous breeds such as the Namaqua Afrikaner.

3.5. Conclusions:

The DNA integrity of samples destined for OvineSNP50 chip genotyping needs to be ensured to attain high quality SNP genotype calls. SNP ascertainment bias was evident in the frequency distribution of minor alleles and the percentage of polymorphic loci that was observed across the four South African sheep breeds investigated in this study. This most likely stems from the selection of SNPs incorporated into the SNP chip as well as the limited number of African breeds included in SNP discovery. Although this will not necessarily compromise genotyping results or affect conclusions drawn from these results, it is nonetheless important to recognise the effect ascertainment bias may have on inferences relating to population genetic estimates or genomic selection strategies. The number of polymorphic loci and the MAF distribution of the South African Merino and SAMM seem to be comparable to those of international Merino breeds. The results of the Dorper followed the same trend as those of the South African Merino and SAMM. Therefore, the utility the OvineSNP50 chip for genotyping Dorper, South African Merino and other Merino-type breeds, has been demonstrated. However, studies utilising the chip to genotype indigenous breeds or breeds not extensively included in SNP discovery, may be hampered by the effects of SNP ascertainment bias, as was evident from the Namaqua Afrikaner results in the current study. The number of informative SNP loci may also be less in indigenous breeds in comparison to commercial breeds. Scope exists for further studies examining the use of the *ovine* SNP chip in additional South African sheep breeds as

well as the characterisation of breed diversity of commercial and indigenous South African sheep using whole-genome SNP data.

3.6. Acknowledgements:

Those responsible for the husbandry, care and sampling of the Cradock, Elsenburg, Grootfontein and Nortier resource flocks are gratefully acknowledged. The EBVs and inbreeding coefficient used to select individuals for inclusion in the current study was calculated by Prof. Schalk Cloete.

3.7. References:

Albrechtsen A, Nielsen FC, Nielsen R (2010) Ascertainment biases in SNP chips affect measures of population divergence. Molecular Biology and Evolution 27: 2534-2547

Baloche G, Legarra A, Sallé G, Larrogue H, Astruc J-M, Robert-Granié (2014) Assessment of accuract of genomic predction for French Lacaune dairy sheep. Journal of Dairy Science 97: 1107-1116

Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R (2005) Ascertainment bias in studies of human genome-wide polymorphism. Genome Research 15: 1496-1502

Cloete SWP, Olivier JJ (2010) South African sheep and wool industry. In: Cottle DJ (ed.) The international sheep and wool handbook, pg. 95-112. Nottingham University press, Nottingham, England

Daetwyler HD, Hickey JM, Henshall JM, Dominik S, Gredler B, Van der Werf JHJ, Hayes BJ (2010) Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. Animal Production Science 50: 1004-1010

Demars J, Fabre S, Sarry J, Rossetti R, Gilbert H, Persani L, Tosser-Klopp G, Mulsant P, Nowak Z, Drobik W, Martyniuk E, Bodin L (2013) Genome-wide association studies identify two novel *BMP15* mutations responsible for an atypical hyperprolificacy phenotype in sheep. Public Library of Science Genetics 9: e1003482

Fan B, Du Z-Q, Gorbach DM, Rothschild MF (2010) Development and application of high-density SNP arrays in genomic studies of domestic animals. Asian-Australian Journal of Animal Science 2: 833-847

Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, Hansen MS, Steemers F, Butler SL, Deloukas P, Galver L, Hunt S, McBride C, Bibikova M, Rubano T, Chen J, Wickham E, Doucet D, Chang W, Campbell D, Zhang B,

Kruglyak S, Bentley D, Haas J, Rigault P, Zhou L, Stuelpnagel J, Chee MS (2004) Highly parallel SNP genotyping. Cold Spring Harbor Symposia on Quantitative Biology LXVIII: 69-78

Garbe, JR, Da Y (2008) Pedigraph: A software tool for the graphing and analysis of large complex pedigrees. User manual Version 2.4. Department of Animal Science, University of Minnesota, USA

Gilmour AR, Thompson R, Cullis BR (1995) Average information REML, an efficient algorithm for variance parameter estimation in linear mixed models. Biometrics 51: 1440-1450

Gutiérrez-Gil B, Arranz JJ, Pong-Wong R, García-Gámez E, Kijas J, Wiener P (2014) Application of selection mapping to identify genomic regions associated with dairy production in sheep. Public Library of Science ONE: e94623

Heslot N, Rutkoski J, Poland J, Jannink J-L, Sorrells ME (2013) Impact of marker ascertainment bias on genomic selection accuracy and estimates of genetic diversity. Public Library of Science ONE 8: e74612

Illumina (2005) GenomeStudioTM Genotyping Module v1.0 User Guide, pg. 1-168. Accessed January 2014, from: www.illumina.com

llumina (2010a) OvineSNP50 genotyping BeadChip. Data sheet: DNA analysis. Accessed January 2014, from: www.illumina.com

Illumina (2010b) Infinium genotyping analysis. Technical note: DNA analysis. Accessed January 2014, from: www.illumina.com

International Sheep Genomics Consortium, Archibald AL, Cockett NE, Dalrymple BP, Faraut T, Kijas JW, Maddox JF, McEwan JC, Hutton Oddy V, Raadsma HW, Wade C, Wang J, Wang W, Xun X (2010) The sheep genome reference sequence: A work in progress. Animal Genetics 41: 449-453

Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, Tang D, McEwan J, Cockett N, Oddy HV, Nicholas FW, Raadsma H, for the International Sheep Genomics Consortium (2009) A genome-wide survey of SNP variation reveals the genetic structure of sheep breeds. Public Library of Science ONE 4: e4668

Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J, Dalrymple B, other members of the International Sheep Genomics Consortium10 (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. Public Library of Science ONE Biology 10: e1001258

Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Zhang S, Zhou X, Wei C, Zhao F, Du L (2013) Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. BioMed Central Genomics 14: 229

López Herráez D, Bauchet M, Tang K, Theunert C, Pugach I (2009) Genetic variation and recent positive selection in worldwide human populations: evidence from nearly 1 million SNPs. Public Library of Science ONE 4: e7888

Magee DA, Park SDE, Scraggs E, Murphy AM, Doherty ML, Kijas JW, International Sheep Genomics Consortium, MacHugh DE (2010) Technical note: High fidelity of whole-genome amplified sheep (*Ovis aries*) deoxyribonucleic acid using a high-density single nucleotide polymorphism array-based genotyping platform. Journal of Animal Science 88: 3183-3186

Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, OÇonnell J, Moore SS, Smith TPL, Sonstegard TS, Van Tassell CP (2009) Development and characterization of a high density SNP genotyping assay for cattle. Public Library of Science ONE 4: e5350

Meadows JRS, Chan EKF, Kijas JW (2008) Linkage disequilibrium compared between five populations of domestic sheep. BioMed Central Genetics 9: 61

Miller JM, Poissant J, Kijas JW (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. Molecular Ecology Resources 11: 314-322

Milne C (2000) The history of the Dorper sheep. Small Ruminant Research 36: 99-102

Morin PA, Luikart G, Wayne RK, SNP workshop group (2004) SNPs in ecology, evolution and conservation. Trends in Ecology and Evolution 19: 208-216

Nielsen R, Signorovitch J (2003) Correcting for ascertainment bias when analysing SNP data: applications to the estimation of linkage disequilibrium. Theoretical Population Biology 63: 245-255

Pare G (2010) Genome-wide association studies - data generation, storage, interpretation, and bioinformatics. Journal of Cardiovascular Translational Research 3: 183-188

Phua SH, Brauning R, Baird HJ, Dodds KG (2014) Identifying chromosomal selection-sweep regions in facial eczema selection-line animals using an ovine 50K-SNP array. Animal Genetics 45: 240-247

Quaas RL (1976) Computing the diagonal elements and inverse of a large numerator relationship matrix. Biometrics 32: 949-953

Qwabe SO, Van Marle-Köster E, Visser C (2013) Genetic diversity and population structure of the endangered Namaqua Afrikaner sheep. Tropical Animal Health and Production 45: 511-516

Shariflou MR, Wade CM, Kijas JW, McCulloch R, Windsor PA, Tammen I, Nicholas FW (2012) Brachygnathia, cardiomegaly and renal hypoplasia syndrome (BCRHS) in Merino sheep maps to a 1.1-megabase region on ovine chromosome OAR2. Animal Genetics 44: 231-233

Therkildsen NO, Hemmer-Hansen J, Als TD, Swain DP, Morgan MJ, Trippels EA, Palumbi SR, Meldrup D, Nielsen EE (2013) Microevolution in time and space: SNP analysis of historical DNA reveals dynamic signatures of selection in Atlantic cod. Molecular Ecology 22: 2424-2440

Våge DI, Husdal M, Kent MP, Klemetsdal G, Boman IA (2013) A missense mutation in *growth differentiation factor 9* (*GDF9*) is strongly associated with litter size in sheep. BioMed Central Genetics 14: 1

Zhao X, Dittmer KE, Blair HT, Thompson KG, Rothschild MF, Garrick D (2011) A novel nonsense mutation in the *DMP1* gene identified by a genome-wide association study is responsible for inherited rickets in Corriedale sheep. Public Library of Science ONE 6: e21739

Chapter 4:

Identification of SNP loci under selection in a divergently selected South African Merino flock

4.1. Introduction:

Reproduction traits are highly variable between and within sheep breeds (Celi & Bush 2010). Several highly proliferative sheep breeds and lines, such as the Booroola Merino, have litter sizes of more than 2.0 in comparison to less than 1.3 for the same trait in other Merino lines (De Graaf 2010). Multiple birth rate depends on ovulation rate and embryo survival and forms part of the composite trait, net-reproduction rate, as defined by Olivier (1999) and Snowder and Fogarty (2009). It is therefore not surprising that highly proliferative lines commonly exhibit an increased ovulation rate that underlies an increase in reproductive output (Juengel *et al*. 2013).

The Booroola Merino was the first highly proliferative line in which a single causative mutation relating to reproduction was identified (Davis *et al.* 1982, Mulsant *et al*. 2001). Ovulation rate and litter size are influenced by the autosomal dominant fecundity mutation, *FecB,* and results in an additional 1.5 ova per ovulation for each copy of the mutation in the Booroola Merino. Individuals that are homozygous for the *FecB* mutation will therefore produce 3 extra ova per ovulation. A concomitant increase in litter size of 1.0 and 1.5 occurs in *FecB* heterozygotes and homozygotes, respectively. The Booroola gene is the result of a mutation on chromosome 6 in the *bone morphogenetic protein 1B receptor* gene (*BMPR1B*). The origin of this mutation has been traced back to the Garole sheep from Asia that was introduced to Australia in the 1700's. The Garole and related breeds have been shown to be homozygous for the *FecB* mutation (Davis 2005, McNatty *et al*. 2005). Other prolific Asian breeds, such as the Javanese Thin-tail and Chinese Hu also carry this mutation (Notter 2008).

Mutations in two other genes have also been linked to an increase in sheep ovulation rate. The *FecX* and *FecG* mutations results in an extra 1.0 and 1.4 ova per ovulation, respectively (Davis 2005), although this may vary depending on the specific mutations present (Table 4.1) (Juengel *et al*. 2013). *FecX* is the result of a mutation in the *bone morphogenetic protein 15* gene (*BMP15*) on the X-chromosome; whereas the *FecG*

mutation is in the *growth differentiation factor 9* gene (*GDF9*) on chromosome 5. The *FecX* mutation was first identified in Romney sheep and named the *Inverdale* gene and has since also been found in the Belclare and Cambridge sheep lines. These two lines have also been found to carry the *FecG* mutation (Davis 2005, McNatty *et al*. 2005). Ewes that are homozygous for the two mutations exhibit abnormal ovarian development and are sterile. Breeding programmes utilising the effect of these two mutations therefore require careful planning to avoid the production of homozygous, and therefore, sterile ewes (Notter 2008).

The abovementioned three genes are part of the transforming growth factor beta (TGFB) superfamily and play a role in the regulation of ovulation in sheep. BMP15 and GDF9 are produced by the oocyte and regulate granulose cell formation, while BMPR1B is a TGFB superfamily receptor that interacts with BMP15. Several causative mutations in *BMP15* and *GDF9* and a single one in *BMPR-IB* have been linked to increases in ovulation (Table 4.1) (Juengel *et al*. 2011, 2013).

Additional lines carrying presumably unique mutations in major genes affecting ovulation rate have also been reported. The causative mutation in the Woodlands, Wishart and Davidsdale lines has however not yet been identified (Davis 2005, Juengel *et al*. 2011). The causative mutation in the case of the Woodlands line seems to be on the X-chromosome and is likely located in genes forming part of the TGFB superfamily. An interaction with *BMP15* and *BMPR1B* occurs, but the exact gene(s) and pathways involved are still not clear (Juengel *et al*. 2011). More genes and pathways involved in increased ovulation rate may therefore be identified within these highly proliferative lines.

Table 4.1: Mutations in the *BMPR1B*, *BMP15* and *GDF9* genes affecting ovulation rate in sheep.

Gene	Identified in (Line/Breed)	Allele	Coding residue	Amino acid change
BMPR1B	Booroola	*FecB*	249	Gln-Arg
BMP15	Belclare, Cambridge	*FecX^{G}*	239	Gln-STOP
	Belclare	*FecX^{B}*	367	Ser-Ile
	Inverdale	*FecX^{I}*	299	Val-Asp
	Hanna	*FecX^{H}*	291	Glu-STOP
	Lacuane	*FecX^{L}*	321	Cys-Tyr
	Rasa Arag	*FecX^{R}*	154	Frame shift
GDF9	Belclare, Cambridge	*FecG^{H}*	395	Ser-Phe
	Thoka	*Thoka*	427	Ser-Arg
	Santa Inês	*FecG^{I}*	345	Phe-Cys

Adapted from Juengel *et al.* (2013).

Divergent selection of the Elsenburg Merino flock for the ability to rear multiple offspring has resulted in a High line and a Low line that differ markedly in their reproductive output (see Chapters 1 and 2 for a detailed discussion of the Elsenburg Merino flock, Cloete *et al*. 2004a, 2007). The causative mutations and/or quantitative trait loci (QTL) responsible for the difference in reproductive traits between these two lines have, to date, not been identified. These lines could serve as a model for identifying the genomic regions underlying reproductive traits in the South African Merino and to determine whether mutations identified in the aforementioned highly proliferative lines are also segregating in this flock. The Elsenburg Merino flock does not have any direct pedigree links with any of the well-known highly proliferative lines, such as the Booroola, Inverdale or Cambridge lines. It is therefore assumed that accidental or intentional introgression of these known mutations, associated with an increased ovulation rate in other highly proliferative breeds, has not occurred in this flock.

Due to the importance of ovulation rate for increases in reproductive potential, it was relevant to investigate whether regions associated with variation in highly proliferative lines, have undergone selection within the Elsenburg Merino flock. As several mutations in single genes (*BMP15* and *GDF9*) have been shown to affect ovulation

rate, the possibility of novel mutations in *BMPR1B*, *BMP15* and *GDF9*, affecting ovulation rate in the Elsenburg Merino flock, should be considered. However, ovulation rate is a complex trait relying on the functioning and interaction of several gene products (Juengel *et al*. 2013) and therefore unidentified mutations or polymorphisms in numerous other major genes may also affect ovulation rate. The Elsenburg Merino lines do not only differ with regard to ovulation rate (as seen in the number of lambs born), but also for other components of net-reproductive rate that include production, robustness and behavioural traits (Cloete & Scholtz 1998, Cloete *et al.* 2003, 2004b, 2005a, 2005b, 2005c, 2009, 2010) and therefore numerous genes may be responsible for the differentiation seen between the lines.

In the current study, a whole-genome SNP genotyping strategy was employed to identify regions underlying the differences in reproduction potential between the Elsenburg Merino lines. The genotype data was partitioned according to the 27 *ovine* chromosomes to pinpoint chromosomal regions subject to selection and to determine whether the aforementioned genes and surrounding genomic regions may be under selection and therefore play a role in the differential reproductive output of the Elsenburg Merino flock. An F_{ST} outlier method was used to identify SNPs subject to selection by respectively applying Bayesian and frequentist approaches to the data.

4.2. Materials and Methods:

4.2.1. Sample selection:

Sample selection, collection, genotyping and quality assessment are described in Chapter 3. In short, 112 samples (Table 4.2) were selected from the Elsenburg Merino flock from individuals with fairly accurate BLUP estimated breeding values for number of lambs weaned per parity (NLW) and total weight of lamb weaned per parity (TWW). The samples selected were from animals born between 2002 and 2010 and represented the extremes of the phenotypic distribution for NLW and TWW. Pedigree information was considered to minimise the inbreeding and relatedness in the sampling cohort and to reduce substructure in the samples. A t-test was performed to confirm whether significant differences exist between the phenotypes of the two lines.

Table 4.2: The total number of samples grouped according to sex and selection line.

	Ewes	Rams	Total
High line	45	19	64
Low line	40	8	48
Total	85	27	112

4.2.2. Sample collection and genotyping:

Blood samples were stored at -20°C for a prolonged period and were thawed at least once during this time. Samples were also thawed for application to bloodcards that were forwarded to GeneSeek Inc. (Lincoln, NE, USA) for genotyping using the OvineSNP50 BeadChip. The samples used in the current study therefore underwent at least two freeze-thaw cycles.

Samples and loci included for further analyses had a >0.25 GenCall score; >0.5 GenTrain score; >0.01 MAF; >0.85 call rate and a sample call rate >0.85. These quality control cut-offs were selected to ensure data quality and are similar to those used in other *ovine* SNP chip studies (Illumina 2010, Miller *et al*. 2011, Kijas *et al*. 2012, Våge *et al*. 2013). It should be noted that the loci call rate and sample call rate cut-off (>0.85) was slightly lower than the values used in other studies (see Chapter 3 for a full discussion). This was done to ensure that a greater number of samples could be included in downstream analyses, thereby allowing a more representative sample of the Elsenburg Merino flock to be included in the current study.

4.2.3: Statistical analyses:

To investigate the possibility that specific chromosomal regions known to be associated with variation in ovulation rate, were under selection in the current study, the genotypic data were partitioned by chromosome. A factorial component analysis was conducted in Genetix version 4.05 (Belkhir *et al*. 2004) to assess and visualise the multi-factorial variance and genetic differences between the lines. Markers subject to selection were identified by means of two F_{ST} outlier approaches using a Bayesian and a frequentist method implemented in BayeScan version 2.1 (Foll & Gaggiotti 2008)

and Lositan version 1.0 (Antao *et al*. 2008), respectively. The F_{ST} parameter is defined as the difference in the average expected heterozygosity of the subpopulations and the expected heterozygosity of the total population. The F_{ST} value can be calculated as

$$F_{ST} = \frac{H_T - H_S}{H_T},$$

where H_T and H_S represents the expected heterozygosity of the total population and subpopulation, respectively. The F_{ST} value thus provides an indication of genetic differentiation between groups and locus-specific F_{ST} values that differ significantly from the mean are indicative of a locus under selection pressure (Holsinger & Weir 2009, Duforet-Frebourg *et al*. 2014).

It is assumed that loci unaffected by selection will exhibit a neutral pattern of variation, whereas loci under selection will have an atypical pattern. The Fdist2 method described by Beaumont and Nichols (1996) is used to identify F_{ST} outliers in Lositan. This method is based on a symmetrical island model and therefore assumes drift-migration equilibrium. Loci under selection are identified by comparing the F_{ST} value of each locus with that of the mean F_{ST} value averaged across the whole sample set (Antao *et al*. 2008, Manel *et al*. 2009). In contrast, the hierarchical Bayesian method employed in BayeScan estimates population-specific F_{ST} values. This approach therefore allows for differences in population structure and does not assume migration-drift equilibrium (Foll & Gaggiotti 2008, Manel *et al*. 2009). The default parameter settings of Lositan and BayeScan were employed and 100 000 simulated loci were generated in Lositan and 5 000 MCMC-simulated loci, based on the distribution of F_{ST} values, were generated in BayeScan. A correction for multiple testing was implemented in Bayescan and Lositan by applying a 0.1 significance threshold to the false discovery rate (q-value).

The SNPs indicated to be under selection by both methods, were investigated further. The genomic locations and neighbouring genes of these SNPs were considered for their role in reproduction. The *Ovis aries* genome (Oar_v3.1, available from http://www.ncbi.nlm.nih.gov/genome?term=ovis%20aries) was used to determine whether markers were located within annotated gene sequences while the nearest upstream or downstream location of annotated gene sequences were also noted.

4.3. Results and discussion:

4.3.1: Genotype quality control:

Ninety-one individuals (Table 4.3) and 23 780 SNPs, or 43% of the total SNPs genotyped, met the quality control criteria and were included for downstream analyses.

Table 4.3: The number of samples that met quality control measures grouped according to sex and selection line.

	Ewes	Rams	Total
High line	38	16	54
Low line	30	7	37
Total	68	23	91

The number of markers meeting quality control measures was relatively low in comparison with other *ovine* SNP studies where most loci had a call rate of more than 98% (Kijas *et al*. 2012, Demars *et al*. 2013, Liu *et al*. 2013, Våge *et al*. 2013). Possible causes of the low overall call rate are explained at length in Chapter 3. Shortly, OvineSNP50 genotyping results from other Merino lines have yielded relatively high percentages of markers meeting quality control measures. Kijas *et al*. (2012) reported that several Merino lines had >82% polymorphic loci meeting quality control measures when genotyped with the 50K chip. Although local adaptation and other population genetic trends of the Elsenburg Merino flock could have resulted in a lower number of polymorphic loci, this does not account for the extent to which the current data differs from those of other Merino lines. The most plausible explanation is that the DNA quality of the genotyped samples may have been compromised by extended storage and repeated cycles. Low DNA quality could lead to difficulty in automated scoring of SNPs during analyses of SNP chip results and would thus lead to ambiguous calls that would subsequently be excluded during quality control. Therkildsen *et al*. (2013) reported that automated SNP genotyping of historical samples of which the DNA integrity may have been compromised, yielded less useful samples and SNP loci meeting quality control measures. Indeed, genotyping of an independent set of *ovine* samples stored at -80˚C without exposure to freeze-thaw cycles, yielded much higher call rates (discussed in Chapter 3).

4.3.2. Line effects:

The mean (SD) for the High line ewes for EBVs for NLW and TWW amounted to 0.32 (0.08) and 8.56 (1.46), respectively. In the Low line ewes the mean (SD) of the EBVs for NLW and TWW were -0.12 (0.07) and -3.00 (1.80), respectively. A t-test confirmed significant differences in the phenotypic values of the two lines for NLW and TWW, consistent with the report of Cloete *et al*. (2004). The factorial component plot for each of the 26 autosomal chromosomes and the X-chromosome indicated two distinct clusters representing the divergently selected lines (Figure 4.1). The High and Low lines can therefore be considered phenotypically and genetically distinct as a result of several generations of selection.

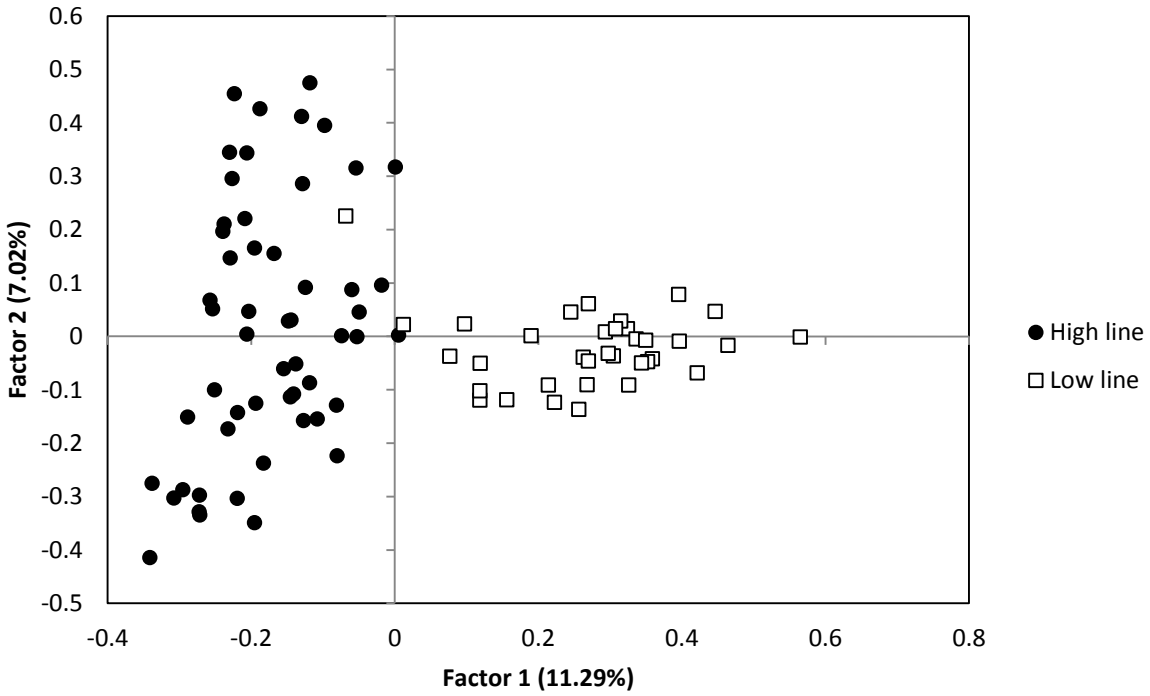


Figure 4.1: A factorial component plot of chromosome 14, indicating two distinct clusters. Dark circles represent individuals from the High line and white squares that of the Low line. Similar results were observed for all 27 *ovine* chromosomes.

4.3.3. Chromosome-specific markers under selection:

Prior to correcting for multiple testing, 46 SNPs subject to directional selection were identified by the Bayesian-based analysis, while the frequentist method indicated 1 476 markers under directional and balancing selection. These numbers were reduced to 31 SNPs and 961 SNPs, respectively, after correction for multiple testing. The overall percentage of markers identified to be subject to selection was 0.13%

using the Bayesian method and 4.04% using the frequentist method (Appendix Table A4.1). All markers identified by the Bayesian approach implemented in BayeScan were also identified by the frequentist method implemented in Lositan (Figure 4.2). The disparity between the numbers of markers identified by the two methods has been noted in other related studies (Manel *et al*. 2009, Narum & Hess 2011, Therkildsen *et al*. 2013). The Bayesian method implemented in BayeScan has been shown to be less prone to type I errors (false positives) compared to the frequentist Fdist2 method employed in Lositan. This could be the cause of the discrepancy in the number of markers identified by the two different approaches. Given that the Fdist2 method does not consider population-specific demographics, inflated F_{ST} values may be generated for some loci. This is especially true when gene flow is asymmetrical across populations. By calculating population-specific F_{ST} values, the Bayesian model compensates for population-specific artefacts and does not assume drift-migration equilibrium (Manel *et al*. 2009). As the two Elsenburg Merino lines have been reproductively isolated and may have experienced dissimilar levels of selection pressure, the Bayesian model implemented in BayeScan, may be more appropriate for this data set.

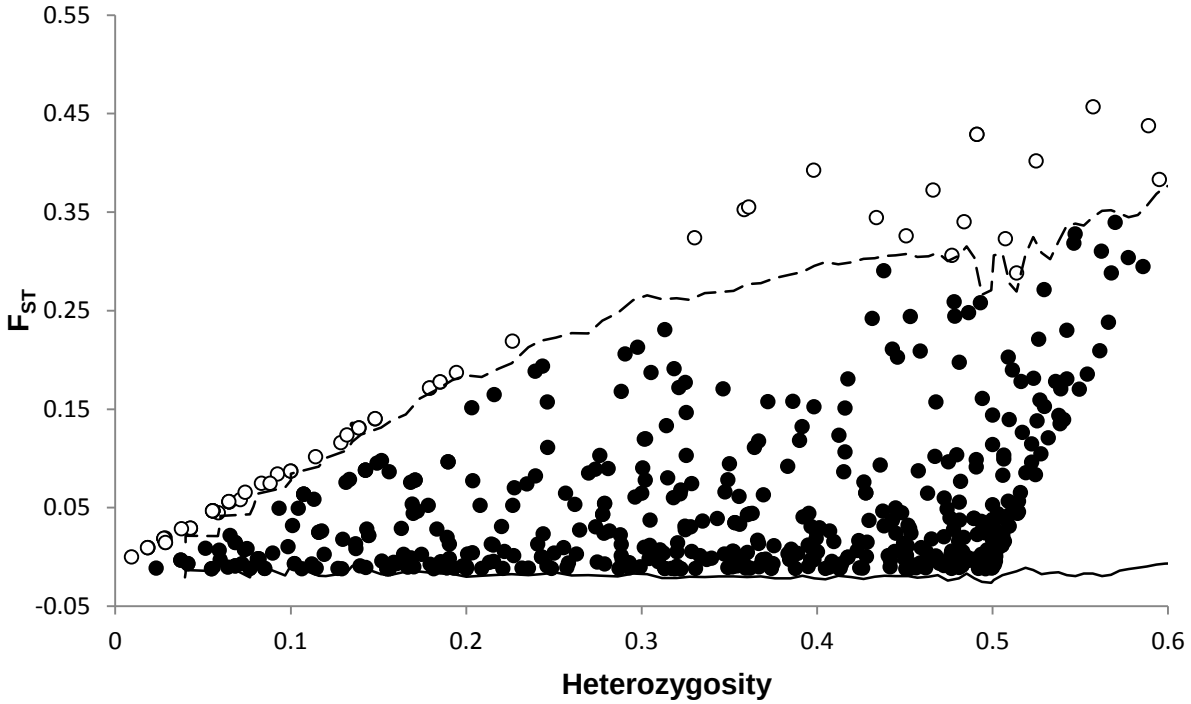


Figure 4.2: The distribution of F_{ST} values as a function of heterozygosity for SNPs located on chromosome 22 as calculated using Lositan. Loci under directional selection are represented by open circles located above the 95% percentile (dashed line) and markers under balancing selection are below the solid line.

The Fdist2 method identified the largest percentage of markers subject to directional and/or balancing selection on chromosomes 8 and 9; respectively 6.93% and 6.67%. The same method identified the smallest percentage of markers subject to directional and/or balancing selection on chromosomes 11 and 19; respectively 1.67% and 1.66%. Using the Bayesian method, chromosomes 12 and 21 were indicated to carry the largest percentage of markers under directional selection; respectively 0.37% and 0.47% (Appendix Table A4.2). Several chromosomes without any markers under balancing or directional selection were also identified (Figure 4.3, Appendix Table A4.1).

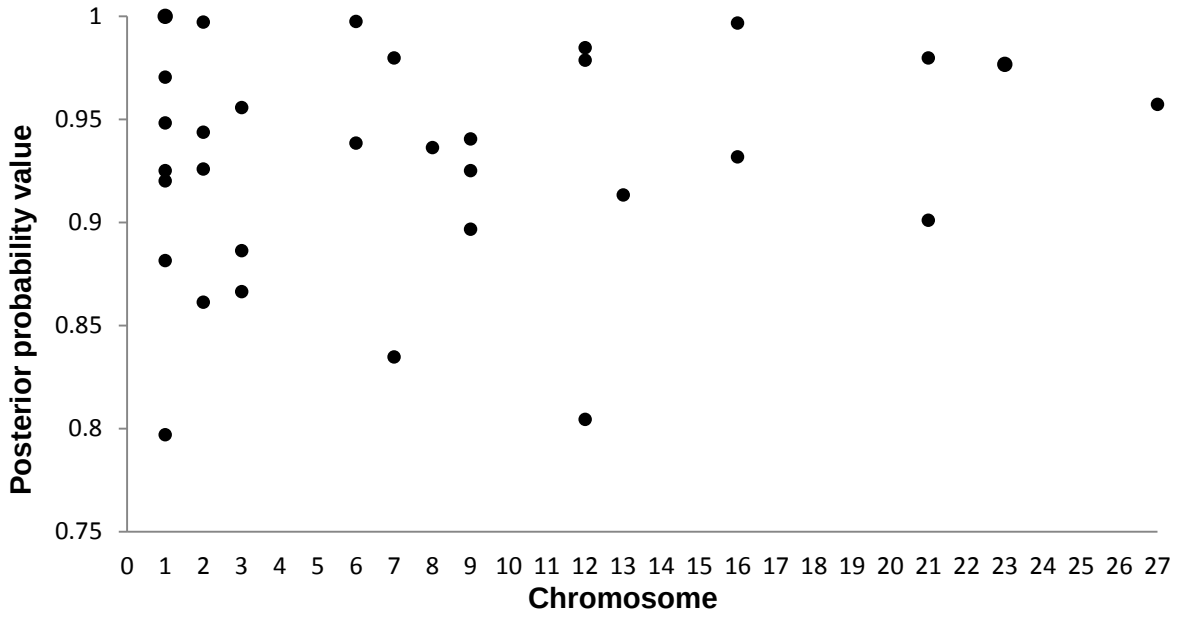


Figure 4.3: The posterior probability values generated in BayeScan for the markers across all chromosomes (X-chromosome referred to as chromosome 27). Posterior probability values, indicative of directional and/or balancing selection, ≥0.76, are shown.

The 31 markers identified by both methods were all proposed to be under directional selection. Several of these markers were found to be located in or near annotated genes. Twelve markers were located within the sequence of annotated genes and are listed in Table 4.4. The nearest gene upstream or downstream of the remaining 19 markers as well as the distance between the marker and the gene(s) has been listed in Table 4.4. The distance between the latter 19 markers and an annotated gene sequence were between 4 678bp and 4 464 080bp. The mean distance between a marker locus and gene was 389 071bp. The largest distances between markers and genes were observed in cases where the marker was located at the end of a chromosome (for example, s40378.1 on chromosome 21).

The location of the 31 markers under selection did not overlap with any of the genes associated with variation in ovulation rate on chromosome 5 and 6 and the X-chromosome identified in previous studies (Table 4.5). The markers identified on chromosome 6 and the X-chromosome are between 5 800kb and 53 000kb removed from the *BMPR1B* and *BMP15* genes. None of the markers are in close proximity to QTL associated with the total number of lambs born or ovulation rate either (http://www.animalgenome.org/cgi-bin/QTLdb/OA/index, Hu *et al*. 2013). Therefore, the F_{ST} outlier methodology used in the current study did not identify non-neutral SNP

markers in the proximity of the *BMPR1B*, *BMP15* and *GDF9* genes nor any other annotated genomic regions previously identified to be associated with the total number of lambs born or ovulation rate.

The 54 241 SNPs included on the *ovine* 50K SNP chip have a mean distance of approximately 51kb between each marker. As less than half of these markers met quality control measures for the current study, it is assumed that a mean distance of approximately 100kb therefore exists between the markers. Consequently, linkage disequilibrium between genomic regions subject to selection and SNP markers would need to extend beyond 100kb to ensure detection using the current methodology. However, linkage disequilibrium (LD) dissipates rapidly at distances >30kb for most sheep breeds (Kijas *et al*. 2014) and therefore the marker density in the current study is insufficient to identify all genomic regions subject to selection. Therefore, even though the current study could not identify markers under selection within the *BMPR1B*, *BMP15* and *GDF9* genes, this does not exclude the involvement of mutations in these genes playing a role in the variation in reproductive potential in the Elsenburg Merino flock. Further studies utilising denser SNP panels are necessary for more accurate identification of regions under selection before the involvement of mutations in the *BMPR1B*, *BMP15* and *GDF9* genes can be unequivocally excluded for this flock.

The *cylicin basic protein of sperm head cytoskeleton 2* (*CYLC2*) and the *corticotropin releasing hormone* (*CRH*) genes are located in the proximity of non-neutral markers listed in Table 4.4. Due to their apparent connection with reproduction, these two genes and their possible role in the variation of reproduction traits will be discussed further. The latter genes are 1 058 070bp and 114 082bp removed from the OAR2_19076972.1 and OAR9_4346649.1 markers, respectively. The *CRH* gene is not located adjacent to the OAR9_4346649.1 marker; in fact, the *ribosome biogenesis regulator homolog* (*Saccharomyces cerevisiae*) (*RRS1)* and *alcohol dehydrogenase, iron containing, 1* (*ADHFE1)* genes are within 47 000bp of the marker. However, considering the aforementioned mean distance between markers and the large distances (≥1 000kb) reported for putative regions under selection in other *ovine* studies (Moradi *et al.* 2012, Gutiérrez-Gill *et al.* 2014), the proximity of *CRH* may nonetheless be relevant.

The *CYLC2* gene product plays a direct role in the structural formation of the sperm head (Hess *et al*. 1995) and the involvement of this gene product in reproduction may be related to the ability of sperm to fertilise ova. A study of the morphometric sperm characteristics of the Elsenburg Merino rams indicated that the lines differed with regard to sperm head morphology (Boshoff 2014). The latter study found that the sperm head was broader and rounder in Low line individuals, while the acrosome surface was larger in High line individuals. Motility traits, such as the percentage motile sperm and the sperm velocity also differed significantly between the two lines. The aforementioned study therefore supports the involvement of male reproductive traits, specifically sperm cell characteristics, in the difference in reproductive output between the lines. Whether the *CYLC2* gene product has a functional role in female reproduction traits is uncertain at this stage.

The CRH is key to the mammalian endocrine stress response through the hypothalamo-pituitary-adrenal axis (HPAA) as well as the autonomic and behavioural stress responses by the central nervous system (Seasholtz *et al*. 2002). In mammals the hypothalamus' secretion of CRH stimulates the release of adrenocorticotropic hormone (ACTH) by the anterior pituitary, resulting in the secretion of corticosteroids, such as cortisol from the adrenal glands. The HPAA has a direct effect on stress coping ability and also influences energy metabolism, food intake, immune responses, blood pressure, anxiety, fertility, sexual libido and the ability to learn (Koob & Heinrichs 1999, Smagin *et al*. 2001, Seasholtz *et al*. 2002). The CRH's involvement in parturition in sheep has been confirmed and it is accepted that an increase in foetal CRH, induces the secretion of ACTH and results in a cortisol surge leading to the onset of parturition (Chan *et al*. 1998). The CRH is therefore a determinant of the length of gestation and timing of parturition and it is not surprising that it is highly expressed in the placental tissue (Wadhwa *et al*. 2004).

Other studies on the Elsenburg Merino flock have indicated that the High line has a superior stress coping ability in comparison to the Low line. The High line exhibited less stress indicator behaviours during an arena test with a human as a stressor (Cloete *et al*. 2010) and High line rams produced higher levels of cortisol when experiencing physiological stress by an insulin challenge (Hough 2010). High line

ewes give birth earlier in the lambing season (Cloete *et al.* 2004a, 2005c) and High line progeny also have shorter parturitions (Cloete & Scholtz 1998, Cloete *et al.* 2003). Meat quality traits, such as pH, also indicate that High line individuals experienced less stress prior to slaughter (Cloete *et al*. 2004b). The High line's higher slaughter weight (Cloete *et al.* 2005a), taken together with the aforementioned characteristics of the High line, also suggests a line that copes better with stress and therefore exhibits higher food intake and/or more efficient energy metabolism.

A study by Hough (2012) found differences in the stress response of the Elsenburg Merino lines, depending on the isoform of the P450 17-hydroxylase/17,20-lyase (*CYP17)* gene present. The *CYP17* gene product is involved in the release of glucocorticoids, such as cortisol and corticosterone that are involved in the stress coping mechanism. Deficiencies in this gene product is characterised by abnormal sexual development, such as sexual infantilism, impaired virilisation and pseudohermaphroditism (Yanase *et al.* 1991, Kater & Biglieri 1994, Yanase *et al.* 1995, Storbeck *et al*. 2008). None of the markers indicated in Table 4.4, however, are in close proximity to the *CYP17* gene located on chromosome 22 (position: 22 557 949bp to 22 558 089bp). Nonetheless, the corticoid-pathways seems to be involved, either through its effect on sexual development, parturition or stress coping mechanisms in the differences in reproductive potential between the Elsenburg Merino lines.

Table 4.4: Genes located in the proximity of putative markers subject to selection.

Chr	Marker	Marker position (bp)	Gene	Gene range (bp)	Distance from marker (bp)
1	OAR1_114418178.1	114 418 178	TRNAC-ACA (tRNA)	114 406 400 – 114 413 500	4 678
			PBX1	114 539 578 – 114 869 959	121 400
1	OAR1_138571443.1	138 571 443	TRNAS-GGA (tRNA)	139 093 930 – 139 094 002	522 487
			TUBA1C-like (pseudogene)	137 979 577 – 137 988 676	582 767
1	OAR1_52263906.1	52 263 906	ST6GALNAC3 ST6	51 732 419 – 52 367 927	
1	OAR1_52312510.1	52 312 510	ST6GALNAC3 ST6	51 732 419 – 52 367 927	
1	s39542.1	156 652 168	PROS1	156 719 835 – 156 786 394	67 667
			EPHA3	155 656 947 – 156 064 253	587 915
1	s57436.1	28 009 525	GLIS1	27 768 357 – 28 016 168	
2	OAR2_19076972.1	19 076 972	SMC2	18 869 225 – 18 918 862	158 110
			CYLC2	20 135 042 – 20 162 469	1 058 070
2	OAR2_39793505.1	39 793 505	CDCA2	39 869 240 – 39 910 606	75 735
			KCTD9	39 911 688 – 39 993 477	118 183
2	OAR2_66083138.1	66 083 138	MAMDC2	65 933 180 – 66 097 234	
2	s53687.1	160 590 062	RPL26L1-like (pseudogene)	160 579 110 – 160 623 341	
			ACVR2A	160 457 581-160 548 526	41 536
3	OAR3_182821715.1	182 821 715	KIAA1551	182 636 486-182 667 748	153 967
3	s28031.1	185 982 335	CCDC91	185 938 352 – 186 263 491	
			C3H9orf50	6 846 882 – 6 864 293	212 990
3	s33128.1	110 605 308	ATXN7L3B	110 403 915 – 110 404 750	200 558
			KCNC2	110 831 468 – 110 841 826	226 160
6	OAR6_23745238.1	23 745 238	PPP3CA	23 527 781 – 23 850 412	
6	s08703.1	49 249 020	PCDH7-like (mRNA)	49 839 231 – 50 304 642	590 211
7	OAR7_19668788.1	19 668 788	NEO1	19 509 385 – 19 742 776	
7	OAR7_45129063.1	45 129 063	VPS13C	45 080 689 – 45 260 975	
8	OAR8_30441759.1	30 441 759	RTN4IP1	30 416 044 – 30 467 678	
9	OAR9_43466490.1	43 466 490	RRS1	43 508 982 – 43 510 624	42 492
			ADHFE1	43 513 058 – 43 545 520	46 568
			CRH	43 350 804 – 43 352 408	114 082

			TRIM55	43 303 233 – 43 349 247	117 243
9	OAR9_91934740.1	91 934 740	YWHAG-like (pseudogene)	91 680 870 – 91 681 532	253 208
			ZNF134-like	92 522 186 – 92 523 241	587 446
9	s57595.1	84 939 559	RUNX1T1	84 898 418 – 85 022 806	
12	OAR12_63154765.1	63 154 765	GLT25D2	63 073 044 – 63 187 516	
12	s25288.1	4 510 994	C4BPA-like	4 470 248 – 4 477 532	33 462
			CD55	4 582 491 – 4 600 516	71 497
12	s44950.1	4 530 169	CD55	4 582 491 – 4 600 516	52 322
			C4BPA-like	4 470 248 – 4 477 532	52 637
13	OAR13_80614774_X.1	80 614 775	ZNF217	80 692 179 – 80 712 632	77 404
			TSHZ2	80 349 024 – 80 436 174	178 601
16	OAR16_68784953.1	68 784 953	IRX1	69 235 560 – 69 269 592	450 607
			ADAMTS16	67 708 850 – 67 884 104	900 849
16	s03359.1	67 937 140	ADAMTS16	67 708 850 – 67 884 104	53 036
			IRX1	69 235 560 – 69 269 592	1 298 420
21	s36706.1	31 219 545	ETS1	31 044 523 – 31 116 703	102 842
			FLI1	31 356 351 – 31 417 035	136 806
21	s40378.1	54 522 768	MUC2	49 984 510 – 50 058 688	4 464 080
23	s33116.1	8 046 999	CCDC102B	8 237 383 – 8 441 143	190 384
			DOK6	7 486 798 – 7 762 162	284 837
X	OARX_104046252.1	104 046 252	RPL23A (pseudogene)	104 078 031 – 104 078 512	31 779
			HNRNPK-like	104 003 751 – 104 005 980	40 272

Chr: Chromosome

Table 4.5: The genomic position of genes cited in literature for their involvement with variation in ovulation rate in sheep.

Chromosome	Gene	Position (bp)
5	*GDF9*	418 405 000 – 418 435 000
6	*BMPR1B*	29 361 995 – 29 587 033
X	*BMP15*	50 970 938 – 50 977 427

4.4. Conclusions:

Ovulation rate is key to reproductive output and it is therefore not surprising that several highly proliferative breeds or lines carry mutations in genes affecting ovulation rate. The current study failed to identify markers in the vicinity of three genes (*GDF9, BMPR1B*, *and BMP15)* involved in increased ovulation rate of highly proliferative lines. Chromosome-specific partitioning of the data did not identify specific chromosomes with comparatively large numbers of markers subject to selection. However, it did facilitate the identification of 31 loci, across 13 of the 27 chromosomes, associated with genomic regions subject to selection. The latter supports the premise that reproductive traits are under the control of numerous loci spread throughout the genome. Two of the markers identified to be subject to selection are located near genes involved with the HPAA, stress coping ability and sperm characteristics and these traits may be involved in the differences in reproductive output of the Elsenburg Merino lines. Indeed, other studies have confirmed that differences in these traits occur between the Elsenburg Merino lines.

In future, DNA quality will need to be ensured to guarantee the reliability of genotype results, thereby also yielding more useful genotypes. A denser panel of SNP markers with greater coverage of the *ovine* genome will be beneficial in identifying genomic regions under selection that may have remained undiscovered by the current study.

4.5. References:

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a F_{ST}-outlier method. BioMed Central Bioinformatics 9: 323

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society: Biological Sciences 263: 1619-1626

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05: logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier, France

Boshoff NH (2014) The influence of genotype on sperm motility and sperm head morphometry of Merino (*Ovis aries*) sheep. MSc dissertation, Stellenbosch University, South Africa

Celi P, Bush R (2010) Pregnancy, lambing and survival. In DJ Cottle (ed.), International sheep and wool handbook, pg. 233-268. Nottingham: Nottingham University Press, Nottingham, England

Chan EC, Falconer J, Madsen G, Rice KC, Webster EL, Chrousos GP, Smith R (1998) A corticotropin-releasing hormone type I receptor antagonist delays parturition in sheep. Endocrinology 139: 3357-3360

Cloete JJE, Cloete SWP, Hoffman LC, Fourie JE (2004b) Slaughter traits of Merino sheep divergently selected for multiple rearing ability. South African Journal of Animal Science. 34: 189-196

Cloete JJE, Cloete SWP, Hoffman LC (2005b) Behaviour of Merinos divergently selected for multiple rearing ability in response to external stimuli. Small Ruminant Research 60: 227-236

Cloete JJE, Cloete SWP, Hoffman LC (2010) Behaviour of Merinos divergently selected for multiple rearing ability in response to external stimuli. Proceedings of the World Congress on Genetics Applied to Livestock Production, August 2010, Leipzig, Germany

Cloete SWP, Scholtz AJ (1998) Lamb survival in relation to lambing and neonatal behaviour in medium wool Merino lines divergently selected for multiple rearing ability. Australian Journal of Experimental Agriculture 38: 801-811

Cloete SWP, Van Wyk JB, Scholtz AJ, Gilmour AR (2003) Genetic and environmental parameters for lambing behaviour in Merino lines divergently selected for ewe maternal rearing ability. Proceedings of the New Zealand Society of Animal Production 63: 169-172

Cloete SWP, Gilmour AR, Olivier JJ, Van Wyk JB (2004a) Genetic and phenotypic trends and parameters in reproduction, greasy fleece weight and live weight in Merino lines divergently selected for multiple rearing ability. Australian Journal of Experimental Agriculture 44: 745-754

Cloete SWP, Olivier JJ, Van Wyk JB, Schoeman SJ, Erasmus GJ (2005a) Genetic parameters and trends for hogget traits in Merino lines divergently selected for multiple rearing ability. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 16: 24-27

Cloete SWP, Scholtz AJ, Cloete JJE, Van Wyk JB (2005c) The ability of Merino ewes and lambs to reunite after separation, as affected by divergent selection for ewe multiple rearing capacity. Australian Journal of Experimental Agriculture 45: 1131-1137

Cloete SWP, Olivier JJ, Olivier WJ (2007) Genetic change in South African Merino resource flocks. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 17: 320-323

Cloete SWP, Misztal I, Olivier JJ (2009) Genetic parameters and trends for lamb survival and birth weight in a Merino flock divergently selected for multiple rearing ability. Journal of Animal Science 87: 2196-2208

Davis GH, Montgomery GW, Alison AJ, Kelly RW, Bray AR (1982) Segregation of a major gene influencing fecundity in progeny of Booroola sheep. New Zealand Journal of Agricultural Research 25: 525-529

Davis GH (2005) Major genes affecting ovulation rate in sheep. Genetics Selection Evolution 37: S11-S23

De Graaf SP (2010) Reproduction. In: Cottle DJ (ed.) The international sheep and wool handbook, pg. 189-222. Nottingham University press, Nottingham. England

Demars J, Fabre S, Sarry J, Rossetti R, Gilbert H, Persani L, Tosser-Klopp G, Mulsant P, Nowak Z, Drobik W, Martyniuk E, Bodin L (2013) Genome-wide association studies identify two novel *BMP15* mutations responsible for an atypical hyperprolificacy phenotype in sheep. Public Library of Science Genetics 9: e1003482

Duforet-Frebourg N, Bazin E, Blum MGB (2014) Genome scans for detecting footprints of local adaptation using a Bayesian factor model. Molecular Biology and Evolution 31: 2483-2495

Foll M, Gaggiotti OE (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics 180: 977-993

Gutiérrez-Gil B, Arranz JJ, Pong-Wong R, García-Gámez E, Kijas J, Wiener P (2014) Application of selection mapping to identify genomic regions associated with dairy production in sheep. Public Library of Science ONE 9: e94623

Hess H, Heid H, Zimbelmann R, Franke WW (1995) The protein complexity of the cytoskeleton of bovine and human sperm heads: The identification and characterization of cylicin II. Experimental Cell Research 218: 174-182

Holsinger KE, Weir BS (2009) Genetics in geographically structures populations: defining, estimating and interpreting FST. Nature Reviews Genetics 10: 639-650

Hough D, Swart P, Cloete JJE, Cloete SWP (2010) Searching for SNPs that affect sheep robustness: *CYP17* SNP affects behavioural responses to psychological stress. Proceedings for the Advancement of Animal Breeding and Genetics 20: 503-506

Hough D (2012) Comparison of two CYP17 isoforms: implications for cortisol production in the South African Merino. PhD dissertation, Stellenbosch University, South Africa

Hu Z-L, Park CA, Wu X-L, Reecy JM (2013) Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Research 41: D871-D879

Illumina (2010) Infinium genotyping analysis. Technical note: DNA analysis. Accessed January 2014, from: www.illumina.com

Juengel JL, Connell AR, French MC, Proctor LE, Wheeler R, Farquhar PA, Dodds KG, Galloway SM, Johnstone PD, Davis GH (2011) Identification of a line of sheep carrying a putative autosomal gene increasing ovulation rate in sheep that does not appear to interact with mutations in the transforming growth factor beta superfamily. Biology of Reproduction 85: 113-120

Juengel JL, Davis GH, McNatty KP (2013) Using sheep lines with mutations in single genes to better understand ovarian function. Reproduction 146: 111-123

Kater CE, Biglieri EG (1994) Disorders of steroid 17α-hydroxylase deficiency. Endocrinology and Metabolism Clinics of North America 23: 341-357

Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J, Dalrymple B, other members of the International Sheep Genomics Consortium10 (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. Public Library of Science Biology 10: e1001258

Kijas JW, Porto-Neto L, Dominik S, Reverter A, Bunch R, McCulloch R, Hayes BJ, Brauning R, McEwan J, International Sheep Genomics Consortium (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. Animal Genetics 45: 754-757

Koob GF, Heinrichs SC (1999) A role for corticotropin-releasing factor and urocortin in behavioral responses to stressors. Brain Research 848: 141-152

Liu J, Zhang L, Xu L, Ren H, Lu J, Zhang X, Zhang S, Zhou X, Wei C, Zhao F, Du L (2013) Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. BioMed Central Genomics 14: 229

Manel S, Conord C, Després L (2009) Genome scan to assess the respective role of host-plant and environmental constraints on the adaptation of a widespread insect. BioMed Central Evolutionary Biology 9: 288

McNatty KP, Galloway SM, Wilson T, Smith P, Hudson NL, O'Connell A, Bibby AH, Heath DA, Davis GH, Hanrahan JP (2005) Physiological effects of major genes affecting ovulation rate in sheep. Genetics Selection Evolution: 37 S25-S38

Miller JM, Poissant J, Kijas JW (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. Molecular Ecology Resources 11: 314-322

Moradi MH, Nejati-Javaremi A, Moradi-Shahrbabak M, Dodds KG, McEwan JC (2012) Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. BioMed Central Genetics 13:10

Mulsant P, Lecerf F, Fabre S, Schibler L, Monget P, Lanneluc I, Pisselet C, Riquet J, Monniaux D, Callebaut I, Cribiu E, Thimonier J, Teyssier J, Bodin L, Cognié Y, Chitour N, Elsen J-M (2001) Mutation *in bone morphogenetic protein receptor-IB* is associated with increased ovulation rate in Booroola Merino ewes. Proceedings of the National Academy of Sciences 98: 5104-5109

Narum SR, Hess JE (2011) Comparison of FST outlier test for SNP loci under selection. Molecular Ecology Resources 11: 184-194

Notter DR (2008) Genetic aspects of reproduction in sheep. Reproduction of Domestic Animal 43: 122-128

Olivier JJ (1999) The South African Merino performance testing scheme. In: Rising to the challenge – Breeding for the 21st century customer. Beef Industry and CRC for Premium Quality Wool Industry Symposia. Supplement to the Proceedings of the Association for the Advancement of Animal Breeding and Genetics 13: 119-124

Seasholtz AF, Valverde RA, Denver RJ (2002) Corticotropin-releasing hormone-binding protein: biochemistry and function from fishes to mammals. Journal of Endocrinology 175: 89-97

Smagin GN, Heinrichs SC, Dunn AJ (2001) The role of CRH in behavioral responses to stress. Peptides 22: 713-724

Snowder GD, Fogarty NM (2009) Composite trait selection to improve reproduction and ewe productivity: A review. Animal Production Science 49: 9-16

Storbeck K-H, Swart AC, Snyman MA, Swart P (2008) Two *CYP17* genes in the South African Angora goat (*Capra hircus*) – the identification of three genotypes that differ in copy number and steroidogenic output. FEBS Journal 272: 3934-3943

Therkildsen NO, Hemmer-Hansen J, Als TD, Swain DP, Morgan MJ, Trippels EA, Palumbi SR, Meldrup D, Nielsen EE (2013) Microevolution in time and space: SNP analysis of historical DNA reveals dynamic signatures of selection in Atlantic cod. Molecular Ecology 22: 2424-2440

Våge DI, Husdal M, Kent MP, Klemetsdal G, Boman IA (2013) A missense mutation in *growth differentiation factor 9* (*GDF9*) is strongly associated with litter size in sheep. BioMed Central Genetics 14:1

Wadhwa PD, Garite TJ, Porto M, Glynn L, Chicz-DeMet A, Dunkel-Schetter C, Sandman CA (2004) Placental corticotropin-releasing hormone (CRH), spontaneous preterm birth, and fetal growth restriction: A prospective investigation. American Journal of Obstetrics and Gynecology 191: 1063-1069

Yanase T, Simpson ER, Waterman MR (1991) 17α-hydroxylase/17,20-lyase deficiency: from clinical investigation to molecular definition. Endocrine Review 12: 91-108

Yanase T (1995) 17α-hydroxylase/17,20-lyase defects. Journal of Steroid Biochemistry and Molecular Biology 53: 153-157

Chapter 5:

The identification of stretches of homozygosity in a South African Merino flock divergently selected for reproduction traits

5.1. Introduction:

Selection within a population or species refers to the differential contribution of certain individuals to the genetic composition of the next generation (Frankham *et al*. 2002). In natural populations, individuals with an enhanced ability for survival and reproduction will produce more offspring and as a result the phenotype and genotype of these individuals will be inherited preferentially and be present at a higher frequency in the next generation (Frankham *et al*. 2002, Nielsen 2005). The latter is referred to as natural selection, whereas artificial selection refers to the influence of humans on the selection process. When humans are responsible for the preferential selection and breeding of individuals, based either on phenotypic or genetic characteristic, the favoured phenotype and/or genotype will also be found at a higher frequency in the population (Price 2002). Artificial selection of animals has taken place since the domestication of the dog more than 14 000 years ago (Vilà *et al.* 1997, Leonard *et al*. 2002) and has resulted in the development of numerous domestic species as well as the large diversity in breeds of most domestic animals (FAO 2000, Neff & Rine 2006, DAD-IS 2014).

The effect of selection can be observed in changes in phenotypic and genotypic characteristics of livestock species. When selection favours an extreme phenotype, the phenotypic mean shifts in the direction of the extreme and this is referred to as directional selection (Frankham *et al*. 2002). In contrast, balancing selection results in an increase in phenotypic variability as both extremes of the phenotypic range is favoured (Nielsen 2005). Certain alleles may also be favoured or infer an advantage for reproduction and/or survival and these alleles therefore reach fixation in the population and this is referred to as positive selection. If a deleterious allele is present, negative selection tends to eliminate this allele from the population. Positive and negative selection therefore decreases variability in the genome by increasing the occurrence of a certain favourable allele or decreasing the frequency of a deleterious allele (Nielsen 2005, Walsh 2008). Regions surrounding loci subject to selection may

also experience a reduction in variability as a result of linkage with those regions under selection. This is referred to as a selective sweep and is characterised by a loss of heterozygosity of selectively neutral regions as a result of selection pressure on nearby loci (Luikart *et al.* 2003, Nielsen 2005, Walsh 2008). Recent directional selection results in a deviation from the neutral model and may manifest in a change in local allele frequencies, the occurrence of extended regions of homozygosity, shared haplotype homozygosity, increased levels of linkage disequilibrium (LD) and differentiation between populations (Tang *et al*. 2007, Walsh 2008, Moradi *et al.* 2012, Randhawa *et al*. 2014).

The development of high-throughput single nucleotide polymorphism (SNP) genotyping platforms for livestock has enabled genotyping thousands of loci evenly distributed throughout the genome and has facilitated studies of genomic regions underlying quantitative traits, and population genetic studies on the evolution, domestication and breed formation of livestock (Hu *et al*. 2009, The Bovine HapMap Consortium 2009, Daetwyler *et al*. 2010, Fan *et al*. 2010, Snelling *et al.* 2010, Bolormaa *et al.* 2011, Kijas *et al*. 2012, Zhang *et al*. 2012, McTavish *et al*. 2013, Decker *et al*. 2014). Most livestock species have undergone multiple selection events during domestication and subsequent selective breeding and therefore serve as an ideal platform to identify signatures of selection. Identification of these regions and the effect of selection on the genome allows for a greater understanding of the mechanisms involved in adaptation, domestication and trait selection of livestock (Neff & Rine 2006, Barendse *et al.* 2009, Kijas *et al*. 2012, Gutiérrez-Gil *et al.* 2014, Randhawa *et al*. 2014). Whole-genome selection sweep analyses have been utilised in *ovine* studies to identify genomic regions involved with fat deposition in fat-tailed sheep (Moradi *et al.* 2012), disease resistance (Phua *et al.* 2014), milk production (Gutiérrez-Gil *et al*. 2014), polledness and muscling (Randhawa *et al*. 2014), parasite resistance (McRae *et al.* 2014), production traits, and global signatures of selection across sheep populations (Kijas *et al.* 2012, Fariello *et al.* 2014).

The Elsenburg Merino flock has undergone divergent selection for more than eight generations for the ability to raise multiple offspring and has resulted in a High and Low line that differ markedly for several reproduction and robustness traits (see Chapters 1 and 2 for a comprehensive description of the history and selection of the

Elsenburg Merino flock). Due to the selection pressure applied to this flock and the subsequent change in the reproductive trait means, the allele frequencies within or adjacent to regions subject to selection may exhibit selection signatures. The Elsenburg Merino flock therefore serves as an ideal platform to investigate the effect of artificial selection on the genome as well as identifying putative genomic regions underlying the divergence in reproduction and robustness traits of this flock. A F_{ST} outlier approach was used to identify SNP markers that have been subject to selection in the Elsenburg Merino flock in Chapter 4. The aim of the current study was therefore to identify genomic regions that have been subject to selection as a result of the divergent selection pressure applied to the number of lambs weaned by identifying regions of homozygosity indicative of recent directional selection.

5.2. Materials and Methods:

5.2.1. Sample selection:

Sample selection, collection, genotyping and quality assessment are described in Chapter 3. In short, 112 samples (Table 5.1) were selected from the Elsenburg Merino flock from individuals born between 2002 and 2010, with fairly accurate BLUP estimated breeding values for number of lambs weaned per parity (NLW) and total weight of lamb weaned per parity (TWW). The samples selected represented the extremes of the phenotypic distribution for NLW and TWW. Pedigree information was considered to minimise bias due to inbreeding and relatedness in the sampling cohort and to reduce substructure in the samples. Fourty-six High line and 29 Low line samples could be used in downstream analyses following quality control.

5.2.2. Genotyping and quality control:

Blood samples were stored at -20°C for a prolonged period and were thawed at least once during this time. Samples were also thawed for application to bloodcards that were forwarded to GeneSeek Inc. (Lincoln, NE, USA) for genotyping using the OvineSNP50 BeadChip. The samples used in the current study therefore underwent at least two freeze-thaw cycles. GenomeStudio version 1.0 software (Genotyping Module, Illumina) was used to convert signal intensity data into genotype calls and calculate quality statistics. The sample set exhibited poor genotype call quality (as discussed in Chapter 3) as the DNA quality of some samples were poor. In an attempt to increase the number of SNPs that could be included in the current analysis, samples

that exhibited call rates ≤95% were excluded. In addition, the call rate cut-off for SNP loci was also increased to 95% to match the increase in the sample call rate and ensure reliability of the calls. Other quality control measures remained the same as in previous analyses (refer to Chapters 3 and 4); >0.25 GenCall score; >0.5 GenTrain score; and >0.01 MAF.

5.2.3. Data analyses:

Input files containing pedigree, genotype and SNP information (PED and MAP files) were created using a PLINK version 1.07 (Purcell *et al.* 2007) input report plug-in version 2.1.3 for GenomeStudio Software. The SNP loci in strong LD were pruned from the data sets of the High and Low lines using the --indep-pairwise 50 5 0.5 command in PLINK which calculates pairwise LD within a 50 SNP window and removes one SNP from a pair where the LD exceeds 0.5 before moving on 5 SNPs and repeating the procedure. Linkage disequilibrium (r^2) was calculated for all SNP pairs using the --r2 command. The observed heterozygosity and inbreeding coefficient (F_{IS}) was calculated separately for the two lines from LD pruned SNP genotype data. A sliding window approach was implemented in PLINK to identify regions of homozygosity in the High line, Low line and a combination sample. The combination sample comprised an equal number (n=29 per line) of High and Low line samples selected at random to identify stretches of homozygosity present in both lines. Default values in PLINK are applicable to high density human SNP genotype data and therefore several values were tested (see Table 5.2) to determine the appropriate parameters for identifying runs of homozygosity (ROH) in the current 50K *ovine* data set. One heterozygous locus and five missing calls were allowed in each window for all parameter sets. Regions that overlapped across individuals within a group were identified using the --homozyg-group command in PLINK. The --homozyg-match 0.90 command was used to ensure allelic matching of overlapping region within a group. The consensus region of overlapping ROH that were present in 95% of samples in the three respective groups was further investigated using the *Ovis aries* genome assembly v3.0 (http://www.ncbi.nlm.nih.gov/genome) and compared to F_{ST} results presented in Chapter 4.

5.3. Results:

A total of 75 samples had a call rate >95% (see Table 5.1 for the distribution of samples between the lines and sexes) and the average call rate of these samples was 99.40%. After removal of SNP loci with a GenCall score ≤0.25, a GenTrain score ≤0.5 and a call rate ≤95%, 40 167 SNPs remained. A further 3 750 SNPs were removed due to a MAF≤0.01 as well as 257 unmapped SNPs and thereafter a total of 36 162 SNPs remained for downstream analyses.

Table 5.1: The number of samples included in the analyses grouped according to selection line and sex.

	Ewes	Rams	Total
High line	31	15	46
Low line	25	4	29
Total	56	19	75

As LD differed slightly between the lines, LD pruning was conducted separately for each sample group. After removal of SNPs in strong LD, 16 796, 13 430 and 22 958 SNP loci remained for further analyses of the High line, Low line and combined sample, respectively. The pairwise LD results indicated a decline in LD with an increase in genomic distance between loci (Figure 5.1). The High line exhibited LD decay at somewhat shorter distances than the Low line, thereby indicating that LD extends over slightly larger distances in the Low line.

The High and Low lines exhibited negative mean inbreeding coefficient values (SD) of -0.06 (0.04) and -0.08 (0.05), respectively (Appendix Table A5.1). A total of 36% of SNPs included after quality control and LD pruning were heterozygous in the two lines (Appendix Table A5.1).

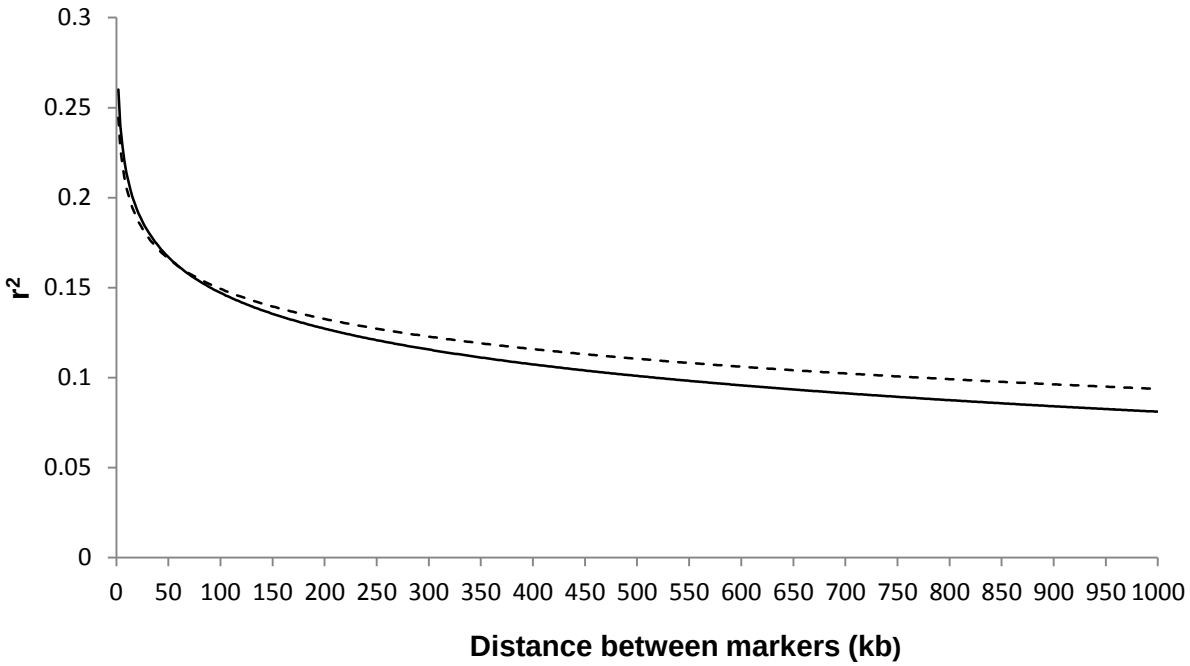


Figure 5.1: Linkage disequilibrium calculated for the High line (continuous line) and Low line (dashed line).

It was not possible to identify any ROH using the default ROH parameters in PLINK and therefore the parameters were adjusted to appropriate values for the current data. Initially, the sliding window was reduced from 5 000kb and 50 SNPs to 500kb and five SNPs and thereafter the number of homozygous SNPs, the size of a homozygous stretches, and SNP density within a homozygous stretch was systematically reduced to five SNPs, 10kb and 1SNP/500kb, respectively (see Table 5.2). The aforementioned parameters allowed the identification of numerous segments of homozygosity in individual samples, however, little overlap was observed between samples and therefore the number of SNPs and the size of homozygous stretches were increased to ensure that ROH occurred in all samples. When the following parameters were used to determine homozygous stretches, numerous ROH could be identified of which several overlapped between individuals: 100kb containing at least ten SNPs with a density of 1SNP every 500kb.

Table 5.2: The parameter values tested using PLINK to establish a functional set of values to identify runs of homozygosity in the current data set.

SNPs	Kb	Window: SNP	Window: Kb	Density	Result
100	1000	50	5000	50	No matches
50	1000	50	5000	50	No matches
10	10	50	5000	50	No matches
5	10	5	500	500	Many matches
100	1000	5	500	500	Matches in <90% of samples
50	1000	5	500	500	Matches in <90% of samples
50	100	5	500	500	Matches in <90% of samples
50	10	5	500	500	Matches in <90% of samples
10	100	5	500	500	Matches in <95% of samples

(Density refers to the distance between SNPs)

Putative ROH that overlapped between >95% of individuals in the three groups tested, are indicated in Table 5.3 along with the genes that occur within the identified regions in Table 5.4. Nine ROH located on five respective chromosomes were identified in the Low line, while two regions on two chromosomes and one region on chromosome 13 were identified in the High line and combination group, respectively. The union region refers to the combination of all the homozygous SNP stretches of all individuals around a specific genomic location, while the consensus regions only refers to an area within the union region that overlap across individuals within a group. A consensus region will therefore most likely encompass fewer SNPs and stretch over smaller genomic regions than the union regions. Due to the commonality of the consensus regions between all samples within the three groups, these regions were considered to be the putative targets of selection and were therefore investigated in more detail. At most, one or two annotated genes were identified for most of the consensus regions; however, the consensus region identified in the combination group for chromosome 13, spanned a region of ten genes and four pseudogenes. The mean (SD) number of SNPs within a consensus region was 2.91 (2.23) and the mean size was 356.09 (681.51)kb, whereas the mean number of SNPs within a union region was 107.67 (92.84) and stretched over 18 231.89 (9 439.44)kb.

The union region of a putative ROH identified on chromosome 9 in the High line overlapped with the location of a SNP locus (at position 43 466 490bp) identified to be subject to selection using F_{ST} outlier methodology (in Chapter 4).

Table 5.3: The size and location of regions of homozygosity identified within the Low line, High line and a combined sample, respectively.

Region	Chr	SNP ID (Start)	SNP ID (End)	Start (bp)	End (bp)	Size (kb)	Number of SNPs
			Low line				
Con	1	OAR1_174847452.1	OAR1_174976013.1	174 847 452	174 976 013	128.56	3
Union	1	OAR1_162623538.1	s67438.1	162 623 538	180 257 958	17 634.40	73
Con	3	OAR3_51459716.1	OAR3_52003414.1	51 459 716	52 003 414	543.70	4
Union	3	OAR3_47351324.1	OAR3_55315847.1	47 351 324	55 315 847	7 964.52	27
Con	3	OAR3_26841095.1	OAR3_27391943.1	26 841 095	27 391 943	550.85	6
Union	3	OAR3_16804510.1	s07119.1	16 804 510	37 867 080	21 062.60	112
Con	6	OAR6_76793301.1	OAR6_79180962.1	76 793 301	79 180 962	2 387.66	8
Union	6	OAR6_68847351.1	OAR6_83937289.1	68 847 351	83 937 289	15 089.90	52
Con	6	OAR6_33453920.1	OAR6_33453920.1	33 453 920	33 453 920	0	1
Union	6	OAR6_29577816.1	OAR6_39003764.1	29 577 816	39 003 764	9 425.95	39
Con	6	s44353.1	s20716.1	33 132 578	33 142 436	9.86	2
Union	6	OAR6_29577816.1	OAR6_39003764.1	29 577 816	39 003 764	9 425.95	39
Con	20	OAR20_9220483.1	OAR20_9220483.1	9 220 483	9 220 483	0	1
Union	20	OAR20_3067949.1	s22096.1	3 067 949	14 602 357	11 534.40	61
Con	20	s43094.1	s73139.1	8 007 630	8 060 613	52.98	2
Union	20	OAR20_3067949.1	s22096.1	3 067 949	14 602 357	11 534.40	61
Con	23	s45643.1	s45643.1	3 874 491	3 874 491	0	1
Union	23	OAR23_393569_X.1	OAR23_19193102.1	393 570	19 193 102	18 799.50	81
			High line				
Con	9	s41167.1	s51056.1	34085 524	34 128 718	43.19	2
Union	9	s23772.1	OAR9_43708493.1	15928980	43 708 493	27 779.50	186
Con	14	OAR14_31350096.1	OAR14_31350096.1	31350096	31 350 096	0	1
Union	14	OAR14_2934499.1	s69327.1	2934499	35 454 622	32 520.10	238
			Combined				
Con	13	OAR13_52630089.1	OAR13_53186320.1	52 630 089	53 186 320	556.23	4
Union	13	s75440.1	s54595.1	38 396 670	74 407 997	36 011.30	323

Chr: Chromosome; SNP ID: SNP identifier; Con: Consensus

Table 5.4: The location of annotated genes situated within consensus regions of homozygosity in the Low line, High line and a combination sample, respectively (continues on next page)**.**

Chr	Start (bp)	End (bp)	Gene	Gene location
			High line	
1	174 847 452	174 976 013	PLCXD2	174 810 638-174 873 256
3	51 459 716	52 003 414	CTNNA2	50 474 109-51 867 876
3	26 841 095	27 391 943	OSR1	27 161 502-27 168 291
6	76 793 301	79 180 962	LPHN-3-like (pseudogene)	76 991 900-77 568 519
			EEF1A1-like (pseudogene)	77 503 848-77 505 236
			PK-like (pseudogene)	77 603 225-77 611 754
			TECRL	79 660 320-79 792 332
6	33 453 920	33 453 920	FAM190A	33 661 151-34 495 012
6	33 132 578	33 142 436	FAM190A	33 661 151-34 495 012
20	9 220 483	9 220 483	TCP11	9 185 364-9 216 909
20	8 007 630	8 060 613	LEMD2	8 009 715-8 024 749
			MLN	8 033 225-8 040 616
23	3 874 491	3 874 491	ZNF407	3 529 795-3 890 758
			Low line	
9	34 085 524	34 128 718	ST18	33 984 577-34 120 466
14	31 350 096	31 350 096	RPL13A-like (pseudogene)	31 465 158-31 500 659
			CDH11	32 063 965-32 152 357

Chr: Chromosome

Table 5.4: The location of annotated genes situated within consensus regions of homozygosity in the Low line, High line and a combination sample, respectively (continued)**.**

Chr	Start (bp)	End (bp)	Gene	Gene location
			Combined	
13	52 630 089	53 186 320	PTPN1-like (pseudogene)	52 719 190-52 739 409
			PTPN1-like (pseudogene)	52 760 944-52 787 706
			PTPN1-like (pseudogene)	52 835 171-52 854 366
			MCR3-like (pseudogene)	52 886 162-52 895 619
			PSFS1-like	5 2921 394-52 936 554
			TRNAC-GCA	52 967 204-52 967 275
			PCMTD2	52 971 948-52 986 197
			MYT1	53 005 310-53 040 574
			NPBWR2	53 097 320-53 098 330
			OPRL1	53 103 104-53 110 493
			C13H20orf201	53 116 179-53 116 953
			RGS19	53 120 751-53 125 975
			TCEA2	53 127 405-53 134 947
			PRPF6	53 135 458-53 193 679

Chr: Chromosome

5.4. Discussion:

As mentioned in Chapter 3 and 4, the stringency of genotype quality control measures depended on the ultimate aim of the respective studies. More emphasis was placed on larger numbers of samples for the F_{ST} outlier methodology employed in Chapter 4 to accurately capture SNP allele frequency differences between groups. In the current study, a greater emphasis was however placed on genome coverage for selection sweep analyses that would enable identifying the presence and extent of ROH. The exclusion of samples with a call rate of ≤95% (in comparison to ≤85% used in Chapter 3 and 4) resulted in a 23% increase in the number of SNP loci that could be incorporated into further analyses. The number of SNPs used in the current study amounted to approximately 67% of the total number of SNPs assayed by the OvineSNP50 chip. The genotype results of 46 High line and 29 Low line individuals could be used for further analyses as compared to 54 High line and 37 Low line individuals used in Chapter 3 and 4.

High levels of LD in SNP chip genotype data may reflect SNP ascertainment bias and can influence population estimates, such as genetic diversity and genetic distance between populations. Therefore, SNP loci in high LD were removed from the current data set (López Herráez *et al*. 2009, Kijas *et al*. 2012). This resulted in a loss of approximately 50% of SNP loci, however, other *ovine* studies using SNP chip data also reported a >50% loss in the number of loci available for downstream analyses after LD pruning (Kijas *et al.* 2012).

The overall extent of LD within the High and Low lines was similar to results reported for international sheep breeds where LD was found to dissipate at distances >30kb (Kijas *et al.* 2014). International Merino samples exhibit high levels of genetic diversity (Kijas *et al.* 2012) and therefore it is not surprising that the Merino exhibits LD decay over shorter distances than less diverse breeds such as the Border Leicester (Kijas *et al*. 2014). The LD decay in the current sample set was intermediate to that of International Merino samples and that of less diverse breeds, such as the Border Leicester and Poll Dorset samples by Kijas *et al*. (2014). The fact that both lines and specifically the Low line exhibits higher levels of LD than expected for Merinos in general, may be the result of a limited sample number as well as potential loss of genetic diversity in these lines due to the ongoing selection trial.

The mean inbreeding coefficient was negative for both lines and indicates that the sample cohort is similar to an outbred population where inbreeding does not occur at high levels. This is not surprising considering that care was taken to select the least related individuals with the lowest inbreeding coefficient values based on pedigree information for inclusion in the study.

Samples were divided into the High line, Low line and a combination sample that included equal numbers of the aforementioned High line and Low line samples. This was done to ensure that discrete regions of homozygosity could be detected in the separate lines as well as regions of homozygosity shared between the lines. Regions of homozygosity were identified using a sliding window approach to identify homozygous stretches that overlapped in >95% of the samples included in the three groups. Twelve respective regions of homozygosity on eight chromosomes were identified in >95% of individuals within the three respective groups tested. Nine ROH were identified on eight chromosomes in the Low line, two ROH were identified on two chromosomes in the High line and one ROH was identified on a single chromosome in the combined sample. The ROH consensus regions (the region overlapping >95% of samples) contained an average of approximately three SNPs and stretched over an average distance of approximately 360kb. One or two annotated genes were identified in most consensus regions, except for the regions on chromosome 13, where several genes were identified. Some genes were found to have an obvious role in reproduction traits such as the *t-complex protein 11* (*TCP11*) gene that has been shown to play a role in sperm morphology and fertilisation in humans (Liu *et al*. 2011). The *cadherin-11* (*CDH11*) and *suppression of tumorigenicity 18 breast carcinoma zinc-finger protein* (*ST18*) have been shown to suppress tumerogenesis in humans (Jandrig *et al.* 2004, Marchong *et al.* 2010). The *ST18* gene has also been shown to regulate apoptosis and inflammation (Sarig *et al.* 2012). These genes may potentially play a role in regulating cell proliferation and cell death that relate to reproductive performance during ova or foetal development.

The consensus regions of the 12 ROH identified in the current study did not overlap with any of the SNP loci identified as subject to selection in Chapter 4; however, the union region of a single ROH did overlap with the location of a SNP locus identified in

Chapter 4 to be under selection. The SNP locus is located on chromosome 9 at position 43 466 490bp. In Chapter 4, it was reported that the *corticotropin releasing hormone* (*CRH*) gene was in relatively close proximity to the aforementioned SNP locus on chromosome 9. However, this gene is located >110kb away from the marker and >9 000kb away from the ROH identified in the current study. Nonetheless, other studies have reported large regions (>1000kb) under selection in *ovine* samples (Moradi *et al*. 2012, Gutiérrez-Gil *et al.* 2014, Moioli *et al.* 2013) and therefore the aforementioned distances between markers and putative regions under selection may be relevant to the current study.

The twelve ROH identified within the three groups tested in the current study overlapped with several sheep QTL as listed on the sheep QTL database (www.animalgenome.org/cgi-bin/QTLdb/OA/index, Hu *et al.* 2013). Reproduction trait QTL that overlapped with the ROH identified in the current study, included aseasonal reproduction (Position: 168.6-197.3 Mb) as reported by Mateescu and Thonney (2010). The aforementioned study made use of the number of oestrus cycles, level of progesterone, pregnancy and lambing status, and the number of lambs born to identify QTL associated with aseasonal reproduction by means of genotyping 120 microsatellite markers in Dorset-Finnish crossbred sheep. A QTL on chromosome 1 exceeded the 1% experiment-wide significance level, whereas QTL on chromosome 3 and 20 exceeded the threshold at the 1% chromosome-wide level. Other QTL identified on chromosome 12, 17, 19 and 24, were significant at the 5% chromosome-wide level. Although ROH were identified on chromosome 1, 3 and 20, in the current study, only the region on chromosome 1 overlapped with the QTL identified for maximum progesterone level during the prebreeding season by Mateescu and Thonney (2010). The location of QTL for several milk traits also overlapped with ROH identified in the current study. The study by Raadsma *et al.* (2009a) on Awassi and Merino crossbreds indentified QTL for milk yield (Position: 41.4 to 76.9Mb), milk lactose yield (Position: 39.0 to 72.1Mb) and milk fat percentage (Position: 38.6 to 75.5 Mb) on chromosome 3 that overlapped with the ROH identified for the Low line in the current study. Other milk QTL have been reported on chromosome 20 (Position 2.4 to 38.4Mb) by Mateescu and Thonney (2010) and on chromosome 14 (Position 33.8 to 33.9Mb) by Crisà *et al.* (2010) and both these regions overlap with ROH identified within the High and Low lines, respectively. A QTL for testis weight on chromosome

20 (Position: 2.4 to 22.8 Mb) reported by Fullard *et al*. (2006) overlapped with the aforementioned milk traits as well as the ROH identified in the Low line in the current study. Other QTL for traits such as body weight at 56 weeks (Chromosome 3, Position: 14.1 to 74.6Mb, Raadsma *et al.* 2009b), bone weight in carcass and carcass fat percentage (Chromosome 1, Position: 1168.6 to 197.3Mb, Cavanagh *et al*. 2010) also overlapped ROH identified in the Low line. A QTL associated with intestinal parasite resistance, overlapped with a ROH identified in the Low line at chromosome 6 (Position: 68.0 to 85.1 Mb, Gutiérrez-Gil *et al.* 2009) and chromosome 20 (Position: 7.2-12.4Mb, Davies *et al.* 2006). A recent study by Gutiérrez-Gil *et al.* (2014) identified six regions indicative of a selection sweep in European dairy sheep breeds by employing F_{ST} and reduced heterozygosity approaches. Two of these regions overlapped with the union region of ROH on chromosome 3 and 13, and overlapped with the consensus region of the ROH on chromosome 6 identified in the Low line. It is of interest to note that other studies on the milk production and composition of the Elsenburg Merino ewes, reported a significant difference in the milk yield of the High and the Low line with the High line producing 17% more milk than the Low line (Cloete *et al.* 2011).

The fact that QTL for several reproduction traits overlap the ROH identified in the current study may indicate that several traits add to the overall reproduction ability of sheep. This corresponds with previous suggestions that net-reproduction is a composite trait (Olivier 1999, Snowder & Fogarty 2009) and that reproduction, similar to production traits, is a complex trait controlled by many genes (Goddard & Hayes 2009). It is therefore not surprising that genes associated with sperm cell characteristics and stress response, and QTL associated with milk yield, testis weight and even internal parasite resistance were found to be located near putative regions under selection identified in the current study. It is clear that reproduction and specifically the number of lambs born is influenced by several factors and that several genes are involved in variation of this trait. Other studies of phenotypic traits have also indicated differences in the yearling weight and lamb survival (Cloete *et al.* 2005, 2009) as well as other traits, such as the animal's reaction to humans in an arena test (Cloete *et al.* 2010) and susceptibility to breech blowfly strike (Scholtz *et al*. 2010) between the High and Low lines. The fact that ROH overlapped with QTL for production traits that have not been shown to differ significantly between the Elsenburg Merino lines,

may indicate the role of these traits in reproduction performance or may be due to indirect selection facilitated by the selection on reproduction traits (Brooks & Endler 2001, Weigel *et al.* 2006).

Runs of homozygosity may arise due to inbreeding (Bosse *et al.* 2012) and have also been employed to map autozygous regions underlying disease phenotypes (Nalls *et al.* 2009, Lencz *et al*. 2007). In the current study, sample selection for the least inbred and least related individuals coupled with strict ROH cut-off values and the ability to compare ROH between lines and groups were used to ensure that the identified ROH were related to recent directional selection. Nonetheless, ROH identified in the current study may be the result of genetic drift, inbreeding, genomic regions of low recombination, historic selection or indirect selection (Walsh 2008) and therefore further studies taking population demographic history and haplotype extent into account may be beneficial in identifying additional regions under selection in the Elsenburg Merino lines (McRae *et al*. 2014).

5.5. Conclusions:

Twelve regions of homozygosity were identified within the Elsenburg Merino lines and are most likely the result of recent divergent directional selection. Several of these regions have been shown to overlap with QTL for reproduction traits or regions under selection for milk traits in other studies. Putative QTL for production traits and internal parasite resistance have also been shown to overlap with ROH identified in the current study. These results have confirmed the complexity of the genetic architecture of reproduction traits and the fact that several genes distributed throughout the genome are involved with variation in reproduction traits in the Elsenburg Merino flock. Future studies concerned with identifying genomic regions underlying reproduction traits may benefit from denser SNP panels and analyses approaches taking population demographic history and extended haplotypes into account.

5.6. References:

Barendse W, Harrison BE, Bunch RJ, Thomas MB, Turner LB (2009) Genome wide signatures of positive selection: The comparison of independent samples and the identification of regions associated to traits. BioMed Central Genomics 10: 178

Bolormaa S, Neto LR, Zhang YD, Bunch RJ, Harrison BE, Goddard ME, Barendse W (2011) A genome-wide association study of meat and carcass traits in Australian cattle. Journal of Animal Science 89: 2297-2309

Bosse M, Megens H-J, Madsen O, Paudel Y, Frantz LAF, Schook LB, Crooijmans RPMA, Groenen MAM (2012) Regions of homozygosity in the *porcine* genome: Consequence of demography and the recombination landscape. Public Library of Science Genetics 8: e1003100

Brooks R, Endler JA (2001) Direct and indirect sexual selection and quantitative genetics of male traits in guppies (*Poecilia reticulate*). Evolution 55: 1002-1015

Cavanagh CR, Jonas E, Hobbs M, Thomson PC, Tammen I, Raadsma HW (2010) Mapping quantitative trait loci (QTL) in sheep. III. QTL for carcass composition traits derived from CT scans and aligned with a meta-assembly for sheep and cattle carcass QTL. Genetics Selection Evolution 42: 36

Cloete JJE, Cloete SWP, Hoffman LC (2010) Behaviour of Merinos divergently selected for multiple rearing ability in response to external stimuli. Proceedings of the World Congress on Genetics Applied to Livestock Production, August 2010, Leipzig, Germany

Cloete SWP, Olivier JJ, Van Wyk JB, Schoeman SJ, Erasmus GJ (2005) Genetic parameters and trends for hogget traits in Merino lines divergently selected for multiple rearing ability. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 16: 24-27

Cloete SWP, Misztal I, Olivier JJ (2009) Genetic parameters and trends for lamb survival and birth weight in a Merino flock divergently selected for multiple rearing ability. Journal of Animal Science 87: 2196-2208

Cloete SWP, Snyman MA, Scholtz AJ (2011) Genetic and environmental parameters of milk production and milk composition in South African Merinos. Proceedings of the Association of the Advancement of Animal Breeding and Genetics 19: 419-422

Crisà A, Marchitelli C, Pariset L, Contarini G, Signorelli F, Napolitano F, Catillo G, Valentini A, Moioli B (2010) Exploring polymorphisms and effects of candidate genes on milk fat quality in dairy sheep. Journal of Dairy Science 93: 3834-3845

DAD-IS (Domestic Animal Diversity Information System) (2014) Food and Agriculture Organization of the United Nations. Accessed May 2014, from: http://dad.fao.org/

Daetwyler HD, Hickey JM, Henshall JM, Dominik S, Gredler B, Van der Werf JHJ, Hayes BJ (2010) Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. Animal Production Science 50: 1004-1010

Davies G, Stear MJ, Benothman M, Abuagob O, Kerr A, Mitchell S, Bishop SC (2006) Quantitative trait loci associated with parasitic infection in Scottish blackface sheep. Heredity 96: 252-258

Decker JE, McKay SD, Rolf MM, Kim J, Molina Alcalá A, Sonstegard TS, Hanotte O, Götherström A, Seabury CM, Praharani L, Babar ME, Correia de Almeida Regitano L, Yildiz MA, Heaton MP, Liu W-S, Lei C-Z, Reecy JM, Saif-Ur-Rehman M, Schnabel RD, Taylor JF (2014) Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. Public Library of Science Genetics 10: e1004254

Fan B, Du Z-Q, Gorbach DM, Rothchild MF (2010) Development and application of high-density SNP arrays in genomic studies of domestic animals. Asian-Australian Journal of Animal Sciences 23: 833-847

FAO (Food and Agriculture Organisation of the United Nations) (2000) World watch list for domestic animal diversity, 3rd edn, Scherf B (ed.), Rome

Fariello M-I, Servin B, Tosser-Klopp G, Rupp R, Moreno C, International Sheep Genomics Consortium, Cristobal MS, Boitard S (2014) Selection signatures in worldwide sheep populations. Public Library of Science ONE 9: e103813

Frankham R, Ballou JD, Briscoe DA (2002) Evolution in large populations I. Natural selection and adaptation. In: Introduction to conservation genetics, pg. 126-152. Cambridge University Press, Cambridge, UK

Fullard K, Margawati PC, Thomsom PC, Raadsma HW (2006) QTL for testis size in ram lambs as an indicator of ovulation rate in an Indonesian Thin tail by Merino resource flock. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, 13-18 August, Belo Horizonte, Brazil

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Reviews Genetics 10: 381-391

Gutiérrez-Gil B, Pérez J, Alvarez L, Martínez-Valladares M, De la Fuente LF, Bayón Y, Meana A, San Primitivo F, Rojo-Vázquez FA (2009) Quantitative trait loci for resistance to trichostrongylid infection in Spanish Churra sheep. Genetics Selection Evolution 41: 46

Gutiérrez-Gil B, Arranz JJ, Pong-Wong R, García-Gámez E, Kijas J, Wiener P (2014) Application of selection mapping to identify genomic regions associated with dairy production in sheep. Public Library of Science ONE 9: e94623

Hu X, Gao Y, Feng C, Liu Q, Wang X, Du Z, Wang Q, Li N (2009) Advanced technologies for genomic analysis in farm animals and its application for QTL mapping. Genetica 136: 371-386

Hu Z-L, Park CA, Wu X-L, Reecy JM (2013). Animal QTLdb: An improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Research 41: D871-D879

Jandrig B, Seitz S, Hinzmann B, Arnold W, Micheel B, Koelble K, Siebert R, Schwartz A, Ruecker K, Schlag PM, Scherneck S, Rosenthal A (2004) *ST18* is a breast cancer tumor suppressor gene at human chromosome 8q11.2 Oncogene 23: 9295-9302

Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J, Dalrymple B, other members of the International Sheep Genomics Consortium (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. Public Library of Science Biology 10: e1001258

Kijas JW, Porto-Neto L, Dominik S, Reverter A, Bunch R, McCulloch R, Hayes BJ, Brauning R, McEwan J, The International Sheep Genomics Consortium (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. Animal Genetics45: 754-757

Lencz T, Lambert C, De Rosse P, Burdick KE, Morgan TV, Kane JM, Kucherlapati R, Malhotra AK (2007) Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proceedings of the National Academy of Sciences 104: 19942-19947

Leonard JA, Wayne RK, Wheeler J, Valadez R, Guillén S, Vilà C (2002) Ancient DNA evidence for Old World origin of New World dogs. Science 298: 1613-1616

Liu Y, Jiang M, Li C, Yang P, Sun H, Tao D, Zhang S, Ma Y (2011) Human t-complex protein 11 (TCP11), a testis-specific gene product, is a potential determinant of the sperm morphology. Tohuku Journal of Experimental Medicine 224: 111-117

López Herráez D, Bauchet M, Tang K, Theunert C, Pugach I (2009) Genetic variation and recent positive selection in worldwide human populations: Evidence from nearly 1 million SNPs. Public Library of Science ONE 4: e7888

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: From genotyping to genome typing. Nature Reviews Genetics 4: 981-994

Marchong MN, Yurkowski C, Ma C, Spencer C, Pajovic S, Brenda L. Gallie (2010) Cdh11 acts as a tumor suppressor in a murine retinoblastoma model by facilitating tumor cell death. Public Library of Science Genetics 6: e1000923

Mateescu RG, Thonney ML (2010) Genetic mapping of quantitative trait loci for aseasonal reproduction in sheep. Animal Genetics 41: 454-459

McRae KM, McEwan JC, Dodds KG Gemmell NJ (2014) Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. BioMed Central Genomics 15: 637

McTavish EJ, Decker JE, Schnabel RD, Taylor JF, Hillis DH (2013) New World cattle show ancestry from multiple independent domestication events. Proceedings of the National Academy of Sciences 110: E1398-E1406

Moioli B, Scatà MC, Steri R, Napolitano F, Catillo G (2013) Signatures of selection identify loci associated with milk yield in sheep. BioMed Central Genetics 14: 76

Moradi MH, Nejati-Javaremi A, Moradi-Shahrbabak M, Dodds KG, McEwan JC (2012) Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. BioMed Central Genetics 13:10

Nalls MA, Guerreiro RJ, Simon-Sanchez J, Bras JT, Traynor BJ, Launer L, Hardy J, Singleton AB (2009) Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. Neurogenetics 10: 183-190

Neff MW, Rine J (2006) A fetching model organism. Cell 124: 229-231

Nielsen R (2005) Molecular signatures of natural selection. Annual Review of Genetics 30: 197-218

Olivier JJ (1999) The South African Merino performance testing scheme. In: Rising to the challenge – Breeding for the 21st century customer. Beef Industry and CRC for Premium Quality Wool Industry Symposia. Supplement to the Proceedings of the Association for the Advancement of Animal Breeding and Genetics 13: 119-124

Phua SH, Brauning R, Baird HJ, Dodds KG (2014) Identifying chromosomal selection-sweep regions in facial eczema selection-line animals using an ovine 50K-SNP array. Animal Genetics 45: 240-247

Price EO (2002) Artificial selection. In: Animal domestication and selection, pg. 43-50. CABI publishing, New York, USA

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ, Sham PC (2007) PLINK: A tool set for whole-

genome association and population-based linkage analyses. The American Journal of Human Genetics 81: 559-575

Raadsma HW, Jonas E, McGill D, Hobbs M, Lam MK, Thomson PC (2009a) Mapping quantitative trait loci (QTL) in sheep. II. Meta-assembly and identification of novel QTL for milk production traits in sheep. Genetics Selection Evolution 41: 45

Raadsma HW, Thomson PC, Zenger KR, Cavanagh C, Lam MK, Jonas E, Jones M, Attard G, Palmer D, Nicholas FW (2009b) Mapping quantitative trait loci (QTL) in sheep. I. A new male framework linkage map and QTL for growth rate and body weight. Genetics Selection Evolution 41: 34

Randhawa IAS, Khatkar MS, Thomson PC, Raadsma HW (2014) Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. BioMed Central Genetics 15: 34

Sarig O, Bercovici S, Zoller L, Goldberg I, Indelman M, Nahum S, Israeli S, Sagiv N, De Morentin HM, Katz O, Baum S, Barzilai A, Trau H, Murrell DF, Bergman R, Hertl M, Rosenberg S, Nöthen MM, Skorecki K, Schmidt E, Zillikens D, Darvasi A, Geiger D, Rosset S, Ibrahim SM, Sprecher E (2012) Population-specific association between a polymorphic variant in *ST18*, encoding a pro-apoptotic molecule, and *Pemphigus Vulgaris*. Journal of Investigative Dermatology 132: 1798-1805

Scholtz AJ, Cloete SWP, Van Wyk JB, Kruger ACM, Van der Linde TCD (2010) Influence of divergent selection for reproduction on the occurrence of breech strike in mature Merino ewes. Animal Production Science 50: 203-209

Snelling WM, Allan MF, Keele JW, Kuehn LA, McDaneld T, Smith TPL, Sonstegard TS, Thallman RM, Bennett GL (2010) Genome-wide association study of growth in crossbred beef cattle. Journal of Animal Science 88: 837-848

Snowder GD, Fogarty NM (2009) Composite trait selection to improve reproduction and ewe productivity: A review. Animal Production Science 49: 9-16

Tang K, Thornton K, Stoneking M (2007) A new approach for using genome scans to detect recent positive selection in the human genome. Public Library of Science 5: e171

The Bovine HapMap Consortium (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science: 324: 528-532

Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, Honeycutt RL, Crandall KA, Lundeberg J, Wayne RK (1997) Multiple and ancient origins of the domestic dog. Science 276: 1687-1689

Walsh B (2008) Using molecular markers for detecting domestication, improvement, and adaptation genes. Euphytica 161:1-17

Weigel KA (2006) Prospects for improving reproductive performance through genetic selection. Animal Reproduction Science 96: 323-330

Zhang H, Wang A, Wang S, Li H (2012) Progress of genome wide association study in domestic animals. Journal of Animal Science and Biotechnology 3: 26-36

Chapter 6:

Summary and conclusions

6.1. Summary of research rationale and main findings:

The projected increase in the global food demand (FAO 2013) and the threat of climate change (Nardone *et al*. 2010) looms over livestock industries. The adaptability and continued production of livestock animals in challenging and changing environmental conditions (also referred to as robustness) has thus become increasingly important to ensure efficient and sustainable livestock farming (Knapp 2005, Star *et al*. 2008, Klopčič & Kuipers 2009, Strandberg 2009). Reproduction ability contributes to the overall robustness of livestock (Klopčič & Kuipers 2009, Strandberg 2009), in addition to being a determining factor in efficiency and profitability of sheep farming (Olivier 1999, De Graaf 2010). Increases in reproductive potential, as seen in traits such as number of lambs born (NLB), number of lambs weaned (NLW) and total weight of lamb weaned (TWW), results in efficient land use and an increase in the number of slaughter lambs, which translates into higher profit margins. Increasing the reproductive potential of sheep is, however, complicated by the fact that reproductive traits are notoriously difficult to quantify as most of these traits are lowly heritable, sex-limited and only measureable after participation in at least one reproductive opportunity (Safari *et al.* 2005, Notter 2012). Traditional selective breeding strategies are constrained by the low heritability of reproduction traits and together with the costly and complex nature of trait recording results in relatively slow industry genetic gains. The aim of the current study was to identify genomic regions underlying reproduction and robustness traits by exploiting signatures of selection in a divergently selected South African Merino flock. This was achieved by employing a whole-genome single nucleotide polymorphism (SNP) genotyping strategy using the OvineSNP50 chip. Following this, the second aim was to elucidate quantitative loci (QTL) and/or genes associated with selection for reproduction to shed more light on the underlying mechanisms associated with variation in reproductive traits. The main research findings are summarised in sections 6.1.1 to 6.1.4.

6.1.1. Variation in reproduction traits of the Elsenburg Merino flock:

The Elsenburg Merino flock has been divergently selected for the ability to raise multiple offspring and has resulted in a High and Low line that differ markedly with regard to the NLB, NLW, TWW and lamb survival (Cloete *et al*. 2007, 2009), and therefore served as an ideal study cohort for investigating the effect of selection for reproduction traits. Analyses of NLB, NLW and TWW indicated an increase in the High line, while a general decrease occurred in the Low line since the selection experiment commenced in 1985. Although both lines have clearly exhibited a phenotypic response to selection pressure, the response was less pronounced in the Low line and more variation in reproduction traits still exists within this line. The difference in the selection response and variation within the lines may be due to the decrease in the number of animals in the Low line during the course of the selection experiment. This is a direct consequence of the selection pressure applied to this flock and has resulted in fewer offspring being born and weaned to the Low line. This has led to a reduction in the number of selection candidates and may therefore have resulted in a greater percentage of the offspring being selected for breeding; thereby reducing the selection pressure applied to this line. The small sample size in the Low line may also have influenced the degree of variation within this line as measures of variation are sensitive to the number of samples tested and a small sample size may therefore inflate these values. Nonetheless, selection for a reduced NLW is in direct contradiction with the continued survival of the line and therefore variation in reproduction traits may have been maintained through either natural selection or indirect (unintentional) selection.

The F1 crosses of the High and Low lines exhibited intermediate values for NLB, NLW and TWW that were similar to the midparent value of the two lines as would be expected in the absence of heterosis. Unexpected results were obtained when the mean trait values of the backcrosses to either the High or the Low lines were compared to that of the parental groups. The backcrosses to the Low line exhibited mean raw values for NLB, NLW and TWW that were greater than that of both parental groups, while the backcrosses to the High line exhibited mean trait values that were below the average of the parental groups. The genetic background of the parental lines may therefore have played a role in the reproductive performance of hybrid offspring and this may in turn indicate that the interaction of several genes is responsible for variation in reproductive traits in these lines. Information relating to the genetic architecture of

the genome was thus needed to identify the genes and/or quantitative trait loci (QTL) responsible for variation in reproduction traits and to determine the mechanisms involved with variation in reproductive traits. The SNP genotyping strategy described in Chapters 3 to 5 (summarised in sections 6.1.2 to 6.1.4) therefore enabled investigation of the genomic regions potentially underlying variation in reproduction traits in the Elsenburg Merino flock.

6.1.2. Evaluation of the OvineSNP50 genotyping array in four South African sheep breeds:

The OvineSNP50 Beadchip enables high-throughput genotyping of more than 54 000 SNPs across the *ovine* genome in a time- and cost-effective manner. The OvineSNP50 chip was selected to obtain genomic information that could aid in the identification of genomic regions under selection in the Elsenburg Merino lines. Due to possible SNP ascertainment bias (Kijas *et al.* 2012), the utility of the OvineSNP50 chip needed to be determined for South African sheep before use in downstream analyses. As the use of the OvineSNP50 chip may extend to other South African sheep breeds in future, samples from the Elsenburg Merino flock together with additional South African Merino, Blackheaded Dorper, South African Mutton Merino (SAMM) and Namaqua Afrikaner samples were also assessed. Genotyping results indicated that the samples from the Elsenburg Merino flock had low call rates and call quality values. These samples underwent multiple freeze-thaw cycles and it was concluded that poor DNA quality due to incorrect sample storage was the cause of low call rates and call quality. However, the number of polymorphic loci and minor allele frequencies (MAFs) of additional South African Merino samples were comparable to those reported for international Merino breeds. The percentage polymorphic loci and MAF of the SAMM were slightly less than the corresponding values for its progenitor breed, the German Merino, and the additional South African Merino samples used in the current study. Values similar to those of the SAMM were observed for the number of polymorphic loci and MAF of the Dorper samples. The Namaqua Afrikaner samples exhibited the lowest number of polymorphic loci and MAF value relative to the other breeds included in the current study. The results obtained for the Namaqua Afrikaner are however, comparable to those reported for Namaqua Afrikaners sheep included in the International sheep HapMap project (Kijas *et al.* 2012, www.sheephapmap.org). The

Dorper, SAMM and South African Merino samples had MAFs>0.3 for most loci, whereas the Namaqua Afrikaner samples exhibited a larger percentage of loci with rare alleles in comparison to the three other breeds. A similar trend has been reported for BovineSNP50 (Illumina) genotyping results where African and indicine cattle breeds exhibited a greater percentage of SNP loci with low MAFs and less polymorphic loci in comparison to commercial taurine breeds (Matukumalli *et al.* 2009). The limited number of individuals and breeds representing African and other indigenous breeds during SNP discovery, were most likely the cause of the aforementioned genotype results. Matukumalli *et al.* (2009) therefore reports the BovineSNP50 chip's utility in European and other commercial breeds, but warns of reduced power in African and indicine breeds. Similarly, the OvineSNP50 chip will be of greater use in commercial and/or American and European sheep breeds, such as the Merino, in comparison to the Namaqua Afrikaner and other related breeds. The SNP ascertainment bias may affect inferences relating to population genetic estimates or genomic selection strategies and may therefore hamper studies on breeds not extensively included during SNP discovery.

6.1.3. Identification of genomic regions subject to selection by the use of an F_{ST} outlier approach:

In literature, mutations in three genes, *GDF9, BMPR1B*, *and BMP15*, located on chromosome 5, 6 and the X-chromosome, have been identified as the causative mutations resulting in an increased reproductive output of highly proliferative lines (Davis 2005, McNatty *et al*. 2005, Juengel *et al*. 2013). To determine whether the aforementioned genes and surrounding genomic regions may have been under selection and associated with the differential reproductive output of the High and Low lines of the Elsenburg Merino flock, the genotype data were partitioned according to the 27 *ovine* chromosomes to allow pinpointing of chromosomal regions subject to selection. An F_{ST} outlier approach was used to identify SNPs subject to selection by respectively applying a Bayesian and frequentist method. A total of 31 SNP loci located on 13 chromosomes, were identified to be putatively subject to selection. Although SNPs were identified on chromosome 6 and the X-chromosome, these loci were not in close proximity to the aforementioned mutations associated with highly proliferative lines. The fact that putative loci under selection are located across several genomic locations, confirms the premise that several QTL across the genome may

influence variation in reproduction traits. Two loci subject to selection were in close proximity to genes associated with sperm morphology and stress response, respectively. Other studies of the Elsenburg Merino flock have confirmed that significant differences in sperm morphometric characteristics (Boshoff 2014) and stress coping ability (Cloete *et al.* 2005a, Cloete *et al*. 2010, Hough *et al*. 2010) occurs between the lines.

6.1.4. *Identification of regions of homozygosity:*

Recent directional selection may result in a selection sweep that manifests in extended regions of homozygosity (ROH) in the genome (Nielsen 2005, Walsh 2008). The SNP genotype data of the Elsenburg Merino flock was analysed to detect genomic regions under selection by identifying ROH. Accurate identification of ROH is dependent on sufficient SNP coverage and therefore, call rate stringency was increased for the current genotype data set, thereby excluding a greater number of underperforming samples, while increasing the number of SNPs by 23%. The SNP genotype data from the High line, Low line and a combined sample of the two lines were screened for regions of homozygosity using a sliding window approach. Regions spanning at least 100kb and ten consecutive homozygous SNPs (allowing for one heterozygous and five missing calls) were matched across individuals in the three respective groups. Nine, two and one homozygous region(s) overlapping in >95% of individuals were identified in the Low line, High line and combined sample, respectively. Although the total (union) ROH across all individuals in a group spanned a mean distance of 18 231kb, the overlapping (consensus) region across individuals had a mean distance of 356kb. Some of the annotated genes in the proximity of the putative ROH had functions relating to sperm morphology and fertilisation (Liu *et al.* 2011) as well as cell proliferation and apoptosis (Marchong *et al.* 2010, Sarig *et al.* 2012, Jandrig *et al.* 2004). The location of several *ovine* QTL reported in literature, overlapped with union and consensus regions identified in the current study. Overlap was seen for QTL related to aseasonal reproduction (Mateescu & Thonney 2010), testis weight (Fullard *et al*. 2006), milk traits (Raadsma *et al.* 2009a, Crisà *et al.* 2010, Mateescu & Thonney 2010), parasite resistance (Davies *et al.* 2006, Gutiérrez-Gil *et al.* 2009), as well as production traits, such as body weight at 56 weeks (Raadsma *et al.* 2009b) and carcass characteristics (Cavanagh *et al*. 2010). Two regions indicative of selection

sweep in European dairy sheep overlapped with two union regions and a consensus region identified in the current study (Gutiérrez-Gil *et al.* 2014).

It is of interest to note that other studies concerned with the phenotypic differences between the Elsenburg Merino lines have indicated significant differences in the milk production (Cloete *et al.* 2011), yearling weight, carcass characteristics and lamb survival (Cloete *et al*. 2004, 2005b, 2009) as well as other traits, such as the animal's reaction to humans in an arena test (Cloete *et al.* 2010), stress coping ability (Cloete *et al.* 2005b, Hough 2012) and susceptibility to breech blowfly strike (Scholtz *et al*. 2010). These results have confirmed the complexity of the genetic architecture of reproduction traits and that several components may add to the net-reproduction of the Elsenburg Merino flock.

6.2. Limitations of the current study:

Other studies have indicated linkage disequilibrium to decay at distances <30kb (Kijas *et al.* 2012, Kijas *et al.* 2014) in the *ovine* genome and therefore the ~50kb distances between SNPs included on the OvineSNP50 chip may already be insufficient to capture all variation present. The exclusion of large numbers of SNP loci in the current study resulted in the mean distance between SNPs increasing beyond ~50kb and may have resulted in inadequate SNP coverage to identify all regions affected by the selection pressure applied to the Elsenburg flock. Therefore, many regions subject to selection may have remained undiscovered. Ensuring the DNA quality of samples destined for whole-genome genotyping is therefore vital in ensuring adequate SNP coverage and a representative sample number. Future studies employing genotype data obtained from samples with adequate DNA quality will benefit from the increased SNP coverage of the genome.

Due to the unavailability of samples from males and Low line individuals, unequal numbers of male and female, as well as High and Low line individuals were included in the current study. Some studies have reported the exclusion of markers on the X-chromosome when including male and female individuals in the sample cohort (Riggio *et al.* 2013, Randhawa *et al.* 2014), however, other studies based on selection sweep analysis in sheep have not necessarily done so (Moioli *et al*. 2013, Gutiérrez-Gil *et al.* 2014). Chromosome-specific partitioning of the genotype data (in Chapter 4) and the

inclusion of a combined sample of an equal number of High and Low line individuals (in Chapter 5) limited the influence of gender-specific chromosomal differences and overrepresentation of High line individuals. The potential bias towards homozygosity in male X-chromosome data was remedied by enforcing a minimum cut-off of >95% for overlap within a group of samples (in Chapter 5). Exclusion of X-chromosome results or partitioning of male and female individuals prior to F_{ST} analyses may be considered in future.

During quality control analyses, some previous *ovine* studies have excluded SNP loci that were not in Hardy-Weinberg equilibrium (Hayes *et al.* 2012, Moradi *et al.* 2012, Demars *et al.* 2013). The Elsenburg Merino flock defies certain assumptions of Hardy-Weinberg equilibrium, most notably, the absence of selection. Loci under selection were also sought in the current study and therefore, loci were not excluded based on Hardy-Weinberg equilibrium.

Genetic drift and inbreeding may result in the fixation of alleles and the occurrence of long stretches of homozygosity (Frankham *et al.* 2002, Keller *et al.* 2011, Bosse *et al*. 2012) and therefore the methodology used in the current study may have been subject to the identification of drift and inbreeding instead of regions under recent selection. Sheep husbandry practices, such as the use of a limited number of breeding rams and the limited number of animals maintained within the respective lines, also contribute to the potential for the occurrence of drift and inbreeding. However, great care was taken to select animals that were least inbred and least related. In fact, inbreeding coefficients calculated from SNP data (in Chapter 5) were negative for both the High and Low lines, thereby indicating an outbred sample cohort. By grouping the lines together, shared ROH could be identified that may result from historical inbreeding, drift and/or selection. However, ROH identified in the High and Low lines did not overlap with regions indicated in the combination sample. The fact that ROH overlapped with regions identified for QTL relating to reproduction, production and parasite resistance, may also indicate the validity of the results of the current study.

Domestic sheep are divided into more than 2 400 breeds and comprise 17% of all known livestock breeds (DAD-IS 2014). Domestic sheep have also been shown to have higher levels of genetic diversity than most other livestock species (Kijas *et al.*

2012) and the extent of LD, SNP distribution and allele frequencies may differ significantly between breeds (Kijas *et al.* 2012, Kijas *et al*. 2014). As a result, the genomic regions identified in the current study may not be directly applicable to variation in reproduction traits in other sheep lines or breeds. In future, the transferability of the current study results may be evaluated in other closely related lines or breeds. However, it is anticipated that results of the current study may not be transferrable to more distantly related breeds due to differences in genetic architecture.

6.3. Conclusions and future research directions:

Although annotated genes with putative roles in reproduction were identified, the exact mechanism of involvement with variation in reproduction traits could not be determined for all regions and genes. However, the fact that putative regions of homozygosity identified in the current study overlap with QTL for several reproduction, milk, production and parasite resistance traits, does shed some light on the possible role of genes found within regions identified in the current study. The overlap between QTL for production and parasite resistance with putative ROH identified in the current study may indicate that several, seemingly unrelated, traits add to the overall net-reproduction. Divergent selection pressure placed on the ability to raise multiple offspring has resulted in the divergence between the High and Low line for several other traits as well, as reviewed by Hough *et al.* (2013). This possibly indicates that indirect selection of several other correlated traits may have occurred. If a positive correlation exists between increases in reproduction and undesirable traits, this may influence marker-assisted breeding strategies for reproduction traits. However, selection pressure placed on an increase in the number of lambs weaned in the High line has resulted in increases in favourable traits such as stress coping ability, mothering ability, milk yield, hogget weight, weaning weight and resistance to breech strike (Cloete *et al.* 2003, 2005a, 2005b, 2009, 2010, Scholtz *et al*. 2010). It is therefore not expected that selection for an increase in reproductive output should result in indirect selection for potentially unfavourable traits. Nonetheless, a holistic selection strategy taking several important traits, such as robustness, reproduction and production into account while maintaining genetic diversity, is necessary to breed animals with the ability to produce and reproduce in an efficient manner and thereby ensure profitable and sustainable farming (Knap *et al*. 2005, Strandberg 2009, Clark

et al. 2013). The experimental design of the current study did not allow the quantification of the SNP effects on the variation in reproduction traits and it is anticipated that the genomic regions identified in the current study may only account for a small percentage of the observed variation in reproduction traits (Goddard & Hayes 2009). Studies with a greater coverage of the genome and larger sample size may identify numerous additional genomic regions associated with variation in reproduction traits.

Internationally, several commercial livestock industries have moved away from marker-assisted selection strategies and have invested in genomic selection of several commercially important traits (Fan *et al.* 2010). Genomic selection utilises a whole-genome SNP panel (such as the OvineSNP50 chip) to quantify variation in several traits of interest and thereby enable the accurate prediction of the genomic breeding value of genotyped selection candidates. Genomic selection strategies are well suited to the selection for quantitative traits that may be influenced by numerous QTL spread throughout the genome, each contributing a small percentage of the overall variation to the trait of interest (Goddard & Hayes 2009). Genomic breeding practices have already been implemented in the Australian and New Zealand sheep industries and increased accuracy of genomic breeding values in comparison to pedigree-based BLUP estimated breeding values have been reported (Swan *et al*. 2012, Rowe *et al.* 2013, Auvrey *et al.* 2014). A selection strategy based on a single set of SNP markers associated with reproduction traits, therefore seems inadvisable at this stage.

The current study added to the knowledge relating to the effect of selection on the *ovine* genome and possible genomic regions and genes involved with variation in reproduction traits in the South African Merino. It is expected that the results obtained from the current study will encourage interest in whole-genome selection strategies for a combination of reproduction, production and welfare traits for South African sheep breeds. This study also confirmed the value of the Elsenburg Merino selection experiment and highlights the potential for selective breeding, facilitated by the high levels of genetic diversity within the South African Merino breed.

Future genomic studies on the Elsenburg Merino flock will benefit from ensuring the DNA quality of samples destined for whole-genome SNP genotyping. Also, studies

utilising a denser SNP marker panel may uncover additional regions subject to selection in the Elsenburg lines. Use of the 600K *ovine* SNP chip (Kijas *et al*. 2014) or even whole-genome sequencing (Jiang *et al.* 2014) may offer further insight in the effect of selection on the genome of the Elsenburg Merino line and may facilitate the identification of additional genomic regions and genes involved with variation in reproduction and other robustness traits in this flock. Future genotyping results of additional Elsenburg Merino samples may also facilitate the imputation (Hayes *et al.* 2012) of genotype calls that were excluded from the current study cohort due to low quality scores.

Due to the limited number of individuals and breeds used during SNP discovery, OvineSNP50 genotyping results of indigenous African sheep breeds may be influenced by SNP ascertainment bias. Scope therefore exists for the identification of a whole-genome SNP panel customised for genotyping indigenous South African or African breeds in future. Such unique genetic resources are often highly adapted to specific low-input environments (Eisler *et al*. 2014). This ability to adapt to low-input environments may validate the need to better comprehend the genetic architecture of such breeds. However, an endeavour of this nature is costly and therefore the outcomes, such as the potential for whole-genome selection strategies or conservation efforts, need to justify the input cost.

Reproduction and robustness traits play an important role in the sheep breeding, however, relatively little is known about genes and/or QTL involved with variation in these traits. The current study has added to the knowledge of the genetic architecture of reproduction and robustness traits of South African Merino sheep. These results may be used in future to guide decisions relating to selection strategies of South African sheep breeds in general.

6.4. References:

Auvray B, McEwan JC, Newman SAN, Lee M, Dodds KG (2014) Genomic prediction of breeding values in the New Zealand sheep industry using a 50K SNP chip. Journal of Animal Science 92: 4375-4389

Boshoff NH (2014) The influence of genotype on sperm motility and sperm head morphometry of Merino (*Ovis aries*) sheep. MSc dissertation, Stellenbosch University, South Africa

Bosse M, Megens H-J, Madsen O, Paudel Y, Frantz LAF, Richard LB, Crooijmans MA, Groenen MAM (2012) Regions of homozygosity in the *porcine* genome: consequence of demography and the recombination landscape. Public Library of Science Genetics 8: e1003100

Cavanagh CR, Jonas E, Hobbs M, Thomson PC, Tammen I, Raadsma HW (2010) Mapping Quantitative Trait Loci (QTL) in sheep. III. QTL for carcass composition traits derived from CT scans and aligned with a meta-assembly for sheep and cattle carcass QTL. Genetics Selection Evolution 42: 36

Clark SA, Kinghorn BP, Hickey JM, Van der Werf JHJ (2013) The effect of genomic information on optimal contribution selection in livestock breeding programs. Genetics Selection Evolution 45: 44

Cloete JJE, Cloete SWP, Hoffman LC Fourie JE (2004) Slaughter traits of Merino sheep divergently selected for multiple rearing ability. South African Journal of Animal Science 34: 189-196

Cloete JJE, Cloete SWP, Hoffman LC (2005a) Behaviour of Merinos divergently selected for multiple rearing ability in response to external stimuli. Small Ruminant Research 60: 227-236

Cloete JJE, Cloete SWP, Hoffman LC (2010) Behaviour of Merinos divergently selected for multiple rearing ability in response to external stimuli. Proceedings of the World Congress on Genetics Applied to Livestock Production, August 2010, Leipzig, Germany

Cloete SWP, Van Wyk JB, Scholtz AJ, Gilmour AR (2003) Genetic and environmental parameters for lambing behaviour in Merino lines divergently selected for ewe maternal rearing ability. Proceedings of the New Zealand Society of Animal Production 63: 169-172

Cloete SWP, Olivier JJ, Van Wyk JB, Schoeman SJ, Erasmus GJ (2005b) Genetic parameters and trends for hogget traits in Merino lines divergently selected for multiple rearing ability. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 16: 24-27

Cloete SWP, Olivier JJ, Olivier WJ (2007) Genetic change in South African Merino resource flocks. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 17: 320-323

Cloete SWP, Misztal I, Olivier JJ (2009) Genetic parameters and trends for lamb survival and birth weight in a Merino flock divergently selected for multiple rearing ability. Journal of Animal Science 87: 2196-2208

Cloete SWP, Snyman MA, Scholtz AJ (2011) Genetic and environmental parameters of milk production and milk composition in South African Merinos. Proceedings of the Association of the Advancement of Animal Breeding and Genetics 19: 419-422

Crisà A, Marchitelli C, Pariset L, Contarini G, Signorelli F, Napolitano F, Catillo G, Valentini A, Moioli B (2010) Exploring polymorphisms and effects of candidate genes on milk fat quality in dairy sheep. Journal of Dairy Science 93: 3834-3845

DAD-IS (Domestic Animal Diversity Information System) (2014) Food and Agriculture Organization of the United Nations. Accessed May 2014, from: http://dad.fao.org/

Davies G, Stear MJ, Benothman M, Abuagob O, Kerr A, Mitchell S, Bishop SC (2006) Quantitative trait loci associated with parasitic infection in Scottish blackface sheep. Heredity 96: 252-258

Davis GH (2005) Major genes affecting ovulation rate in sheep. Genetics Selection Evolution 37: S11-S23

De Graaf SP (2010) Reproduction. In: Cottle DJ (ed.) The international sheep and wool handbook, pg. 189-222. Nottingham University press, Nottingham. England

Demars J, Fabre S, Sarry J, Rossetti R, Gilbert H, Persani L, Tosser-Klopp G, Mulsant P, Nowak Z, Drobik W, Martyniuk E, Bodin L (2013) Genome-wide association studies identify two novel *BMP15* mutations responsible for an atypical hyperprolificacy phenotype in sheep. Public Library of Science Genetics 9: e1003482

Eisler MC, Lee MRF, Tarlton JF, Martin GB, Beddington J, Dungait JAJ, Greathead H, Liu J, Mathew S, Miller H, Misselbrook T, Murray P, Vinod VK, Van Saun R, Winter M (2014) Steps to sustainable livestock. Nature 507: 32-34

Fan B, Du Z-Q, Gorbach DM, Rothchild MF (2010) Development and application of high-density SNP arrays in genomic studies of domestic animals. Asian-Australian Journal of Animal Sciences 23: 833-847

FAO (Food and Agriculture Organisation of the United Nations) (2013) FAO Statistical yearbook: World food and agriculture. ISBN 978-92-5-107396-4

Frankham R, Ballou JD, Briscoe DA (2002) Evolution in large populations I. Natural selection and adaptation. In: Introduction to conservation genetics, pg. 126-152. Cambridge University Press, Cambridge, UK

Fullard K, Margawati PC, Thomsom PC, Raadsma HW (2006) QTL for testis size in ram lambs as an indicator of ovulation rate in an Indonesian Thin tail by Merino resource flock. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, 13-18 August, Belo Horizonte, Brazil

Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Reviews Genetics 10: 381-391

Gutiérrez-Gil B, Arranz JJ, Pong-Wong R, García-Gámez E, Kijas J, Wiener P (2014) Application of selection mapping to identify genomic regions associated with dairy production in sheep. Public Library of Science ONE 9: e94623

Gutiérrez-Gil B, Pérez J, Alvarez L, Martínez-Valladares M, de la Fuente LF, Bayón Y, Meana A, San Primitivo F, Rojo-Vázquez FA (2009) Quantitative trait loci for resistance to trichostrongylid infection in Spanish Churra sheep. Genetics Selection Evolution 41: 46

Hayes BJ, Bowman PJ, Daetwyler HD, Kijas JW, Van der Werf JHJ (2012) Accuracy of genotype imputation in sheep breeds. Animal Genetics 43: 72-80

Hough D, Swart P, Cloete JJE, Cloete SWP (2010) Searching for SNPs that affect sheep robustness: *CYP17* SNP affects behavioural responses to psychological stress. Proceedings for the Advancement of Animal Breeding and Genetics 20: 503-506

Hough D (2012) Comparison of two CYP17 isoforms: implications for cortisol production in the South African Merino. PhD dissertation, Stellenbosch University, South Africa

Hough D, Swart P, Cloete S (2013) Exploration of the hypothalamic-pituitary-adrenal axis to improve animal welfare by means of genetic selection: Lessons from the South African Merino. Animals 3: 442-474

Jandrig B, Seitz S, Hinzmann B, Arnold W, Micheel B, Koelble K, Siebert R, Schwartz A, Ruecker K, Schlag PM, Scherneck S, Rosenthal A (2004) *ST18* is a breast cancer tumor suppressor gene at human chromosome 8q11.2. Oncogene 23: 9295-9302

Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, Faraut T, Wu C, Muzny DM, Li Y, Zhang W, Stanton J-A, Brauning R, Barris WC, Hourlier T, Aken BL, Searle SMJ, Adelson DL, Bian C, Cam GR, Chen Y, Cheng S, DeSilva U, Dixen K, Dong Y, Fan G, Franklin IR, Fu S, Fuentis-Utrilla P, Guan R, Highland MA, Holder ME, Huang G, Ingham AB, Jhangiani SN, Kalra D, Kovar CL, Lee SL, Liu, W, Liu X, Lu C, Lv T, Mathew T, McWilliam S, Menzies M, Pan S, Robelin D, Servin B, Townley D, Wang W, Wei B, White SN, Yang X, Ye C, Yue Y, Zeng P, Zhou Q, Hansen JB, Kristiansen K, Gibbs RA, Flicek P, Warkup CC, Jones

HE, Oddy VH, Nicholas FW, McEwan JC, Kijas JW, Wang Y, Worley KC, Archibald AL, Cockett N, Xu X, Wang W, Dalrymple BP (2014) The sheep genome illuminates biology of the rumen and lipid metabolism. Science 344: 1168-1173

Juengel JL, Davis GH, McNatty KP (2013) Using sheep lines with mutations in single genes to better understand ovarian function. Reproduction 146: 111-123

Keller MC, Visscher PM, Goddard ME (2011) Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics 189: 237-249

Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, Servin B, McCulloch R, Whan V, Gietzen K, Paiva S, Barendse W, Ciani E, Raadsma H, McEwan J, Dalrymple B, other members of the International Sheep Genomics Consortium10 (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. Public Library of Science ONE Biology 10: e1001258

Kijas JW, Porto-Neto L, Dominik S, Reverter A, Bunch R, McCulloch R, Hayes BJ, Brauning R, McEwan J, the International Sheep Genomics Consortium (2014) Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. Animal Genetics 45: 754-757

Klopčič M, Kuipers A (2009) Perception of robustness traits in breeding goals for dairy cattle in a new EU country. In: Klopčič M, Reents R, Philipsson J, Kuipers A (eds.) Breeding for robustness. EAAP publication 126: 249-262

Knapp PW (2005) Breeding robust pigs. Australian Journal of Experimental Agriculture 45: 763-773

Liu Y, Jiang M, Li C, Yang P, Sun H, Tao D, Zhang S, Ma Y (2011) Human t-complex protein 11 (TCP11), a testis-specific gene product, is a potential determinant of the sperm morphology. Tohuku Journal of Experimental Medicine 224: 111-117

Marchong MN, Yurkowski C, Ma C, Spencer C, Pajovic S, Gallie BL (2010) Cdh11 acts as a tumor suppressor in a murine retinoblastoma model by facilitating tumor cell death. Public Library of Science Genetics 6: e1000923

Mateescu RG, Thonney ML (2010) Genetic mapping of quantitative trait loci for aseasonal reproduction in sheep. Animal Genetics 41: 454-459

Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, OÇonnell J, Moore SS, Smith TPL, Sonstegard TS, Van Tassell CP (2009) Development and characterization of a high density SNP genotyping assay for cattle. Public Library of Science ONE 4: e5350

McNatty KP, Galloway SM, Wilson T, Smith P, Hudson NL, O'Connell A, Bibby AH, Heath DA, Davis GH, Hanrahan JP (2005) Physiological effects of major genes affecting ovulation rate in sheep. Genetics Selection Evolution 37: S25-S38

Moioli B, Scatà MC, Steri R, Napolitano F, Catillo G (2013) Signatures of selection identify loci associated with milk yield in sheep. BioMed Central Genetics 14: 76

Moradi MH, Nejati-Javaremi A, Moradi-Shahrbabak M, Dodds KG, McEwan JC (2012) Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. BioMed Central Genetics 13: 10

Nardone A, Ronchi B, Lacetera N, Ranieri MS, Bernabucci U (2010) Effects of climate changes on animal production and sustainability of livestock systems. Livestock Science 130: 57-69

Nielsen R (2005) Molecular signatures of natural selection. Annual Review of Genetics 30: 197-218

Notter DR (2012) Genetic improvement of reproductive efficiency of sheep and goats. Animal Reproduction Science 130: 147-151

Olivier JJ (1999) The South African Merino performance testing scheme. In: Rising to the challenge – Breeding for the 21st century customer. Beef Industry and CRC for Premium Quality Wool Industry Symposia. Supplement to the Proceedings of the Association for the Advancement of Animal Breeding and Genetics 13: 119-124

Raadsma HW, Jonas E, McGill D, Hobbs M, Lam MK, Thomson PC (2009a) Mapping quantitative trait loci (QTL) in sheep. II. Meta-assembly and identification of novel QTL for milk production traits in sheep. Genetics Selection Evolution 41: 45

Raadsma HW, Thomson PC, Zenger KR, Cavanagh C, Lam MK, Jonas E, Jones M, Attard G, Palmer D, Nicholas FW (2009b) Mapping quantitative trait loci (QTL) in sheep. I. A new male framework linkage map and QTL for growth rate and body weight. Genetics Selection Evolution 41: 34

Randhawa IAS, Khatkar MS, Thomson PC, Raadsma HW (2014) Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. BioMed Central Genetics 15: 34

Riggio V, Matika O, Pong-Wong, Stear MJ, Bishop SC (2013) Genome-wide association and regional heritability mapping to identify loci underlying variation in nematode resistance and body weight in Scottish Blackface lambs. Heredity 110: 420-429

Rowe JB; Gill S, Banks RG; Van der Werf JHJ (2013) Genomics for the Australian sheep industry: from design to delivery. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 20: 14-17

Safari E, Fogarty NM, Gilmour AR (2005) A review of genetic parameter estimates for wool, growth, meat and reproduction traits in sheep. Livestock Production Science 92: 271-289

Sarig O, Bercovici S, Zoller L, Goldberg I, Indelman M, Nahum S, Israeli S, Sagiv N, De Morentin HM, Katz O, Baum S, Barzilai A, Trau H, Murrell DF, Bergman R, Hertl M, Rosenberg S, Nöthen MM, Skorecki K, Schmidt E, Zillikens D, Darvasi A, Geiger D, Rosset S, Ibrahim SM, Sprecher E (2012) Population-specific association between a polymorphic variant in *ST18*, encoding a pro-apoptotic molecule, and *Pemphigus Vulgaris*. Journal of Investigative Dermatology 132: 1798-1805

Scholtz AJ, Cloete SWP, Van Wyk JB, Kruger ACM, Van der Linde TCD (2010) Influence of divergent selection for reproduction on the occurrence of breech strike in mature Merino ewes. Animal Production Science 50: 203-209

Star L, Ellen ED, Uitdehaag K, Brom WA (2008) A plea to implement robustness into a breeding goal: poultry as an example. Journal of Agriculture and Environmental Ethics 21: 109-125

Strandberg E (2009) The role of environmental sensitivity and plasticity in breeding for robustness: Lessons from evolutionary genetics. In: Klopčič M, Reents R, Philipsson J, Kuipers A (eds.) Breeding for robustness. EAAP publication 126: 17-34

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

Walsh B (2008) Using molecular markers for detecting domestication, improvement, and adaptation genes. Euphytica 161: 1-17

Appendix:

Table A2.1: The mean number of lambs born (NLB), number of lambs weaned (NLW) and total weaning weight of lambs per joining (TWW)values for the High and Low lines as well as the standard errors of the mean.

	High line							Low line						
		Mean			Standard error				Mean			Standard error		
Year	n	NLB	NLW	TWW	NLB	NLW	TWW	n	NLB	NLW	TWW	NLB	NLW	TWW
1985	19	1.03	0.81	18.76	0.24	0.19	4.42	18	1.11	0.66	15.76	0.27	0.16	3.82
1986	27	0.97	0.74	17.66	0.19	0.15	3.46	29	1.11	0.66	15.76	0.27	0.16	3.82
1987	28	1.09	0.81	18.86	0.21	0.16	3.63	29	1.05	0.75	17.81	0.20	0.18	4.32
1988	24	1.21	0.84	18.73	0.25	0.18	3.91	21	0.93	0.66	14.48	0.18	0.12	2.74
1989	25	1.27	0.98	20.81	0.26	0.20	4.25	13	1.06	0.71	15.01	0.24	0.16	3.36
1990	24	1.30	0.95	21.12	0.27	0.20	4.40	15	1.13	0.67	14.82	0.33	0.19	4.28
1991	35	1.43	1.20	26.92	0.25	0.21	4.62	14	1.29	0.91	18.26	0.34	0.24	4.88
1992	24	1.43	0.93	20.74	0.30	0.19	4.32	17	1.16	0.95	18.97	0.32	0.26	5.26
1993	35	1.39	1.14	25.21	0.24	0.20	4.32	19	1.11	0.68	14.59	0.28	0.17	3.65
1994	27	1.46	1.11	22.61	0.29	0.22	4.44	19	0.96	0.70	14.39	0.23	0.17	3.39
1995	26	1.30	1.13	23.63	0.26	0.23	4.73	17	1.30	0.84	15.62	0.31	0.20	3.68
1996	37	1.44	0.98	19.95	0.24	0.16	3.32	14	1.12	0.87	16.59	0.28	0.22	4.15
1997	20	1.25	0.88	18.12	0.29	0.20	4.16	17	1.05	0.76	14.35	0.29	0.21	3.98
1998	25	1.35	1.05	20.28	0.28	0.21	4.14	15	0.82	0.68	13.54	0.21	0.17	3.38
1999	21	1.25	1.03	22.37	0.28	0.23	5.00	15	0.57	0.42	8.33	0.15	0.11	2.23
2000	21	1.40	1.26	27.98	0.31	0.28	6.26	8	0.81	0.59	12.96	0.22	0.16	3.46
2001	42	1.17	1.02	22.79	0.18	0.16	3.56	7	0.64	0.40	7.86	0.24	0.15	2.97
2002	20	1.30	1.12	26.49	0.30	0.26	6.08	4	0.99	0.74	13.84	0.40	0.30	5.65
2003	25	1.34	1.20	27.39	0.27	0.24	5.59	11	0.50	0.25	5.87	0.29	0.14	3.39
2004	22	1.38	1.08	23.33	0.30	0.24	5.09	4	0.95	0.64	13.21	0.30	0.20	4.18
2005	29	1.24	1.00	22.56	0.23	0.19	4.26	5	0.67	0.44	9.71	0.38	0.25	5.61
2006	34	1.45	1.09	22.67	0.25	0.19	3.95	11	1.06	0.82	15.92	0.53	0.41	7.96
2007	23	1.21	0.92	19.98	0.26	0.20	4.26	3	0.94	0.73	13.58	0.30	0.23	4.29
2008	38	1.55	1.30	28.92	0.04	0.03	0.65	3	1.08	0.92	15.76	0.77	0.65	11.14
2009	35	1.47	1.07	25.03	0.25	0.18	4.29	6	1.33	1.00	20.39	0.94	0.71	14.42

Table A2.2: The mean EBVs number of lambs born (NLB), number of lambs weaned (NLW) and total weaning weight of lambs per joining (TWW)values for the High and Low lines as well as the standard errors of the mean.

	High line							Low line						
		Mean			Standard error				Mean			Standard error		
	n	BVNL B	BVNL W	BVTW W	BVNL B	BVNL W	BVTW W	n	BVNL B	BVNL W	BVTW W	BVNL B	BVNL W	BVTW W
1985	19	-0.08	-0.02	-0.71	-0.02	0.00	-0.17	18	-0.07	-0.07	-1.84	-0.02	-0.02	-0.45
1986	27	-0.05	-0.02	-0.65	-0.01	0.00	-0.13	29	-0.04	-0.03	-0.76	-0.01	-0.01	-0.14
1987	28	0.01	0.04	1.02	0.00	0.01	0.20	29	-0.12	-0.07	-1.91	-0.02	-0.01	-0.36
1988	24	0.06	0.03	0.59	0.01	0.01	0.12	21	-0.01	-0.03	-1.19	0.00	-0.01	-0.27
1989	25	0.08	0.08	1.55	0.02	0.02	0.32	13	-0.06	-0.08	-2.09	-0.02	-0.02	-0.60
1990	24	0.10	0.07	1.88	0.02	0.01	0.39	15	0.03	-0.01	-0.82	0.01	0.00	-0.22
1991	35	0.10	0.11	3.20	0.02	0.02	0.55	14	-0.14	-0.08	-2.22	-0.04	-0.02	-0.62
1992	24	0.13	0.07	1.64	0.03	0.01	0.34	17	-0.08	-0.11	-3.04	-0.02	-0.03	-0.76
1993	35	0.14	0.15	4.11	0.02	0.02	0.71	19	-0.16	-0.12	-3.21	-0.04	-0.03	-0.76
1994	27	0.20	0.12	2.87	0.04	0.02	0.56	19	0.02	-0.09	-2.67	0.01	-0.02	-0.63
1995	26	0.18	0.17	4.54	0.04	0.03	0.91	17	-0.12	-0.10	-2.61	-0.03	-0.02	-0.65
1996	37	0.24	0.16	4.16	0.04	0.03	0.69	14	-0.10	-0.10	-2.56	-0.03	-0.03	-0.71
1997	20	0.20	0.16	4.32	0.05	0.04	0.99	17	-0.13	-0.11	-2.53	-0.03	-0.03	-0.63
1998	25	0.26	0.19	4.37	0.05	0.04	0.89	15	-0.28	-0.18	-4.39	-0.07	-0.05	-1.17
1999	21	0.20	0.19	5.18	0.04	0.04	1.16	15	-0.16	-0.13	-2.62	-0.04	-0.04	-0.70
2000	21	0.19	0.22	6.03	0.04	0.05	1.35	8	-0.22	-0.17	-4.36	-0.08	-0.06	-1.65
2001	42	0.16	0.17	4.54	0.02	0.03	0.71	7	-0.07	-0.10	-2.68	-0.03	-0.04	-1.09
2002	20	0.18	0.18	5.27	0.04	0.04	1.21	4	-0.27	-0.19	-4.82	-0.15	-0.11	-2.78
2003	25	0.17	0.20	5.35	0.03	0.04	1.09	11	-0.12	-0.10	-2.78	-0.04	-0.03	-0.88
2004	22	0.26	0.23	6.06	0.06	0.05	1.32	4	-0.19	-0.13	-4.07	-0.11	-0.08	-2.35
2005	29	0.18	0.19	5.26	0.03	0.04	0.99	5	-0.12	-0.10	-2.68	-0.06	-0.05	-1.34
2006	34	0.29	0.23	6.17	0.05	0.04	1.07	11	-0.15	-0.10	-2.71	-0.05	-0.03	-0.86
2007	23	0.17	0.17	5.04	0.04	0.04	1.07	3	-0.11	-0.06	-1.99	-0.08	-0.04	-1.41
2008	38	0.19	0.18	5.08	0.03	0.03	0.84	3	-0.09	-0.07	-1.86	-0.07	-0.05	-1.32
2009	35	0.16	0.16	4.64	0.03	0.03	0.78	6	-0.10	-0.11	-2.77	-0.04	-0.05	-1.24

Table A2.3: The mean number of lambs born (NLB), number of lambs weaned (NLW) and total weaning weight of lambs per joining (TWW)values for the High and Low lines, crossbreds and backcrosses as well as the standard deviation and standard error of the mean.

			Means			Standard deviations			Standard errors			Variance coefficients		
Group	n	Mean NOL	NLB	NLW	TWW	NLB	NLW	TWW	NLB	NLW	TWW	NLB	NLW	TWW
A	206	3.00	1.39	1.09	24.31	0.51	0.46	10.03	0.04	0.03	0.70	36.53	42.33	41.27
B	43	4.00	1.06	0.81	15.70	0.50	0.41	7.83	0.16	0.12	2.42	47.49	50.60	49.84
C	43	4.00	1.16	0.90	17.22	0.42	0.38	7.56	0.06	0.06	1.17	36.18	42.74	43.90
D	39	4.00	1.18	1.00	22.50	0.37	0.36	7.87	0.06	0.06	1.28	30.99	36.11	34.97
CD	82	4.00	1.17	0.95	19.73	0.39	0.37	8.11	0.04	0.04	0.90	33.56	39.57	41.09
E	19	3.00	1.29	1.05	20.69	0.50	0.61	12.27	0.12	0.14	2.89	38.35	58.17	59.31
F	19	3.00	1.14	0.98	23.04	0.49	0.41	9.17	0.12	0.10	2.16	43.06	41.32	39.80
EF	38	3.00	1.22	1.02	21.86	0.49	0.51	10.75	0.08	0.08	1.77	40.47	50.46	49.17
G	20	3.00	1.24	1.01	21.25	0.45	0.47	9.99	0.10	0.11	2.29	36.36	46.33	47.00
H	23	3.00	1.20	0.86	18.00	0.50	0.53	11.26	0.11	0.11	2.40	41.83	61.52	62.52
HG	43	3.00	1.22	0.93	19.51	0.47	0.49	10.56	0.07	0.08	1.63	38.42	53.30	54.12

NOL: Number of lambing opportunities; CD, EF and GH represent the combined results of these groups.

Table A4.1: The number of markers identified to be under selection by the Fdist2 and Bayesian method in Lositan and BayeScan, respectively.

		Lositan			BayeScan	
Chr	Total number of markers	Directional selection	Balancing selection	%	Directional selection	%
1	2349	95	7	4.34	6	0.26
2	2336	94	1	4.07	4	0.17
3	2308	116	1	5.07	3	0.13
4	1040	41	0	3.94	0	0.00
5	1065	19	0	1.78	0	0.00
6	908	41	0	4.52	2	0.22
7	1023	29	1	2.93	2	0.20
8	736	49	2	6.93	1	0.14
9	869	56	2	6.67	3	0.35
10	702	27	0	3.85	0	0.00
11	659	11	0	1.67	0	0.00
12	820	17	2	2.32	3	0.37
13	946	45	5	5.29	1	0.11
14	642	22	1	3.58	0	0.00
15	755	43	0	5.70	0	0.00
16	618	28	1	4.69	2	0.32
17	584	14	0	2.40	0	0.00
18	706	30	1	4.39	0	0.00
19	603	9	1	1.66	0	0.00
20	549	15	0	2.73	0	0.00
21	425	18	2	4.71	2	0.47
22	547	26	0	4.75	0	0.00
23	455	13	1	3.08	1	0.22
24	445	14	0	3.15	0	0.00
25	465	15	1	3.44	0	0.00
26	395	21	0	5.32	0	0.00
X	650	15	6	3.23	1	0.15
Unmapped	180	3	0	1.67	0	0.00
Total	23780	926	35	4.04	31	0.13

Chr: Chromosome

Table A4.2: The posterior probability and F_{ST} values calculated in BayeScan for putative markers under selection.

Chromosome	Marker	Probability	F_{ST}
1	OAR1_114418178.1	0.92519	0.25627
1	OAR1_138571443.1	0.88158	0.24305
1	OAR1_52263906.1	0.97059	0.29619
1	OAR1_52312510.1	0.94839	0.28897
1	s39542.1	0.79716	0.21888
1	s57436.1	0.92018	0.27642
2	OAR2_19076972.1	0.9972	0.33688
2	OAR2_39793505.1	0.86137	0.26916
2	OAR2_66083138.1	0.94379	0.26028
2	s53687.1	0.92599	0.28955
3	OAR3_182821715.1	0.86657	0.2335
3	s28031.1	0.95579	0.28855
3	s33128.1	0.88638	0.26509
6	OAR6_23745238.1	0.93859	0.21772
6	s08703.1	0.9976	0.2931
7	OAR7_19668788.1	0.9798	0.26575
7	OAR7_45129063.1	0.83477	0.24353
8	OAR8_30441759.1	0.93639	0.28563
9	OAR9_43466490.1	0.94059	0.23188
9	OAR9_91934740.1	0.89678	0.25024
9	s57595.1	0.92519	0.22867
12	OAR12_63154765.1	0.9848	0.25412
12	s25288.1	0.80456	0.20006
12	s44950.1	0.9788	0.24228
13	OAR13_80614774_X.1	0.91338	0.22282
16	OAR16_68784953.1	0.93179	0.23337
16	s03359.1	0.9968	0.30881
21	s36706.1	0.90118	0.27949
21	s40378.1	0.9798	0.27956
23	s33116.1	0.9768	0.26723
X	OARX_104046252.1	0.95739	0.31019

Table A5.1: Observed heterozygosity (He (obs)) and inbreeding coefficient (F_{IS}) calculated per line after LD pruning.

High line			Low line		
Animal ID	He (obs)	F_{IS}	Animal ID	He (obs)	F_{IS}
1020212	0.39	-0.14	1020693	0.37	-0.09
1020282	0.40	-0.17	1020763	0.40	-0.20
1020452	0.36	-0.07	1021433	0.38	-0.14
1020902	0.37	-0.08	1031123	0.39	-0.17
1021222	0.36	-0.06	1031813	0.34	0.00
1021452	0.36	-0.06	1040303	0.36	-0.09
1030062	0.35	-0.04	1040743	0.36	-0.08
1030082	0.36	-0.06	1040773	0.36	-0.07
1030212	0.36	-0.05	1050513	0.35	-0.05
1031932	0.35	-0.04	1050773	0.38	-0.13
1040542	0.33	0.04	1051073	0.37	-0.10
1040592	0.37	-0.10	1051733	0.36	-0.08
1041072	0.36	-0.05	1052153	0.37	-0.10
1041122	0.36	-0.06	1060423	0.35	-0.06
1050122	0.37	-0.09	1060563	0.36	-0.07
1050292	0.34	-0.01	1060913	0.37	-0.09
1050302	0.34	0.01	1061333	0.36	-0.07
1050592	0.35	-0.03	1062093	0.36	-0.07
1050812	0.37	-0.10	1062143	0.38	-0.12
1050852	0.36	-0.07	1062253	0.35	-0.05
1051482	0.36	-0.07	1062343	0.33	0.01
1051852	0.37	-0.10	1071603	0.36	-0.08
1052142	0.38	-0.13	1072143	0.34	-0.03
1060222	0.34	-0.01	1072203	0.34	-0.03
1060252	0.35	-0.02	1080133	0.35	-0.04
1060472	0.37	-0.09	1081593	0.36	-0.08
1060572	0.37	-0.09	1081863	0.35	-0.04
1060582	0.37	-0.08	1082613	0.34	-0.03
1061112	0.37	-0.10	1090493	0.37	-0.12
1061622	0.34	0.01	**Mean (SD)**	**0.36 (0.02)**	**-0.08 (0.05)**
1071022	0.34	0.00			
1071392	0.35	-0.04			
1080492	0.35	-0.03			
1080612	0.34	0.00			
1080652	0.36	-0.05			
1080772	0.35	-0.04			
1080932	0.37	-0.08			
1080942	0.36	-0.05			
1081212	0.35	-0.04			
1081242	0.36	-0.06			
1081642	0.36	-0.06			
1081652	0.36	-0.07			
1082752	0.37	-0.09			
1090292	0.37	-0.08			
1091862	0.35	-0.04			
1093442	0.35	-0.04			
Mean (SD)	**0.36 (0.01)**	**-0.06 (0.04)**			