Genetic characterization of smallholder sheep flocks in the Western Cape, South Africa.

by

Gabriella Venter

Thesis presented in partial fulfilment of the requirements for the degree of
Master of Science (Agriculture)

at
Stellenbosch University
Animal Science, Faculty of AgriSciences

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions and conclusions arrived at are those of the author and are not necessarily to be attributed to the NRF.

Supervisor: Dr Annelin Molotsi

March 2020

DECLARATION

By submitting this thesis 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.

Date: March 2020

SUMMARY

The arid environment found in South Africa, coupled with the changing climate could present new obstacles that need to be accounted for in farming activities and practises. Smallholder farmers experience a number of limiting factors which emphasised the need to develop an easy and cost-effective approach that should be implemented for the improvement of breeding strategies and selection decisions that will improve the production outputs as well as the overall genetic gain for smallholder farmers.

Extensive molecular genetic studies have been done on a number of South African sheep breeds, including genetic diversity and population structure studies. There are, however, very few studies that identify CNVs in sheep breeds and then determine possible correlation between identified CNVs and desirous traits.

The aim of this study was to conduct a pedigree analysis to facilitate better breeding management decisions for smallholder farmers, and to identify preliminary CNVs that could assist with the selection of superior animals that may be more adaptable or have superior production outputs, such as superior fertility, desirous body conformation, higher growth rates or even superior feed conversion ratio (FCR). These production outputs depend on the goals set by each individual farmer with regards to his flock.

A pedigree was constructed of a smallholder flock from Beaufort-West that consisted of 48 Dorper individuals. Quality control and pruning was carried out on the individuals in *Plink.* One individual was removed due to missing genotype, and a second individual was removed by the *SEQUOIA* package in *RStudio*. Eleven dams were assigned, and four sires were assigned as parents to individuals. Ten dams were assigned as possible relatives with likely parent-offspring (PO) relationships, and nine sires were assigned as possible relatives with likely parent-offspring (PO) relationships.

Diversity statistics which were calculated are the inbreeding coefficient, LD and the runs of homozygosity for the sample populations. These sample populations include two smallholder Dorper populations and a Dorper and Namaqua Afrikaner (Namafr) population. The one smallholder population was from a farm outside of Beaufort-West (Dorpersm 2) and the other smallholder flock was from the Ebenheaser community (Dorpersm 1). The Dorper and Namafr populations were from the Nortier research farm. The diversity statistics were calculated in *Plink* and a PCA was constructed in *Rstudio* using the *pcadapt* package. The Namafr population had the highest LD according to the r^2 statistic (0,4805 $\pm$ 0,2476) followed by Dorper population (0,4156 $\pm$ 0,2000), then the two smallholder Dorper populations, Dorpersm 1 (0,3673 $\pm$ 0,1714) and Dorpersm 2 (0,3898 $\pm$ 0,1774). The smallholder Dorper population from the Ebenheaser community had the highest inbreeding coefficient (-0,0029 $\pm$ 0,0409), followed by the smallholder Dorper population from the Beaufort-West community (-0,0265 $\pm$ 0,0768) then the Dorper population (0,0535 $\pm$ 0,0000) and lastly the Namafr with the lowest inbreeding coefficient (-0,0652 $\pm$ 0,0399). Two runs of homozygosity (ROH) were discovered in the Dorper population, 105 ROH were discovered in the Dorpersm 1 population, 22 ROH were discovered in the Dorpersm 2 population and no ROH were discovered in the Namafr population. The

constructed PCA identified two distinct clusters, one Namafr cluster and one Dorper cluster. The Dorper cluster showed sub-clusters between the Dorper population from the Nortier farm and the smallholder Dorpers.

Preliminary CNVs were identified in a smallholder Dorper population from the Beaufort-West community. CNVs were determined using the *PennCNV* program. Gene annotation and classification studies were carried out in DAVID on candidate genes that overlapped the identified genes. A total of 206 CNVs were identified in 36 individuals. Candidate genes that were found to overlap with identified CNVs included interferon genes, which are responsible for immune defence mechanisms as well as a number of other genes responsible for biological functions including transport, metabolic precursors, neurogenesis, signalling as well as bone and cartilage matric composition. CNVs are genetic changes/mutations on the DNA level that could result in phenotypic variation between individuals. Copy number detection could thus be an important aspect in understanding the underlying genetic control of phenotypes observed in sheep flocks. Using CNVs that could be advantageous for certain desired traits for selection purposes could be economically beneficial to farmers.

Genetic diversity statistics of a population as well as the possible mutations, such as copy number variations, are important genetic information that could be incorporated into breeding strategies to ensure optimal production and genetic gain.

OPSOMMING

Die droë omgewing in Suid-Afrika, tesame met die veranderende klimaat, kan nuwe hindernisse inhou wat in die boerdery-aktiwiteite en -praktyke in ag geneem moet word. Kleinboere ervaar 'n aantal beperkende faktore wat die noodsaaklikheid beklemtoon om ‘n maklike en koste-effektiewe benadering te ontwikkel wat geïmplementeer moet word vir die verbetering van teel strategieë en seleksie besluite wat die produksie uitsette en die algemene genetiese wins vir kleinboere sal verbeter.

Uitgebreide genetiese molekulêre studies is al op ‘n aantal Suid-Afrikaanse skaaprasse gedoen, insluitend genetiese diversiteit en populasie struktuur. Daar is egter baie min studies wat genetiese variante in skaaprasse identifiseer en moontlike korrelasie tussen geïdentifiseerde genetiese variante en begerige eienskappe bepaal.

Die doel van hierdie studie was om 'n stamboom analise te doen om beter teel bestuur besluite vir kleinboere te fasiliteer, en om voorlopige CNVs te identifiseer wat kan help met die seleksie van verhewe diere wat meer aanpasbaar is of beter produksie-uitsette het, soos superieure vrugbaarheid, gewenste liggaamsbou, hoër groeitempo of selfs 'n beter voeromsetverhouding (FCR). Hierdie produksie-uitsette hang af van die doelwitte wat elke individuele boer ten opsigte van sy kudde stel. Elf ooie en vier ramme is as ouers aan individue toegewys. Tien ooie is as moontlike familielede met waarskynlik ouer-nakomeling verhoudings toegewys, en nege ramme is as moontlike familielede met waarskynlike ouer-nakomeling verhoudings toegewys.

‘n Stamboom was saamgestel van ‘n kleinboer skaap trop uit Beaufort-Wes wat uit 48 Dorper-individue bestaan. Kwaliteitskontrole is op die individue in *Plink* uitgevoer. Een individu is verwyder weens die ontbrekende genotipe, en 'n tweede individu is deur die *SEQUOIA*-pakket in RStudio verwyder. Elf ooie en vier ramme is as ouers aan individue toegewys. Tien ooie is as moontlike familielede met waarskynlik ouer-nakomeling verhoudings toegewys, en nege ramme is as moontlike familielede met waarskynlike ouer-nakomeling verhoudings toegewys.

Die diversiteit statistieke wat bereken is sluit in die teel koëffisiënt, onewewigtigheid koppeling en die mate van homosigositeit van die kudde. Die populasies sluit in twee kleinboer Dorper kuddes, ‘n Dorper kudde en ‘n Namaqua Afrikaner (Namafr) kudde. Die een Dorper kleinboer kudde kom van 'n plaas buite Beaufort-Wes (Dorpersm 2) en die ander kleinvee-kudde kom van die Ebenheaser-gemeenskap (Dorpersm 1). Die Dorper- en Namafr-kudde kom van die Nortier-navorsing plaas. Die diversiteit statistieke is in *Plink* bereken en ‘n hoof komponentanalise was in *Rstudio* opgestel met behulp van die *pcadapt*-pakket.

Die Namafr kudde het die hoogste onewewigtigheid koppeling volgens die r^2-statistiek (0,4805 ± 0,2476) gevolg deur die Dorper kudde (0,4156 ± 0,2000), daarna die twee kleinboer Dorper kuddes, Dorpersm 1 (0,3673 ± 0 , 1714) en Dorpersm 2 (0,3898 ± 0,1774). Die Dorper kudde van die kleinboer uit die Ebenheaser-gemeenskap het die hoogste teel koëffisiënt (-0,0029 ± 0,0409), gevolg deur die Dorper kudde uit die Beaufort-Wes-gemeenskap (-0,0265 ± 0,0768) en dan die Dorper kudde (0,0535 ± 0,0000) en laastens die

Namafr met die laagste teel koëffisiënt (-0,0652 ± 0,0399). Twee lopies van homosigositeit (ROH) is in die Dorper kudde ontdek, 105 ROH is in die Dorpersm 1-kudde gevind, 22 ROH in die Dorpersm 2-kudde en geen ROH is in die Namafr-kudde gevind nie. Die gekonstrueerde hoof komponentanalise (PCA) het twee afsonderlike groepe geïdentifiseer, een Namafr-groepering en een Dorper-groepering. Die Dorper-groepering het onderlinge tussen die Dorper-kudde van die Nortier-navorsing plaas en die kleinboer Dorpers getoon.

Voorlopig aantal kopie variante (CNV) was geïdentifiseer in 'n kleinboer Dorper kudde uit die Beaufort-Wes-gemeenskap. CNVs is met behulp van die PennCNV-program bepaal. Gene aantekening- en klassifikasie studies is in DAVID uitgevoer op kandidaat gene wat die geïdentifiseerde gene oorvleuel. Altesaam was 206 CNVs is in 36 individue geïdentifiseer. Kandidaat gene wat gevind was wat oorvleuel het met geïdentifiseerde CNVs, sluit interferon-gene in, wat verantwoordelik is vir immuun verdedigingsmeganismes, asook 'n aantal ander gene wat verantwoordelik is vir biologiese funksies, insluitend vervoer, metaboliese voorlopers, nekrogeniese, seine, as ook been- en kraakbeen matriek samestelling. CNVs is genetiese veranderinge / mutasies op die DNA-vlak wat kan lei tot fenotipiese variasie tussen individue. Die identifisering van kopie nommers kan dus 'n belangrike aspek wees in die begrip van die onderliggende genetiese beheer van fenotipes wat by skaap kuddes waargeneem kan word. Die gebruik van CNVs wat voordelig kan wees vir sekere eienskappe vir keurings doeleindes, kan ekonomies voordelig wees vir boere.

Statistieke oor genetiese diversiteit van 'n kudde, sowel as die moontlike mutasies, soos kopiegetalvariasies, is belangrike genetiese inligting wat in teel strategieë geïnkorporeer kan word om optimale produksie en genetiese wins te verseker.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude and appreciation to the following persons and institutions:

Firstly, to my supervisor, Dr Annelin Molotsi, thank you for your guidance with every aspect of this project. Thank you for all your encouragement and all the assistance with the data analysis and troubleshooting.

Thank you to the National Research Foundation for the financial support and for making this project possible.

Thank you to the smallholder farmer in Beaufort-West for the genetic material from his flock that made this study a possibility.

I would like to thank my family; my parents, siblings and grandparents. Thank you for all your patience through this period and process in my life. All your encouragement and prayers have helped me through this.

Lastly to everyone who kept me in your prayers, thank you for your support. I hope you know how much you have helped and meant to me.

PREFACE

This thesis is presented as a compilation of six chapters.

Chapter 1	**General Introduction**
Chapter 2	**Literature review**
Chapter 3	**Research chapter 1** Pedigree analysis of a smallholder sheep flock in Beaufort-West using an Ovine50KSNP BeadChip
Chapter 4	**Research chapter 2** Diversity statistics of South African smallholder sheep and sheep from a South African research flock using an Ovine50KSNP BeadChip
Chapter 5	**Research chapter 3** Identification of CNVs found in South African smallholder sheep using an Ovine50KSNP BeadChip
Chapter 6	**General conclusion and recommendations**

TABLE OF CONTENT

LIST OF ABBREVIATIONS

AnGR	Animal Genetic Resources
QTL	Quantitative trait loci
GEBV	Genomic estimated breeding value
PA	Parent average
EBV	Estimated breeding value
BLUP	Best Linear Unbiased Prediction
TBV	True breeding value
BV	Breeding values
PE	Prediction equations
GS	Genomic selection
PWWT	Post-weaning weight
MAS	Marker assisted selection
MQTL	Market quantitative trait locus
IBD	Identical by decent
AR	Average relatedness
Ne	Effective population size
N_{ef}	Inbreeding effective size
N_{ev}	Variance effective size
MDS	Multidimensional scaling
SNP	Single nucleotide polymorphism
GWAS	Genome-wide association studies
CNV	Copy number variation
CNP	Copy number polymorphism
SV	Structural variant
CNVR	Copy number variation region
CGH	Comparative genomic hybridization
aCGH	Array comparative genomic hybridization
LOH	Loss of heterozygosity
NMI	Non-Mendelian inheritance
HWE	Hardy-Weinberg Equilibrium
NGS	Next-generation sequencing
mRNA	messenger RNA
ERV	Endogenous retroviruses

OPA	Ovine pulmonary adenocarcinoma
JSRV	Jaagsiekte sheep retrovirus
exJSRV	Exogenous jaagsiekte sheep retrovirus
enJSRV	Endogenous jaagsiekte sheep retrovirus
NAHR	Nonallelic homologous recombination
LD	Linkage disequilibrium
GO	Gene ontology
Ms	Microsatellite
RFLP	Restriction fragment length polymorphisms
ISTS	Immotile, short-tail sperm defect
LLR	Log-likelihood Ratio
PO	Parent-offspring
FS	Full-sibling
HS	Half-sibling
MS	Maternal siblings (full or half)
PS	Paternal siblings (full or half)
GG	Grandparent-grand-offspring
FA	Full aunt/uncle – niece/nephew
HA	Half aunt/uncle – niece/nephew
GGG	Great-grandparent-great-grand-offspring
CC	Full cousins
U	Unrelated
OH	Opposite homozygotes
IBD	Identical-by-descent
IBS	Identical-by-state
ISAG	International Society for Animal Genetics
ME	Mendelian errors

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

1.1 BACKGROUND

The South African sheep industry consists of commercial, emerging commercial and smallholder farmers (Molotsi *et al.,* 2017). Smallholder farmers can often be classified as subsistence farmers (Chamberlin, 2007), and often have limited resources available to them (DAFF, 2012). According to Motiang & Webb (2016), a large number of people (82%) rely on cattle for income while 55% rely on other livestock including sheep. Limitations, such as lack of capital, infrastructure as well as a limited knowledge of agricultural production practises that subsistence farmers often face could negatively influence their income due to possible lowered production outputs. Production outputs could be negatively influenced by a lack of correct breeding knowledge needed to optimise production such as which superior individuals to select. Disease or sickness could also negatively affect production if the farmer does not have available capital for medical treatment.

One method to ensure optimal production outputs is through the use of effective and sound breeding management programmes. Some considerations that could assist in constructing breeding programmes include; pedigree information, production records, health records as well as the available genetics and genetic structure of the flock in question.

According to UNESCO Institute for Statistics (2018), sub-Saharan Africa has the highest out-of-school rates. This indicates the educational gap that may be a major problem in many communities. Due to this and a number of other limitations that smallholder communities face, it is often assumed that smallholder farmers do not keep livestock records, or the records are sporadic. Basic livestock records that should be kept include animal production records and pedigree information. Production and breeding or pedigree records will provide some basic information that may assist the farmer to make informed breeding management decisions, namely which individuals to select for breeding (Cadmus, 1949; Mavrogenis & Papachristoforou, 1990). Production records include reproduction as well as production trait records. Reproduction traits that are important include fertility as well as whether the ewe might have had single or multiple offspring, mothering ability as well as milk production. Production records that could be advantageous for management decisions include growth records, feed conversion ratio (FCR), fleece weight and length as well as a number of other traits that may be important to farmer. These production and reproduction records allow the farmer to identify individuals that may have above average production and may be essential in increasing the flock production outputs through improved breeding strategies according to the desired production goals set by the farmer.

Knowledge of the available genetic resources and the genetic structure of the flock could be used for constructing breeding programmes suited to the goals of the farmer. The genetic resources and genetic structure of the flock includes genetic information regarding inbreeding, linkage disequilibrium (LD), runs of

homozygosity (ROH), minor allele frequency (MAF), heterozygosity as well as the presence of mutations or genetic variation within the population. Due to the improvement of high throughput techniques, new structural variants are being identified that have the potential to be used for genomic selection alongside traditional breeding systems to advance genetic improvement (Liu *et al.*, 2013). These structural mutations are the permanent incorporation of random errors in the DNA that results in differences between ancestral and descendant copies of DNA (Hamilton, 2011a). Changes to DNA can be a point mutation where a single nucleotide is affected such as single nucleotide polymorphisms (SNP) (Nowak *et al.,* 2009), or it could affect a number of nucleotides, namely copy number variation (CNV) (Liu *et al.*, 2010; Liu *et al.*, 2013; Yang *et al.*, 2018). Mutations often affect more than a single nucleotide (Henrichsen *et al.,* 2009; Wang *et al.,* 2012), which emphasises the need to study and understand the possible effects of these larger mutations known as CNVs. Knowledge of mutations and genetic variation present in the flock could be important for breeding programmes since they could be either beneficial or detrimental (Hamilton, 2011a).

Genomic selection is a type of marker-assisted selection that uses genetic markers and genetic information to select individuals (Goddard & Hayes, 2007). Genomic selection has become a widely used method for genetic improvement, and could be used alongside traditional breeding strategies used by smallholder farmers to improve the overall production within farming systems.

1.2 PROBLEM STATEMENT

A large area of South Africa is classified as arid (Engelbrecht & Engelbrecht, 2016). This arid environment coupled with the changing climate could present new obstacles that need to be accounted for in farming activities and practises (Thomas *et al.,* 2007). The changing environment could include an increase in temperature in certain areas and a decrease in temperatures in other areas, as well as changes in the expected rainfall for different regions, resulting in possible floods or drought in areas. This changing environment highlights the need for animals that are able to adapt to the expected climate change (Pilling & Hoffmann, 2011), while maintaining or improving their production outputs.

Due to the ever-changing climate, more adaptable and robust animals will be needed that will be able to survive and produce in the extreme climates that are currently being experienced and could also lie ahead. Individuals that are more adaptable and robust are able to survive and produce in less than favourable environments. Some examples of adaptation traits are that they are able to withstand higher temperatures before physiological problems occur, they are also often able to survive longer with no or significantly less water than needed by other individuals. This adaptation will only be possible if there is genetic diversity within the flock. To have genetic diversity within the flock there needs to be individuals that are both homozygous and heterozygous for adaptation traits, meaning that a full understanding of possible variants in the genome is necessary in order to select for adaptable individuals.

A number of limitations, namely; the inability to enter the formal market, limited resources, lack of capital, scarcity of feed as well as lack of infrastructure may all have a negative influence on the current production

systems being used by smallholder sheep farmers (Ellis, 1991; Tibbo, 2006; DAFF, 2012). Therefore, an easy and cost-effective approach needs to be implemented that will assist with the improvement of breeding strategies and selection decisions of the dams and sires of the next generation that will result in improvement of the production outputs as well as an overall genetic gain for smallholder farmers.

Selection is one of the most important tools available to farmers to change the population gene frequency. Selection methods, traditional or genomic, are not commonly used in smallholder farming systems. Therefore, there is very little, or no genetic progress being made for important economic traits due to the lack of structured breeding programmes with specific goals and outcomes in the smallholder farming set-up.

Extensive molecular genetic studies have been done on a number of South African sheep breeds. These studies include genetic diversity and population structure of the sheep breeds, as well as pedigree construction of smallholder sheep (The International Sheep Genomics Consortium *et al.,* 2010; Li *et al.,* 2009; Oravcová & Krupa, 2011; Soma *et al.,* 2012). However, there are currently very few studies that identify CNVs in sheep breeds or that focus on possible correlations between identified CNVs with economically relevant traits or even disease resistance. This lack of knowledge means that there is a large part of the sheep's biological functions and genome that is still unknown to animal breeders and could possibly have beneficial outcomes. Identifying CNVs that might be correlated to economically relevant traits or characteristics and then selecting individuals based on these genetic variations could result in increased production outputs if incorporated into breeding programmes correctly.

Animals that are adaptable and robust would be beneficial on smallholder farms, as the adaptable individuals may be able to produce better than the average individual in the low-input systems often employed by smallholder farmers and communities.

1.3 RATIONALE OF STUDY (JUSTIFICATION)

The construction of the pedigree for the smallholder flock will provide the farmer with valuable parental and relationship in the population that could assist the farmer with breeding and management decisions. Genetic population information could inform the farmer as to the degree of inbreeding in the population, ensuring that the farmer can make an informed decision with regards to which individuals to breed to each other as well as which animals to possibly replace. Knowledge of the genetic resources available in the flock will assist the farmer to construct effective breeding programmes that are in line with the goals of the farmer. Linkage disequilibrium has often been used for selection purposes for specific traits desired by the farmer. A full account of the available genetic resources in the flock can be determined through estimations of genetic diversity statistics.

Identifying mutations, namely CNVs, that could provide production, reproduction or even adaptation advantages will have a beneficial economic outcome for the farmer as well as result in a possible increase of the overall production outputs. Production outputs could be increased through identifying CNVs in sheep

that increase reproduction (Davis *et al.*, 1992) and disease resistance (Klymiuk *et al.*, 2003; Viginier *et al.*, 2012). CNVs have been identified in cattle that increase production outputs through milk production (Aguilar *et al.*, 2017), growth traits (Zhou *et al.*, 2016) and even parasite resistance (Pickering, 2017). Further studies are needed that identify CNVs that could affect additional production traits and thus increase production outputs.

1.4 STUDY AIMS

The aim of this study was to characterise the genetic diversity and possible preliminary CNVs present in smallholder sheep that could facilitate breeding management programmes using the relevant population genetic information.

1.5 RESEARCH OBJECTIVES

1) To construct a pedigree and estimate IBD relationships of a smallholder sheep flock from Beaufort-West with an OvineSNP50 Beadchip using *Plink* and *pcadapt* package in RStudio.

2) Estimate diversity statistics of two smallholder Dorper populations, a Dorper and Namaqua Afrikaner research population in order to compare the inbreeding, LD, heterozygosity and the runs of homozygosity between the populations.

3) To identify copy number variation and perform gene functional classification in a smallholder flock from Beaufort-West with an OvineSNP50 BeadChip and PennCNV software.

1.6 RESEARCH PAPER OUTLINE

- This study consists of six chapters, chapter one gives a brief background, the problem statement, as well as the aims and objective of the thesis.
- Chapter two consists of a literature review that provides background information and the reasoning behind some of the ideas mentioned in this thesis.
- Chapter three is the first research chapter and focuses on the construction of a pedigree of smallholder sheep as well as the calculation of pairwise IBD relationships in order to develop a more complete knowledge of the pedigree of the studied flock.
- Chapter four is the second research chapter and estimates diversity statistics of four sheep populations in order to compare the diversity within and between the different populations.
- Chapter five is the third and last research chapter and aims to identify CNVs in smallholder sheep.
- Chapter six is the last chapter and contains a general discussion as well as the conclusion of the study and recommendations.

1.7 REFERENCES

Aguilar, M., Ponce, S. I., López, F. J., Padilla, E., Peláez, C.G., Bagnato, A. & Strillacci, M., 2017. Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers. *J. Anim. Breed. Genet.*, *134*, 49–59. https://doi.org/10.1111/jbg.12238

Cadmus, W., 1949. Sheep Records for Greater Profits. Agricultural Experiment Station, Circular. 182.

Chamberlin, J., 2007. Defining Smallholder Agriculture in Ghana: Who Are Smallholders, What Do They Do and How Are They Linked With Markets?

DAFF., 2012. A framework for the development of smallholder farmers through cooperative development. https://doi.org/10.1016/j.jcsr.2006.03.008

Davis, G.H., Dodds, K.G. & Bruce, G.D. 1992. Combined effect of the Inverdale and Booroola prolifacacy genes on ovulation rate in sheep. *Proc. Assoc. Advmt. Anim. Breed. Genet.*, *13*, 74–77.

Ellis, F., 1991. Peasant Economics: Farm Households and Agrarian Development. Am. Anthropol. 93. https://doi.org/10.2307/635095

Engelbrecht, C. & Engelbrecht, F., 2016. Shifts in Köppen-Geiger climate zones over southern Africa in relation to key global temperature goals. Theor Appl Climatol. 123(1–2), 247–261. https://doi.org/10.1007/s00704-014-1354-1

Goddard, M. & Hayes, B., 2007. Genomic selection. *J. Anim. Breed. Genet.*, *124*, 323–330.

Hamilton, M., 2011. Population genetics. Chichester: John Wiley & Sons.

Henrichsen, C., Chaignat, E. & Reymond, A., 2009. Copy number variants, diseases and gene expression. Hum. Mol. Genet. 18(R1), 1–8. https://doi.org/10.1093/hmg/ddp011

International Sheep Genomics Consortium., Archibald, A., Cockett, N., Dalrymple, B., Faraut, T., Kijas, J. & Xun, X., 2010. The sheep genome reference sequence: A work in progress. Anim. Genet. 41(5), 449–453. https://doi.org/10.1111/j.1365-2052.2010.02100.x

Kim, E., Elbeltagy, A., Aboul-Naga, A., Bischkowsky, B., Sayre, B., Mwacharo, J. & Rothschild, M., 2016. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity. 116(3), 255–264. http://dx.doi.org/10.1038/hdy.2015.94

Klymiuk, N., Muller, M., Brem, G. & Aigner, B., 2003. Characterization of Endogenous Retroviruses in Sheep. *J. Virol.*, *77*(20), 11268–11273. https://doi.org/10.1128/jvi.77.20.11268-11273.2003

Li, M., Strandén, I. & Kantanen, J., 2009. Genetic diversity and pedigree analysis of the Finnsheep breed. J. Anim. Sci. 87, 1598–1605. https://doi.org/10.2527/jas.2008-0848

Liu, G., Hou, Y., Zhu, B., Cardone, M., Jiang, L., Cellamare, A. & Keele, J., 2010. Analysis of copy number variations among diverse cattle breeds. Genome Res. 20, 693–703. https://doi.org/10.1101/gr.105403.110

Liu, J., Zhang, L., Xu, L., Ren, H., Lu, J., Zhang, X. & Du, L.., 2013. Analysis of copy number variations in the sheep genome using 50K SNP BeadChip array. BMC Genom. 14(229), 1–11.

Mavrogenis, A. & Papachristoforou, C., 1990. Use of Part Lactation Records for Selection in Chios Sheep and Damascus Goats. In *Agricultural Research Institute Ministry of Agriculture and Natural Resources*. https://doi.org/Doi 10.1016/S0038-0717(99)00179-0

Molotsi, A., Dube, B., Oosting, S., Marandure, T., Mapiye, C., Cloete, S.W.P. & Dzama, K., 2017. Genetic traits of relevance to sustainability of smallholder sheep farming systems in South Africa. *Sustainability,* 9(1225), 1–18. https://doi.org/10.3390/su9081225

Motiang, D. & Webb, E., 2016. The influence of household characteristics on cattle off-take rates in the North West Province of South Africa. Livestock Research for Rural Development. 28(6). Retrieved from https://repository.up.ac.za/bitstream/handle/2263/56472/Motiang_Influence_2016.pdf?sequence=3&isAllowed=y

Nowak, D., Hofmann, W. & Koeffler, H., 2009. Genome-wide mapping of copy number variations using SNP arrays. Transfus Med Hemoth. 36, 246–251.

Oravcová, M. & Krupa, E., 2011. Pedigree Analysis of the Former Valachian Sheep. Slovak J. Anim. Sci. 44, 6–12.

Pickering, L., 2017. *The Relationship between Copy Number Variations and Tick Resistance in South African Nguni cattle*. Stellenbosch Univeristy.

Pilling, D. & Hoffmann, I., 2011. Climate change and animal genetic resources for food and agriculture: State of knowledge, risks and opportunities. In Background Study Paper. https://doi.org/10.1016/S0140-6736(53)90577-7

Redon, R., Ishikawa, S., Fitch, K., Feuk, L., Perry, G., Andrews, T. & Hurles, M., 2006. Global variation in copy number in the human genome. Nature. 444, 444–454.

Soma, P., Kotze, A., Grobler, J. & van Wyk, J., 2012. South African sheep breeds: Population genetic structure and conservation implications. Small Rumin. Res. 103, 112–119. https://doi.org/10.1016/j.smallrumres.2011.09.041

Thomas, D., Twyman, C., Osbahr, H. & Hewitson, B., 2007. Adaptation to climate change and variability: farmer responses to intra-seasonal precipitation trends in South Africa. Clim. Change. 83, 301–322. https://doi.org/10.1007/s10584-006-9205-4

Tibbo, M., 2006. Productivity and Health of Indigenous sheep breeds and crossbreds in the central Ethiopian highlands (Swedish University). Retrieved from http://pub.epsilon.slu.se/id/eprint/1142

UNESCO Institute for Statistics., 2018. One in Five Children, Adolescents and Youth is Out of School. Retrieved from http://uis.unesco.org/sites/default/files/documents/fs48-one-five-children-adolescents-youth-out-school-2018-en.pdf

Viginier, B., Dolmazon, C., Lantier, I., Lantier, F., Archer, F., Leroux, C. & Terzian, C., 2012. Copy number variation and differential expression of a protective endogenous retrovirus in sheep. *PLoS One*, *7*(7), e41965. https://doi.org/10.1371/journal.pone.0041965

Wang, J., Jiang, J., Fu, W., Jiang, L., Ding, X., Liu, J. & Zhang, Q., 2012. A genome-wide detection of copy number variations using SNP genotyping arrays in swine. BMC Genom. 13(273), 1–10.

Yang, L., Xu, L., Zhou, Y., Liu, M., Wang, L., Kijas, J. & Liu, G., 2018. Diversity of copy number variation in a worldwide population of sheep. Genomics. 110, 143–148. https://doi.org/10.1016/j.ygeno.2017.09.005

Zhou, Y., Utsunomiya, Y., Xu, L., Hamidi, E., Bickhart, D., Alexandre, P. & Liu, G., 2016. Genome-wide CNV analysis reveals variants associated with growth traits in Bos indicus. BMC Genom. 17(419), 1–9. http://dx.doi.org/10.1186/s12864-016-2461-4

CHAPTER 2
Literature Review

2.1 INTRODUCTION

A large portion of the world's poor come from agriculturally based rural households, and therefore, depend primarily on agriculture as a source of income (Markelova *et al.*, 2009). Anderson (2003) estimated that around 1.96 billion people rely on livestock to supply some or all of their daily needs through meat products or for income. *Figure 2.*1 indicates the sheep consumption in countries all over the world.

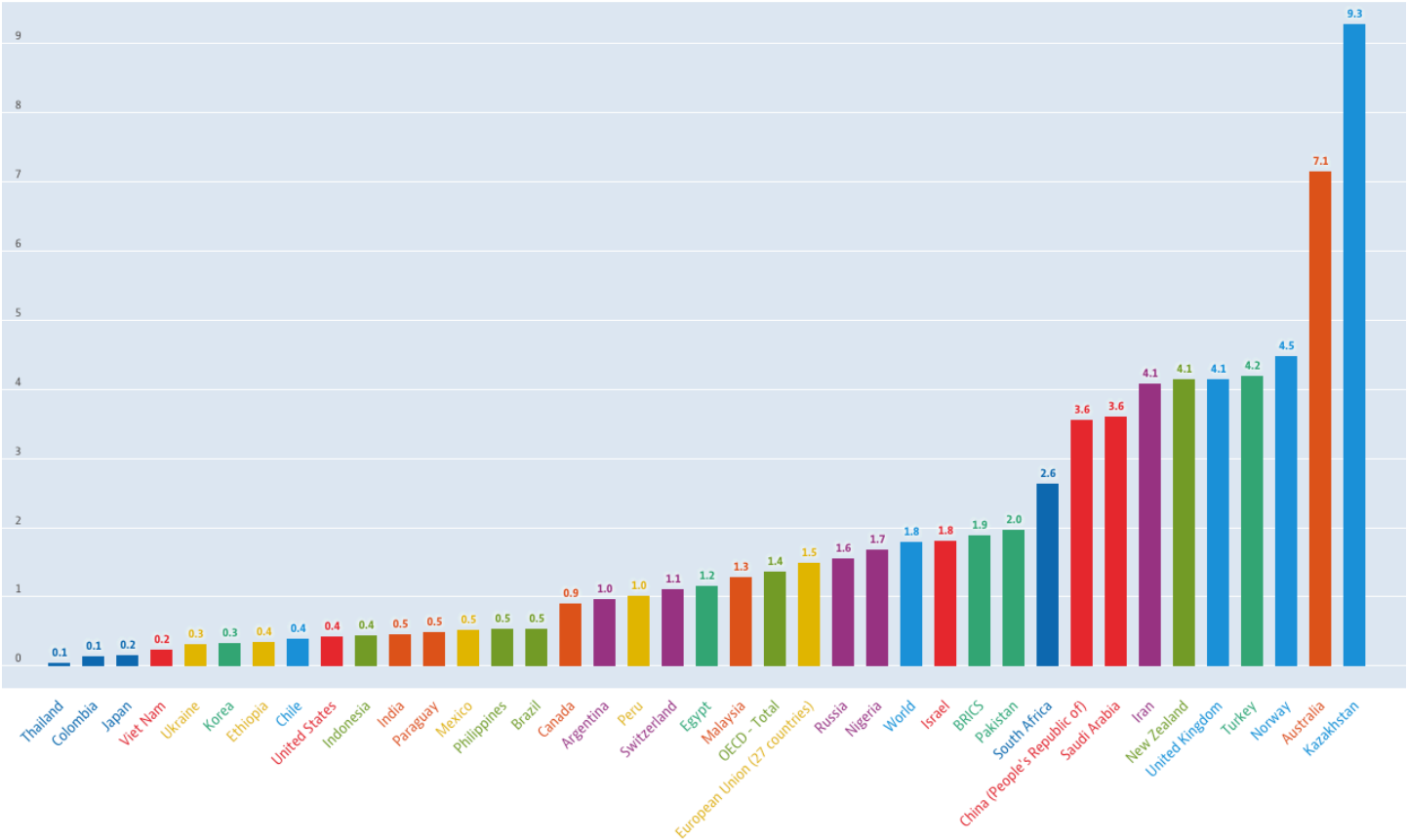

Figure 2.1.1 *Sheep meat consumption for countries all over the world, kg/capita (OECD, 2020)*

Due to the fact that a large number of people depend on livestock for income or as a source of food, as seen in *figure 2.1.1*, attention needs to be given to increase livestock production in order to provide income and a food source to rural households. Livestock production is expected to increase not only with regards to products but also by region (OECD & FAO, 2019). Global sheep meat production is predicted to grow by 14% while the sheep herds in Africa are expected to grow by approximately 2% p.a. however, the animal numbers are believed to increase faster in emerging and developing regions (OECD & FAO, 2019). *Figure 2.1.*2 indicates the sources of meat production growth by region.

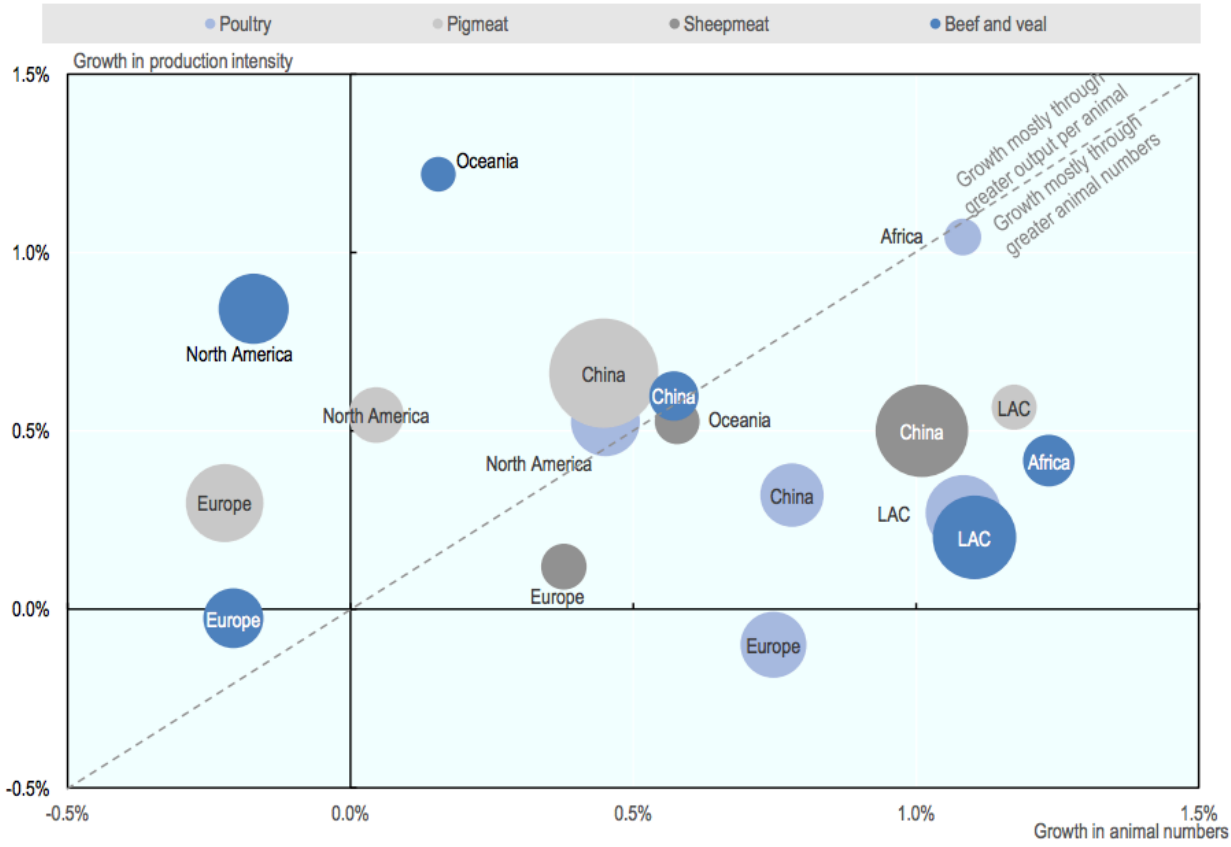

Figure 2.1.2 *Sources of meat production growth, by region* (OECD & FAO, 2019)

Sheep production includes the production of meat (lambs), milk as well as wool. Some factors that could have an influence on production is nutrition, genetics, production and reproduction records, environment and infrastructure (Hynd, 1989; Cloete *et al.*, 2002; Sadiq *et al.*, 2003; Warner *et al.*, 2010). One of the ways that production systems can increase production outputs and production efficiency is through genetics (Rosati *et al.*, 2002). Maintaining genetic diversity within the flock and the total population is important In order to increase production and select individuals that may be more adaptable to changes in production systems, climatic changes or any future selection purposes that may present themselves (Franklin, 1980).

A large area of South Africa is classified as arid (Engelbrecht & Engelbrecht, 2016), resulting in extensive ruminant livestock farming being an economically viable farming method. The South African sheep industry consists of commercial, emerging commercial and smallholder farmers (Molotsi *et al.,* 2017). *Figure 2.1.3* illustrates the 2018 preliminary distribution of sheep production in South Africa across the nine provinces with the Eastern Cape having the highest number of sheep. On average flock sizes range between 50 to 1800 animals, and are kept mainly for mutton and wool production (DAFF, 2018).

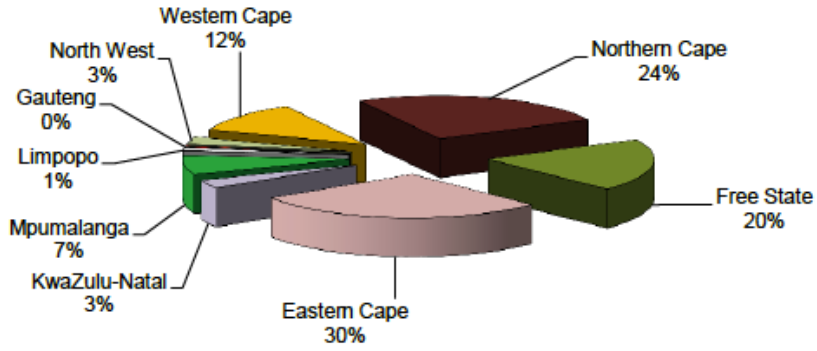

Figure 2.1.3 *Preliminary distribution of sheep by province in South Africa* (DAFF, 2018)

Some of the more prominent sheep breeds present in South Africa that are registered by the South African Studbook Association include; Merino and Dohne Merino (wool), Dorper, black and white-headed (meat), South African Mutton Merino (dual purpose), Dormer, Ile de France, Merino Landsheep, white-wooled Afrino (dual purpose), Van Rooy, Meatmaster, Damara and Suffolk (Cloete & Olivier, 2010).

2.2 SOUTH AFRICA'S SMALLHOLDER SHEEP INDUSTRY

2.2.1 CHALLENGES IN SMALLHOLDER COMMUNITIES

Smallholder farming communities face multiple issues, including scarcity of feed, climate change, diseases and a lack of infrastructure, limited resources as well as the knowledge gap (Tibbo, 2006). A large area of South Africa is classified as arid (Engelbrecht & Engelbrecht, 2016). Climate change has been recognised as having a significant effect on farming. Due to this there are definite concerns regarding the impact climate change will have on rural communities and smallholder farmers with regards to farming activities and livestock that depend on the ecosystem (Thomas & Twyman, 2007). Climate change can involve a wide range of challenges that will present themselves including; temperature fluctuations, changes in expected rainfall, reduced or disrupted feed supplies as well as disease outbreaks due to changes in disease epidemiology, which could make present methods of combating certain diseases obsolete (Pilling & Hoffmann, 2011). A change in temperature could lead to possible heat stress in livestock, which in turn will cause the animals to eat less resulting in possible weight loss and eventual loss in body condition if the high temperatures continue (Marai *et al.*, 2007). Changes in rainfall may negatively affect pasture quality thus disrupting feed supplies resulting in increased production costs due to the need for additional feed to be bought in (Rust & Rust, 2013). Due to these expected changes in the climate, genetic variability is needed in animals to allow future adaptation to possible harsh environments that may only present themselves at a later time.

The second challenge that smallholders face is that smallholder farmers are often considered to be only partially involved in the formal markets since they produce mainly for the informal market and for subsistence (Bojanic & Ellis, 2006). However, they are limited in their ability to fully enter the formal market on a competitive level due to both physical and non-physical reasons (DAFF, 2012). Non-physical reasons include the consumers demand and preferences which smallholders may not be familiar with and thus may

not be producing like certain cuts of meat. Physical limitations that could affect the smallholder from participating in the formal market could include the distance to the formal market. Smallholder farmers that live far from the formal markets and that do not have a way of getting their products there will not be able to sell their products (Chamberlin, 2007). The ability of smallholders to increase their profit depends solely on their capability to participate in the market competitively (Markelova *et al.*, 2009).

According to DAFF (2012) smallholder farmers are farmers that have limited resources on a small-based plot and mainly farm for subsistence and may have one or two cash crops or livestock that they introduce into the formal and informal market for an added income. Chamberlin (2007) suggests that land size, wealth and access to markets are all indicators of smallholder farmers, and that a subsistence farmer can also be classified as a smallholder farmer. A smallholder farm is often run by the family themselves instead of hiring external labour (DAFF, 2012). Small-scale and backyard livestock production allows the poor to earn an income from animals that can graze on common property pastures or are fed household waste (Delgado *et al.*, 1999). In smallholder communities livestock provides cash income either from the sale of the animal or through secondary products such as milk, wool or meat (Kosgey & Okeyo, 2007). Livestock can also be classified as assets or as a means for insurance, so the animal can be sold and the cash used for possible emergencies (Anderson, 2003). Smallholder farmers often have limited opportunities to increase their income, often due to a lack of capital.

The educational gap that has been identified in sub-Saharan Africa also adds to the challenges smallholder farmers face with regards to the knowledge resource gap (UNESCO Institute for Statistics, 2018). The absence of knowledge in smallholder communities regarding the implementation of correct production systems and effective breeding programmes could have a negative impact on animal production outputs including meat, milk or even wool products (Kosgey & Okeyo, 2007). Limited knowledge regarding new technologies available also limit smallholder farmers with regards to utilising new methods or these new technologies that may assist in increasing production. New technologies can include genomic selection which smallholder farmers may be unfamiliar with which could be a challenge. The high cost associated with many genomic technologies also means that many smallholder farmers will not be able to utilise the technology to improve their production due to the costs involved.

2.2.2 PROSPECTS IN SMALLHOLDER COMMUNITIES

In South Africa there are not many livestock production systems that would be able to produce in the arid climate that covers a large area of the country (Engelbrecht & Engelbrecht, 2016). Due to these large unfavourable environments sheep farming has become more popular because of the low space and maintenance requirements of sheep, as well as the low capital investment needed for buildings and other resources for their upkeep (Ademosun, 1994). These low input requirements for sheep make them a good alternative to large livestock like cattle for many smallholder farmers with limited resources and finances.

There are thus opportunities available for smallholder communities to get involved in sheep production due to the low initial capital needed and the low-input requirements.

There are a number of opportunities available to advance the South African sheep industry, one of them being through the improvement of the production systems of smallholder farms using various technologies, namely; reproductive technologies, genomics, information and communication technology (ict,) etc. (Chiwawa, 2019; Molotsi *et al.*, 2019). There is potential for genomic technology that could assist in incorporating advantageous genes into resource-poor production systems, improving livestock product traceability as well as identifying genes that may be associated with disease resistance, adaptability and a number of other relevant traits that may contribute to food security (van Marle-Koster *et al.*, 2015). However, a number of countries in the South African Development Communities (SADC) are in various agricultural and infrastructural developmental stages which limits the implementation of certain advanced technologies due to limited financial viability and infrastructural support (van Marle-Koster *et al.*, 2015). Therefore, government and other research institutes should prioritise allocating funding to support breeding programmes incorporating the use of some of these technologies.

Research institutes such as universities, the Agricultural Research Council (ARC) and other agricultural based organisations have the opportunity to fill the knowledge gap between the latest research and the smallholder farmers that don't have access to these sources of knowledge (Iñiiguez, 2011).

2.3 BREEDING PROGRAMMES AND GENETIC MANAGEMENT

2.3.1 ROLE IN THE INDUSTRY

Many developing countries make use of low-input production systems, and so it often happens that there are no breeding programmes or schemes within smallholder communities (Kosgey *et al.*, 2006; Mueller *et al.*, 2015). Effective breeding programmes facilitate genetic gain and thus result in increased production outputs and increased profits for the farmer (Richards *et al.*, 2010).

Breeding programmes should be aimed at improving economically relevant or beneficial traits within the population at a genetic level. Breeding strategies depend on how the traits of interest are inherited, the selection pressure applied in the system as well as the generation interval (Mavrogenis, 1995). Having clear-cut criteria for a breeding programme is difficult due to various production systems employed by numerous farmers as well as the different production goals in smallholder communities (Kosgey *et al.*, 2006). One of the limitations of creating a successful breeding system is the absence of reliable genetic and environmental parameter estimates (Cloete *et al.,* 2000; Kosgey & Okeyo, 2007). The lack of parameters can be due to an absence of reliable pedigree and performance information. Genetic resources can play an important role in future management strategies to preserve the remaining sheep breeds that are in existence (Qwabe *et al.,* 2012). Correct documentation of genetic resources is a key aspect with regards to animal or population conservation, especially the genetic diversity and uniqueness of the sheep populations as well as the development of breeding strategies for production purposes (Soma *et al.*, 2012). Record keeping and

pedigree information coupled with genetic analyses are important tools that can be used to assist with selection of superior animals within a flock.

Effective selection programmes depend on the population structure and require the evaluation of the genetic variability within the population, as well as the genetic structure and the gene flow of the population in order to enlarge the basis for selection (Gowane *et al.*, 2013; Duru, 2017).

2.3.2 CURRENT BREEDING PROGRAMMES

Four main livestock production systems have been identified by Waters-Bayer & Bayer (1992); 1) fulltime livestock keepers that depend on livestock as their primary living, 2) livestock keepers who also have some crops, but livestock remain their main source of livelihood, 3) crop farmers who keep some livestock and 4) the landless who keep some livestock as a side-line to their main source of income. A standard breeding programme will not suite every farmer due to the various production systems that are employed by farmers as stated by Waters-Bayer & Bayer (1992). Kosgey *et al.* (2006) found that one reason so many breeding strategies failed was due to the fact that the strategies were often designed by scientists and implemented by development agencies without considering the needs of the farmers or the long-term impacts of the breeding programme. For breeding programmes to be effective it needs to be compatible with the socio-cultural aspects of the farmers and have objectives and outputs that are in-line with the goals of the farmers as well as the preferences and needs of the market (Kosgey *et al.*, 2006). Due to the increasing need for meat in the growing population, smallholder farmers have the potential to contribute to meeting the national food needs and so alleviating some of the pressure put on the commercial markets by supplying some of their products into the formal market (Waters-Bayer & Bayer, 1992; Wiggins & Keats, 2013).

Breeding programmes need to be aimed at improving production while maintaining genetic diversity within populations to allow for selection and to ensure that individuals can adapt to changes in the environment.

2.3.3 BREEDING METHODS

2.3.3.1 Artificial vs Natural Selection

Selection is one of the most important tools available to animal breeders to change population gene frequencies. There are two types of selection; natural selection which is mediated by reproduction and survival of the fittest, and artificial selection which is mediated by reproduction. Artificial selection has continuously added to phenotypic variation between populations, while natural selection fuels the generation of biodiversity on earth (Guan *et al.*, 2016). Phenotypic changes over time are often as a result of adaptive evolution, which is mainly driven by selection, natural or artificial, mutation and genetic drift (Orr, 1998).

Natural selection occurs if the fitness for the three genotypes aren't equal (Nielsen, 2005). Negative selection is when a trait or mutation is selected against because it is disadvantageous to the individual, while

positive selection favours a trait or mutation because of the advantage it provides the individual (Nielsen, 2005). Due to artificial selection, sheep have become specialized in a wide range of purposes, including meat, milk and wool production (Kijas *et al.*, 2009). Breeds that have undergone extensive genetic selection for specific traits, often have superior genetic potential that can be utilised under efficient management systems, correct breeding systems and selection (Anderson, 2003). A decrease in genetic diversity can thus be contributed to extensive genetic selection, drift, inbreeding and the loss of founder alleles due to selection (Vozzi *et al.*, 2007).

Genetic improvement through artificial selection is influenced by a number of factors including the selection differential. The selection differential (S) is the difference between the base population mean and the mean of the selected parents. The selection differential is influenced by the proportion of individuals selected and the standard deviation (σ_p) of the phenotype (Falconer & Mackay, 2009). *Figure 2.3.1* is a visual representation of how selection intensity and phenotypic variation has an influence on the final number of individuals selected.

P = proportion individuals selected (shaded area)

(σ_p) = Standard deviation of phenotype (bell-shaped curve)

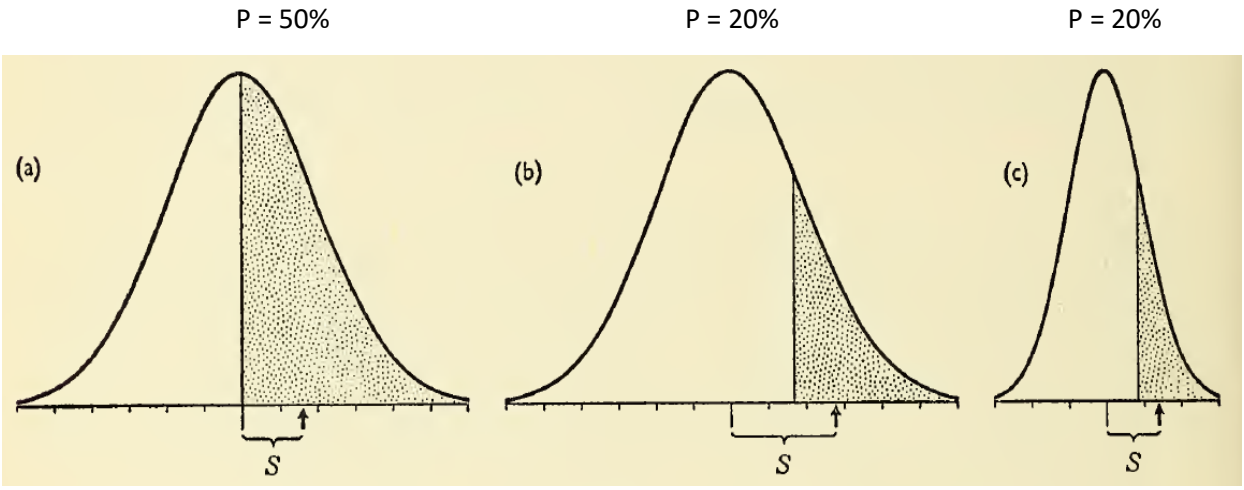

Figure 2.3.1 *Diagrams show how the selection differential, S, depends on the proportion of the population selected, and on the variability of the character (Falconer & Mackay, 2009) (a) < S, > P, > (σ_p). (b) > S, < P, > (σ_p). (c) < S, < P, < (σ_p)*

The rate at which a change in genotypes can be implemented depends on the initial frequency of the allele and the selection pressure that is applied to change the gene frequencies. Selection can be carried out using high or low selection pressure. The intensity of selection, *I*, depends only on the proportion of individuals selected. The selection intensity can be described by the following equation:

$I = \frac{S}{\sigma_p}$ (Mühlenbein, 1997).

Using a high selection pressure will result in the breeder only using a few superior individuals to breed, while with a low selection pressure many individuals are selected to breed. With a high selection intensity,

only a few individuals are replaced, so the flock stays constant and older animals are kept longer so the average age of the flock is high. High selection intensity could also result in a decreased genetic variability and could thus possibly lead to inbreeding. With a low selection intensity, more of the older animals are replaced resulting in a lower average age of the flock and a shorter generation interval. Intense selection leads to rapid genetic improvement but at the same time leads to a reduced effective population size (Goddard, 1992).

2.3.3.2 Traditional vs Genomic Selection

Selection can be based on an individual's phenotype using traditional selection or based on the genotype using genomic selection. The majority of smallholder farmers make use of phenotypic or traditional selection where they select individuals based on the physical characteristics of the animal. Traditional selection makes use of phenotype and pedigree information to determine breeding values using Best Linear Unbiased Prediction (BLUP) (Goddard & Hayes, 2007). Farmers will often select individuals that may have the best looking wool or individuals that have the highest growth or the best body conformation based on their own opinion or the opinion of an outside professional. Phenotypic selection is often an effective selection method with a variety of visible characteristics that can be selected for. Douch *et al.* (1996) suggested a number of parameters that selection could be based on in order to select animals that are nematode-resistant. This principle could be employed in other selection programmes in order to select individuals for traits desired by the farmer such as reproduction traits or even production traits. Hatcher & Atkins (2007) determined that selecting young ewes based on their body weight (BW) could have a positive effect on the reproduction of the ewes in their lifetime.

Phenotypically selecting for a trait of interest, such as wool quality, will indirectly affect the genotype of the individual. By visually determining which animals have superior wool quality and selecting them for further breeding will result in genetic improvement within the flock with regards to wool quality. Individuals with superior wool quality will differ genetically from individuals that have low wool quality (Karam, 1959).

Genomic selection uses genomic technology in order to select individuals based on the desired genotypes of the animals that the farmer is wanting to select for (Hayes *et al.*, 2009; Van Der Werf, 2009). Genomic selection makes use of traditional selection methods in addition to genotypic data namely; SNPs and QTL to determine breeding values using Genomic Best Linear Unbiased Prediction (GBLUP) (Fernando *et al.*, 2007; Goddard & Hayes, 2007). Genomic estimated breeding values (GEBV) are expected to be more accurate than pedigree-based estimated breeding values (EBV), because GEBV exploits LD and LE between markers and quantitative trait loci (QTL) along with pedigree relationships between individuals to calculate the breeding values (Habier *et al.,* 2007; Wolc *et al.*, 2011). Using pedigree information alongside genomic information such as determining the absence or presence of linkage between traits and markers will assist farmers in the selection of individuals with desirous traits that will be productive in the particular production system.

Genomic selection has positively influenced genetic gain in many flocks and has assisted in the improvement and genetic gain of several traits in a number of breeds and flocks (Van Der Werf, 2009; van der Werf *et al.*, 2014; Rupp *et al.*, 2016). Genomic selection results in a shortening of the generation interval because the markers can be genotyped at birth or even before (Goddard & Hayes, 2007). Genomic selection will also eliminate the need for progeny testing, thus reducing the costs by 92% and resulting in an increase in the genetic change by a factor of two as found by Schaeffer (2006). Genomic selection can identify traits that are of economic importance and may only appear at a later stage in the animals' life, since the traits may be sex-related or only measurable after the death of the individual e.g., intra-muscular fat content (Mohammadi *et al.*, 2013).

2.3.3.3 Gene Flow and Selection

Gene flow has been found to effect genetic gain (Gizaw *et al.*, 2014a). This effect could be positive or negative, depending on the breeding programme. Gene flow can negatively affect genetic gain if it occurs randomly with no specific goal, as it could work against selection within the flock by randomly adding new genes to the gene pool. Gene flow could also positively affect genetic gain if desired genes or traits are added to the gene pool of the flock in a structured programme to enable selection for these genes and thus to ensure genetic gain (Kosgey *et al.*, 2006; Gizaw *et al.*, 2014a). Correct gene flow and selection strategies could be used together to ensure genetic gain on smallholder farms as well as a way to control inbreeding within the flock.

Some smallholder farmers often unintentionally practise undesirable selection due to the fact that they often sell the faster growing young males, because they will fetch a higher market price (Gizaw *et al.*, 2014b). This could possibly result in negative selection because all the best individuals are being sold instead of being bred with to increase the flock genetic average. Mating practices are also often uncontrolled within the smallholder community and take place year-round; thus, all males mate with the females in a communal village mating system resulting in no selection occurring. This will result in an increase in diversity where mating is random between individuals and flocks and an absence of selection.

In a flock where random mating does not occur and the flock is 'closed', there is a chance of a decrease in genetic variability within the flock and possibly a decrease in heterozygosity, thus inbreeding (Selvaggi *et al.*, 2010). A possible way to increase genetic variability among and within populations could be to exchange rams between herds in a structured mating programme. Gene flow can be described as the successful incorporation of alleles into a population due to the movement (migration) of individuals into the population (Hamilton, 2011a). Gene flow can also be used in order to incorporate a possible trait of interest that may not be present in the flock. For smallholder farmers that are in close community with other smallholder farmers exchanging genetic material is relatively easy. Exchange of genetic material will only be effective if mating's between flocks are recorded and if there is a certain level of genetic variation between the flocks (Gizaw *et al.*, 2014a; b). Exchange of genetic material between flocks that operate under a communal village

mating system may not have any effect or It could have a negative effect, as the flocks may be genetically similar and could probably have a relatively high average relatedness (Gizaw *et al.*, 2014a). If done correctly, this exchange of genetic material can assist with genetic gain for each farmer and ensure a decrease or absence of inbreeding depression. This exchange of genetic material or gene flow is also possible for commercial and stud farmers with the improved reproduction technology that is now more easily available. Encouraging smallholder farmers and commercial farmers to exchange genetic material in a structured breeding system, could result in a larger base from which genetic material could be exchanged. This could ensure that South African sheep genetics is used in the production systems instead of genetics from a population not from South Africa that will not be adapted to the South African climate and so not reach its full genetic production potential. Negussie *et al.* (2002) states that selection within the indigenous breeds would be a better means for genetic improvement than importing exotic germplasm, as it makes use of environmental adaptation of the indigenous breeds and could result in a sustainable genetic improvement in productivity. Due to the occasional replacement of locally adapted breeds with 'higher' producing breeds, crossbreeding has become a widely used breeding system, but has not always been successful due to the incompatibility of these exotic breeds with the low-input production systems often employed by smallholder farmers (Kosgey *et al.*, 2006).

Attention should be given to ensure there is not a loss of genetic diversity due to the increased exchange of genetic material if such an exchange should be considered. Eteqadi *et al.* (2015) suggests a way to decrease inbreeding and so inversely increase genetic diversity by increasing the number of breeding males and implementing more frequent replacement of certain males and females. Increasing the number of breeding males could ensure a lowered relatedness among the individuals within the flock since less offspring would be related, and thus lowered inbreeding within the flock. More frequent replacement of certain individuals would only be effective if correct records are kept to ensure the farmer knows which individuals may be superior for certain desirous traits (Richards *et al.*, 2010). Correct records could then possibly indicate which individuals possess the desired traits that are important for optimal production and that are in line with the farmers breeding goals.

2.3.4 IMPORTANT TRAITS

In order to ensure improvement of production outputs, animal production traits should be in line with the market demand and should have economic market value. The best way to increase profits in the sheep industry is by increasing reproduction, growth rate and the carcass quality of the sheep (Safari *et al.*, 2005; Zishiri *et al.*, 2013a; b). The production traits most commonly measured in sheep include bodyweight, ultrasound scanning of muscle and fat on live animals in all breeds, and wool weight and quality, in Merinos (Duguma *et al.*, 2002a; Swan *et al.*, 2012; Zishiri *et al.*, 2013b). Production traits that can be measured in individuals before selection age generally have breeding values of moderate to high accuracy. However, several hard-to-measure traits where accuracies of breeding values are low include; adult wool production,

parasite resistance, reproduction, and carcass and eating-quality traits (Swan *et al.*, 2012). Growth traits can include birth weight, daily gain from birth to weaning, daily gain from weaning till the end of the artificial rearing, daily gain in the finishing period of lambs, daily gain in rearing of breeding animals and mature weight of ewes (Fogarty, 2009; Wolfová *et al.*, 2009).

Certain reproduction traits should be focused on such as; fertility, mothering ability and fecundity. Improved ewe productivity and longevity is a major objective in the sheep industry (Duguma *et al.*, 2002b). Ewe lifetime productivity being measured by the total number of lambs born, number weaned and total weight weaned; without these the whole production system will not be able to produce sheep with the desired production traits (Duguma *et al.*, 2002b; Zishiri *et al.*, 2013b). Ewe fertility was defined as the proportion of the actual number of times ewes lambed over the lambing chances that were afforded. If ewes or rams aren't fertile there will be no progeny and therefore no animals to sell. Ewes that don't have good mothering ability will produce lambs that might not survive or reach their full genetic potential (Martin *et al.*, 2004). Overall reproduction rate is defined as the number of lambs weaned per ewe mated (Cloete *et al.,* 2000). Wolfová *et al.* (2009) compared a number of traits and found that litter size had the highest economic importance, with high values also observed for lamb survival traits, productive lifetime and conception rate of ewes. Many of these reproduction traits are important to select for in order to improve productivity. However, most of the traits mentioned require accurate record keeping in order to include them in the breeding programme. Absence of records for the number of lambs born to a ewe or the number of lambs a ewe weaned means that there is no way to determine the actual productivity of that ewe and so the farmer will be unable to select for specific traits such as ewe productivity without the necessary records.

All the production traits mentioned above are important to optimise production, however, farmers can not only focus on some traits and ignore others. Fogarty (1995), Snyman *et al.* (1998) and Safari *et al.* (2007) discovered correlations between production traits, wool traits and reproduction traits. An overall understanding of traits is important as certain traits may be positively correlated, while other traits are negatively correlated (Safari *et al.*, 2005).

In an effort to improve production, species that might be considered more productive may eventually lead to the replacement of locally adapted breeds/species, resulting in a trade-off of adaption traits for production traits, leading to a loss of Animal Genetic Resources (AnGR) (Anderson, 2003). Due to the changing climate adaptation traits should be preserved where possible. Local or indigenous breeds that are more suited to the climate of South Africa should thus be used (Nardone *et al.*, 2010). Attention should be given to these indigenous breeds to genetically improve their production traits through breeding strategies and programmes without detrimentally affecting their adaptation traits.

Indigenous and locally developed breeds often have important genetic resources that include adaptation/fitness traits that allow for selection and adaptation of breeds during times of biological stress such as famine, drought or disease epidemics (Buduram, 2004; Peters *et al.*, 2010). Robustness is the ability of an individual to express high-production potential in a wide variety of environmental conditions, and has

become a specific breeding goal in the context of sustainable farm animal breeding (Mormède *et al.*, 2011). Knap (2005) defined robust pigs as pigs that combine high production potential with resistance to external stressors, allowing for unhindered expression of high production in a wide variety of environmental conditions. Robustness is defined by Friggens *et al.* (2017) as being the ability of an individual to face environmental restrictions, while carrying out the necessary functions that favour its future ability to reproduce. According to the robustness definitions stated by Friggens *et al.* (2017), the ability to reproduce includes growth traits to reach sexual maturity, longevity as well as the ability to avoid either natural death or culling in the population. This includes disease resistance as well as the ability to produce above a certain minimal level to avoid culling. Some of these adaptation traits that have developed over time include; tolerance/resistance to many diseases, tolerance to varying availability of feed and water, climatic tolerance, adaptation to low input management conditions and lastly the ability to survive and to regularly reproduce and produce over long periods of time (Hammond, 2000; Nsoso *et al.*, 2004). Many breeders have learned that vigour, fecundity and other traits of fitness decline at a rate proportional to the degree of random genetic drift, and this is inversely proportional to the population size (Soulé, 1980).

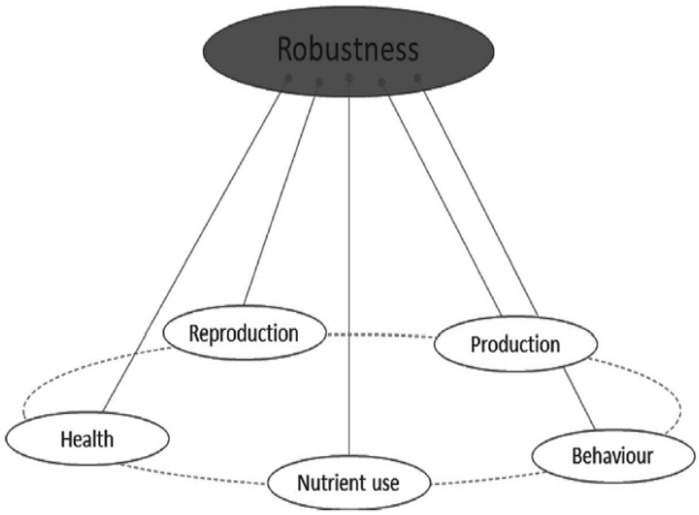

Figure 2.3.4.1 *A schematic representation of the various features operating at underlying levels, that combine to build robustness in the individual* (Friggens *et al.*, 2017)

As seen in *Figure 2.3.4.1,* there are many functional components/traits that contribute to overall robustness including; 1) disease resistance or resilience implies a good health status; 2) reproduction includes an individual's ability to reproduce well and regularly or at an interval desired by the production system, (semen quality, fertility, ease of birth, heat); 3) nutrient use includes resource acquisition and utilization, so basically traits that are related to feed intake, feed conversion, digestion and metabolism; 4) the animal is able to produce in accordance to the farmers expectations (weight or number of weaned offspring, fibre diameter in wool); 5) the behavioural component includes good social interactions, both with the farmer as

well as other animals, these traits allow for easy handling, milk-ability, parental care and aggressiveness (Friggens *et al.*, 2017).

Due to genotype x environment (G x E) interactions, an individual may have certain production outputs in one environment but when it is moved to another environment it may have higher or lower production due to the ability to adapt to the environment it was moved to (Burgueño *et al.*, 2012). If animals can only be evaluated for their 'own' environment, the two populations will diverge, however, if the animals can be evaluated for production in both environments, then the animals suited to either environment can be selected from one population and the overall effective population size will be smaller (Goddard, 1992). This will ensure that adaptable individuals are selected and thus will ensure that further generations will be adaptable to different environments.

2.4 GENOMIC TECHNOLOGIES

Genomic selection aims at increasing production and product quality and reducing the production costs by increasing selection accuracy through direct selection of genotypes instead of the observed phenotype (Mormède *et al.*, 2011). One of the aims of livestock genomics research is to identify genetic differences that are responsible for variation in phenotypic traits, especially those of economic significance; characterizing genetic variation in livestock species is an important step towards finally linking genes or genomic regions with phenotypes that could be economically beneficial (Stothard *et al.*, 2011).

2.4.1 MOLECULAR AND GENOMIC TECHNOLOGIES

The development of the ovine genome map including molecular markers and genes has assisted with the identification of genetic regions that influence and control traits of interest in sheep such as; fertility, reproduction, growth rate and efficiency, milk production, carcass quality and composition, wool characteristics and disease resistance (Bidwell *et al.,* 2009). The construction of the sheep reference genome was done using Sanger sequencing, however, the development of next-generation sequencing has provided additional sequence data from a number of sequencing platforms that can now be included in the construction of the reference genome (Sheep Genomics Consortium *et al.*, 2010). Whole-genome sequencing has identified candidate genes that may be associated with adaptation in extreme environments (Yang *et al.*, 2016). This could be valuable information for future breeding programmes with the changing climate being experienced if animals can be selected that may possess these candidate genes that could affect adaptation traits (Rust & Rust, 2013).

Microsatellites are sequences made up of a single sequence motif, no more than six bases long, that is tandemly repeated. Microsatellites have become a popular molecular marker with high polymorphism (Zane *et al.*, 2002). Microsatellites have been used in evolutionary studies to determine the evolutionary history of different species, as well as to determine genetic diversity between different sheep breeds (Ellegren *et al.*, 1997; Diez-Tascón *et al.*, 2000a; Farid *et al.*, 2000; Stahlberger-Saitbekova *et al.*, 2001). One of the main

disadvantages of microsatellites is that they need to be separated *de novo* from species that are examined for the first time (Zane *et al.*, 2002). However, microsatellites that acquired numerous point mutations would eventually degrade to a non-repetitive sequence, but an intermediate state is also possible where a sequence is made up of a few intermixed motifs, which show little sign of a tandem arrangement (Hancock, 1999). Kashi & Soller (1999) stated that changes in microsatellite repeat number can cause quantitative variation in protein function and gene activity, and can even affect organismal physiology and development.

Microsatellites have commonly been used as genetic markers in studies of parentage and kinship, however, single nucleotide polymorphisms (SNPs) have attracted more attention within genomic technologies and practises (Arranz *et al.*, 1998; Rosa *et al.*, 2013; Al-Atiya, 2015; Weinman *et al.*, 2015).

Single nucleotide polymorphisms (SNPs) are genetic variants on the DNA caused by a spontaneous germline mutation of single nucleotides (Nowak *et al.,* 2009). A SNP occurrence is illustrated in *Figure 2.4.1.1.* Generally, SNPs occur in two alleles, and are distinguished into synonymous or non-synonymous SNPs depending on whether or not they change the amino acid sequence if they lie in a coding region of a gene. However, SNPs lying in non-coding regions of the genome may have an impact on splicing processes or transcription factor binding and hence varying phenotypes (Nowak *et al.,* 2009).

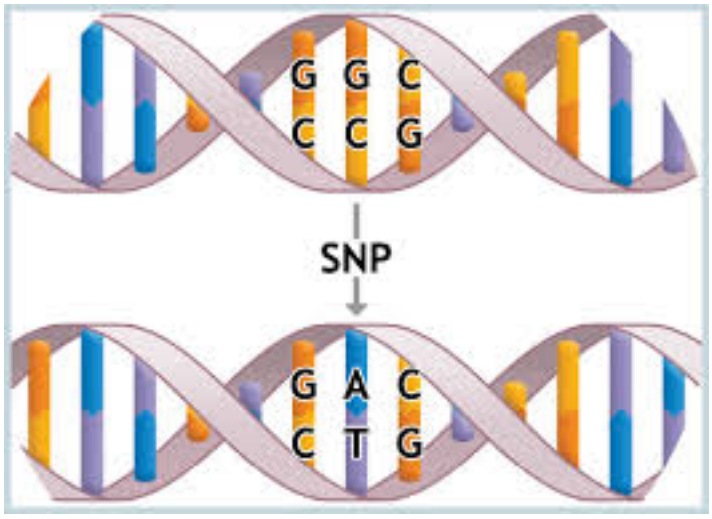

Figure 2.4.1.1 *Basic representation of a SNP (google)*

SNPs are considered the most used genetic marker and provide a large amount of information with regards to the effects of point mutations (Beckmann *et al.,* 2007). Koboldt *et al.* (2006) determined that SNPs are not spaced randomly across the genotype, but instead are clustered together. Due to the fact that variants tend to cluster together, there is a likelihood that a region surrounding a target SNP possibly contains a neighbouring variant (Koboldt *et al.,* 2006). Genotyping has become a widely used method to determine sheep population genetic structure and diversity as well as associations with parasite resistance, production traits and complex disorders in the human genome (Syvanen, 2005; Kijas *et al.*, 2009, 2012; Zhang *et al.*, 2013; Al-Mamun *et al.*, 2015; Benavides *et al.*, 2015). Van Raden (2008) states that information obtained

from genotyping is equivalent to approximately 20 daughters with phenotypic records for dairy cattle. This suggests possible future studies to determine the value of genotypic information and whether it can be used in association with phenotypic records or if genotypic information can completely replace phenotypic information.

Genomic diversity is often maintained by genetic recombination during meiosis. However, chromosomal segments tend to be transmitted as blocks referred to as haplotypes (The International HapMap Consortium, 2005, 2007; Sabeti *et al.*, 2007). This implies that single genomic markers like SNPs in a certain haplotype can be used to predict the genotype of the surrounding region, which is described as linkage disequilibrium (Nowak *et al.,* 2009). Single nucleotide polymorphisms have widely been used as genetic markers due to its genomic abundance and amenability to cost effective high throughput genotyping (Kijas *et al.*, 2009). SNPs affect a single nucleotide pair, however, the genomic abundance of SNPs (over 10 million) means they are the most common source of polymorphic changes (Beckmann *et al*., 2007).

Due to the development of high throughput techniques, genetic structural variations have been found, namely copy number variations (CNV) (Liu *et al.*, 2013). Copy number variants is a structural difference in specific areas of DNA (Thapar & Cooper, 2013). These structural differences are visually shown in *Figure 2.4.1.2.*

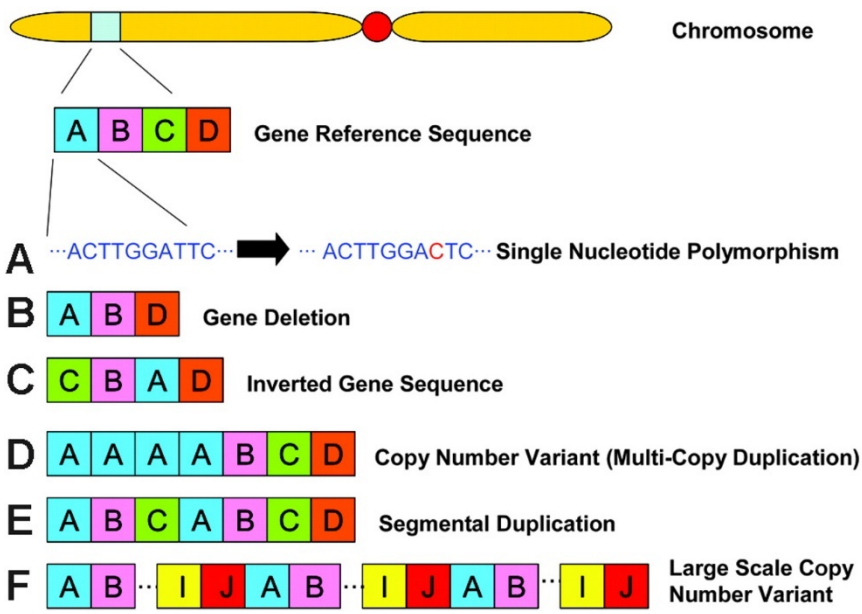

Figure 2.4.1.2 *Basic visual representation of the structural difference between a SNP and a CNV (google)*

These structural variants (SV) can be as a result of mutations, namely; insertions, deletions, inversions and translocations (Liu *et al.*, 2010; Mills *et al.*, 2011), as seen in *Figure 2.4.1.3.* Copy number variations can range in size from 1kb to several mega-bases long (Redon *et al.*, 2006; Ramayo-Caldas *et al.*, 2010). As detection of these polymorphisms improve, a more accurate length will be able to be determined (Zöllner & Teslovich, 2009). CNVs cover a larger genomic region compared to SNPs and so have the potential to have a larger effect on reproduction, production and adaptation. In this review the term variant will be used instead

of polymorphisms to describe copy number changes, as the term polymorphisms is generally used for genetic variants that have a minor allele frequency of ≥ 1% in a given population (Freeman *et al.*, 2006).

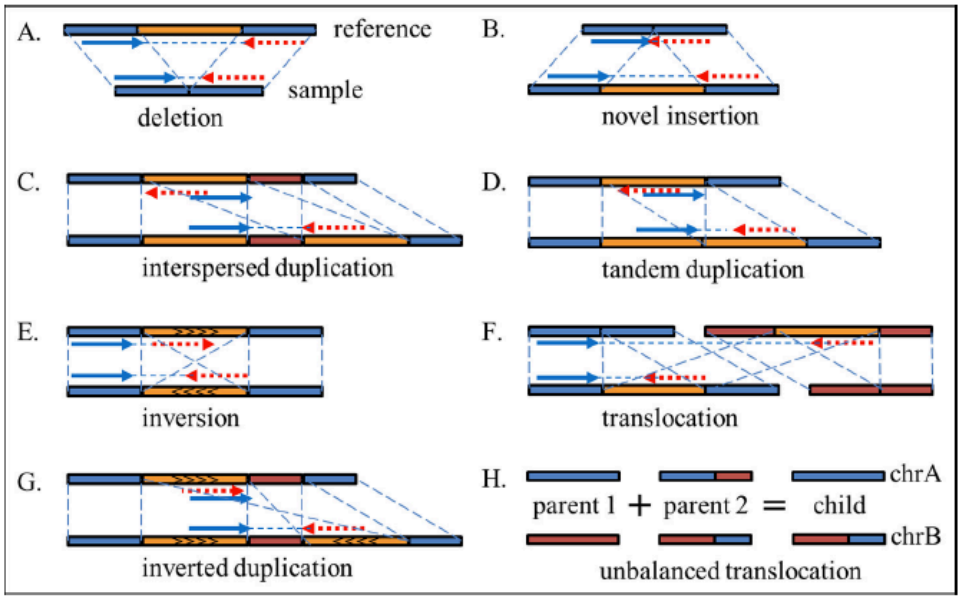

Figure 2.4.1.3 *Different types of SVs and discordantly mapped reads. Blue arrow representing reads from the 50 end and red arrow representing reads from the 3' end. The first line of each SV type in A to G represents the reference genome sequence and the last line represents the sequence in the sample. The orange-coloured sequence is the sequence being deleted, inserted, duplicated or inverted. H shows a compound event leading to an unbalanced translocation* (Guan & Sung, 2016)

According to Khusainova *et al.* (2015) and Jenkins *et al.* (2016) deletions exhibit stronger selective pressure and are better phylogenetic markers of population relationships than duplication polymorphisms. According to Zöllner & Teslovich (2009), the frequency distribution of CNVs, with its strong excess of rare variants, can be interpreted as a signal of purifying selection acting on CNV loci, or even as a signal of population growth. However, based on Conrad *et al.* (2006), deleted regions are relatively gene-poor, which is consistent with the action of purifying selection against deletions. As seen by the large number of genes affected by deletions, it seems highly possible that deletions may be important in the genetic basis of complex traits. Therefore, it is important to continue in the development of high-resolution techniques for studying deletion variation, including genetic association studies (Conrad *et al.*, 2006). Sebat *et al.* (2004) found that SNP genotyping methods assume that every individual has two copies of each locus, while studies of copy number variation assume that individuals vary in their copy number across the genome. CNVs cover a larger genomic region than SNPs and so may have a larger effect through the possible alteration of gene structure and dosage, as well as altering gene regulation (Henrichsen *et al.*, 2009; Wang *et al.*, 2012). As the selection acting on CNVs is more pronounced than that found in SNPs, it can be assumed that CNVs will have a greater functional impact than SNPs, which could negatively affect the reproduction fitness of carriers as well as the production and adaptation fitness of individuals (Zöllner & Teslovich, 2010).

Genetic variants may not directly be responsible for certain genetic diseases, but their presence may lead to unequal recombination or rearrangement of chromosomes that result in specific diseases (Giglio *et al.*,

2001; Osborne *et al.*, 2001; Iafrate *et al.*, 2004; Shaw & Lupski, 2004). These changes in DNA could also influence the expression of traits due to a change in the gene's genomic location or chromatin environment, this is known as position effect (Feuk *et al.,* 2006a). Copy number variants influence the expression of genes in their vicinity, and genes within CNVR often show lower expression and more specific spatial expression patterns than genes mapping elsewhere (Henrichsen *et al.*, 2009b). Gene expression is the basis of many important biological functions in cells, and the influence of SNP and CNVs are an indication of the nature of the mutational and natural selection processes that contribute to genetic diversity and divergence (Stranger *et al.*, 2007). Some CNVs have been found to have phenotypic effects in a number of species including sheep, horses, chickens and cattle (Norris & Whan, 2008; Pielberg *et al.*, 2008; Wright *et al.*, 2009; Zhou *et al.*, 2016). Identifying individuals that possibly possess CNVs that may positively affect desirable economically relevant traits could result in improved production outputs and thus be of economic advantage. Studies need to be carried out to determine whether CNVs are a reliable method to identify individuals that may be a carrier for a recessive disease or disorder that may only be expressed in later generations, or to identify individuals that may have traits that are economically relevant (Henshall *et al.,* 2010).

Copy number variants provide information that complements existing CNV information and can be applied to SNP-based genome wide association and selection studies and is a topic in need of further studies that can provide more information about CNV related to economically beneficial phenotypes (Ma *et al.*, 2015). Known CNVs can be genotyped in case-control populations with similar methods to the SNP-based association studies, and applying association studies on CNVs may provide relevant information with regards to traits that may be economically beneficial to farmers (Winchester *et al.,* 2009).

2.4.2 MOTIVATION FOR MOLECULAR TOOLS

Genomics can assist with the selection of individuals with desired traits that are only expressed at a later age or are not easily passed on to the next generation but are of economic significance (Molotsi *et al.,* 2017b). Genomic selection results in faster genetic gain because you can select individuals at an earlier age than with phenotypic selection. Using genetic evaluations will assist in better genetic gain as well as greater selection accuracy for traits that may have limited phenotypes (Tribout *et al.,* 2013). Significant genetic progress can be made through increasing the accuracies of hard-to-measure traits and traits that are 'late maturing' by using genomic information (Swan *et al.*, 2012). Using production and progeny testing to make selection decisions means that accurate and complete records need to be kept to ensure precise selection decisions are made that will benefit the farmer economically and will result in genetic progress within the flock.

2.4.3 APPLICATION OF GENOMICS

2.4.3.1. Pedigree Inference

Pedigree data coupled with molecular markers can be useful with regards to selection and genetic management within a population at a low cost (Paiva *et al.*, 2011). Due to the fact that many smallholder farmers do not keep records, accurate pedigree information for the flock may be unavailable or incomplete. Multiple-sire mating systems or even extensive mating systems could limit the accuracy of parental assignment and records in production systems (Rupp *et al.*, 2016). Errors in pedigree could have negative effects on the rate of genetic gain within the population as well as result in a decreased accuracy of the calculated breeding values (Sanders *et al.*, 2006). However, correct parentage assignment based on DNA could possibly improve genetic gain due to the increasing accuracy of the pedigree (Heaton *et al.*, 2014).

Genomic technologies have become useful to determine the correct pedigree of individuals where incomplete or inaccurate records have been kept. Marker information can be used as an alternative if pedigree information is unavailable, and could even prove to be more accurate than pedigree information if sufficiently dense markers are used (Hayes & Goddard, 2008). More than 90% of individuals were assigned to the correct breed when microsatellites where used to classify individual animals (Farid *et al.*, 2000). Since, genomic technology can also be used for parentage assignment, it suggests the possibility of using genomic technologies to accurately prove the pedigree of registered individuals such as stud individuals (Alford *et al.*, 1994; Clarke *et al.*, 2014). However, to make this technology available to more producers a reduction in the cost of the microsatellites or SNP panels are needed to ensure more farmers can make use of the benefits associated with the technology such as increased genetic gain through correct pedigree assignment. Additional funding or subsidies from the government could assist farmers with being able to make use of this technology, as well as funding from private and public development agencies. Further studies could also test the potential of using smaller panels for basic pedigree testing that may be more affordable to certain farmers.

2.4.3.2 Genomic Selection

Genome-wide association studies (GWAS) have been used to find genetic variants, single nucleotide polymorphisms, that have been linked/associated with many diseases (Hindorff *et al.*, 2009; Ku *et al.*, 2010). Kijas *et al.* (2009) suggests that genome-wide association analysis may provide a way of highlighting genetic regions and mutations that may influence disease phenotypes and production traits. This suggests that association studies may be economically beneficial for farmers if certain disease genotypes can be identified as well as certain superior production and reproduction genotypes. Genomic prediction of BV for animals has become an additional method to select individuals that possess superior genetic potential that can be utilised (Habier *et al.*, 2007b; Goddard & Hayes, 2009; Daetwyler *et al.*, 2010). Genomic prediction equations (PE) assists with the implementation of genomic selection (GS). Pannier *et al.* (2014) discovered the effect of using BV for meat quality measurements in Australian lambs. It was found that increasing the post-weaning weight

(PWWT) BV, resulted in a decreased tenderness, overall liking, juiciness and flavour scores at two cuts at the Kirby site as well as a decrease on average the tenderness scores within the topside samples at two sites. This study is an example of how BV can aid in genomic selection to produce animals that are more economically beneficial to the farmer as well as desired by the consumer.

Genomic information has the potential to change selection processes and increase genetic gain and should be used in combination with phenotypic data and pedigree information which will increase the accuracy of EBVs and selection indexes (Swan *et al.*, 2012). GEBV accuracy increases when more phenotypic records are considered at certain heritability's of traits, thus further emphasising the need to use phenotypic, pedigree and genomic information together in order to set up breeding programmes to improve production (Burgueño *et al.*, 2012). Screening superior individuals for SNPs associated with specific reproduction or production traits such as parasite resistance or increased ovulation rates will have a definite economic advantage for the farmer as well as ensure a greater genetic gain (Zhang *et al.*, 2013).

PARASITE RESISTANCE

Host resistance is possibly due to changes to the host-parasite relationship where the balance has been shifted to favour the parasite due to modern husbandry practises that have allowed highly susceptible individuals to survive in their respective population (Beh & Maddox, 1996). Production costs, the inevitable development of anthelmintic-resistant parasites and the growing concerns of potentially adverse environmental impacts of grazing industries, prompted the search for non-chemotherapeutic methods for parasite control. The use of antibiotics and chemotherapeutic methods have enabled susceptible individuals to survive and produce in the population resulting in a decrease in selection for resistant individuals. The use of chemotherapeutic methods may have the ability to hide the disease phenotype in individuals, thus providing an inaccurate count of individuals that appear to be resistant but are actually susceptible, but do not show the disease symptoms. Genetic selection of resistant individuals presented a cheap, environmentally friendly way to control gastrointestinal parasites, while also following the trend for sustainable agriculture (Beh & Maddox, 1996; Benavides *et al.*, 2015).

Phenotypically, resistance to parasites is a continuously variable and complex trait and can statistically be described as a quantitative trait. The number of genes that determine quantitative traits vary and could be a result of the combined small effects of many genes, or they could be due to major genes whose effects are obscured by genetic and phenotypic heterogeneity modified by environmental variation (Beh & Maddox, 1996). Successful genetic marker development requires knowledge of the precise genetic basis of parasite resistance. Phenotypic heterogeneity may be a result of incomplete penetrance of a gene; meaning that some individuals who inherit a predisposing allele may not manifest the disease, resulting in the individual dying before the disease is expressed (Beh & Maddox, 1996).

Selection for resistance to certain internal parasites is an important breeding goal for future parasite control within flocks or on farms. While there are phenotypic markers available to assist with determining the level of infection in sheep, the phenotype is not always a reliable way to determine susceptibility,

especially when the disease expression may be controlled by a number of small effects of many genes. For this reason genetic markers could be an effective way to incorporate resistance into selection programmes (Beh & Maddox, 1996). Variation in resistance to internal parasites was found based on the faecal egg counts that were recorded and were used as an indication for disease susceptibility (Stewart & Miller, 1937; Mcewan *et al.*, 1992). This observed variation in disease susceptibility has led to attention being given to the development of genetic markers for parasite resistance. The variation in disease resistance found in both within-breed and between-breed, indicates the possibility of using this genetic variation to construct effective breeding programmes to select for resistant individuals for specific parasites. The variability of the faecal egg score heritability estimate of ewe lambs (0.34 ± 0.19) observed by Watson *et al.* (1986), further emphasises the possibility of parasite resistance selection. A favourable genetic correlation with faecal egg count (FEC), body weight and scrotal traits was found by Matebesi-Ranthimo *et al.* (2014). This suggests that rams with low levels of gastro-intestinal nematodes possibly have higher testis measurements, resulting in a possible effect on the flock's fertility. This verifies the need to determine the disease resistance status of native and exotic breeds in the various production systems.

Native sheep breeds generally have a higher parasite resistance than exotic breeds (Piedrafita *et al.*, 2010). This suggests that in many production systems the replacement of exotic breeds with native breeds is a not a viable option; however, it may be possible to utilise resistance genes from native breeds through introgression using backcrossing or transgenesis into more productive breeds. For this to be possible, a selection marker is needed that can accurately and cheaply assess the resistance status of individuals (Beh & Maddox, 1996). Studies have been aimed at identifying regions in the genome, which present resistance or susceptibility to parasites (Beh & Maddox, 1996). These experiments involve genotyping the entire genome of sheep families using enough hypervariable DNA markers to tag all chromosomes, the presence of a gene with significant effect on the trait of interest can then be determined using least squares analysis (Beh & Maddox, 1996).

BOOROOLA and INVERDALE

Ewes that possess the Booroola gene have been found to have high ovulation rates and litter sizes; this is due to the ($FecB^{B}$) allele of the major autosomal gene (FecB) that influences the number of ovulations per oestrous cycle in the ewe (Mulsant *et al.*, 2001). Montgomery *et al.* (1993) reported linkage of the FecB locus to a syntenic group, including two anonymous ovine microsatellite markers and restricted fragment length polymorphisms (RFLPs) for *SPP1* and *EGF*. The Inverdale gene (FecX) which also has an influence on ovulation, has been located on the X chromosome; and a *FecX* mutation could possibly result in the complete loss of *BMP15* function, which in turn could affect the action of growth differential factor 9 (GDF9) (Galloway *et al.*, 2000). A study conducted by Davis *et al.* (1992) found that the effects for both the Inverdale and Booroola gene were multiplicative for ovulation rate. Suggesting that the Booroola gene, in the presence or absence of the Inverdale gene increased the ovulation rate by 90%, while the Inverdale gene, in the presence or absence of the Booroola gene increased the ovulation rate by 44% (Davis *et al.*, 1992).

The ability of identifying these reproduction genes in individuals within a flock could be economically beneficial to the farmer. Increased ovulation in ewes will result in a larger number of offspring and thus more profits for the farmer.

2.4.3.3 Genetic Diversity of Populations

The genetic diversity of many sheep breeds has been analysed on a global scale (Peter *et al.*, 2007). Soma *et al.* (2012) conducted population structure analysis on 20 different sheep breeds and found that indigenous breeds have a uniqueness which could be a result of local adaptations over time. This further supports the need for world-wide population structure analysis to improve the conservation strategies of the remaining sheep breeds in existence, both indigenous and foreign.

HETEROZYGOSITY

The degree of heterozygosity in a population can be a good indication of the genetic variation within the population (Swan *et al.*, 2012). The Hardy-Weinberg law describes the consequences of random mating on allele and genotype frequencies (Gillespie, 1998). When a population has no external forces acting on it (mutation, migration and selection) causing a change in the gene frequencies, and random mating occurs we can say the population is in Hardy-Weinberg Equilibrium (Hamilton, 2011b). The gene frequency equation is: p2 + 2pq q2 = 1. Populations with genotype frequencies that do not fit Hardy-Weinberg expectations are evidence that one or more evolutionary processes/forces are acting on it to determine the genotype frequencies. Forces that cause a change in gene frequency within a population are migration, mutation, selection, genetic drift and the founder effect.

GENETIC STRUCTURE

Farmers need more than just pedigree and performance records to make effective management and breeding decisions. A good understanding of the population genetic structure and diversity is also needed to identify factors that may have had an influence on the genetic history of the population (Valera *et al.*, 2005). Genetic distances explain the differences between populations in terms of number of mutations, differences in allele frequencies or genetic drift (Eding & Bennewitz, 2007). Knowing the genetic history of the flock will ensure the farmer can make informed future management and breeding decisions. Population structure will also give a good indication of the animal genetic resources (AnGR) available that could assist with the improvement of production efficiency.

Genetic variability can be defined as the variability of alleles and genotypes found in the population (Boichard *et al.*, 1997). Wright's F-statistic is an effective way to measure the pattern of population subdivision through F_{IS}, F_{ST}, and F_{IT}. Barros *et al.* (2017) explains the F-statistics as the F_{ST}, which determines the loss of heterozygosity (reduction in diversity) in a subpopulation compared to the total population, F_{IS}, determines the loss of heterozygosity within the subpopulation and the F_{IT}, which estimates the loss of heterozygosity of the entire population. F_{ST} compares the variance of allele frequencies between the populations (Holsinger & Weir, 2009). $F_{ST} > 0.25$ indicates great genetic differentiation between the

populations, and can thus be assumed that allele frequencies are different; F_{ST} < 0.25 indicates moderate to little genetic differentiation, meaning that the allele frequencies within each population are similar (Hamilton, 2011c). Large F_{ST} values at a locus are indicators of a differentiation between populations, which suggests directional selection, while a small F_{ST} value at a locus indicates that the population is homogenous which suggests balancing or directional selection in both populations (Vitti *et al.*, 2013). Directional selection indicates selection against the dominant phenotype, which results in a loss of alleles because the dominant allele is not shielded from natural selection in the heterozygote state, however, balancing selection indicates heterozygote advantage, and so maintains both alleles in the population (Nielsen, 2005). Within- and among-breed genetic diversity is important; within-breed genetic diversity assists with breed management, while among-breed genetic diversity assists with the identification of divergent breeds that may have distinct genotypes (Handley *et al.*, 2007).

INBREEDING, F_{IS}

An individual can be considered inbred if its parents are more closely related than any two other individuals picked randomly from the population (Keller & Waller, 2002; Mokhtari *et al.*, 2014). Individuals that are inbred have a higher chance of homozygosity due to identical by decent (IBD) at all loci across the genome compared to an individual from an outbred population (Overall *et al.*, 2005).

Linebreeding also makes use of inbreeding with the objective of producing individuals that are closely related to an admired ancestor that displays desirous traits (Lush, 1933). Allele frequencies stay the same during inbreeding, but there is a change in the genotype frequencies. Inbreeding increases the frequency of homozygous genotypes at the same tempo that heterozygous genotypes decrease, causing a change in the genetic structure of the population and favouring homozygosity in the gene pool (Hamilton, 2011b; Vostry *et al.*, 2018). This results in a decrease in genetic variability, thus decreasing possible future genetic gains (Goddard, 2009). Decreased genetic variability and an increase in homozygosity in a population may have a negative impact on fitness-related traits and may result in an increase in the expression of phenotypic defects, recessive genotypes as well as lethal genotypes (Fernández *et al.*, 1995; Falconer & Mackay, 2009). Lethal genotypes are often hidden because they are carried in the population in a heterozygous state. Although inbreeding results in the purging of a population of some of its deleterious genes, some of the populations do not survive such intensive inbreeding. Populations that do survive the high intensity of inbreeding are often at a disadvantage with lower fitness and random changes in morphological, physiological and behavioural traits (Soulé, 1980). An example of the occurrence of a recessive lethal genotype would be in the Karakul sheep where the dominant grey colour is lethal when homozygous (WRWR) (Schoeman, 1998). Black lambs cause a significant loss in the sheep industry due to the demand for white wool (Fleet, 2006). Black coat colour has been found to be recessive and is influenced by the differential expression of the ASIP gene (Royo *et al.*, 2008). Recessive genotypes often have an economically detrimental effect on the sheep industry as shown with the economic losses caused by a recessive genotype such as black coat colour.

The level of inbreeding for an individual is influenced by a number of factors, namely; the ratio of males to females, the mating system, reproduction ability as well as the population size, while the inbreeding estimates are influenced by the depth and completeness of pedigree and the selection intensity (Barczak *et al.*, 2009). A high level of inbreeding could possibly lead to inbreeding depression (Leroy, 2014). One of the factors believed to have a major influence on inbreeding depression is the increase in homozygosity for deleterious recessive alleles due to inbreeding (Charlesworth & Charlesworth, 1999). Inbreeding depression causes a decrease in the value of a trait associated with inbreeding (Wright, 1984), thus affecting performance in the population (Gholizadeh & Ghafouri-Kesbi, 2016). The increase in homozygosity and loss of allelic diversity can lead to lowered production and fitness for individuals that are inbred (Selvaggi *et al.*, 2010). Keller & Waller (2002) and Dorostkar *et al.* (2012) suggest that inbreeding depression often has a significant effect on traits such as production, survival, reproduction, growth, birth weight as well as disease resistance and health. This lower production and reproduction indices within populations can be related to the fact that inbred animals are less flexible to environmental changes (Barros *et al.*, 2017). A certain level of genetic diversity needs to be maintained in order to ensure that the future generations will be able to respond to selection pressure, environmental changes that might occur as well as possible changes in market conditions, new knowledge of human nutrition requirements or even consumer demands (van Wyk *et al.*, 1993, 2009; FAO, 2000).

Leroy (2014) suggests that some populations or traits may be affected by inbreeding depression to a larger extent than others, and states that this could be due to an interaction between inbreeding and the environment. Changes to natural habitat (e.g., due to climate change) causes stress to certain individuals or populations resulting in the possible decline in the populations size and an increase in the inbreeding levels (Reed *et al.*, 2012). In domestic animals inbreeding depression could decrease the response to selection, as well as the expected potential genetic gain of economically important traits (Selvaggi *et al.*, 2010). Inbreeding depression often has an unstated assumption that most of the effects of inbreeding affect fitness, however Amos *et al.* (2001) states that many individuals in a population are often unaffected by inbreeding depression, especially once they have successfully reached adulthood.

Figure 2.7.1.1 illustrates the effect of interaction between the environment and an inbred population will have on the fitness of the population.

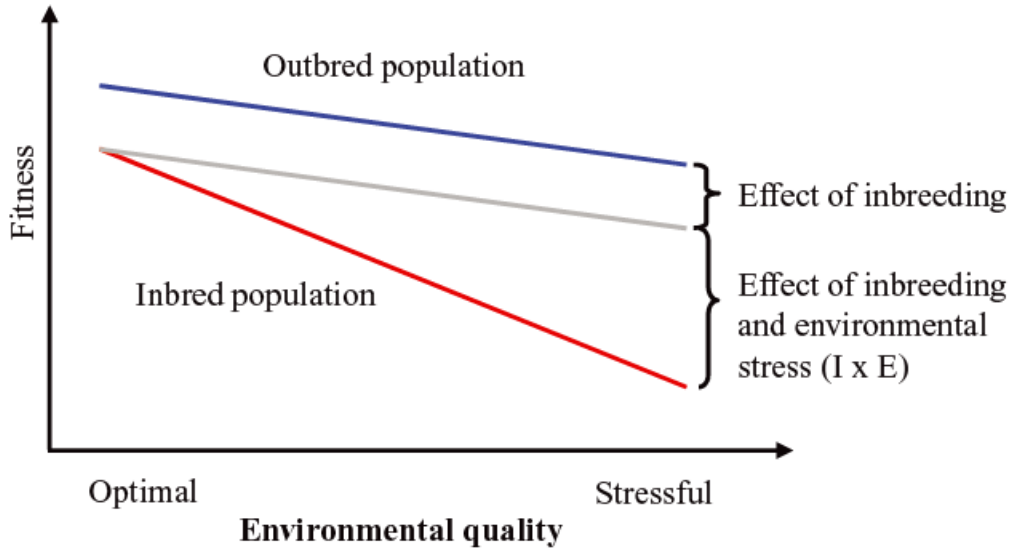

Figure 2.4.3.1 *Assuming inbreeding is independent of environment, the reduction in fitness as a result of reduced environmental quality will be equal for outbred and inbred populations. The blue and grey lines illustrate fitness of an outbred and an inbred population, respectively, in the absence of inbreeding-environment interactions. Inbreeding depression is, however, often more severe under stressful environmental conditions. Thus, the red line illustrates fitness of an inbred population assuming inbreeding-environment interactions.* (Reed *et al.*, 2012)

In general, Reed *et al.* (2012) states that inbreeding-stress interactions can result from 1) the effect of exposure to stress on the expression of deleterious alleles (genetic level) and/or, 2) the phenotypic effects caused by the expression of genetic load that affects resistance to stress (phenotypic level).

Inbreeding levels are lower when pedigree information is incomplete and the Ne is overestimated (Vostry *et al.*, 2018). According to the FAO of the United Nations guidelines, (FAO, 1998), the rate of inbreeding should not exceed 1%. Inbreeding has become a problem due to a decrease of gene flow as well as the reduction in population size as a result of urbanisation and a decrease in available land for agriculture leading to mating of close relatives, resulting in a reduction in fitness due to inbreeding. One proposed way to overcome inbreeding and inbreeding depression could be through the development of reproductive biotechnologies that could facilitate genetic exchange between populations at a domestic as well as an international level (Selvaggi *et al.*, 2010). Correct record-keeping along with structured breeding programmes could all contribute to decreasing or preventing inbreeding within a population.

EFFECTIVE POPULATION SIZE

The effective population size, N_e, is the number of reproducing individuals that would lead to the actual increase in inbreeding if they all contributed equally to the next generation (van Wyk *et al.,* 2009; Leroy *et al.*, 2013). Thus, N_e is a function of the relative increase in inbreeding or the variance of gene frequency from one generation to another (Boichard *et al.,* 1997). Inbreeding leads to an increase in homozygosity, which means that the effective population size can give an indication of the likely loss of heterozygosity across the alleles in the population.

As mentioned by Handley *et al.* (2007), a decrease in the effective population size is due to a number of reasons; less rams are being used for breeding due to artificial insemination and improved transport,

production systems that focus mainly on a select few breeds to the detriment of rare or important breeds. Some breeds have also been lost by introgression into commercial populations, further decreasing the population size. This decrease in effective population size causes a decrease in genetic variability and a loss of possible advantageous traits. The FAO of the United Nations (FAO, 1998) suggest a minimum N_e of 50 for genetic diversity conservation, while Mace & Lande (1991) estimates the effective population as roughly a fifth of the total population. Franklin (1980) proposed the 50/500 concept. This concept states that the short-term effective population size should not be less than 50, while the long-term minimum effective population size should be 500.

Knowing the effective population size of the flock will allow the farmer to keep the number of animals that will prove to be the most efficient. The N_e will also provide the farmer with the suggested number of animals to be kept in the flock to ensure genetic gain and to ensure there is no undesired inbreeding within the population (Hanrahan *et al.*, 1973). Knowledge of the correct effective population size will ensure the farmer will not have to keep unnecessary animals which would decrease the production costs due to less animals and thus increase the expected profits.

RUNS OF HOMOZYGOSITY

SNPs have aided in the identification of Runs Of Homozygosity (ROH) which could detect identical by descent (IBD) chromosomal regions (Marras *et al.*, 2014). ROH are stretches of homozygous genotypes which could be due to the inheritance of the same haplotype from both parents (Al-Mamun *et al.*, 2015). ROH have been suggested to be influenced by inbreeding as well as selection (Zhang *et al.*, 2015a). We can thus assume that longer lengths of ROH could be due to recent selection that occurred in the population, while shorter lengths of ROH could be due to recombination.

2.4.3.4 Mutations

Polymorphisms are responsible for a large portion of genetic variation. Mutations are the permanent incorporation of random errors in the DNA that results in differences between the ancestral and descendent copies of DNA (Hamilton, 2011a). Changes in the DNA could result in genetic variants that may influence physiological, biochemical, morphological and pathological processes and functions in the population (Feuk *et al.,* 2006a). This could thus affect gene function and possibly the reproductive fitness of organisms. Some mutations do not have an effect since the translocation of the DNA results in the synthesis of the same protein, this is called a synonymous mutation; while other mutations that result in different amino acid sequences are known as nonsynonymous mutations (Hamilton, 2011a). Mutations can be advantageous for survival and reproduction or they can result in reduced survival and reproduction capabilities, while some mutations are neutral, and thus have no positive or negative effect on the individual (Nielsen, 2001, 2005). A mutation that is selectively favoured over the ancestral allele will be spread in the population through natural selection until it becomes fixed in the population; while the new allele is being fixed, polymorphisms in the flanking region may be wiped out, leading to a reduced genetic variation in this region (Schlötterer &

Wiehe, 1999). Micro-mutationalism is the view that beneficial mutations fixed by the process of natural selection have small effects and therefore, that the process of adaptation is marked by gradual genetic change (Hamilton, 2011a). Mutations thus result in a diverged line in the population and an increased genetic variance. This divergence and increased genetic variance can be beneficial with regards to the individual's ability to adapt to changes in the environment over time. SNPs and CNVs are genetic structural variants that could be due to mutations on the genome of individuals (Liu *et al.*, 2010; Mills *et al.*, 2011).

Due to the challenges of finding and typing copy number variants, and the scarcity of the basic knowledge of their location and molecular structures, a possible method to identify CNVs may be to rely on SNPs to serve as markers by linkage disequilibrium for common variants throughout the genome (McCarroll & Altshuler, 2007). SNPs have been found to be in linkage disequilibrium with a number of CNVs (de Bakker *et al.*, 2005; Hinds *et al.*, 2006; Locke *et al.*, 2006).

Some complex diseases have been associated with copy number variants, such as glomerulonephritis (Aitman *et al.*, 2006), autism (Sebat *et al.*, 2007) and HIV-1 susceptibility (Gonzalez *et al.*, 2005) in humans. Aldred *et al*. (2005) found that β-Defensin copy number is correlated with expression level, allowing the possibility that different expression levels could lead to varying susceptibility to infectious diseases. Increase in copy number variants have been linked with increased levels of messenger RNA (mRNA) (Linzmeier & Ganz, 2005). A correlation between the genomic copy number of DEFB4 and levels of its messenger RNA (mRNA) transcript was found by Hollox *et al.* (2003). The peptides encoded by these genes are strong anti-microbial agents and are effective against certain clinical pathogens such as *Pseudomonas aeruginosa* and *Staphylococcus aureus.* Thus, these variants affect the mRNA transcript and ultimately influence the immune system functioning. In a study conducted by Hou *et al.* (2012), they found CNVs that were associated with and possibly contributed to parasite resistance in Angus cattle, due to the genes these CNVs were found to overlap. This CNV association with disease resistance in Angus raises the interest of possible future studies to determine whether there may be CNV associations with diseases in other species, namely sheep.

Endogenous retroviruses have been found in the genomes of pigs, sheep, cattle, horses, chickens and other domestic animals (Cousens *et al.*, 1999; Klymiuk *et al.*, 2002; Lee *et al.*, 2002; Palmarini *et al.*, 2002; Garcia-Etxebarria *et al.*, 2014). Ovine pulmonary adenocarcinoma (OPA), also known as Jaagsiekte, is a contagious lung tumour in sheep caused by Jaagsiekte sheep retrovirus (JSRV) (Demartini *et al*., 1988; York *et al.*, 1992). Jaagsiekte sheep retrovirus is usually a disease in adult sheep from the ages between one and four years, and often causes coughing, weight loss and a watery nasal discharge. Jaagsiekte sheep retrovirus causes the tumours in the lungs to increase in size and results in the tumours excreting fluids in the lungs which affects respiration (Griffiths *et al*., 2010). While the Jaagsiekte sheep retrovirus exJSRV and its endogenous counterpart enJSRV co-exist in sheep, copy variants have been found at the R-to-W amino acid change that is able to block exJSRV budding from the cells, providing a potential protective role for the host (Viginier *et al.*, 2012). From these studies we can assume that there is a potential to use CNVs as a selection

method for certain disease susceptibility genotypes that will be beneficial to individuals and thus be economically significant for farmers.

Zhou *et al.* (2016) identified a total of 231 CNVs in *Bos indicus* cattle, where 17 of them were significantly associated with growth traits and seven being moderate to highly correlated with growth traits. Three of these CNVs were significantly associated with all seven of the growth traits identified by (Zhou *et al.*, 2016). A positive correlation between transcriptional expression and copy numbers in Chinese cattle indicate the promising impact of the MYH3 gene copy number variants on growth traits due to the effect on muscle development (Xu *et al.*, 2014b). In a study carried out by Durán Aguilar *et al.* (2017), 24 and 47 copy number variation regions were significantly associated with estimated breeding values for somatic cell score. Copy number variations could possibly be used to select dams for milk traits that will be beneficial to their lambs and ensure greater lamb weaning weights because the dams were able to produce good quality and quantity milk for the lambs. A total of 34 CNVs were significantly associated with at least one milk production trait in Holsteins, demonstrating the need for further studies in sheep to determine if it could be a viable option to select for in sheep (Xu *et al.*, 2014a).

According to Bidwell *et al.* (2009), sheep possess the most identified genetic mutations associated with reproduction compared to any livestock species, and many of these mutations have been mapped to specific chromosomal regions (Montgomery *et al.*, 1994; Mulsant *et al.*, 2001; Rohrer, 2004; McNatty *et al.*, 2005). Several contributing mutations for reproductive traits like Booroola and Inverdale have been identified, however, there are many mutations for other traits of interest that are still unknown (Galloway *et al.*, 2000; Mulsant *et al.*, 2001; Souza *et al.*, 2001; Wilson *et al.*, 2001). Fertility is an important trait with regards to production, and a CNV deletion has been identified that causes a recessive embryonic lethal mutation which negatively influences the cattle industry (Kadri *et al.*, 2014). However, this mutation is negatively correlated to milk yield, resulting in a negative genetic correlation between fertility and milk yield. Based on the findings in the study conducted by Shin *et al.* (2010), it is stated that future research might allow the examination of the genetic effects of CNVs on numerous economic traits.

Knowing how selection for desired traits occurred over time will provide a better understanding of how to effectively select for these specific traits. It is thus also important to have a complete understanding of how CNVs evolved over time or how their selection occurred. Knowing how CNVs evolved and allowed animals to adapt to certain conditions will allow for more informed decision making with regards to selection for breeding animals as well as traits of economic value in certain environments. The CNVs present in different sheep breeds need to be studied in order to determine the effect the CNVs could have had on the adaption of certain breeds to specific environments and whether adaptable individuals can be selected based on CNVs. This could be determined by doing a population differential analysis or population clustering on CNVs found in the different sheep breeds. Knowing whether certain CNVs may be predominant in certain sheep populations may assist with creating effective crossbreeding programmes that could be beneficial to farmers by producing economically relevant traits in the extreme environments. Feuk *et al*. (2006) carried

out a gene ontology (GO) analysis that showed statistically significant enrichment of genes that are involved in general defence responses, and are an indication that genes involved in structural variation may have effects on the response to external pressure. These genes are believed to be more 'flexible' resulting in a greater potential to evolve quickly, and so might be important for gene and organismal evolution.

Xu *et al.* (2017) observed clear distinctions in CNV prevalence between diverse groups, as well as a large proportion of CNVR shared by multiple groups; several lineage-specific CNVRs were identified showing different patterns of distribution among the seven sheep groups. Copy number variable genes are a well-established cause of gene family member differentiation and a common mechanism underpinning evolutionary change (Liu *et al.*, 2019). These lineage-specific CNVRs that were identified by Xu *et al.* (2017) allow the possibility of using CNVRs and CNVs to determine the evolutionary history of the different sheep populations and so better understand the selection signatures and selection pressures that may have resulted in these different CNVRs. In a study involving Qinchuan cattle (Zhang *et al.*, 2015b), functional analysis indicated that most of the genes in the CNVRs were involved in environmental stress, thus allowing the possibility of using CNVs or CNVRs to determine the evolutionary history and adaptation or selection direction of a population. Studies are needed to determine the relevance copy number variations have among different sheep breeds and the effect it could have on selection for production and disease resistance traits in sheep (Fontanesi *et al.*, 2011). To fully understand the part CNVs play in disease as well as possibly adaptation/fitness traits, the biological processes that creates CNVs needs to be fully identified, improve the sensitivity and specificity of the methods used to identify CNVs, and develop statistical methods that fully leverage CNV signals that exist in data obtained from genome wide genotyping arrays as well as next-generation sequencing technologies (Zöllner & Teslovich, 2010).

2.5 SOUTH AFRICAN SHEEP

The Dorper sheep has been found to be a highly favoured breed in South Africa due to its adaptability to less than optimal environments. The Dorper was established due to the demand for good slaughter lambs in the harsh climate of South Africa (Cloete *et al.*, 2000) and has proven itself with its adaptiveness and robustness. The Namaqua Afrikaner originates from the dry, harsh areas of the North West Cape and southern Namibia where they were kept by the Nama people (Qwabe *et al.,* 2013). The Namaqua Afrikaner has been found to be a very hardy and prolific breed due to their relatively high reproductive performance recorded under extensive conditions (Snyman *et al.,* 1993). Snyman (2010) defines hardiness as the ability of an animal to economically produce and reproduce under adverse environmental conditions. The Dorper and the Namaqua Afrikaner are two indigenous breeds that have the potential to be good producing breeds that are adapted to the climate of South Africa. Correct breeding programmes are thus needed to ensure these two breeds meet their production potential. To enable this, further studies are needed to determine the genetic structure of the populations as well as to identify possible mutations, such as CNVs that may be beneficial for adaptation or production purposes.

2.5.1 SOUTH AFRICAN SHEEP GENETIC STRUCTURE

To increase efficiency, breed replacement and crossbreeding has become an alternative considered by many farmers. Documenting sheep population diversity is becoming more important due to the increasing popularity of crossbreeding (Hanotte & Jianlin, 2005). Crossbreeding results in the mixing of the breeds involved in the cross, meaning that certain genetic characteristics of each individual gets diluted or possibly lost, leading to a decrease in genetic diversity and a loss of genetic animal resources. However, crossbreeding often results in heterosis or hybrid vigour. Hybrid vigour is a measure of the increased productivity associated with the crossing of two different pure breeds, hybrid vigour is the amount by which the offspring will differ from the average of the two parental breeds. Higher heterosis in offspring is often observed when the parental breeds are more genetically diverse. Individuals will only show heterosis if their gene frequencies differ. Traits that have low heritability often have higher heterosis, such as growth rate, viability and fertility. The presence of heterosis in crosses suggests that the breeds differ in gene frequency, and that some detrimental recessive alleles may have a higher than expected frequency in some breeds (Goddard, 2009a). One of the major challenges of managing animal genetic resources is ensuring that the animals are genetically adjusted or adapted to the production system and environment in which they have to produce, while maintaining an adequate level of genetic variability to ensure possible adaption in the future if needed (Pilling & Hoffmann, 2011).

Knowledge of the sheep population diversity may assist with local breed selection that may have higher production outputs than an exotic breed, thus ensuring there is no loss of adaption animal genetic resources. The effective population size of different sheep breeds has decreased substantially due to stud breeding and intensive management (Peter *et al.*, 2007). Biosecurity and quality control have also become more important due to the exchange of genetics, thus increasing the need for accurate genetic evaluations world-wide (Neser *et al.,* 2000).

2.6 CONCLUSION

The arid climate that makes up a large portion of South Africa, is less than ideal for a number of livestock species, however, sheep have been able to survive and produce relatively well under these harsh environmental conditions. The sheep production that occurs in South Africa consists of smallholder farmers and commercial farmers. There is still a lot of potential within the smallholder communities to increase their production through correct breeding programmes and correct and accurate record-keeping. Recording measurements of individuals will allow the farmer to identify the top and bottom producers in the flock and so allow for further management decisions to be made with regards to selection of superior individuals and the replacement of below average animals (Richards *et al.*, 2010). Different breeding practises are used by farmers all over the country depending on their production goals. However, farmers in smallholder communities do not have specific breeding or selection programmes. The absence of breeding and selection programmes could result in little to no genetic gain within the flock.

There are specific traits that may be of interest to farmers that could assist in increasing production outputs as well as genetic gain within their flock. Reproduction traits are important to ensure that there will be a next generation. Important reproduction traits that should be considered in breeding programmes include fertility traits, multiple offspring traits as well as mothering ability to ensure the offspring will survive. Production traits are just as important and should be aligned with the goals of the farmer as well as the demands of the market. Some production traits in the sheep industry include wool quality, carcass conformation as well as fat content.

With the improvement of genomic technology there are a lot of options available eg., genomic selection, for smallholder farmers to increase production and thus increase profits. Genomic technologies, namely MAS, can result in an increased genetic gain as well as increased production within the population. Individuals with adaptation/robustness traits need to be identified and included in breeding programmes to ensure animals are able to adapt to the changing climate. Genomic selection will assist with the selection of individuals that may be genetically superior for specific traits that are in line with the breeding and production goals of the farmer.

Structural variants have been found to have resistance to gastrointestinal nematodes in Angus cattle (Hou *et al.*, 2012), a potential protective role from Jaagsiekte in sheep (Viginier *et al.*, 2012), associations with growth traits in cattle (Zhou *et al.*, 2016) as well as association with milk production traits in Holsteins (Xu *et al.*, 2014a). Further studies are needed to determine whether there are associations between production or reproduction traits and structural variants in sheep.

Constructing effective breeding programmes is only possible if the farmer has a complete and accurate knowledge of the genetic resources within the flock. Genetic resources that are important for breeding programmes are the genetic structure and diversity within the flock. Knowledge of any possible variants within the population could also assist with the construction of breeding programmes.

2.7 REFERENCES

Ademosun, A., 1994. Constraints and Prospects for small ruminant research and development in Africa. Page 268 in Proceedings of the Second Biennial Conference of teh African Small Ruminant Research Network. Lebbie, S.H., Rey, B., Irungu, E., eds. Addis Ababa, Ethiopia.

Aguilar, M., Ponce, S.I., López, F.J., Padilla, E., Peláez, C.G., Bagnato, A. & Strillacci, M., 2017. Genome-wide association study for milk somatic cell score in holstein cattle using copy number variation as markers. J. Anim. Breed. Genet. 134, 49–59 https://doi.org/10.1111/jbg.12238.

Aitman, T., Dong, R., Vyse, T., Norsworthy, P., Johnson, M., Smith, J., Mangion, J., Roberton-Lowe, C., Marshall, A., Petretto, E., Hodges, M., Bhangal, G., Patel, S., Sheehan-Rooney, K., Duda, M., *et al.,* 2006. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature. 439, 851–855 https://doi.org/10.1038/nature04489.

Al-Atiya, R., 2015. The power of 28 microsatellite markers for parentage testing in sheep. Electron J Biotechn. 18, 116–121 https://doi.org/10.1016/j.ejbt.2015.01.001.

Al-Mamun, H., Clark, S., Kwan, P. & Gondro, C., 2015. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. Genet. Sel. Evol. 47, 1–14 https://doi.org/10.1186/s12711-015-0169-6.

Aldred, P.M., Hollox, E. & Armour, J.A., 2005. Copy number polymorphism and expression level variation of the human α-defensin genes DEFA1 and DEFA3. Hum. Mol. Genet. 14, 2045–2052 https://doi.org/10.1093/hmg/ddi209.

Alford, R., Hammond, H., Coto, I. & Caskey, C., 1994. Rapid and efficient resolution of parentage by amplification of short tandem repeats. Am. J. Hum. Genet. 55, 190–195.

Amos, W., Wilmer, J., Fullard, K., Burg, T., Croxall, J., Bloch, D. & Coulson, T., 2001. The influence of parental relatedness on reproductive success. Proc. R. Soc. Lond. B 268, 2021–2027 https://doi.org/10.1098/rspb.2001.1751.

Anderson, S., 2003. Animal genetic resources and sustainable livelihoods. Ecol. Econ. 45, 331–339 https://doi.org/10.1016/S0921-8009(03)00088-0.

Arranz, J., Bayón, Y. & San Primitivo, F., 1998. Genetic relationships among Spanish sheep using microsatellites. Anim. Genet. 29, 435–440 https://doi.org/10.1046/j.1365-2052.1998.296371.x.

Bae, J., Cheong, H., Kim, L., NamGung, S., Park, T., Chun, J., Kim, J., Pasaje, C.F., Lee, J. & Shin, H., 2010. Identification of copy number variations and common deletion polymorphisms in cattle. BMC Genom. 11 https://doi.org/10.1186/1471-2164-11-232.

de Bakker, P.I., Yelensky, R., Pe'er, I., Gabriel, S., Daly, M. & Altshuler, D., 2005. Efficiency and power in genetic association studies. Nat. Genet. 37, 1217–1223 https://doi.org/10.1038/ng1669.

Barczak, E., Wolc, A., Wójtowski, J., Ślósarz, P. & Szwaczkowski, T., 2009. Inbreeding and inbreeding depression on body weight in sheep. J. Anim. Feed Sci. 18, 42–50 https://doi.org/10.22358/jafs/66366/2009.

Barros, E., de A. Brasil, L., Tejero, J., Delgado-Bermejo, J. & Ribeiro, M., 2017. Population structure and genetic variability of the Segureña sheep breed through pedigree analysis and inbreeding effects on growth traits. Small Rumin. Res. 149, 128–133 https://doi.org/10.1016/j.smallrumres.2017.02.009.

Beckmann, J.S., Estivill, X. & Antonarakis, S.E., 2007. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat. Rev. Genet. 8, 639–646.

Beh, K. & Maddox, J., 1996. Prospects for Development of Genetic Markers for Resistance to Gastrointestinal Parasite Infection in Sheep. Int. J. Parasitol. 26, 879–897 https://doi.org/10.1016/S0020-7519(96)80060-9.

Benavides, M., Sonstegard, T., Kemp, S., Mugambi, J., Gibson, J., Baker, R., Hanotte, O., Marshall, K. & Van Tassell, C., 2015. Identification of novel loci associated with gastrointestinal parasite resistance in a red Maasai x Dorper backcross population. PLoS One. 10, 1–20 https://doi.org/10.1371/journal.pone.0122797.

Bidwell, C., Cockett, N., Maddox, J. & Beever, J., 2009. Genome Mapping and Genomics in Domestic Animals (N. Cockett & C Kole, Eds.). 3rd ed. Springer.

Boichard, D., Maignel, L. & Verrier, É., 1997. The value of using probabilities of gene origin to measure genetic variability in a population. Genet. Sel. Evol. 29, 5–23 https://doi.org/10.1186/1297-9686-29-1-5.

Buduram, P., 2004. Genetic Characterization of Southern African Sheep Breeds Using DNA Markers. MSc thesis, Univ. Free State.

Burgueño, J., de los Campos, G., Weigel, K. & Crossa, J., 2012. Genomic Prediction of Breeding Values when Modeling Genotype × Environment Interaction using Pedigree and Dense Molecular Markers. Crop Sci. 52, 707–719 https://doi.org/10.2135/cropsci2011.06.0299.

Cadmus, W., 1949. Sheep Records for Greater Profits. Agric. Exp. Station. Circ. 182.

Chamberlin, J., 2007. Defining Smallholder Agriculture in Ghana : Who Are Smallholders , What Do They Do and How Are They Linked With Markets ?

Charlesworth, B. & Charlesworth, D., 1999. The genetic basis of inbreeding depression. Genet. Res., Camb. 74, 329–340 https://doi.org/10.1017/S0016672399004152.

Chiwawa, T., 2019. Integration of information and communication technology (ICT) in the development of a recordkeeping database of smallholder sheep farming systems in the Western Cape Province of South Africa.

Clarke, S., Henry, H., Dodds, K., Jowett, T.W., Manley, T., Anderson, R. & McEwan, J., 2014. A high throughput single nucleotide polymorphism multiplex assay for parentage assignment in New Zealand sheep. PLoS One. 9, e93392 https://doi.org/10.1371/journal.pone.0093392.

Cloete, S.W. & Olivier, J., 2010. South African sheep and wool Industries. Pages 95–104 in International Sheep and Wool Handbook.

Cloete, S.W., Scholtz, A., Gilmour, A. & Olivier, J. 2002. Genetic and environmental effects on lambing and neonatal behaviour of Dormer and SA Mutton Merino lambs. Livest. Prod. Sci. 78, 183–193 https://doi.org/10.1016/S0301-6226(02)00117-3.

Cloete, S.W., Snyman, M. & Herselman, M., 2000. Productive performance of Dorper sheep. Small Rumin. Res. 36, 119–135 https://doi.org/10.1016/S0921-4488(99)00156-X.

Conrad, D., Andrews, T., Carter, N., Hurles, M. & Pritchard, J., 2006. A high-resolution survey of deletion polymorphism in the human genome. Nat. Genet. 38, 75–81 https://doi.org/10.1038/ng1697.

Cousens, C., Minguijon, E., Dalziel, R., Ortin, A., Garcia, M., Park, J., Gonzalez, L., Sharp, J. & de las Heras, M., 1999. Complete Sequence of Enzootic Nasal Tumor Virus, a Retrovirus Associated with Transmissible Intranasal Tumors of Sheep. J. Virol. 73, 3986–93.

Daetwyler, H., Hickey, J., Henshall, J., Dominik, S., Gredler, B., Van Der Werf, J.H. & Hayes, B., 2010. Accuracy of estimated genomic breeding values for wool and meat traits in a multi-breed sheep population. Anim. Prod. Sci. 50, 1004–1010 https://doi.org/10.1071/AN10096.

DAFF., 2012. A framework for the development of smallholder farmers through cooperative development.

DAFF., 2018. Trends in the Agricultural Sector.

Davis, G.H., Dodds, K.G. & Bruce, G.D., 1992. Combined effect of the Inverdale and Booroola prolifacacy genes on ovulation rate in sheep. Proc. Assoc. Advmt. Anim. Breed. Genet. 13, 74–77.

Delgado, C., Rosegrant, M., Steinfeld, H., Ehui, S. & Courbois, C., 1999. Live stock to 2020: The Next Food Revolution. Outlook Agric. 30.

Demartini, J., Rosadio, R. & Lairmore, M., 1988. The Etiology and Pathogenesis of Ovine Pulmonary Carcinoma (Sheep Pulmonary Adenomatosis). Vet. Microbiol. 17, 219–236 https://doi.org/10.1016/0378-1135(88)90067-3.

Diez-Tascón, C., Littlejohn, R., Almeida, P.A. & Crawford, A., 2000. Genetic variation within the Merino sheep breed: Analysis of closely related populations using microsatellites. Anim. Genet. 31, 243–251 https://doi.org/10.1046/j.1365-2052.2000.00636.x.

Dorostkar, M., Arough, H., Shodja, J., Rafat, S., Rokouei, M. & Esfandyari, H., 2012. Inbreeding and inbreeding depression in Iranian Moghani sheep breed. J. agric. sci. technol. 14, 549–556.

Douch, P.G., Green, R., Morris, C., Mcewan, J. & Windon, R., 1996. Phenotypic markers for selection of nematode-resistant sheep. Int. J. Parasitol. 26, 899–911 https://doi.org/10.1016/S0020-7519(96)80062-2.

Duguma, G., Schoeman, S.J., Cloete, S.W.P. & Jordaan, G.F., 2002a. Genetic parameter estimates of early growth traits in the tygerhoek Merino flock. S. Afr. J. Anim. Sci. 32, 66–75 https://doi.org/10.4314/sajas.v32i2.3747.

Duguma, G., Schoeman, S., Cloete, S.W. & Jordaan, G., 2002b. Genetic and environmental parameters for ewe productivity in Merinos. S. Afr. J. Anim. Sci. 32, 154–159 https://doi.org/10.4314/sajas.v32i3.3740.

Duru, S., 2017. Pedigree analysis of the Turkish Arab horse population: structure, inbreeding and genetic variability. Animal. 11, 1449–1456 https://doi.org/10.1017/s175173111700009x.

Eding, H. & Bennewitz, J., 2007. Measuring genetic diversity in farm animals.Pages 103–130 in Utilisation and conservation of farm animal genetic resources. Oldenbroek, K., ed. Wageningen Academic Publishers, Netherlands.

Ellegren, H., Moore, S., Robinson, N., Byrne, K., Ward, W. & Sheldon, B., 1997. Microsatellite evolution - A reciprocal study of repeat lengths at homologous loci in cattle and sheep. Mol. Biol. Evol. 14, 854–860 https://doi.org/10.1093/oxfordjournals.molbev.a025826.

Ellis, F., 1991. Peasant Economics: Farm Households and Agrarian Development. Am. Anthr. 93 https://doi.org/10.2307/635095.

Engelbrecht, C. & Engelbrecht, F., 2016. Shifts in Köppen-Geiger climate zones over southern Africa in relation to key global temperature goals. Theor Appl Clim. 123, 247–261 https://doi.org/10.1007/s00704-014-1354-1.

Eteqadi, B., Hossein-Zadeh, N. & Shadparvar, A., 2015. Inbreeding effects on reproductive traits in Iranian Guilan sheep. Trop Anim Heal. Prod. 47, 533–539 https://doi.org/10.1007/s11250-015-0758-y.

Falconer, D. & Mackay, T.F., 2009. Introduction to Quantitative Genetics. 4. ed. Pearson, Prentice Hall, Harlow.

FAO., 1998. Management of small populations at risk.

FAO., 2000. World watch list for domestic animal diversity (BD Scherf, Ed.). 3rd ed. Rome.

Farid, A., O'Reilly, E., Dollard, C. & Kelsey, C., 2000. Genetic analysis of ten sheep breeds using microsatellite markers. Can. J. Anim. Sci. 80, 9–17 https://doi.org/10.4141/A99-086.

Fernández, A., Toro, M. & López-Fanjul, C., 1995. The effect of inbreeding on the redistribution of genetic variance of fecundity and viability in Tribohum castaneum. Heredity. (Edinb). 75, 376–381 https://doi.org/10.1038/hdy.1995.149.

Fernando, R., Habier, D., Stricker, C., Dekkers, J. C. & Totir, L., 2007. Genomic selection. Acta Agric Scand A Anim Sci. 57, 192–195 https://doi.org/10.1080/09064700801959395.

Feuk, L., Carson, A. & Scherer, S., 2006a. Structural variation in the human genome. Nat. Rev. Genet. 7, 85–97 https://doi.org/10.1038/nrg1767.

Feuk, L., Marshall, C., Wintle, R. & Scherer, S., 2006b. Structural variants: changing the landscape of chromosomes and design of disease studies. Hum. Mol. Genet. 15, 57–66 https://doi.org/10.1093/hmg/ddl057.

Fleet, M., 2006. Development of black pigmented skin spots and pigmented wool fibres in a Merino flock—causes, field observations, and wool measurement. Aust. J. Exp. Agric. 57, 751 https://doi.org/10.1071/ar05032.

Fogarty, N., 1995. Genetic parameters for live weight, fat and muscle measurements, wool production and reproduction in sheep: a review. Anim. Breed. Abstr. 63, 101–143.

Fogarty, N., 2009. MEAT SHEEP BREEDING – WHERE WE ARE AT AND FUTURE CHALLENGES. Proc. Assoc. Advmt. Anim. Breed. Genet. 18, 414–421.

Fontanesi, L., Beretti, F., Martelli, P., Colombo, M., Dall'Olio, S., Occidente, M., Portolano, B., Casadio, R., Matassino, D. & Russo, V., 2011. A first comparative map of copy number variations in the sheep genome. Genomics. 97, 158–165 https://doi.org/10.1016/j.ygeno.2010.11.005.

Franklin, I., 1980. Evolutionary change in small populations. Pages 135–149 in Conservation Biology: An evolutionary-ecological perspective. Soule, M., Wilcox, B., eds. Sinauer Associates, Inc., Massachusetts.

Freeman, J.L., Altshuler, D.M., Jones, K.W., Aburatani, H., Perry, G.H., Redon, R., Scherer, S.W., Lee, C., McCarroll, S A., Carter, N.P., Tyler-Smith, C., Feuk, L. & Hurles, M.E., 2006. Copy number variation: new insights in genome diversity. Genome Res. 16, 949–961 https://doi.org/10.1101/gr.3677206.

Friggens, N., Blanc, F., Berry, D. & Puillet, L., 2017. Review: Deciphering animal robustness. A synthesis to facilitate its use in livestock breeding and management. Animal. 11, 2237–2251 https://doi.org/10.1017/S175173111700088X.

Galloway, S., McNatty, K., Cambridge, L., Laitinen, M.P., Juengel, L., Jokiranta, T., McLaren, R., Luiro, K., Dodds, K., Montgomery, G., Beattie, A., Davis, G. & Ritvos, O., 2000. Mutations in an oocyte-derived growth factor gene (BMP15) cause increased ovulation rate and inferility in a dosage-sensitive manner. Nat. Genet. 25, 279–283.

Garcia-Etxebarria, K., Sistiaga-Poveda, M. & Jugo, B., 2014. Endogenous Retroviruses in Domestic Animals. Curr.Genom. 15, 256–265 https://doi.org/10.2174/1389202915666140520003503.

Gholizadeh, M. & Ghafouri-Kesbi, F., 2016. Inbreeding depression in growth traits of Baluchi sheep. Small Rumin. Res. 144, 184–190 https://doi.org/10.1016/j.smallrumres.2016.09.012.

Giglio, S., Broman, K., Matsumoto, N., Calvari, V., Gimelli, G., Neumann, T., Ohashi, H., Voullaire, L., Larizza, D., Giorda, R., Weber, J., Ledbetter, D. & Zuffardi, O., 2001. Olfactory Receptor–Gene Clusters, Genomic-Inversion Polymorphisms, and Common Chromosome Rearrangements. Am. J. Hum. Genet. 68, 874–883 https://doi.org/10.1086/319506.

Gillespie, J., 1998. Population Genetics: A Concise Guide. The Johns Hopkins University Press, Baltimore, Maryland.

Gizaw, S., van Arendonk, J.A., Valle-Zárate, A., Haile, A., Rischkowsky, B., Dessie, T. & Mwai, A., 2014a. Breeding programmes for smallholder sheep farming systems: II. Optimization of cooperative village breeding schemes. J. Anim. Breed. Genet. 131, 350–357 https://doi.org/10.1111/jbg.12102.

Gizaw, S., Rischkowsky, B., Valle-Zárate, A., Haile, A., van Arendonk, J.A., Mwai, A. & Dessie, T., 2014b. Breeding programs for smallholder sheep farming systems: I. Evaluation of alternative designs of breeding schemes. J.

Anim. Breed. Genet. 131, 341–349 https://doi.org/10.1111/jbg.12101.

Goddard, M., 1992. Optimal Effective Population Size for the Global Population of Black and White Dairy Cattle. J. Dairy Sci. 75, 2902–2911 https://doi.org/10.3168/jds.s0022-0302(92)78052-7.

Goddard, M., 2009. Fitness traits in animal breeding programs. In Adaptation and Fitness in Animal Populations: Evolutionary and Breeding Perspectives on Genetic Resource Management. van der Werf, J., Hans-Ulrich, G., Richard, F., Gondro, C., eds. Springer.

Goddard, M. & Hayes, B., 2007. Genomic selection. J. Anim. Breed. Genet. 124, 323–330.

Goddard, M. & Hayes, B., 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10, 381–391 https://doi.org/10.1038/nrg2575.

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R., Freedman, B., Quinones, M., Bamshad, M., Murthy, K., Rovin, B., Bradley, W., Clark, R., Anderson, S., *et al.,* 2005. The Influence of CCL3L1 Gene-Containing Segmental Duplications on HIV-1/AIDS Susceptibility. Science. 307, 1434–1440.

Gowane, G., Prakash, V., Chopra, A. & Prince, L.L., 2013. Population structure and effect of inbreeding on lamb growth in Bharat Merino sheep. Small Rumin. Res. 114, 72–79 https://doi.org/10.1016/j.smallrumres.2013.06.002.

Griffiths, D., Martineau, H. & Cousens, C., 2010. Pathology and Pathogenesis of Ovine Pulmonary Adenocarcinoma. J. Comp. Path. 142, 260–283 https://doi.org/10.1016/j.jcpa.2009.12.013.

Guan, D., Luo, N., Tan, X., Zhao, Z., Huang, Y., Na, R., Zhang, J. & Zhao, Y., 2016. Scanning of selection signature provides a glimpse into important economic traits in goats (Capra hircus). Sci. Rep. 6, 1–7 https://doi.org/10.1038/srep36372.

Guan, P. & Sung, W., 2016. Structural variation detection using next-generation sequencing data: A comparative technical review. Methods. 102, 36–49 https://doi.org/10.1016/j.ymeth.2016.01.020.

Habier, D., Fernando, R. & Dekkers, J.C., 2007a. The impact of genetic relationship information on genome-assisted breeding values. Genetics. 177, 2389–2397 https://doi.org/10.1534/genetics.107.081190.

Habier, D., Fernando, R. & Dekkers, J.C., 2007b. The impact of genetic relationship information on genome-assisted breeding values. Genet. 177, 2389–2397 https://doi.org/10.1534/genetics.107.081190.

Hamilton, M., 2011a. Population genetics. John Wiley & Sons, Chichester.

Hamilton, M., 2011b. Population structure and gene flow. Pages 105–153 in Population Genetics. Wiley-Blackwell.

Hamilton, M., 2011c. Genotype Frequencies. Pages 9–52 in Population Genetics. Wiley-Blackwell.

Hammond, K., 2000. A Global Strategy for the Development of Animal Breeding Programmes in Lower-Input Production Environments. Pages 85–94 in Animal breeding and animal genetic resources: proceedings of Workshop 7 on Sustainable animal production. Glodek, P., Groeneveld, E., eds. FAL Agricultural Research, Brunswick.

Hancock, J., 1999. Microsatellites and other simples sequences: genomic context and mutational mechanisms. Pages 1–9 in Microsatellites: evolution and application. Goldstein, D.., Schlötterer, C., eds. Oxford University Press, Inc., New York.

Handley, L.J., Byrne, K., Santucci, F., Townsend, S., Taylor, M., Bruford, M. & Hewitt, G., 2007. Genetic structure of European sheep breeds. Heredity. (Edinb). 99, 620–631 https://doi.org/10.1038/sj.hdy.6801039.

Hanotte, O. & Jianlin, H., 2005. Genetic Characterization of Livestock Populations and Its Use in Conservation Decision-Making. Pages 131–136 in The role of biotechnology. Ruane, J., Sonnino, A., eds. Food and Agricultural Organization of the United Nations, Turin, Italy.

Hanrahan, J., Eisen, E. & Legates, J., 1973. Effects of Population Size and Selection Intensity on Short-term Response to

Selection for Posweaning gain in Mice. Genet. 73, 513–530.

Hatcher, S. & Atkins, K., 2007. The impact of phenotypic selection for clean fleece weight on reproduction. Proc. Assoc. Advmt. Anim. Breed. Genet. 17, 260–263.

Hayes, B., Bowman, P., Chamberlain, A. & Goddard, M., 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92, 433–43 https://doi.org/10.3168/jds.2008-1646.

Hayes, B. & Goddard, M., 2008. Technical note: Prediction of breeding values using marker-derived relationship matrices. J. Anim. Sci. 86, 2089–2092 https://doi.org/10.2527/jas.2007-0733.

Heaton, M., Leymaster, K., Kalbfleisch, T., Kijas, J., Clarke, S., McEwan, J., Maddox, J., Basnayake, V., Petrik, D., Simpson, B., Smith, T.P., Chitko-Mckown, C. & The International HapMap Consortium., 2014. SNPs for parentage testing and traceability in globally diverse breeds of sheep. PLoS One. 9, e94851 https://doi.org/10.1371/journal.pone.0094851.

Henrichsen, C., Chaignat, E. & Reymond, A., 2009a. Copy number variants, diseases and gene expression. Hum. Mol. Genet. 18, 1–8 https://doi.org/10.1093/hmg/ddp011.

Henrichsen, C., Vinckenbosch, N., Zöllner, S., Chaignat, E., Pradervand, S., Schütz, F., Ruedi, M., Kaessmann, H. & Reymond, A., 2009b. Segmental copy number variation shapes tissue transcriptomes. Nat. Genet. 41, 424–429 https://doi.org/10.1038/ng.345.

Henshall, J., Whan, V. & Norris, B., 2010. Reconstructing CNV genotypes using segregation analysis: Combining pedigree information with CNV assay. Genet. Sel. Evol. 42, 1–8 https://doi.org/10.1186/1297-9686-42-34.

Hindorff, L., Sethupathy, P., Junkins, H., Ramos, E., Mehta, J., Collins, F. & Manolio, T., 2009. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Pages 9362–9367 in Proc. Natl. Acad. Sci. U.S.A. IEEE Comput. Soc.

Hinds, D., Kloek, A., Jen, M., Chen, X. & Frazer, K., 2006. Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat. Genet. 38, 82–85 https://doi.org/10.1038/ng1695.

Hollox, E.J., Armour, J.A.L. & Barber, J.C.K., 2003. Extensive Normal Copy Number Variation of a β-Defensin Antimicrobial-Gene Cluster. Am. J. Hum. Genet. 73, 591–600.

Holsinger, K. & Weir, B., 2009. Genetics in geographically structured populations: defining, estimating and interpreting FST. Nat. Rev. Genet. 10, 639–50 https://doi.org/10.1038/nrg2611.

Hou, Y., Liu, G., Bickhart, D., Matukumalli, L., Li, C., Song, J., Gasbarre, L., Van Tassell, C. & Sonstegard, T., 2012. Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes in Angus cattle. Func Integr Genomics. 12, 81–92 https://doi.org/10.1007/s10142-011-0252-1.

Hynd, P., 1989. Effects of nutrition on wool follicle cell kinetics in sheep differing in efficiency of wool production. Aust. J. Agric. Res. 40, 409–417 https://doi.org/10.1071/AR9890409.

Iafrate, A., Feuk, L., Rivera, M., Listewnik, M., Donahoe, P., Qi, Y., Scherer, S. & Lee, C., 2004. Detection of large-scale variation in the human genome. Nat. Genet. 36, 949–951 https://doi.org/10.1038/ng1416.

Iñiiguez, L., 2011. The challenges of research and development of small ruminant production in dry areas. Small Rumin. Res. 98, 12–20 https://doi.org/10.1016/j.smallrumres.2011.03.010.

International HapMap Consortium., 2005. A haplotype map of the human genome. Nature. 437, 1299–1320.

International Hapmap Consortium, Frazer, K., Ballinger, D., Cox, D., Hinds, D., Stuve, L., Gibbs, R., Belmont, J., Boudreau, A., Hardenbol, P., Leal, S., Pasternak, S., Wheeler, D., Willis, T., Yu, F., *et al.,* 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature. 449, 851–861 https://doi.org/10.1038/nature06258.

International Sheep Genomics Consortium, Archibald, A., Cockett, N., Dalrymple, B., Faraut, T., Kijas, J., Maddox, J., McEwan, J., Oddy, V., Raadsma, H., Wade, C., Wang, J., Wang, W., & Xun, X. 2010. The sheep genome reference sequence: A work in progress. Anim. Genet. 41, 449–453 https://doi.org/10.1111/j.1365-2052.2010.02100.x.

Jenkins, G., Goddard, M., Black, M., Brauning, R., Auvray, B., Dodds, K., Kijas, J., Cockett, N. & McEwan, J., 2016. Copy number variants in the sheep genome detected using multiple approaches. BMC Genom. 17, 1–14 https://doi.org/10.1186/s12864-016-2754-7.

Kadri, N., Sahana, G., Charlier, C., Iso-Touru, T., Guldbrandtsen, B., Karim, L., Nielsen, U., Panitz, F., Aamand, G., Schulman, N., Georges, M., Vilkki, J., Lund, M. & Druet, T., 2014. A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic Red Cattle: Additional Evidence for the Common Occurrence of Balancing Selection in Livestock. PLoS Genet. 10, e1004049.

Karam, H., 1959. Selecting Yearling Rahmani Sheep. J. Anim. Sci. 18, 1452–1459.

Kashi, Y. & Soller, M., 1999. Functional roles of microsatellites and minisatellites.Pages 10–23 in Microsatellites: evolution and application. Goldstein, D., Schlötterer, C., eds. Oxford University Press, Inc., New York.

Keller, L. & Waller, D., 2002. Inbreeding effects in wild populations. Trends Ecol. Evol. 17, 230–241.

Kijas, J., Lenstra, J., Hayes, B., Boitard, S., Neto, L. R., Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J. & Dalrymple, B., 2012. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 10, e1001258 https://doi.org/10.1371/journal.pbio.1001258.

Kijas, J., Townley, D., Dalrymple, B., Heaton, M., Maddox, J., McGrath, A., Wilson, P., Ingersoll, R., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Oddy, V., Nicholas, F. & Raadsma, H., 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS One. 4, e4668 https://doi.org/10.1371/journal.pone.0004668.

Klymiuk, N., Mu, M., Brem, G. & Aigner, B., 2002. Characterization of Porcine Endogenous Retrovirus y pro-pol nucleotide sequences. J. Virol. 76, 11738–11743 https://doi.org/10.1128/JVI.76.22.11738–11743.2002.

Knap, P., 2005. Breeding robust pigs. Aust. J. Exp. Agric. 45, 763–773.

Koboldt, D., Miller, R. & Kwok, P., 2006. Distribution of Human SNPs and Its Effect on High-Throughput Genotyping. Hum. Mutat. 27, 249–254 https://doi.org/10.1002/humu.20286.

Kosgey, I.S., Baker, R.L., Udo, H.M. & Van Arendonk, J.A., 2006. Successes and failures of small ruminant breeding programmes in the tropics: A review. Small Rumin. Res. 61, 13–28 https://doi.org/10.1016/j.smallrumres.2005.01.003.

Kosgey, I. & Okeyo, A., 2007. Genetic improvement of small ruminants in low-input, smallholder production systems: Technical and infrastructural issues. Small Rumin. Res. 70, 76–88 https://doi.org/10.1016/j.smallrumres.2007.01.007.

Ku, C., Loy, E., Pawitan, Y. & Chia, K., 2010. The pursuit of genome-wide association studies: Where are we now? J. Hum. Genet. 55, 195–206 https://doi.org/10.1038/jhg.2010.19.

Lee, J., Webb, G., Allen, R.D. & Moran, C., 2002. Characterizing and Mapping Porcine Endogenous Retroviruses in Westran Pigs. J. Virol. 76, 5548–5556 https://doi.org/10.1128/jvi.76.11.5548-5556.2002.

Leroy, G., 2014. Inbreeding depression in livestock species: Review and meta-analysis. Anim. Genet. 45, 618–628 https://doi.org/10.1111/age.12178.

Leroy, G., Mary-Huard, T., Verrier, E., Danvy, S., Charvolin, E. & Danchin-Burge, C., 2013. Methods to estimate effective

population size using pedigree data: Examples in dog, sheep, cattle and horse. Genet. Sel. Evol. 45, 1–10.

Linzmeier, R. & Ganz, T., 2005. Human defensin gene copy number polymorphisms: Comprehensive analysis of independent variation in α- and β-defensin regions at 8p22-p23. Genomics. 86, 423–430 https://doi.org/10.1016/j.ygeno.2005.06.003.

Liu, G., Hou, Y., Zhu, B., Cardone, M., Jiang, L., Cellamare, A., Mitra, A., Alexander, L., Coutinho, L., Dell'Aquila, M., Gasbarre, L., Lacalandra, G., Li, R., Matukumalli, L., Nonneman, D., de A. Regitano, L., Smith, T.P., *et al.,* 2010. Analysis of copy number variations among diverse cattle breeds. Genome Res. 20, 693–703 https://doi.org/10.1101/gr.105403.110.

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. BMC Genom. 14, 1–11.

Liu, M., Zhou, Y., Rosen, B., Van Tassell, C., Stella, A., Tosser-Klopp, G., Rupp, R., Palhière, I., Colli, L., Sayre, B., Crepaldi, P., Fang, L., Mészáros, G., Chen, H. & Liu, G., 2019. Diversity of copy number variation in the worldwide goat population. Heredity. (Edinb). 122 https://doi.org/10.1038/s41437-018-0150-6.

Locke, D., Sharp, A., Mccarroll, S., Mcgrath, S., Newman, T., Cheng, Z., Schwartz, S., Albertson, D., Pinkel, D., Altshuler, D. & Eichler, E., 2006. Linkage Disequilibrium and Heritability of Copy-Number Polymorphisms within Duplicated Regions of the Human Genome. J. Hum. Genet. 79, 275–290.

Lush, J., 1933. Linebreeding. Bulletin 25.

Ma, Y., Zhang, Q., Lu, Z., Zhao, X. & Zhang, Y., 2015. Analysis of copy number variations by SNP50 BeadChip array in Chinese sheep. Genomics. 106, 295–300 https://doi.org/10.1016/j.ygeno.2015.08.001.

Mace, G. & Lande, R., 1991. Assessing Extinction Threats: Toward a Reevaluation of IUCN Threatened Species Categories. Conserv. Biol 5, 148–157.

Marai, I.F., El-Darawany, A., Fadiel, A. & Abdel-Hafez, M.A., 2007. Physiological traits as affected by heat stress in sheep-A review. Small Rumin. Res. 71, 1–12 https://doi.org/10.1016/j.smallrumres.2006.10.003.

Markelova, H., Meinzen-Dick, R., Hellin, J. & Dohrn, S., 2009. Collective action for smallholder market access. Food Policy. 34, 1–7 https://doi.org/10.1016/j.foodpol.2008.10.001.

van Marle-Koster, E., Visser, C., Makgahlela, M. & Cloete, S.W., 2015. Genomic technologies for food security: A review of challenges and opportunities in Southern Africa. Food Res. Int. 76, 971–979 https://doi.org/10.1016/j.foodres.2015.05.057.

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., Williams, J. & MacCiotta, N.P., 2014. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim. Genet. 46, 110–121 https://doi.org/10.1111/age.12259.

Martin, G., Milton, J.T., Davidson, R., Hunzicker, G.E., Lindsay, D. & Blanche, D., 2004. Natural methods for increasing reproductive efficiency in small ruminants. Anim. Reprod. 82–83, 231–245 https://doi.org/10.1016/j.anireprosci.2004.05.014.

Matebesi-Ranthimo, P.A.M., Cloete, S.W.P., van Wyk, J.A. & Olivier, J.J., 2014. Relationships of faecal worm egg count with body weight and male fertility in South African Merinos. Proc. 10th World Congr. Genet. Appl. to Livest. Prod., 1–4.

Mavrogenis, A., 1995. Breeding systems and selection strategies for sheep improvement in Cyprus. CIHEAM. 26, 17–26.

McCarroll, S. & Altshuler, D., 2007. Copy-number variation and association studies of human disease. Nat. Genet. 39, S37–S42 https://doi.org/10.1038/ng2080.

Mcewan, J., Mason, P., Baker, R., Clarke, J., Hickey, S. & Turner, K., 1992. Effect of selection for productive traits on internal parasite resistance in sheep Breeding. Proc. New Zeal. Soc. Anim. Prod. 52, 53–56 https://doi.org/10.13140/2.1.1059.1362.

McNatty, K., Galloway, S., Wilson, T., Smith, P., Hudson, N., O'Connell, A., Bibby, A., Heath, D., Davis, G., Hanrahan, J. & Juengel, J., 2005. Physiological effects of major genes affecting ovulation rate in sheep. Genet. Sel. Evol. 37, S25-s38 https://doi.org/10.1051/gse:2004029.

Mills, R., Walter, K., Stewart, C., Handsaker, R., Chen, K., Alkan, C., Abyzov, A., Yoon, S., Ye, K., Cheetham, R., Chinwalla, A., Conrad, D., Fu, Y., Grubert, F., Hajirasouliha, I., Hormozdiari, F., Iakoucheva, L., *et al.,* 2011. Mapping copy number variation by population-scale genome sequencing. Nature. 470, 59–65 https://doi.org/10.1038/nature09708.

Mohammadi, H., Shahrebabak, M. & Sadeghi, M., 2013. Association Between Single Nucleotide Polymorphism in the Ovine DGAT1 Gene and Carcass Traits in Two Iranian Sheep Breeds. Anim Biotechnol. 24, 159–167 https://doi.org/10.1080/10495398.2013.763816.

Mokhtari, M., Shahrbabak, M., Esmailizadeh, A., Shahrbabak, H. & Gutierrez, J., 2014. Pedigree analysis of Iran-Black sheep and inbreeding effects on growth and reproduction traits. Small Rumin. Res. 116, 14–20 https://doi.org/10.1016/j.smallrumres.2013.10.008.

Molotsi, A., Dube, B., Oosting, S., Marandure, T., Mapiye, C., Cloete, S.W. & Dzama, K., 2017. Genetic traits of relevance to sustainability of smallholder sheep farming systems in South Africa. Sustainability. 9, 1–18 https://doi.org/10.3390/su9081225.

Molotsi, A., Halimani, T., Monau, P. & Dzama, K., 2019. Technology as a transformative tool in livestock production. Pages 124–131 in Transforming Agriculture in Southern Africa: Constraints, Technologies,Policies and Processes. Sikora, R., Terry, E., Vlek, P.L., Chitja, J., eds.

Montgomery, G., Crawford, A., Penty, J., Dodds, K., Ede, A., Henry, H., Pierson, C., Lord, E., Galloway, S., Schmack, A., Sise, J., Swarbrick, P., Hanrahan, V., Buchanan, F. & Hill, D., 1993. The ovine Booroola fecundity gene (FecB) is linked to markers from a region of human chromosome 4q. Nat. Genet. 4, 410–414.

Montgomery, G., Lord, E., Penty, J., Dodds, K., Broad, T., Cambridge, L., Sunden, S.L., Stone, R. & Crawford, A., 1994. The Booroola Fecundity (FECB) Gene Maps to Sheep Chromosome 6. Genomics. 22, 148–153.

Mormède, P., Foury, A., Terenina, E. & Knap, P., 2011. Breeding for robustness: The role of cortisol. Animal. 5, 651–657 https://doi.org/10.1017/S1751731110002168.

Mueller, J., Rischkowsky, B., Haile, A., Philipsson, J., Mwai, O., Besbes, B., Zárate, A., Tibbo, M., Mirkena, T., Duguma, G., Sölkner, J. & Wurzinger, M., 2015. Community-based livestock breeding programmes: Essentials and examples. J. Anim. Breed. Genet. 132, 155–168 https://doi.org/10.1111/jbg.12136.

Mühlenbein, H., 1997. The Equation for Response to Selection and Its Use for Prediction. Evol. Comput. 5, 303–347.

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., Cognie, Y., Chitour, N. & Elsen, J., 2001. Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in Booroola Merino ewes. Proc. Natl. Acad. Sci. U.S.A 98, 5104–5109 https://doi.org/10.1073/pnas.091577598.

Nardone, A., Ronchi, B., Lacetera, N., Bernabucci, U. & Ranieri, M., 2010. Effects of climate changes on animal production and sustainability of livestock systems. Livest. Sci. 130, 57–69 https://doi.org/10.1016/j.livsci.2010.02.011.

Negussie, E., Abegaz, S. & Rege, J.E., 2002. Genetic Trend and Effects of Inbreeding on Growth Performance of Tropical

Fat-Tailed Sheep. In 7th World Congress on Genetics Applied to Livestock Production. Montpellier, France.

Neser, F.W., Erasmus, G. & Van Wyk, J., 2000. Genetic studies on the South African Mutton Merino: Growth traits. S. Afr. J. Anim. Sci. 30, 172–177.

Nielsen, R., 2001. Statistical tests of selective neutrality in the age of genomics. Heredity. (Edinb). 86, 641–7.

Nielsen, R., 2005. Molecular Signatures of Natural Selection. Annu. Rev. Genet. 39, 197–218 https://doi.org/10.1146/annurev.genet.39.073003.112420.

Norris, B. & Whan, V., 2008. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. Genome Res. 18, 1282–1293 https://doi.org/10.1101/gr.072090.107.

Nowak, D., Hofmann, W. & Koeffler, H., 2009. Genome-wide mapping of copy number variations using SNP arrays. Transfus Med Hemoth. 36, 246–251.

Nsoso, S., Podisi, B., Otsogile, E., Mokhutshwane, B. & Ahmadu, B., 2004. Phenotypic Characterization of Indigenous Tswana Goats and Sheep Breeds in Botswana: Continuous Traits. Trop Anim Heal. Prod 36, 789–800.

OECD, & FAO., 2019. OECD-FAO Agricultural Outlook 2019-2028. OECD, Rome.

Orr, H., 1998. The Population Genetics of Adaptation: the Distribution of Factors Fixed During Adaptive Evolution. Evol. Int. J. 52, 935–949.

Osborne, L., Li, M., Pober, B., Chitayat, D., Bodurtha, J., Mandel, A., Costa, T., Grebe, T., Cox, S., Tsui, L. & Scherer, S., 2001. A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. Nat. Genet. 29, 321–325 https://doi.org/10.1038/ng753.

Overall, A.D., Byrne, K., Pilkington, J. & Pemberton, J., 2005. Heterozygosity, inbreeding and neonatal traits in Soay sheep on St Kilda. Mol. Ecol. 14, 3383–3393 https://doi.org/10.1111/j.1365-294X.2005.02682.x.

Paiva, S., Facó, O., Faria, D., Lacerda, T., Barretto, G., Carneiro, P.L., Lobo, R.N. & McManus, C., 2011. Molecular and pedigree analysis applied to conservation of animal genetic resources: the case of Brazilian Somali hair sheep. Trop Anim Heal. Prod. 43, 1449–1457 https://doi.org/10.1007/s11250-011-9873-6.

Palmarini, M., Hallwirth, C., York, D., Murgia, C., de Oliveira, T., Spencer, T. & Fan, H., 2002. Molecular Cloning and Functional Analysis of Three Type D Endogenous Retroviruses of Sheep Reveal a Different Cell Tropism from That of the Highly Related Exogenous Jaagsiekte Sheep Retrovirus. J. Virol. 74, 8065–8076.

Pannier, L., Gardner, G., Pearce, K., Mcdonagh, M., Ball, A., Jacob, R. & Pethick, D., 2014. Associations of sire estimated breeding values and objective meat quality measurements with sensory scores in Australian lamb. Meat Sci. 96, 1076–1087 https://doi.org/10.1016/j.meatsci.2013.07.037.

Peter, C., Bruford, M., Perez, T., Dalamitra, S., Hewitt, G. & Erhardt, G., 2007. Genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds. Anim. Genet. 38, 37–44 https://doi.org/10.1111/j.1365-2052.2007.01561.x.

Peters, F., Kotze, A., van der Bank, F., Soma, P. & Grobler, J., 2010. Genetic profile of the locally developed Meatmaster sheep breed in South Africa based on microsatellite analysis. Small Rumin. Res. 90, 101–108 https://doi.org/10.1016/j.smallrumres.2010.02.005.

Piedrafita, D., Raadsma, H., Gonzalez, J. & Meeusen, E., 2010. Increased production through parasite control: Can ancient breeds of sheep teach us new lessons? Trends Parasitol. 26, 568–573 https://doi.org/10.1016/j.pt.2010.08.002.

Pielberg, G., Golovko, A., Sundström, E., Curik, I., Lennartsson, J., Seltenhammer, M., Druml, T., Binns, M., Fitzsimmons, C., Lindgren, G., Sandberg, K., Baumung, R., Vetterlein, M., Strömberg, S., Grabherr, M., Wade, C., *et al.,* 2008. A

cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. Nat. Genet. 40, 1004–1009.

Pilling, D. & Hoffmann, I., 2011. Climate change and animal genetic resources for food and agriculture: State of knowledge, risks and opportunities. Rome.

Qwabe, S., van Marle-Köster, E. & Visser, C., 2013. Genetic diversity and population structure of the endangered Namaqua Afrikaner sheep. Trop Anim Heal. Prod. 45, 511–516 https://doi.org/10.1007/s11250-012-0250-x.

Ramayo-Caldas, Y., Castelló, A., Pena, R., Alves, E., Mercadé, A., Souza, C., Fernández, A., Perez-Enciso, M. & Folch, J., 2010. Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. BMC Genom. 11.

Redon, R., Ishikawa, S., Fitch, K., Feuk, L., Perry, G., Andrews, T., Fiegler, H., Shapero, M., Carson, A., Chen, W., Cho, E., Dallaire, S., Freeman, J., González, J., Gratacòs, M., Huang, J., Kalaitzopoulos, D., *et al.,* 2006. Global variation in copy number in the human genome. Nature. 444, 444–454.

Reed, D., Fox, C., Enders, L. & Kristensen, T., 2012. Inbreeding-stress interactions: Evolutionary and conservation consequences. Ann. N.Y. Acad. Sci. 1256, 33–48 https://doi.org/10.1111/j.1749-6632.2012.06548.x.

Richards, J., Semple, S. & Atkins, K., 2010. Practical application of precision sheep management. Reproduction. 357, 53–56.

Rohrer, G., 2004. An overview of genomics research and its impact on livestock reproduction. Reprod. Fertil. Dev. 16, 47–54 https://doi.org/10.1071/rd03077.

Rosa, A.J., Sardina, M., Mastrangelo, S., Tolone, M. & Portolano, B., 2013. Parentage verification of Valle del Belice dairy sheep using multiplex microsatellite panel. Small Rumin. Res. 113, 62–65 https://doi.org/10.1016/j.smallrumres.2013.03.021.

Rosati, A., Mousa, E., Van Vleck, L. & Young, L., 2002. Genetic parameters of reproductive traits in sheep. Small Rumin. Res. 43, 65–74 https://doi.org/10.1016/S0921-4488(01)00256-5.

Royo, L., Álvarez, I., Arranz, J., Fernández, I., Rodríguez, A., Pérez-Pardal, L. & Goyache, F., 2008. Differences in the expression of the ASIP gene are involved in the recessive black coat colour pattern in sheep: Evidence from the rare Xalda sheep breed. Anim. Genet. 39, 290–293 https://doi.org/10.1111/j.1365-2052.2008.01712.x.

Rupp, R., Mucha, S., Larroque, H., McEwan, J. & Conington, J., 2016. Genomic application in sheep and goat breeding. Anim. Front. 6, 39–44 https://doi.org/10.2527/af.2016-0006.

Rust, J. & Rust, T., 2013. Climate Change and Livestock Production, A review with emphasis on Africa. S. Afr. J. Anim. Sci. 43 https://doi.org/http://dx.doi.org/10.4314/sajas.v43i3.3.

Sabeti, P., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E., McCarroll, S., Gaudet, R., Schaffner, S. & Lander, E., 2007. Genome-wide detection and characterization of positive selection in human populations. Nature. 449, 913–918 https://doi.org/10.1038/nature06250.

Sadiq, G., Ishaq, M. & Saddozai, S., 2003. Estimation of Cost and Revenue and Analysis of the Different Factors Affecting Livestock Production. Sarhad J. Agric. 19, 579–584.

Safari, E., Fogarty, N. & Gilmour, A., 2005. A review of genetic parameter estimates for wool, growth, meat and reproduction traits in sheep. Livest. Prod. Sci. 92, 271–289 https://doi.org/10.1016/j.livprodsci.2004.09.003.

Safari, E., Fogarty, N., Gilmour, A., Atkins, K., Mortimer, S., Swan, A., Brien, F., Greeff, J. & Van Der Werf, J.H., 2007. Genetic correlations among and between wool, growth and reproduction traits in Merino sheep. J. Anim. Breed. Genet. 124, 65–72 https://doi.org/10.1111/j.1439-0388.2007.00641.x.

Sanders, K., Bennewitz, J. & Kalm, E., 2006. Wrong and missing sire information affects genetic gain in the Angeln dairy

cattle population. J. Dairy Sci. 89, 315–321 https://doi.org/10.3168/jds.S0022-0302(06)72096-3.

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

Schlötterer, C. & Wiehe, T., 1999. Microsatellites, a neutral marker to infer selective sweeps. Pages 238–248 in Microsatellites: evolution and application. Goldstein, D., Schlötterer, C., eds. Oxford University Press, Inc., New York.

Schoeman, S., 1998. Genetic and environmental factors influencing the quality of pelt traits in Karakul sheep. S. Afr. J. Anim. Sci. 28, 125–139.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Jendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y., Hicks, J., Spence, S., Lee, A., Puura, K., Lehtimäki, T., *et al.,* 2007. Strong Association of De Novo Copy Number Mutations with Autism. Science. 316, 445–449.

Sebat, J., Troge, J., Young, J., Walker, M., Chi, M., Navin, N., Lucito, R., Healy, J., Hicks, J., Ye, K., Reiner, A., Gilliam, T., Trask, B., Patterson, N., Zetterberg, A. & Wigler, M., 2004. Large-Scale Copy Number Polymorphism in the Human Genome. Science. 305, 525–528.

Selvaggi, M., Dario, C., Peretti, V., Ciotola, F., Carnicella, D. & Dario, M., 2010. Inbreeding depression in Leccese sheep. Small Rumin. Res. 89, 42–46 https://doi.org/10.1016/j.smallrumres.2009.12.005.

Shaw, C. & Lupski, J., 2004. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum. Mol. Genet. 13, R57–R64 https://doi.org/10.1093/hmg/ddh073.

Snyman, M., 2010. Namaqua Afrikaner Sheep.

Snyman, M., Erasmus, G., Van Wyk, J. & Olivier, J., 1998. Genetic and phenotypic correlations among production and reproduction traits in Afrino sheep. S. Afr. J. Anim. Sci. 28, 74–81 https://doi.org/10.4314/sajas.v28i2.44256.

Snyman, M., Olivier, J. & Cloete, J.A., 1993. Productive and reproductive performance of Namaqua Afrikaner sheep. Karoo Agric. 5, 21–24.

Soma, P., Kotze, A., Grobler, J. & van Wyk, J., 2012. South African sheep breeds: Population genetic structure and conservation implications. Small Rumin. Res. 103, 112–119 https://doi.org/10.1016/j.smallrumres.2011.09.041.

Soulé, M., 1980. Thresholds for Survival: Maintaining Fitness and Evolutionary Potential. Pages 151–169 in Conservation Biology: An evolutionary-ecological perspective. Soulé, M., Wilcox, B., eds. Sinauer Associates, Inc., Massachusetts.

Souza, C., MacDougall, C., MacDougall, C., Campbell, B., McNeilly, A. & Baird, D., 2001. The Booroola (FecB) phenotype is associated with a mutation in the bone morphogenetic receptor type 1 B (BMPR1B) gene. J. Endocrinol. 169, R1–R6 https://doi.org/10.1677/joe.0.169r001.

Stahlberger-Saitbekova, N., Schläpper, J., Dolf, G. & Gaillard, C., 2001. Genetic relationships in Swiss sheep breeds based on microsatellite analysis. J. Anim. Breed. Genet. 118, 379–387 https://doi.org/10.1046/j.1439-0388.2001.00312.x.

Stewart, M. & Miller, R., 1937. Resistance of sheep of different breeds to infestation by Ostertagia Circumcincta. J. Agric. Res. 56, 923–930.

Stothard, P., Choi, J., Basu, U., Sumner-Thomson, J., Meng, Y., Liao, X. & Moore, S., 2011. Whole genome resequencing of Black Angus and Holstein cattle for SNP and CNV discovery. BMC Genom. 12, 1–14.

Stranger, B., Forrest, M., Dunning, M., Ingle, C., Beazley, C., Thorne, N., Redon, R., Bird, C., Grassi, A., Lee, C., Tyler-smith, C., Carter, N., Scherer, S., Tavaré, S., Deloukas, P., Hurles, M. & Dermitzakis, E., 2007. Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. Science. 315, 848–853

https://doi.org/10.1126/science.1136678.

Sudmant, P., Mallick, S., Nelson, B., Hormozdiari, F., Krumm, N., Huddleston, J., Coe, B., Baker, C., Nordenfelt, S., Bamshad, M., Jorde, L., Posukh, O., Sahakyan, H., Watkins, W., Yepiskoposyan, L., Abdullah, M., *et al.,* 2015. Global diversity, population stratification, and selection of human copy-number variation. Science. 349, 1181.

Swan, A., Johnston, D., Brown, D., Tier, B. & Graser, H., 2012. Integration of genomic information into beef cattle and sheep genetic evaluations in Australia. Anim. Prod. Sci. 52, 126–132 https://doi.org/10.1071/AN11117.

Syvanen, A., 2005. Toward genome-wide snp genotyping. Nat. Genet. 37, S5–S10 https://doi.org/10.1038/ng1558.

Thapar, A. & Cooper, M., 2013. Copy number variation: What is it and what has it told us about child psychiatric disorders? J. Am. Acad. Child Adolesc. Psychiatry. 52, 772–774.

Thomas, D., Twyman, C., Osbahr, H. & Hewitson, B., 2007. Adaptation to climate change and variability: farmer responses to intra-seasonal precipitation trends in South Africa. Clim. Chang. 83, 301–322 https://doi.org/10.1007/s10584-006-9205-4.

Tibbo, M., 2006. Productivity and Health of Indigenous sheep breeds and crossbreds in the central Ethiopian highlands.

Tribout, T., Larzul, C. & Phocas, F., 2013. Economic aspects of implementing genomic evaluations in a pig sire line breeding scheme. Genet. Sel. Evol. 45, 1–16.

UNESCO Institute for Statistics., 2018. One in Five Children , Adolescents and Youth is Out of School.

Valera, M., Molina, A., Gutiérrez, J., Gómez, J. & Goyache, F., 2005. Pedigree analysis in the Andalusian horse: Population structure, genetic variability and influence of the Carthusian strain. Livest. Prod. Sci. 95, 57–66 https://doi.org/10.1016/j.livprodsci.2004.12.004.

Viginier, B., Dolmazon, C., Lantier, I., Lantier, F., Archer, F., Leroux, C. & Terzian, C., 2012. Copy number variation and differential expression of a protective endogenous retrovirus in sheep. PLoS One. 7, e41965 https://doi.org/10.1371/journal.pone.0041965.

Vitti, J., Grossman, S. & Sabeti, P., 2013. Detecting Natural Selection in Genomic Data. Annu. Rev. Genet. 47, 97–120 https://doi.org/10.1146/annurev-genet-111212-133526.

Vostry, L., Milerski, M., Schmidova, J. & Vostra-Vydrova, H., 2018. Genetic diversity and effect of inbreeding on litter size of the Romanov sheep. Small Rumin. Res. 168, 25–31 https://doi.org/10.1016/j.smallrumres.2018.09.004.

Vozzi, P., Marcondes, C., Bezerra, L.A. & Lôbo, R., 2007. Pedigree analyses in the Breeding Program for Nellore Cattle. Genet. Mol. Res. 6, 1044–1050.

Wang, J., Jiang, J., Fu, W., Jiang, L., Ding, X., Liu, J. & Zhang, Q., 2012. A genome-wide detection of copy number variations using SNP genotyping arrays in swine. BMC Genom. 13, 1–10.

Warner, R., Greenwood, P., Pethick, D. & Ferguson, D., 2010. Genetic and environmental effects on meat quality. Meat Sci. 86, 171–183 https://doi.org/10.1016/j.meatsci.2010.04.042.

Waters-Bayer, A. & Bayer, W., 1992. The role of livestock in rural economy. Nomad. People. 31, 3–18.

Watson, T., Baker, R. & Harvey, T., 1986. Genetic variation in resistance or tolerance to internal nematode parasites in strains of sheep at Rotomahana. Pages 23–26 in Proceedings of the New Zealand Society of Animal Production.

Weinman, L., Solomon, J. & Rubenstein, D., 2015. A comparison of single nucleotide polymorphism and microsatellite markers for analysis of parentage and kinship in a cooperatively breeding bird. Mol. Ecol. Resour. 15, 502–511 https://doi.org/10.1111/1755-0998.12330.

Van Der Werf, J.H., 2009. Potential Benefit of Genomic Selection in Sheep. Proc. Assoc. Advmt. Anim. Breed. Genet. 18, 38–41 https://doi.org/10.1111/j.1365-2478.2006.00545.x.

van der Werf, J.H., Banks, R., Clark, S., Lee, S., Daetwyler, H., Hayes, B. & Swan, A., 2014. Genomic Selection in Sheep Breeding Programs. In Proceedings of the 10th World Congress of Genetics Applied to Livestock Production Genomic. Vancouver.

Wiggins, S. & Keats, S., 2013. Smallholder agriculture's contribution to better nutrition.

Wilson, T., Wu, X., Juengel, J., Ross, I., Lumsden, J., Lord, E., Dodds, K., Walling, G., McEwan, J., O'Connell, A., McNatty, K. & Montgomery, G., 2001. Highly Prolific Booroola Sheep Have a Mutation in the Intracellular Kinase Domain of Bone Morphogenetic Protein IB Receptor (ALK-6) That Is Expressed in Both Oocytes and Granulosa Cells1. Biol. Reprod. Suppl. 64, 1225–1235 https://doi.org/10.1095/biolreprod64.4.1225.

Winchester, L., Yau, C. & Ragoussis, J., 2009. Comparing CNV detection methods for SNP arrays. Br. Funct Genomic Proteomic. 8, 353–366 https://doi.org/10.1093/bfgp/elp017.

Wolc, A., Habier, D., Fernando, R., Garrick, D., Dekkers, J.C., Arango, J., Settar, P., Fulton, J., O'Sullivan, N. & Preisinger, R., 2011. Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. Genet. Sel. Evol. 43, 1–8.

Wolfová, M., Wolf, J. & Milerski, M., 2009. Calculating economic values for growth and functional traits in non-dairy sheep. J. Anim. Breed. Genet. 126, 480–491 https://doi.org/10.1111/j.1439-0388.2009.00815.x.

Wright, S., 1984. Evolution and the genetics of populations, volume 3: experimental results and evolutionary deductions. University of Chicago press.

Wright, D., Boije, H., Meadows, J.R., Bed'hom, B., Gourichon, D., Vieaud, A., Tixier-Boichard, M., Rubin, C., Imsland, F., Hallböök, F. & Andersson, L., 2009. Copy Number Variation in Intron 1 of SOX5 causes the Pea-comb Phenotype in Chickens. PLoS Genet. 5, e1000512 https://doi.org/10.1371/journal.pgen.1000512.

van Wyk, J., Erasmus, G. & Konstantinov, K., 1993. Inbreeding in the Elsenburg Dormer sheep stud. S. Afr. J. Anim. Sci. 23, 77–80.

van Wyk, J., Fair, M. & Cloete, S.W., 2009. Case study: The effect of inbreeding on the production and reproduction traits in the Elsenburg Dormer sheep stud. Livest. Sci. 120, 218–224 https://doi.org/10.1016/j.livsci.2006.10.005.

Xu, L., Cole, J., Bickhart, D., Hou, Y., Song, J., VanRaden, P., Sonstegard, T., Van Tassell, C. & Liu, G., 2014a. Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. BMC Genom. 15, 1–10.

Xu, Y., Shi, T., Cai, H., Zhou, Y., Lan, X., Zhang, C., Lei, C., Qi, X. & Chen, H., 2014b. Associations of MYH3 gene copy number variations with transcriptional expression and growth traits in Chinese cattle. Gene. 535, 106–111 https://doi.org/10.1016/j.gene.2013.11.057.

Yang, J., Li, W., Lv, F., He, S., Tian, S., Peng, W., Sun, Y., Zhao, Y., Tu, X., Zhang, M., Xie, X., Wang, Y., Li, J., Liu, Y., Shen, Z., Wang, F., Liu, G., Lu, H., Kantanen, J., Han, J., Li, M. & Liu, M., 2016. Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. Mol. Biol. Evol. 33, 2576–2592 https://doi.org/10.1093/molbev/msw129.

Yang, L., Xu, L., Zhou, Y., Liu, M., Wang, L., Kijas, J., Zhang, H., Li, L. & Liu, G., 2018. Diversity of copy number variation in a worldwide population of sheep. Genomics. 110, 143–148 https://doi.org/10.1016/j.ygeno.2017.09.005.

York, D., Vigne, R., Verwoerd, D. & Querat, G., 1992. Nucleotide Sequence of the Jaagsiekte Retrovirus, an Exogenous and Endogenous Type D and B Retrovirus of Sheep and Goats. J. Virol. 66, 4930–9.

Zane, L., Bargelloni, L. & Patarnello, T., 2002. Strategies for microsatellite isolation: A review. Mol. Ecol. 11, 1–16 https://doi.org/10.1046/j.0962-1083.2001.01418.x.

Zhang, Q., Guldbrandtsen, B., Bosse, M., Lund, M. & Sahana, G., 2015a. Runs of homozygosity and distribution of functional variants in the cattle genome. BMC Genom. 16, 1–16 https://doi.org/10.1186/s12864-015-1715-x.

Zhang, L., Liu, J., Zhao, F., Ren, H., Xu, L., Lu, J., Zhang, S., Zhang, X., Wei, C., Lu, G., Zheng, Y. & Du, L., 2013. Genome-Wide Association Studies for Growth and Meat Production Traits in Sheep. PLoS One. 8, e66569 https://doi.org/10.1371/journal.pone.0066569.

Zhang, Q., Ma, Y., Wang, X., Zhang, Y. & Zhao, X., 2015b. Identification of copy number variations in Qinchuan cattle using BovineHD Genotyping Beadchip array. Mol Genet Genomics. 290, 319–327 https://doi.org/10.1007/s00438-014-0923-4.

Zhou, Y., Utsunomiya, Y., Xu, L., Hamidi, E., Bickhart, D., Alexandre, P., Rosen, B., Schroeder, S., Carvalheiro, R., Neves, H.H., Sonstegard, T., van Tassell, C., Bento, J., Ferraz, S., Fukumasu, H., Garcia, J. & Liu, G., 2016. Genome-wide CNV analysis reveals variants associated with growth traits in Bos indicus. BMC Genom. 17, 1–9.

Zishiri, O., Cloete, S.W., Olivier, J. & Dzama, K., 2013a. Genetic parameter estimates for subjectively assessed and objectively measured traits in South African Dorper sheep. Small Rumin. Res. 109, 84–93 https://doi.org/10.1016/j.smallrumres.2012.07.014.

Zishiri, O., Cloete, S.W., Olivier, J. & Dzama, K., 2013b. Genetic parameters for growth, reproduction and fitness traits in the South African Dorper sheep breed. Small Rumin. Res. 112, 39–48 https://doi.org/10.1016/j.smallrumres.2013.01.004.

Zöllner, S. & Teslovich, T., 2009. Using GWAS Data to Identify Copy Number Variants Contributing to Common Complex Diseases. Stat. Sci. 24, 530–546 https://doi.org/10.1214/09-sts304.

CHAPTER 3
Pedigree analysis and IBD relationships of a smallholder sheep flock in Beaufort-West using an OvineSNP50 BeadChip

3.1 ABSTRACT

The aim of this study was to construct a pedigree diagram of 48 sheep from a smallholder farm in Beaufort-West to assist with breeding and management decisions. Blood samples were genotyped with an OvineSNP50 BeadChip. Data underwent quality control in *Plink* which resulted in 44767 variants and 47 individuals that passed the filters that could be used for pedigree construction using the *SEQUOIA* package in *RStudio*. Eleven dams were assigned and four sires were assigned as parents to individuals. Ten dams were assigned as possible relatives with likely parent-offspring (PO) relationships, and nine sires were assigned as possible relatives with likely parent-offspring (PO) relationships. The sample population included two rams that were the probable sires. For the four individuals that had sires assigned, it was found that ram 9943 was assigned as the sire of these four individuals, while ram 16008 was only a likely relative to a number of individuals but was not assigned as a parent.

Pairwise relationships were also determined for the individuals in *Plink* using the *--genome* function. Familial relationships were categorized as parent-offspring (PO), full-sib (FS), half-sib (HS), grandparent-grand-offspring (GG) and Full aunt/ uncle-niece/ nephew (FA) based on the IBD proportion (PI_HAT value). The genetic information obtained from the pedigree construction and pairwise relationships will, therefore, assist the farmer to make more informed breeding and management decisions with regards to which individuals to select as the parents of the next generation. Understanding the relationships between individuals will provide the farmer with information that could assist in limiting unintended inbreeding in the flock. Pedigree information will also indicate which individuals contributed to the next generation and could thus assist with selection of next generation individuals as well as the removal of animals that did not contribute genetically to the flock.

Keywords: pedigree, SEQUOIA, sheep, smallholder, Dorper

3.2 INTRODUCTION

Lamb and mutton are estimated to have the second highest increase in consumption by the year 2050 (Williams, 2009). Due to this, new innovative and efficient ways will have to be developed and carried out to increase production outputs. A basic way to increase production would be to ensure a complete understanding of the animal genetic resources that are being used and is available in order to ensure that individuals are able to reach their full genetic potential (Anderson, 2003). Accurate and complete pedigree and genetic records is a cost-effective and easy way to gather information on the genetic resources available that could possibly assist to improve genetic gains through improved breeding strategies.

Pedigree information and accurate record keeping is often the only tool smallholder farmers have available to them to make good breeding and management decisions, since smallholder farmers generally cannot afford genetic tests or genotyping to be done on their sheep. Without pedigree or production information good breeding decisions cannot be made since there is no way to separate the above average individuals from the average individuals (Richards *et al.*, 2010). Genetic advances will only be possible if individuals with high genetic potential are selected to be bred. If all animals are used in a breeding system it could result in low to no genetic gain.

Smallholder farmers with financial burdens that have smaller flocks may be placed in situations where they may have to sell potential breeding stock in order to meet certain financial obligations (Motiang & Webb, 2016). The animals that would likely be sold first would be the individuals that are market ready or are the closest to being market ready and would thus generate the highest profit. The individuals that would be market ready before the other individuals may have an above average genetic potential compared to the rest of the flock, and should be used as the next generation breeding stock, however, they are often sold first because of their above average size and growth. When these animals are sold, the farmers are often practising undesirable selection. When farmers sell their fastest growing or heavier animals, they are essentially taking away the above average genetic resources available to them. Farmers should select the individuals that are in line with the breeding objectives of the farm and would thus have the best genetics to reach these production goals (Richards *et al.*, 2010). These animals should be selected based on the desired products, their response to the environment as well as the farming systems being employed by the farmer. Faster growing and heavier animals may be desired in one farming system that has an optimum environment, while a different farmer may decide to focus on adaptability due to an unfavourable environment. Certain farmers select for size and growth, while others select for adaptability or reproduction, depending on the farmers breeding objectives. These breeding decisions cannot be made without correct and accurate records.

Pedigree data coupled with molecular markers can be useful with regards to selection and genetic management within a population at a low cost (Paiva *et al.*, 2011). Selection has improved with the development of genomic selection. Marker-assisted selection is a form of genomic selection where markers, such as SNPs, covering the whole genome are used so that quantitative trait loci (QTL) are in linkage disequilibrium with at least one marker (Goddard & Hayes, 2007). Genomic selection could assist with the identification of animals that possess a higher than average genetic potential. Using genomic selection, individuals can be selected at a younger age, because progeny testing is no longer necessary and traits that may be late maturing can be selected for at a young age (Rohrer, 2004), thus decreasing the generation interval. Genomic selection also assists with selecting for traits that may be sex-linked. Genomic selection allows farmers to select for economically beneficial traits such as disease or parasite resistance (Beh & Maddox, 1996), production traits or even reproduction traits such as the Booroola gene (Montgomery *et al.*, 1994; Mulsant *et al.*, 2001).

SNPs also provide a way to construct the pedigree of a flock when there is a lack of pedigree records and is based on parental and sibship reconstruction using likelihood-based methods (Huisman, 2017). Full probability parentage analysis and sibship reconstruction has advanced to the degree that they are seen as an appropriate method and solution to some parentage analysis problems (Jones *et al.*, 2010). *Figure 3.1* describes the probability of a parent-offspring relationship being excluded at a locus with three equifrequent codominant alleles.

Huisman (2017) presents an algorithm that compares likelihoods for seven different relationship alternatives as seen in *Table 3.1*, including their inbred derivatives, speeded up by steps to exclude unlikely relatives.

Table 3.1 *Genealogical relationships considered by* Huisman, (2017) *and their mean pairwise relatedness r in the absence of inbreeding or additional relationships between a pair of individuals*

	Relationship	**Code**	**Mean *r***
H_1	Parent-offspring	PO	½
H_2	Full-siblings	FS	½
H_3	Half-siblings	HS	¼
	Maternal siblings (full or half)	MS	½ or ¼
	Paternal siblings (full or half)	PS	½ or ¼
H_4	Grandparent-grand-offspring	GG	¼
H_5	Full aunt/ uncle-niece/ nephew	FA	¼
H_{6a}	Half aunt/ uncle-niece/ nephew	HA	1/8
H_{6b}	Great-grandparent-great-grand-offspring	GGG	1/8
H_{6c}	Full cousins	CC	1/8
H_0	Unrelated	U	0

In the maximum likelihood approach, information about the distribution of sib family sizes in the sample (Wang, 2004), age, demographic data and genetic data may be used to limit the possible relatives, since an individual cannot be the parent of an individual unless they have some allele in common at every locus (Thompson & Meager, 1987). Thompson & Meager (1987), states that sib log-likelihood is a more useful statistic than parent log-likelihood in order to detect parents, however, a multivariate approach including other log-likelihood statistics may prove to be more effective at detecting parents from genetic data. This demonstrates the possible effectiveness of *SEQUOIA*, as it uses the log-likelihood statistics of seven relationships to determine the most likely parent of individuals (Huisman, 2019a).

One solution to ensure that one indeed maximises the total likelihood is to calculate for each set of candidate relatives the likelihoods under many possible alternative relationships (Huisman, 2017).

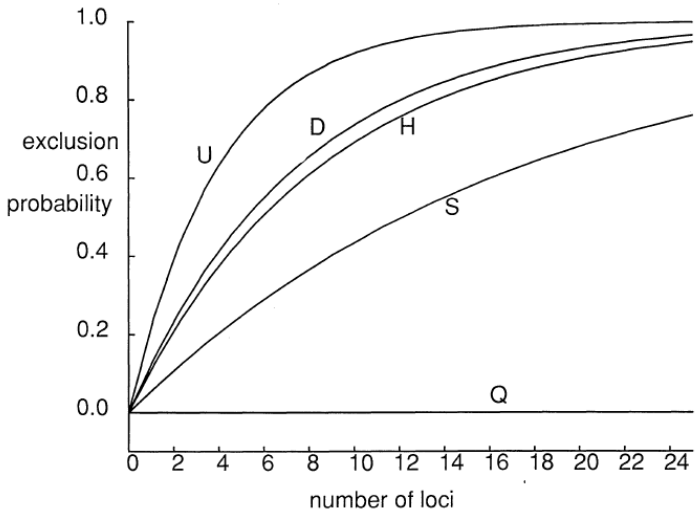

Figure 3.1 *Probabilities of exclusion of a parent-offspring relationship (Q) by loci with three equifrequent codominant alleles, for each of five true pairwise relationships, (U) unrelated pair, (D) double first cousins, (H) pair of half sibs, (S) pair of full sibs* (Thompson & Meager, 1987)

SEQUOIA assigns parents based on log-likelihood ratios. The log-likelihood ratio is the likelihood of an individual being the parent relative to the likelihood of another individual being the parent (Marshall *et al.*, 1998).

For each pair of possible relatives, the likelihoods are determined of them being parent-offspring (PO), full siblings (FS), half siblings (HS), grandparent-grand-offspring (GG), full avuncular (niece/nephew – aunt/uncle) (FA), half avuncular/great-grandparental/cousins (HA), or unrelated (U). Assignments are made if the log-likelihood ratio (LLR) between the focal relationship and the most likely alternative exceed the *Tassign* threshold (Huisman, 2017). Huisman (2019a) defines *Tassign* as the minimum LLR required to accept the proposed relationship relative to the next most likely relationship.

Additionally, *SEQUOIA* determines the opposite homozygote (OH) of each assigned parent. Two individuals have opposing homozygous loci when one individual is homozygous for one allele and the other individual is homozygous for the other allele (Calus *et al.,* 2011). Offspring receive 50% of their genetic material from one parent and 50% from the other parent, meaning that an individual has 50% of each parent's genetic material. It can thus be assumed that the genes that were passed from one parent to the offspring should be similar and should, therefore, be identical-by-descent (IBD). So, offspring and parents should have fewer opposing homozygotes than would be found between two individuals that are unrelated. It stands to reason that a parent-offspring pair would have more alleles in common than a pair that are unrelated and the parent-offspring pair would be expected to have fewer opposing homozygotes compared to an unrelated pair of individuals.

According to Hayes (2011), when using SNP data, an individual can be excluded as a possible parent if, at a locus, the individual and the prospective parent are both homozygous but for different alleles ("opposing homozygotes"). However, with this approach genotyping errors could possibly result in true parents having a small number of opposing homozygotes with their progeny (Hayes, 2011), however, this concern can be

overcome if the genotyping error distribution is known. Strucken *et al.* (2016) suggested that composite sheep breeds need approximately 220 markers for correct parental assignment and approximately 4400 SNPs are needed to eliminate false-positive parental assignments.

The International Society for Animal Genetics (ISAG) recommends a minimum of 100 SNP markers for parentage testing, as well as a maximum of 1% acceptable genotyping mismatches with regards to true parent-offspring relations where only one parents genotype is available (Strucken *et al.*, 2016). In a study conducted by Montgomery *et al.* (2005), the estimated genotyping error was approximately 0.05%, while Saunders *et al.* (2007) calculated an error rate of approximately 0.1%.

In addition to using the *SEQUOIA* package and SNPs to determine pedigree other SNP methods are also available to determine the pedigree of a flock such as estimates of pairwise IBD. Related individuals inherit alleles through common ancestors and the regions that are in common are generally identical by descent (IBD) (Powell *et al.,* 2010; Clark *et al.,* 2013). For any two individuals, identity-by-state (IBS) can be found at a given locus with three possible results; the two individuals can have two different alleles (IBS0) or they could share one common allele (IBS1) or they could have two (IBS2) alleles in common, individuals that share one or two alleles at a certain locus may have inherited the alleles from a common ancestor, meaning that the alleles are identical-by-descent (Stevens *et al.*, 2011). There is also the possibility that two individuals could have alleles in common due to the fact that it might be commonly found in the population (IBS) due to possible previous inbreeding which would increase the frequency of a certain allele within the population without the two individuals having to be directly related. Estimates of pairwise IBD for individuals is useful to identify individuals that look more similar to each other than would be expected by chance in a random sample (Purcell *et al.*, 2007).

The aim of this study was to construct the pedigree of the smallholder sheep flock in Beaufort-West and to assign familial relationships, that would assist in the development of an effective breeding programme that could result in an increase in production outputs and a more complete knowledge of the pedigree and relationships within the sheep population on the smallholder farm.

3.3 MATERIALS AND METHODS

3.3.1 DESCRIPTION OF STUDY LOCATION

According to the Köppen-Geiger climate classification, Beaufort-West falls under the arid cold desert (BWk) climate classification, which is described as arid, desert and cold (Engelbrecht & Engelbrecht, 2016), with most of the arid zone receiving less than 500mm rainfall annually.

The sheep are kept on natural grazing with some supplements provided due to the fact that the area is currently experiencing a drought. Additional supplements were provided as the farmer deemed necessary.

The sheep were selected based on their ability to adapt and thrive in the harsh environment that they live in. The Karoo area has sparse vegetation, resulting in the sheep having to walk long distances to find feed and water. Due to the Karoo climate, sheep need to be able to handle the heat and be able to regulate their

body temperature efficiently to ensure that other physiological and metabolic processes are not compromised. All these factors are taken into account by the farmer when choosing which animals to keep for breeding and which to sell. These factors all contribute to the selection criteria for adaptation.

3.3.2 RESEARCH POPULATION AND SAMPLING METHOD

Blood samples from 48 individuals were collected by a certified animal technician on the smallholder farm. Samples were collected from the jugular vein into EDTA vacutainer tubes from all the animals. Blood samples were collected from two rams > 8 years old, 24 ewes ranging in age from 2 years to > 8 years, 19 ewe lambs and four ram lambs. Descriptive statistics of the flock can be found in the *Supplementary Material, 5 Descriptive Statistics, Table 1.5.1.* Blood samples were transported in a cooler box back to Stellenbosch, where it was immediately placed in the freezer at -4°C until it was removed for DNA extraction.

Ethical clearance was obtained from the Departmental Ethics Committee for Research on Animals (DECRA), approval numbers of R12/53 for smallholder flocks from the Western Cape Department of Agriculture.

3.3.3 DNA EXTRACTION AND GENOTYPING

DNA was extracted using a Sbeadex livestock kit; catalogue number 44701 and 44702. Concentration and quality of the extracted DNA was recorded using a Nanodrop Machine from the Central Analytical Facility (CAF) at University of Stellenbosch. DNA concentrations and quality are recorded in the *Supplementary Material, Table 1.1*.

Extracted DNA samples were run using FlashGel™ Starter Kit to confirm the presence of DNA. The DNA electrophoresis process is recorded in the *Supplementary Material, 3) Lonzo Gel, with Figure 1.3.1 – Figure 1.3.7* illustrating the DNA runs.

Extracted DNA was genotyped at the ARC Biotechnology Platform in Pretoria, with the assistance of qualified personnel.

The genotyping process took three days and included a number of steps: on the first day, double stranded DNA samples were quantified using the Qubit protocol and are recorded in the *Supplementary material, 2) Qubit Protocol,* with the Qubit results recorded in *Table 1.1* under *1) DNA extraction*. The DNA samples were then denatured and neutralized for the amplification process. The DNA was incubated at 37°C for 20-24 hours for the amplification to take place. On day two the amplified DNA was fragmented through a controlled enzymatic process, after which the DNA samples were precipitated and then resuspended. The resuspended DNA samples were loaded onto BeadChips which were then placed in Hyb Chambers. It was then incubated in the Illumina Hybridization oven for 16-24 hours at 48°C. On day three the BeadChips were washed in preparation of the XStain HD BeachChip process. Labelled nucleotides were incorporated into the extended primers in an extension reaction, after which the labelled extension primers underwent a multi-layer staining process. The iScan reader used a laser to excite the fluor of the single-base extension product on the

BeadChips and the light emissions were then recorded as high-resolution images and then imported into the Illumina Genome Studio Genotyping Module (Illumina, 2008).

3.3.4 ANALYSIS

Quality Control and Pedigree Construction

Quality control and pruning was conducted in *Plink* using the functions --mind, --maf and --geno (Purcell *et al.*, 2007), with threshold values of mind = 0.1, maf = 0.05 and geno = 0.1. Quality control was performed on 48 sheep, and *Plink* removed sample 136 due to missing genotype data. The genotyping rate for the remaining samples was 0.9495. During pruning 3279 variants were removed due to missing genotype data (--geno) and 6195 variants were removed due to minor allele threshold (--maf). A total of 44767 variants and 47 sheep passed the filters and quality control. After pruning and quality control 47 samples were used to construct the pedigree in RStudio (R, 2011) using the *SEQUOIA* package, which utilizes likelihood-based methods for parental and sibship reconstruction (Huisman, 2019b). Sample 7275 was rejected by *SEQUOIA* during pedigree construction in RStudio, resulting in 46 samples being included for pedigree construction.

Quality Control and Pairwise IBD Relationship

Quality control and pruning was conducted in *Plink* using the functions --mind, --maf, --geno and --hwe (Purcell *et al.*, 2007), with threshold values of mind = 0.1, maf = 0.05, geno = 0.1 and hwe = 0.01. Quality control was carried out on the 48 individuals, one individual (136) was removed due to missing genotype data (--mind). Due to missing genotype data (--geno), 3279 variants were removed, 242 variants were removed due to non-conformity toHardy-Weinberg exact test (--hwe) which could be an indication of genotyping errors (Weir, 2013), and 6183 variants were removed due to minor allele thresholds (--maf). After pruning a total of 44537 variants and 47 sheep passed quality control and the set thresholds to be further analysed to determine pairwise IBD relationships.

Plink (Purcell *et al.*, 2007) was used to determine the pairwise IBD relationships between the individuals included in the study using the *--genome* function. Of the 44537 variants used to determine IBD relationships, 1204 variants on non-autosomes were excluded from the IBD calculation.

3.4 RESULTS AND DISCUSSION

3.4.1 PEDIGREE DIAGRAMS

Pedigree diagrams were constructed based on the sires and dams that were assigned to the individuals using the *SEQUOIA* package in RStudio. *SEQUOIA* assigned dams to 11 individuals and sires to four individuals. Other individuals were identified as probable relatives, but were not assigned as the dam or sire of the individual using the *-GetMaybeRel* function in *SEQUOIA.* The *-MaybeTrio* function provided non-assigned possible parent-parent-offspring trios to the individuals studied (Huisman, 2019a). A basic probable pedigree diagram based on the information provided by the SEQUOIA package can be found in the *Supplementary Material, 4) Genotyping, Table 1.4.1*.

Legend:

= Assigned male individual
= Assigned female individual
= MaybeRel male individual
= MaybeRel female individual
= MaybeTrio male individual
= MaybeTrio female individual

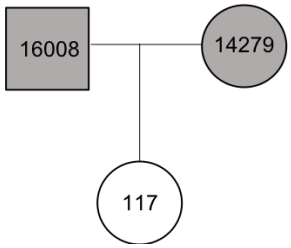

Figure 3.2 *Pedigree diagram of individual 117*

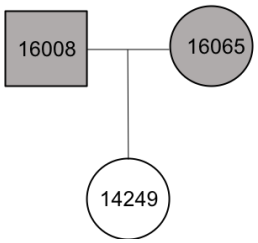

Figure 3.3 *Pedigree diagram of individual 14249*

Figure 3.2 depicts the pedigree diagram of individual 117. Ram 16008 and ewe 14279 were assigned as likely relatives to 117. According to the *GetMaybeRel* data frame, ram 16008 and ewe 14279 likely have a parent-offspring (PO) relationship with 117. Ram 16008 had a log-likelihood ratio (LLR) of 460.20 as the probable sire and 14279 had an LLR of 480.48 as the probable dam. Ewe lamb 117 and her non-assigned dam had 45 opposing homozygotes while ewe lamb 117 and her non-assigned sire had 33 opposing homozygotes meaning that the dam was homozygous for one allele, while the ewe lamb was homozygous for the other allele and the same for the non-assigned sire (Calus *et al.,* 2011). This shows that the ewe lamb was homozygous for 45 alleles while the non-assigned dam was homozygous for the exact opposite alleles.

Figure 3.3 describes the pedigree diagram of 14249. Ram 16008 and ewe 16065 were assigned as likely relatives to 14249. According to the *GetMaybeRel* data frame, ram 16008 and ewe 16065 likely have a parent-offspring (PO) relationship with 14249. Ram 16008 had an LLR of 544.49 as the probable sire and 16065 had an LLR of 509.89 as the probable dam. Ewe lamb 14249 and her non-assigned dam had 39 opposing homozygotes while 14249 and her non-assigned sire had 54 opposing homozygotes.

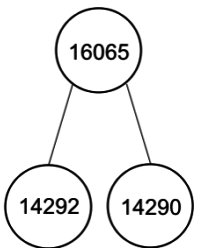

Figure 3.4 *Pedigree diagram of 16065 offspring*

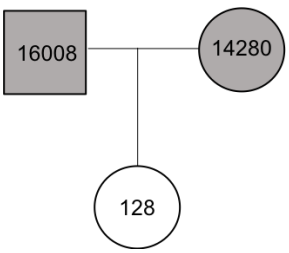

Figure 3.5 *Pedigree diagram of individual 128*

Figure 3.4 depicts the pedigree diagram of ewe 16065s offspring. Ewe 16065 was assigned as the dam of ewe 14292 and ewe 14290. Ewe 16065 had an LLR of 191.14 as the assigned dam for ewe 14292, and an LLR of 432.68 as the assigned dam of ewe 14290. Ewe 16065 and ewe 14292 had 61 opposing homozygotes while 16065 and 14290 had 42 opposing homozygotes.

Figure 3.5 illustrates the pedigree of ewe lamb 128. Ram 16008 and ewe 14280 were assigned as likely relatives to 128. According to the *GetMaybeRel* data frame, ram 16008 and ewe 14280 have a likely parent-offspring (PO) relationship to 128. Ram 16008 had an LLR of 457.66 as the probable sire and 14280 had an LLR of 535.61 as the probable dam. Ewe lamb 128 and her non-assigned dam had 48 opposing homozygotes while 128 and her non-assigned sire had 34 opposing homozygotes.

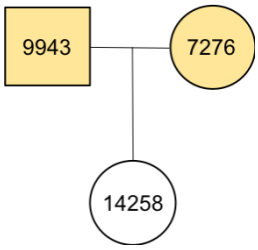

Figure 3.6 *Pedigree diagram of individual 14258*

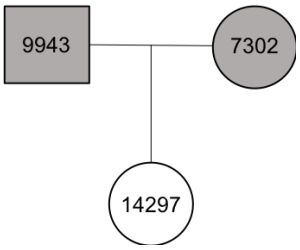

Figure 3.7 *Pedigree diagram of individual 14297*

Figure 3.6 describes the pedigree diagram of ewe lamb 14258. Ram 9943 and ewe 7276 were assigned as likely relatives to 14258. According to the *MaybeTrio* data frame, ram 9943 was assigned as parent2 and ewe 72276 was assigned as parent1 to individual 14258. Ram 9943 had an LLR of 485.27 as parent1 and ewe 7276 had an LLR of 621.27 as parent1. Ram 9943 and ewe 7276 also had a combined LLRpair of 1364.45. Ewe lamb 14258 and parent1 (7276) had 42 opposing homozygotes, while 14258 and parent2 (9943) had 33 opposing homozygotes. *MaybeTrio* also reported the number of Mendelian errors (ME) between the offspring and the parent pair as 76.

Figure 3.7 depicts the pedigree diagram of ewe lamb 14297. Ram 9943 and ewe 7302 were assigned as likely relatives to 14297. According to the *GetMaybeRel* data frame, ram 9943 and ewe 7302 have a likely parent-offspring (PO) relationship to 14297. Ram 9943 had an LLR of 465.34 as the probable sire and 7302 had an LLR of 708.17 as the probable dam. Ewe lamb 14297 and her non-assigned dam had 35 opposing homozygotes while 14297 and her non-assigned sire had 30 opposing homozygotes.

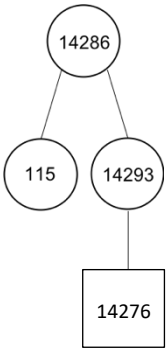

Figure 3.8 *Pedigree diagram of individuals 115, 14293 and 14276*

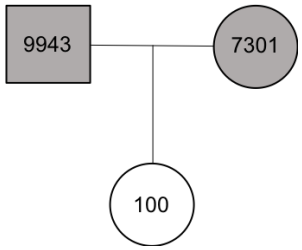

Figure 3.9 *Pedigree diagram of individual 100*

Figure 3.8 illustrates the pedigree diagram of the offspring of ewe 14286. Ewe 14286 was assigned as the dam of ewe lamb 115 and ewe 14293 with an LLR of 530.02 and 435.90 respectively. Ewe 14286 and ewe lamb 115 had 64 opposing homozygotes, while ewe 14286 and ewe 14293 had 52 opposing homozygotes. Ewe 14293 was assigned as the dam of ram lamb 14276 and had an LLR of 702.05 with 36 opposing homozygotes. We can thus assume that 14286 would be the grand-dam of ram lamb 14276.

Figure 3.9 depicts the pedigree diagram of individual 100. Ram 9943 and ewe 7301 were assigned as likely relatives to individual 100. According to the *GetMaybeRel* data frame, ram 9943 and ewe 7301 have a likely parent-offspring (PO) relationship to 100. Ram 9943 had an LLR of 453.34 as the probable sire and 7301 had a LLR of 646.07 as the probable dam. Ewe lamb 100 and her non-assigned probable dam had 32 opposing homozygotes while 100 and her non-assigned probable sire had 37 opposing homozygotes.

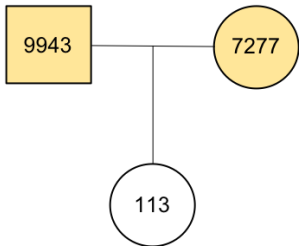

Figure 3.10 *Pedigree diagram of individual 113*

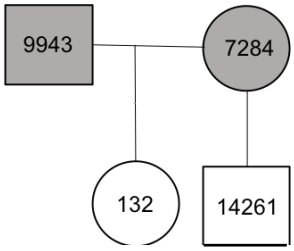

Figure 3.11 *Pedigree diagram of individual 132 and 14261*

Figure 3.10 describes the pedigree diagram of individual 113. Ram 9943 and ewe 7277 were assigned as likely relatives to 113. According to the *MaybeTrio* data frame, ram 9943 was assigned as parent2 and ewe 7277 was assigned as parent1 to individual 113. Ram 9943 had an LLR of 459.55 as parent2 and ewe 7277 had a LLR of 617.20 as parent1. Ram 9943 and ewe 7277 also had a combined LLRpair of 1440.37. Ewe lamb 113 and parent1 (7277) had 36 opposing homozygotes, while 113 and parent2 (9943) had 25 opposing homozygotes. *MaybeTrio* also reported the number of Mendelian errors (ME) between the offspring and the parent pair as 62.

Figure 3.11 illustrates the pedigree diagram of individuals 132 and 14261. Ram 9943 and ewe 7284 were assigned as likely relatives to individual 132. According to the *GetMaybeRel* data frame, ram 9943 and ewe 7284 have a likely parent-offspring (PO) relationship with ewe 132. Ram 9943 had an LLR of 424.01 as the probable sire and 7284 had a LLR of 669.36 as the probable dam. Ewe 132 and her non-assigned probable dam had 37 opposing homozygotes while 132 and her non-assigned probable sire had 25 opposing homozygotes. Ewe 7284 was also assigned as a likely relative to ram lamb 14261. According to the *GetMaybeRel* data frame, ewe 7284 and ram lamb 14261 likely have a parent-offspring (PO) relationship. Ewe 7284 had an LLR of 652.75 as the probable dam. Ram lamb 14261 and ewe 7284 had 33 opposing homozygotes.

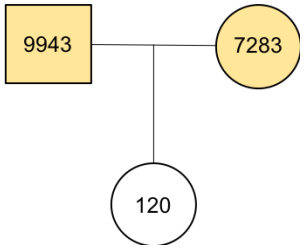

Figure 3.12 *Pedigree diagram of individual 120*

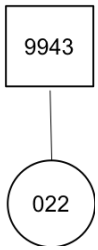

Figure 3.13 *Pedigree diagram of individual 022*

Figure 3.12 describes the pedigree diagram of individual 120. Ram 9943 and ewe 7283 were assigned as likely relatives to 120. According to the *MaybeTrio* data frame, ram 9943 was assigned as parent2 and ewe 7283 was assigned as parent1 to individual 120. Ram 9943 had an LLR of 422.60 as parent2 and ewe 7283 had a LLR of 580.60 as parent1. Ram 9943 and ewe 7283 also had a combined LLRpair of 1417.44. Ewe lamb 120 and parent1 (7283) had 29 opposing homozygotes, while 120 and parent2 (9943) had 36 opposing homozygotes. *MaybeTrio* also reported the number of Mendelian errors (ME) between the offspring and the parent pair as 65.

Figure 3.13 depicts the pedigree diagram of individual 022. Ram 9943 was assigned as the sire of ewe lamb 022. Ram 9943 had an LLR of 126.52 as the assigned sire for ewe lamb 022. Ram 9943 and 022 had 48 opposing homozygotes.

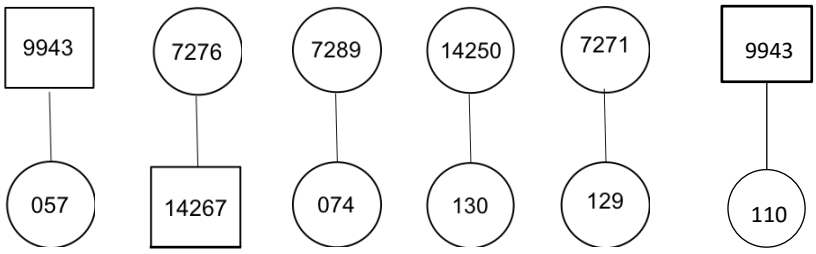

Figure 3.14 *Pedigree diagram of the individuals 057, 14267, 074, 130, 129 and 110*

Figure 3.14 illustrates the pedigree diagram of individuals; 057, 14267, 074, 130, 129 and 110. Ram 9943 was assigned as the sire of ewe lamb 057. Ram 9943 had an LLR of 130.31 as the assigned sire for ewe lamb 057. Ram 9943 and 057 had 60 opposing homozygotes. Ram 9943 was also assigned as the sire for ewe lamb 110. Ram 9943 had a LLR of 426.22 as the assigned sire for ewe lamb 110. Ram 9943 and ewe lamb 110 had 38 opposing homozygotes.

Ewe 7276 was assigned as the dam of the ram lamb 14267. Ewe 7276 had an LLR of 535.80 as the assigned dam for the ram lamb 14267, while ewe 7276 and 14267 had 34 opposing homozygotes.

Ewe 7289 was assigned as the dam of the ewe lamb 074. Ewe 7289 had an LLR of 541.21 as the assigned dam for the ewe lamb 074. Ewe 7289 and 074 had 68 opposing homozygotes.

Ewe 14250 was assigned as the dam of the ewe lamb 130. Ewe 14250 had an LLR of 531.14 as the assigned dam for the ewe lamb 130, while ewe 14250 and 130 had 74 opposing homozygotes.

Ewe 7271 was assigned as the dam of the ewe lamb 129. Ewe 7271 had an LLR of 518.61 as the assigned dam for the ewe lamb 129. Ewe 7271 and 129 had 31 opposing homozygotes.

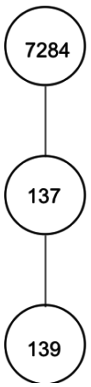

Figure 3.15 *Pedigree diagram of 7284 offspring and grand-offspring*

Figure 3.15 describes the pedigree diagram of the offspring and grand-offspring of the ewe 7284. Ewe 7284 was assigned as the dam of ewe 137 with an LLR of 416.18. Ewe 7284 and ewe 137 had 44 opposing homozygotes. Ewe 137 was assigned as the dam of ewe lamb 139 and had an LLR of 833.83 and 62 opposing homozygotes. It can thus be assumed that 7284 would be the grand-dam of ewe lamb 139, because ewe 137 is the dam of 139 and ewe 7284 is the dam of ewe 137.

A pedigree diagram was constructed for 46 individuals. Eleven dams were assigned to individuals and four sires were assigned to individuals. Ten dams were assigned as likely relatives with a likely parent-offspring (PO) relationship with the individual. Nine sires were assigned as likely relatives with a likely parent-offspring (PO) relationship to the individuals being studied. *SEQUOIA* assigned three likely parent pairs to three individuals. The samples collected included two rams that were believed to be the sires of a number of the studied individuals. According to *SEQUOIA* the sire that were assigned as a definite sire of a number of individuals was ram 9943. Ram 16008 was not assigned as a sire to any individual. Ram 16008 was assigned

as a likely relative with a likely parent-offspring relationship to three individuals. This information brings to attention the possibility that ram 16008 might not be contributing to the gene pool as much as would be expected. This could possibly be due to physical reasons such as injuries that might have occurred, there is also the possibility that ram 9943 could be more dominant and thus be preventing ram 16008 from contributing to the gene pool. This could suggest that the farmer should possibly consider replacing ram 16008 with a ram that might be more effective in the reproduction process and so contribute more to the gene pool and may have a higher genetic potential to produce more offspring.

The fact that ram 9943 is contributing to the gene pool to this extent could raise some concerns. With ram 9943 siring so many offspring compared to ram 16008, there is a possibility that this could result in inbreeding down the line if ram 9943 is used as a sire for the next couple of generations, since he would then be mating his daughters. With ram 9943 siring so many offspring it could also result in a decrease in genetic variability as the inbreeding within the flock possibly increases.

The ewes and rams that were sampled and genotyped can be considered the founders since the flock was started before mating. As many of the individuals are the founders of the population, a limited family structure/pedigree was expected. This was found to be accurate, since a limited number of individuals had assigned parents and even fewer individuals had both parents assigned.

A number of individuals were not assigned a sire or a ram which is to be expected since they are the founders and their parents may not be present in the flock. Due to the fact that these individuals make up the founders of the population, it is important to have a complete record of the dams and sires of the offspring. Ensuring records are kept will assist with the construction of breeding programmes and will provide a more complete understanding of the populations genetic structure if the complete pedigree of the flock is known. Knowledge of which rams sired which ewes will assist in the construction of an effective breeding programme to ensure that inbreeding is kept low by not breeding daughters to sires or rams to their dams. If more than one sire is used to service the ewes a programme can be drawn up to ensure that related individuals will not be mated to each other. However, this will only be possible if complete and accurate records are kept. If no records are kept genotyping and further pedigree analysis would be an effective way to determine the pedigree of the individuals within the flock.

3.4.2 PAIRWISE IBD RELATIONSHIPS

Table 3.2 provides a summary of the calculated PI_HAT values for each individual as found in the *--genome* output. The PI_HAT value is the proportion IBD [P(IBD=2) + 0.5*P (IBD=1)]. Tentative relationships between individuals were proposed based on the proportion IBD (PI_HAT). Relationships with a PI_HAT below 2,5 were not considered. Possible familial relationships were categorized according to *Table 3.*1. Relationships determined using *SEQUOIA* package were confirmed with the PI_HAT values for the relationships. The function *--genome* identified a full-sib relationship and five parent-offspring relationships that were not identified using the *SEQUOIA* package. A number of half-sib relationships were identified according to the

PI_HAT values, three possible HS/GG/FA relationships, two possible HS/FA relationships and two GG relationships were proposed.

In order to ensure a more accurate assignment of familial relationships, the PI_HAT values were used and the Z0, Z1 and Z0 values were also considered to confirm the possible familial relationships. Z0 is when no alleles are shared between two individuals, Z1 is when two individuals share one allele in common and Z2 means that two individuals have two alleles in common. The Z0:Z1:Z2 ratio for a parent-offspring pair, unrelated pair, twin pair and full-sib pair would be; (0:1:0), (1:0:0), (0:0:1) and (0,25:0,5:025) respectively. The Z0:Z1:Z2 ratio was used in addition to the PI_HAT value to estimate and confirm the familial relationships between the respective individuals.

As seen in *Table 3.2* there are a number of parent-offspring relationships as well as half-sib relationships. There are also two grandparent-grand-offspring relationships, one of which coincides with the grandparent-grand-offspring relationship determined using the *SEQUOIA* package in *RStudio*. IBD analysis provides a more complete indication of the familial relationships between individuals within the flock. Based on the PI_HAT values, the majority of the individuals that were included in the analysis were closely related as most of the relationships were parent-offspring relationships and sib relationships. These familial relationships were expected since the ewes and rams are the founders of the flock and any lambs present within the flock are expected to be the offspring of these individuals thus increasing the level of relatedness within the flock.

The familial relationships determined using the IBD relationships can be used along with the pedigree determined using *SEQUOIA* to ensure an accurate pedigree record for the individuals within the flock.

Table 3.2 *Summary containing PI_HAT values of each individual from the --genome Plink output (* 0,25<PI_HAT<0,5; ** PI_HAT >0,5 and *** PI_HAT >0,5), including the possible familial relationship (Rel)*

IID1	IID2	PI_HAT	Rel
22	120	0,25*	HS
22	14270	0,2657*	HS
22	14297	0,3114*	HS
22	57	0,2783*	HS
22	14276	0,293*	HS
22	14258	0,2594*	HS
22	100	0,2774*	HS
22	113	0,2659*	HS
57	14276	0,2647*	HS
57	14258	0,277*	HS
57	113	0,2601*	HS
74	7289	0,5**	PO
100	113	0,2631*	HS
110	22	0,3055*	HS
110	120	0,3077*	HS
110	14270	0,2775*	HS
110	14297	0,2634*	HS
110	57	0,2584*	HS
110	14258	0,2867*	HS
110	113	0,2599*	HS
115	74	0,2776*	HS/GG/FA
115	14293	0,2589*	HS
115	14286	0,5**	PO
117	14249	0,2827*	HS
117	16008	0,504***	PO
117	14279	0,503***	PO
120	14270	0,3096*	HS
120	14297	0,2689*	HS
120	57	0,2838*	HS
120	14276	0,3034*	HS
120	14258	0,2717*	HS
120	100	0,2523*	HS
120	113	0,3101*	HS
122	9943	0,5**	PO
122	22	0,2584*	HS
122	120	0,2939*	HS
122	14270	0,2574*	HS
122	14297	0,2581*	HS
122	57	0,2847*	HS
122	14276	0,2584*	HS
122	14258	0,2712*	HS
122	100	0,2682*	HS
128	14280	0,5**	PO
129	7271	0,5135***	PO
130	115	0,278*	HS/GG/FA
130	74	0,2687*	HA/GG/FA
130	14250	0,5**	PO
132	14261	0,5006***	FS
132	122	0,2597*	HS
132	137	0,2741*	HS
132	9943	0,5**	PO
132	110	0,2658*	HS
132	22	0,2562*	HS
132	120	0,2745*	HS
132	7284	0,5015***	PO
132	14297	0,3009*	HS
132	14276	0,2895*	HS
132	14258	0,3043*	HS
132	100	0,2811*	HS
137	14290	0,3013*	HS/FA
137	7284	0,5**	PO
139	137	0,5**	PO
139	7284	0,2569*	GG

IID1	IID2	PI_HAT	Rel
7275	110	0,5081***	PO/FS
7276	14258	0,512***	PO
7277	113	0,5172***	PO
7283	120	0,5416***	PO
7290	7301	0,2816*	HS/FA
7301	100	0,5088***	PO
9943	110	0,5123***	PO
9943	22	0,5**	PO
9943	120	0,5438***	PO
9943	14270	0,5324***	PO/FS
9943	14297	0,5282***	PO
9943	57	0,5**	PO
9943	14276	0,5**	PO/FS
9943	14258	0,5107***	PO
9943	100	0,5169***	PO
9943	113	0,5218***	PO
14249	16008	0,5**	PO
14249	16065	0,5113***	PO
14249	14292	0,3071*	HS
14258	100	0,3164*	HS
14258	113	0,2654*	HS
14261	14267	0,2511*	HS
14261	9943	0,5016***	PO
14261	110	0,2924*	HS
14261	22	0,3114*	HS
14261	120	0,2974*	HS
14261	14270	0,2738*	HS
14261	7284	0,5021***	PO
14261	14297	0,378*	HS
14261	57	0,3023*	HS
14261	14276	0,3079*	HS
14261	14258	0,2863*	HS
14261	100	0,287*	HS
14261	113	0,2677*	HS
14267	7276	0,5237***	PO
14267	9943	0,4993*	PO
14267	110	0,31*	HS
14267	22	0,2573*	HS
14267	14270	0,3318*	HS
14267	14297	0,2622*	HS
14267	14258	0,4767*	HS
14267	100	0,3174*	HS
14270	14297	0,311*	HS
14270	57	0,2602*	HS
14270	14276	0,3015*	HS
14270	14258	0,2603*	HS
14270	100	0,382*	HS
14276	14258	0,2713*	HS
14276	100	0,269*	HS
14276	113	0,2748*	HS
14286	14276	0,3248*	GG
14292	14290	0,3565*	HS
14293	14286	0,5**	PO
14293	14276	0,5**	PO
14297	7302	0,5**	PO
14297	57	0,2751*	HS
14297	14276	0,2881*	HS
14297	14258	0,2553*	HS
14297	100	0,2828*	HS
14297	113	0,2628*	HS
16008	128	0,5**	PO
16065	14292	0,5368***	PO
16065	14290	0,5165***	PO

3.5 CONCLUSION

Using both *SEQUOIA* and pairwise IBD relationships, a more complete pedigree of the individuals within the flock could be constructed. Pedigree records will assist the farmer to make more informed breeding decisions and constructing effective breeding programmes. The pedigree of a small flock is important to ensure there is no unintentional inbreeding. Sibling relationships might not be important for the breeding decisions of the breeder if all the offspring are sold, however, the information will provide a more encompassing knowledge to the farmer should any of the offspring be used as subsequent breeding stock. Knowing the sibling relationships will assist in preventing unnecessary inbreeding if certain individuals are kept as replacement ewes or young rams.

Knowing the sires of the individuals will alert the farmer as to which ram sired more offspring and so provide relevant information with regards to which ram might need to be replaced due to the fact that the ram might not be reproducing as much as expected, possibly due to physical or reproductive limitations. Reproductive limitations could be due to possible sterility, the ram may not be able to ejaculate or the ram may even have some or other reproductive disease that could negatively affect the breeding process with the ewes, resulting in a lowered conception rate and a loss of profit. If multiple sires are used to service the ewes a correct breeding programme can be set up in order to breed certain ewes to certain rams to ensure there is no inbreeding or for selection purposes.

Knowing the dams and sires of individuals will provide the farmer with an advantage if the offspring of a certain parent pair outperform other individuals. The farmer will then be able to ensure that certain dams are mated to certain sires that the farmer knows will produce outstanding offspring that may possess above average genetic potential and result in desirous production outputs. Known pedigree will also allow a farmer to incorporate controlled inbreeding in his flock in order to make use of potential hybrid vigour, which could have a possible advantageous effect on the production outputs.

Pedigree information combined with the genetic structure and diversity statistics of the flock will alert the farmer as to how much inbreeding he can allow in his flock if he only has one ram servicing the ewes or how to rotate the rams among the ewes to ensure the minimum inbreeding. A combination of pedigree information, the genetic structure of the flock as well as the genetic diversity within the flock can also be important for selection purposes with regards to the breeding goals of the farmer.

Further studies are needed to determine the possibility of using a smaller marker panel to determine pedigree for flocks where meeting the cost of a 50K SNP chip is not possible. Finding a cheaper method to attain the same data will be beneficial for many smallholder farmers with regards to constructing breeding and selection programmes. Studies are needed to determine the possibility of developing a smaller chip that is able to match the accuracy of a larger panel.

3.6 REFERENCE

Anderson, S., 2003. Animal genetic resources and sustainable livelihoods. Ecol. Econ. 45, 331–339 https://doi.org/10.1016/S0921-8009(03)00088-0.

Beh, K. & Maddox, J., 1996. Prospects for Development of Genetic Markers for Resistance to Gastrointestinal Parasite Infection in Sheep. Int. J. Parasitol. 26, 879–897 https://doi.org/10.1016/S0020-7519(96)80060-9.

Calus, M.P., Mulder, H. & Bastiaansen, J.W., 2011. Identification of Mendelian inconsistencies between SNP and pedigree information of sibs. Genet. Sel. Evol. 43, 1–12 https://doi.org/10.1186/1297-9686-43-34.

Clark, S., Kinghorn, B. & van der Werf, J.H., 2013. Comparisons of identical by state and identical by descent relationship matrices derived from SNP markers in genomic evaluation. Proc. Assoc. Advmt. Anim. Breed. Genet. 20, 261–265.

Engelbrecht, C. & Engelbrecht, F., 2016. Shifts in Köppen-Geiger climate zones over southern Africa in relation to key global temperature goals. Theor Appl Clim. 123, 247–261 https://doi.org/10.1007/s00704-014-1354-1.

Goddard, M. & Hayes, B., 2007. Genomic selection. J. Anim. Breed. Genet. 124, 323–330.

Hayes, B., 2011. Technical note: Efficient parentage assignment and pedigree reconstruction with dense single nucleotide polymorphism data. J. Dairy Sci. 94, 2114–2117 https://doi.org/10.3168/jds.2010-3896.

Huisman, J., 2017. Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. Mol. Ecol. Resour. 17, 1009–1024 https://doi.org/10.1111/1755-0998.12665.

Huisman, J., 2019a. Package 'sequoia'.

Huisman, J., 2019b. sequoia: Reconstruction of multi-generational pedigrees from SNP data. , 1–46.

Jones, A., Small, C., Paczolt, K. & Ratterman, N., 2010. A practical guide to methods of parentage analysis. Mol. Ecol. Resour. 10, 6–30 https://doi.org/10.1111/j.1755-0998.2009.02778.x.

Marshall, T., Slate, J., Kruuk, L. E. & Pemberton, J., 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7, 639–655.

Montgomery, G., Campbell, M., Dickson, P., Herbert, S., Siemering, K., Ewen-White, K., Visscher, P. & Martin, N., 2005. Estimation of the rate of SNP genotyping errors from DNA extracted from different tissues. Twin Res. Hum. Genet. 8, 346–352 https://doi.org/10.1375/twin.8.4.346.

Montgomery, G., Lord, E., Penty, J., Dodds, K., Broad, T., Cambridge, L., Sunden, S.L., Stone, R. & Crawford, A., 1994. The Booroola Fecundity (FECB) Gene Maps to Sheep Chromosome 6. Genomics. 22, 148–153.

Motiang, D. & Webb, E., 2016. The influence of household characteristics on cattle off-take rates in the North West Province of South Africa. Livest. Res. Rural Dev. 28.

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., Cognie, Y., Chitour, N. & Elsen, J., 2001. Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in Booroola Merino ewes. Proc. Natl. Acad. Sci. U.S.A 98, 5104–5109 https://doi.org/10.1073/pnas.091577598.

Paiva, S., Facó, O., Faria, D., Lacerda, T., Barretto, G., Carneiro, P.L., Lobo, R.N. & McManus, C., 2011. Molecular and pedigree analysis applied to conservation of animal genetic resources: the case of Brazilian Somali hair sheep. Trop Anim Heal. Prod. 43, 1449–1457 https://doi.org/10.1007/s11250-011-9873-6.

Powell, J., Visscher, P. & Goddard, M., 2010. Reconciling the analysis of IBD and IBS in complex trait studies. Nat. Rev.

Genet. 11, 800–805.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Sklar, P., Maller, J., de Bakker, P.I., Daly, M. & Sham, P., 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. 81, 559–575 https://doi.org/10.1086/519795.

R., 2011. The R User Conference.in The R User Conference. Covetry, UK.

Richards, J., Semple, S. & Atkins, K., 2010. Practical application of precision sheep management. Reproduction. 357, 53–56.

Rohrer, G., 2004. An overview of genomics research and its impact on livestock reproduction. Reprod. Fertil. Dev. 16, 47–54 https://doi.org/10.1071/rd03077.

Saunders, I., Brohede, J. & Hannan, G., 2007. Estimating genotyping error rates from Mendelian errors in SNP array genotypes and their impact on inference. Genomics. 90, 291–296 https://doi.org/10.1016/j.ygeno.2007.05.011.

Stevens, E., Heckenberg, G., Roberson, E.D., Baugher, J., Downey, T. & Pevsner, J., 2011. Inference of relationships in population data using identity-by-descent and identity-by-state. PLoS Genet. 7, e1002287 https://doi.org/10.1371/journal.pgen.1002287.

Strucken, E., Lee, S., Lee, H., Song, K., Gibson, J. & Gondro, C., 2016. How many markers are enough? Factors influencing parentage testing in different livestock populations. J. Anim. Breed. Genet. 133, 13–23 https://doi.org/10.1111/jbg.12179.

Thompson, E. & Meager, T., 1987. Parental and Sib Likelihoods in Genealogy Reconstruction. Biometrics. 43, 585–600 https://doi.org/105.28.118.141.

Wang, J., 2004. Sibship Reconstruction from Genetic Data with Typing Errors. Genetics. 166, 1963–1979.

Weir, B., 2013. Interpreting Whole-Genome Marker Data. Stat Biosci. 5 https://doi.org/10.1007/s12561-013-9090-0.

Williams, N., 2009. Feeding the future world. Curr. Biol. 19, R968–R969 https://doi.org/10.1016/j.cub.2009.10.045

CHAPTER 4
Diversity statistics of South African smallholder sheep and sheep from a South African research flock using an OvineSNP50 BeadChip

4.1 ABSTRACT

The aim of this study was to determine genetic diversity statistics in order to understand the genetic structure in the sheep populations as well as to compare the statistics between the four sample populations. The sample populations included 45 purebred Dorpers and 43 Namaqua Afrikaner (Namafr) sheep from the Nortier research farm, 176 Dorpers (Dorpersm 1) from the Ebenheaser smallholder community and 48 Dorpers (Dorpersm 2) from a smallholder community in Beaufort-West. The Namafr had the highest Linkage Disequilibrium (LD), 0.9866, the Dorper had the second highest LD, 0.9861, and the two smallholder Dorpers with an LD of 0.3898 and 0.3673 for the Dorpersm 2 and the Dorpersm 1 respectively. The inbreeding coefficients of each population was calculated with the Dorpersm 1 having the highest inbreeding value of -0.29%, followed by the Dorpersm 2, the Dorper and lastly the Namafr with inbreeding values of -2.65%, -5.35% and -7% respectively. Runs of homozygosity (ROH) identified 2 ROH in the Dorper, 105 ROH in the Dorpersm 1, 22 ROH in the Dorpersm 2 and no ROH were found in the Namafr population. A Principal Component Analysis was carried out based on the combined populations indicated two separate clusters for the Namafr and the Dorpers. There was relative variation found in the Dorpersm 1 and Dorpersm 2 compared to the Dorper purebred population. The information obtained from the analyses performed on the populations will provide the farmer with accurate and informative genetic information for the populations, which will assist with breeding and management decisions as well as the possible conservation of the genetic resources available in the population.

Keywords: genetic diversity, inbreeding, runs of homozygosity, linkage disequilibrium, principal component analysis, Nama Afrikaner, Dorper, smallholder

4.2 INTRODUCTION

Selection is one of the most important tools a farmer has available to genetically improve his population (Mavrogenis, 1995). Selection used by farmers is artificial selection, where the farmer determines which genotypes will be mated to each other and the desirable traits to select for and improve. An effective selection programme depends on three factors, namely; the degree of inheritance of the desirous traits, the selection pressure enforced and the generation interval (Mavrogenis, 1995). Selection can be applied at different intensities (Falconer & Mackay, 2009); a high selection intensity means the farmer chooses a select

few individuals that are superior to the other individuals to be bred, while a low selection intensity includes many individuals to be bred, meaning some individuals may be selected that may not have as high a genetic potential as actually desired in the selected individuals. Selection cannot be practised if there are no records for the individuals and no reliable evaluation procedures in place (Mavrogenis, 1995). In order to ensure that selection of individuals is effective, a complete understanding of the genetic diversity and structure of the population is needed (Mavrogenis, 1995).

Population parameters can be calculated that will provide a good overview of the genetic structure of the population and the animal genetic resources available within the population (Sheikhlou & Abbasi, 2016). This could assist in developing an effective breeding strategy and making informed selection decisions based on the genetic resources available within the population as well as the breeding goals of the farmer (Toro & Caballero, 2005). Some of these parameters include; the degree of linkage disequilibrium (LD), the inbreeding coefficient, the genetic structure of the population, the degree of heterozygosity within the population as well as the relatedness of the individuals within the population.

The degree of heterozygosity within a population is also a good indication of the genetic variation within the population. Gregorius (1978) suggests that heterozygosity measures the genotypic variation and may refer to Hardy-Weinberg proportions. The fixation index, F, compares the observed heterozygotes to the Hardy-Weinberg expected heterozygotes in populations under random mating (Hamilton, 2011a). $F = \frac{H_e - H_o}{H_e}$, where H_e is the expected heterozygote, and H_o is the observed heterozygote. A negative fixation index indicates an excess of heterozygotes relative to Hardy-Weinberg expectations, while a positive fixation index indicates an excess of homozygotes relative to Hardy-Weinberg expectations (Wakefield, 2010; Hamilton, 2011a). Expected heterozygosity is often intended as an indirect methods of detecting variation of adaptive polygenic traits (Toro & Caballero, 2005), where a high level of heterozygosity indicates more genetic variability and low levels of heterozygosity indicates less genetic variability and a small N_e (Al-Mamun *et al.*, 2015).

Zhu *et al.* (2013) defines LD as the non-random association between alleles at different loci. Gibson *et al.* (2006) describes LD as the tendency of alleles to be inherited together above what would be expected under random segregation. According to Meadows *et al. (*2008), LD can also be described as the ability of an allele from one marker to predict the allelic status at a second marker. The level of LD is generally influenced by non-genetic factors as well as genetic factors including; allele frequency, rate of recombination, selection, non-random mating, rate of mutation, genetic drift and population structure (McRae *et al.*, 2002; Zhu *et al.*, 2013). The LD is often used for selection purposes such as marker-assisted selection, genomic selection as well as association studies, which all depend on the LD present in a population (Al-Mamun *et al.,* 2015). This suggests that certain traits can be linked to their specific genes making it easier for a farmer to select individuals that may have the desired genes. Selecting individuals based on physical characteristics will lead to the indirect selection of genes that are associated with these traits that are being selected. These traits may be beneficial for reproduction or production, thus resulting in an economic advantage for the farmer.

Inbreeding occurs when individuals that are closely related are bred to each other. The inbreeding coefficient of an individual is the portion of its gene pairs that are expected to be homozygous, that would be heterozygous in an average non-inbred animal (Keller & Waller, 2002). Inbreeding results in an increase in the frequency of homozygous genotypes (Hamilton, 2011a). This increase in homozygous genotypes means there is a decrease in genetic variability within the population, suggesting that there are no or few outlier individuals that could be selected for their superior traits. Breeding within a population that has little to no genetic diversity means that there will be no genetic gain or improvement within the flock as all the individuals could be genetically similar. This decrease in genetic variability means that no new or advantageous genotypes will appear in the population, except through possible genetic mutations. This lack of genetic variability within the population will eventually lead to the inability of the population to adapt to the changing environment as well as a possible decrease in future genetic gains (Goddard, 2009).

Inbreeding within the population could result in inbreeding depression which could have a negative impact on fitness-related traits or even the expression of phenotypic defects as well as lethal genotypes (Fernández *et al.,* 1995; Falconer & Mackay, 2009).

The inbreeding coefficient determines the extent to which a population is possibly inbred, and it is determined by comparing the observed and expected homozygous genotypes (Purcell *et al.*, 2007). The inbreeding coefficient can be influenced by the intensity of selection pressure that is applied by the farmer (Hedrick, 2013). Inbreeding has been reported to have both positive and negative effects on different production traits, growth traits as well as reproduction traits (Lamberson *et al.*, 1982; Ercanbrack & Knight, 1991; van Wyk *et al.*, 1993, 2009; Analla *et al.*, 1999; Dorostkar *et al.*, 2012; Gowane *et al.*, 2013). Inbreeding has also resulted in a reduction in birth weight as well as weaning weight (Doney, 1957; van Wyk *et al.*, 2009; Selvaggi *et al.*, 2010; Eteqadi *et al.*, 2014; Mokhtari *et al.*, 2014; Gholizadeh & Ghafouri-Kesbi, 2016). Increases and reductions in production traits like weight at first service, weight at first kidding as well as average daily gain have been found to be influenced by inbreeding (Khan *et al.*, 2007; Pedrosa *et al.*, 2010; Hossein-Zadeh, 2013; Yavarifard *et al.,* 2014). This indicates the importance of knowing whether a flock is inbred or not as well as the extent of inbreeding if there is inbreeding. Knowledge of the extent of possible inbreeding will assist in determining whether there is a need for a new or revised breeding plan to ensure that the inbreeding does not result in possible inbreeding depression, phenotypic defects or even lethal alleles. These could result in major losses to the production system through the loss of genetic resources and economic loss.

The genetic structure of the population can be described by the variability of alleles and genotypes in the population (Boichard *et al.,* 1997). Wright's F-statistic is an effective method to determine the populations subdivision using the following fixation indexes; F_{IS}, F_{ST} and F_{IT}. F_{ST} determines the loss of heterozygosity within the subpopulation compared to the total population, while F_{IS} determines the loss of heterozygosity within the subpopulation and F_{IT} determines the loss of heterozygosity of the entire population (Barros *et al.*, 2017). A balance between homozygous and heterozygous genotypes is needed to ensure there is a large

enough variance in the gene pool for future selection decisions within the subpopulations but also across the entire population of the breed. There is a need to know the current status of the genetic resources and genetic variance available and how to use them for maximum and efficient production, but also to be able to set up specialized breeding and conservation strategies (Kevorkian *et al.*, 2010).

A number of genetic diversity studies have been conducted on sheep worldwide. Genetic diversity studies on five populations of Australian sheep, determined low levels of linkage disequilibrium (LD) compared to other studies, and genetic distances that were modest compared to other studies (Al-Mamun *et al.*, 2015). However, the high genetic diversity was observed within the breeds and runs of homozygosity (ROH) suggested there was strong recent selection on two chromosome regions (Al-Mamun *et al.*, 2015). The genetic structure of three Sicilian dairy sheep breeds were investigated; Valle del Belice, Comisana and the Pinzirita (Mastrangelo *et al.*, 2014). The average LD (r^2) was calculated as (0.155 ± 0.204) for the Valle del Belice, (0.156 ± 0.208) for the Comisana and (0.128 ± 0.188) for the Pinzirita breed. Mastrangelo *et al.* (2014) stated that although the Valle del Belice had the lowest genetic diversity compared to the other breeds, it possessed high genetic differentiation within the breed. However, there were inbred individuals in the same flock as indicated by the high inbreeding coefficient for the Valle del Belice. It was found that the Pinzirita breed had the highest genetic diversity of the three breeds (Mastrangelo *et al.*, 2014). The population structure of 18 native Welsh breeds were investigated, and an average pairwise F_{ST} between breeds was reported as 0.107. The 18 native breeds formed four subpopulations, high levels of haplotype sharing was also found between the native Welsh breeds and other European breeds (Beynon *et al.*, 2015). The genetic diversity and population structure of four South African breeds was also investigated by Sandenbergh *et al.* (2015). The investigated Namaqua Afrikaner (NA) possessed the least number of polymorphic loci and was found to be the least genetically diverse breed compared to the other three breeds that were investigated. The South African Merino displayed high diversity compared to results found in international Merinos. A principal component analysis (PCA) reported four definite clusters (Sandenbergh *et al.*, 2015).

Other genetic diversity studies have been conducted on sheep populations from around the world (Worley *et al.*, 2004; Gizaw *et al.*, 2007; Handley *et al.*, 2007; Kevorkian *et al.*, 2010; Naqvi *et al.*, 2017) as well as in South Africa (Peters *et al.*, 2010; Molotsi *et al.*, 2017b). There is, however, still a need for additional genetic diversity studies that focus specifically on sheep breeds that are native to South Africa and may more adapted to the South African climate.

Climate change has been recognised as having a significant effect on farming practises and there are definite concerns regarding the impact it will have on rural communities and smallholder farms with regard to farming activities and livestock that depend on the ecosystem (Thomas & Twyman, 2007). A full understanding and knowledge of the genetic resources we have available in South Africa is needed in order to develop effective breeding strategies to facilitate adaptation and robustness in individuals in these smallholder communities to ensure continuous production irrespective of the climate (Thomas *et al.*, 2007) This further emphasises the need for additional genetic studies of the sheep breeds native to South Africa.

Genotyping a wide range of individuals is an effective method to determine a complete and accurate record of the genetics available in a population.

The aim of this study was to calculate population parameters that will assist in understanding the genetic structure and diversity of the studied sheep flocks and determine the genetic resources available to the farmer, thereby increasing production outputs and improving efficiency.

4.3 MATERIALS AND METHODS

4.3.1 DESCRIPTION OF STUDY LOCATION

Blood samples were collected from 48 smallholder Dorpers on a farm in Beaufort-West. This sample group will be identified as Dorpersm 2 for the remainder of this thesis. A description of the study location can be found in chapter three of this thesis under 3.3.1 Description of study location.

Further genotypic data from 264 sheep were used from a study conducted by Molotsi *et al.* (2017). These blood samples were collected from the Nortier Research farm and the Ebenheaser smallholder community, a description of the study location can be found in (Molotsi *et al.*, 2017b). A full ID list of the individuals used can be found in the *Supplementary Material, 6) Sample ID's, Table 1.6.1*.

4.3.2 RESEARCH POPULATION AND SAMPLING METHOD

Blood sample collection and a description of the research population can be found in chapter three of this thesis under 3.3.2 Research population and sampling method.

Additional genotypic data was provided by Molotsi *et al.* (2017b), and further information with regards to the research population and sampling method can be found in (Molotsi *et al.*, 2017b). Genotyping samples received included 176 smallholder Dorpers collected from the Ebenheaser smallholder community (Dorpersm 1), 45 Dorper and 43 Namaqua Afrikaner individuals that were collected from the Nortier research farm.

4.3.3 STATISTICAL ANALYSIS

Quality control and pruning was carried out on each sample population separately. The samples were pruned according to set threshold values in *Plink* (Purcell *et al.*, 2007). For quality control purposes --mind 0.1 was set to remove individuals with excessive missing genotype data, this means that individuals with more than 10% missing genotypes were pruned from the sample list. A --geno 0.1 threshold was set to include only SNPs with a 90% genotyping rate, while --maf 0.05 was set to exclude SNPs with a minor allele frequency below 0.05, and lastly a --hwe threshold of 0.001 was set in order to exclude markers that failed to conform to the Hardy-Weinberg test at this significant threshold, which could be an indication of genotyping errors (Weir, 2013). The average Minor Allele Frequency (MAF) was also calculated for each population using the --freq function in *Plink*.

Genetic information was calculated using *Plink* (Purcell *et al.*, 2007); these include the Linkage Disequilibrium (LD) of the population --r2, the inbreeding coefficient –het and the runs of homozygosity --homozyg.

The binary files of each sample population were combined in *Plink* after pruning and quality control using the --merge-list function. The combined file was imported into RStudio (R, 2011), where a PCA was constructed using the *pcadapt* package (Luu *et al.*, 2019).

4.4 RESULTS AND DISCUSSION

Before pruning each sample, population contained 54241 variants. Quality control was carried out on all sample populations and individuals that did not meet the threshold values set were pruned. From the Nortier research farm, 43 Namaqua Afrikaner sheep underwent quality control, 1369 variants were removed due to missing genotype data (--geno), 59 variants were removed due tonon-conformity to Hardy-Weinberg exact tests (--hwe) and 16279 variants were removed due to minor allele thresholds (--maf). A total of 36534 variants and 43 sheep passed the filters and quality control.

From the Ebenheaser smallholder community 176 Dorpers (Dorpersm 1) underwent quality control. Due to missing genotype data (--geno), 3242 variants were removed, 201 variants were removed due to non-conformity to Hardy-Weinberg exact tests (--hwe) and 5101 variants were removed due to minor allele thresholds (--maf). Finally, 45697 variants and 176 sheep passed filters and quality control.

From the Beaufort-West smallholder community 48 smallholder Dorpers (Dorpersm 2) underwent quality control. One sheep (136) was removed due to missing genotype data (--mind), 3279 variants were removed due to missing genotype data (--geno), 242 variants were removed due to non-conformity to Hardy-Weinberg exact tests (--hwe) and 6183 variants were removed due to minor allele thresholds (--maf). A total of 44537 variants and 47 sheep passed the filters and quality control.

From the Nortier research farm, 45 Dorpers underwent quality control. One sheep (108170) was removed due to missing genotype data (--mind), 1405 variants were removed due to missing genotype data (--geno), 44 variants were removed due to non-conformity to Hardy-Weinberg exact tests and 8656 variants were removed due to minor allele thresholds (--maf). A total of 44136 variants and 44 sheep passed filters and quality control to be analysed further.

A number of genetic parameters were calculated for each population. The genetic information that was calculated include the linkage disequilibrium, inbreeding coefficient, genotype counts, Hardy-Weinberg test statistics and the runs of homozygosity for each sample population. These statistics were all calculated using specialised functions in *Plink* (Purcell *et al.*, 2007).

4.4.1 GENOTYPING RATE, POLYMORPHIC SNPs AND MAF

The call rate for a given SNP is defined as the proportion of individuals in the study for which the corresponding SNP information is not missing, and is thus the proportion of genotypes per marker with non-missing data (Anderson *et al.*, 2010; Reed *et al.*, 2015).

The Namafr population had the highest genotyping rate of 0.9866, followed by the Dorper population, 0.9861, and then the Dorpersm 1 population, 0.9413, while the Dorpersm 2 population had the lowest genotyping rate of 0.9405. The genotyping rates of the Namafr population and the Dorper population were similar while the Dorpersm 1 and Dorpersm 2 population were more similar. The lower genotyping rate observed in both the smallholder populations could possibly be due to random genetic mutations that occurred close to the marker (Pompanon *et al.*, 2005). It would thus be expected that the smallholder populations could possibly have higher genetic variability compared to the more purebred populations of the Namafr and the Dorper.

The polymorphic SNPs for the Dorpersm 1 was the highest, 84.25%, followed by the Dorpersm 2 with a polymorphic SNP of 82.11%, then the Dorper population, 81.37%. The Namafr had the lowest polymorphic SNPs compared to the other populations 67.35%. The low polymorphic SNPs observed in this study is similar to the low polymorphic SNPs for the Namafr in a study conducted by Sandenbergh *et al.* (2016).The number of polymorphic SNPs within a population can be an indication of the suitability of the specific SNP panel for further genome-wide association studies within the population (Molotsi *et al.*, 2017c). The low polymorphic % observed for the Namafr population can be due to the Namafr and other African fat-tailed breeds being under-represented for the development of the Ovine50SNP Beadchip compared to other breeds (Molotsi, 2017). The low polymorphic SNPs observed in the Namafr population could possibly be due to ascertainment bias during the SNP discovery process (Kijas *et al.*, 2009; Sandenbergh *et al.*, 2016). Ascertainment bias occurs when data has not been collected randomly with regards to the observed data patterns (Nielsen & Signorovitch, 2003). This could suggest that the markers on the Ovine50SNP may not be well-suited for identifying SNPs in Namafr. Kijas *et al.* (2009) performed Multidimensional scaling (MDS) on the sample populations included in his study, which demonstrates the clustering of the populations from Asia, Africa and other countries. It is evident that the African populations are clustered together away from the other clusters from Asia and the other countries. This suggests that there is a genetic difference between the African sheep populations and the other sheep populations from around the world. This genetic difference between the populations could also explain the low polymorphic SNPs observed in the Namafr since markers that may be present in sheep populations from other countries may not necessarily be present in African sheep populations. Further studies are needed to determine the validity of constructing Ovine50SNP BeadChips that are specially designed for certain countries to ensure a more accurate representation of the SNP data, since different populations will have specific markers. There is thus a need for a specialised BeadChip for genotyping purposes.

The Dorpersm 1 population had the highest mean MAF of 0.2869, followed by the Dorper population, 0.2810, then the Dorpersm 2 population with a mean MAF of 0.2805 and finally the Namafr population with the lowest mean MAF of 0.2743. Low MAF could possibly be an indication of lower genetic variation within the population. We can thus assume that the Namafr population may have a lower genetic diversity compared to the other breeds in the study.

Table 4.1 *Diversity stats including mean genotyping rate, mean LD, mean inbreeding coefficient and the observed and expected heterozygosity of the Dorpersm 2, dorper, Dorpersm 1 and Namafr sample populations, (n) is the sample size after pruning*

	Sample Size (n)	Polymorphic SNPs after pruning	Polymorphic %	Mean MAF	Genotyping rate	LD (r^2)	Inbreeding coefficient (F_{IS})	Observed (HET)	Expected (HET)
Dorpersm 2	47	44537	82.11%	0.2805	0.9405	0.3898	- 2.65%	0.3795	0.3697
Dorper	44	44136	81.37%	0.2810	0.9861	0.4156	- 5.35%	0.3820	0.3627
Dorpersm1	176	45697	84.25%	0.2869	0.9413	0.3673	- 0.29%	0.3778	0.3767
Namafr	43	36534	67.35%	0.2743	0.9866	0.4805	- 7%	0.3860	0.3541

4.4.2 HETEROZYGOSITY

As observed in *Table 4.1* the observed heterozygosity for all four the sample populations were relatively similar, and the expected heterozygosity for the three Dorper populations were similar, while the Namaqua Afrikaner had a lower expected heterozygosity. The higher observed heterozygosity is expected for the Dorpersm 1 and Dorpersm 2 populations, since smallholder communities generally do not have structured breeding programmes for their flocks, which could cause an increase in heterozygotes observed compared to the expected heterozygotes, resulting in a possible increase in genetic variability. The parent individuals from the Dorpersm 2 population can be considered the founders of the smallholder population. If these individuals are the founders of the population, we can assume there is relative genetic diversity within the flock since they have not had the chance to mate each other as they are the founder generation. This high expected genetic variability could, therefore, be a cause for the high observed heterozygosity observed in the Dorpersm 2 population.

The Namafr population and the Dorper population both come from the Nortier Research farm suggesting these two populations are probably relatively isolated. However, taking into account the conservation strategy programme that is in place for the Namafr population there is a possibility that there could have been structured gene flow to facilitate the conservation programme and to ensure that inbreeding does not become a problem in the population (Qwabe *et al.*, 2013). This isolate breaking through deliberate gene flow could result in a temporary excess of heterozygotes. This could possibly be a cause for the high observed heterozygosity observed in the Namafr population.

Sandenbergh *et al.* (2015) determined observed heterozygosity for four South African breeds that were similar to the observed heterozygosity in this study. A higher observed heterozygosity compared to the expected heterozygosity was also observed in a study conducted by Molotsi *et al.* (2017b).

The genetic variability observed in the sample populations indicate genetic potential for possible selection according to desired goals. Taking into account all the studies mentioned above with regard to expected and observed heterozygosity, it is clear that South African sheep breeds generally have a high genetic variability. This suggests that South African breeds have the potential for possible higher genetic gain, as well as the ability to better adapt to the climate change that is expected. Therefore, it is important to use native sheep breeds for production systems in order to ensure the individuals are robust and able to adapt to the climate found in South Africa.

4.4.3 LINKAGE DISEQUILIBRIUM

The Dorpersm 1 had the lowest LD (r^2) value, 0.3673, followed by the Dorpersm 2 with the second lowest LD, 0.3898, then the Dorper, 0.4156 and finally the Namafr population with the highest LD of 0.4805. In a Churra sheep population, the average LD for SNPs more than 50kb apart was 0.003 (García-Gámez *et al.*, 2012). This LD was similar to the LD determined for both the smallholder Dorper populations. Kijas *et al.* (2014) calculated the LD for SNPs 70kb apart for five populations namely; Merino ewes (r^2 = 0.080), Merino sires (r^2 = 0.083), Poll Dorset (r^2 = 0.166), Suffolk (r^2 = 0.111) and Border Leicester (r^2 = 0.224). The LD for all five populations were much lower than the LD determined for the populations in this study. Variation in LD could also be contributed to recombination, admixture and bottlenecks due to breed selection (Meadows *et al.,* 2008).

The high LD observed in all the populations in this study compared to the LD determined in the other studies could be due to movement of individuals between subpopulations which results in gene flow which thus causes an increase in LD if the allele frequencies among the subpopulations differ (Nei *et al.,* 1973; Slatkin, 2008). Gene flow often occurs in smallholder communities since farmers often "borrow" the neighbours rams, they may also get replacement ewes from neighbours (Mueller *et al.*, 2015). This could explain the high LD observed in the smallholder populations. The high LD in the Namafr population could also be due to gene flow, if gene flow is part of the conservation programmes for the Namafr population (Qwabe *et al.*, 2013).

The limited or structured migration in the Namafr and Dorper could lead to less effective recombination due to the possible decrease of heterozygotes and thus an increase in linkage disequilibrium among local populations (Ohta, 1982). The limited migration could explain the high LD in the Namafr and Dorper populations. Ascertainment bias is also a possibly explanation for the high LD observed in the Namafr population as explained in section *4.4.1) Genotyping rate, polymorphic SNPs and MAF*.

4.4.4 F-STATISTICS

F_{IS}

Namafr had an inbreeding coefficient, F_{IS}, of -7%. This was the lowest among the four groups, the second lowest was the Dorper with an inbreeding coefficient, F_{IS}, of -5.35%. The Dorpersm 2 population had the

second highest inbreeding coefficient, F_{IS}, of -2.65% and the Dorpersm 1 had the highest inbreeding coefficient, F_{IS}, of -0.29%. The average inbreeding coefficient determined for each sample population in this study is lower than the average inbreeding coefficients determined in other studies conducted on South African sheep flocks (van Wyk *et al.*, 1993, 2009; Swanepoel *et al.*, 2007; Sandenbergh *et al.*, 2015; Molotsi *et al.*, 2017) as well as a number of other flocks from around the world (Diez-Tascón *et al.*, 2000b; Goyache *et al.*, 2003; Paiva *et al.*, 2011; Dorostkar *et al.*, 2012; Eteqadi *et al.*, 2015; Vostry *et al.*, 2018).

All the sample populations in this study had a negative F_{IS}, which indicates an excess of heterozygotes relative to the Hardy-Weinberg expectations (Hamilton, 2011c). These low inbreeding coefficients could be as a result of random matings occurring between individuals ensuring a high genetic diversity and a low inbreeding coefficient within the populations.

The higher inbreeding coefficient determined for the Dorpersm 2 compared to the Dorpersm 1 population could possibly be due to the fact that the individuals included in the study from the Dorpersm 2 population are the founders of the population on the smallholder farm. Inbreeding values can also be influenced by the high selection intensity applied to rams, which could result in an increase in the average relationship coefficients among rams (Li *et al.,* 2009). If a limited number of rams were used and the number of founders are relatively small there could be a higher level of inbreeding when the offspring are taken into consideration (Barczak *et al.*, 2009). This could be why the Dorpersm 2 population has a slightly higher inbreeding coefficient compared to the Dorpersm 1 population.

Smallholder farmers often do not keep accurate or complete records which leads to a number of unknown parents for individuals which may result in an underestimation of inbreeding levels and so also the underestimation of losses due to inbreeding (Lutaaya *et al.*, 1999).

F_{ST}

The F_{ST} determines the heterozygosity within the subpopulation compared to the total population. F_{ST} could be an indication of genetic differentiation or population substructure among different populations (Gizaw *et al.*, 2007). *Figure 4.3* displays the pairwise F_{ST} that was calculated in *Rstudio* using the *hierfstat* package (Goudet & Jombart, 2015).

Table 4.2 *Four sample populations pairwise F_{ST} according to Nei., (1987)*

	Dorper	**Dorpersm 1**	**Dorpersm 2**	**Namafr**
Dorper	-	0.0231	0.0964	0.2212
Dorpersm 1	0.0231	-	0.0516	0.1762
Dorpersm 2	0.0964	0.0516	-	0.2022
Namafr	0.2212	0.1762	0.2022	-

A global F_{ST} was determined for the four populations (F_{ST} = 0.0913), with pairwise F_{ST} ranging between 0.0231 – 0.2212. The F_{ST} values between the Namafr and all three the Dorper populations is the largest. This

indicates the most genetic differentiation is between the Namafr and the Dorper breed. This is expected since the Namafr and Dorper are two different breeds with the Dorper being a composite breed derived from a cross between the Dorset Horn and the Black-headed Persian, while the Namaqua Afrikaner is a fat-tailed breed that was kept by the Nama people (Cloete *et al.*, 2000; Qwabe *et al.*, 2013). The F_{ST} between the Namafr and the Dorper is the largest indicating the largest genetic differentiation between these two breeds, while the F_{ST} between the Namafr and the two Dorpersm populations is slightly smaller. It was expected that the Namafr and smallholder Dorpers would have a slightly lower genetic differentiation due to possible crossbreeding and the possible lack of structured breeding programmes often encountered in smallholder communities. Smallholder communities do not focus on ensuring the animals stay purebred, but focus mainly on production. This suggests that the smallholder populations may have been bred to some Namafr individuals or individuals that may have Namafr in their ancestry, thus decreasing the genetic differentiation.

According to the F_{ST} there is negligible to little genetic differentiation between the three Dorper populations. This was expected since they are from the same breed. The purebred Dorper population was genetically more similar to the Dorpersm 1 population (F_{ST} = 0.0231) than the Dorpersm 2 population (F_{ST} = 0.0964). It can thus be assumed that the Dorper population is more genetically similar to the Dorpersm 1 population than the Dorpersm 2 population. This was expected since the Dorper population from the Nortier research farm is geographically closer to the smallholder Dorpers from the Ebenheaser community than the smallholder Dorpers from the Beaufort-West farm. It can thus be assumed that geographic distance could influence differentiation between populations.

4.4.5 PRINCIPAL COMPONENT ANALYSIS

Scree plots were constructed in RStudio using the *pcadapt* package (Luu *et al.*, 2019). Scree plots with K = 10 and K = 4 were produced, with K equal to the number of subpopulations assumed. Both the scree plots indicated that the variance was explained up to three factors in *Figure 4.1* and F*igure 4.2*.

A principal component analysis (PCA) was carried out on the combined sample populations to determine any segregation between the sample populations. The PCA visually displayed the variation between the different sample populations in *Figure 4.3*.

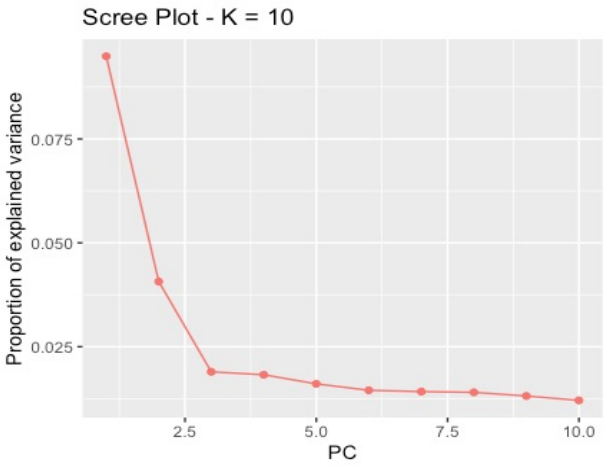

Figure 4.1 *Scree plot of the combined sample populations, K = 10.*

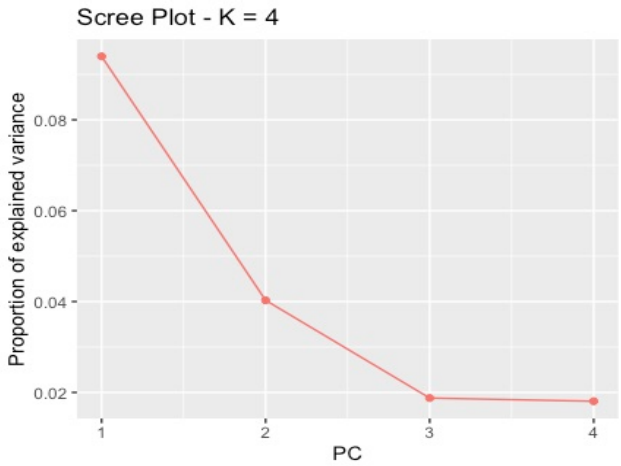

Figure 4.2 *Scree plot of the combined sample populations, K = 4.*

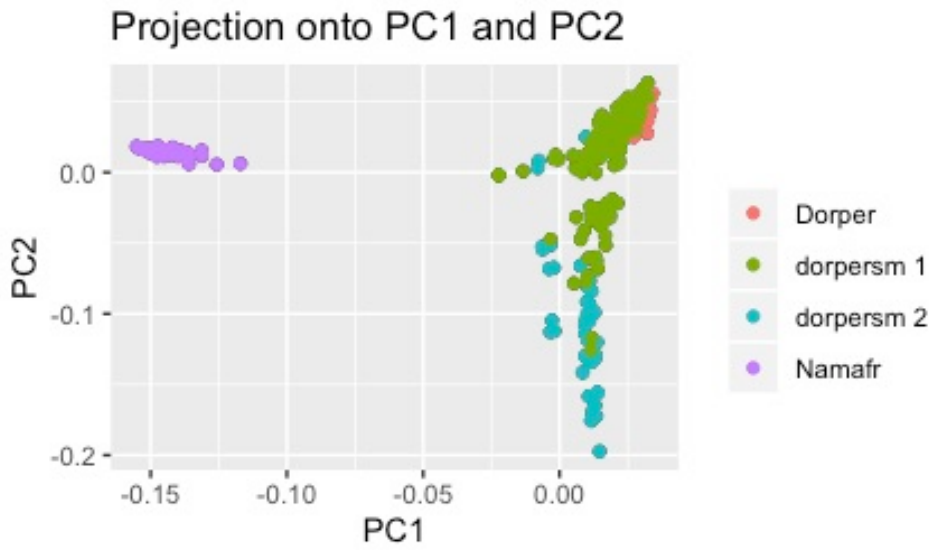

Figure 4.3 *PCA plot of the combined sample populations with the first two principal components*

As seen in *Figure 4.3*, there are two distinct clusters. However, the mixed Dorper, Dorpersm 1 and Dorpersm 2 cluster on the right is relatively spread out vertically. From the PCA we can see that the Namafr has a distinct cluster, while the Dorper, Dorpersm 1 and Dorpersm 2 cluster together.

The Namafr cluster is expected as the Namafr is a separate breed from the Dorper and there would be significant genetic difference between the two breeds after years of divergence and evolution. This was verified by the high F_{ST} calculated between the Namafr and all the Dorper populations in section 4.4.4 F-Statistics of this chapter.

The Dorper cluster includes individuals from the Dorper population, individuals from the Dorpersm 1 and the Dorpersm 2. The Dorper population is very tightly clustered. This verifies that the population is a purebred flock and there is low genetic variation within the population due to the tight cluster of the individuals. Some of the Dorpersm 1 individuals cluster close to the Dorper population, implying that the individuals clustered

close to the Dorper population may be genetically similar to a certain degree, and there is a degree of variation within the Dorpersm 1 population due to the spread of the data.

The Dorpersm 2 population does not appear to cluster close to the Dorper population, except a few individuals that are located close to the Dorper cluster, indicating some definite genetic diversity between the two populations. The Dorpersm 2 individuals do not cluster close together but instead spread out. This absence of clustering may indicate the high genetic variation present within the population.

The PCA illustrated in *Figure 4.5* is similar to the PCA illustrated by Molotsi *et al.* (2017c). The Namafr population clusters separately from the other breeds while the Dorper and smallholder Dorpers cluster together. The smallholder Dorpers show relative genetic diversity due to the spread of data points in the PCA in both studies. The Namafr populations are clustered close together in both studies suggesting low genetic variability within the population.

Sandenbergh *et al.* (2015) presented a PCA of four South African sheep breeds consisting of a Namafr population, Dorper population, SAMM population and three Merino populations. In the PCA, the Namafr and Dorper populations clustered close together similar to the clustering illustrated in *Figure 4.5*.

4.4.6 RUNS OF HOMOZYGOSITY

The Dorpersm 2 population had a total of 22 ROH, with four of the ROH being less than 5 Mbp long indicating possible older haplotype relatedness, and 18 ROH between 5 – 10 Mbp long. The Dorper flock had a total of two ROH, one of these was <5 Mbp and the other was between 5 – 10 Mbp long. The Dorpersm 1 population had the most ROH compared to the other sample populations, with a total of 105 ROH that were identified. Of the total 105 identified ROH, 14 ROH were <5 Mbp long, 89 ROH were between 5 -10 Mbp long and two ROH were >10 Mbp long. The majority of ROH identified in the Dorpersm 1 population and Dorpersm 2 population were between 5 – 10 Mbp long. No ROH were identified in the Namafr population.

Runs of homozygosity (ROH), are uninterrupted lengths of homozygous genotypes in an individual that can be due to the inheritance of identical haplotypes from their parents (Keller *et al.*, 2011; Purfield *et al.*, 2012; Zhang *et al.*, 2015a). Related matings can result in the inheritance of haplotypes that are Identical By Descent (IBD), which results in possible homozygous stretches across the genome of the offspring (Bosse *et al.*, 2012; Peripolli *et al.*, 2017). Selection will also result in haplotypes that are associated with certain desired traits to be passed on to the next generation thus increasing ROH in the offspring of individuals were selection occurred for the desired traits (Leocard, 2009; Purfield *et al.*, 2017). Identifying long and uninterrupted ROH through analysis of SNPs facilitates the identification of chromosomal regions that may be identical by descent or that may have been selected for (Marras *et al.*, 2014).

The two smallholder populations possessed the most ROH between the four breeds that were studied. Runs of homozygosity could be indicators of selection. It can thus be assumed that the smallholder populations may have been under selection. Smallholder farmers often have low-input systems, which could indicate that the sheep in a low-input production system would experience selection, intentional or

unintentional, to make it more adaptable to less than favourable environments. Further studies are thus needed to determine whether these ROH are indeed due to possible selection for adaptation/robust traits.

Figure 4.4 displays the chromosome on which the ROH were identified in each of the populations. As seen in the graph, the majority of the identified ROH occurred on chromosome 2 in both the smallholder populations. The chromosomes with the next highest number of ROH were chromosome 8, 9 and 12. Further studies are needed to determine why the majority of the ROH were identified in the smallholder populations. Studies are also needed to determine why the majority of the ROH occurred on chromosome 2 followed by chromosome 8, 9 and 12. There is also the possibility that certain desirous traits are found on chromosome 2, and if selection for any of these traits occurred it could explain why the majority of the ROH occurred on chromosome 2.

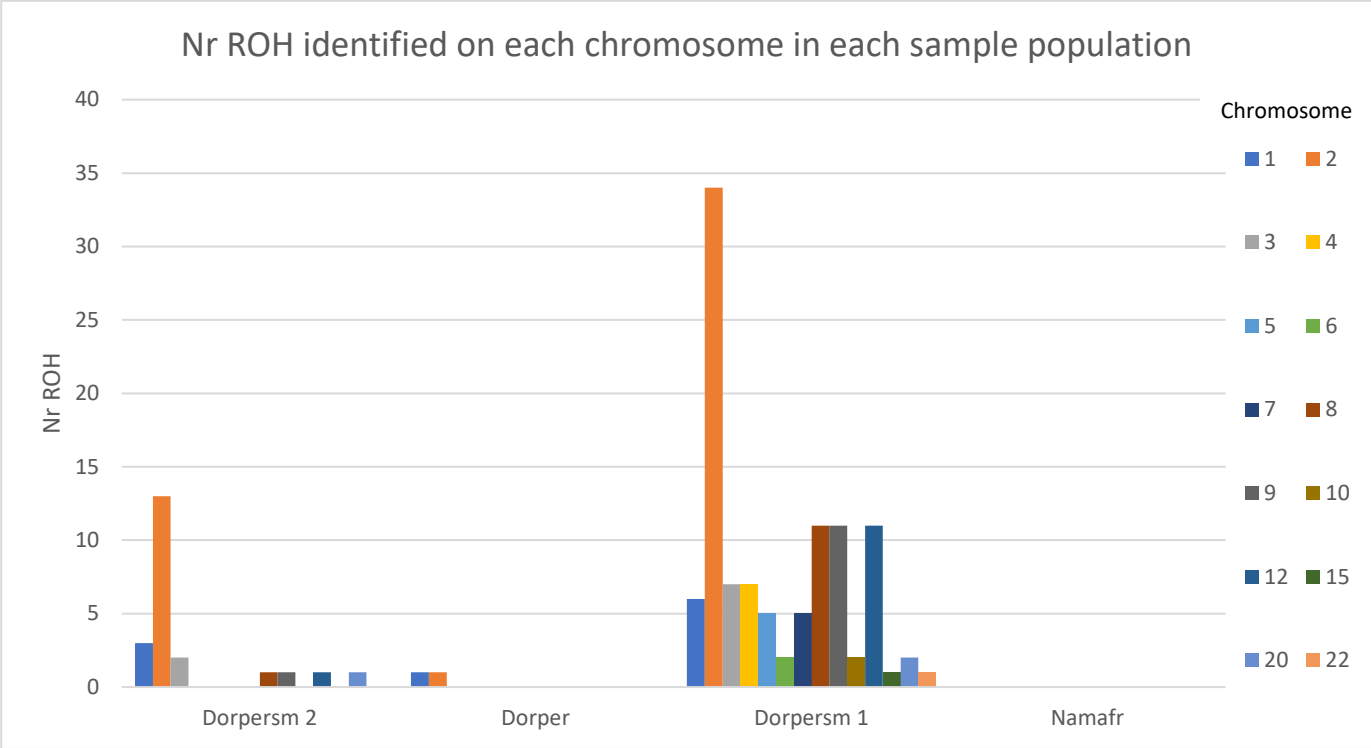

Figure 4.4 *The ROH identified on each chromosome in all four populations*

Kirin *et al.* (2010) and Al-Mamun *et al.* (2015) states that short runs of homozygosity indicate older haplotype relatedness, while longer runs of homozygosity reveal more recent relatedness between individuals. It can thus be assumed that short and medium length ROH were under selection further back than ROH that are longer in length where selection could have occurred more recently (Gibson *et al.*, 2006; Marras *et al.*, 2014; Zhang *et al.*, 2015a).

Figure 4.5 displays the number of ROH per sample population as well as the length of each ROH. There were only two runs of homozygosity longer than 10Mbp, which could possibly indicate that there are no recent relatedness events in any of the populations, which is consistent with the low inbreeding coefficient values determined. The majority of the identified ROH in the two smallholder populations and the Dorper

population are less than 10 Mbp long, which could be classified as shorter ROH. These shorter ROH could indicate older haplotype selection events, suggesting that some selection may have occurred in the populations previously.

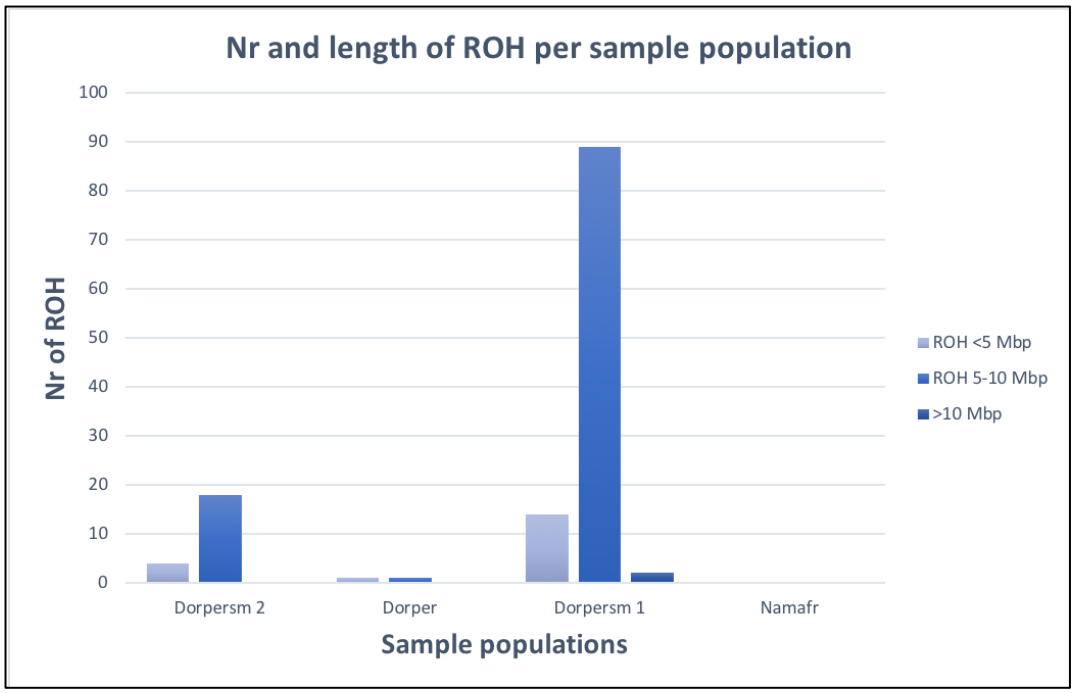

Figure 4.5 *The number and length of ROH for each sample population*

One of the causes for the different ROH lengths can be due to inheritance and the recombination process that often breaks down the haplotypes before they can be passed on to the offspring (Broman & Weber, 1999). However, individuals that are related or have undergone selection have many haplotypes in common resulting in the longer ROH, while individuals that are not related and have not undergone selection will not have as many haplotypes in common or the haplotypes would be cut short due to possible recombination events resulting in shorter ROH.

The ROH results are similar to those estimated by Mastrangelo *et al.* (2018), where the average length of the identified ROH for the breeds under investigation was 4.55 Mbp and ranged from 3.85 Mbp for the Biellese sheep to 5.51 Mbp in the Leccese sheep. A study conducted by Purfield *et al.* (2017), found that the majority of the ROH detected were less than 10 Mbp long, with very few ROH that were > 20 Mbp long.

Runs of homozygosity provide the farmer with information regarding possible relatedness in his flock as well as possible selection events that occurred previously or are occurring in the flock. Certain ROH that have remained in the population longer could be because the ROH may be associated with traits that could have a fitness advantage or be beneficial for production purposes and has thus remained in the population due to selection.

4.5 CONCLUSION

A complete and full understanding of the genetic resources available in the flock is important in order to make sound and efficient breeding and management decisions. There are a number of diversity statistics that will provide the farmer with a good understanding of the genetic components within the population. Knowledge of possible inbreeding present in the flock is important to know, as it could have an effect on production outputs as well as certain breeding decisions.

As shown by the analysis, the populations display low inbreeding. The Dorper and Namafr population indicated low within-population variation, while the Dorpersm 1 and Dorpersm 2 indicated relatively higher genetic variation within each population. This higher genetic variation within the smallholder populations could indicate the absence of set breeding strategies, and will ensure the populations have the ability to adapt to any possible climatic changes due to the high genetic variation. The genetic variability within each population was further verified by an excess of heterozygosity within the populations relative the expected heterozygosity. This excess heterozygosity could suggest genetic variation within the population.

High genetic differentiation was observed between the Namafr population and all the Dorper populations, which was expected due to the Namafr and Dorper being a different breed. The Dorper and Dorpersm 1 population had a lower F_{ST} than the Dorper and the Dorpersm 2 population. The Dorper and Dorpersm 1 population are geographically closer which could possibly explain the lowered genetic differentiation between these two populations. The F_{ST} range observed between the three Dorper populations (0.0231 – 0.0964), could indicate relatively high genetic differentiation between the Dorper populations. This suggests high genetic variability which implies that the Dorper breed as a whole has the genetic potential for selection regarding production traits as well as adaptation traits which would be beneficial to the farmer.

Although the Namafr breed presented the lowest inbreeding coefficient, the LD, heterozygosity as well as the PCA and F_{ST} all indicate low genetic diversity within the population which is consistent with other genetic diversity studies relating to the Namafr that also suggest low genetic diversity within the Namafr breed.

Runs of homozygosity in smallholder flock should be studied further, since ROH could possibly indicate selection for adaptation or robustness traits since smallholder sheep are often in extensive low-input production systems.

All these analyses provide a relatively complete description of the genetic material available in each population, and thus will assist with breeding and management decisions, as well as the conservation of genetic resources currently available.

4.6 REFERENCES

Al-Mamun, H., Clark, S., Kwan, P. & Gondro, C., 2015. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. Genet. Sel. Evol. 47, 1–14 https://doi.org/10.1186/s12711-015-0169-6.

Analla, M., Montilla, J. & Serradilla, J., 1999. Study of the variability of the response to inbreeding for meat production in Merino sheep. J. Anim. Breed. Genet. 116, 481–488 https://doi.org/0931–2668/99/1606–0481.

Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P. & Zondervan, K.T., 2010. Data quality control in genetic case-control association studies. Nat Protoc. 5, 1564–1573 https://doi.org/10.1038/nprot.2010.116.

Barczak, E., Wolc, A., Wójtowski, J., Ślósarz, P. & Szwaczkowski, T., 2009. Inbreeding and inbreeding depression on body weight in sheep. J. Anim. Feed Sci. 18, 42–50 https://doi.org/10.22358/jafs/66366/2009.

Barros, E., de A. Brasil, L., Tejero, J., Delgado-Bermejo, J. & Ribeiro, M., 2017. Population structure and genetic variability of the Segureña sheep breed through pedigree analysis and inbreeding effects on growth traits. Small Rumin. Res. 149, 128–133 https://doi.org/10.1016/j.smallrumres.2017.02.009.

Beynon, S., Slavov, G., Farré, M., Sunduimijid, B., Waddams, K., Davies, B., Haresign, W., Kijas, J., MacLeod, I., Newbold, C., Davies, L. & Larkin, D., 2015. Population structure and history of the Welsh sheep breeds determined by whole genome genotyping. BMC Genet. 16 https://doi.org/10.1186/s12863-015-0216-x.

Boichard, D., Maignel, L. & Verrier, É., 1997. The value of using probabilities of gene origin to measure genetic variability in a population. Genet. Sel. Evol. 29, 5–23 https://doi.org/10.1186/1297-9686-29-1-5.

Bosse, M., Megens, H., Madsen, O., Paudel, Y., Frantz, L., Schook, L., Crooijmans, R.P.M. & Groenen, M.A., 2012. Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape. PLoS Genet. 8, e1003100 https://doi.org/10.1371/journal.pgen.1003100.

Broman, K. & Weber, J., 1999. Long Homozygous Chromosomal Segments in Reference Families from theCentre d'E´tude du Polymorphisme Humain.pdf. Am. J. Hum. Genet. 65, 1493–1500.

Cloete, S.W., Snyman, M. & Herselman, M., 2000. Productive performance of Dorper sheep. Small Rumin. Res. 36, 119–135 https://doi.org/10.1016/S0921-4488(99)00156-X.

Diez-Tascón, C., Littlejohn, R., Almeida, P.A. & Crawford, A. 2000. Genetic variation within the Merino sheep breed: analysis of closely related populations using microsatellites. Anim. Genet. 31, 243–251.

Doney, J., 1957. Effects of inbreeding on four families of Peppin Merinos. Aust. J. Exp. Agric., 299–312 https://doi.org/10.1071/AR9570299.

Dorostkar, M., Arough, H., Shodja, J., Rafat, S., Rokouei, M. & Esfandyari, H., 2012. Inbreeding and inbreeding depression in Iranian Moghani sheep breed. J. agric. sci. technol. 14, 549–556.

Ercanbrack, S. & Knight, A., 1991. Effects of inbreeding on reproduction and wool production of Rambouillet, Targhee, and Columbia ewes. J. Anim. Sci. 69, 4734–4744.

Eteqadi, B., Hossein-Zadeh, N. & Shadparvar, A., 2014. Population structure and inbreeding effects on body weight traits of Guilan sheep in Iran. Small Rumin. Res. 119, 45–51 https://doi.org/10.1016/j.smallrumres.2014.03.003.

Eteqadi, B., Hossein-Zadeh, N. & Shadparvar, A., 2015. Inbreeding effects on reproductive traits in Iranian Guilan sheep. Trop Anim Heal. Prod. 47, 533–539 https://doi.org/10.1007/s11250-015-0758-y.

Falconer, D. & Mackay, T.F., 2009. Introduction to Quantitative Genetics. 4. ed. Pearson, Prentice Hall, Harlow.

Fernández, A., Toro, M. & López-Fanjul, C., 1995. The effect of inbreeding on the redistribution of genetic variance of fecundity and viability in Tribohum castaneum. Heredity. 75, 376–381 https://doi.org/10.1038/hdy.1995.149.

García-Gámez, E., Sahana, G., Gutiérrez-Gil, B. & Arranz, J., 2012. Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. BMC Genet. 13, 1–11 https://doi.org/10.1186/1471-2156-13-43.

Gholizadeh, M. & Ghafouri-Kesbi, F., 2016. Inbreeding depression in growth traits of Baluchi sheep. Small Rumin. Res. 144, 184–190 https://doi.org/10.1016/j.smallrumres.2016.09.012.

Gibson, J., Morton, N. & Collins, A., 2006. Extended tracts of homozygosity in outbred human populations. Hum. Mol. Genet. 15, 789–795 https://doi.org/10.1093/hmg/ddi493.

Gizaw, S., Van Arendonk, J.A., Komen, H., Windig, J. & Hanotte, O., 2007. Population structure, genetic variation and morphological diversity in indigenous sheep of Ethiopia. Anim. Genet. 38, 621–628 https://doi.org/10.1111/j.1365-2052.2007.01659.x.

Goddard, M., 2009. Fitness traits in animal breeding programs.in Adaptation and Fitness in Animal Populations: Evolutionary and Breeding Perspectives on Genetic Resource Management. van der Werf, J., Hans-Ulrich, G., Richard, F., Gondro, C., eds. Springer.

Goudet, J. & Jombart, T., 2015. Package 'hierfstat'. Evolution (N. Y). https://doi.org/10.2307/2411227.

Gowane, G., Prakash, V., Chopra, A. & Prince, L.L., 2013. Population structure and effect of inbreeding on lamb growth in Bharat Merino sheep. Small Rumin. Res. 114, 72–79 https://doi.org/10.1016/j.smallrumres.2013.06.002.

Goyache, F., Gutiérrez, J., Fernández, I., Gomez, E., Alvarez, I., Díez, J. & Royo, L., 2003. Using pedigree information to monitor genetic variability of endangered populations: The Xalda sheep breed of Asturias as an example. J. Anim. Breed. Genet. 120, 95–105 https://doi.org/10.1046/j.1439-0388.2003.00378.x.

Gregorius, H., 1978. The Concept of Genetic Diversity and Its Formal Relationship to Heterozygosity and Genetic Distance. Math. Biosci. 41, 253–271 https://doi.org/10.1016/0025-5564(78)90040-8.

Hamilton, M., 2011a. Population genetics. John Wiley & Sons, Chichester.

Hamilton, M., 2011b. Population structure and gene flow. Pages 105–153 in Population Genetics. Wiley-Blackwell.

Handley, L.J., Byrne, K., Santucci, F., Townsend, S., Taylor, M., Bruford, M. & Hewitt, G., 2007. Genetic structure of European sheep breeds. Heredity (Edinb). 99, 620–631 https://doi.org/10.1038/sj.hdy.6801039.

Hedrick, P., 2013. High inbreeding in sheep or erroneous estimation? J. Hered. 104, 298–299 https://doi.org/10.1093/jhered/ess139.

Hossein-Zadeh, N., 2013. Inbreeding Effects on Average Daily Gains and Kleiber Ratios in Iranian Moghani Sheep. Iran. J. Appl. Anim. Sci. 3, 545–551.

Keller, M., Visscher, P. & Goddard, M., 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. Genetics. 189, 237–249 https://doi.org/10.1534/genetics.111.130922.

Keller, L. & Waller, D., 2002. Inbreeding effects in wild populations. Trends Ecol. Evol. 17, 230–241.

Kevorkian, S.E.M., Georgescu, S.E., Manea, M.A., Zaulet, M., Hermenean, A.O. & Costache, M., 2010. Genetic diversity using microsatellite markers in four Romanian autochthonous sheep breeds. Rom. Biotechnol. Lett. 15, 5059–5065.

Khan, M., Ali, A., Hyder, A. & Chatta, A., 2007. Effect of inbreeding on growth and reproduction traits of Beetal goats. Arch. Anim. Breed. 50, 197–203.

Kijas, J., Porto-Neto, L., Dominik, S., Reverter, A., Bunch, R., McCulloch, R., Hayes, B., Brauning, R. & McEwan, J., 2014.

Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. Anim. Genet. 45, 754–757 https://doi.org/10.1111/age.12197.

Kijas, J., Townley, D., Dalrymple, B., Heaton, M., Maddox, J., McGrath, A., Wilson, P., Ingersoll, R., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Oddy, V., Nicholas, F. & Raadsma, H., 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS One. 4, e4668 https://doi.org/10.1371/journal.pone.0004668.

Kirin, M., McQuillan, R., Franklin, C., Campbell, H., Mckeigue, P. & Wilson, J., 2010. Genomic runs of homozygosity record population history and consanguinity. PLoS One. 5, e13996 https://doi.org/10.1371/journal.pone.0013996.

Lamberson, W., Thomas, D. & Rowe, K., 1982. The effects of inbreeding in a flock of Hampshire sheep. J. Anim. Sci. 55, 780–786.

Leocard, S., 2009. Selective sweep and the size of the hitchhiking set. Ann. Appl. Probab. 41, 731–764 https://doi.org/10.1239/aap/1253281062.

Li, M., Strandén, I. & Kantanen, J., 2009. Genetic diversity and pedigree analysis of the Finnsheep breed. J. Anim. Sci. 87, 1598–1605 https://doi.org/10.2527/jas.2008-0848.

Lutaaya, E., Misztal, I., Bertrand, J. & Mabry, J., 1999. Inbreeding in populations with incomplete pedigrees. J. Anim. Breed. Genet. 116, 475–480.

Luu, K., Blum, M., Privé, F., Bazin, E. & Duforet-frebourg, N., 2019. Package ' pcadapt '. 1–7 https://doi.org/10.1111/1755-0998.12592.

Marras, G., Gaspa, G., Sorbolini, S., Dimauro, C., Ajmone-Marsan, P., Valentini, A., Williams, J. & MacCiotta, N.P., 2014. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim. Genet. 46, 110–121 https://doi.org/10.1111/age.12259.

Mastrangelo, S., Ciani, E., Sardina, M., Sottile, G., Pilla, F. & Portolano, B., 2018. Runs of homozygosity reveal genome-wide autozygosity in Italian sheep breeds. Anim. Genet. 49, 71–81 https://doi.org/10.1111/age.12634.

Mastrangelo, S., Di Gerlando, R., Tolone, M., Tortorici, L., Sardina, M. & Portolano, B., 2014. Genome wide linkage disequilibrium and genetic structure in Sicilian dairy sheep breeds. BMC Genet. 15, 1–10 https://doi.org/10.1186/s12863-014-0108-5.

Mavrogenis, A., 1995. Breeding systems and selection strategies for sheep improvement in Cyprus. CIHEAM. 26, 17–26.

McRae, A., McEwan, J., Dodds, K., Wilson, T., Crawford, A. & Slate, J., 2002. Linkage disequilibrium in domestic sheep. Genetics. 160, 1113–1122.

Meadows, J.R., Chan, E.F. & Kijas, J., 2008. Linkage disequilibrium compared between five populations of domestic sheep. BMC Genet. 9, 1–10 https://doi.org/10.1186/1471-2156-9-61.

Mokhtari, M., Shahrbabak, M., Esmailizadeh, A., Shahrbabak, H. & Gutierrez, J., 2014. Pedigree analysis of Iran-Black sheep and inbreeding effects on growth and reproduction traits. Small Rumin. Res. 116, 14–20 https://doi.org/10.1016/j.smallrumres.2013.10.008.

Molotsi, A., 2017. Genomics as part of an integrated study of smallholder sheep farming systems in the Western Cape , South Africa.

Molotsi, A., Taylor, J., Cloete, S.W., Muchadeyi, F., Decker, J., Sandenbergh, L. & Dzama, K. 2017a. Preliminary genome-wide association study for wet-dry phenotype in smallholder ovine populations in South Africa. S. Afr. J. Anim. Sci. 47 https://doi.org/10.4314/sajas.v47i3.9.

Molotsi, A., Taylor, J., Cloete, S.W., Muchadeyi, F., Decker, J., Whitacre, L., Sandenbergh, L. & Dzama, K., 2017b. Genetic

diversity and population structure of South African smallholder farmer sheep breeds determined using the OvineSNP50 beadchip. Trop Anim Heal. Prod. 49, 1771–1777 https://doi.org/10.1007/s11250-017-1392-7.

Mueller, J., Rischkowsky, B., Haile, A., Philipsson, J., Mwai, O., Besbes, B., Zárate, A., Tibbo, M., Mirkena, T., Duguma, G., Sölkner, J. & Wurzinger, M., 2015. Community-based livestock breeding programmes: Essentials and examples. J. Anim. Breed. Genet. 132, 155–168 https://doi.org/10.1111/jbg.12136.

Naqvi, A., Mahmood, S., Vahidi, S.M., Abbas, S., Utsunomiya, Y., Garcia, J. & Periasamy, K., 2017. Assessment of genetic diversity and structure of major sheep breeds from Pakistan. Small Rumin. Res. 148, 72–79 https://doi.org/10.1016/j.smallrumres.2016.12.032.

Nei, M. & Wen-Hsiung, L., 1973. Linkage disequilibrium in subdivided populations. Genetics. 75, 213–219.

Nei, M., 1987. *Molecular evolutionary genetics*. Columbia university press.

Nielsen, R. & Signorovitch, J., 2003. Correcting for ascertainment biases when analyzing SNP data: Applications to the estimation of linkage disequilibrium. Theor. Popul. Biol. 63, 245–255 https://doi.org/10.1016/S0040-5809(03)00005-4.

Ohta, T., 1982. Linkage disequilibrium with the island model. Genetics. 101, 139–155.

Paiva, S., Facó, O., Faria, D., Lacerda, T., Barretto, G., Carneiro, P.L., Lobo, R.N. & McManus, C., 2011. Molecular and pedigree analysis applied to conservation of animal genetic resources: the case of Brazilian Somali hair sheep. Trop Anim Heal. Prod. 43, 1449–1457 https://doi.org/10.1007/s11250-011-9873-6.

Pedrosa, V., Santana, L., Oliveira, P., Eler, J. & Ferraz, J.B., 2010. Population structure and inbreeding effects on growth traits of Santa Inês sheep in Brazil. Small Rumin. Res. 93, 135–139 https://doi.org/10.1016/j.smallrumres.2010.05.012.

Peripolli, E., Munari, D., Silva, M.V.G., Lima, A.L., Irgang, R. & Baldi, F., 2017. Runs of homozygosity: current knowledge and applications in livestock. Anim. Genet. 48, 255–271 https://doi.org/10.1111/age.12526.

Peters, F., Kotze, A., van der Bank, F., Soma, P. & Grobler, J., 2010. Genetic profile of the locally developed Meatmaster sheep breed in South Africa based on microsatellite analysis. Small Rumin. Res. 90, 101–108 https://doi.org/10.1016/j.smallrumres.2010.02.005.

Pompanon, F., Bonin, A., Bellemain, E. & Taberlet, P., 2005. Genotyping errors: Causes, consequences and solutions. Nat. Rev. Genet. 6, 847–859 https://doi.org/10.1038/nrg1707.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Sklar, P., Maller, J., de Bakker, P.I., Daly, M. & Sham, P., 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. 81, 559–575 https://doi.org/10.1086/519795.

Purfield, D., Berry, D., McParland, S. & Bradley, D., 2012. Runs of homozygosity and population history in cattle. BMC Genet. 13, 1–11.

Purfield, D., McParland, S., Wall, E. & Berry, D., 2017. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. PLoS One. 12, 1–23 https://doi.org/10.1371/journal.pone.0176780.

Qwabe, S., van Marle-Köster, E. & Visser, C., 2013. Genetic diversity and population structure of the endangered Namaqua Afrikaner sheep. Trop Anim Heal. Prod. 45, 511–516 https://doi.org/10.1007/s11250-012-0250-x.

R., 2011. The R User Conference. In The R User Conference. Covetry, UK.

Reed, E., Nunez, S., Kulp, D., Qian, J., Reilly, M. & Foulkes, A., 2015. A guide to genome-wide association analysis and post-analytic interrogation. Stat. Med. 34, 3769–3792 https://doi.org/10.1002/sim.6605.

Sandenbergh, L., Cloete, S.W., Roodt-Wilding, R., Snyman, M. & Bester-van der Merwe, A., 2016. Evaluation of the

ovineSNP50 chip for use in four South African sheep breeds. S. Afr. J. Anim. Sci. 46, 90–93 https://doi.org/10.4314/sajas.v46i1.11.

Sandenbergh, L., Cloete, S.W., Roodt-Wilding, R., Snyman, M. & Van der Merwe, A., 2015. Genetic Diversity and Population Structure of four South African sheep breeds. Proc. Assoc. Advent. Breed. Gent. 21, 294–297.

Selvaggi, M., Dario, C., Peretti, V., Ciotola, F., Carnicella, D. & Dario, M., 2010. Inbreeding depression in Leccese sheep. Small Rumin. Res. 89, 42–46 https://doi.org/10.1016/j.smallrumres.2009.12.005.

Sheikhlou, M. & Abbasi, M., 2016. Genetic diversity of Iranian Lori-Bakhtiari sheep assessed by pedigree analysis. Small Rumin. Res. 141, 99–105 https://doi.org/10.1016/j.smallrumres.2016.07.009.

Slatkin, M., 2008. Linkage disequilibrium - Understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. 9, 477–485 https://doi.org/10.1038/nrg2361.

Swanepoel, J., Van Wyk, J., Cloete, S.W. & Delport, G., 2007. Inbreeding in the Dohne Merino breed in South Africa. S. Afr. J. Anim. Sci. 37, 176–179.

Thomas, D., Twyman, C., Osbahr, H. & Hewitson, B., 2007. Adaptation to climate change and variability: farmer responses to intra-seasonal precipitation trends in South Africa. Clim. Chang. 83, 301–322 https://doi.org/10.1007/s10584-006-9205-4.

Toro, M. & Caballero, A., 2005. Characterization and conservation of genetic diversity in subdivided populations. Phil. Trans. R. Soc. B. 360, 1367–1378 https://doi.org/10.1098/rstb.2005.1680.

Vostry, L., Milerski, M., Schmidova, J. & Vostra-Vydrova, H., 2018. Genetic diversity and effect of inbreeding on litter size of the Romanov sheep. Small Rumin. Res. 168, 25–31 https://doi.org/10.1016/j.smallrumres.2018.09.004.

Wakefield, J., 2010. Bayesian methods for examining Hardy-Weinberg equilibrium. Biometrics. 66, 257–265 https://doi.org/10.1111/j.1541-0420.2009.01267.x.

Weir, B., 2013. Interpreting Whole-Genome Marker Data. Stat Biosci. 5 https://doi.org/10.1007/s12561-013-9090-0.

Worley, K., Strobeck, C., Arthur, S., Carey, J., Schwantje, H., Veitch, A. & Coltman, D., 2004. Population genetic structure of North American thinhorn sheep (Ovis dalli). Mol. Ecol. 13, 2545–2556 https://doi.org/10.1111/j.1365-294X.2004.02248.x.

van Wyk, J., Erasmus, G. & Konstantinov, K., 1993. Inbreeding in the Elsenburg Dormer sheep stud. S. Afr. J. Anim. Sci. 23, 77–80.

van Wyk, J., Fair, M. & Cloete, S.W., 2009. Case study: The effect of inbreeding on the production and reproduction traits in the Elsenburg Dormer sheep stud. Livest. Sci. 120, 218–224 https://doi.org/10.1016/j.livsci.2006.10.005.

Yavarifard, R., Ghavi Hossein-Zadeh, N. & Shadparvar, A., 2014. Population genetic structure analysis and effect of inbreeding on body weights at different ages in Iranian Mehraban sheep. J Anim Sci Technol. 56, 1–9 https://doi.org/10.1186/s40781-014-0031-3.

Zhang, Q., Guldbrandtsen, B., Bosse, M., Lund, M. & Sahana, G., 2015. Runs of homozygosity and distribution of functional variants in the cattle genome. BMC Genom. 16, 1–16 https://doi.org/10.1186/s12864-015-1715-x.

Zhu, M., Zhu, B., Wang, Y., Wu, Y., Xu, L., Guo, L., Yuan, Z., Zhang, L., Gao, X., Gao, H., Xu, S. & Li, J., 2013. Linkage disequilibrium estimation of Chinese beef simmental cattle using high-density SNP panels. Asian-Australas. J. Anim. Sci. 26, 772–779 https://doi.org/10.5713/ajas.2012.12721.

CHAPTER 5
Functional classification of genes overlapping CNVs identified in South African smallholder sheep using the OvineSNP50 Beadchip

5.1 ABSTRACT

Smallholder farmers often make use of low-input systems, suggesting that robust and adaptable individuals are needed in these systems that have good production and reproduction in these low-input systems. One of the reasons certain individuals may be more adaptable or have higher production outputs could be due to the presence of advantageous mutations or genetic structural variants. Genetic variants, namely copy number variations (CNVs), are structural changes to the DNA and are larger than a single nucleotide. In this study, 47 sheep were investigated for the presence of CNVs. A total of 206 CNVs passed quality control. These CNVs were compared to the NCBI RefSeq Ovis aries: Oar_v4.0 to identify candidate genes located within or overlapping the copy number variations identified. Gene annotation analysis was carried out on the identified candidate genes. Gene annotation assigned the candidate genes to two gene groups. The first gene group were protein coding genes responsible for interferons that are the natural defences individuals have against viral and bacterial infection. The second gene group was found to be responsible for a variety of biological functions including transport, metabolic precursors, neurogenesis, signalling as well as bone and cartilage matrix composition along with a number of other important functions.

This indicates that CNVs could have various effects on important biological process which could possibly influence an individual's survival or even production and reproduction. This highlights the need for CNV studies to determine the influence of these CNVs and how they can be utilised in breeding programmes to improve adaptation and production outputs.

Keywords: genomics, genetics, sheep, copy number variations, adaptability, smallholder, Dorper

5.2 INTRODUCTION

Population structure and cluster analysis are visual representations of how populations have diverged over time, due to adaptation or even selection. With climate change, animals need to adapt to a changing environment in order to survive. To ensure effective production, an adaptable breed is needed that will not negatively be effected by changes to the environment (Friggens *et al.,* 2017; Mormède *et al.,* 2011). While many indigenous and locally developed breeds, such as the Meatmaster, Namaqua Afrikaner and Dorper, are naturally better suited to certain environments, selection still provides a way to move the population mean

in a more beneficial genotypic direction, which could be more suited to certain production systems or even certain climates or environments (Peters *et al.*, 2010; Soma *et al.*, 2012) .

Understanding how certain desired traits evolved enables breeders to select for certain production characteristics more effectively. Kijas *et al.* (2009) suggests that genome-wide association analysis may provide a way of highlighting genetic regions and mutations that may influence disease phenotypes and production traits. According to Bidwell *et al.* (2009), sheep possess the most identified genetic mutations associated with reproduction compared to any livestock species, and many of these mutations have been mapped to specific chromosomal regions (Montgomery *et al.*, 1994; Mulsant *et al.*, 2001; Rohrer, 2004; McNatty *et al.*, 2005).

Genetic variation is often as a result of mutations that occur at the DNA level. Mutations are the permanent incorporation of random errors in the DNA that results in differences between ancestral and descendent copies of DNA (Hamilton, 2011a). The ultimate source of genetic variability is mutations, but the probability that a new mutation will reach a reasonable frequency depends on the mutation rate, the population size and the current mode of selection for the trait (Franklin, 1980). Mutations can result in the divergence of a population, however, it also has the positive effect of increasing the genetic variance within a population, which is especially important with regards to surviving and adapting to a changing environment (Hamilton, 2011d). While some mutations may have negative effects, certain mutations could be advantageous in a certain environment or for specific production, and so natural or artificial selection occurs due to the advantageous or disadvantageous mutation (Orr, 1998; Nielsen, 2005; Feuk *et al.*, 2006b; Hamilton, 2011a). Determining whether these mutations are present in a flock is important genetic knowledge which could be beneficial if incorporated into an effective breeding strategy that is in line with the production goals.

Due to the improvement of these high throughput techniques, new structural variants have been identified, copy number variations (CNVs) (Liu *et al.*, 2013). These are structural differences in the DNA (Thapar & Cooper, 2013), and are a result of mutations in the DNA code. These copy number variations range in size from 1 kb to several mega-bases long suggesting they have the potential to affect a much larger portion of the genome than SNPs, which only affect a single nucleotide pair (Redon *et al.*, 2006; Ramayo-Caldas *et al.*, 2010). A variable number of nucleotides make up a gene, and often a variable number of genes are responsible for a specific phenotype (Glazier *et al.*, 2002). The presence of CNVs has led to unequal recombination or rearrangement of chromosomes which has been found to result in some diseases (Giglio *et al.*, 2001; Osborne *et al.*, 2001; Iafrate *et al.*, 2004; Shaw & Lupski, 2004). Gene expression is the basis of many important biological functions in cells, and the influence of SNP and CNV variants are an indication of the nature of the mutational and natural selection processes that contribute to genetic diversity and divergence (Stranger *et al.*, 2007).

Significant correlation was identified between the copy number of *DEFB4* and its mRNA expression levels in lymphoblastoid cell lines (Hollox *et al.*, 2003). The resulting peptides are effective antimicrobial agents,

and *DEFB4* has effectively linked the innate and adaptive immune responses by acting as a cytokine in humans. Further studies are needed to determine which immune responses could be influenced or associated with CNVs in sheep. In a study conducted by Hou *et al.* (2012), CNVs were identified that were associated with and possibly contributed to parasite resistance in Angus cattle. This CNV association with disease resistance in Angus raises the interest of possible future studies to determine whether there may be CNV associations with diseases in other species, namely sheep (Suaŕez-Vega *et al.*, 2013). This also increases the interest in determining whether CNVs may also have an influence on desired production traits in sheep. Screening individuals for CNVs that could possibly be associated with production traits such as carcass characteristics, reproduction traits like increased ovulation or multiple progeny, or disease resistance and internal parasite resistance will have a beneficial influence on the selection of individuals. Being able to select individuals that possess these desirous traits will have an impact on production inputs and outputs as well as the possibility of genetic gain. Studies are needed to find possible CNVs in the sheep genome that may be economically beneficial to farmers and will assist with increasing reproduction, production and genetic gain. Since there has been limited studies in sheep to identify CNVs, this preliminary identification of CNVs in a smallholder sheep flock from Beaufort-West should urge further CNV association studies to identify beneficial CNVs for production systems in South Africa.

5.3 MATERIALS AND METHODS

5.3.1 DESCRIPTION OF STUDY LOCATION

Blood samples were collected from 48 individuals from a smallholder farm just outside of Beaufort-West. A thorough description of the study location is provided in chapter three of this thesis under 3.3.1) Description of study location. The Dorper, Namafr and Dorpersm 1 populations from chapter 4 were not included in CNV identification due to a lack of genotype outputs from the genotyping process, that are needed for further CNV analysis.

5.3.2 RESEARCH POPULATION AND SAMPLING METHOD

The 48 blood samples were collected by a certified animal technician. A full description of the research population and sampling method used to collect the blood samples can be found in chapter three of this thesis under 3.3.2 Research population and sampling method.

5.3.3 DNA EXTRACTION AND GENOTYPING

DNA was extracted using a Sbeadex livestock kit. Extracted DNA was genotyped at the Agricultural Research Council (ARC) lab in Pretoria with the assistance of qualified personnel. A description of the genotyping process can be found in chapter three of this thesis under 3.3.3 DNA extraction and Genotyping.

5.3.4 STATISTICAL ANALYSIS

Plink

Quality control and pruning was conducted in *Plink* using the functions --mind, --maf, --geno, --hwe (Purcell *et al.*, 2007), with thresholds of mind = 0.1, maf = 0.05, geno = 0.1 and hwe = 0.001. Quality control was performed on 48 sheep, however *Plink* removed sample 136 due to missing genotype data. The genotyping rate for the remaining samples was 0.9495. During pruning, 3279 variants were removed due to missing genotype data (--geno), 242 variants were removed due to Hardy-Weinberg exact test (--hwe) and 6183 variants were removed due to minor allele threshold (--maf). A total of 44537 variants and 47 sheep passed the filters and quality control thresholds to be used in further analysis.

PennCNV

CNV analysis of the 47 samples was carried out with the PennCNV program (Wang et al., 2007) using a Perlscript. PennCNV utilises the Hidden Markov Model (HMM) in order to identify CNVs. HMM is a statistical technique that follows a Markov process, where the probability of observing a specific state at a specific point in time depends on the state at previous time points; and so to detect CNVs, a first-order HMM is used that assumes the hidden copy number state at each SNP only depends on the copy number state of nearby markers (Wang et al., 2007, 2008).

Input files

The Log R ratio, SNP allelic ratio distribution (B allele frequency), chromosome and position were retrieved from the genotyping report provided by Illumina Genotyping procedure. Signal intensity files were created for each sample using the *split_illumina_report.pl* script in PennCNV. A pfb file was created for the combined samples using the *compile_pfb.pl* script. Only SNPs that passed the quality control in *PLINK* were included in the construction of the pfbfile for the CNV analysis.

CNV Detection

CNVs were detected in the combined samples using the *detect_cnv.pl* script and the *–test* option. CNVs were determined by default for all chromosomes except for the X-chromosome where the *-chrx* option was used to identify CNVs on the X-chromosome in PennCNV. A total of 368 CNVs were detected on the autosomes while 57 CNVs were detected on the sex chromosomes.

CNV Filter and Quality Control

These detected CNVs were then filtered in PennCNV using the *filter_cnv.pl* script. Samples were filtered according to the following thresholds; a standard deviation (SD) of LRR <0.30, a BAF drift <0.01 and a waviness factor value between -0.05 and 0.05. LRR is the Log R Ratio which is a measure of the normalized intensity of the SNP, and is calculated as the $\log_2(R_{observed}/R_{expected})$, and the $R_{expected}$ is calculated from the linear interpolation of the canonical genotype clusters (Wang et al., 2007). R is the signal intensity A + signal intensity B, where each SNP has two alleles which is referred to as the A and B allele (Wang *et al.*, 2008). LRR signifies the relative abundance of genomic DNA that surrounds the SNP and is expected to correlate with the copy number status (Lin *et al.,* 2013). BAF is the B Allele Frequency which can be described as the

normalized measure of the allelic intensity ratio of allele A and allele B (Wang *et al.*, 2007), and reflects the relative abundance of the B allele intensity (Lin *et al.*, 2013). After filtering 206 CNVs passed the set thresholds on the autosomes and the sex chromosomes.

Gene Annotation and Classification

CNVs that passed the filters and quality control were then compared with the NCBI RefSeq Ovis aries: Oar_v4.0 (Pruitt *et al.,* 2007) (http://www.ncbi.nlm.nih.gov/RefSeq/) to identify possible genes (candidate genes) that overlapped with the identified copy number variations. Identified candidate genes were then analysed in DAVID Bioinformatics Resources (Dennis *et al.*, 2003) (https://david.ncifcrf.gov/). Some definitions for the analysis carried out in DAVID include; annotation category is a group of annotation sources that collect similar biological questions such as "pathways", the annotation source is an independent database in a category such as "BioCarta Pathways" and lastly, term is a detailed item in an annotation source, such as the p53 signalling pathway in BioCarta (https://david.ncifcrf.gov/content.jsp?file=functional_annotation.html). Functional annotation clustering and gene functional classification was carried out on the candidate genes that overlapped with identified CNVs. The GeneID.txt of these candidate genes was submitted on the DAVID platform and the gene list was identified as ENTREZ_GENE_ID, and the species was set as *Ovis aries*. Thresholds for the functional annotation clustering and gene functional clustering were set at the default values for the respective tests.

For the functional annotation clustering tool the thresholds were as follows; Similarity Term Overlap = 3, Similarity Threshold = 0.50, Initial Group Membership = 3, Final Group Membership = 3, Multiple Linkage Thresholds = 0.50 and EASE = 1.0. Gene functional classification thresholds were as follows; Similarity Term Overlap = 4, Similarity Threshold = 0.35, Initial Group Membership = 4, Final Group Membership = 4 and Multiple Linkage Threshold = 0.50.

5.4 RESULTS AND DISCUSSION

5.4.1 CNV IDENTIFICATION

A total of 44537 variants and 47 sheep passed the filters and quality control thresholds in *Plink*. These variants were then used for further CNV analysis in PennCNV. Detection of CNVs in PennCNV identified 368 CNVs on the autosomes of the 47 individuals and 57 CNVs on the sex-chromosomes, resulting in a total identification of 425 CNVs before filtering the 47 samples.

Samples were filtered resulting in a total of 206 CNVs that passed filtering and quality control, 56 on the sex-chromosome and 150 on the autosomes. The proportion of CNVs identified per individual is shown in the *Supplementary Material, 7) CNV Analysis, Figure 1.7.1.* CNVs from autosomes as well as X-chromosome were combined and used for further analysis. *Table 5.1* describes the copy number, the number of CNVs, the average length as well as the chromosome these CNVs are found on.

***Table 5.1**. The number of CNVs and the average length of CNVs on the various chromosomes of the sheep genome.*

*Copy Number	Number CNVs	Average Length (kb)	Chromosomes
0	-	-	-
1	42	160,48	1, 2, 4, 6, 8, 9, 10, 12, 15, 16, 17, 18, 22, X
2	47	2047,24	X
3	116	183,02	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 15, 16, 17, 18, 20, 21, 22, X
4	1	150,35	1

**Copy number; 0=deletion of two copies, 1=deletion of one copy, 2=copy-neutral with LOH, 3=single copy duplication, 4=double copy duplication*

The copy number indicates the type of polymorphism that caused the identified CNV (Table 5.1). CNVs that are copy neutral with Loss of heterozygosity (LOH), indicates the region has normal LRR values, but without the heterozygote cluster (Wang *et al.*, 2007). As seen in *Table 5.1* the majority of the CNVs (116) are copy number 3, indicating a single copy duplication while 47 CNVs are copy number 2 which classifies it as copy-neutral with LOH. All the copy number 2 events identified occurred on the X-chromosome. This suggests that loss of heterozygosity is possibly occurring on the X-chromosome. Zhu *et al.* (2015) identified selection signatures on the X-chromosome for a number of genes. Further association studies are needed to determine whether this LOH is due to selection of specific traits.

***Table 5.2** CNVs identified in (n=36) individuals on all chromosomes and autosomes*

	CNVs	Length (kb)
Total for autosomes and sex chromosomes	206	124341,21
Average	5,72	603,60
Total excluding X-Chromosome	150	27058,66
Average per individual excluding X-Chromosome	4,17	180,39

Table 5.2 describes the total sheep used for CNV identification as well as the total CNVs identified and the total length of all the identified CNVs. It was found that each individual had 5,72 CNVs on average with an average length of 603,60kb. The average length of the CNVs on chromosome X were the longest compared to the other chromosomes. The longer CNVs identified on the X-chromosome can be due to selection of traits that may be present on the chromosome (Zhu *et al.*, 2015). As selection occurs, the haplotypes will be passed on to the next generation, which could result in longer lengths of haplotypes in the offspring (Kirin *et al.*, 2010).

Table 5.3 *The number of CNVs identified in different studies and their average length*

Study	Nr individuals	Nr CNVs	Average length of CNVs
This study	36	150	180,39
Fontanesi *et al.* (2011)	11	186	73,9
Liu *et al.* (2013)	329	256	144,6
Ma *et al.* (2015)	160	173	117,82
Yang *et al.* (2018)	2254	24 558	140,76

Table 5.3 compares the number of CNVs and the average length of CNVs in this study with identified CNVs and the average length determined for CNVs in four other studies.

Yang *et al.* (2018) reported definite differences in the CNV occurrence between the populations from different geographical areas, indicating that there has possibly been adaptation or divergence of certain breeds according to the environment the breeds are found in. We can thus assume that due to the different selection pressures experienced in different countries under different environments, certain CNVs may be more prevalent or have undergone selection to make the individual more suited to its environment. Some CNVs may be beneficial to certain production systems or certain environmental areas such as South Africa if identified CNV are associated with beneficial production or adaptation traits. However, CNV association studies in sheep are needed to determine the validity of selecting for beneficial CNVs.

Copy number variations were found on 20 out of the 27 chromosomes in the sheep genome as shown by *Figure 5.1* that illustrates the distribution of the identified CNVs across the different chromosomes in the ovine genome. Further studies are needed to determine if CNVs are affecting genes on the X-chromosome of sheep since in this study the majority of CNVs appear to occur on the X-chromosome. Determining whether specific production or reproduction traits are being influenced by CNVs on the X-chromosome is important genetic information that could assist with breeding strategies and selection decisions, since genes affecting ovulation rate have been mapped to the X-chromosome in sheep (Davis *et al.*, 1991, 2001).

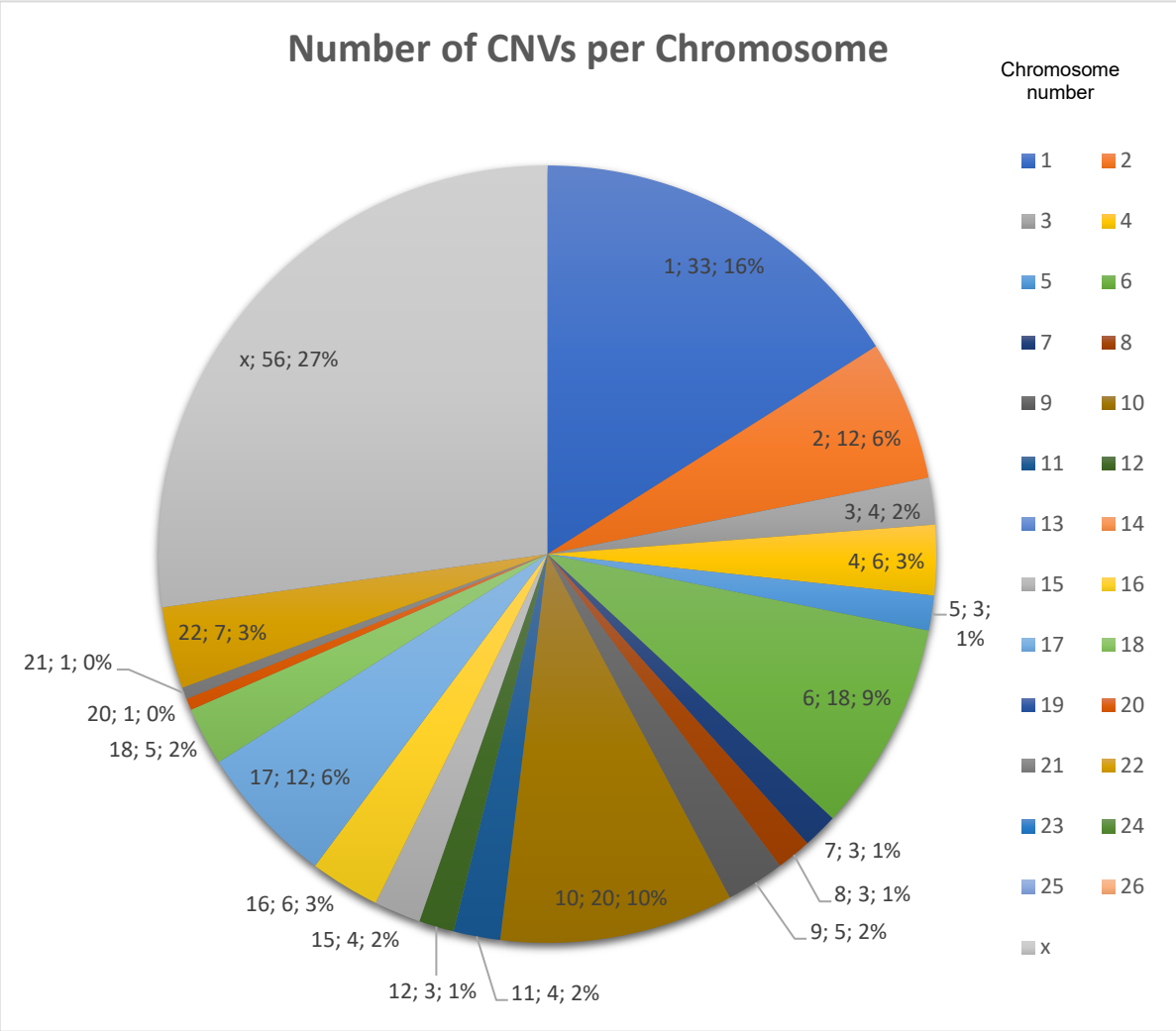

Figure 5.1 *Distribution of CNVs across the chromosomes in the ovine genome.*

Table 5.4 *Distribution of the average length of the CNVs identified in this study and two other studies*

	<100kb	100kb – 300kb	300kb – 500kb	500kb – 1000kb	>1000kb
CNV lengths identified in this study	25%	45%	6%	11%	13%
CNV lengths identified by Liu *et al.* (2013)	25%	58%		-	-
CNV lengths identified by Ma *et al.* (2015)	53,76%	45,66%		0,58%	-

Table 5.4 illustrates the proportion and length of the CNVs identified in this study and two other studies. The number of CNVs in this study that are between 100 kb – 500 kb make up 51% of the identified CNVs which is similar to the CNVs of the same length identified in other studies. This study identified more CNVs that are longer than 500kb compared to the other two studies. The longer average length of CNVs identified in this study indicates that these CNVs could possibly have an effect on a large portion of the genome and

thus more genes and so more phenotypes since a variable number of nucleotides often make up a gene (Glazier *et al.*, 2002; Redon *et al.*, 2006; Ramayo-Caldas *et al.*, 2010).

Studies are needed to determine the reason why this study identified more CNVs that are longer than 500kb. Further studies could assist in determining whether CNVs with longer average length are more prevalent in smallholder populations, or South African sheep in general, since the studies conducted by Liu *et al.* (2013) and Ma *et al.* (2015) both included Dorper populations.

5.4.2 GENE ANNOTATION

A total of 331 candidate genes were identified that overlapped the identified CNVs. Illustrations of the identified CNVs on each chromosome can be found under the *Supplementary Material, 8 CNV Identification, Figure 1.8.1 – Figure 1.8.20.*

Of the 331 Gene ID's that were submitted for analysis in DAVID, 6 Gene ID's were unmapped, namely; 5605336, 1121246, 5605334, 1107484, 1107046 and 1111497. Genes were grouped into 20 clusters with enrichment scores ranging between 0.08 - 2.95. A total of 73 terms were not clustered due to possible orphan or irrelevant genes, the genes may be below the similarity thresholds or there may be too few genes to form a functional group based on the minimal final cluster threshold (https://david.ncifcrf.gov/content.jsp?file=functional_annotation.html). Each cluster contains a group of terms that have similar biological meaning due to sharing similar gene members.

Table 5.5 *The enrichment score of functional annotated clusters as identified using the GO annotation/ DAVID*

Functional Annotation Clusters	Enrichment Score	Functional Annotation Clusters	Enrichment Score
Cluster 1	2.95	Cluster 11	0.45
Cluster 2	1.77	Cluster 12	0.35
Cluster 3	1.52	Cluster 13	0.34
Cluster 4	1.21	Cluster 14	0.25
Cluster 5	1.16	Cluster 15	0.21
Cluster 6	1.09	Cluster 16	0.19
Cluster 7	0.71	Cluster 17	0.16
Cluster 8	0.54	Cluster 18	0.14
Cluster 9	0.51	Cluster 19	0.13
Cluster 10	0.47	Cluster 20	0.08

Table 5.5 provides the enrichment scores of each annotation cluster. The enrichment score is an indication of the importance of the cluster or gene group in the total gene list. The enrichment score gives an indication as to which group of genes are more prevalent in the study compared to background genes (population genes), it can thus be assumed that the enriched genes may have important functions in the study (Huang *et*

al., 2009). Cluster one had the highest enrichment score and the terms that were included in cluster one are; chloride channel calcium-activated, calcium-activated chloride channel protein, intracellular calcium activated chloride channel activity, pancreatic secretion, VWA, renin secretion and von Willebrand type A. Genes associated with these terms had the highest enrichment score and could possibly have a higher prevalence in this study than would be expected compared to the population genes. These genes may thus have functional importance and further studies are needed to determine the importance of the genes associated with the enriched terms within the study population.

Heatmaps for each cluster were also produced. The heatmap for the functional annotation cluster 5 can be seen in *Figure 5.2*. The genes on the right-hand side of the heatmap are the genes that were submitted into DAVID from the gene list, while the terms below the heatmap are the annotation terms. The heatmap illustrates which genes are associated with which annotation terms. The green blocks indicate which genes are associated with an annotation term, while the black block indicates that no association has been reported between the respective gene and annotation term. In cluster 5 the interferon alpha-H and -2 and the interferon beta-2 were associated with the majority of the annotation terms within the cluster, suggesting they are functionally similar to the annotation term they are associated with (Huang *et al.*, 2009).

As seen in *Figure 5.2* the majority of the genes have not been reported to be associated with the term annotation. It is possible that more studies are needed in order to confirm or deny associations between genes and annotation terms, which confirms the need for further CNV association studies to fill the knowledge gap with regard to the unreported possible associations between a number of genes and association terms. Heatmaps for the other 19 clusters can be found in the *Supplementary Material, 9) Gene Annotation, Figure 1.9.1 – Figure 1.9.19*.

Gene functional classification was carried out on the gene list to determine the major gene functional groups in the submitted gene list. A total of 41 genes were assigned to two cluster groups, while 279 genes from the submitted gene list were not included in the output. The gene functional classification tool assigned the genes included in the output to two group clusters. The first gene group had an enrichment score of 1.1 while the second gene group had an enrichment score of 0.45. In *Figure 5.3.a and Figure 5.3.b* the heatmaps for the two gene functional classification clusters are illustrated. Almost all the genes are associated with the annotation terms in both cluster one and two, however, there are a couple of genes in both clusters that have not yet been associated with the respective annotation terms.

The gene functional classification tool in DAVID was used to better display the major biological functions the candidate genes were responsible for. Gene group one consisted of 28 term records that are responsible for similar biological processes, and four of the candidate genes are included in this gene group (Huang *et al.*, 2009). The genes assigned to gene group one were all interferon alpha and beta genes. A list of the exact genes in gene group one can be found in the *Supplementary Material, 9) Gene Annotation, Table 1.9.1.*

Interferons are ubiquitous cytokines (Preedy & Hunter, 2011) produced by mononuclear cell types in response to infections caused by DNA or RNA viruses (Mays, 1981). They are responsible for the stimulation

or inhibition of different genes that encode proteins involved in antiviral and bacterial defence mechanisms, inflammation, apoptosis, adaptive immunity, angiogenesis as well as a number of other processes in humans (De Andrea *et al.*, 2002; Mendelsohn, 2008; Meyer, 2009). Many genes have not been fully investigated in sheep species, and for this reason many of the gene functions stated in this thesis are based on the human genome. Thus, it is evident in this study that CNVs occur in regions where genes are responsible for immune defence mechanisms as well as a number of other functions as illustrated in *Figure 5.3.b.*

Further association studies are needed to determine whether CNVs found in these regions affect the regular functioning of these identified genes, as well as the possible beneficial or detrimental effects that these genetic alterations in the form of deletions and duplication could have on the functioning of the immune system of the individual. Understanding the possible effect CNVs could have on the immune systems of individuals through their effect on interferons could lead to the eventual formation of selection programmes for individuals that could be resistant to certain viral or bacterial infections. This selection of resistant individuals based on CNVs would be economically and genetically beneficial to livestock farmers through the improvement of livestock production and could assist in the welfare of animals.

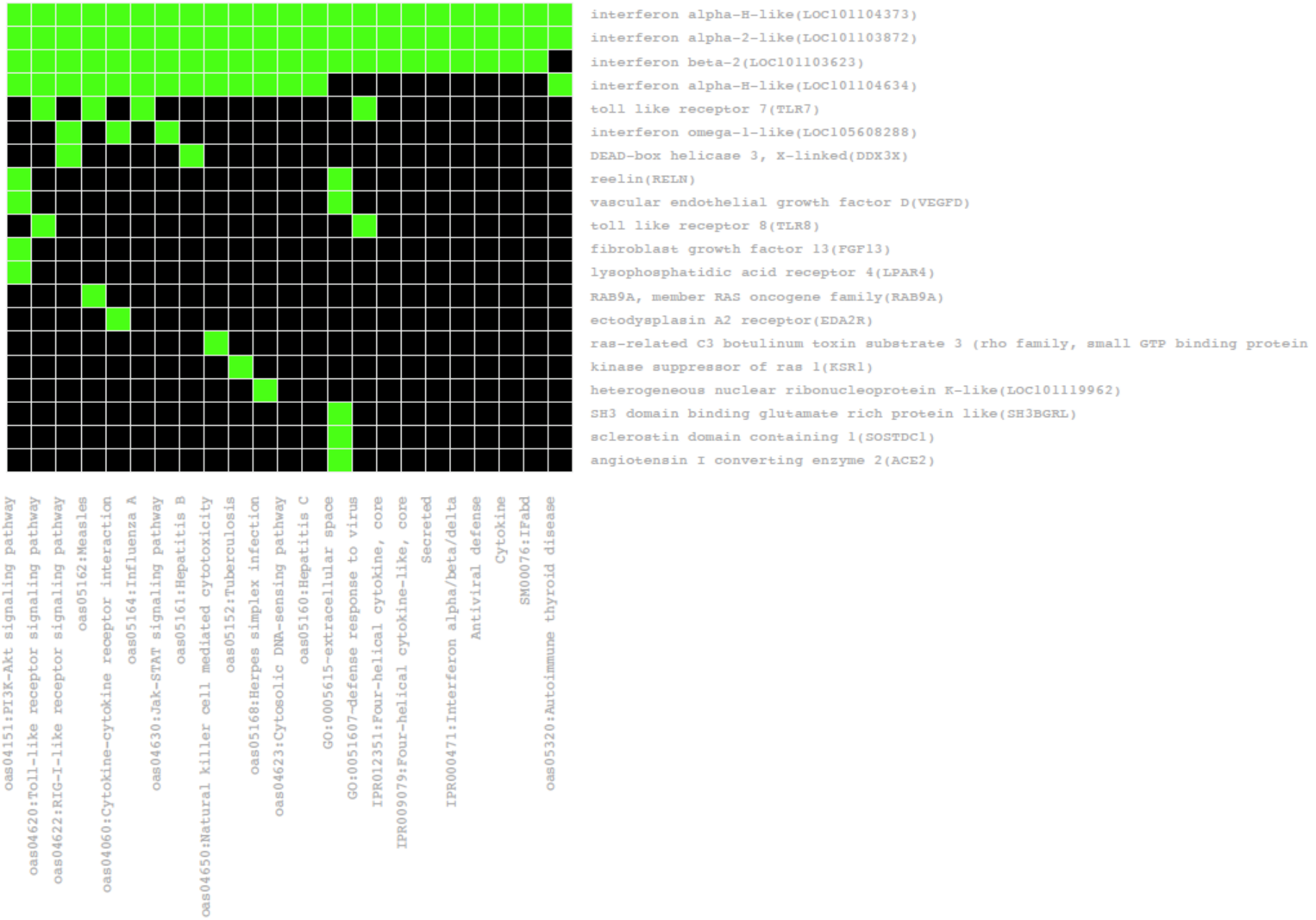

Figure 5.2 *Heatmap of cluster 5 from the functional annotation clustering tool in DAVID for genes overlapping with identified CNVs. (■ = corresponding gene-term association positively reported, ■ = corresponding gene-term association not reported yet.)*

Figure 5.3.a *Heatmaps of the gene functional classification clusters for gene group 1 (■ = corresponding gene-term association positively reported, ■ = corresponding gene-term association not reported yet.)*

A full list of the genes assigned to gene group two can be found in the *Supplementary Material, 9) Gene Annotation, Table 1.9.1.* The genes in the gene group two were found to be responsible for transport, sensory perception, metabolic precursors, neurogenesis, blastocyst hatching, smooth muscle cell contraction, epithelial cell proliferation as well as bone and cartilage matrix composition along with a number of other functions. *TMC1* is a gene reported to be part of a gene family that encodes transmembrane proteins (Kurima *et al.*, 2002). While the function of the *TMC1* gene region is still unknown, an in-frame deletion of 57 amino acids has been found that includes part of the TMC domain, which is encoded by the *Tmcl*^{dn} allele of the deafness (*dn)* mutant mouse strain (*l*) (Labay *et al.,* 2010). This 57 amino acid deletion could possibly be caused by a copy number deletion, so more investigation is needed to determine the possible effects CNV mutations could have on the transmembrane channel-like gene 1 (TMC1) that have been found to cause dominant or recessive hearing loss in humans and mice. Determining whether this mutation could have a similar effect in sheep could be beneficial to survival in some production systems. Sheep that are deaf due to possible mutations of the *TMC1* could have a lower survival due to the fact they will not be able to hear predators or other dangers (Jolly *et al.*, 2004). This is especially relevant in some low-input smallholder systems as well as extensive production systems where sheep graze large areas with very little human contact.

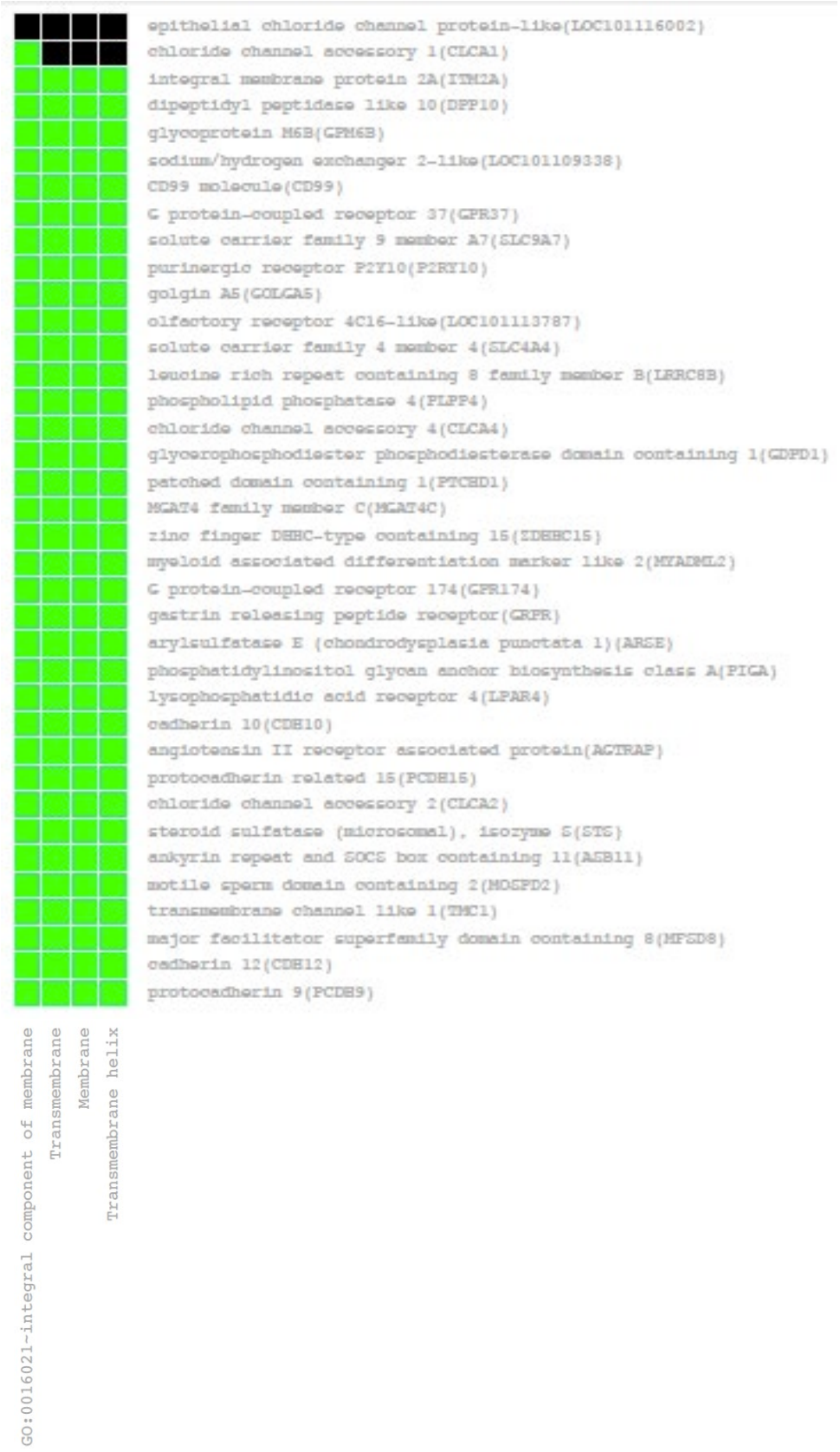

Figure 5.3.b *Heatmaps of the gene functional classification clusters for gene group 2 (■ = corresponding gene-term association positively reported, ■ = corresponding gene-term association not reported yet.)*

Copy number variations were found to overlap the genetic region that codes for the steroid sulfatase (STS). Steroid sulfatase belongs to the sulfatase family which serve as metabolic precursor for estrogens, androgens and cholesterol (Reed *et al.*, 2005). STS is responsible for hydrolysis of estrone sulfate and dehydroepiandrosterone sulfate to estrone and dehydroepiandrosterone, and can be converted to steroids with estrogenic properties which have been found to stimulate tumour growth (Reed *et al.,* 2005). Further investigation should be carried out to determine whether CNVs could possibly influence the expression and functioning of the steroid sulfatase which in turn could result in a decrease of tumour growth that would be advantageous to the welfare of sheep and thus have a beneficial effect on sheep production.

CNVs were found in the region of the *ITM2A* (integral membrane protein 2A) gene. *ITM2A* is a novel type II integral membrane protein that plays an important role in osteo- and chondrogenic differentiation (Coli *et al.,* 2001). Deleersnijder et al. (1996) identified the gene *ITM2A* as a useful marker of osteogenic differentiation, and has been used as marker genes for chondrogenic/osteoblastic cells in bone formation (Tuckermann *et al.,* 2000). On extensive smallholder farms based in the Karoo, sheep travel long distances for feed and water. This means that sheep need to have sturdy legs in order to cover the large distances without injury. Identifying CNVs that could possibly affect osteo- and chondrogenic differentiation could be beneficial in selecting individuals that are more robust for less than ideal environments.

The gastrin releasing peptide receptor (GRPR) was found to overlap with identified CNVs. Whitley *et al.* (1998; 2000) discovered gastrin-releasing peptides (GRP) in the endometrium of pregnant sheep and women, suggesting a hormonal role for the endometrial GRP-like peptide in foetal development. CNVs have been identified in the region that codes for the gastrin releasing peptide receptor, meaning that CNVs could possibly influence the effect of the gastrin releasing peptide through possible genetic changes to the receptor coding which could affect its functioning. Fertility is one of the most important aspects of farming. Identifying CNVs that could have a positive effect on foetal development or any part of the gestation period could be economically beneficial to the farmer since more sheep will be fertile and more lambs will be produced.

Copy number variations are an important source of genomic structural variation that needs to be identified and further studied to determine its validity as markers to investigate phenotypic and economic traits that may be beneficial to producers as well as consumers (Salomón-Torres *et al.*, 2015).

5.5 CONCLUSION

CNVs in diploid organisms can cause a deletion on the one chromosome while a duplication could occur on the other homologous chromosome, therefore, it is important to know the chromosome specific copy number for the development of linkage and association tests of CNVs (Wang *et al.,* 2008).

In this study, a number of CNVs were identified, and 41 genes that overlapped the identified CNVs were assigned to two gene groups using a gene functional classification tool. Gene group one included interferon genes that play an important role in the immune defence mechanism, while gene group two consisted of a number of different genes that are important for biological functions such as transport, metabolic precursors,

signalling and even bone and cartilage matrix composition along with a number of other important function. As one of the identified gene groups included a number interferons that play an important role in immune defence, there is an area for further study to determine whether the identified CNVs could have positively influenced the health status of the individuals on the farm, as they did not receive daily health care due to the extensive production system. Identifying CNVs that have an influence on the health status of individuals will have economic benefits for the farmer as less animals would be sick and less money would need to be spent on medication or other medical procedures.

Further studies in CNV identification is required to identify genes that are affected by CNVs and to determine the possible effect they could have on complex and desired traits for production. Additional studies will also determine whether these CNVs could be used for selection of superior individuals with regards to adaptation to specific production systems or environments or for certain desirous production outputs that would assist in improving the sheep industry.

5.6 REFERENCES

De Andrea, M., Ravera, R., Gioia, D., Gariglio, M. & Landolfo, S., 2002. The interferon system: An overview. Eur J Paediatr Neurol. 6 https://doi.org/10.1053/ejpn.2002.0573.

Bidwell, C., Cockett, N., Maddox, J. & Beever, J., 2009. Genome Mapping and Genomics in Domestic Animals (N. Cockett & C Kole, Eds.). 3rd ed. Springer.

Coli, R., Anderson, S.L., Volpi, S.A. & Rubin, B.Y., 1999. Genomic organization and chromosomal localization of the Itm2a gene. Mamm. Genome. 10, 81–89 https://doi.org/10.1016/S0378-1119(01)00737-5.

Davis, G., Dodds, K., Wheeler, R. & Jay, N., 2001. Evidence That an Imprinted Gene on the X Chromosome Increases Ovulation Rate in Sheep. Biol. Reprod. 64, 216–221 https://doi.org/10.1095/biolreprod64.1.216.

Davis, G., McEwan, J., Fennessy, P., Dodds, K. & Farquhar, P., 1991. Evidence for the Presence of a Major Gene Influencing Ovulation Rate on the X Chromosome of Sheep. Biol. Reprod. 44, 620–624.

Deleersnijder, W., Hong, G., Cortvrindt, R., Poirier, C., Tylzanowski, P., Pittois, K., Van Marck, E. & Merregaert, J., 1996. Isolation of markers for chondro-osteogenic differentiation using cDNA library subtraction. J. Biol. Chem. 271, 19475–19482 https://doi.org/10.1074/jbc.271.32.19475.

Dennis, G., Sherman, B., Hosack, D., Yang, J., Gao, W., Lane, H. & Lempicki, R., 2003. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 4 https://doi.org/10.1186/gb-2003-4-5-p3.

Feuk, L., Marshall, C., Wintle, R. & Scherer, S., 2006. Structural variants: changing the landscape of chromosomes and design of disease studies. Hum. Mol. Genet. 15, 57–66 https://doi.org/10.1093/hmg/ddl057.

Fontanesi, L., Beretti, F., Martelli, P., Colombo, M., Dall'Olio, S., Occidente, M., Portolano, B., Casadio, R., Matassino, D. & Russo, V., 2011. A first comparative map of copy number variations in the sheep genome. Genomics. 97, 158–165 https://doi.org/10.1016/j.ygeno.2010.11.005.

Franklin, I., 1980. Evolutionary change in small populations. Pages 135–149 in Conservation Biology: An evolutionary-ecological perspective. Soule, M., Wilcox, B., eds. Sinauer Associates, Inc., Massachusetts.

Friggens, N., Blanc, F., Berry, D. & Puillet, L., 2017. Review: Deciphering animal robustness. A synthesis to facilitate its use in livestock breeding and management. Animal. 11, 2237–2251 https://doi.org/10.1017/S175173111700088X.

Giglio, S., Broman, K., Matsumoto, N., Calvari, V., Gimelli, G., Neumann, T., Ohashi, H., Voullaire, L., Larizza, D., Giorda, R., Weber, J., Ledbetter, D. & Zuffardi, O., 2001. Olfactory Receptor–Gene Clusters, Genomic-Inversion Polymorphisms, and Common Chromosome Rearrangements. Am. J. Hum. Genet. 68, 874–883 https://doi.org/10.1086/319506.

Glazier, A., Nadeau, J. & Aitman, T., 2002. Genetics: Finding genes that underline complex traits. Science. 298, 2345–2349 https://doi.org/10.1126/science.1076641.

Hamilton, M., 2011a. Population genetics. John Wiley & Sons, Chichester.

Hamilton, M., 2011b. Molecular evolution. Pages 235–282 in Population Genetics. Wiley-Blackwell.

Hollox, E.J., Armour, J.A.L. & Barber, J.C.K., 2003. Extensive Normal Copy Number Variation of a β-Defensin Antimicrobial-Gene Cluster. Am. J. Hum. Genet. 73, 591–600.

Hou, Y., Liu, G., Bickhart, D., Matukumalli, L., Li, C., Song, J., Gasbarre, L., Van Tassell, C. & Sonstegard, T., 2012. Genomic regions showing copy number variations associate with resistance or susceptibility to gastrointestinal nematodes

in Angus cattle. Func Integr Genomics. 12, 81–92 https://doi.org/10.1007/s10142-011-0252-1.

Huang, D., Sherman, B. & Lempicki, R., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 4, 44–57 https://doi.org/10.1038/nprot.2008.211.

Iafrate, A., Feuk, L., Rivera, M., Listewnik, M., Donahoe, P., Qi, Y., Scherer, S. & Lee, C., 2004. Detection of large-scale variation in the human genome. Nat. Genet. 36, 949–951 https://doi.org/10.1038/ng1416.

Jolly, R., Blair, H., & Johnstone, A., 2004. Genetic disorders of sheep in New Zealand: A review and perspective. N Z Vet J. 52, 52–64 https://doi.org/10.1080/00480169.2004.36405.

Kijas, J., Townley, D., Dalrymple, B., Heaton, M., Maddox, J., McGrath, A., Wilson, P., Ingersoll, R., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Oddy, V., Nicholas, F. & Raadsma, H., 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS One. 4, e4668 https://doi.org/10.1371/journal.pone.0004668.

Kirin, M., McQuillan, R., Franklin, C., Campbell, H., Mckeigue, P. & Wilson, J., 2010. Genomic runs of homozygosity record population history and consanguinity. PLoS One. 5, e13996 https://doi.org/10.1371/journal.pone.0013996.

Kurima, K., Peters, L., Yang, Y., Riazuddin, S., Ahmed, Z., Naz, S., Arnaud, D., Drury, S., Mo, J., Makishima, T., Ghosh, M., Menon, P.S., Deshmukh, D., Oddoux, C., Ostrer, H., Khan, S., Riazuddin, S., Deininger, P., *et al.,* 2002. Dominant and recessive deafness caused by mutations of a novel gene, TMC1, required for cochlear hair-cell function. Nat. Genet. 30, 277–284 https://doi.org/10.1038/ng842.

Labay, V., Weichert, R., Makishima, T. & Griffith, A., 2010. Topology of Transmembrane Channel-like Gene 1 Protein. Biochemistry. 49, 8592–8598 https://doi.org/10.1021/bi1004377.

Lin, C., Naj, A. & Wang, L., 2013. Analyzing copy number variation using SNP array data: Protocols for calling CNV and association tests. Curr Protoc Hum Genet. 79, 1–17 https://doi.org/10.1002/0471142905.hg0127s79.

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. BMC Genom. 14, 1–11.

Ma, Y., Zhang, Q., Lu, Z., Zhao, X. & Zhang, Y., 2015. Analysis of copy number variations by SNP50 BeadChip array in Chinese sheep. Genomics. 106, 295–300 https://doi.org/10.1016/j.ygeno.2015.08.001.

Mays, L., 1981. Genetics: A Molecular Approach. Macmillan Publishing Co., Inc., New York.

McNatty, K., Galloway, S., Wilson, T., Smith, P., Hudson, N., O'Connell, A., Bibby, A., Heath, D., Davis, G., Hanrahan, J. & Juengel, J., 2005. Physiological effects of major genes affecting ovulation rate in sheep. Genet. Sel. Evol. 37, S25-s38 https://doi.org/10.1051/gse:2004029.

Mendelsohn, J., 2008. Interferons. Pages 739–752 in The Molecular Basis of Cancer. Fourth. Saunders/Elsevier, Philadelphia.

Meyer, O., 2009. Interferons and autoimmune disorders. Jt. Bone Spine. 76, 464–473 https://doi.org/10.1016/j.jbspin.2009.03.012.

Montgomery, G., Lord, E., Penty, J., Dodds, K., Broad, T., Cambridge, L., Sunden, S.L., Stone, R. & Crawford, A., 1994. The Booroola Fecundity (FECB) Gene Maps to Sheep Chromosome 6. Genomics. 22, 148–153.

Mormède, P., Foury, A., Terenina, E. & Knap, P., 2011. Breeding for robustness: The role of cortisol. Animal. 5, 651–657 https://doi.org/10.1017/S1751731110002168.

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., Cognie, Y., Chitour, N. & Elsen, J., 2001. Mutation in bone morphogenetic protein receptor-IB is associated with increased ovulation rate in Booroola Merino ewes. Proc.

Natl. Acad. Sci. U.S.A 98, 5104–5109 https://doi.org/10.1073/pnas.091577598.

Nielsen, R., 2005. Molecular Signatures of Natural Selection. Annu. Rev. Genet. 39, 197–218 https://doi.org/10.1146/annurev.genet.39.073003.112420.

Orr, H., 1998. The Population Genetics of Adaptation: the Distribution of Factors Fixed During Adaptive Evolution. Evol. Int. J. 52, 935–949.

Osborne, L., Li, M., Pober, B., Chitayat, D., Bodurtha, J., Mandel, A., Costa, T., Grebe, T., Cox, S., Tsui, L. & Scherer, S., 2001. A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. Nat. Genet. 29, 321–325 https://doi.org/10.1038/ng753.

Peters, F., Kotze, A., van der Bank, F., Soma, P. & Grobler, J., 2010. Genetic profile of the locally developed Meatmaster sheep breed in South Africa based on microsatellite analysis. Small Rumin. Res. 90, 101–108 https://doi.org/10.1016/j.smallrumres.2010.02.005.

Preedy, V. & Hunter, R., 2011. Cytokines (J Hiscott & C Ware, Eds.). Boca Raton.

Pruitt, K., Tatusova, T. & Maglott, D., 2007. NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 35, 61–65.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Sklar, P., Maller, J., de Bakker, P.I., Daly, M. & Sham, P., 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. 81, 559–575 https://doi.org/10.1086/519795.

Ramayo-Caldas, Y., Castelló, A., Pena, R., Alves, E., Mercadé, A., Souza, C., Fernández, A., Perez-Enciso, M. & Folch, J., 2010. Copy number variation in the porcine genome inferred from a 60 k SNP BeadChip. BMC Genom. 11.

Redon, R., Ishikawa, S., Fitch, K., Feuk, L., Perry, G., Andrews, T., Fiegler, H., Shapero, M., Carson, A., Chen, W., Cho, E., Dallaire, S., Freeman, J., González, J., Gratacòs, M., Huang, J., Kalaitzopoulos, D., *et al.,* 2006. Global variation in copy number in the human genome. Nature. 444, 444–454.

Reed, M., Purohit, A., Woo, L.W., Newman, S. & Potter, B.V., 2005. Steroid sulfatase: Molecular biology, regulation, and inhibition. Endocr. Rev. 26, 171–202 https://doi.org/10.1210/er.2004-0003.

Rohrer, G., 2004. An overview of genomics research and its impact on livestock reproduction. Reprod. Fertil. Dev. 16, 47–54 https://doi.org/10.1071/rd03077.

Salomón-Torres, R., González-Vizcarra, V., Medina-Basulto, G., Montaño-Gómez, M., Mahadevan, P., Yaurima-Basaldúa, V., Villa-Angulo, C. & Villa-Angulo, R., 2015. Genome-wide identification of copy number variations in Holstein cattle from Baja California, Mexico, using high-density SNP genotyping arrays. Genet. Mol. Res. 14, 11848–11859 https://doi.org/10.4238/2015.

Shaw, C. & Lupski, J., 2004. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum. Mol. Genet. 13, R57–R64 https://doi.org/10.1093/hmg/ddh073.

Soma, P., Kotze, A., Grobler, J. & van Wyk, J., 2012. South African sheep breeds: Population genetic structure and conservation implications. Small Rumin. Res. 103, 112–119 https://doi.org/10.1016/j.smallrumres.2011.09.041.

Stranger, B., Forrest, M., Dunning, M., Ingle, C., Beazley, C., Thorne, N., Redon, R., Bird, C., Grassi, A., Lee, C., Tyler-smith, C., Carter, N., Scherer, S., Tavaré, S., Deloukas, P., Hurles, M. & Dermitzakis, E., 2007. Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes. Science. 315, 848–853 https://doi.org/10.1126/science.1136678.

Suaŕez-Vega, A., Gutieŕrez-Gil, B., Cuchillo-Ibañez, I., Saéz-Valero, J., Pérez, V., Garciá-Gámez, E., Benavides, J. & Arranz, J.J., 2013. Identification of a 31-bp deletion in the RELN gene causing Lissencephaly with Cerebellar Hypoplasia

in sheep. PLoS One. 8, e81072 https://doi.org/10.1371/journal.pone.0081072.

Thapar, A. & Cooper, M., 2013. Copy number variation: What is it and what has it told us about child psychiatric disorders? J. Am. Acad. Child Adolesc. Psychiatry. 52, 772–774.

Tuckermann, J., Pittois, K., Partridge, N., Merregaert, J. & Angel, P., 2000. Collagenase-3 (MMP-13) and Integral Membrane Protein 2a (Itm2a) are marker Genes of Chondrogenic/Osteoblastic cells in bone formation: Sequential Temporal, and Spatial Expression of Itm2a, Alkaline Phosphatase, MMP-13, and Osteocalcin in the mouse. J Bone Min. Res. 15, 1257–1265.

Wang, K., Chen, Z., Tadesse, M., Glessner, J., Grant, S., Hakonarson, H., Bucan, M. & Li, M., 2008. Modeling genetic inheritance of copy number variations. Nucleic Acids Res. 36, e138 https://doi.org/10.1093/nar/gkn641.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F., Hakonarson, H. & Bucan, M., 2007. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 17, 1665–1674 https://doi.org/10.1101/gr.6861907.

Whitley, J., Giraud, A., Mahoney, A., Clarke, I. & Shulkes, A., 2000. Tissue-specific regulation of gastrin-releasing peptide synthesis, storage and secretion by oestrogen and progesterone. J. Endocrinol. 166, 649–658.

Whitley, J., Shulkes, A., Salamonsen, L., Vogiagis, D., Familari, M. & Giraud, A., 1998. Temporal expression and cellular localization of a gastrin-releasing peptide-related gene in ovine uterus during the oestrous cycle and pregnancy. J. Endocrinol. 157, 139–148.

Yang, L., Xu, L., Zhou, Y., Liu, M., Wang, L., Kijas, J., Zhang, H., Li, L. & Liu, G., 2018. Diversity of copy number variation in a worldwide population of sheep. Genomics. 110, 143–148 https://doi.org/10.1016/j.ygeno.2017.09.005.

Zhu, C., Fan, H., Yuan, Z., Hu, S., Zhang, L., Wei, C., Zhang, Q., Zhao, F. & Du, L., 2015. Detection of selection signatures on the X chromosome in three sheep breeds. Int. J. Mol. Sci. 16, 20360–20374 https://doi.org/10.3390/ijms160920360

CHAPTER 6
Conclusion and Recommendations

6.1 CONCLUSION

The aim of this study was to characterise the genetic diversity and possible preliminary CNVs present in smallholder sheep that could facilitate breeding management programmes using the relevant population genetic information.

Through the *SEQUOIA* package in Rstudio, a better understanding of the familial relationships between the individuals within the flock was achieved. Eleven dams were assigned and four sires were assigned as parents to individuals. Ten dams were assigned as possible relatives with likely parent-offspring (PO) relationships, and nine sires were assigned as possible relatives with likely parent-offspring (PO) relationships. More individuals were assigned as dams and sires when the assigned pedigree from the *SEQUOIA* package was used alongside the pedigree assigned by the pairwise IBD estimations. Combining the two methods provided a more complete pedigree of the flock. Many of the individuals in the parent generation are the founders of the population. Due to this there were some individuals that had no assigned parents.

Knowledge of the familial relationships will assist the farmer in making informed breeding decisions with regards to which individuals can be bred without influencing the inbreeding within the population. The farmer will be able to breed individuals according to the desired outcome with regard to increasing or decreasing the inbreeding. In communal breeding systems or multi-sire systems, pedigree assignment using genomic data will be effective in determining which rams contributed more to the gene pool (Gizaw *et al.*, 2014a; b). Knowledge of which rams sired which individuals could assist with selection decisions. As indicated by the constructed pedigree, ram 9943 sired more offspring than ram 16008. This will notify the farmer to possible problems regarding the rams. Rams that do not contribute to the gene pool could be due to the fact that the ram may have a physical problem, or there could be a problem with his sperm regarding motility or even concentration. This will ensure the farmer will know that a replacement ram may be needed.

Continuous record-keeping and pedigree information is important for genetic gain and to limit the breeding of related individuals. Record-keeping is a good alternative if genomic assignment of the pedigree is unavailable to smallholder farmers. Further studies are also needed to determine the possibility of a smaller SNP panel for genomic assignment of pedigree that would be cheaper and more financially accessible to smallholder farmers.

Genetic diversity parameters for four sheep populations were calculated in order to compare the genetic diversity within and between the four flocks. These diversity parameters included inbreeding, linkage disequilibrium and heterozygosity. This provided a better understanding of the genetic diversity found within and between the different populations. Genetic diversity and variation are important for future selection

decisions. High genetic variation provides a large genetic base from which different traits can be selected that are in line with the breeding and production goals of the farmer. In populations that have low genetic variation and individuals are genetically similar or identical there will not be any individuals that may be superior for certain traits eg., production traits (wool production). With no superior individuals, selection for a trait will be ineffective or result in little to no genetic gain for the trait of interest because all the individuals will have the same production. Farmers thus need to ensure they are aware of the genetic diversity within their flock in order to fully utilise the genetic potential of the individuals within the flock. Flocks that possess low genetic variation for traits desired by the farmer, should have breeding programmes that will increase of the genetic variation of the trait of interest. This can be done by keeping less replacement ewes from the flock and selecting replacement ewes from a different breeder or farmer. A different ram could also contribute to increasing the variation through gene flow.

As represented by the PCA and the calculated F_{ST} values, the Namafr population did not cluster close to the Dorper population, which was expected due to the different breeds. The PCA and F_{ST} indicated the Dorper subpopulations clustered together as expected since they are all from the same breed. The F_{ST} of the Dorper subpopulations indicate a relatively large genetic differentiation between the subpopulations and thus suggest a large genetic base from which adaptation and production traits could be selected.

CNVs were identified in the smallholder flock from Beaufort-West in order to study genes that might be affected by CNVs due to the overlap of these genes with the respective CNVs.

CNVs present in a smallholder flock were identified and functional clustering and gene classification analysis was performed on genes that overlapped the identified CNVs. Gene classification analysis identified two groups of genes that overlapped with the CNVs. Gene group one contained interferons which have important immune defence functions. The second gene group contained genes with a number of functions including; transport, metabolic precursors, sensory perception, smooth muscle contraction, bone and cartilage matrix composition along with a number of other important functions. Functional clustering assigned genes to 20 clusters each with important functions.

Further research is needed to identify CNVs that may be associated with specific production or reproduction traits that could be economically beneficial to farmers. Additional studies should be able to determine whether CNVs can be incorporated into breeding programmes as a kind of marker for marker assisted selection (MAS) (van der Werf, 2007).

6.2 RECOMMENDATIONS

Complete breeding and pedigree records should be kept by farmers to ensure effective breeding decisions are made. Smaller and cheaper SNP panels for pedigree assignment should be investigated that could be used by smallholder farmers. Funding from the government and developmental agencies should be encouraged that will assist with the advancement of the smallholder production systems through the use of new technologies such as genomics. There is also a need for research institutes to assist and partner with

smallholder farmers to come up with innovative strategies that could help farmers overcome the challenges they may be facing, and to improve the genetic gain and production outputs through correct breeding programmes and strategies.

Increasing the number of individuals included in future genetic diversity studies for the Namafr would provide a more accurate estimate of the inbreeding coefficient within the population. The development of a custom ovine SNP panel should be investigated that is focused on South African indigenous breeds. This could increase the accuracy of genotyping indigenous breeds as well as decrease the occurrence of possible ascertainment bias due to the indigenous breeds being under-represented during the development of the Ovine50SNP Beadchip (Kijas *et al.*, 2009; Sandenbergh *et al.*, 2016; Molotsi, 2017).

Some studies have investigated CNVs in world-wide sheep populations (Yang *et al.*, 2018), however, there are still many CNVs that have not been investigated or identified to date. Identifying CNVs in more smallholder populations will give a more accurate representation of possible CNVs that may be present in smallholder flocks. CNVs in these populations may be associated with adaptation traits or robustness traits due to the low-input production systems often used in smallholder communities. Association studies are needed to determine what possible effect identified CNVs can have on gene functions and then to further determine whether these identified CNVs are positively or negatively correlated to the overlapping gene. Knowledge of how these CNVs possibly affect gene functions will provide a better understanding of whether CNVs could be incorporated into breeding programmes as markers for MAS.

6.3 REFERENCES

Gizaw, S., van Arendonk, J.A., Valle-Zárate, A., Haile, A., Rischkowsky, B., Dessie, T. & Mwai, A., 2014a. Breeding programmes for smallholder sheep farming systems: II. Optimization of cooperative village breeding schemes. J. Anim. Breed. Genet. 131, 350–357 https://doi.org/10.1111/jbg.12102.

Gizaw, S., Rischkowsky, B., Valle-Zárate, A., Haile, A., van Arendonk, J.A., Mwai, A. & Dessie, T., 2014b. Breeding programs for smallholder sheep farming systems: I. Evaluation of alternative designs of breeding schemes. J. Anim. Breed. Genet. 131, 341–349 https://doi.org/10.1111/jbg.12101

Kijas, J., Townley, D., Dalrymple, B., Heaton, M., Maddox, J., McGrath, A., Wilson, P., Ingersoll, R., McCulloch, R., McWilliam, S., Tang, D., McEwan, J., Cockett, N., Oddy, V., Nicholas, F. & Raadsma, H., 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLoS One. 4, e4668 https://doi.org/10.1371/journal.pone.0004668.

Molotsi, A., 2017. Genomics as part of an integrated study of smallholder sheep farming systems in the Western Cape , South Africa.

Sandenbergh, L., Cloete, S.W., Roodt-Wilding, R., Snyman, M. & Bester-van der Merwe, A., 2016. Evaluation of the ovineSNP50 chip for use in four South African sheep breeds. S. Afr. J. Anim. Sci. 46, 90–93 https://doi.org/10.4314/sajas.v46i1.11.

Yang, L., Xu, L., Zhou, Y., Liu, M., Wang, L., Kijas, J., Zhang, H., Li, L. & Liu, G., 2018. Diversity of copy number variation in a worldwide population of sheep. Genomics. 110, 143–148 https://doi.org/10.1016/j.ygeno.2017.09.005.

SUPPLEMENTARY MATERIAL

1 DNA EXTRACTION

DNA was extracted using a sbeadex livestock kit; catalogue number 44701 and 44702. The first step was the Lysis protocol, where protocol A: Blood (standard) was used, followed by the extraction protocol and finally elution of the DNA. Any changes to the protocol are recorded below:

Extraction 1 - 50 µl eluate was used in step 3.3.4.1.

Extraction 2 - 30 µl eluate was used in step 3.3.4.1.

Extraction 3 - 20 µl eluate was used in step 3.3.4.1.

Extraction 4 - 150 µl blood was used from each sample in step 3.3.2.1.1. 150 µl Binding buffer SB and 5 µl sbeadex particles + EDTA was used instead of the standard binding mix in step 3.3.3.1. 20 µl eluate was used in step 3.3.4.1.

Extraction 5 - Samples were incubated at 56°C for 40 minutes in step 3.3.2.1.2. 20 µl eluate was used in step 3.3.4.1. Elution of DNA was carried out at 56°C in step 3.3.4.3.

Extraction 6 - Samples were incubated at 56°C for an hour in step 3.3.2.1.2. 20 µl eluate was used in step 3.3.4.1. Elution of DNA was carried out at 56°C for 15 minutes in step 3.3.4.3.

Table 1.1 *Smallholder sample DNA quality and concentration as measured by Nano machine and DNA concentration determined through the qubit protocol (ng/ul = the concentration DNA found in sample, 260/280= used as an indicator of the quality of the DNA found in the sample)*

Sample ID	ng/ul	260/280	Qubit
7275	62.7	1.93	1.97
130	71.5	1.84	59.1
117	61.2	1.80	5.32
115	33.9	1.85	1.10
132	73.2	1.64	36.2
74	47.1	2.19	1.32
14249	65.48	1.85	9.05
139	79.2	1.79	61.9
16008	55.0	1.91	4.49
16065	47.8	2.03	18.1
7277	41.1	1.76	12.0
14261	50.0	1.87	2.99
7283	33.2	1.95	-
14267	85.7	1.79	18.6
122	36.1	1.97	2.41
14255	21.2	1.81	-
7276	90.0	1.78	15.2
137	53.8	1.84	26.2
9943	39.6	1.77	-
14293	53.6	1.79	35.7

136	19.5	2.61	-
7290	36.7	2.09	-
110	59.4	2.00	58.2
128	34.3	1.73	4.73
14250	55.5	1.77	1.63
22	58.6	1.75	19.9
14280	46.31	1.78	1.12
14292	20.16	1.83	-
14290	55.8	1.79	33.8
14286	57.9	1.71	-
120	25.3	2.00	-
14270	87.2	1.86	58.8
7299	31.7	1.81	-
7284	40.33	2.05	13.0
14297	108.8	1.81	84.6
129	46.6	1.91	15.2
123	40.8	1.87	-
7289	54.4	1.76	37.0
7301	58.5	1.81	-
7302	24.1	2.09	-
57	44.1	1.71	18.3
7271	73.7	1.80	3.36
14276	55.5	1.81	1.06
14295	121.3	1.75	80.6
14258	56.9	1.85	15.1
100	40.69	1.45	-
14279	96.60	1.95	35.9
113	78.2	1.81	36.3

2 QUBIT PROTOCOL

The Qubit® dsDNA HS (High Sensitivity) Assay Kit was used to determine the qubit concentrations of 48 DNA samples. Each tube was labelled on the lid so as not to interfere with the sample read. 2µl DNA was added to 198µl prepared standard working solution to equal a final volume of 200µl. Samples were then incubated at room temperature for two minutes.

3 LONZO GEL

Extracted DNA samples were run on a gel to confirm the presence of DNA using a FlashGel™ DNA Starter Kit. Solutions were prepared according to FlashGel protocol. A 100-4000bp DNA ladder was used. A total of 2 µl distilled water was added to the well on the gel plate, and 2µl DNA solution was mixed with 3µl dye. The voltage was set at 375V and 400A. Seven runs were performed with differing times and dye concentrations; however, the voltage and amps were kept constant for all runs performed. *Figures 1.3.1 – 1.3.7* illustrate the

results of the performed FlashGel, the individual ID is indicated at the top of the figure above the run, with the DNA concentration in brackets below the individual ID as determined by the Nano machine.

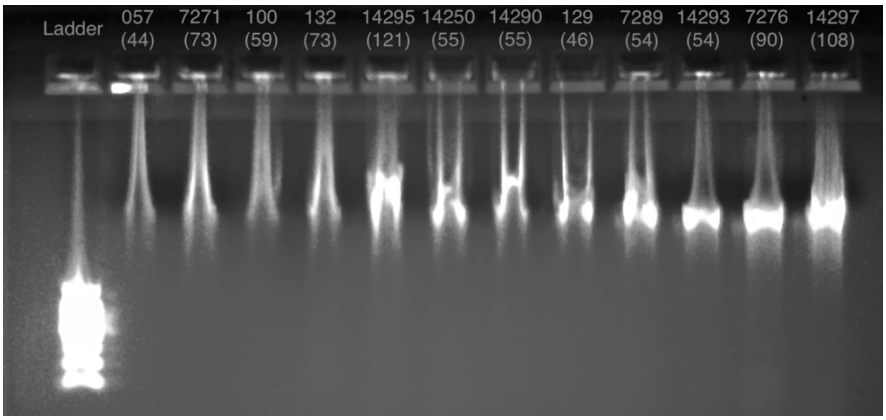

Figure 1.3.1 *FlashGel 1, 100% dye was used, run for five minutes*

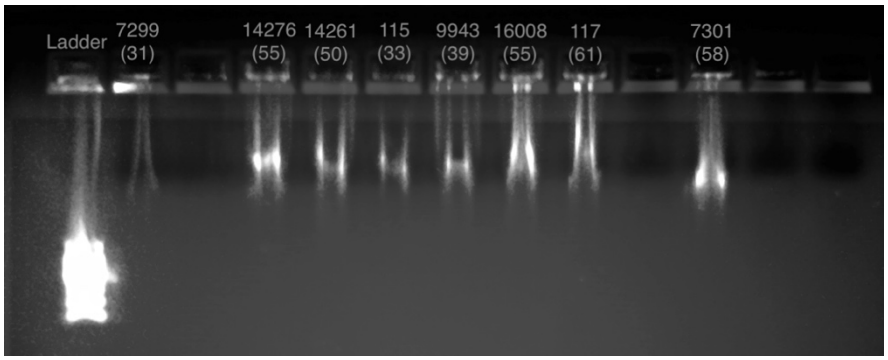

Figure 1.3.2 *FlashGel 2, 40% dye was used, run for four minutes*

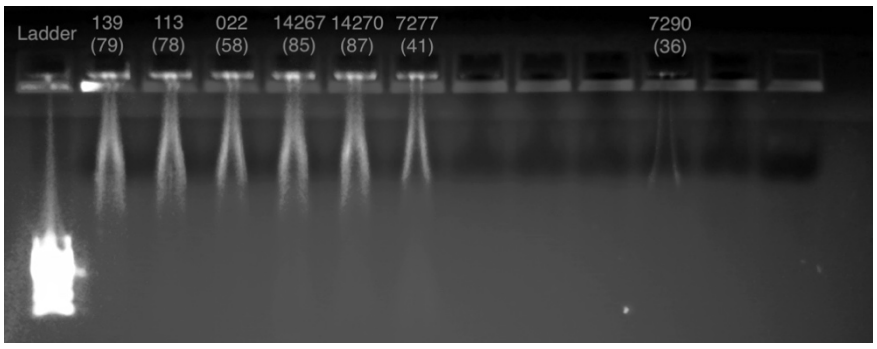

Figure 1.3.3 *FlashGel 3, 60% dye was used, run for four minutes*

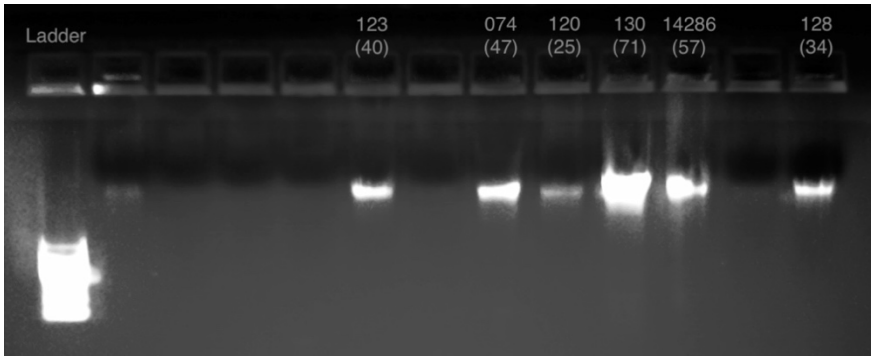

Figure 1.3.4 *FlashGel 4, 100% dye was used, run for four minutes*

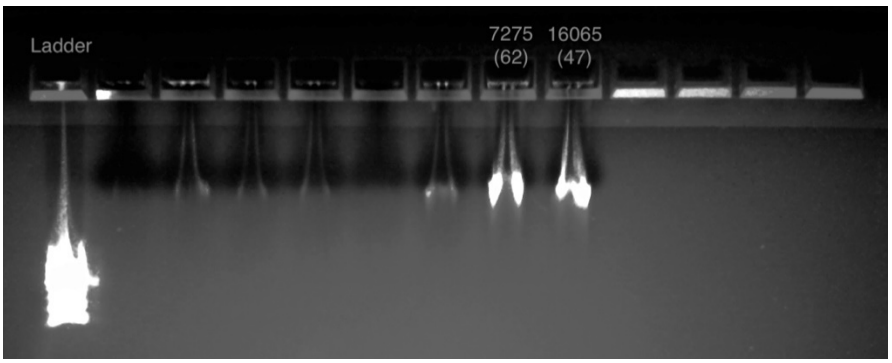

Figure 1.3.5 *FlashGel 5, 100% dye was used, run for four minutes*

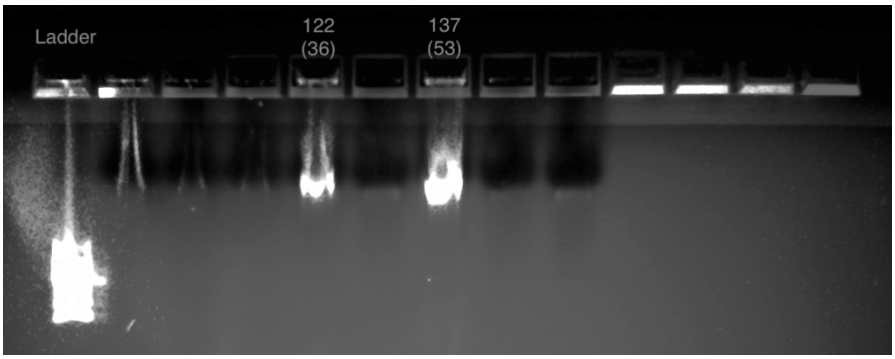

Figure 1.3.6 *FlashGel 6, 100% dye was used, run for four minutes*

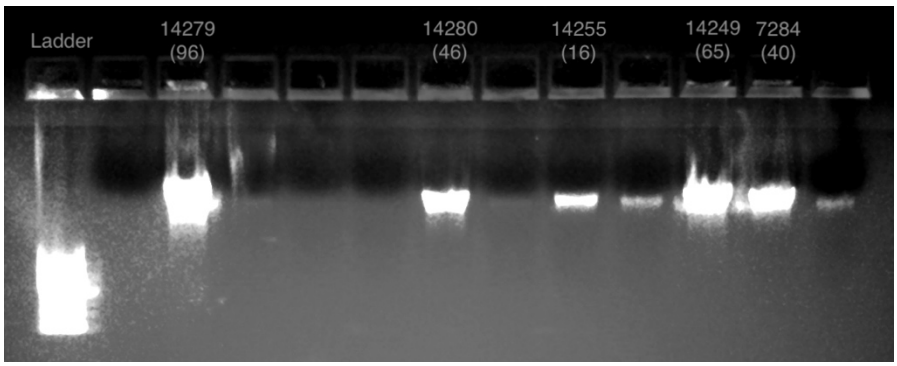

Figure 1.3.7 *FlashGel 7, 100% dye was used, run for four minutes*

4 GENOTYPING

Extracted DNA samples were genotyped at the ARC lab in Pretoria with the assistance of qualified personnel.

Table 1.4.1 *Pedigree assigned by the Sequoia package in RStudio* (Huisman, 2019b)*, including sire ID, dam ID and assigned offspring ID*

Sire ID	Dam ID	Individual ID
16008	14279	117
	16065	14249
	14280	128
9943	7284	132
	-	122
	-	110
	-	22
	7283	120
	7302	14297
	-	57
	7301	100
	7277	113
	137	139
-	7276	14258
-		14267
-	7284	14261
-		137
-	14286	14293
-	16065	14292
-		14290
-	7271	129
-	14293	14276

5 DESCRIPTIVE STATISTICS

Descriptive statistics were carried out on all the individuals included in the pedigree construction. Descriptive analysis was carried out in excel using the data analysis tool, and is illustrated in *Table 1.5.1*.

Table 1.5.1 *Descriptive statistics of the smallholder flock (n=48)*

Age	
Mean	1,5217
Standard Error	0,2577
Median	0,5
Mode	0
Standard Deviation	1,7479
Sample Variance	3,0551
Kurtosis	-1,2031
Skewness	0,5755
Range	5
Minimum	0
Maximum	5
Sum	70
Count	46
Confidence Level (95,0%)	0,5191

6 SAMPLE ID'S

Table 1.6.1 *List of the individuals from each sample population that was analysed in chapter 3-5 before pruning*

Dorpersm 2	Dorper	Namafr	Dorpersm 1	Dorpersm 1	Dorpersm 1	Dorpersm 1
7275	104012D	103024N	24	13433	110142	110409
130	106027D	104003N	13003	13434	110143	110410
117	106030D	104011N	13010	13437	110147	110411
115	106034D	104053N	13012	13438	110150	110412
132	107037D	105020N	13015	13439	110155	110414
74	107092	105031N	13018	13441	110159	110422
14249	108016	108050N	13020	13442	110160	110429
139	108077	108051	13023	13443	110166	110440
16008	108078	108162N	13024	13445	110169	110445
16065	108085D	108204	13026	13446	110171	110446
7277	108143	108205	13027	13459	110177	110448
14261	108220D	108293N	13033	13462	110193	110460
7283	109003	108296	13035	13470	110195	110467
13422	109016D	109095	13039	13477	110198	110473
14267	109030D	109096	13040	13480	110200	110542
122	109040D	109120N	13041	13481	110201	110684
14255	109041	109139N-B	13042	13483	110205	111580
7276	109043	109174N	13045	13499	110211	121286
137	109087D	109176N	13047	14301	110214	m041

9943	109091	109230N	13052	14302	110215	
14293	109097	109247	13053	14304	110216	
7290	109111D	109249N	13055	14305	110220	
110	109118D	109251N	13056	14306	110223	
128	109126	109258	13057	14307	110227	
14250	109156D	109266	13058	14309	110228	
22	109190D	109266N	13060	107084	110234	
14280	109191D	109267N	13143	107092	110235	
14292	109201	109272	13297	110017	110239	
14290	109210	109276	13382	110019	110241	
14286	109220D	109278	13383	110031	110246	
120	109227D	109286	13401	110038	110248	
14270	109228D	110029	13403	110050	110267	
7299	110012	110106	13404	110081	110271	
7284	110042	110139	13405	110083	110275	
14297	110045	110140	13406	110090	110277	
129	110046	110192	13407	110091	110278	
123	110133	110204	13413	110092	110280	
7289	110135	110208	13414	110096	110283	
7301	110165	110259	13415	110097	110290	
7302	110176	111195	13416	110098	110293	
57	110211	111196	13417	110099	110295	
7271	111104	111199	13419	110101	110304	
14276	111178	111280	13420	110104	110305	
14295	111263		13421	110105	110306	
14258			13422	110108	110318	
100			13425	110110	110322	
14279			13426	110115	110323	
113			13428	110122	110327	
110186			13430	110129	110335	
110258			13431	110130	110337	
110340			13432	110131	110341	

7 CNV ANALYSIS

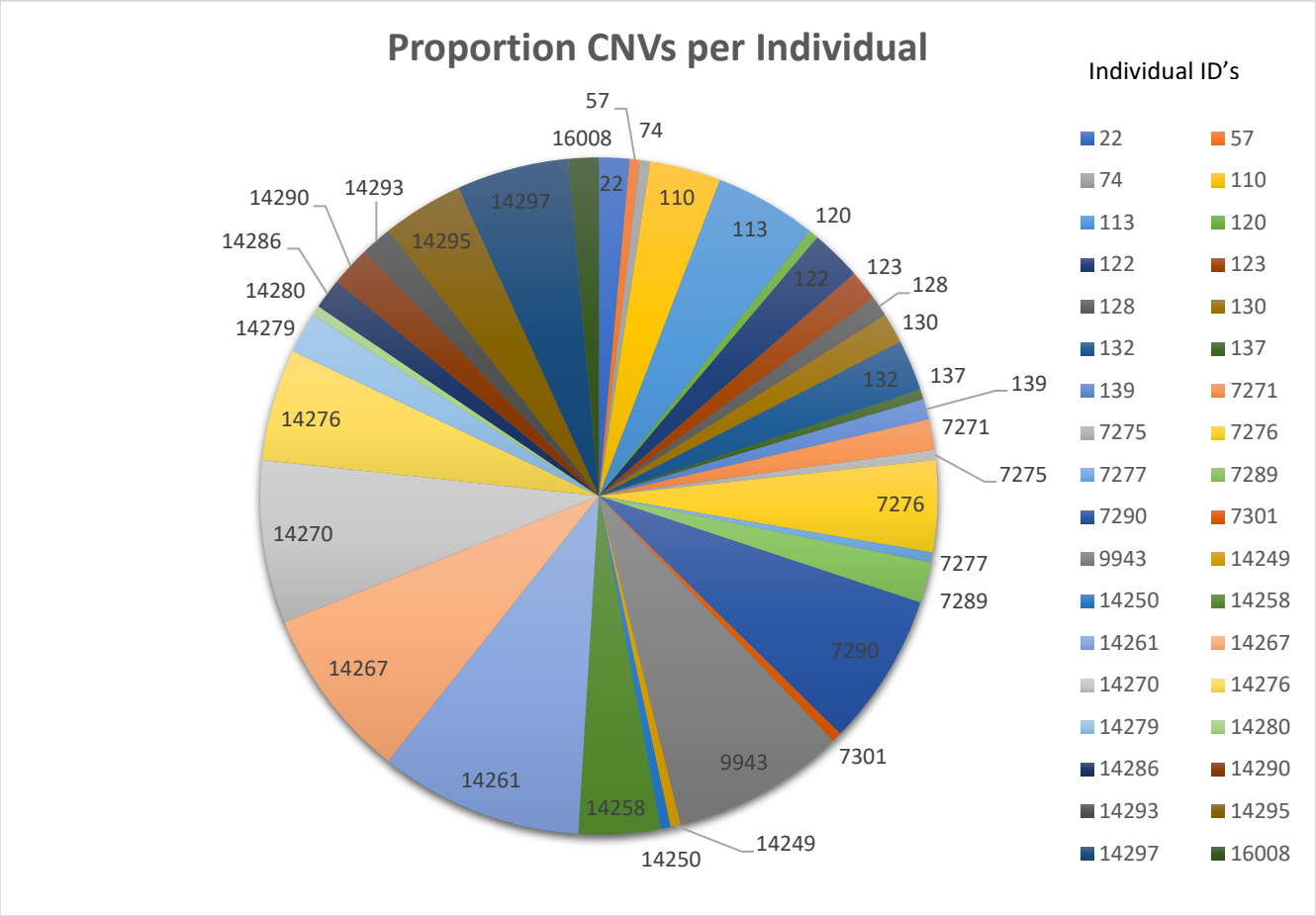

Figure 1.7.1 *The proportion of CNVs identified per individual*

8 CNV IDENTIFICATION

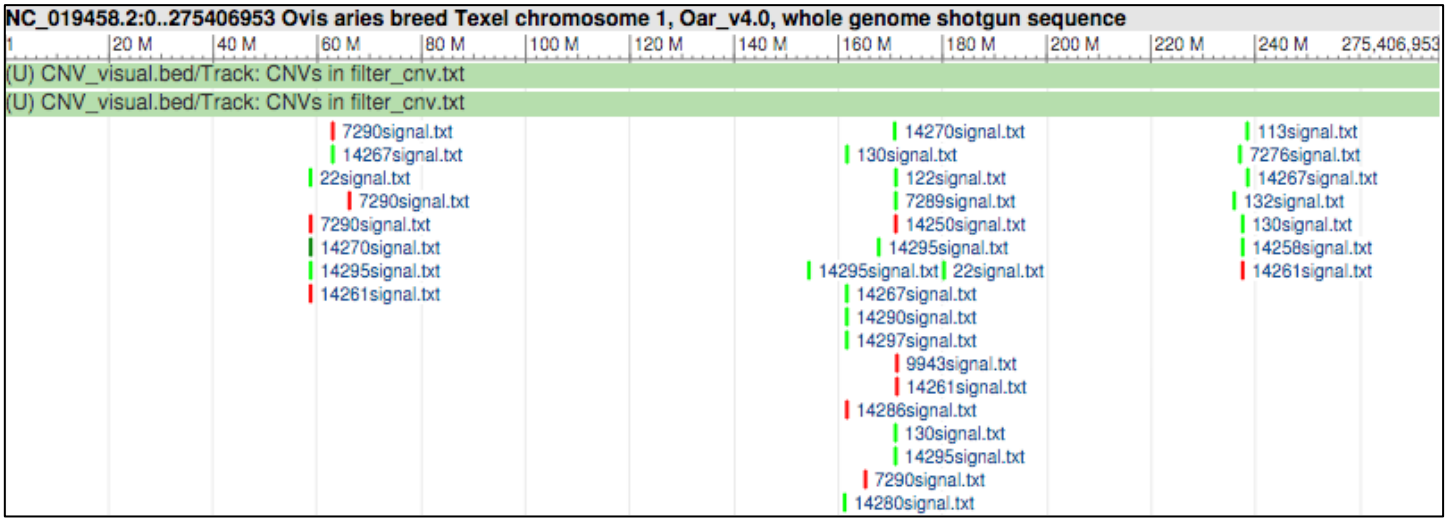

Figure 1.8.1 *CNVs found on chromosome 1*

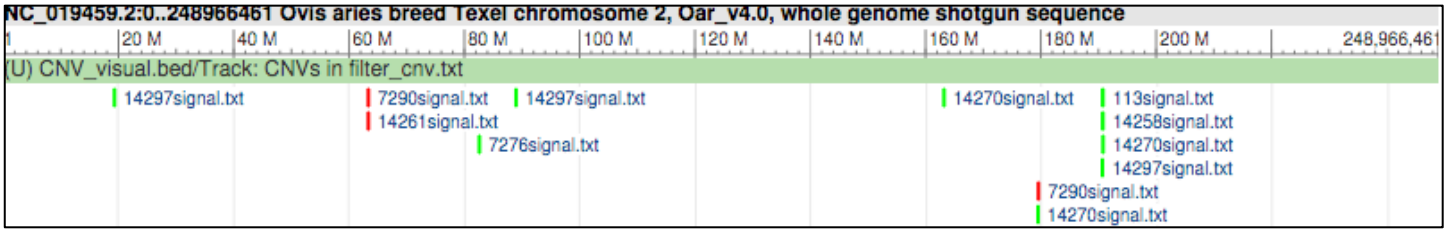

Figure 1.8.2 *CNVs found on chromosome 2*

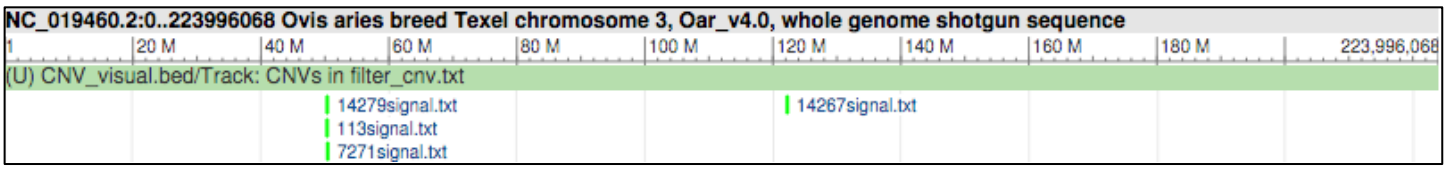

Figure 1.8.3 *CNVs found on chromosome 3*

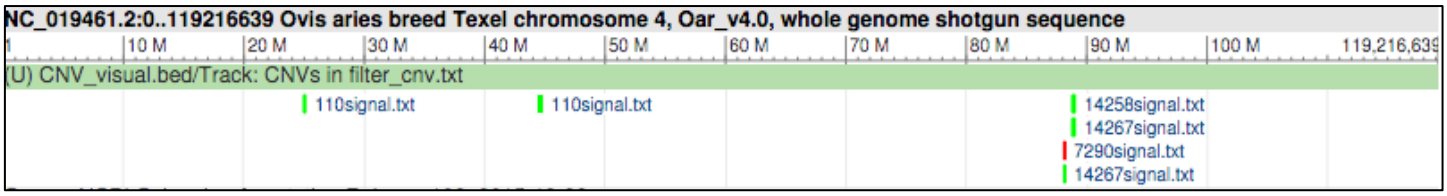

Figure 1.8.4 *CNVs found on chromosome 4*

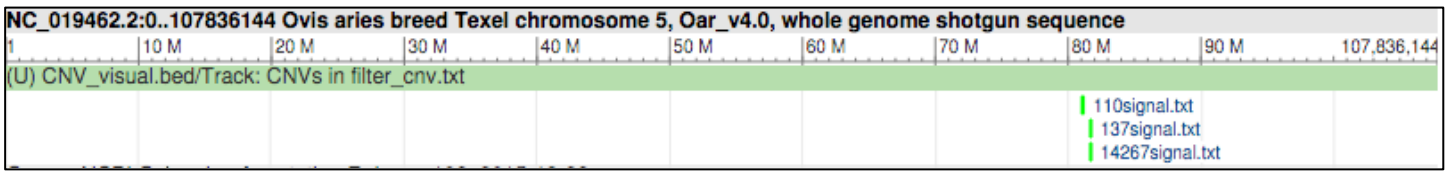

Figure 1.8.5 *CNVs found on chromosome 5*

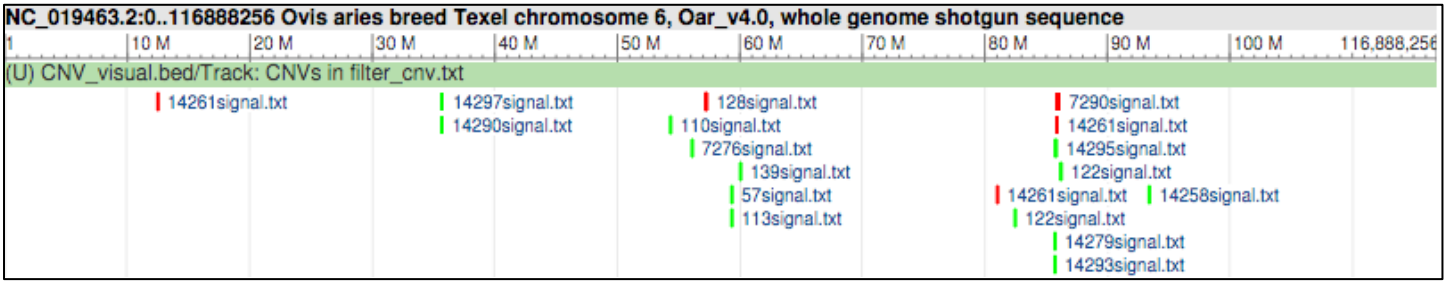

Figure 1.8.6 *CNVs found on chromosome 6*

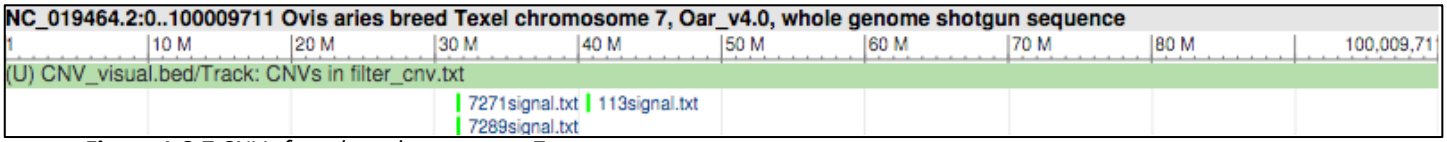

Figure 1.8.7 *CNVs found on chromosome 7*

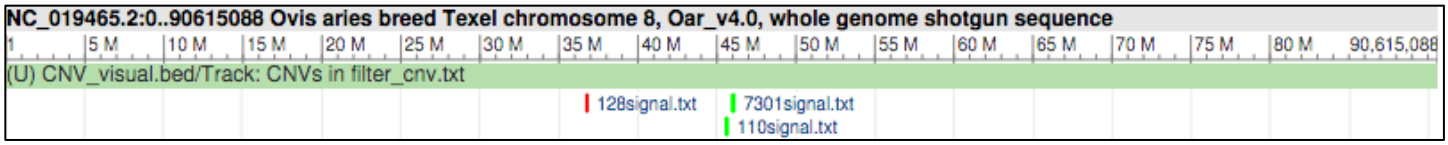

Figure 1.8.8 *CNVs found on chromosome 8*

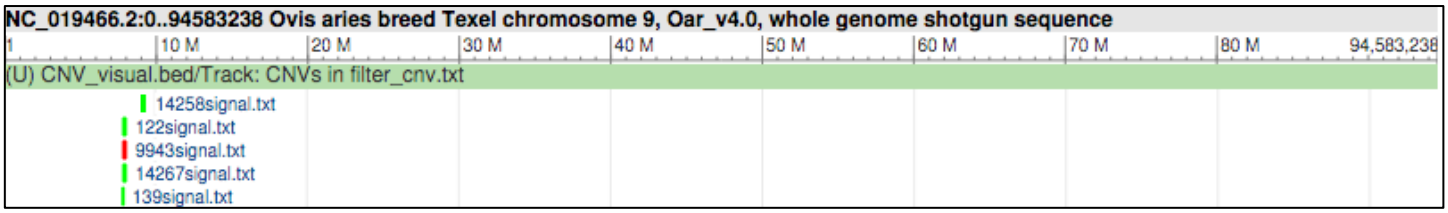

Figure 1.8.9 *CNVs found on chromosome 9*

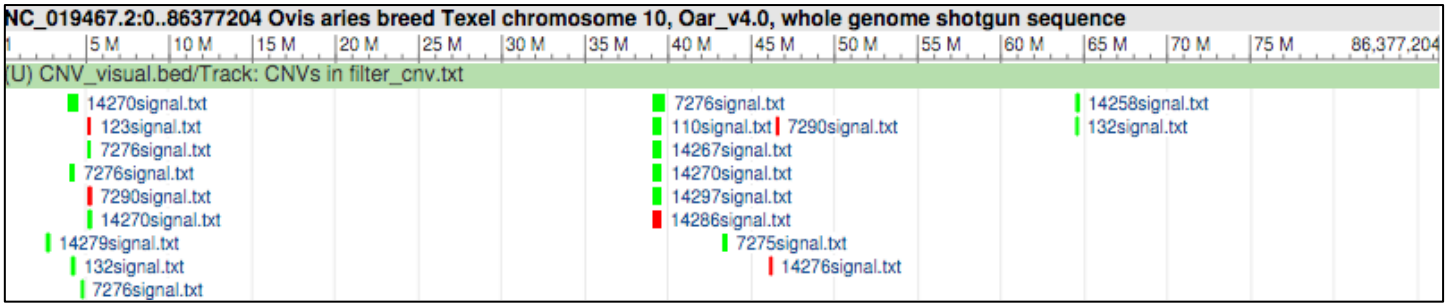

Figure 1.8.10 *CNVs found on chromosome 10*

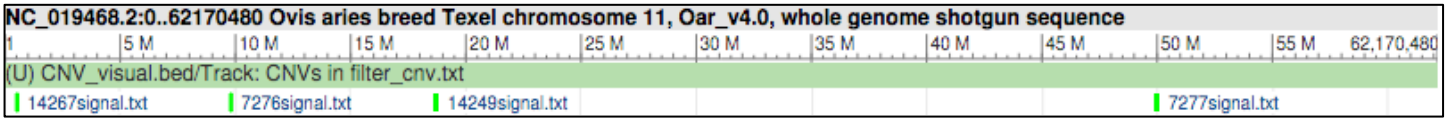

Figure 1.8.11 *CNVs found on chromosome 11*

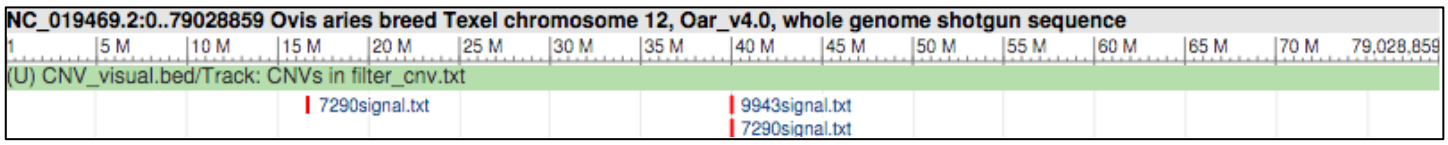

Figure 1.8.12 *CNVs found on chromosome 12*

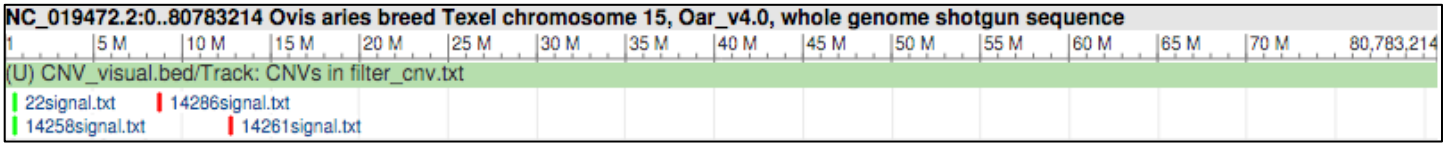

Figure 1.8.13 *CNVs found on chromosome 15*

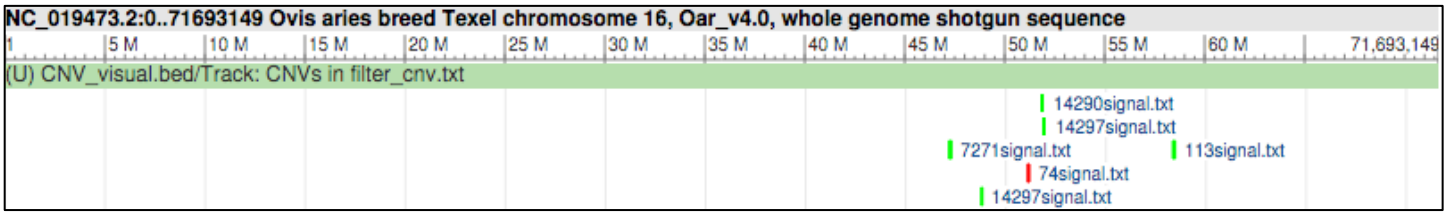

Figure 1.8.14 *CNVs found on chromosome 16*

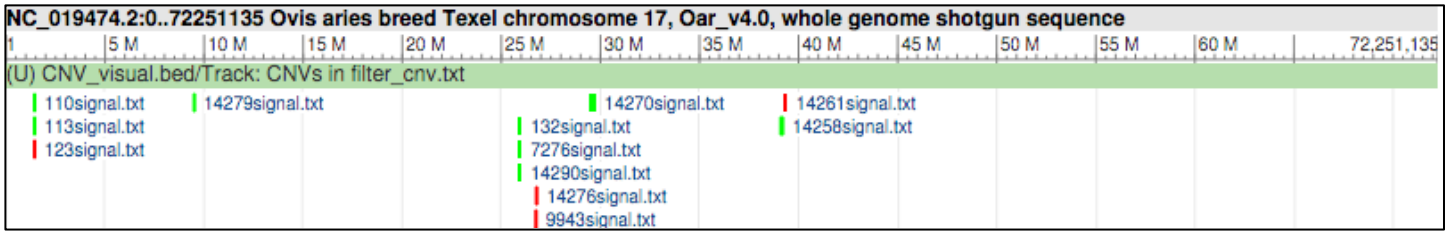

Figure 1.8.15 *CNVs found on chromosome 17*

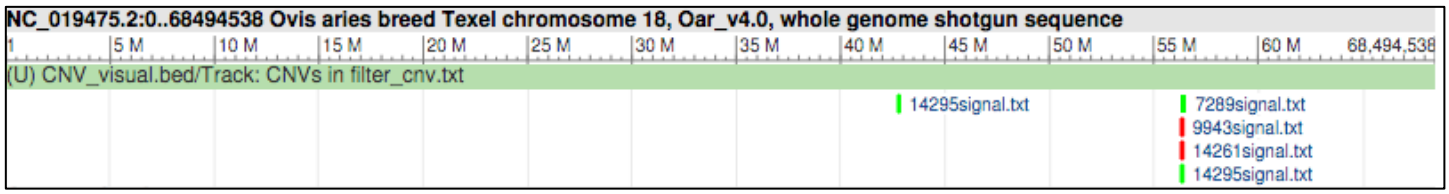

Figure 1.8.16 *CNVs found on chromosome 18*

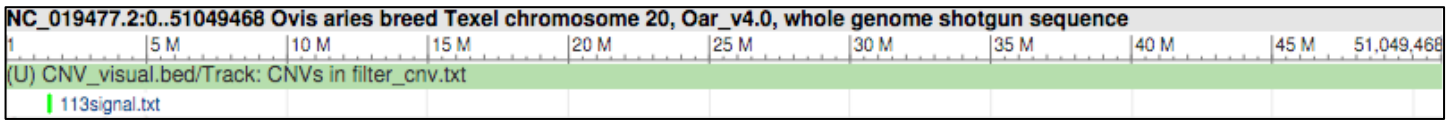

Figure 1.8.17 *CNVs found on chromosome 20*

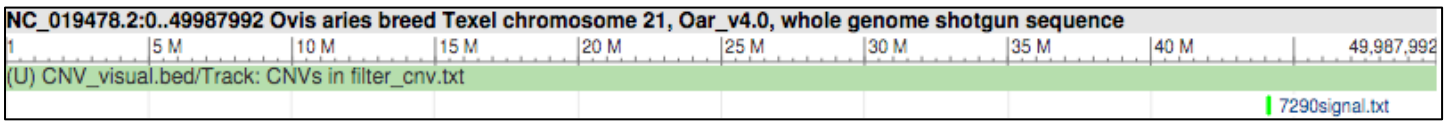

Figure 1.8.18 *CNVs found on chromosome 21*

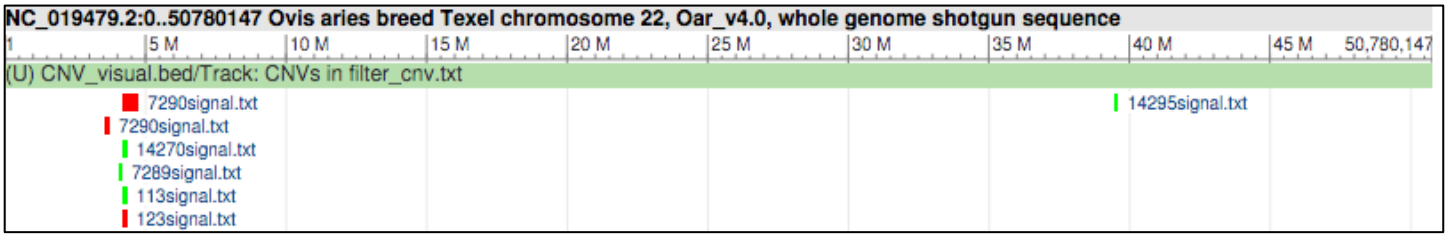

Figure 1.8.19 *CNVs found on chromosome 22*

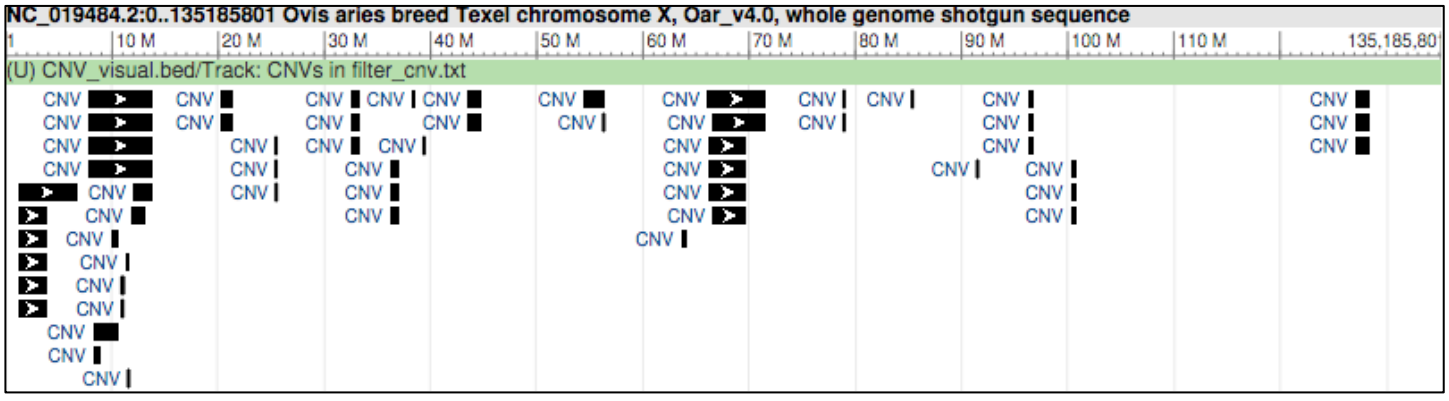

Figure 1.8.20 *CNVs found on chromosome X*

9 GENE ANNOTATION

Legend:

= corresponding gene-term association positively reported

= corresponding gene-term association not reported yet

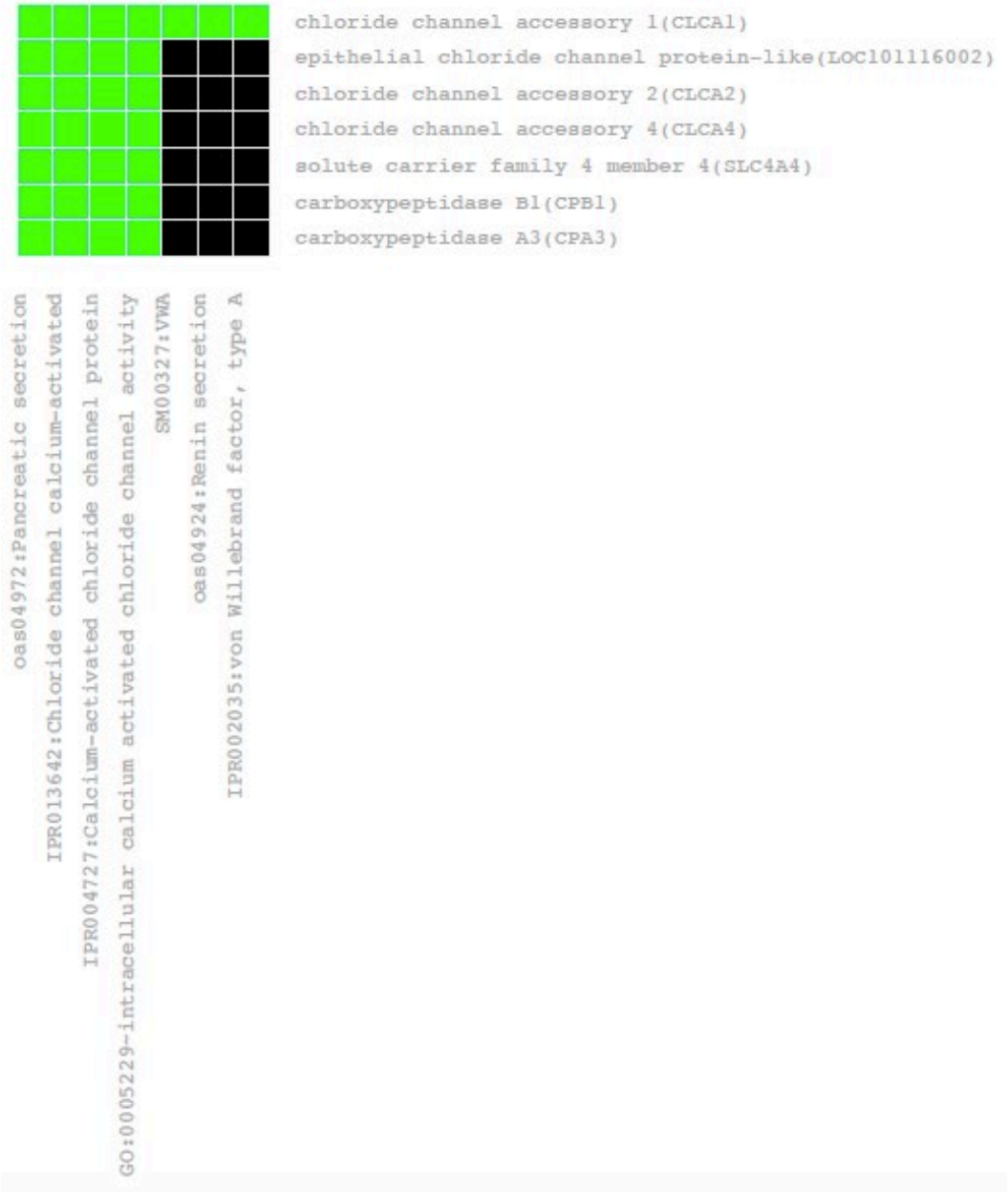

Figure 1.9.1 *Heatmap of the functional annotation cluster group 1*

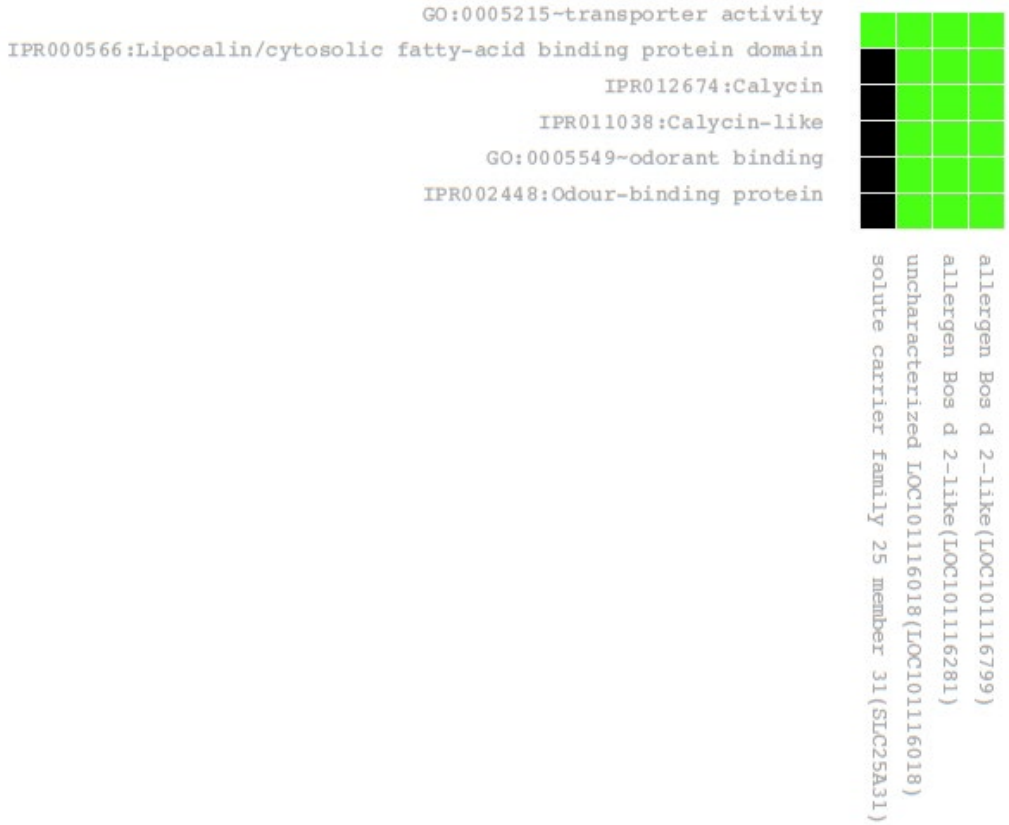

Figure 1.9.2 *Heatmap of the functional annotation cluster group 2*

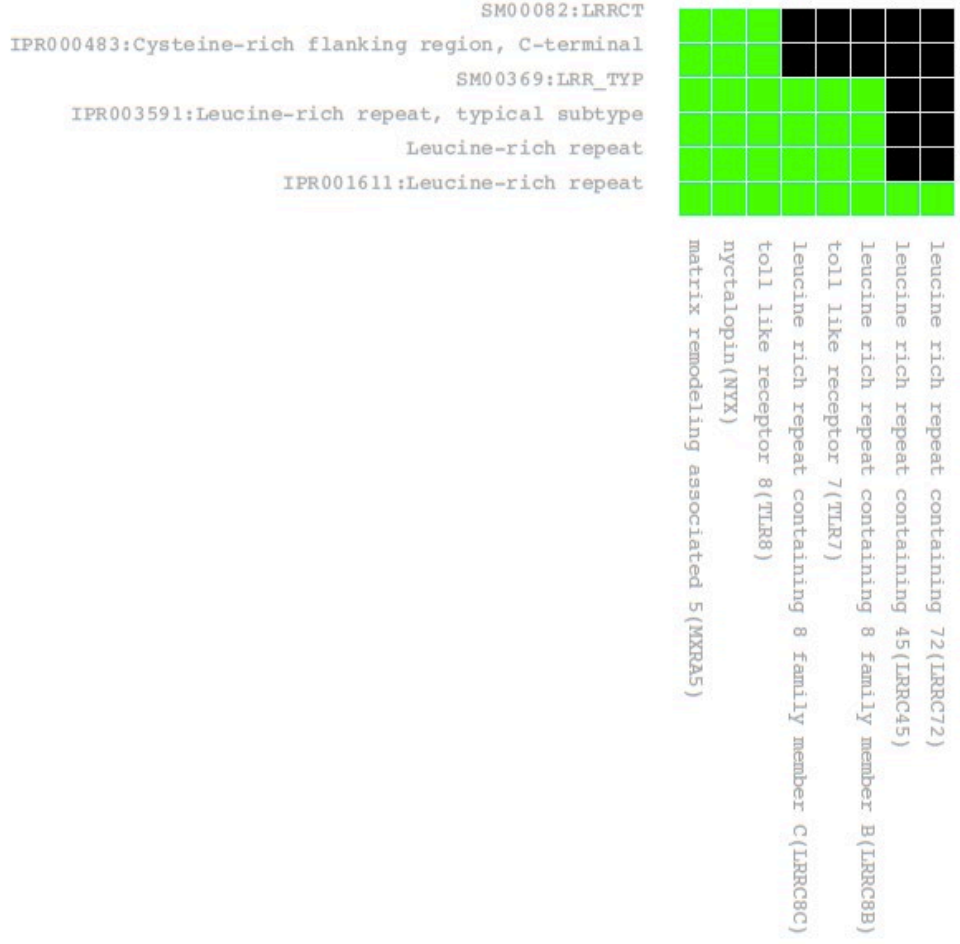

Figure 1.9.3 *Heatmaps of the functional annotation cluster group 3*

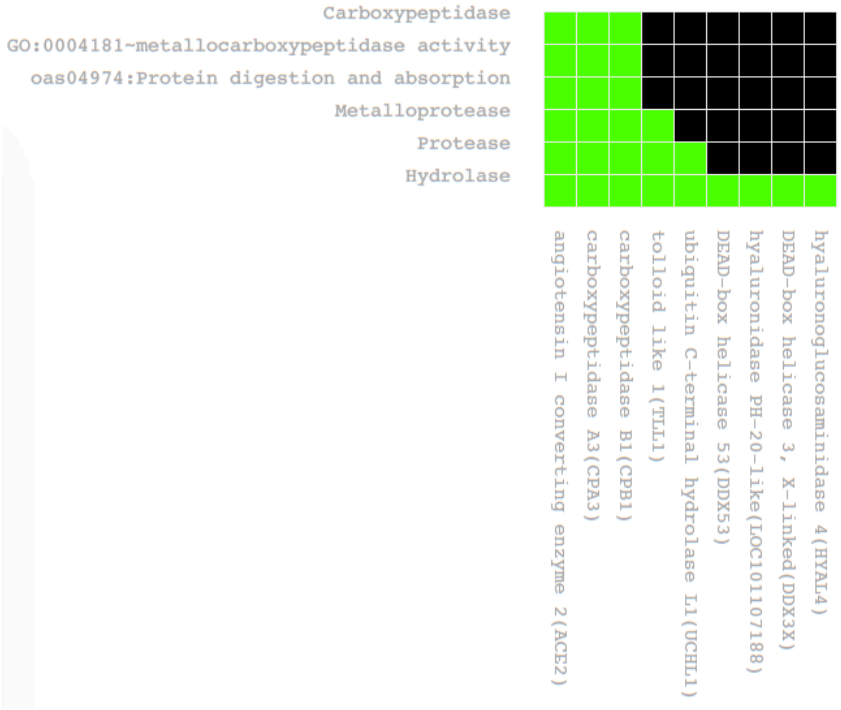

Figure 1.9.4 *Heatmaps of the functional annotation cluster group 4*

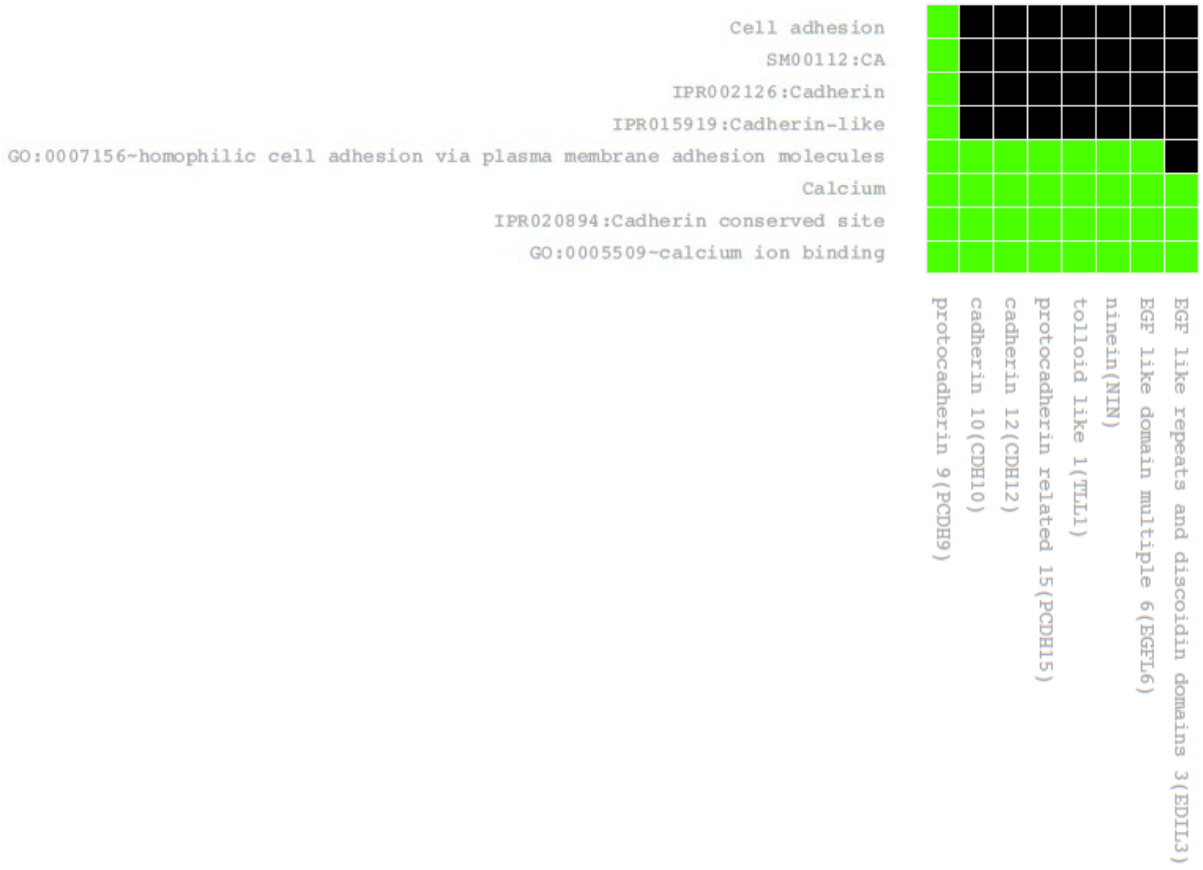

Figure 1.9.5 *Heatmap of the functional annotation cluster group 6*

Figure 1.9.6 *Heatmap of the functional annotation cluster group 7*

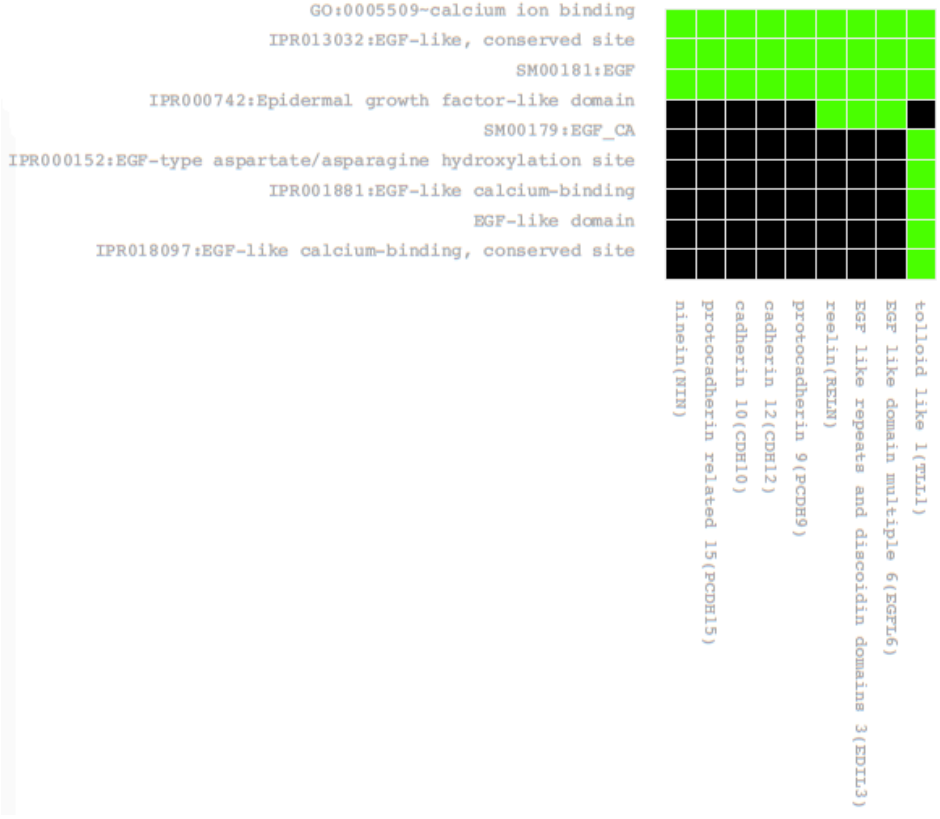

Figure 1.9.7 *Heatmap of the functional annotation cluster group 8*

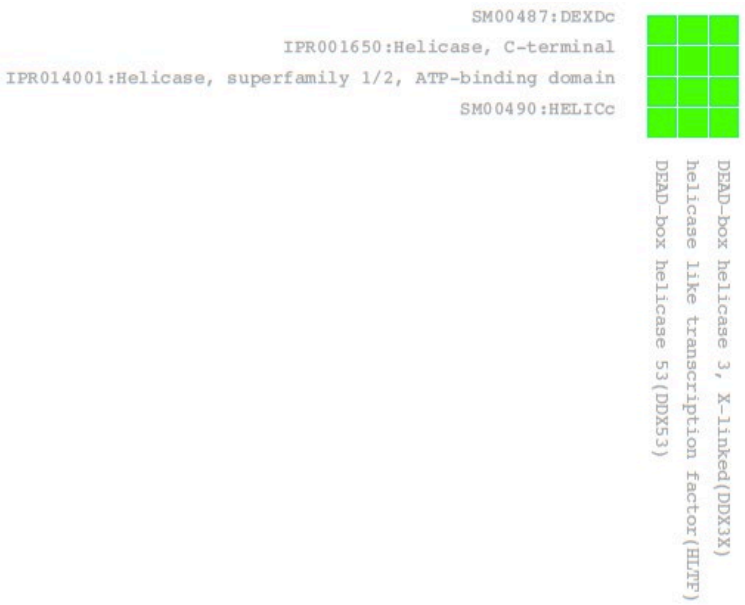

Figure 1.9.8 *Heatmap of the functional annotation cluster group 9*

Figure 1.9.9 *Heatmap of the functional annotation cluster group 10*

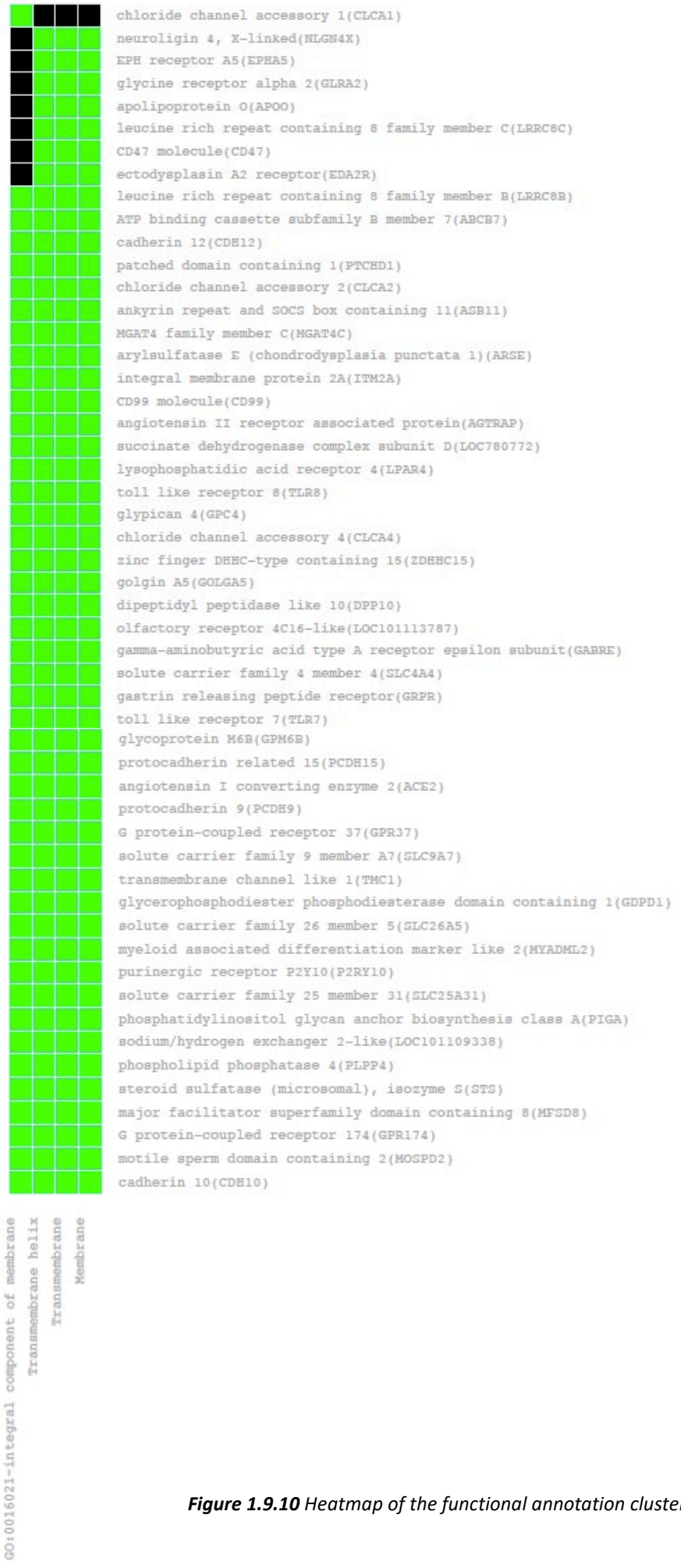

Figure 1.9.10 *Heatmap of the functional annotation cluster group 11*

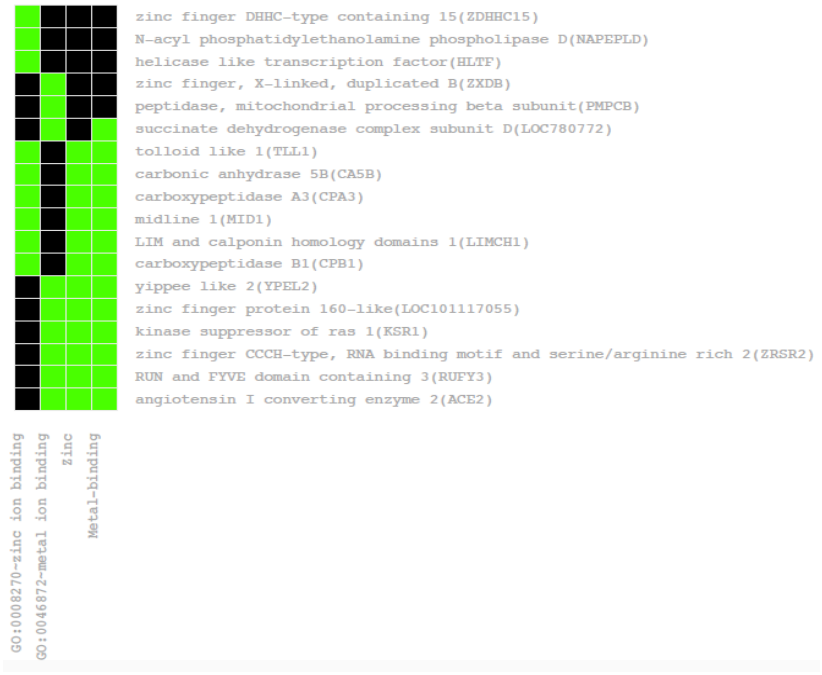

Figure 1.9.11 *Heatmap of the functional annotation cluster group 12*

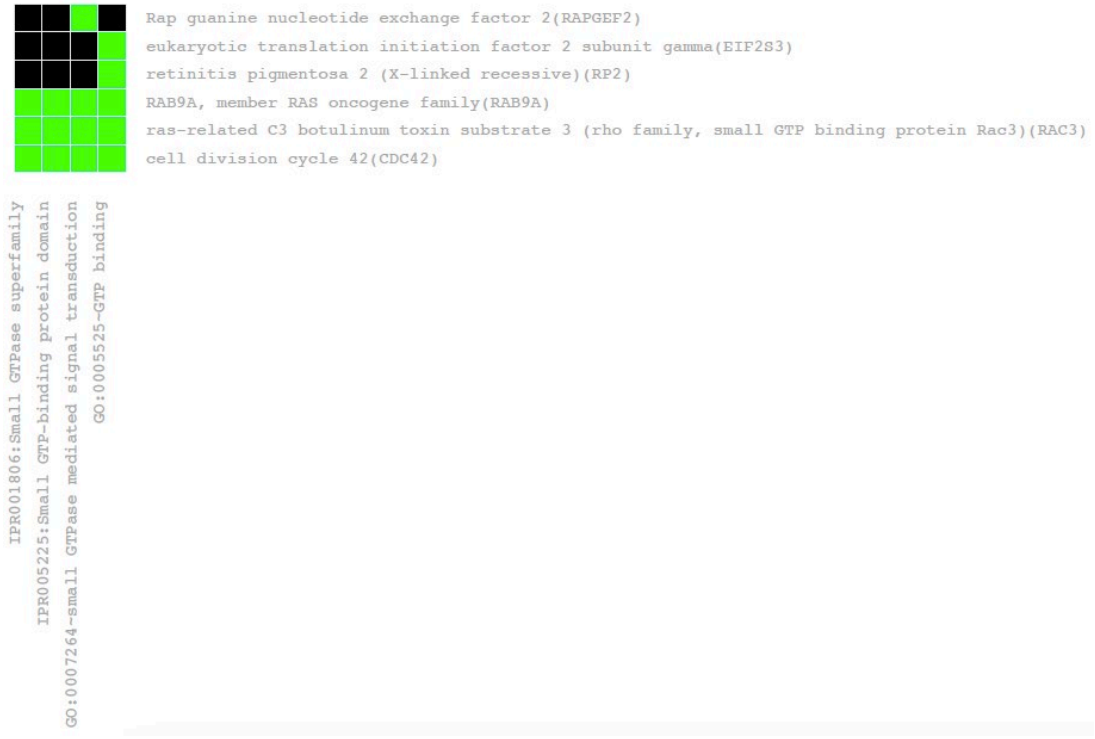

Figure 1.9.12 *Heatmap of the functional annotation cluster group 13*

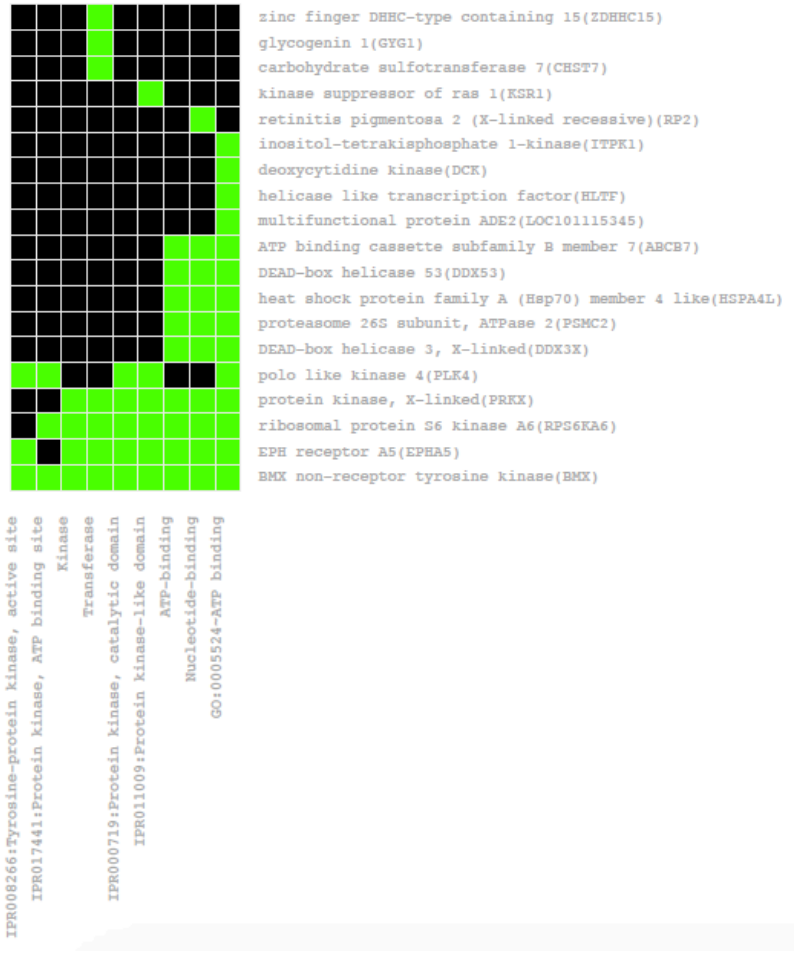

Figure 1.9.13 *Heatmap of the functional annotation cluster group 14*

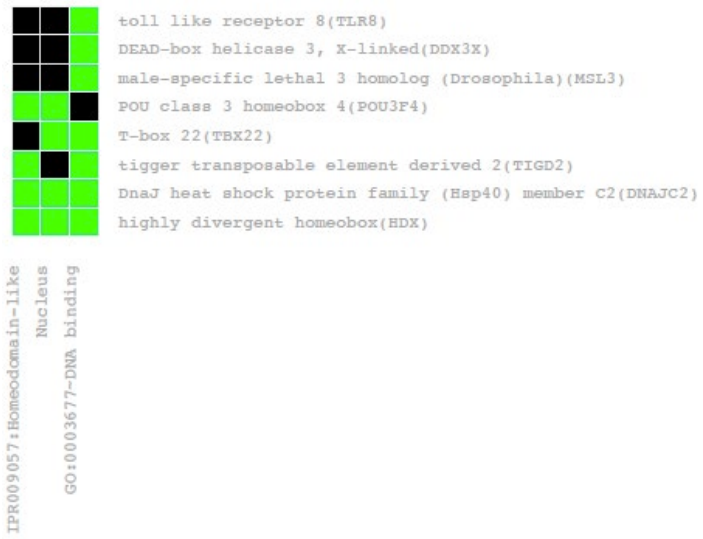

Figure 1.9.14 *Heatmap of the functional annotation cluster group 15*

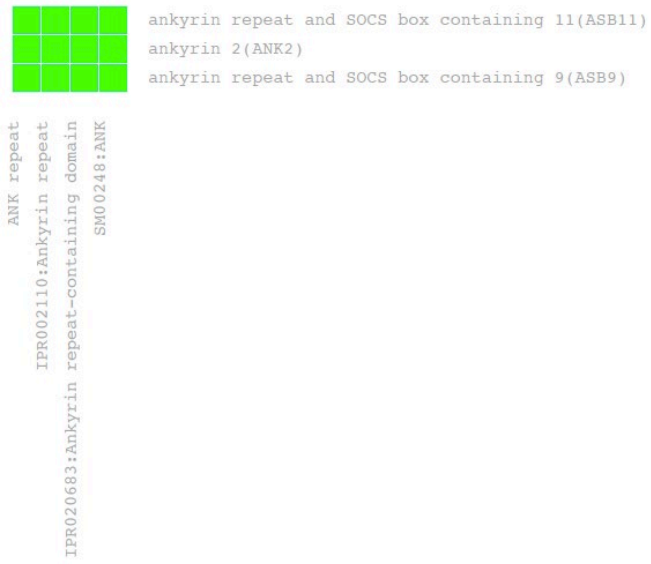

Figure 1.9.15 *Heatmap of the functional annotation cluster group 16*

Figure 1.9.16 *Heatmap of the functional annotation cluster group 17*

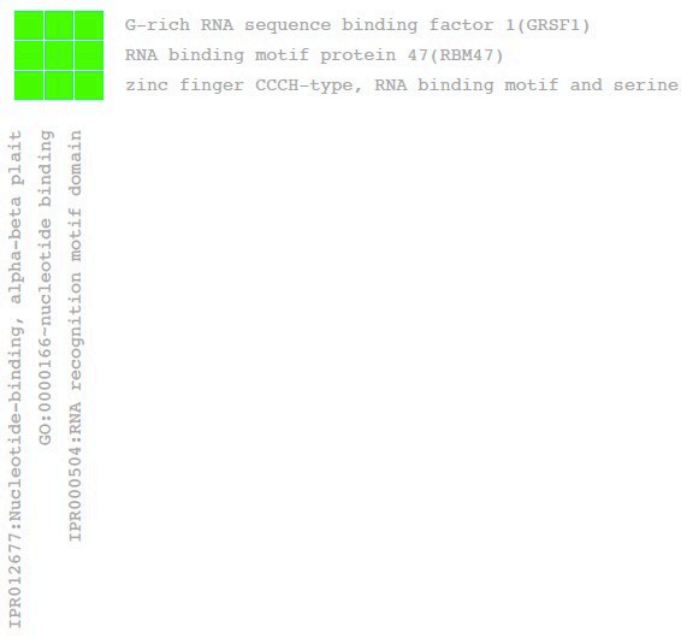

Figure 1.9.17 *Heatmap of the functional annotation cluster group 18*

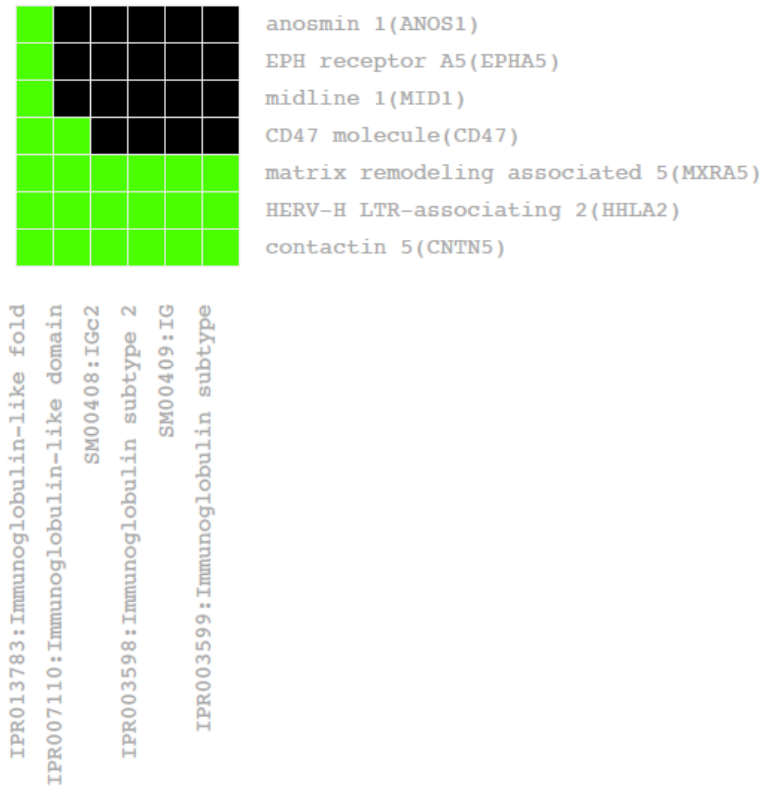

Figure 1.9.18 *Heatmap of the functional annotation cluster group 19*

Figure 1.9.19 *Heatmap of the functional annotation cluster group 20*

Table 1.9.1 *Genes assigned to two gene groups through gene functional classification*

Gene Group 1	
101104373	interferon alpha-H-like (LOC101104373)
101103623	interferon beta-2(LOC101103623)
101104634	interferon alpha-H-like (LOC101104634)
101103872	interferon alpha-2-like (LOC101103872)
Gene Group	
101116002	epithelial chloride channel protein-like (LOC101116002)
101120845	protocadherin 9(PCDH9)
101107256	integral membrane protein 2A(ITM2A)
101109248	dipeptidyl peptidase like 10(DPP10)
101115328	chloride channel accessory 1(CLCA1)
101120549	glycoprotein M6B(GPM6B)
101109338	sodium/hydrogen exchanger 2-like (LOC101109338)
101115513	CD99 molecule (CD99)
101107437	G protein-coupled receptor 37(GPR37)
101118929	solute carrier family 9 member A7(SLC9A7)
101119114	purinergic receptor P2Y10(P2RY10)
101114646	golgin A5(GOLGA5)
101113787	olfactory receptor 4C16-like (LOC101113787)
494440	solute carrier family 4 member 4(SLC4A4)
101121383	leucine rich repeat containing 8 family member B(LRRC8B)
101109316	phospholipid phosphatase 4(PLPP4)
101115586	chloride channel accessory 4(CLCA4)
101103993	glycerophosphodiester phosphodiesterase domain containing 1(GDPD1)
101112540	patched domain containing 1(PTCHD1)
101108733	MGAT4 family member C(MGAT4C)
101115857	zinc finger DHHC-type containing 15(ZDHHC15)
101112912	myeloid associated differentiation marker like 2(MYADML2)
101119371	G protein-coupled receptor 174(GPR174)
100142675	gastrin releasing peptide receptor (GRPR)
101115759	arylsulfatase E (chondrodysplasia punctata 1) (ARSE)
101102455	phosphatidylinositol glycan anchor biosynthesis class A(PIGA)
101118506	lysophosphatidic acid receptor 4(LPAR4)
101117364	cadherin 10(CDH10)
101111567	angiotensin II receptor associated protein (AGTRAP)
101115841	protocadherin related 15(PCDH15)
101115078	chloride channel accessory 2(CLCA2)
101118590	steroid sulfatase (microsomal), isozyme S(STS)
101101864	ankyrin repeat and SOCS box containing 11(ASB11)
101122573	motile sperm domain containing 2(MOSPD2)
101114592	transmembrane channel like 1(TMC1)
101121711	major facilitator superfamily domain containing 8(MFSD8)
101118385	cadherin 12(CDH12)