Landscape genomic approach to investigate genetic adaptation in South African indigenous goat populations

by

Khanyisile Mdladla

Submitted in fulfilment of the academic requirements of

Doctor of Philosophy

in Genetics

School of Life Sciences

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Pietermaritzburg

South Africa

December 2016

PREFACE

The research contained in this thesis was completed by the candidate while based in the Biotechnology Platform, Agricultural Research Council and the Discipline of Genetics, School of Life Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa. The research was financially supported by University of KwaZulu-Natal, National Research Foundation-Department of Science and Technology (NRF-DST) and the Agricultural Research Council.

The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.

_____

Signed: Edgar Farai Dzomba

Date:

_____

Signed: Farai Catherine Muchadeyi

Date:

DECLARATION 1: PLAGIARISM

Note that two declaration sections are required if there are papers emanating from the dissertation/thesis. The first (obligatory) declaration concerns plagiarism and the second declaration specifies your role in the published papers.

I, Khanyisile Mdladla declare that:

(i) the research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;

(ii) this dissertation has not been submitted in full or in part for any degree or examination to any other university;

(iii) this dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;

(iv) this dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written but the general information attributed to them has been referenced;

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

(v) where I have used material for which publications followed, I have indicated in detail my role in the work;

(vi) this dissertation is primarily a collection of material, prepared by myself, published as journal articles or presented as a poster and oral presentations at conferences. In some cases, additional material has been included;

(vii) this dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.

_____

Signed: Khanyisile Mdladla

Date:

DECLARATION 2: PUBLICATIONS

My role in each paper and presentation is indicated. The * indicates corresponding author.

Chapter 2

1. Mdladla, K., Dzomba, E.F., and *Muchadeyi, F.C. 2016. The potential of landscape genomics approaches in the characterization of indigenous goat genetic resources: A South African Perspective. Submitted to Small Rumuinant Research.

Chapter 3

2. Mdladla, K., Dzomba, E.F. and *Muchadeyi, F.C. 2017. Characterization of the village goat production systems in the rural communities of the Eastern Cape, KwaZulu-Natal, Limpopo and North West Provinces of South Africa. *Tropical Animal Health and Production*. DOI 10.1007/s11250-017-1223-x.

Chapter 4

3. Mdladla, K., Dzomba, E.F. and *Muchadeyi, F.C. 2016. Seroprevalence of *Ehrlichia ruminantium* antibodies and its associated risk factors in indigenous goats of South Africa. *Preventive Veterinary Medicine* **125**, 99-105.

Chapter 5

4. Mdladla, K., Dzomba, E.F., and *Muchadeyi, F.C. 2014. Linkage disequlibrium and population structure of South African indigenous goat populations using genome wide SNP data. A poster presented during the 47th South African Animal Science (SASAS), held on the 6-8th July 2014 in Pretoria, South Africa. Presented by Mdladla K.
5. Mdladla, K., Dzomba, E.F., Huson, H.J., and *Muchadeyi, F.C. 2014. The 47th South African Animal Science (SASAS), held on the 6 to 8th July 2014 in Pretoria, South Africa. Work was presented at the student session held on the 6th July 2014. Presented by Mdladla K.
6. Mdladla, K., Dzomba, E.F., Huson, H.J., and *Muchadeyi, F.C. 2014. Linkage disequlibrium and population structure of South African indigenous goat populations using genome wide SNP data.

A poster presented during the 34th International Society for Animal Genetics Conference, held on 27th July 2014 to 1st of August 2014 in Xian, China. Presented by Mdladla K.

7. Mdladla, K., Dzomba, E.F., Huson, H.J., and *Muchadeyi, F.C. 2014. Linkage disequlibrium and population structure of South African indigenous goat populations using genome wide SNP data. The 34th International Society for Animal Genetics Conference, held on 27th July 2014 to 1st of August 2014 in Xian, China. Work was presented at the Livestock Genomics for Developing Countries Workshop, held on 29th of July 2014. Presented by Mdladla K.
8. Mdladla, K., Dzomba, E.F., Huson, H.J., and *Muchadeyi, F.C. 2014. Linkage disequlibrium and population structure of South African indigenous goat populations using genome wide SNP data. The 34th International Society for Animal Genetics Conference, held on 27th July 2014 to 1st of August 2014 in Xian, China. Work was presented at the ISAG-FAO Workhop on Animal Genetic Diversity, held on 29th of July 2014. Presented by *Muchadeyi F.C.
9. Mdladla, K., Dzomba, E.F., Huson, H.J., and *Muchadeyi, F.C. 2015. Population structure and Breed relations of South African indigenous goat ecotypes using genome-wide SNP data. The 23rd International Plant and Animal Genome research, held on 10th to 14th January 2015 the in San Diego, CA. Work was presented at the International Goat Genome Consortium Workshop held on 12th of January 2015. Presented by Mdladla K.
10. Mdladla, K., Dzomba, E.F., Huson, H.J., and *Muchadeyi, F.C. 2015. Population structure and Breed relations of South African indigenous goat ecotypes using genome-wide SNP data. A poster presented at the 23rd International Plant and Animal Genome research, held on 10th to 14th January 2015 the in San Diego, CA. Presented by Mdladla K.
11. Mdladla, K., Dzomba, E.F., Huson, H.J. and *Muchadeyi, F.C. 2016. Population genomic structure and linkage disequilibrium analysis of South African goat breeds using genome-wide SNP data. *Animal Genetics* **47**, 471-482.

Chapter 6

12. Mdladla, K., Dzomba, E.F., and *Muchadeyi, F.C. 2016. Landscape genomics approach in understanding adaptation of indigenous goats to their environment. The 35th International Society for Animal Genetics Conference, held on 23rd July 2016 to 27th of July 2016 in Utah, Salt Lake City, USA. Work was presented at the Applied Sheep and Goat Genetics Workhop, held on 26th of July 2016. Presented by Mdladla K.

13. Mdladla, K., Dzomba, E.F., and *Muchadeyi, F.C. 2016. Landscape genomics approach in understanding adaptation of indigenous goats to their environment. A poster presented during 35th International Society for Animal Genetics Conference, held on 23rd July 2016 to 27th of July 2016 in Utah, Salt Lake City, USA. Presented by Mdladla K.
14. Mdladla, K., Dzomba, E.F., and *Muchadeyi, F.C. 2016. A landscape genomics approach to explain the genetic diversity in South African indigenous goats. A poster presented during 4th South African Genetics Society (SAGS) and South African Society for Bioinformatics (SASBi) Joint Congress, held on 20th September 2016 to 23rd of September 2016 in Durban, KwaZulu-Natal. Presented by Mdladla K.

_____

Signed: Khanyisile Mdladla

Date:

EXTENDED ABSTRACT

Traditional indigenous goat breeds have evolved the crucial ability to adapt and survive in challenging environments and ecological conditions associated with their geographical origins. The environments are heterogeneous and production conditions fluctuate with seasons. Although indications of local adaptation have been reported in most indigenous goats and other livestock species, there are many unknowns regarding the genetic mechanisms that determine the response of indigenous goats to their production and climatic environments. There is potential for these goats to be used for sustainable improvement of productivity in the tropics through combining ecologically adaptive and economically important production traits. Identifying regions that are potentially adaptive in indigenous goats will allow us to explore the adaptive potential for specific environments. This is particularly relevant now when goats and other livestock species are faced with immense production, climatic and other natural selection pressures. South Africa for example is endemic to heartwater and a number of goat populations are frequently exposed and need to adapt. There has been major developments in appropriate genomics technologies such as next-generation sequencing (NGS) and genome wide single nucleotide polymorphism (SNP) BeadChips, and bioinformatics analytical methods, which will facilitate comprehensive characterization of these genetic resources. The study targeted the major goat producing provinces and analysed the diversity in production systems, and the morphological variations of the indigenous goats that they keep, prevalence of heartwater and how genetic diversity has been shaped by the production, climatic and other geographic conditions.

The first objective of this study was to determine the flock composition, the role of goats at household level, and the management practices under which goats are raised in the villages of South Africa. Results obtained from this study were based on the field survey that captured rural community goat raisers' opinion in the form of a questionnaire, and direct observation and measurement of phenotypic characteristics on randomly selected goats. In total, 130 smallholder communal farmers were interviewed from 26 villages from the Eastern Cape (n = 2), KwaZulu-Natal (n = 6), Limpopo (n = 13) and North West (n = 5) provinces of South Africa. The mean flock size per household ranged from 13.2±12.39 in Limpopo to 34.18±28.36 in Eastern Cape, with majority of goats in a flock being adults and females. Goats play a unique role in supporting some of the poorest people in South Africa, and the

study observed that home consumption was the major role of keeping goats. Goats were maintained under scavenging regime with no supplemental feeding in Limpopo, while a few farmers in Eastern Cape, KwaZulu-Natal and North West provinces provided supplemental feed to their animals. Parasites and diseases such as heartwater, ORFV (scabs), gall sickness, mange, and diarrhea are major constraints to goat production thus limiting their potential for commercialization in these communities. The studied goats were highly diverse presenting a variety of phenotypic characters and body size measurements within and between populations. White was the most dominant color among the Tswana and Xhosa, while black was dominant in the Venda, Zulu and Tankwa populations. The canonical discriminant analysis showed that most Zulu, Xhosa, Tankwa, Tswana individuals were assigned to their source population (90.41%, 82.93%, 74.07% and 57.50%) whilst 32 (39.51%) Venda individuals were classified as Zulu individuals. Lowest percentage (> 50%) of individuals of commercial breeds (Boer, Kalahari Red and Savanna) and Venda population were correctly classified into their source population. The first principal component (PC1) explained 43.2% of the observed genetic variation and PC2 accounted for 29.3%, showing admixed individuals. Weak population structuring basis of breed or population was observed with the highest numbers of individuals of the commercial breeds admixed and the lowest observed for Tankwa breed. These results support the population overlaps and incorrect assignment of individuals in the discriminatory analysis. The diversity patterns observed may suggest the effect of production system and that these populations share similar genetic identities thus reflect their adaptation to different production systems and local conditions.

The second aim of the study was to investigate the seroprevalence of antibodies to *Ehrlichia ruminantium,* causative bacteria for heartwater, and its associated risk factors. Sera was collected from 686 goats of the commercial meat type (n = 179), mohair type (n = 9), non-descript indigenous goats from Eastern Cape (n = 56), KwaZulu-Natal (n = 209), Limpopo (n = 111), North West (n = 61) and Northern Cape (n = 11) provinces and a feral Tankwa goat (n = 50). The sera was tested for the presence of immunoglobulin G (IgG) antibodies to antigens of *E. ruminantium* using the indirect fluorescent-antibody test (IFAT). The overall seroprevalence was 64.87% (445/686) with highest seroprevalence reported in endemic provinces of Limpopo, KwaZulu-Natal, North West and Eastern Cape while the lowest was in the non-endemic regions of the Northern Cape province. Increased risk for seroprevalence was observed for non-endemic regions, commercial production system, and for tick infested

goats. At breed level, the risk was lowest for Angora, Tankwa and Kalahari Red, whilst higher in the Indigenous and Savanna than the Boer breed. This analysis managed to identify risk factors of goats raised in different regions of the country, which is important in the development of disease control strategies and local goat improvement programs in the country.

The third aim of the study was to investigate the genetic diversity, population structure, and breed relations, linkage disequilibrium, effective population size and persistence of gametic phase in goat populations of South Africa. Three locally developed meat type breeds, the Boer (n = 33), Savanna (n = 31), and Kalahari Red (n = 40), a feral breed of Tankwa (n = 25), unimproved non-descript village ecotypes (n = 100) from Eastern Cape (n = 20), KwaZulu-Natal (n = 30), Limpopo (n = 30) and North West (n = 20), and Nguni (n = 10) were genotyped using the Illumina Goat SNP50K panel. The proportion of SNPs with minor allele frequencies (MAF > 0.05) ranged from 84.22% in Tankwa to 97.58% in Xhosa ecotype and a mean value of 0.32 ± 0.13 was observed across the nine South African goat populations. Principal Component Analysis, ADMIXTURE and pair-wise F_{ST}, identified Tankwa as a genetically distinct population and supported clustering of the populations according to their production systems and historical origins. Genome-wide F_{ST} identified 101 markers potentially under positive selection in the Tankwa breed compared to their domestic counterparts. Average linkage disequilibrium (r^2) was highest in the Tankwa (0.25±0.26) and lowest in the village ecotypes (range = 0.09±0.12 to 0.11±0.14). An effective population size of less than 150 was observed for all the populations' 13 generations ago. The estimated correlations for all the breed pairs, except Savanna and Tswana populations, were lower than 0.80 for all marker distances greater than 200kb. South African indigenous goat ecotypes have dispersed across the country through historic human migrations and settlement of ethnic tribes, formal and informal trading networks, and expansion of agriculture. They therefore carry the genetic legacy of past historic events, and adaptation to local production systems and environments. The observed genetic variation present opportunities for implementing within population genetic improvement programs for better overall performance and productivity.

The fourth aim of the study was to investigate the types and roles of climatic and geographical variables shaping patterns of genetic divergence and extent of local adaptation in the studied goat populations. Firstly, the study investigated the distribution of genetic variation across

geographical regions. Results showed that the isolated population of the Tankwa breed formed a separate and homogeneous genetic cluster restricted to the Northern Cape. The village goats on the other hand clustered in a way that suggested sharing genetic components along longitudinal components resulting in weak population sub-structure between the Venda and Zulu goats that were differentiated from the Tswana and Xhosa goats. The commercial breeds represented genetic clusters that were not restricted to any geographic location. Further analysis were conducted to test for significant associations between genetic variations and environmental variables (geographic location, altitude, annual and seasonal trend of temperature, and rainfall) using a multivariate redundancy analysis (RDA). The combined effect of climatic and geographic variables explained 14.34% (R^2_{adj} = 9%) of the total genetic variation. Variance partitioning showed that climate variation among sampling sites explained significantly more of the genomic variation compared to geographical distances. Analysis using correlation-based landscape genomic approaches (Spatial Analysis Method and Latent Factor Mixed Models) were used to identify loci associated with environmental and geographic factors. SAM identified a total of 443 (0.92%) SNPs while LFMM identified 370 (0.77%) significantly associated SNPs. In addition, the RDA also allowed for detection of outlier SNPs and identified a total of 149 outlier SNPs correlated to geographic and climatic factors using the first axis. Gene annotation and pathway analysis indicated that a total of 395, 220 and 97 genes identified for SAM, LFMM and RDA, respectively, were involved in numerous pathways important for environmental adaptation including roles in metabolism, response to diseases and heat-response indicating their importance for survival in local environments. Overall, the study contributed valuable understanding of the impact of environmental factors on the genetic variation of South African goat populations. Thus, besides adding insights into the conservation of South African indigenous genetic resources, especially the village goats, these results provided preliminarily evidence that the hypothesis that both the production system and climatic factors maintain the genetic diversity in these goats. Overall, the study offers new insights into genetic adaptations to production environments in goats, and provides a valuable resource for future programs to preserve the adaptive potential of local breeds through genetic improvement programs.

ACKNOWLEDGMENTS

I would like to express my gratitude to the following for their contributions to the completion of this study:

the study promoters, Mr. EF Dzomba and Dr FC Muchadeyi, introducing me to landscape genetics, their continued support, advice and guidance for the past four years. The doors that you've opened for me can only better the young researcher in me.

the Agricultural Research Council Biotechnology Platform, Department of Science and Technology, National Research Foundation, University of KwaZulu-Natal are acknowledged for the financial support of this study.

a sincere appreciation extends to all the goat farmers who allowed the use of their animals for this study and welcomed the research team to their farms and houses expecting nothing in return.

thanks are also due to extension workers of Department of Agriculture (Limpopo, Eastern Cape, and North West provinces) and Mdukatshane Rural Development volunteers for their co-operation during sample collection.

a special thank you to the Veterinary Services and Research from the Northern Cape Department of Agriculture, Kalahari Kid Corporation, Koopmansfontein Research Institute, and Cedara Agricultural College-Pasture Science Department for allowing us to sample.

the collaborators, US Department of Agriculture-Agricultural Research Service (USDA-ARS), African Goat Improvement Network (AGIN), and Heather Huson (Cornell University) for their technical support.

I would like to express my gratitude to my lab sisters, Ms. Petunia Malatji and Ms. Keabestwe Ncube for their dedication to my project during sampling and being there for me throughout my study. A special acknowledgement to Khulekani Khanyile and Sthembile Olga Makina for their assistance in genotyping and data analysis; the students and staff of the Agricultural Research Council Biotechnology Platform for sharing their expertise and long hours in the lab with me.

To my family:

Thokozani Hadebe, my dear husband, for his love, sacrifices and support;

Okuhle Simba Wangamkela Hadebe, my son, for your smile that warms my heart;

my Mom (Ntomb'kaS'mawuza), Nokuthula, Nonjabulo, Nosipho, Noxolisa, Sindiswa and Thulisile for being wonderful mothers to my Son (Okuhle 'Simba' Hadebe) in my absence. May God bless you in abundance for helping me through my studies;

my brothers S'fiso, and Bhekuyise; nephews (Too many to count) for being there as pillars. I love you;

to my late father, Sobula, you've always encouraged me to study. This is your dream, Sangom' esabhula ngonyezi.

Lord Almighty whom I have drawn strength and faith.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1: INTRODUCTION

1.1 Rationale for the research (nature and scope)

The value of goats in the rural areas of South Africa and other developing countries is enormous. Goats are used to address issues of food security at household level, and to serve as sources of cash income amongst other roles. Local indigenous goats are preferred in communal production systems due to their adaptability to prevailing environmental and production conditions. Adapted breeds offer opportunities for sustainable improvement of smallholder and village livestock production (Baker & Gray 2004; Kosgey *et al.* 2006). Local goat populations therefore constitute valuable and unexploited genetic resources for small-scale farms.

The history of South African indigenous goats is interesting, the wild Bezoar (*Capra aegagragus*) having contributed to the genetic pool of the majority of today's domestic goat (*Capra hircus*) including the South African indigenous goat populations (Ncube 2016). Recent studies proposed that they entered the African continent through North Africa (Egypt) before migrating to the southern parts of the continent during the 5th century. Over 6.2 million goats are found in South Africa, with majority kept under communal rural setting, rather than in the commercial sector. Eastern Cape, KwaZulu-Natal and Limpopo have the highest number of smallholder farmers, while Northern Cape has the highest number of commercial farmers. The unimproved veld ecotype populations (Nguni, Xhosa lobbed and Northern Cape Speckeld) are found in the major agro-ecological zones where they represent the main goat type in the communal areas (Bester *et al.* 2009). Due to limited management input, characteristic of the communal production system, natural environment represent one of the major selection pressure in these populations. These include climatic conditions, exposure to disease and parasites, poor quality and scarce feed resources as well as ill-defined selection by humans to meet different socio-economic and cultural needs. Consequently, they have evolved unique adaptation to their agro-ecological environments in which they occur. Unimproved veld goats have been reported to show a degree of resistance to tick-borne disease such as heartwater (Donkin *et al.* 1992), as well tolerance to poor quality forage (Silanikove *et al.* 1993). Other adaptative strategies for unimproved Veld goats vary from morphological characteristics such as varied coat colours and small frame to phenotypic traits

including the presence and absence of horns, and wattles. Morphological variation could be quite attractive for screening overall adaptive genetic diversity (Toro & Caballero 2005). A comprehensive description of the phenotypic characteristic of the three unimproved Veld ecotype populations have been previously described (Morrison 2007). It is hypothesised that phenotypic responses to selection pressures lead to non-random or directional changes in the allelic and genotypic frequencies of a population. Thus, each ecotype would comprise of a unique set of genes or genomic regions that have responded to selection to provide the population with an addded advantage to survive and optimally produce in the specific agro-ecological zones or regions. The diversity and mechanisms controlling the diversity and adaptation are not well understood in South Africa and other developing countries' goat populations and should not be ignored.

The unimproved veld ecotype goat genetic resource has provided the material for the successful breed improvement programs of the developed high performance meat type goat breeds of the Boer, Kalahari Red and Savanna goats mostly preferred by commercial farmers. These are kept under intensive production environment thus are often buffered from natural selection pressures. Despite this, the Boer goat has been reported to have the ability to survive drought conditions without supplementary feeding and resist and survive diseases such as blue tongue, prussic acid poisoning and enterotoxaemia (Malan 2000).

Lack of proper breeding definition is a major threat to indigenous genetic resources of South Africa as it undermines opportunities for efficient utilization of the goat genetic resources (Hassen *et al.* 2012). The lack of well defined breed structure especially in the village production system challenge comparisons of performance and diversity between breeds or populations. The term 'local' goats in South Africa is interchangeably used with 'indigenous', 'traditional' and to a less extent 'village/non-descript' without well defined characteristics or parameters differentiating them into breeds or population. These goat breeds exhibit quite distinct phenotypic variation in size and other conformation traits such as coat color, presence or shape of horns and ears (Morrison 2007). Local/indigenous/village/non-descript goat populations are also defined as ecotypes based on their distribution in different agro-ecological zone or farming regions. Ecotypes are considered unique genotypes defined by a combination of separate adaptive, natural and artificial selection and other evolutionary forces that influences genetic evolution. Thus, it is important to understand the genetic structure of

the South Africa goat populations and their genetic relations in order to unravel genetic processes responsible for their adaptation and improved productivity. In the current study, the term 'non-descript village ecotype', will be used for the populations sampled in the villages, in addition to the local people's indigenous language for each ecotype (i.e. Zulu for goats sampled in the KwaZulu-Natal etc.). In addition, in the village goat farming system, breeding is often uncontrolled and the communal production system allows for indiscriminate breeding resulting in uncontrolled crossbreeding and mixed genotypes of goats (Gwaze *et al.* 2009a). Further, it is hypothesized that in pursuit of higher meat yields, smallholder goat farmers often either buy and raise the Boer and other commercial breeds or crossbreed these with their local goats. Conservation strategies require a good knowledge of the population genetic structure of these local ecotype populations, as well as an assessment of their adaptation to local environment, to provide recommendations regarding their future management and utilization.

South Africa, with its diverse agro-ecological and climatic features, experience huge temperature and rainfall differences among the agro-ecological zones. It contains high range of altitudes between the sea level (0m) in the coastal areas to 3658m above sea level inland (Dent *et al.* 1987). Western Cape is a winter rainfall region, while the rest of the zones are summer-rainfall regions. Majority of the country experiences highly variable rainfall in the rainy season, and dry periods are often combined with high temperatures (Sivakumar 1992). These seasonal and agro-ecological features in turn affect pasture quality and quantity, and disease distribution and burden. As an example for seasonal profiles of diseases is the gastrointestinal parasites such as *Haemonchus contortus*, the burden of which is more pronounced in the summer rainfall regions of South Africa (Vatta *et al.* 2001b). Another example that is a huge challenge to smallholder farming sector is the tick-borne disease due to the lack of or limited of tick control is heartwater. In the cases reported for heartwater to the South African National Department of Agriculture between 1993 and 2014, the mortality rate due to heartwater was 32.12% of the reported cases with difference varying largely between regions, and times of the year. Eastern Cape had the highest number of cases, followed by Limpopo, North West, Mpumalanga, KwaZulu-Natal, and Gauteng with only one case reported. Close and strong relationships exist between the climate, natural resources, and livestock populations particularly in systems where there is little or no human input into the production system as characterstics of the village goat production systems.

Landscape genomics correlates genetic variation patterns with geographical and environmental characteristics of production systems to scan for the presence of signatures of selection (Luikart *et al.* 2003; Joost *et al.* 2007) that define ecotypes. This approach detects the markers that underlie adaptive genetic variation, thus is crucial in indigenous goats as it offers tools to identify the genotypes suitable for given production environments. The varied topography and environmental heterogeneity may be the major reasons behind the highly diversified ecotype populations across South Africa, thus presents a good model to explore the environmental adaptation of these populations. Currently, there is a growing body of literature demonstrating the feasibility of landscape genomics in detecting loci related to adaptation in goats. For instance, Pariset *et al.* (2009) detected association of 27 loci to environmental parameters using Spatial Analysis Method (SAM). Colli *et al.* (2014) reported significant correlations between AFLP markers and climatic variables (diurnal temperature range, frequency of precipitation, relative humidity and solar radiation) indicating the evidence of natural selection in European and Western Asian Goat Breeds. In recent years, the advances in goat genetics and the availability of high throughput genomic tools combined with genome scans has enabled the development of a Goat 50K BeadChip (Tosser-Klopp *et al.* 2014) which can be used to facilitate the investigation of adaptive genomic divergence at a genome-wide coverage. Genome-wide single nucleotide polymorphisms can be useful in identifying loci under selection for a particular breed or population that can be modeled to understand genetic adaptation of populations. Such information is useful in identifying environmentally suitable populations and aid in designing production system specific breed improvement programs.

This study therefore, hypothesized that diversity of South Africa goats is defined by the variations in production landscape and that the goat population genetic structures will be correlated to the geography and the environmental characteristics of the local production systems.

The hypotheses to be tested in this study were:

- The goat production landscape in the communal areas of different South African provinces are similar.
- Genetic diversity in South African indigenous goat populations is high, and village goat populations are not sub-structured by agro-ecological zones.

- Environmental factors have no significant effects on genetic diversity, population genetic structure and differentiation amongst goat populations.

1.3 Aims

The main aim of the study was to define the production landscapes of goat populations in the major goat producing provinces of South Africa, unravel the genetic diversity of indigenous goats and further employ landscape genomic tools to investigate the factors shaping genetic variation and adaptation of South African goats.

1.4 Objectives

Specific research objectives were to:

- Determine the local goat production systems in the communal areas of the Eastern Cape, Kwazulu-Natal, Limpopo and North West provinces of South Africa and identify production challenges as well as improvement opportunities.
- Investigate phenotypic diversity of local goat populations found in communal areas of the Eastern Cape, Kwazulu-Natal, Limpopo and North West and investigate population clustering using morphometric attributes.
- Investigate the sero-prevalance of *Ehrlichia ruminantium* and the associated risk factors in South African goat populations.
- Investigate genetic diversity and population genetic structure within and between indigenous goat population using Illumina Goat 50K SNP BeadChip.
- Determing genomic regions under selection and the associated geographic and climatic factors shaping adaptation of indigenous goats to their environment, using a landscape genomic approach.

1.5 Outline of dissertation/thesis structure

The thesis starts with a general introduction and literature review followed by four experimental chapters written in a manuscript format. Two of the chapters (Chapter 4 and Chapter 5) have been published in international journals. Chapter 4 titled ‘Sero-prevalence of *Ehrlichia ruminantium* antibodies and its associated risk factors in indigenous goats of South

Africa' was published in the Journal Preventive Veterinary Medicine, whilst Chapter 5 titled 'Genome-wide assessment of South African goat breed genetic diversity reveals population structure and admixture' was published in the Journal Animal Genetics. Chapter 3 has been accepted for publication in the Tropical Animal Health and Production. The final discussion and study conclusions are contained in Chapter 7. The thesis chapters are summarized as:

Chapter 1: Introduces the rationale of the study by highlighting research gaps on which the hypothesis to be tested were formed. It also states the main aim and objectives of the study.

Chapter 2: Reviews current knowledge on the South African goat genetic resource, the characteristics and challenges of different production systems, the environmental landscape and explores the potential of landscape genomics in studying the adaptive potential of indigenous goats in the country.

Chapter 3: Focuses on the village goat production system characteristics, production challenges and the morphological variations among the indigenous goat populations in the rural communities of South Africa.

Chapter 4: Profiles the heartwater seroprevalence of the sampled areas using the serologoical test Indirect Flourescent Antibody (IFA) test.

Chapter 5: The genetic diversity, linkage disequilibrium and population structure for the different breeds using the Goat 50K BeadChip was invetsigated as well as identify loci under selection in the Tankwa goat breed.

Chapter 6: Focuses on the geographical genetic population structure, investigating the contribution of environmental and geographical variables on the genetic structure and unravelling the genes that have adaptive potential in indigenous goat populations using a landscape genomic approach.

Chapter 7, provides a critical discussion of the study and its findings, presents the study conclusions and gives recommendations and future prospects of the study.

1.6 Ethical statement of the study

Ethical clearance for the use of animals for this study was obtained from the Agriculture Research Council - Animal Production Institute Animal Ethics Committee. Blood samples were collected by Mr. Teko Isaac (Chief Animal Technician- State Veterinary Office).

CHAPTER 2: THE POTENTIAL OF LANDSCAPE GENOMICS APPROACH IN THE CHARACTERIZATION OF ADAPTIVE GENETIC DIVERSITY IN INDIGENOUS GOAT GENETIC RESOURCES: A SOUTH AFRICAN PERSPECTIVE

Accepted in the Journal Small Ruminant Research

ABSTRACT

South African indigenous goat populations are renowned for their adaptability to local agro-ecological conditions and suitability for low input production systems. This is particularly important in a country where a majority of the goats are kept under marginalized communal production systems. Adaptation of animal genotypes to the environment is central to the coping of local livestock populations to changing climatic and production objectives and is a key parameter to use in conservation and improvement programs. South Africa has successfully developed world-renowned commercial meat breeds, the Boer, Kalahari Red and Savanna, all from the non-descript indigenous goat populations. In addition, South Africa has a feral Tankwa goat population that is found in the Northern Cape province. Despite its success stories and unique goat populations, majority of the South African goats are non-descript indigenous veld goats whose genetic diversity is yet to be fully characterized and harnessed. Advances in genomics and population genetics have led to substantial progress in identifying local genetic adaptation of livestock species. Landscape genomics merges the competing effects of production system, geographical, and environment landscapes with adaptive variation. The review gives an overview of current knowledge on the South African indigenous goat genetic resources and the major goat production systems across the different agro-ecological zones of the country. The review then explores the opportunities and potential of landscape genomic methods and how they can complement traditional phenotypic and genetic characterization approaches. The factors that make it feasible to undertake a landscape genomics analysis for South African livestock such as goats are discussed. The developments in genomics and statistical genomics and the opportunities they bring to goat landscape genomics research is presented.

Keywords: *Goats, genetic resource, landscape genomics, genetic adaptation*

2.1 Introduction

South African indigenous goats are mainly kept under traditional management practices in communal areas (Lebbie 2004; Webb & Mamabolo 2004; Bester *et al.* 2009; du Toit *et al.* 2014). Majority of the goats are non-descript and raised in hyper-variable production systems. They present heterogeneous phenotypic and possibly genotypic characteristics and are adapted to the agro-ecological zones that they inhabit. However, their production potential has not been fully harnessed owing to a dearth in research and development work around their utilization. The full genomic architecture of these goat populations is under investigation and research on the effect of natural and/or artificial selection on goat genetic diversity is on-going. In recent years, demographic, production and evolutionary events that impact on local adaptation of indigenous goats have started appearing in literature in South Africa and worldwide. Compared to cattle, goat research attracts limited sources of funding owing to cattle's position as the conventional source of meat compared to goats (Schiere & Kater 2001). Goats are also regarded as livestock for the poverty-stricken and marginalized communities (Sebei *et al.* 2004a). Goat genetic studies are still in their infancy and only a few studies investigating genetic diversity, phenotype-genotype association and signatures of selection are available (Kijas *et al.* 2013; Brito *et al.* 2014; Nicoloso *et al.* 2015; Lashmar *et al.* 2016; Mdladla *et al.* 2016a). None of these studies assessed the genomic adaptation to specific environments nor attempted to untangle the ecological relevance of the genome-wide outlier markers. The adaptive genetic potential of indigenous goats is paramount in ecosystems where harsh and extreme production environments are worsened by climate changes as experienced in South Africa and other tropical developing countries.

The availability and continued falling costs of genome-wide SNP marker panels for goats (Tosser-Klopp *et al.* 2014) and other livestock species provide opportunities for more reliable investigations of genetic diversity and genomic signatures of selection in this species. Further, a combination of population and landscape genomic approaches represents a significant opportunity for investigating variation in response to historic or prevailing environmental conditions, giving an insight into the biological basis for adaptive genetic variation. The significance of landscape genomics lies in its ability to combine the genotypic data, geographic and environmental variables of spatially referenced individuals or populations across different landscapes to identify genomic regions that are involved in local adaptation. The hypotheses behind landscape genomics is that variation in allele frequencies is defined by

geographic, climatic, disease and pathogen profiles. Individual markers that show higher genetic differentiation and subsequently, skewed allele frequencies related to environmental variation, are targeted for genomic association analysis.

This review describes the goat genetic resources of South Africa and describes the production and environmental landscape of the major goat producing provinces. Recent developments in goat genomics are outlined and the potential of landscape genomics in deciphering the genetic merits of South African goat breeds is assessed. The considerations made when undertaking a landscape genomics analysis and the probable opportunities and constraints foreseen for South Africa and most developing countries are highlighted.

2.2 Goat genetic resources in South Africa

South Africa is characterized by a substantial diversity of indigenous goat breeds. There are four registered indigenous goat breeds that consists of three commercial meat type goats, the Boer, Kalahari Red and Savanna, as well as multi-purpose non-descript indigenous veld type goats found mainly in marginalized rural areas (DAFF 2012). Furthermore, South Africa has the Tankwa goat, which is a feral goat breed that is currently reared as conserved flock at Carnarvon Research Station (Mohlatlole *et al.* 2015).

2.2.1 Commercial meat type goats

The development of commercial meat-type goat breeds in South Africa dates back to the early 1900 and is one of the best success stories in livestock improvement to date. The Boer goat is one of the first breeds to be developed and is the leading meat-type breed in South Africa that has a worldwide distribution and utilization (NDA 2012). The development of the Boer goat can be traced to the Dutch farmers of South Africa (Casey & Van Niekerk 1988). The Boer Goat Breed Association describes the earliest breeding stock that was a result of crossbreeding between the shorthaired female goats that had white bodies and light colored heads with a large dapple-colored male goat (www.boergoats.co.za). After generations of selection and improvement, the breed evolved towards a white body and a red head and became known as the “Improved/Ennobled Boer Goat” differentiating them from the original type of goats of the unimproved type. The Improved Boer Goat is the only line or type, which the South African Boer Goat Association registered as a breeding quality animal

(http://www.boerboksa.co.za/). The breed is now distributed in 53 countries worldwide and in Africa, it is found in Lesotho, Swaziland, Botswana, Namibia, Mozambique, Zimbabwe, Angola, Zambia, Malawi, Tanzania, Burundi, Rwanda and in South Africa where it originates from (Rischkowsky & Pilling 2007).

The white Savanna was selected from indigenous goats in 1957 by Messers Cilliers and Sons (Campbell 2003) but was first recognized as a distinct breed resulting in the formation of the Savanna Goat Breeders Society on November 21, 1993. The Kalahari Red was developed by several breeders around the country with different breeding objectives and its developmental history and timeline remains one of the most interesting yet complex of the commercial meat type goats. Kalahari Red goats originated from two lines of the brown lop-eared 'unimproved' indigenous goats in South Africa and Namibia and of the Boer goats (Campbell 2003). The earliest record of the breed dates back in 1970s when Mr. Voster collected red and red-dappled goats in South Africa and Namibia. In 1991, Albie Horn also started selecting indigenous brown and brown and white goats from the former homelands of the Eastern Cape, the Karoo, and Namibia (http://studbook.co.za). The goats were selected based on hardiness, adaptability and conformation. Ewe selection concentrated on good mothering ability, fertility and milk production for the kids (Personal communication with Mr. Albie Horn, a farmer and breeder in Northern Cape). The breed association was formed and registered in 1999.

2.2.2 Non-descript indigenous veld goats

The non-descript indigenous veld goats are native to Southern Africa (Mamabolo 1999; Morrison 2007) and are regarded as an unimproved population that has not been selected for any production traits. These goats are thought to have migrated to Southern Africa about 2500 BC (Epstein 1971; Maree & Plug 1993) with the local Khoikhoi nomads who travelled southwards from Northern Botswana down to the Orange River and later followed two additional routes to reach the Southern and Western Cape (Ehret 1982; Elphick 1985). Natural and, to some extent, human selection coupled with new mutations and the effects of random drifts could have resulted over time, in the modifications and subsequent development of the various non-descript indigenous veld goats presently available in various climatic regions of Southern Africa that are referred to as ecotypes. There are three ecotypes that have been historically described in South Africa which include the Nguni (also referred to as iMbuzi),

Eastern Cape Xhosa, and Northern Cape Speckled (Morrison 2007). The Nguni (iMbuzi) occurs more abundantly than the other ecotypes in South Africa occurring specifically in the high rainfall areas of the Eastern Cape, KwaZulu-Natal, Limpopo, Mpumalanga, and Northern Province (Morrison 2007) of South Africa. The Eastern Cape Xhosa type originated from the Eastern Cape (Morrison 2007) while the Northern Cape 'speckled' type has its origin in the Northern Cape province (Morrison 2007). These ecotypes are believed to be the main genetic resource used in the development of the current commercial meat type goats (Campbell 2003). Currently, the different ecotypes are now found with breeders all over the country (Personal communication with Jonny Morrison from the Indigenous Veld Goat Breeders Society) and form majority of the goat populations in the communal areas (Bester *et al.* 2009).

2.2.3 South African feral goat populations

Feral populations may be sources of genetic variation including primitive traits not present in modern and developed breeds, and novel or rare adaptations (van Vuren & Hedrick 1989). Distribution and density of feral goats in South Africa is limited to the Tankwa goat from the Northern Cape province (http://www.nzg.ac.za). Although sparse information on the history of Tankwa goat and its origin is available, records show that the breed has been freely roaming in the Karoo region of the Northern Cape province for more than 50 years with a population size that ranged between 100 and 300 (Kotze *et al.* 2014). The South African National Parks (SANParks) initiated the removal of feral goats from Tankwa Karoo National Park to Carnarvon Nature Reserve in a move to conserve flora and fauna of the park. The Tankwa feral goats represent an untainted and important goat genetic resource of South Africa (Kotze *et al.* 2014; Mohlatlole *et al.* 2015).

2.3 Challenges in breed definition and impact on inventories on goat genetic resources

The State of the World's Animal Genetic Resources for Food and Agriculture (SoW-AnGR; Rischkowsky & Pilling 2007) states that the characterization and monitoring of livestock genetic resources is key to conservation and proper utilization of breeds and should include the identification and quantitative and qualitative description of breeds or populations including their production systems. The concept of breed needs to be established first, in order to ensure accurate descriptions and to avoid losing relevant data on the genetic resource.

Currently, there are discrepancies in the nomenclature of the indigenous goats of South Africa, which is a challenge when allocating goat populations to breeds and when assessing the information available for the different breeds. The current naming system does not accommodate the thousand of goats found outside these specific locations throughout South Africa and may not reflect the genetic differentiation within and between breeds.

There are currently 15 South African goat genetic resources listed on the Domestic Animal Diversity Information System (DAD-IS & FAO) of FAO (http://dad.fao.org; Accessed June 2014) and 13 on the Domestic Animal Genetic Resources Information System (DAGRIS) of the International Livestock Research Institute (ILRI; http://dagris.ilri.cgiar.org; Accessed June 2014), including those listed in DAD-IS. This list includes the indigenous goat breeds that were either developed for a specific use or those where no improvement has been initiated. Imported breeds are either used as pure breeds or for crossbreeding with local indigenous goats for heterosis and complementarity of traits. In addition, the Animal Improvement Act of South Africa (Act No 62 of 1998) recognizes two groups of goat genetic resources in South Africa as being landrace and exotic breeds. The 'landrace' breed group indicates a specified breed or a kind of animal indigenous to, or developed in, the Republic of South Africa. Under this definition are the non-descript indigenous veld goats, the Tankwa feral and the commercial meat-type goat breeds. The 'exotic' breed group consists of breeds that have been imported from other countries into the Republic of South Africa, which consist predominantly of the dairy and mohair type breeds. Furthermore, descriptors for goats in the Animal Genetic Resources data bank classify South African goats into five groups (http://www.fao.org/; Accessed June 2014).

There are a number of challenges that contribute to poor information on the goat production sector of South Africa and other developing countries. Firstly, the current naming system on these databases as with the ecotype system does not reflect the genetic differentiation of the breeds and is often confusing as names overlap and description of breeds is not uniform. Secondly, the considerable variation observed among the local goat populations, in terms of size and coat colors has led to some inconsistencies in the classification of goat breeds. Initiatives to compile databases on these breeds are important in building a better understanding of the goat diversity in South Africa and most developing countries. One such is the maintained by the Agricultural Research Council. However, most of these initiatives

have been ineffective due to the paucity of information available (Simela & Merkel 2008) and lack of continuity in documenting more breeds and updating with new information. The limited information compiled in these databases indicates the shortage of information on goat genetic resources in South Africa, as much of the diversity that exists in the locally adapted populations still remains unknown and undocumented. Goat statistics are often inaccurate and questionable as they differ even amongst official sources. National surveys do not often account for all production systems or are based on questionnaires administered at household level (Bester *et al.* 2009). The extensive system of production under which communal goats are raised is characterized by a lack of written pedigree and production records which make it difficult to describe breed characteristics (Gebreyesus *et al.* 2013). The number of goat populations in communal areas may not be accurate due to informal exports and exchange of animals between communal farmers and household slaughtering. Not enough surveys have been conducted to characterize breeds and sharing of data collected for the different breeds in the public domain is poor. Although universal and standardized protocols to characterize the phenotypic diversity of populations (FAO 2012) are available however, they need to be adopted and additional surveys are needed to determine the state of South African goat genetic resources.

2.4 South African goat production landscape

South African goat genetic resources consist of various breeds and populations raised under diverse production systems across variable agro-ecological zones. The landscape of these goat genetic resources will therefore be defined by the production systems under which they are raised and the geographical and environmental parameters across the different agro-ecological zones that define the biotic and abiotic factors the animals are exposed to.

2.4.1 Production system landscape of South African goat genetic resources

South African goats exists under two main production systems of commercial and smallholder or communal, which includes village backyard goat production in communal areas (Benhin 2006). A commercial farm is based on large scale farming of goats, while communal farms are predominantly for subsistence. Smallholder communal production accounts for more than 60 % of the total goat inventory in South Africa (du Toit *et al.* 2014). KwaZulu-Natal, Eastern Cape and Limpopo have the highest proportion of the country's smallholder farmers, whilst

commercial farmers are mainly in the Northern Cape (NDA 2012). Management systems and production in commercial systems differ significantly with those of communal farming systems. A proportion of South African goats are kept at research stations and farms as experimental populations. Feral goats constitute another production system in South Africa where goats are either in the wild or exist as captive populations for conservation purposes. Assessment of existing production systems is an important tool for understanding the production constraints and opportunities that exist for the different goat populations.

2.4.1.1 Commercial production system

The commercial goat production system is intensive, with high labor requirements, comprised of moderate to high management levels and the rearing of specialized breeds. In South Africa, the three commercial breeds of the Boer, Kalahari Red and Savanna dominate the meat-producing goat industry (Visser *et al.* 2004). Selection for specific production traits and management interventions have resulted in improvements in their production performance (Almeida *et al.* 2006) but at the cost of reduced genetic diversity due to high inbreeding levels (Kristensen *et al.* 2015; Mdladla *et al.* 2016a). Goats raised under commercial farming are vaccinated and covered by tailor-made biosecurity measures, and as a result, are protected from the different disease and pathogen challenges. Consequently, they do not take advantage of natural immunity, which threatens their ability to be adapted and cope with the harsh and fluctuating environmental conditions as well as the effects of climate change.

2.4.1.2 Communal production system

Smallholder production systems are based on the extensively raised indigenous local populations (Lebbie 2004; Webb & Mamabolo 2004) that play an integral role to meet the multiple social, economic and cultural needs of rural households (Ilatsia *et al.* 2012). Non-descript indigenous veld goat populations remain predominant in villages (Mamabolo 1999; Bester *et al.* 2009; Gwaze *et al.* 2009b) despite the introduction of developed, exotic and crossbred types that have not been able to function fully under communal production systems (Norris *et al.* 2011). Smallholder farmers usually have broad breeding objectives often determined by the multiple uses of goats at household level. Village goat production is characterized by low input, small flock sizes, poor nutrition, and high levels of diseases and parasite infections and low productivity (Van Niekerk & Pimentel 2004; Bester *et al.* 2009;

Gwaze *et al.* 2009b, 2009c). Although the production system is sustainable for the resource poor rural households, output in terms of weight gain and births per year are very low with high mortality rates. Genetic improvement of village goats is not as advanced as in the commercial farming sector and is considered exceptionally rare or non-existence. However, in order to meet the socio-cultural and religious preferences, farmers and livestock keepers often do some sort of selection on phenotypic traits. This type of selective breeding practice has been used to maintain and increase the frequencies of particular phenotype(s) for the preferred phenotypic traits and not so much for production. Smallholder farmers may have specific considerations and choices of specific coat colors (Manton 2005; Gwaze *et al.* 2009a) and body frame (Gwaze *et al.* 2009a; Marume *et al.* 2013), which they then use in their selection for breeding animals. The selection is however unorganized, with no institutional capacity to record and measure the selection pressure applied and monitor the response realized.

2.4.1.3 Feral wild system

According to Biodiversity Act 2004 (Act No, 10 of 2004), feral goats are invasive mammal species in South Africa. Feral goats populations do not experience human interventions, have no specific production goals and survival is determined by adaptation to local environmental conditions (Sponenberg *et al.* 2014). Breeding and access to resources such as food, water and shelter is without human intervention. The topography of the South African Karoo creates natural environmental barriers to gene flow and population migration in the form of mountain ranges and vegetation. The Tankwa feral goat populations may have evolved or maintained adaptation specific to their environment that have been lost in domesticated breeds during domestication and breed formation. The capturing of these populations into conservation units may have adversely reduced their genetic diversity since only a few animals were relocated and could ultimately reduce their ability to survive in captivity.

2.4.2 Environmental landscape of South African goats

South Africa is located at longitude 16°- 32° East and latitude 22°-34° South and is divided into nine administrative provinces and four agro-ecological zones (Gbetibouo & Hassan 2005) varying in rainfall distribution and average temperature. Goats are widely distributed across the major geographical regions and agro-ecological zones (Lebbie 2004). Eastern Cape, Limpopo and KwaZulu-Natal provinces hold the largest goat population, encompassing for

approximately 72% of the total live goats (6.2 million; FAOSTAT 2013). The largest agro-ecological zone is the medium-rainfall (400mm-600mm) arid zone covering the Eastern Cape, Free State, Gauteng, Limpopo, and Mpumalanga provinces. The arid zone comprises the whole of the Limpopo, Mpumalanga and the North West and Free State Provinces, the western parts of KwaZulu-Natal and the Eastern Cape, and the northern parts of the Western Cape. This zone keeps majority of the indigenous goats (Lebbie 2004) compared to other zones. KwaZulu-Natal falls in the high-rainfall (800mm) sub-tropical wet zone while Western Cape experiences winter rainfall of 400mm and falls in the sub-tropical winter zone. The desert agro-ecological zone is a low-rainfall area (200mm) in the Northern Cape province with an average temperature of 23°C. It borders Namibia and Botswana and is one of the hottest, driest regions of the country.

South Africa is generally a water scarce country with an average rainfall of about 450mm per year, which is well below the world average of about 860mm per year. Rainfall mostly occurs in the summer (November to March), with winter rainfall (June to August) in the south-west in the Western Cape province. The driest province is the Northern Cape and the wettest is KwaZulu-Natal.

Climatic characteristics within the agro-ecological zones have significant impacts on pasture and feed resources availability and, in turn, the livestock production. In arid areas, vegetation is prone to very little plant turnover (Holmgren *et al.* 2006), thus livestock will often have insufficient or low quality feed (Maree & Casey 1993). Abundance of natural pastures and forage of high quality is often linked to wet seasons (Bakare & Chimonyo 2011). Inadequate and poor quality feed is one main constraint in goat production in South Africa (Kosgey 2004) and other African countries (Mrema & Rannobe 1996; Maleko & Koipapi 2015; Nampanzira *et al.* 2015). In these areas, only animals that are adapted to the local environment will thrive. Therefore, there is a need to understand climate-driven adaptation of indigenous populations and study the spatial patterns along the agro-ecological zones of South Africa.

South Africa exhibits a high degree of climatic, habitat and seasonal heterogeneity, which influences the distribution, and incidences of parasites and other pathogens and risk to disease infection. For example, the tick vector *Ambylomma hebraeum* distribution determines the

heartwater disease prevalence in South Africa. This tick vector has been found in altitudes ranging from 0-1525m with variable rainfall between 300mm and 800mm (Theiler 1948). Mdladla *et al.* (2016b) characterized disease prevalence targeting different breeds, production systems, and provinces, using heartwater as a model disease. Heartwater seroprevalence differed across breeds, age groups, sex and tick infestation status of the goats, with high prevalence rates in non-descript village indigenous goats. Province, breed, tick infestation, endemic status of the region, and the production system were identified as risk factors for heartwater exposure. Disease incidence also varies within the provinces with most cases occurring in the Eastern Cape, Limpopo, and North West, fewer in Mpumalanga, KwaZulu-Natal, and Gauteng while most disease cases are reported during the rainy seasons (South African National Department of Agriculture, Disease database, 2014). Nematode burden is more pronounced in the summer rainfall regions (Vatta *et al.* 2001a; Mbuh *et al.* 2008) and worm populations decline with the lowest numbers being encountered during the dry seasons (Gwaze *et al.* 2009c). These disease profiles and patterns can be used in goat improvement programs for veterinary interventions and farm level input such as providing animals with supplementary feeds or ensuring that they have adequate food to enable coping with disease pressure. However, research to assess the impact of diseases and parasites in the different farming systems in different agro-ecological zones in South Africa is scant. In addition, the epidemiology of diseases at community level is not clearly understood. In most village livestock production systems, diseases are not properly documented in terms of etiology and burden making it difficult to conduct association studies and identify resistant or tolerant populations.

2.5 High phenotypic diversity from a heterogenous landscape

Phenotypic characteristics based on morphological traits can provide a good representation of differences amongst populations. Morrison (2007) described the three main South Africa goat ecotype populations of the Nguni, Eastern Cape Xhosa, and Northern Cape Speckled in detail and outlined their phenotypic characteristics. The coat patterns, colors and hair type of indigenous goats are highly variable indicative of the uncontrolled and indiscriminate mating of stock attributed to their production system. This phenotypic diversity portrays the breadth of the goats' adaptability to the various ecological zones in which they are raised (Yakubu *et al.* 2010b; Hagan *et al.* 2012).

2.6 Genetic diversity of goat populations

Genetic diversity analyses shed light on domestication events, relationships among breeds, within-breed genetic diversity and breed structure and are essential for proper breed utilization and establishment of conservation priorities (Toro *et al.* 2009). In the past, microsatellite markers were instrumental in providing an insight into the genetic structure and variation among South African goat populations (Kotze *et al.* 2004; Visser *et al.* 2004; Kotze *et al.* 2014), and parentage verification in Angora goats (Visser *et al.* 2011). Kotzé *et al.* (2014) observed average heterozygosity of 63% in Kalahari Reds using 18 microsatellites markers nine of which were used in the study by Visser *et al.* (2004). Furthermore, microsatellites have recently been used to study genetic variation of the Tankwa feral goat population, which showed it to be highly divergent from the other farmed populations (Kotze *et al.* 2014). In spite of their common use in most livestock diversity studies, microsatellites are often criticized for their usual location in the non-coding regions of the genome and for not being directly associated with genes that affect phenotypes. This has led to low-density microsatellites finding little application in studies of adaptive genetic diversity of local breeds.

Genome-wide SNP genotyping technology has emerged as a powerful tool for population genetic studies in a variety of indigenous livestock species in Africa (Mbole-Kariuki *et al.* 2014; Khanyile *et al.* 2015; Makina *et al.* 2015). The application of this technology in goats became possible with the advent of the 2.66Gb reference goat genome sequenced from a female Yunnan black goat (Dong *et al.* 2013) as well as the availability of the first genome-wide Goat SNP52K (containing >50,000 probes) genotyping array (Tosser-Klopp *et al.* 2014). Kijas *et al.* (2013) was the first to report use of the array and differentiated goat populations into their breeds providing more evidence for breeds' histories and further identified 10 SNPs on chromosome 1 that were strongly associated with polledness in Boer and Rangeland goats. It has since been used to infer genetic diversity and population structure of Italian goats (Nicoloso *et al.* 2015) and South African commercial dairy and fiber goats (Lashmar *et al.* 2016), implementation in genomic evaluation (Carillier *et al.* 2013), and to investigate patterns of linkage disequilibrium (LD) and consistence of gametic phase in Canadian goats (Brito *et al.* 2014). A study on commercial, non-descript village and feral goat populations of South Africa using the SNP50 panel clearly differentiated the goat populations according to their historical origins and represented the genetic distinctiveness of the Tankwa breed from the domesticated breeds (Mdladla *et al.* 2016a). The clustering of the village goat populations

supported the breed distribution reported by Morrison (2007). The population differentiation suggested further investigation of the association between genetic markers and geographic and environmental data to better unravel the adaptive potential of these goat populations.

2.7 Signatures of selection

Identifying genomic regions that are affected by selection is important to understand the domestication, artificial and natural selection history of goat populations as well provide insights into biological pathways underlying phenotypic and economically important traits and breeding goals (Dillon *et al.* 2014). Kim *et al.* (2016) used two selection approaches ($F_{ST,}$ and integrated Haplotype Score (*iHS*)) to investigate signatures of selection between goats from cold humid temperate versus hot arid tropical regions. The study revealed candidate genes involved in thermo-tolerance (melanogenesis) (*FGF2*, *GNAI3*, *PLCB1*), body size and development (*BMP2*, *BMP4*, *GJA3*, *GJB2*), energy and digestive metabolism (*MYH*, *TRHDE*, *ALDH1A3*), and nervous and autoimmune response (*GRIA1*, *IL2*, *IL7*, *IL21*, *IL1R1*) needed for adaptation to hot arid environmental conditions. A combination of outlier detection methods with the environmental and production landscape analysis presents a good tool to accurately assess the level of association between specific genomic regions and environmental (including geographic) characteristics to make inferences on adaptive genetic structure of populations to different habitats.

2.8 Landscape genomics

Landscape genomics refers to the correlation between genomic data and environmental parameters to infer on the genomic basis of local adaptation (Luikart *et al.* 2003; Joost *et al.* 2007). It also identifies the environmental characteristics that drive divergent selection (Manel *et al.* 2010). Landscape genomics facilitates modeling of different selection pressures to determine those to which the genome has responded resulting in elevation of allele frequencies in those genomic regions responsible for adaptation. Contrary to conventional genome-wide association analysis (GWAS), landscape genomics analysis can still be undertaken using minimum and indirect measures of phenotypic data. Geographical coordinates are often used as proxies for the climatic and production selection pressures animals are exposed to, which will be modeled to investigate genetic adaptation. Such an

analysis is therefore ideal for non-descript indigenous livestock species kept by smallholder farmers.

A wide variety of data types are required for a landscape genomics analysis including geographical boundaries, socio-economic and socio-demographic data, animal husbandry practices within the production systems, population, environmental and climatic and genetic data (Joost 2014). Geographic coordinates, via Geographic Information System (GIS) tools, constitute additional descriptors or variables in the data sets (generally X for longitude and Y for latitude) for interconnection of different thematic databases such as molecular, economic and climatic data. The South African Weather Service and other local geography agencies have coordinated the collection of relatively good quality and numerous meteorological observations (Ziervogel *et al.* 2014) that allow elucidating the impact of climate on animals and reveal the mechanisms shaping adaptation and survival to these environments.

Historically, amplified fragment length polymorphisms (AFLPs) and microsatellite markers were used for landscape genetic studies in goats (Pariset *et al.* 2009; Colli *et al.* 2014). With the recent advances in high throughput sequencing and availability of high-density SNP markers, genetic adaptation analyses can adopt a whole genome approach. The locus-specific Spatial Analysis Method (SAM; Hagan *et al.* 2012) is the most widely used landscape genomic method in combination with other methods (Joost *et al.* 2007; Joost *et al.* 2008). This method assumes that an environmental variable creates selective forces that favor different genetic variants, which can be identified through correlation of marker frequencies with environmental variables. Pariset *et al.* (2009) used low-density AFLP markers and two approaches of the F_{ST} based and landscape genomics method SAM to detect unequivocal associations between genetic and environmental data. Five markers were detected using the F_{ST} based method whereas 16 markers, associated with an environmental parameter, were detected using the SAM method. Only three markers were common between the two methods (Pariset *et al.* 2009).

Methods that control for population structure minimize chances of picking false positives and include BAYENV (Coop *et al.* 2010), BAYENV2 (Günther & Coop 2013), and the latent factor mixed modeling (Frichot *et al.* 2013). Colli *et al.* (2014) observed loci that were significantly associated with environmental variables such as diurnal temperature range,

frequency of precipitation, relative humidity and solar radiation in European and West Asian goat breeds.

2.9 Potential of landscape genomics in South African goat studies and goat improvement

Non-descript indigenous goats constitute a highly variable and valuable genetic resource for marginalized smallholder farming communities. There are a number of selection pressures to which local goat populations must adapt in order to survive and optimally produce in their respective production systems and environments. The ability of indigenous livestock breeds to survive and produce under the harsh conditions prevailing in South Africa is attributed to their unique ability to take advantage of marginal areas and the environment in which they occur. Landscape genomic analysis brings complementary and valuable tools for understanding how selection in low input indigenous goat populations is shaping local adaptation.

It is important, for sustainable goat improvement, especially in village production systems, to understand ecologically adaptive traits of non-descript indigenous veld goat populations. A good understanding of management practices such as husbandry, flock size and the identification of the major production constraints is required. Also important is the establishment of traits under selection to the specific production environment targeted in improvement programs. Landscape genomics allows for the detection of signatures of selection associated with adaptive genetic variation. Such information can complement conventional genetic evaluations for breed improvement.

Landscape genomic studies require sampling of individuals from heterogeneous geographical landscapes across agro-ecological zones, and genotyping these individuals. Studies have profiled the goat production landscapes of South Africa focusing on diseases and pests (Marume *et al.* 2013; Mdladla *et al.* 2016a) and nutritional challenges (Gwaze *et al.* 2009b; Marume *et al.* 2013) to a greater extent. However, the developments in whole genome sequencing strategies and availability of genome wide SNP arrays makes it feasible to screen the genomes of goats and other livestock for genetic variants and associate them with climatic and other environmental variables to identify adaptive loci. A number of South African goat populations have been genotyped using the available Illumina goat SNP50K panel (Lashmar *et al.* 2015; Lashmar *et al.* 2016; Mdladla *et al.* 2016a). Efforts are underway to characterize the genome sequences of the Tankwa feral goat and other goat populations (Mohlatlole *et al.* 2015). Landscape genomics will facilitate an analysis that combines production landscape

factors and genetic markers that are associated with climate and ecologically important factors to investigate animal genotypes best suited for the production system.

Costs for national and regional sampling and genotyping large sample sizes are often prohibitive in developing countries such as South Africa. As such, there is need for regional and international collaborations, pooling resources, for the success of landscape genomics. These initiatives further permit sharing of genomic data from diverse regions and sharing of expertise in bioinformatics tools. Collaborative efforts with research groups like the ADAPTmap (http://bioinformatics.tecnoparco.org/adaptmap/), which is an international effort planned in collaboration with the International Goat Genome Consortium (IGGC; http://www.goatgenome.org) and with the Feed-the-Future program of the USAID (http://www.feedthefuture.gov), can be regarded as possible avenues.

CHAPTER 3: CHARACTERIZATION OF THE VILLAGE GOAT PRODUCTION SYSTEMS IN THE RURAL COMMUNITIES OF THE EASTERN CAPE, KWAZULU-NATAL, LIMPOPO AND NORTH WEST PROVINCES OF SOUTH AFRICA

Published in the Journal Tropical Animal Health and Production DOI 10.1007/s11250-017-1223-x

ABSTRACT

Expansion of goat improvement programs requires exploration of the factors that influence the production system and breeding initiatives. Characterization of goat breeds or populations is crucial in providing information on prevalent goat types and their attributes, and may suffice as a guideline on conservation, development, and selection for improved productivity. This study investigated the existing village goat production systems and phenotypic diversity of different village populations from four South African provinces. The study further investigated the use of phenotypic attributes to classify goats to breeds or populations. Data was collected from 130 households in 26 villages of the Eastern Cape (n = 2 villages), KwaZulu-Natal (n = 6 villages), Limpopo (n = 13 villages), and North West (n = 5 villages) provinces through a survey. Individual interviews and focus group discussions revealed that the mean goat flock size per household was least in Limpopo at 13.2 ±12.40 and highest in Eastern Cape (34.18±28.36). Flocks had more ($p<0.05$) adults than kids and the distribution of breeding animals was biased towards does and less bucks. Goats were kept mainly for meat, selling and for ritual ceremonies. The goat production systems was mainly scavenging. Goat health was a major challenge across households and villages. Qualitative traits such as coat, horn, ears, and wattle characteristics were recorded for populations of village goats (n = 319) and a feral Tankwa breed (n = 50). The dominant coat colour pattern was plain (74.53%) with black as the most common coat color (31.77%). Across breeds, a majority (88.08%) of the goats had horns, and 7.59% had wattles while 56.64% had beard. Adult goats (>3yrs; n = 398) were further analyzed for five quantitative traits, chest girth, height, length, and pin bone, revealing significant ($p <0.05$) breed differences in all. The disciminant function analysis showed the proportion of individuals correctly assigned into their original population

was highest in the Zulu goats (90.41%), then Xhosa village population (82.93%), and Tankwa, whilst none of the Savanna goats were correctly classified. The Zulu and Venda village goats clustered together for the first and second principal components, whilst Tankwa goats clustered with Tswana and Xhosa. Village goat production environments offer considerable scope for improvement. The heterogeneity in phenotypic traits reflects of the role of village production system in the maintenance of animal diversity in local populations.

Keywords: *characterization, morphometric traits, phenotypic diversity, village goats*

3.1 Introduction

Small livestock such as goats owned by the resource limited rural communities (Kristjanson *et al.* 2001) are an important contributor to reduced poverty and improved livelihoods of marginalized families (Peacock *et al.* 2005; Randolph *et al.* 2007). Indigenous goats play a vital role in the livelihood of rural communities through provision of protein mainly from the meat, generation of extra cash income and for religious and/or cultural purposes (Braker *et al.* 2002). Village goat production is an integrated component of nearly all-rural households in communal areas and accounts for more than 60% of the total goat population in South Africa (Shabalala & Mosima 2002; du Toit *et al.* 2014). Kunene and Fossey (2006) projected a 20.20% contribution of livestock to total smallholder household income. Indigenous goats are renowned for their adaptation to harsh environmental conditions of heat and humidity, their ability to use limited forage and resistance to endemic diseases (Casey & Van Niekerk 1988; Donkin *et al.* 1992; Barry & Godke 2001; Morand-Fehr *et al.* 2004; Kunene & Fossey 2006). Goats constitute a valuable genetic resource for low input village households. Challenges in communal goat production are mainly the low levels of inputs that farmers can provide (Mpofu 2002), exposure to diseases and parasites, poor nutrition and harsh climatic conditions (Webb & Mamabolo 2004; Sebei *et al.* 2004b), resulting in low productivity because of the compromised production systems. Indigenous goats are predominant in these marginalized farming areas, which is attributed to their variable genetic characteristics that enable them to strive in low input systems of most smallholder-farming communities.

There are efforts to improve goat production systems for smallholder farmers. In South Africa, the success of goat genetic improvement programs has been limited by the paucity of

information on traits of economic importance such as milk and meat yield and failure to account for the challenges posed by the production system and environmental constraints. Genetic improvement programs would be more successful if they were accompanied by better understanding of the production system, the farmers' breeding objectives and their production challenges. Generally, there is limited information on village goat production in South Africa as most studies target commercial flocks or village communities of smaller geographical extent. The knowledge of goat farming systems in the context of other livestock species is essential for identifying factors to improve selection programs in these areas. The production system dictates the breeds of goats that are kept and defines feasible strategies for genetic improvement.

Breed of goat is regarded as the operation unit (Duchev *et al.* 2006) and the first step in the characterization of local goat genetic resources when assessing morphological qualities among breeds (Delgado *et al.* 2001). Morrison (2007) described three goat "ecotypes" of the Nguni type (iMbuzi), Eastern Cape (Xhosa lob ears) and the Northern Cape (Speckled) in South Africa that are defined by their distinctive phenotypic, however, this does not accommodate all village phenotypes as illustrated in Figure 3.2. South Africa also has a feral goat breed, the Tankwa, which is found in the Northern Cape province of the country and is currently kept under conservation. There are currently limited characterization studies on South African goats (Pieter 2007), with a bias towards established commercial breeds. An understanding of the indigenous goat genetic resources is lacking and there is no comprehensive list of the breeds or goat ecotypes existing in the different goat farming regions. The limitation is evident in the paucity of information in national inventories and discrepancies in the ecotype descriptions therein. Characterization of goat breeds or populations is crucial in providing information on prevalent goat types and their attributes, and may suffice as a guideline on conservation, development, and selection for improved productivity (Huson *et al.* 2014). Uncontrolled breeding that is common in smallholder production systems can potentially diminish breed purity (Yakubu *et al.* 2010a) and subsequently obliterate the expected breed characteristics. Having defined breed or population characteristics that can be used to distinguish between different goat types therefore assumes an important role in these production systems. Allocations of goats to their correct breed and/or population gives insight into the genetic relationships among them which aids in guiding optimal utilization, management and conservation of unique genetic resources.

The goat production system, socio-economic factors and environmental characteristics that influence goat production need to be characterized and understood for successful improvement programs. The morphological characteristics of the local goats and the parameters that farmers use to describe and define their goat populations give some insight into the drivers of genetic improvement through selection in these local farming communities and require characterization. The objectives of this study were a) to investigate village production characteristics and production challenges in different geographic regions of South Africa using a questionnaire survey, b) to assess phenotypic variations among the populations by taking qualitative and quantitative morphological traits and investigate population structure using a set of discriminatory qualitative and quantitative variables. Findings from this study are expected to augment the existing scarce information on the village production system, and phenotypic diversity and improve the current classification of South African goat populations.

3.2 Materials and methods

3.2.1 Description of study sites

The study was conducted in the Eastern Cape, Limpopo, KwaZulu-Natal, and North West provinces of South Africa, the major goat producing provinces in the country (Moloko 2011). A total of 26 villages in the rural communities of the Eastern Cape (n = 2), KwaZulu-Natal (n = 6), Limpopo (n = 13) and North West (n = 5) (Figure 3.1) were randomly selected. Eleven households were selected in the Eastern Cape provinces from the Lower Desi and Hobeni villages in Elliotdale, Amatole District. A total of 46 goat farming households were selected in 6 villages (Coniliva, Guqa-Mkhovini, Mathengwenya, Mbomeni, Ncunjana, and Ngubo) in Tugela Ferry, Umzinyathi District of KwaZulu-Natal. Fifty-five households were sampled from thirteen villages in Vhembe District, Limpopo province. In the North West province, 18 households from two districts, Dr Kenneth Kaunda and Bojanala, participated in the study (APPENDIX A). Selection of households was based on the distribution of goat farmers in these provinces and their willingness to participate in the survey. Local agricultural extension officers assisted in identifying goat farming households.

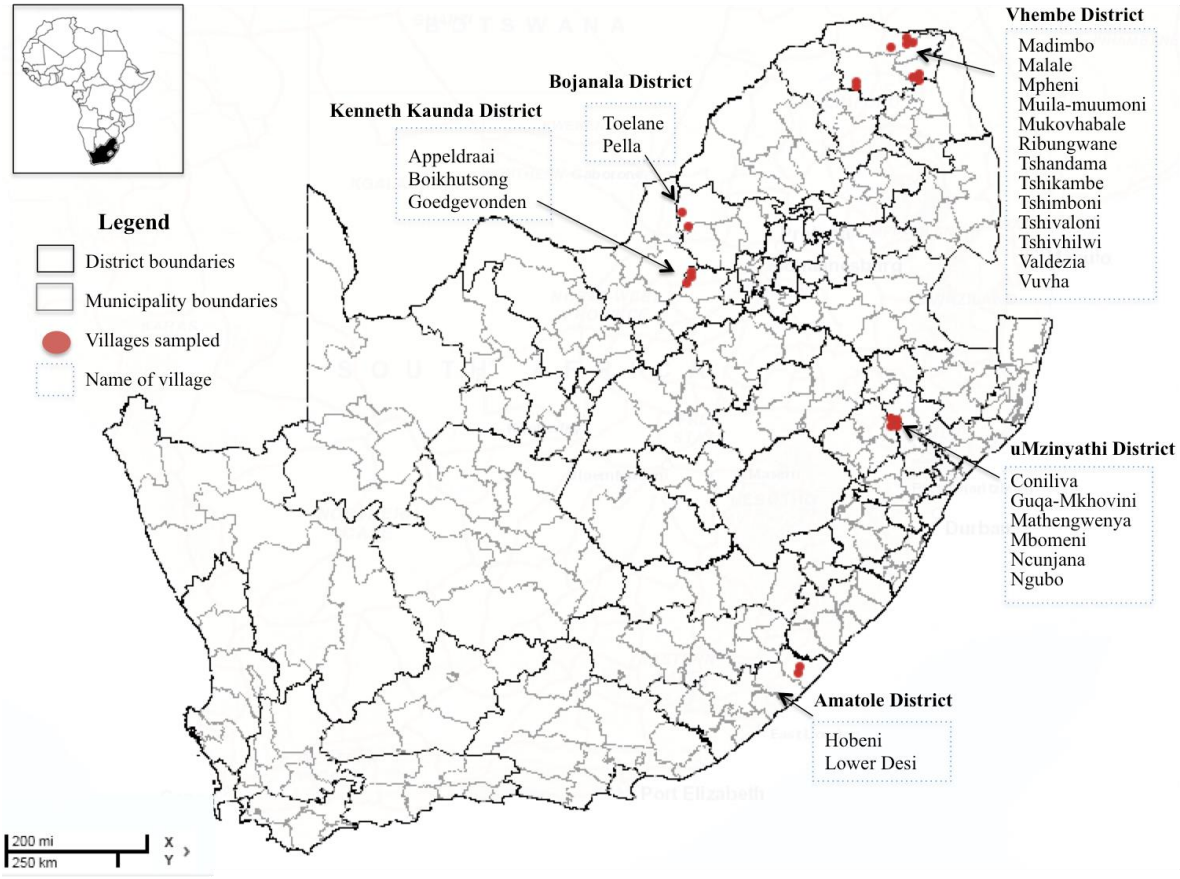


Figure 3.1 A map of South Africa showing the location of the villages included in the study. Names of the villages are outlined in the boxes

3.2.2 Questionnaire survey

Semi-structured questionnaires were used as instruments to collect data from the goat farmers. Data collected included flock structures such as flock size and composition (age and sex), role of goats at household level, management practices focusing on food and water resources, provision of housing, and other livestock species kept by each household. Data was also collected on farmer's perception on goat production constraints that included aspects of nutrition and health. Farmers indicated the type of feeds provided to their goats and the disease symptoms experienced by, their flocks. All goat farmers interviewed did not keep records therefore information on feed resources and prevalent disease symptoms were based on memory recall. Farmers also had limited access to veterinary extension and as a result reported disease symptoms and not the diseases diagnosed. Farmers had information on other

diseases such as heartwater, Orf and gall sickness, which was provided through the Veterinary Department scheduled disease surveillance. Heartwater is endemic in South Africa and to ascertain its occurrence, a separate study focused on seroprevalence of *Ehrlichia ruminantium* antibodies in the studied goat populations results of which are reported in Mdladla *et al.* (2016b).

3.2.3 Goat populations and measurements of qualitative and quantitative traits

Non-descript indigenous goats were selected randomly from each of the households in the different provinces visited. Goat populations, sampled from each province were considered as belonging to a different ecotype and named according to the indigenous language of the respective farming community. Data on five qualitative traits were collected from 369 goats belonging to the Tswana (n = 43) goats from North West, Zulu goats (n = 114) from KwaZulu-Natal, Xhosa goats (n = 52) from Eastern Cape and Venda goats (n = 110) from Limpopo as well as a feral Tankwa (n = 50) goat population from the Carnarvon Research Station in the Northern Cape province. Qualitative traits included coat pattern (which was described as either plain, patchy, spotted or speckled) and coat dominant color. Examples of phenotypic descriptions represented by the different goats are shown in Figure 3.2. Body conformation characteristics such as the presence or absence of horns, a beard and wattles were also documented.

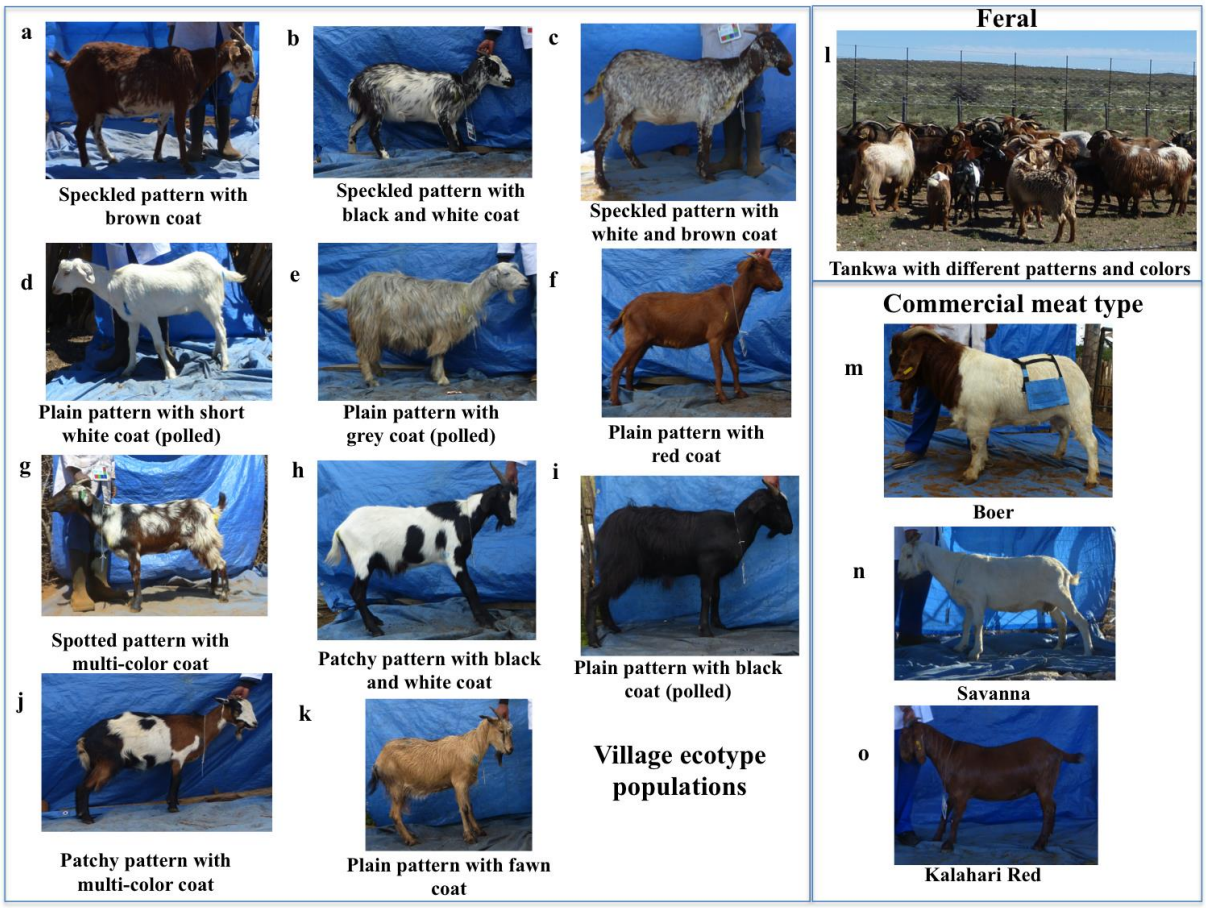


Figure 3.2 Diversity in qualitative characteristics of goats in the study. A-J: village non-descript goats; L: feral Tankwa goat; M-O commercially developed meat type goats (Boer, Savanna, and Kalahari Red)

Five quantitative body measurements were taken from adult goats (>3 years) with teeth count between four to a full mouth (Sisson & Grossman 1953). The body measurements included chest girth (CG) that was defined as the body circumference directly behind the forelegs, chest width (CW) described as points between shoulder bones, height (H) measured as the distance from the front hoof to point of withers top of shoulder blades, length (L) between the points of shoulder to pin bone, and pin bone (PB) which was defined as the distance between rear bones on either side of anus (Figure 3.3). The measurements were taken by the same person on all sampled animals in order to avoid between-individual variations. A measuring tape in centimeters (cm) was used on restrained and calm animals that were standing squarely on all four legs. Once phenotypic measurements were done, the goat was released from the

holding pen to avoid measuring it again. A total of 398 goats belonging to the commercially developed meat type, Boer (*n* = 76), Kalahari Red (*n* = 28), and Savanna (*n* = 32) from various farms across the country; and Tankwa feral (*n* = 27) and the village ecotype populations (Zulu (*n* = 73), Xhosa (*n* = 41), Venda (*n* = 81), and Tswana (*n* = 40) were characterized for quantitative phenotypic attributes.

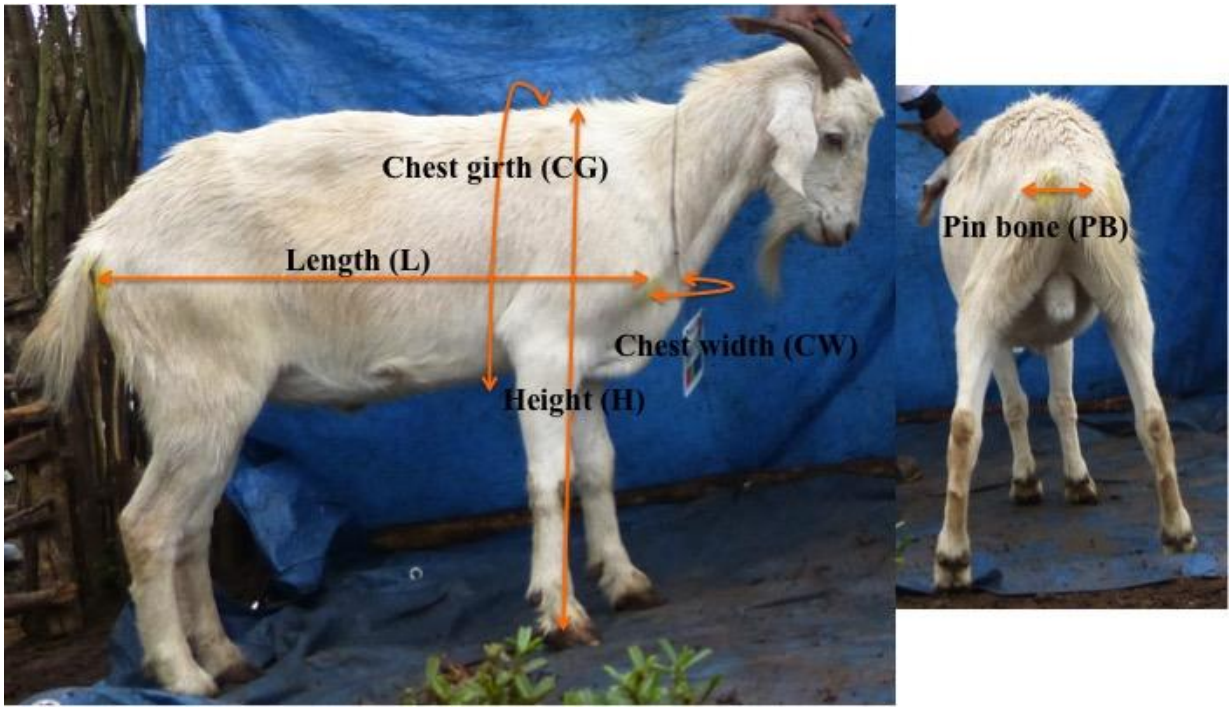


Figure 3.3 Morphological measurements taken from goats in the study. The body measurements included chest girth (CG), chest width (CW), height (H), length (L), and pin bone (PB)

3.2.4 Statistical analysis

Descriptive statistics for the flock composition and management practices were estimated using the PROC UNIVARIATE and PROC FREQ analysis of Statistical Analysis System (SAS institute Inc. 2013). Descriptive statistics for the qualitative characteristics were determined using the PROC FREQ analysis (SAS institute Inc. 2013). General Linear Model (GLM) procedures were used to analyze quantitative data and ascertain the effects of breed, sex and interaction between breed and sex. Where significant differences in means were observed, means were separated using the Least Significant Difference (LSD) at 5% level

significance.

A stepwise discriminatory procedure was applied using PROC STEP DISC (SAS institute Inc. 2013) to rank the quantitative traits according to their discriminatory power to separate breeds or populations. Significant traits were then subjected to canonical discriminant analysis (PROC CANDISC) and discriminant function analysis (PROC DISCRIM) of Statistical Analysis System (SAS institute Inc. 2013) to classify individuals to their original breed/population. Principal component analysis (PCA) was applied using the Excel add-in Multibase package (Numerical Dynamics, Japan) to cluster individuals using the significant body measurements. The analysis was fixed for sex and age.

3.3 Results

3.3.1 Village production system characteristics and management

The mean flock size per household (**Table 3.1**) was lowest in Limpopo (13.2±12.40) followed by North West (19.89±15.81) and was highest in Eastern Cape (34.18±28.36). High variation was observed within provinces as indicated by the high standard deviations. The flock composition followed a similar trend in all provinces and comprised mostly of adult goats and less kids. Bucks accounted for 28.6%, 21.7%, 21.3%, and 19.4% of total adult flocks in North West, Eastern Cape, KwaZulu-Natal and Limpopo provinces, respectively. The ratio of breeding does to breeding buck was 4:1 for Eastern Cape, KwaZulu-Natal and Limpopo and 2:1 for North West. Three (4.34%) and seven (12.7%) households did not own a buck in KwaZulu-Natal and Limpopo provinces, respectively.

Table 3.1 Mean indigenous flock sizes and composition in four South African provinces

Indigenous goat classes		Number of households sampled			
		Eastern Cape (*n* = 11)	**Kwazulu-Natal (*n* = 46)**	**Limpopo (*n* = 55)**	**North West (*n* = 18)**
Age categories	Adult	27.18±21.94	30.69±33.91	10.03±10.32	14.39±11.77
	Kids	7.00±7.54	1.80±3.36	0.85±2.03	5.50±4.99
Sex categories	Does	15.55±15.65	21.85±29.52	8.03±8.34	9.59±8.98
	Buck	5.78±4.84	5.73±4.45	2.32±3.04	4.35±4.39
	Castrates	2.55±2.77	1.59±3.84	0.20±0.61	2.44±3.33
Total mean		34.18 ±28.36	33.87±34.58	13.20±12.40	19.89±15.81

The animals were kept mostly for household meat consumption, selling, and rituals (**Table 3.2**). Majority (92.73%) of the animals in Limpopo were kept extensively, where animals browsed and grazed on natural pastures with 10 (90.91%) and 28 (60.87%) farmers providing supplementary feeds in Eastern Cape and KwaZulu-Natal, respectively. Majority of the goats accessed water from natural water resources (rivers and natural water ponds) in the Eastern Cape (*n* = 11; 100%) and Limpopo (*n* = 51; 92.73%), whilst tap and dams were the main source for animals in KwaZulu-Natal (*n* = 17; 36.96%) and North West (*n* = 15; 83.33%). All goats were provided with housing. Very few households kept goats only and were mainly in Limpopo (*n* = 12; 21.82%). Other livestock species kept with goats were chickens, cattle, sheep and pigs (**Table 3.3**).

Table 3.2 Frequencies (percentage) of household role of goats, and management in the production system per province

Parameters	Study provinces (number of households sampled)			
	Eastern Cape (*n* = 11)	Kwazulu-Natal (*n* = 46)	Limpopo (*n* = 55)	North West (*n* = 18)
Uses				
Meat	7 (63.63)	42 (91.30)	50 (90.91)	15 (83.33)
Selling	11 (100)	35 (76.09)	24 (43.63)	11 (61.11)
Rituals	7 (63.63)	21 (45.65)	0 (0)	2 (13.33)
Water resource				
Natural resources	11 (100)	29 (63.04)	51 (92.73)	3 (16.67)
Taps and dams	0 (0)	17 (36.96)	4 (7.27)	15 (83.33)
Feed practices				
Scavenging	1 (9.09)	18 (39.13)	51 (92.73)	5 (27.78)
Scavenging with supplementation	10 (90.91)	28 (60.87)	4 (7.27)	13 (72.22)

Table 3.3 Frequencies (percentage) of livestock species kept in the househollds per province

Household livestock	Study provinces (number of households sampled)			
	Eastern Cape (*n* = 11)	Kwazulu-Natal (*n* = 46)	Limpopo (*n* = 55)	North West (*n* = 18)
Goats only	0 (0)	1 (2.17)	12 (21.82)	3 (16.67)
Chickens	11 (100)	40 (86.96)	41 (74.55)	11 (61.11)
Cattle	11 (100)	28 (60.68)	11 (20.00)	11 (61.11)
Sheep	9 (81.82)	7 (15.22)	1 (1.82)	6 (33.33)
Pigs	2 (18.18)	0 (0)	0 (0)	1 (5.55)

3.3.2 Goat health issues

Overall, goat health issues were perceived to be the main constraint in village goat production with 60.76 % (n = 79) of the farmers indicating internal parasites to be the most prevalent health challenge followed by external parasites (n = 68; 52.3%). Some diseases were specific to certain provinces and not reported by others. Majority (n = 50; 90.91%) of the respondents in Limpopo reported heartwater as the main disease challenge. Only 47.83% (n = 22) and 18.18% (n = 2) reported heartwater as a constraint in KwaZulu-Natal and Eastern Cape, respectively. Based on feedback reports by veterinary extension workers, farmers in KwaZulu-Natal and Eastern Cape, reported gall sickness and Orf viral infection. While majority of the farmers in Eastern Cape (n = 11; 100%) and North West (n = 17; 94.4%) vaccinated their animals, vaccinations were not very common in KwaZulu-Natal (n = 16; 34.78%) and Limpopo (n = 18; 33.0%).

3.3.3 Goat characterization

As shown in **Table 3.4**, the coat color patterns varied within and amongst the goat populations. Higher number of Tankwa goats were black (36%), followed by brown (28%) and white and red (14%). White was the most common color amongst the Tswana (48.84%) and Xhosa (46.15%) village ecotypes, while black was common in Venda and Zulu ecotype populations (**Table 3.4**). Majority (92.41%) of the goats lacked wattles/toggles, whilst 56.64% of goats had a beard. The beard was observed in both sexes across breeds. A proportion (88.08%) of the goats had horns, while 11.92% were polled. There were no significant differences observed on the presence of horns between male and female goats from the different ecotype populations (**Table 3.4**).

The descriptive statistics of the linear body measurements of adult goats are presented in **Table 3.5**. Analysis of variance revealed significant (p < 0.05) differences in all the morphometric measurements, with higher means observed in Tankwa goats for chest girth (CG; 84.10± 2.54 cm), height (H; 76.26± 2.94 cm), length (L; 83.26±1.41 cm), and pin bone (PB; 11.37±0.34 cm). Zulu goats had the highest mean for the chest width (CW) at 34.95±2.20 cm but the lowest CG (37.17±1.73 cm) compared to the village ecotype populations. Breed effects were highly significant (p < 0.05) for all body measurements, while

the effects of sex were significant ($p < 0.05$) for H, L, and CW. The interaction of breed and sex was highly significant ($p < 0.05$) for PB measures.

Table 3.4 Percentage (%) of indigenous goats showing the various morphological characteristics in the Tankwa and village ecotype populations

Catergory		Number of individuals measured for each population/breed					
		Tankwa (*n* = 50)	**Tswana (*n* = 43)**	**Venda (*n* = 110)**	**Xhosa (*n* = 52)**	**Zulu (*n* = 114)**	**Overall (*n* = 369)**
Coat pattern	**Plain**	50.00	37.21	85.45	75.00	88.60	74.53
	Patchy	46.00	41.86	10.00	15.38	0.00	16.26
	Spotted	4.00	2.33	3.64	3.85	2.63	3.25
	Speckled	0.00	18.60	0.91	5.77	8.77	5.96
Dominant colour	**Black**	36.00	16.28	36.36	13.46	40.35	31.98
	Brown	28.00	16.28	18.18	21.15	14.91	18.70
	Fawn	4.00	2.33	2.73	9.62	5.26	4.61
	Grey	4.00	4.65	4.55	5.77	9.65	6.23
	Red	14.00	11.63	8.18	3.85	13.16	10.30
	White	14.00	48.84	30.00	46.15	16.67	28.18
Wattle	**Presence**	0.00	4.65	4.55	19.23	9.65	7.59
	Absence	100	95.35	95.45	0.77	90.35	92.41
Beard	**Presence**	86.00	53.49	36.36	82.69	52.63	56.64
	Absence	14.00	46.51	63.64	17.31	47.37	43.36
Horn	**Polled**	0.00	44.19	7.27	5.77	12.28	11.92
	Horned	100	55.81	92.73	94.23	87.72	88.08

Table 3.5 Least square means ±standard error, and the level of significance for the effect of breed, sex, the interaction between breed and sex on quantitative trait measurements

¥Quantitative trait	Number of individuals measured for each population/breed								Significance		
	Boer (*n* = 76)	Kalahari Red (*n* = 28)	Savanna (*n* = 32)	Tankwa (*n* = 27)	Tswana (*n* = 40)	Xhosa (*n* = 41)	Venda (*n* = 81)	Zulu (*n* = 73)	B	S	B*S
CG (cm)	79.99±1.73$^{d;e;f}$	75.04±2.96$^{c;d;e}$	79.01±2.67^{d}	84.10±2.54^{f}	71.60±2.0^{c}	76.08±1.88^{e}	51.02±1.66^{b}	37.17±1.73^{a}	***	NS	NS
H (cm)	62.14±2.0	60.98±3.43	62.11±3.10	76.26±2.94^{c}	68.58±2.32^{b}	65.50±2.18$^{a;b}$	63.32±1.93^{a}	66.50±2.01$^{a;b}$	***	***	***
L (cm)	80.21±0.96$^{b;c}$	74.96±1.64^{b}	75.73±1.48$^{b;c}$	83.26±1.41^{c}	71.48±1.11^{a}	68.36±1.04	66.58±0.92	68.07±0.96	***	*	***
CW (cm)	34.38±2.19^{a}	31.423.77^{a}	30.53±3.40^{a}	24.00±3.23	20.53±2.11	19.21±2.40	24.55±2.11	34.95±2.20^{a}	***	**	***
PB (cm)	10.74±0.23^{d}	8.94±0.39$^{b;c}$	7.80±0.35^{a,b}	11.37±0.34^{e}	8.35±0.27^{b}	6.21±0.25^{a}	9.69±0.22$^{c;d}$	9.24±0.23$^{c;d}$	***	NS	***

¥CG = chest girth; BD=body depth; CW=chest width; H= height; L=length; and PB=pin bone. Across rows means with different superscripts were significantly different for breed effects at $p<0.0001$=***; $p<0.001$=**; $p<0.05$=*; and $p>0.05$=Not Significant (NS). B=Breed; S= Sex; and B*S= Breed*Sex interaction.

Using the Stepwise discriminatory procedure, the quantitative traits of CG, CW, PB, H, and L were observed to have significant discriminatory power (**Table 3.6**). All these traits were thus used for further canonical discriminant, discriminant function and principal component analyses.

Table 3.6 Significant traits using stepwise selection for traits to include in the principal component analysis (PCA)

Step	¥Trait	Partial R^2	F Value	*P*-value	Wilks' Lambda	ASCC
1	**CG**	0.7117	137.55	***	0.28828064	0.10167419
2	**PB**	0.3665	32.15	***	0.18262166	0.15297995
3	**CW**	0.2451	18.00	***	0.13785649	0.18562740
4	**H**	0.2150	15.14	***	0.10822072	0.21457047
5	**L**	0.0715	4.25	**	0.10047776	0.22150972

¥ CG = chest girth; BD=body depth; CW=chest width; H= height; L=length; and PB=pin bone. The *p*-values for both Wilks' lambda and ASCC (Average Squared Canonical Correlation) were significant p<0.0001=***; p<0.001=**.

The *F*-value statistics from canonical discriminant analysis suggested that the chest girth had the highest amount of significant discriminative potential, while height had the least significant discriminative power to differentiate the goat populations (**Table 3.7**).

Table 3.7 Analysis of Variance (ANOVA) from Canonical Discriminant Analysis for significant effect of breed/population

*Variable	Total SD	Pooled SD	Between SD	R^2	F Value	*P*-value
CG	20.2138	10.9501	18.2076	0.7117	137.55	***
H	14.6835	13.4379	6.6000	0.1772	12.00	***
L	8.1157	6.3011	5.5336	0.4078	38.37	***
CW	16.1988	14.2923	8.3890	0.2353	17.14	***
PB	1.9183	1.5464	1.2317	0.3617	31.56	***

*cg = chest girth; cw=chest width; h= height; l=length; and pb=pin bone. ***; Significant *p*< 0.0001.

The performance of the discriminant function analysis in classification was evaluated by estimating the probabilities of misclassification for each of the respective populations. Misclassification parameters based on the posterior probability estimates computed by the discriminant function via cross-validation are presented in **Table 3.8**. Sixteen (21.05%) Boer individuals were correctly assigned whilst 21 (27.63%) individuals each were displaced into Tswana and Tankwa goats, respectively. Ten Kalahari Red individuals and 13 (40.63%) of Savanna were assigned in the category of Tswana goats. Majority (90.41%) of the Zulu, Xhosa (82.93%) and Tankwa (74.07%) were correctly assigned into their respective populations. Thirty-two individuals of Venda were classified as Zulu, whilst seven of Zulu were classified as Venda. These misclassifications and overlaps were confirmed by the principal component weights derived from the correlation matrix of the five linear body measurements (Figure 3.4). PC1 accounted for 43.2% of the total variance whilst PC2 accounted for 29.3%. The Tankwa, Xhosa and Tswana goats formed their own cluster separated from commercial meat type breeds cluster and the Venda and Zulu ecotypes. Majority of the commercial meat type breeds clustered together with a few outliers in the Tankwa, Xhosa and Tswana cluster, whilst the Venda and Zulu village goats formed another separate cluster.

Table 3.8 Number and percentage (%) of individual goats assigned into breeds/populations

From Breed	Boer	Kalahari Red	Savanna	Tankwa	Tswana	Venda	Xhosa	Zulu	Total
Boer	16 (21.05)	6 (7.89)	9 (11.84)	21 (27.63)	21 (27.63)	2 (2.63)	1 (1.32)	0 (0.00)	76 (100.00)
Kalahari Red	5 (17.86)	3 (10.71)	3 (10.71)	5 (17.86)	10 (35.71)	1 (3.57)	0 (0.00)	1 (3.57)	28 (100.00)
Savanna	8 (25.00)	1 (3.13)	0 (0.00)	6 (18.75)	13 (40.63)	0 (0.00)	4 (12.50)	0 (0.00)	32 (100.00)
Tankwa	1 (3.70)	0 (0.00)	0 (0.00)	20 (74.07)	6 (22.22)	0 (0.00)	0 (0.00)	0 (0.00)	27 (100.00)
Tswana	1 (2.50)	1 (2.50)	1 (2.50)	6 (15.00)	23 (57.50)	2 (5.00)	6 (15.00)	0 (0.00)	40 (100.00)
Venda	2 (2.47)	1 (1.23)	0 (0.00)	5 (6.17)	19 (23.46)	21 (25.93)	1 (1.23)	32 (39.51)	81 (100.00)
Xhosa	0 (0.00)	0 (0.00)	1 (2.44)	0 (0.00)	6 (14.63)	0 (0.00)	34 (82.93)	0 (0.00)	41 (100.00)
Zulu	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	7 (9.59)	0 (0.00)	66 (90.41)	73 (100.00)
Total	33 (8.29)	12 (3.02)	14 (3.52)	63 (15.83)	98 (24.62)	33 (8.29)	46 (11.56)	99 (24.87)	398 (100.00)
Priors	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	

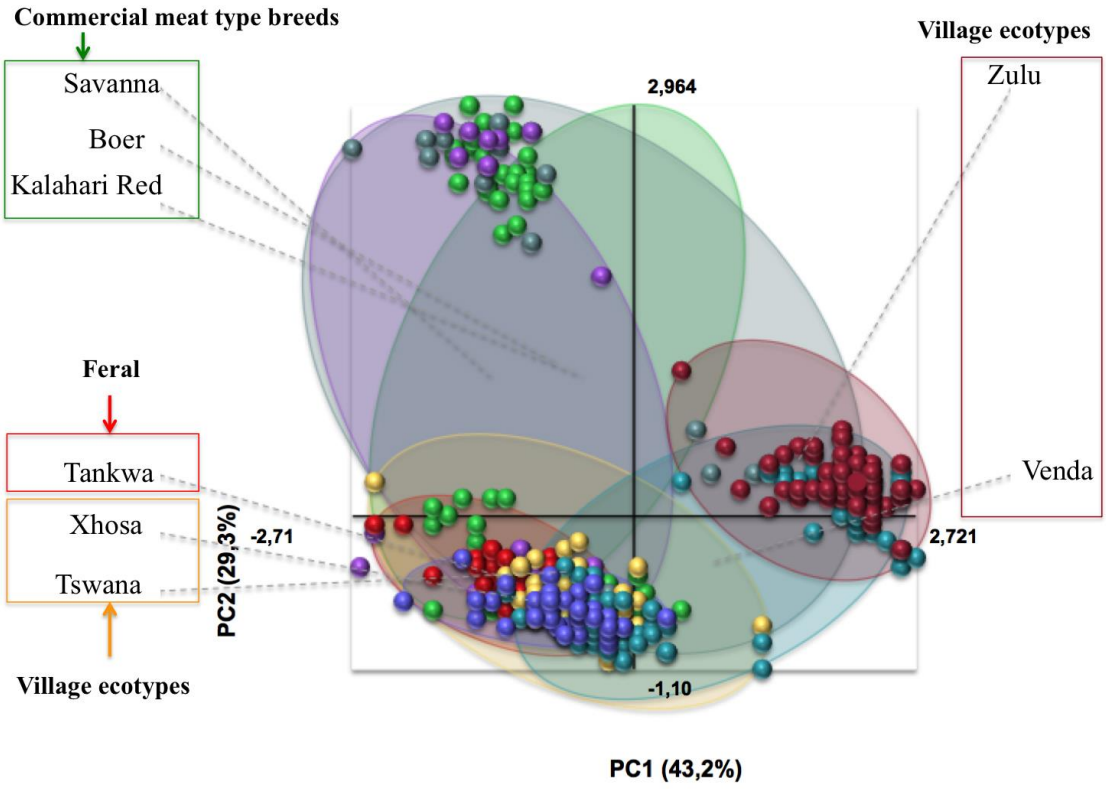


Figure 3.4 Principal Component Analysis using five quantitative traits (Chest girth, Chest width, Height, Length, and Pin bone). Ellipse represents the distribution of individuals across the PCA axis. Green circles represent individuals from the Boer, while grey, purple, red, yellow, blue, maroon, and turquoise were individuals from Kalahari Red, Savanna, Tankwa, Tswana, Zulu and Venda goats, respectively

3.4 Discussion

The mean flock size per farming household was lowest in Limpopo (13.2 ± 12.40) and highest in the Eastern Cape (34.18±28.36), with an overall average of 25.27± 22.78 across provinces. This was higher than the 9.7 goats per household reported in Mozambique (van Niekerk & Pimentel 2004) and the 18.77±11.79 reported by Angaga and Mosimanyana (2001) in Botswana. The number of does per flock outnumbered that of bucks. The buck to doe ratios reported for the different provinces in this study was lower than the 11:1 reported by Bester *et al.* (2009) for other communal areas in South Africa. Greater numbers of does to bucks were

also reported in studies in Mozambique (van Niekerk and Pimentel 2004) and Nigeria (Reynolds & Adediran 1994). Gwaze *et al.* (2009a) observed only 15% and 29% of farmers owned at least one buck in the districts of Alfred Nzo and Amatole, in the Eastern Cape province of South Africa. Our study also reported households that did not own bucks in Limpopo and KwaZulu-Natal provinces. The absence of a buck in a flock is common in small flocks and farmers with these flocks use communal bucks for breeding. Communal herding, which allows breeding does and bucks to mix among flocks, can minimize the levels of inbreeding if population numbers are high (Jaitner *et al.* 2001). The use of communal bucks is common for village production systems in Africa (Gwaze *et al.* 2009a). The castration of young bucks that are not needed for reproduction was observed in the current study and has been reported as a strategy to control breeding and improve meat quality (Gwaze *et al.* 2009a). Often, one or two bucks are retained in the flock for breeding (Webb & Mamabolo 2004) and as a result, the flock composition consists of more castrates than reproducing bucks (Mahanjana & Cronje 2000).

Sustainability of genetic improvement programs depends on the producers' breeding goals and objectives, which are influenced by socio-cultural, economic and geographical factors (Ilatsia *et al.* 2012). The role of goats in these communities determines the traits that farmers select for (Dossa *et al.* 2007) and often forms the basis for the breeding objectives. The main reasons for keeping goats in the study areas were for household meat consumption, selling when in need of income and to a lesser extent socio-cultural rituals. This is in line with observations from previous studies done in Southern Africa countries (Bester *et al.* 2009; Braker *et al.* 2002; Gwaze *et al.* 2009a; Kunene & Fossey 2006; Mahanjana & Cronje 2000; Masika & Mafu 2004). Goat production in the Eastern Cape was mainly for sale of live animals making it more of a commercially orientated production system. Focus group discussions revealed that the size and appearance of the goat as well as the financial needs of the farmer determined the price of goats. In Uganda, animal prices depended on the farmers' need for money and the price the buyer is willing to pay (Ampaire 2011).

Management interventions are strategies used by farmers to improve goat production and these include ensuring adequate and suitable feed and water, disease control, and improved housing. Feed is the most important factor in livestock production (Khan & Usmani 2005) and inadequate and poor quality feeds are often a major constraint to small ruminant

production in smallholder systems (Kosgey 2004). The exclusive scavenging system in 57.69% of the households was lower than the 98% reported in Botswana (Aganga & Mosimanyana 2001). Supplementation with crop residues depended on availability of supplements and was reported in 42.3% of the households. Household leftovers and waste, mostly in the form of kitchen remains and crop residues during harvest, were the most common form of supplementation reported in this study, which is also common in other African countries (Aganga & Mosimanyana 2001; Anaeto *et al.* 2009) due to the high cost of commercial feeds. The use of homegrown supplements to improve milk production and reproductive performance were reported in other parts of South Africa by Kunene and Fossey (2006). Access to water is a challenge for most South African rural areas resulting in majority of the animals having to fend for their own water in rivers and natural ponds. According to Census (2011), about 8.80 % of South African households do not have access to clean piped water and often these households use alternative sources of water such as rivers and wells that are shared with livestock (Sebei *et al.* 2004a). These water sources are unreliable because they often dry up during the dry seasons (Sebei *et al.* 2004a). This study revealed that majority (72.31%) of the farmers depended on unstable water resources. All farmers provided some form of housing structure made from variable but locally available and financially affordable building materials. Some households provided roofing while others did not. Webb and Mamabolo (2004) indicated that farmers were aware of the importance of providing shelter for their animals and the failure to do so or the poor quality materials used were a result of the limited resources available to the farmers (Sebei *et al.* 2004b). The enclosing of livestock in huts or pens protects them from theft and predation (Webb & Mamabolo 2004) and diseases (Devendra & McLeroy 1982; Payne & Wilson 1999). Sebei *et al.* (2004b) reported similar observations for building materials such as wood and corrugated iron for indigenous goats under communal grazing in the Jericho district in the North West province. Variations in the additional livestock species kept were also observed. It is typical of most communal households to keep goats, chickens, cattle, sheep and pigs (De Villiers & Letty 2001; Khan & Usmani 2005; Kunene & Fossey 2006; Gwaze *et al.* 2009b).

Goat health was identified as a major constraint across provinces. Parasite burden, which translates to diseases such as heartwater, was reported in the Eastern Cape, KwaZulu-Natal, and Limpopo. A sero-prevalence analysis in the studied goat populations indicated that heartwater remains a significant disease of village goats, with an overall prevalence rate of

89.54% (Mdladla *et al.* 2016b). Sero-prevalence was 85.90%, 73.77%, and 91.89%, in the Eastern Cape, KwaZulu-Natal, and Limpopo provinces respectively (Mdladla *et al.* 2016b). High prevalence of heartwater disease has been linked to lack of tick control in communal sectors (Bester *et al.* 2009). Households reported that heartwater affected kids mostly and was not as common in adult goats and was less fatal in the older than young age groups. Diarrheal cases were reported to be less frequent in all provinces except in KwaZulu-Natal. Small-scale farmers often do not treat animals for internal parasites nor vaccinate against heartwater (Sebei *et al.* 2004b). Levels of vaccination are generally low in the communal areas (Bester *et al.* 2009) due to the high costs of veterinary interventions that most smallholder farmers cannot afford. In Northern KwaZulu-Natal, goats and sheep were not dipped for tick control (Kunene & Fossey 2006). The high levels of vaccination levels in the Eastern Cape (100%) and North West (94.40%) suggests that farmers in these provinces were aware of the benefits of disease control and could possibly afford it.

Village goat production is based mainly on unimproved indigenous goats (Bester *et al.* 2009) with large variations in morphological appearances, and body size. Commercial breeds were excluded in the analysis for diversity of morphological appearances since pure breeds are used that follow standards defined by Breeders Associations with often no room for phenotypic variation. Morphological differences have important socio-cultural and economic values to the rural communities and as a result, most farmers have specific consideration and choices for goat coat colors (Mahanjana & Cronje 2000; Manton 2005; Gwaze *et al.* 2009b). The finding that white is the dominant color in the Eastern Cape is suggestive of selection towards a white color in these communities. A study by Mahanjana and Cronje (2000) found that white goats were in high demand for sacrificial purposes, and comparatively high prices were paid for them in the Eastern Cape, South Africa. Moreover, qualitative traits may represent some adaptive mechanism developed for adaptation and survival in different environment. In Ghana, dark coat colors have been linked to environmental adaptation (Hagan *et al.* 2012). Black colored goats are believed to have superior adaptation to seasonal cold weather as the dark pigment helps them to absorb heat and warm up faster than other coat colors (Robertshaw 2006). Goats are commonly used for socio-cultural ceremonies in rural communities of South Africa and preferences are made for specific coat colors. We therefore speculate that the village goats from Limpopo and KwaZulu-Natal were mostly black because of the demand for such color in cultural ceremonies. The frequency of wattle/toggles of

7.59% was lower than the 36.5% in West African Dwarf (WAD) goats (Adebayo & Chineke 2011). Varied expression of wattles in the village goats may represent some adaptive mechanisms related to milk yield as observed in Saanen goats (Shongjia *et al.* 1992). Although further conclusions cannot be drawn on the current study regarding wattles/toggles based on the collected data, Yakubu *et al.* (2010a) reported an associated taboo towards toggled village goats which we also speculate might be the case in the communities in the study. Ozoje (2002) correlated the position of wattles with tail length and neck circumference in WAD goats, while Adedeji *et al.* (2011) observed that the presence, shape and location of the wattle had a significant effect on body weight, body length, chest girth and scrotal length of WAD bucks. The frequencies of a beard were generally higher with no significant difference among male and female goats as indicated in the study by Yakubu *et al.* (2010a). In a study by Adebayo and Chineke (2011) beard frequencies differed with sex of the animal with 82% of the bucks and few does (8.3%) exhibiting the trait. The low occurrence of polledness in indigenous goat populations has been reported in Ghana (Hagan *et al.* 2012). The presence of horn is an adaptive feature especially for the Tankwa goat, to fight predation or where animals had to fight competitors for feed and water and even for does during mating.

Morphometric measurements can be used to describe the animals' production status, and breed characteristics (Cam *et al.* 2010). Previous studies have used multivariate analysis of morphological parameters to explain population structure, facilitate breed identification, and estimate genetic variation within and between indigenous goat populations from different agro-ecological zones (Traoré *et al.* 2008; Yakubu *et al.* 2010a; Yakubu *et al.* 2010b; Okpeku *et al.* 2011; Tsegaye *et al.* 2013). Hagan *et al.* (2012) reported a significant influence of age on body measurements with no differences observed for goats 3 years and above. This study therefore analysed body measures in adult goats only. The chest girth has been used to estimate live body weight in several studies (De Villiers *et al.* 2009; Hassen *et al.* 2012), and was observed to have the highest discriminatory power amongst the body measurement. The principal component analysis separated populations into production system (i.e. commercial and village) and strongly indicated an eco-geographical trend of population differentiation and perhaps cultural preferences among the ethnical groups keeping these animals and supports the historic geographic distribution of the respective goat populations (Morrison 2007). The high number of the Venda goats being classified as Zulu goats show a relatively low morphometric differentiation between these goat populations suggesting that they are

homogenous and share similar genetic identities. On average, 45.98% of the goats were correctly classified in this study, whilst Dossa *et al.* (2007) was able to correctly allocate more than 70% of individual goats into their phenotypic groups. A more successful discriminant function was able to correctly classify 100% of West African Dwarf and Red Sokoto goats (Yakubu *et al.* 2010b).

Overall, the study identified a number of production constraints as well as opportunities that could help in developing effective improvement programs for smallholder goats in South African majority of which have also been reported in studies from other African and developing countries. Changes in the current management practices can improve the performance of village goats and in turn their production outputs. There is need for appropriate disease control intervention strategies to reduce mortality and improve productivity of goats. Educating and equipping farmers on management practices to improve production such as modern breeding practices, strategic feeding during less favorable periods and health care can benefit these communities. South Africa has multiple contrasting agro-ecologies that could explain the presence of a diversified goat population. Some of the traits (i.e. coat color) reflect the adaptive fitness under extensive scavenging-type environments. The linear body measurements clustered populations according to their breed and agro-ecological zone affiliation. Information on the diversity and population structure of the village goats is crucial in developing the sustainable breed improvement and conservation strategies for phenotypes adapted to the local production environments.

CHAPTER 4: SEROPREVALENCE OF *EHRLICHIA RUMINANTIUM* ANTIBODIES AND ITS ASSOCIATED RISK FACTORS IN INDIGENOUS GOATS OF SOUTH AFRICA

Published in Preventive Veterinary Medicine 125, 99-105

ABSTRACT

The present study investigated the seroprevalence of antibodies to *Ehrlichia ruminantium* and the associated risk factors in goats from five different farming provinces of South Africa. Sera collected from 686 goats of the commercial meat type (*n* = 179), mohair type (*n* = 9), non-descript indigenous goats from Eastern Cape (*n* = 56), KwaZulu-Natal (*n* = 209), Limpopo (*n* = 111), North West (*n* = 61) and Northern Cape (*n* = 11) provinces and a feral Tankwa goat (*n* = 50) were tested for the presence of immunoglobulin G (IgG) antibodies to antigens of *E. ruminantium* using the indirect fluorescent-antibody test (IFAT). Fifty two percent of these goats had ticks. The overall seroprevalence of antibodies to *E. ruminantium* was 64.87% (445/686) with the highest seroprevalence reported for Limpopo (95.50%) and lowest for Northern Cape (20.29%). Highest seroprevalence for antibodies to *E. ruminantium* was observed in goats from endemic regions (76.09%), and from smallholder production systems (89.54%). High sero-prevalence was also observed in non-descript indigenous goats (85.04%), adult goat (69.62%), in does (67.46%) and goats infested with ticks (85.79%). The logistic model showed a gradient of increasing risk for commercial meat type Savanna (OR = 3.681; CI = 1.335-10.149) and non-descript indigenous (OR = 3.466; CI = 1.57-7.645) compared to Boer goats and for goats from the smallholder production system (OR = 2.582; CI = 1.182-5.639) and those with ticks (OR = 3.587; CI = 2.105-6.112). Results from this study showed that *E. ruminantium* infections were prevalent but were widely and unevenly distributed throughout South Africa. Findings from the study facilitate identification and mapping of risk areas for heartwater and its endeminicity in South Africa and should be taken into consideration for future disease control strategies and local goat improvement programs.

Keywords: *Ehrlichia ruminantium; Heartwater; Goats; Risk factors; Seroprevalence; South Africa*

4.1 Introduction

Heartwater is endemic to most parts of South Africa, and livestock farmed in these regions are subject to constant disease challenge. Veterinarians reported over 800 heartwater outbreaks in goats between 1993 and 2014 (South African National Department of Agriculture, Disease database, 2014) illustrating the burden of the disease in the country. Majority of the indigenous goat populations are raised for subsistence production in rural areas (Coetzee 1998) and tick control is less frequent and erratic in these areas contrary to the commercial sector.

The control of heartwater and other tick-borne diseases has been limited by the lack of population-structured epidemiological information in developing countries including South Africa. South Africa is divided into agro-ecological zones that are defined by variable geographic and climatic conditions (Palmer & Ainslie 2006). These conditions are known to influence tick vector distribution and endeminicity of heartwater (Petney *et al.* 1987; Walker & Olwage 1987; Spickett *et al.* 2011). The goat production system of South Africa is however so heterogeneous that goat populations are reared under both smallholder communal and commercial farming using different goat breeds (Bester *et al.* 2009). The influence of breeds of goats and production systems within each ecological zone on heartwater occurrence has not been investigated. Such interactions could be crucial to the development of effective and sustainable disease control measures.

Genetic variation within and between breeds for resistance to heartwater has been demonstrated by several studies (Matheron *et al.* 1987; Obexer-Ruff *et al.* 2003; Bambou *et al.* 2010). South African indigenous goats and their crosses have been shown to have resistance to the disease (Donkin *et al.* 1992). The need for identifying and promoting use of genetically resistant or more tolerant breeds has therefore been proposed as a practical alternative to the control of heartwater and other livestock diseases especially for resource-limited smallholder farmers.

There are limited studies available on the seroprevalence of heartwater in the different agro-ecological zones and endemic or non-endemic regions of South Africa (Du Plessis *et al.* 1993; De Waal *et al.* 2000) and other African countries (Mahan *et al.* 1998; Bekker *et al.* 2001; Kakono *et al.* 2003; Ahmadu *et al.* 2004; Faburay *et al.* 2005). Most of these studies identified endemic status of the regions as an important risk factor for heartwater infection.

This study investigated the prevalence of antibodies to *Ehrlichia ruminantium* in different goat breeds from different production systems of South Africa using a serological analysis. The indirect fluorescent antibody test (IFAT; Du Plessis 1981) has proven to be a valuable method for detecting the sero-conversion of animals exposed to tick infestation (Du Plessis *et al.* 1984). Overall, the study aimed to identify and quantify the effects of geographic regions, animal and production system associated risk factors for *E. ruminantium* and explore the relationships between these risk factors and prevalence of *E. ruminantium* in South African goats.

4.2 Materials and methods

4.2.1 Study sites

The study was conducted during the summer and spring months (November-December 2012, February 2013 and September-December 2013). Sampling regions were selected from the Eastern Cape, KwaZulu-Natal, Limpopo, Northern Cape and North West provinces of South Africa (Figure 4.1). Geographical coordinates (latitude, longitude and altitude) were assigned to each of the sampled village or farm during sampling. These coordinates were plotted on the base map hosted by the South African National Biodiversity Institute (http://bgis.sanbi.org/vegmap/map. asp; Figure 4.1). Data on outbreaks of heartwater in goats reported by state veterinary extension officers to the South African National Department of Agriculture between 2010 and 2014 were also recorded and are illustrated in Figure 4.1. Further, the sampling regions were characterized in terms of the endeminicity (endemic vs. non-endemic) status of the province which was determined based on the distribution of tick vector *Ambylomma hebraeum* as described in Walker and Olwage (1987) and on heartwater distribution as described by (Turton 1999; Figure 4.1).

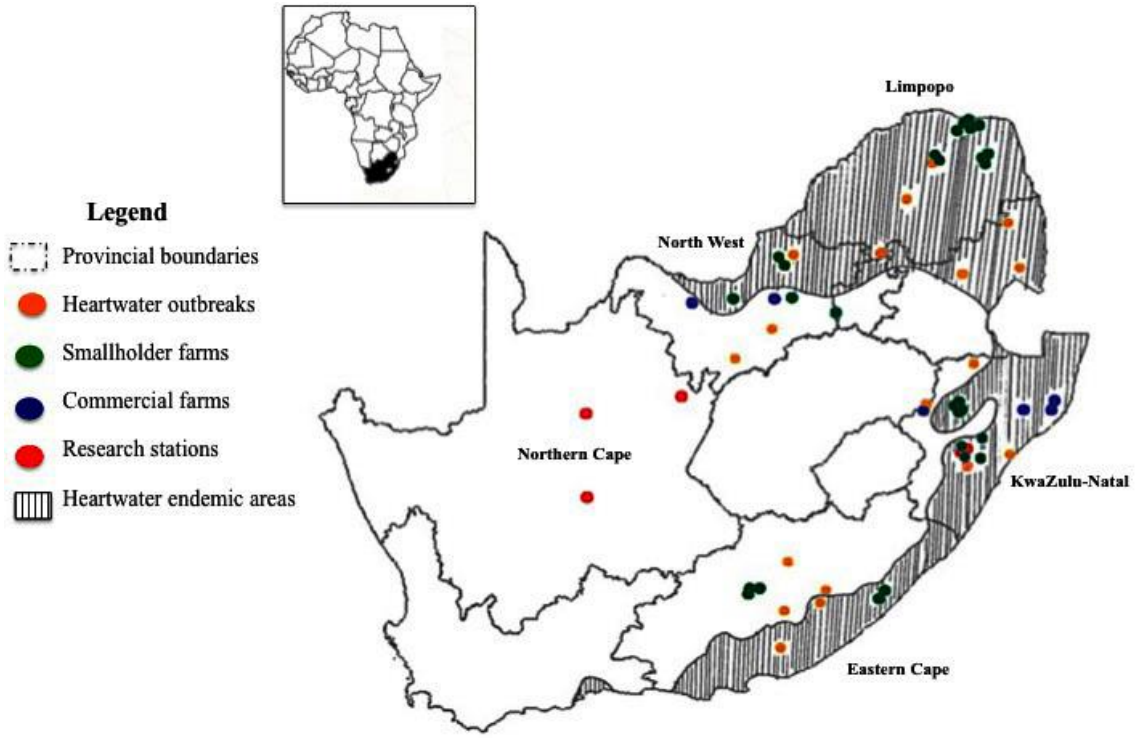


Figure 4.1 Sampling regions of South Africa. The heartwater outbreaks reported for goats between 2010 and 2013 are shown with red dots. Shaded areas illustrate heartwater endemic regions based on vector tick distribution

4.2.2 Goat populations

A total of 686 goats were randomly sampled from major goat farming provinces of the Eastern Cape, KwaZulu-Natal, Limpopo, Northern Cape and North West, and to represent the prevalent goat breeds and production systems (smallholder, commercial, research farms) in South Africa (**Table 4.1**).

Table 4.1 The distribution of goats sampled from smallholder, commercial, and research farms in the study

Province	Production system	Breed/ecotype	Number of goats
Eastern Cape	Smallholder	Boer	8
		Non-descript indigenous	56
		Kalahari Red	3
		Savanna	11
KwaZulu-Natal	Smallholder	Non-descript indigenous	142
	Commercial	Angora	9
		Boer	11
		Non-descript indigenous	56
		Kalahari Red	11
		Savanna	4
	Research	Non-descript indigenous	11
Limpopo	Smallholder	Non-descript indigenous	111
Northern Cape	Research	Boer	36
		Kalahari Red	23
		Savanna	18
		Non-descript indigenous	11
		Tankwa	50
North West	Smallholder	Non-descript indigenous	61
	Commercial	Boer	54
Total			686

4.2.3 Blood Serum Collection

Blood serum samples were collected by jugular venipuncture from each animal into 6 ml vacutainer tubes with serum clot activator (Greiner bio-one GmbH, Austria). The blood samples were kept in iceboxes until they could be refrigerated at 4 °C. The clotted blood samples were centrifuged at 3000 × g for 10 min, and about 2 ml of the serum was collected into cryotubes and stored at -20 °C until further analysis.

4.2.4 Immunoassay

Blood serum samples were submitted to the Parasite, Vectors, and Vector-borne Diseases diagnostic laboratory of the Agricultural Research Council, Ondesterpoort Veterinary Institute for detection of antibodies for *E. ruminantium*. The number of goats serologically positive for *E. ruminantium*, as an indication of heartwater immune status was determined by subjecting the sera, diluted to 1:80, to Indirect Fluorescent Antibody Test (IFAT). Each serum sample was tested in duplicate. Duplicate positive and negative controls were obtained from experimentally infected (Caprine 101) and uninfected sera available in the diagnostic laboratory, respectively. Titers greater than 1:40 were considered positive and acceptable if variation between duplicate samples was less than 10%.

4.2.5 Risk factors

Goats were classified as either *E. ruminantium* positive or *E. ruminantium* negative based on the IFAT. For each of the selected goat, the breed (Angora, Boer, Non-descript indigenous, Kalahari Red, Savanna and Tankwa), the geographic location (Eastern Cape, KwaZulu-Natal, Limpopo, Northern Cape, and North West provinces), endemic status of the region, and the goat production system (smallholder, commercial and research) were recorded. The age of each goat was given as an estimate using the presence and number of permanent incisors according to Sisson and Grossman (1953). Age was then categorised into three groups with Group 1 aged between zero to 12 months; Group 2 aged between one and three years; and Group 3 were adult goats more than 3 years old. Each goat was observed for the absence or presence of ticks.

4.2.6 Statistical analysis

The prevalence of ticks and seroprevalence of *E. ruminantium* was determined as the frequency of goats that were positive for the IFAT using the PROC FREQ of Statistical Analysis System (SAS institute Inc. 2013). The geographic location, endemic status of the region, the goat production system, sex, age group, breed and tick infestation, were treated as explanatory variables for seropositivity to *E. ruminantium*. To determine the risk factors associated with the seropositivity, a Chi-square test was used to select explanatory variables

that were significantly associated with *E. ruminantium* seropositivity. PROC LOGISTIC regression analysis (SAS institute Inc. 2013) was used to estimate the odds ratio (OR) and 95% confidence interval (95% CI) to determine the direction of each risk factor.

4.3 Results

4.3.1 Tick infestation

Tick infestation was recorded across all provinces, breeds, sex of the animal, and production system (Figure 4.2). Overall, 360 (53.35%) animals were found to have ticks. The distribution of tick prevalence across provinces can be divided into three: high in Limpopo (95.50%), intermediate in Eastern Cape (52.60%), KwaZulu-Natal (57.38%) and North West (43.48%) and low in Northern Cape (21.01%) prevalence zones (Figure 4.2a). Tick prevalence among the breeds was highly variable ranging from 69.42% (n = 311) for non-descript indigenous goats to 11.11% (n = 1) for Angora goats (Figure 4.2b). More does (n = 281; 67.46%) had ticks than bucks (n = 85; 47.49%; Figure 4.2c). Highest tick prevalence was observed in the smallholder (n = 392; 77.04%), followed commercial (n = 35; 24.14%) and lowest in the research (n = 29; 19.46%) production systems (Figure 4.2d).

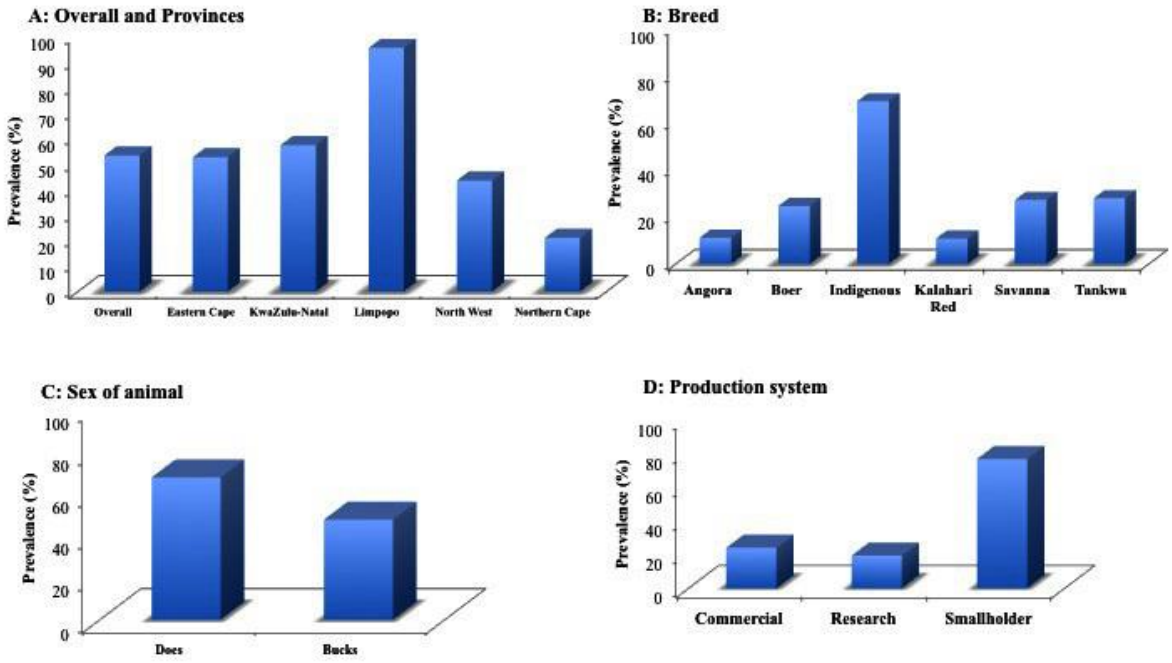


Figure 4.2 Tick prevalence (presence) across provinces, breeds, sex of the animal, and production system

4.3.2 Seroprevalence of E. ruminantium antibodies and risk factors

Of the 686 serum samples tested, 64.87% were positive for antibodies to *E. ruminantium* using the IFAT. Eastern Cape had a seroprevalence of 85.90% whilst KwaZulu-Natal was 73.77%. Highest seroprevalence was observed in the Limpopo province (91.89%) whilst the North West and Northern Cape provinces had the lowest seroprevalence of 59.13%, and 20.29%, respectively. There were significantly higher numbers of seropositive goats in endemic regions (76.09%) compared to those from non-endemic regions (20.29%). Animals from the smallholder farmers were more seropositive than those from other production systems. Does had a significantly (p<0.0169) higher seroprevalence of 67.46% than bucks with a seroprevalence of 57.54%. Seroprevalence was almost similar for young and growing goats (Group 1 and 2), and was high (69.62%) in adult goats (Group 3). Non-descript indigenous goats had the highest seroprevalence (85.04%) across provinces. Of the major commercial meat breeds, Savanna had 22 (66.67%) followed by the Boer with 34 (31.19%) and the Kalahari Red with 8 (21.62%) seropositive animals. Angora and Tankwa goats all

tested negative for antibodies to *E. ruminantium*. The seroprevalence was higher in goats with ticks (85.79%) than those without ticks (40.94%; **Table 4.2**). All seven variables were associated with *E. ruminantium* serological status at the Chi-square cut-off *p*-value of < 0.05 (**Table 4.2**).

Table 4.2 Prevalence of *E. ruminantium* antibodies in goats (*n* = 686; number (%))

Variables	Category	Number	Seroprevalence (%)	Chi-square	
				Chi	*p*
Province	Eastern Cape	78	67 (85.90)	181.1893	<0.0001
	KwaZulu-Natal	244	180 (73.77)		
	Limpopo	111	102 (91.89)		
	North West	115	68 (59.13)		
	Northern Cape	138	28 (20.29)		
Status of region	Endemic	548	417 (76.09)	150.6446	<0.0001
	Non-endemic	138	28 (20.3)		
Production system	Commercial	145	66 (45.52)	267.3426	<0.0001
	Research	149	28 (18.79)		
	Smallholder	392	351 (89.54)		
Sex	Does	507	342 (67.46)	5.7053	0.0169
	Bucks	179	103 (57.54)		
Age group	Group 1	75	41 (54.67)	15.2037	0.0005
	Group 2	137	74 (54.01)		
	Group 3	474	330 (69.62)		
Breed	Angora	9	0 (0)	237.6202	<0.0001
	Boer	109	34 (31.19)		
	Indigenous	448	381 (85.04)		
	Kalahari Red	37	8 (21.62)		
	Savanna	33	22 (66.67)		
	Tankwa	50	0 (0)		
Tick infestation	Present	366	314 (85.79)	138.3409	<.0001
	Absent	320	131 (40.93)		

The logistic regression analysis showed that the risk of acquiring the *E. ruminantium* antibodies was 3.47 times higher (OR = 3.466; 95% CI = 1.57-7.645; p = 0.0021) for non-descript indigenous and 3.68 times (OR = 3.681; 95% CI = 1.335-10.149; p = 0.0118) in Savanna breed than in Boer (**Table 4.3**). Goats from smallholder production system were at a higher risk (2.58 times) of being seropositive compared to commercial production system (OR = 2.582; 95% CI = 1.182-5.639; p = 0.0173). Goat infested with ticks at a higher at risk (3.56 times; OR = 3.587; 95% CI = 2.105-6.112) than goats without ticks (**Table 4.3**). Non-descript indigenous goats from endemic regions had a seroprevalence of 85.13% whilst those from non-endemic regions had 81.82% (Figure 4.3).

Table 4.3 Proc logistic model including the risk factors signifcantly ($p < 0.05$) associated with *E. ruminantium* seropositivity with corresponding likelihood Estemiates, Standard Error, Chi-square, Odds Ratio and 95% Confidence interval (upper and lower limit)

Variables	Category	Likelihood Estimates	Standard Error	Chi-square	Odds Ratio	Lower limit 95%CI	Upper limit 95%CI
Province	*Eastern Cape	-	-	-	-	-	-
	KwaZulu-Natal	-0.2284	0.4755	0.6310	0.796	0.313	2.021
	Limpopo	-0.4293	0.5335	0.4210	0.651	0.229	1.852
	North West	-0.6055	0.4779	0.9562	0.546	0.214	1.393
	Northern Cape	13.3095	242.3	0.2052	>999.999	<0.001	>999.999
Management system	*Commercial	-	-	-	-	-	-
	Research	-14.2059	242.3	0.9532	<0.001	<0.001	>999.999
	Smallholder	0.9485	0.3986	0.0173	2.582	1.182	5.639
Sex	Does	0.4660	0.2650	0.0786	1.594	0.948	2.679
	*Bucks	-	-	-	-	-	-
Age group	*Group 1				-		
	Group 2	-0.1185	0.4530	0.7936	0.888	0.366	2.158
	Group 3	0.5893	0.4227	0.1633	1.803	0.787	4.128
Breed	*Boer				-		
	Angora	-12.7228	262.6	0.9614	<0.001	<0.001	>999.999
	Indigenous	1.2429	0.4037	0.0021	3.466	1.571	7.645
	Kalahari Red	-0.5188	0.5400	0.3367	0.595	0.207	1.715
	Savanna	1.3033	0.5174	0.0118	3.681	1.335	10.149
	Tankwa	-12.3995	109.7	0.9100	<0.001	<0.001	>999.999
Tick infestation	Present	1.2774	0.2719	<.0001	3.587	2.105	6.112
	*Absent	-	-	-	-	-	-

* = reference catergory

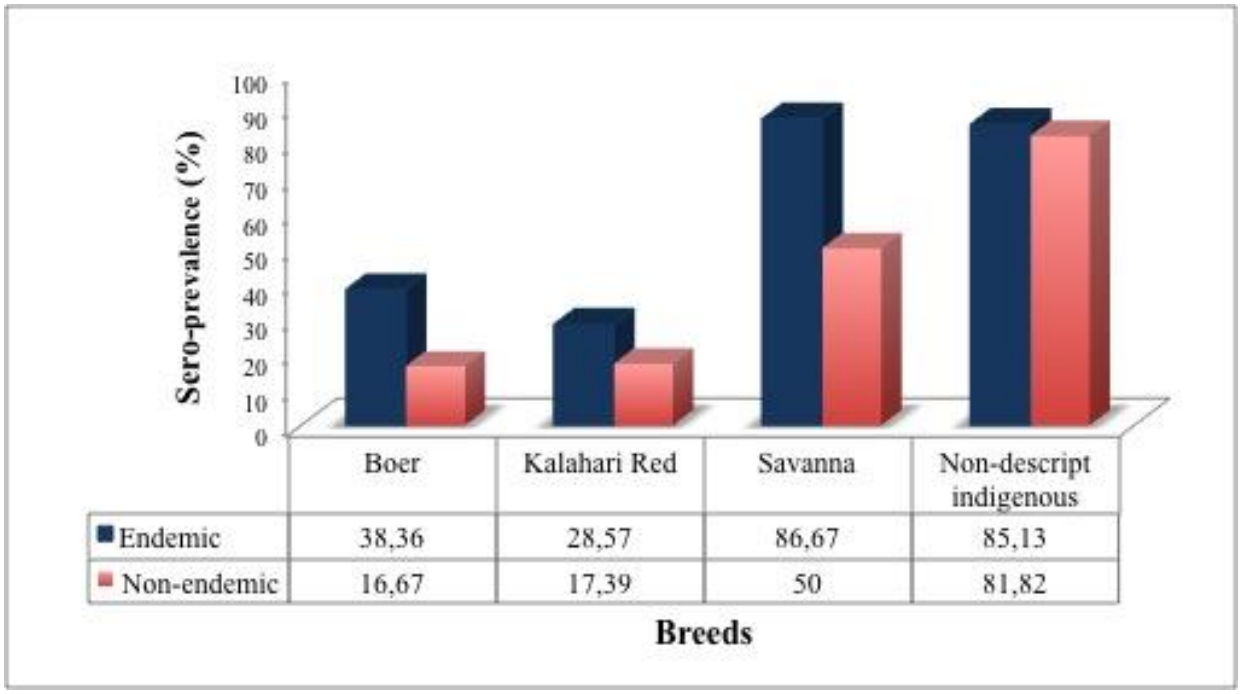


Figure 4.3 Seroprevalence profiles for goat breeds sampled from both endemic and non-endemic regions

4.4 Discussion

The study investigated the seroprevalence and associated risk factors for *E. ruminantium* in South African goats. Over fifty percent of the goats in this study had ticks attached on various body parts. The high prevalence of ticks in smallholder non-descript goats could be due to the less frequent and often absence of dipping as a control measure in this sector. Commercial farmers who supplied goat sera in the study indicated that they practice dipping regularly which was in contrast to the smallholder farmers who either did not have enough resources or depended on irregular support from the government (Chapter 4). Ticks were however also observed on some goats from the commercial farms and research stations indicating the inadequacy of existing control measures. The sampling for this study was done during the high tick burden seasons between September and February (Bryson *et al.* 2002), which might explain the high prevalence of ticks even in the commercial farms. Further analysis on

associations between tick infestation and seroprevalence were limited by the absence of data on tick species and number of ticks per animal.

The seroprevalence analysis in this study revealed high numbers of animals demonstrating antibodies to rickettsia *E. ruminantium*, which is the causative agent of heartwater disease. The findings suggest that a large proportion of the South African goat population could be continuously exposed to *E. ruminantium*. The overall seroprevalence of 64.87% was higher than the 40.1% reported in Zambia (Ahmadu *et al.* 2004; Morrison 2007), and 30.3% in Gambia (Faburay *et al.* 2005) but lower than the 83.74% reported in Zimbabwe (Kakono *et al.* 2003). Within South Africa, seroprevalence was high in goats from Limpopo (91.89%) and least in goats from the Northern Cape (28.29%). None of the goats had clinical symptoms of heartwater. The logistic regression model showed that the effect of province was not a significant risk factor for seropositivity. This is in contrast to findings by Faburay *et al.* (2005) who reported significant ($p > 0.001$) differences between geographical localities. Neither *A. hebraeum* nor heartwater disease outbreaks have been previously reported in the Northern Cape province, which is considered non- endemic to heartwater. However, Allsopp *et al.* (1997) reported the presence of non-pathogenic *E. ruminantium* variants in this region. The specific animals that tested positive for *E. ruminantium* antibodies in the Northern Cape were from the Kalahari Kid Corporation, which is a project supplying improved goat breeds to emergent farmers. Goats in the flocks of this cooperation have been sourced from other provinces in South Africa over the previous years (http://www.kalaharikid.co.za) and it is possible that the animals were acquired from heartwater endemic areas where they had already been exposed to heartwater before moving to the Kalahari Kid Corporation. The movement of animals from endemic to non-endemic regions has been reported to influence the spread of the disease between regions in Zimbabwe (Peter *et al.* 1998) and result in major losses of animals in those regions (Neitz 1967).

In South Africa, heartwater has been considered a disease restricted to the hot and dry regions of Eastern Cape, KwaZulu-Natal, and Limpopo provinces where the tick vector *A. hebraeum* prevails (Figure 4.1). Generally, the distribution of tick-borne diseases is mainly driven by environmental conditions that favor the survival of the tick vector. Tick control in the smallholder farming sector was irregular or non-existent, and ticks were present on at least some animals at each site on sampling days. Goats infested with ticks were associated with

high levels of seroprevalence than those without ticks. A similar observation was made in Cameroon (Awa 1997). Goats infested with ticks were 3.5 times higher at risk of infection than goats without ticks. However, as a limitation of our study, the positive serological results cannot be associated with specific species of the tick. Majority of goats (76.09%) sampled from endemic regions had *E. ruminantium* antibodies indicating the importance of control strategies in these regions. Mahan *et al.* (1998) reported a prevalence of *E. ruminantium* antibodies of about 90% in goats in heartwater-endemic areas of Zimbabwe. Tick infestation and endemic status were significantly associated with variation in antibody prevalence. Similarly, goats were associated with a higher risk of exposure to heartwater infection if they came from non-endemic regions compared to those from endemic regions.

Other risk factors associated with the prevalence of heartwater have been discussed in detail in literature (Bath *et al.* 2005; Swai *et al.* 2008; Swai *et al.* 2009) and the rationale for selecting risk factor variables for analysis in this study is hereby described. Differences in seroprevalence associated to age have been reported in cattle (Swai *et al.* 2008), and sheep and goats (Swai *et al.* 2009). Swai *et al.* (2008), observed a positive correlation in seropositivity with increasing age suggesting innate resistance and high level of exposure. The high proportion of seropositive goats in all age groups in this study suggests presence of maternal antibodies and high degree of infection due to field exposure and continuous tick challenge when maintained under tick infested conditions. It was not feasible to determine the actual age of goats under the age of 3 months due to the method used to estimate age, which limited the interpretations around maternal antibodies.

Swai *et al.* (2009) showed that adult goats were 2.10 (95% CI = 1.13–3.90) times higher at risk of infection than weaners, however, the logistic regression model in this study showed that age was not a significant risk factor. Although the antibody response has been reported to last for 6-12 months post infection there are possibilities of reinforcement and prolonged immunity in populations that are continuously under challenge (Zweygarth *et al.* 2008), which was the assumption of this study.

The effects of the sex of goat were investigated in the context of financial and physical capital asset contributions of male and female goats in small-scale production systems (Sebei *et al.* 2004a) where over 50% of the sampled goats are raised. The relatively high number of

seropositive does to bucks is similar to those reported in a study by Swai *et al.* (2009) in Tanzania. In most smallholder production systems, does are kept on-farm for longer periods with farmers bringing in new bucks more frequently. This potentially extends the exposure of does to ticks and predisposes them to heartwater. In addition, it is also possible that the higher seroprevalence in does is influenced by the presence of more does than bucks in the flocks.

The logistic regression results indicated that breed is a crucial risk factor for exposure to *E. ruminantium* infection. In this study, we found that the seroprevalence was higher in Savanna and non-descript indigenous goats that had a 3.7 times and 3.47 times higher risk of infection, respectively, than the Boer goats. The lack of detection of antibodies in Angora, and Tankwa goats in the study indicates that these goats had not been exposed to the infection at the time of the survey and could be regarded as susceptible populations. Tankwa goats live in a non-endemic area and are exposed to infection through movement of animals from endemic regions. Angoras are highly susceptible to heartwater (Du Plessis *et al.* 1983) and cannot thrive even in areas of low risk without veterinary care. The high seroprevalence in the non-descript indigenous goats suggested that these unimproved and naturally selected animals could be resistant to heartwater (Donkin *et al.* 1992).

Production system determines the risk status of farms to heartwater and the dynamics due to these differences need to be explored in order to establish accurate and effective control strategies (Mukhebi *et al.* 1999). Our results showed a 2.6 times higher risk in the smallholder production system in comparison to the commercial production system regardless of geographical location and endemic status. The high level of antibodies detected in the smallholder production system highlights the economic importance and the need for heartwater control measures in this production system, which is often limited due to non-existent tick control and low vaccination levels (Bester *et al.* 2009). On the other hand, the exposure of the smallholder populations to ticks and heartwater could have facilitated adaptation or tolerance of non-descript indigenous goats to infections.

The study demonstrated that *E. ruminantium* antibodies are widely distributed in South Africa. The distribution is greater in the endemic regions especially in smallholder production systems. The study provides a baseline understanding of the epidemiology of heartwater in the country and adds to the information necessary for designing effective control strategies

targeted for different production systems and populations at high risk of infection. There were limitations to the number of animals sampled from other production systems and breeds. More samples would probably have provided clearer epidemiological data about each production system and breed.

CHAPTER 5: POPULATION STRUCTURE AND ADMIXTURE OF SOUTH AFRICAN GOAT BREEDS INVESTIGATED USING GENOME-WIDE SNP DATA

Published in the Journal Animal Genetics 47, 471-482

ABSTRACT

The sustainability of goat farming in marginal areas of southern Africa depends on local breeds that are adapted to specific agro-ecological conditions. Unimproved non-descript goats are the main genetic resources used for the development of commercial meat-type breeds of South Africa. Little is known about genetic diversity and the genetics of adaptation of these indigenous goat populations. This study investigated the genetic diversity, population structure and breed relations, linkage disequilibrium, effective population size and persistence of gametic phase in goat populations of South Africa. Three locally developed meat-type breeds of the Boer (n = 33), Savanna (n = 31), Kalahari Red (n = 40), a feral breed of Tankwa (n = 25) and unimproved non-descript village ecotypes (n = 110) from four goat-producing provinces of the Eastern Cape, KwaZulu-Natal, Limpopo and North West were assessed using the Illumina Goat 50K SNP BeadChip assay. The proportion of SNPs with minor allele frequencies >0.05 ranged from 84.22% in the Tankwa to 97.58% in the Xhosa ecotype, with a mean of 0.32 ± 0.13 across populations. Principal components analysis, ADMIXTURE and pairwise F_{ST} identified Tankwa as a genetically distinct population and supported clustering of the populations according to their historical origins. Genome-wide F_{ST} identified 101 markers potentially under positive selection in the Tankwa. Average linkage disequilibrium was highest in the Tankwa (r^2 = 0.25 ± 0.26) and lowest in the village ecotypes (r^2 range = 0.09±0.12 to 0.11±0.14). We observed an effective population size of <150 for all populations 13 generations ago. The estimated correlations for all breed pairs were lower than 0.80 at marker distances >100Kb with the exception of those in Savanna and Tswana populations. This study highlights the high level of genetic diversity in South African indigenous goats as well as the utility of the genome-wide SNP marker panels in genetic studies of these populations.

Keywords: *50K genotypes, effective population size, gametic phase, genetic diversity, production system*

5.1 Introduction

Over 63% of the goats in South Africa are so-called indigenous goats. Indigenous goats and other similarly raised livestock species have generally been described as 'ecotypes', inferring that they have evolved, adapted and become part of their specific agro-ecological conditions (Morrison 2007). Three goat ecotypes of the Nguni type (iMbuzi), Eastern Cape (Xhosa lob ears) and Northern Cape (Speckled) have been described according to these eco-regions (Morrison 2007). South Africa is one of the few countries that has successfully developed goat breeds from local populations. The Boer and Savanna were developed from indigenous 'unimproved' goat populations in the Eastern Cape and Northern Cape provinces respectively (Campbell 2003); however, studies suggest that they are genetically closely related (Visser *et al.* 2004; Pieters 2007). On the other hand, the Kalahari Red has a more complex history and is believed to have two bloodlines, one originating from the redheaded Boer goat and the other from brown lop-eared 'unimproved' local goats in South Africa and Namibia (Campbell 2003). Genetic improvement programs have been effective in selecting for the commercial meat-type breeds. However, there is enormous potential for incorporating genomic technologies to improve selection accuracy (Mohlatlole *et al.* 2015).

The genetic contribution of local populations has not been fully exploited regardless of their assumed unique genetic merits and potential role in smallholder and commercial agriculture (Rege 1994). It is hypothesized that the change in production systems toward small- and large-scale commercial agriculture has created a bias toward high-performance breeds (Rege & Gibson 2003) in South Africa and most developing countries. Also, indiscriminate crossbreeding may lead to ecotype populations being genetically eroded by commercial meat-type or exotic breeds, leading to a loss of their unique genetic features (Iniguez 2005).

Unique to South Africa is the Tankwa, a feral goat breed that originated in the Karoo region (Figure 4.1). Tankwa goats resided in the Tankwa Karoo National Park for approximately 50 years and were recently relocated to the Carnarvon Nature Reserve for conservation. The Tankwa has flourished without human management and has adapted to the arid environment that characterizes the Tankwa Karoo National Park and Northern Cape province. The feral population has recently been described using a set of eight microsatellite markers as genetically unique from the rest of the farmed breeds in South Africa (Kotze *et al.* 2014).

There is minimal knowledge about the genetic characteristics of South African goat breeds or ecotypes, or the relationship and divergence between them. Limited studies on South African indigenous goat populations include the use of low-density microsatellite markers and small sample numbers of commercial breeds and/or experimental populations with a bias toward certain production systems or eco-regions (Kotze *et al.* 2004; Visser *et al.* 2004; Pieters 2007; Kotze *et al.* 2014). The usefulness of SNPs in analyzing genetic diversity, population structure and admixture history has been demonstrated in several livestock species (Kijas *et al.* 2012; McCue *et al.* 2012; Ai *et al.* 2013; Mbole-Kariuki *et al.* 2014). The high-density Illumina GoatSNP50K (>50 000 probes) genotyping array (Illumina, Inc.) has been available since 2012, and its utility was recently described (Tosser-Klopp *et al.* 2014). This SNP chip has found application in studying genetic diversity and polledness in multiple breeds (Kijas *et al.* 2013), QTL detection in dairy goats (Maroteau 2013; Palhière *et al.* 2014) and genomic selection in French dairy goats (Carillier *et al.* 2013; Carillier *et al.* 2014).

Here, we report the first analysis of the genomic architecture of South African indigenous goat populations with a focus on intra- and interbreed characteristics and relations. The aim of the study was to determine (i) SNP marker polymorphism and minor allele frequency (MAF) distribution, (ii) genetic diversity and population structure, (iii) signatures of selection, and (iv) the extent of linkage disequilibrium as well as trends in effective population size of the South African Boer, Kalahari Red, Savanna, Tankwa and ecotypes Nguni, Venda, Xhosa, Zulu and Tswana goat populations using the Illumina GoatSNP50K genotyping array. The consistency of gametic phase was also explored to determine the possible use of a multibreed reference population for genomic selection. Information generated in this study is considered a prerequisite for designing the conservation and genetic improvement programs.

5.2 Materials and methods

5.2.1 Breed and ecotype population description

A total of 239 blood samples were collected from goats from locally developed meat-type breeds of the Boer (n = 33), Kalahari Red (n = 40) and Savanna (n = 31); local ecotype populations from smallholder farmers (n = 100); and commercial and research farms in

Kwazulu-Natal (n = 10). Blood samples were also collected from a feral Tankwa (n = 25) goat population from the Carnarvon, Northern Cape (Figure 5.1). The locally developed meat-type breeds were collected from various stud breeders and commercial farms around the country. The non-descript local ecotype populations from smallholder farmers were sampled from villages in the Eastern Cape (n = 20), KwaZulu-Natal (n = 30), Limpopo (n = 30) and North West (n = 20) provinces (Figure 5.1). Samples collected from the villages were named according to the local language of the villagers; thus, goats from Eastern Cape were Xhosa, whereas those from KwaZulu-Natal, Limpopo and North West were named Zulu, Venda and Tswana respectively. Non-descript local ecotype populations from the commercial and research farms in KwaZulu-Natal were referred to as Nguni goats according to (Morrison 2007). The goats used in this study represent the major goat breeds and populations reared in South Africa; pictures of the different goat breeds/populations are presented in Figure 5.1. Blood samples were collected by jugular venipuncture from each animal into 6-ml EDTA vacutainer tubes (Greiner Bio-One, GmbH). The collected blood samples were kept in iceboxes until refrigerated at 4°C. Genomic DNA was extracted using the DNeasy" Blood and Tissue Kit (Qiagen), as per the manufacturer's instructions with a slight modification of increased lysis time to 90 min. DNA quality and quantity were determined using 1% agarose gel electrophoresis (Merck) and Qubit" 3.0 Fluorometer (Life Technologies) respectively.

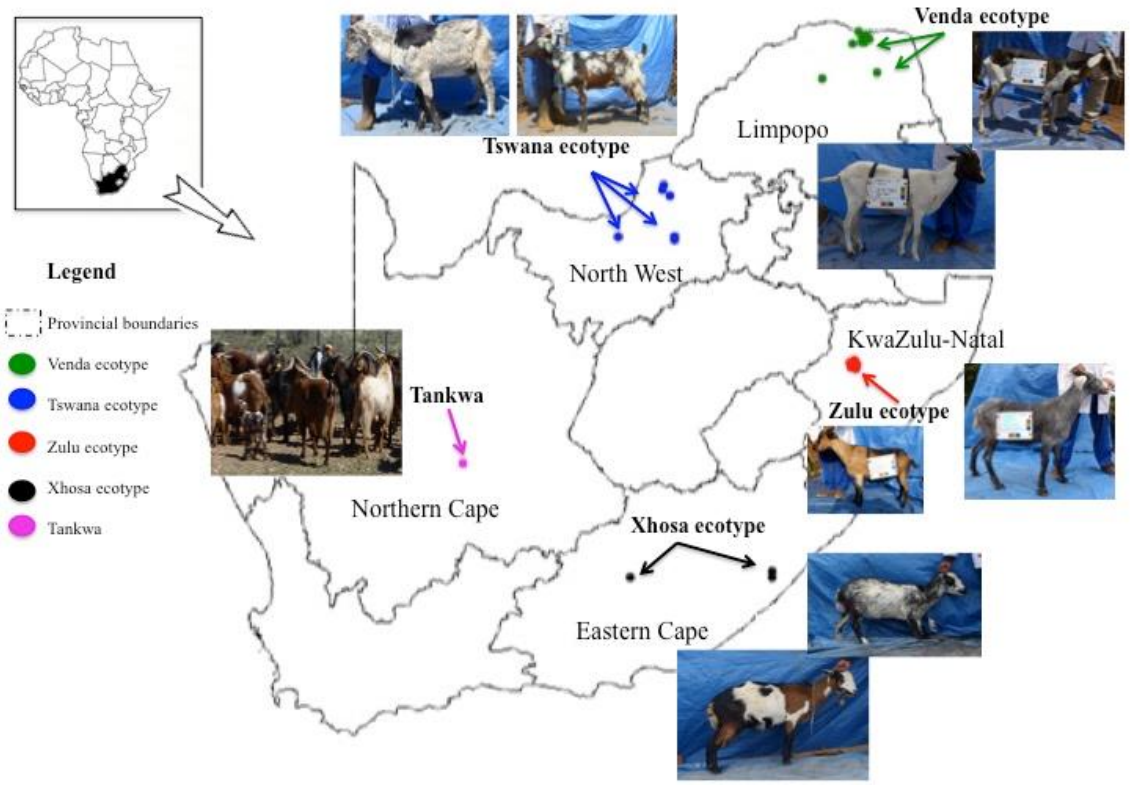


Figure 5.1 Geographic location of non-descript village and the Tankwa goat populations

5.2.2 SNP genotyping and quality control

The 239 goats were genotyped for over 50 000 SNPs on the Illumina Goat 50K SNP BeadChip using the Infinium assay that is compatible with the Illumina HiScan SQ genotyping platform at the Agricultural Research Council-Biotechnology Platform in South Africa. SNP genotypes were called using genotyping module integrated in GenomeStudio™ V2010.1 (Illumina, Inc.).

The SNP marker map file was updated using Golden Helix SNP and Variation Suite (SVS) V8.1 (Golden Helix, Inc. 2012), and information on the chromosome number and chromosomal position was extracted from the CHI_1.0 goat genome assembly (Accessed July 2014) updated in 2014 and available from the International Goat Genome Consortium. Of the 51 829 SNPs on the SNP50K genotyping array, non-autosomal SNPs and those on contigs or

not mapped to the latest reference assembly of the goat genome were excluded, resulting in a set of 49 942 SNPs available for further analyses. PLINK (Purcell *et al.* 2007) was used to prune individuals that failed to pass the inclusion criterion of 95% successful genotypes per individual, resulting in 216 individuals being removed. The 49 942 SNP set was further pruned using different parameters and thresholds that were relevant for the different analysis performed. For within-population genetic diversity, population sub structuring, linkage disequilibrium and pairwise F_{ST} analysis, a series of filters were employed to remove (i) low-quality markers, by removing SNPs with call rate ≤ 0.95%; uninformative markers that had a MAF ≤ 0.05; and (iii) those markers that significantly deviated from Hardy-Weinberg equilibrium (HWE; P < 0.001). Quality control was performed per subpopulation and across populations (**Table 5.1**). Across populations, 44 660 SNPs (87.09%) were available for downstream analysis. Further, in PLINK (Purcell *et al.* 2007), the '-indep-pairwise 50 5 0.2-' command was used to remove one of every pair of SNPs with $r^2 > 0.2$ within 50-SNP sliding windows. This removed 21 426 SNPs. Additionally, an identity-by-distance matrix was computed using the same program, and 11 highly related (identity by distance > 0.45) individuals were removed. A higher number of highly related Tankwa goats (n = 5) were removed, followed by Boer (n = 2), Savanna (n = 2) and Tswana and Kalahari Red with only one individual removed from each breed. The resulting data set of 23 234 SNPs and 205 individuals was used for the population structure analysis.

5.2.3 Minor allele frequencies

Minor allele frequencies and the proportion of polymorphic SNPs were estimated using PLINK (Purcell *et al.* 2007) on 49 942 SNPs under default settings for the remaining individuals. The mean MAF and standard deviation per population were calculated using the PROC MEANS procedure of the Statistical Analysis System (SAS institute Inc. 2013). SNPs were categorized as either fixed (MAF = 0), rare (0 < MAF < 0.01), intermediate (0.01 < MAF < 0.05), or common (0.10 < MAF < 0.5). All SNPs with MAF > 0.05 were considered polymorphic.

5.2.4 Within-population genetic diversity and substructuring

The subset of SNPs and individuals per population and across populations was used to determine allelic richness (*AR*), heterozygosity (H_O), expected heterozygosity (H_E) and inbreeding estimates (F_{IS}), as indicated in **Table 5.1**. The AR was estimated using ADZE 1.0 (Szpiech *et al.* 2008), and H_O, H_E and F_{IS} were estimated using PLINK (Purcell *et al.* 2007). Analysis of molecular variance was performed using the ARLEQUIN (Excoffier *et al.* 2005). The level of population substructure was investigated for (i) the overall population that included all nine populations, (RDA) the ecotype populations and commercial breeds, (iii) the ecotypes (Tswana, Xhosa, Venda, Zulu and the Nguni goats) and (iv) the commercial breeds consisting of the Boer, Kalahari Red and Savanna.

5.2.5 Population structure analysis

For population structure analysis further exclusion of related individuals (IBD>0.45) and SNPs with $r^2 > 0.2$, resulting in a subset of 205 individuals and 23 234 SNPs. A principal components analysis (PCA) was used to assess the population structure using Golden Helix SNP and Variation Suite (SVS) V8.1 (Golden Helix, Inc. 2012). In addition, ADMIXTURE 1.21 was used to infer the most probable number of ancestral populations based on the SNP genotype data (Alexander *et al.* 2009). ADMIXTURE was run from $K = 2$ to $K = 10$, and the optimal number of clusters (*K*-value) was determined as that having the lowest cross-validation error.

5.2.6 Genetic differentiation and detection of signatures of selection

Pairwise F_{ST} was computed to compare diversity between populations (Weir & Cockerham 1984) implemented in Golden Helix SNP and Variation Suite (SVS) V8.1 (Golden Helix, Inc. 2012). To detect the genome-wide pattern of positive selection, we used the outlier loci approach based on the calculation of fixation index (F_{ST}) at different significance levels as a measure of genetic differentiation for each locus between Tankwa and all domesticated breeds using Golden Helix SNP and Variation Suite (SVS) V8.1 (Golden Helix, Inc. 2012). The loci

under selection would be expected to show an allele frequency that deviates from that of neutral loci with an F_{ST} threshold of 0.5.

5.2.7 Mapping selection signatures to genes

Significant SNPs (F_{ST} > 0.5) were annotated to genes using the National Centre for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov) genome browser using the *Capra hircus* genome assembly CHIR v1.0 (Accessed November 2014).

5.2.8 Genome-wide linkage disequilibrium

PLINK 1.07 (Purcell *et al.* 2007) was used to estimate the pairwise r^2 values per chromosome and population (Hill & Robertson 1968). The r^2 value was then estimated using parameters defined by the PLINK option '--r2 --ld-window-kb 2000 --ld-window-r2 0' to calculate linkage disequilibrium (LD) association among SNP pairs up to a distance of 2000Kb. Mean LD and standard deviation were calculated using the PROC MEANS procedure in Statistical Analysis System (SAS institute Inc. 2013). The generalized linear model procedure implemented in SAS was used to determine the effects of breed, chromosome, distance between SNP markers and the interaction between breed and chromosome on r^2 values using the statistical model:

$$Y_{ijk} = \mu + A_i + B_j + (A \times B)_{ij} + bC_k + E_{ijk} \quad (5.1)$$

where Y_{ijk} is the observed r^2; μ is the overall mean; A_i is the fixed effect of breed (i represents Boer, Kalahari Red, Savanna, Tankwa, Nguni, Zulu, Tswana, Venda or Xhosa); B_j is the effect of chromosome (j is chromosome 1–29); $(A \times B)_{ij}$ is the interaction effect of breed with chromosome; C_k is distance between SNP markers (Kb), which was treated as a covariate with a regression coefficient b; and E_{ijk} is the random residual error effect.

Linkage disequilibrium decay was analyzed by first categorizing SNP marker pairs into intermarker distance bins of 0-1, 1-10, 10-20, 20-40, 40-60, 60-100, 100-200, 200-500, 500-1000 and 1000-2000Kb followed by an analysis of mean LD within each bin.

5.2.9 Effective population size

Effective population size (N_e) was estimated based on the known relationship between r^2, N_e and the recombination rate (c) between two loci following eqn 1 of Sved (1971) implemented in SNeP (Barbato *et al.* 2015). We corrected for sample size for each population using the second equation of Sved (1971). The different SNP marker distance bins used for r^2 analysis were used to obtain different estimates of N_e at different time points by calculating the number of generations (t) in the past as 1/2c.

5.2.10 Consistency of gametic phase

Consistency of gametic phase, measured as the correlation between signed *r*-values (gametic phase), was assessed between all pairs of breeds (n = 36) by examining the correlation of LD phase between all breed pairs. Of the SNPs used in LD analysis, a subset of 36 484 SNPs that were common across breeds was used for gametic phase analysis. For each marker pair, the gametic phase was estimated as signed r value using the '*--r2 --ld-window-kb 2000 --r*' command in PLINK (Purcell *et al.* 2007). Gametic phase values were grouped into bins based on intermarker distances of 0-1, 1-10, 10-20, 20-40, 40-60, 60-100, 100-200, 200-500, 500-1000 and 1000-2000Kb. A correlation analysis of the signed r between breed pairs was estimated for each intermarker distance bin using PROC CORR in Statistical Analysis System (SAS institute Inc. 2013). Consistency of gametic phase between breeds was determined by plotting the correlation coefficient of the signed *r* against intermarker distance for each breed pair.

5.3 Results

5.3.1 Quality control statistics

Twenty-three of the 239 goats across breeds had a per-sample call rate below 0.95 and were excluded from analysis (**Table 5.1**). Five goats each from among the Tankwa, Venda and Zulu were excluded, and four were pruned from the Kalahari Red. Overall, 216 individuals

remained after quality control. SNP marker quality control excluded 2951 SNPs across populations. Of these SNPs, 836, 1546 and 961 failed for call rate (<0.05), MAF (<0.05) and deviation from HWE ($P < 0.001$) respectively. The Tankwa had the highest number of SNPs ($n = 8062$) excluded based on low MAF compared with the Xhosa ecotype, which had the least ($n = 1254$) excluded. A common subset of 44 660 SNPs (87.09%) from 216 individuals was used for downstream analysis.

Table 5.1 Number of individuals and SNPs observed per population and across populations after quality control

Population	Number of individuals	Quality control exclusions				Number of SNPs validated for downstream analysis (%)
		Per sample call rate ≤ 0.95	Per marker			
			Call rate ≤ 0.95	MAF ≥ 0.05	HWE >0.001	
Boer	33	2	802	4029	125	45176 (90.46)
Kalahari Red	40	4	1061	2877	266	46027 (92.16)
Savanna	31	2	1111	2308	114	46597 (93.30)
Nguni	10	0	751	2158	0	47276 (94.66)
Tankwa	25	5	733	8062	56	41424 (82.94)
Tswana	20	0	560	1858	65	47704 (95.51)
Venda	30	5	956	2655	71	46563 (93.23)
Xhosa	20	0	748	1254	56	48059 (96.23)
Zulu	30	5	748	2517	85	46893 (93.89)
All	239	23	836	1546	3362	44660 (87.09)

5.3.2 Minor allele frequency

On average, 96.95% of the markers were polymorphic (MAF > 0.05) within the populations (**Table 5.2**). The Zulu ecotype showed the highest proportion of polymorphic SNPs (97.58%), whereas the Tankwa had the lowest proportion of polymorphic SNPs (84.22%).

Table 5.2 Sample size, percentage of polymorphic markers and within population diversity indicators calculated for each breed and whole population

Population	N	% SNPs >0.5	MAF ±SD	*HO*±SD	*HE*±SD	*AR*±SD	F_{IS}
Boer	31	91.98	0.27±0.14	0.36±0.15	0.37±0.14	1.19±0.00	0.12±0.06
Kalahari Red	36	94.29	0.29 ±0.14	0.37±0.14	0.38±0.13	1.82±0.00	0.10±0.08
Savanna	29	95.43	0.30±0.14	0.39±0.15	0.38±0.12	1.93±0.00	0.06±0.07
Tankwa	20	84.22	0.24±0.16	0.35±0.33	0.33±0.16	1.79±0.00	0.15±0.05
Nguni	10	96.06	0.29±0.14	0.41±0.18	0.39±0.12	1.83±0.00	0.01±0.05
Tswana	20	96.64	0.31±0.13	0.40±0.144	0.41±0.10	1.93±0.00	0.03±0.10
Venda	25	95.04	0.29 ±0.13	0.40±0.140	0.40±0.11	1.94±0.00	0.04±0.04
Xhosa	20	95.32	0.31±0.13	0.42±0.14	0.41±0.10	1.92±0.01	0.02±0.05
Zulu	25	97.58	0.30 ±0.13	0.40±0.140	0.40±0.11	1.94±0.01	0.04±0.07
All	216	96.95	0.32 ±0.13	0.38±0.10	0.41±0.10	1.91±0.00	0.07±0.08

N, Number of individuals; MAF, minor allele frequency; AR: allelic richness; HO,observed heterozygosity; HE, expected heterozygosity; F_{IS}, inbreeding coefficient.

5.3.3 Within-population genetic diversity and substructuring

Within-population genetic variability estimates of the different goat populations are shown in **Table 5.2**. The Zulu and Venda ecotypes had the highest number of alleles per locus (*AR* = 1.94), whereas the Boer had the lowest number of alleles per locus (*AR* = 1.19). MAF ranged from 0.24 ± 0.16 in the Tankwa to 0.31 ± 0.13 in the Tswana and Xhosa ecotypes. Observed and expected heterozygosities ranged from the lowest (H_O = 0.35 ± 0.33; H_E = 0.33 ± 0.16) in the Tankwa to the highest (H_O = 0.42 ± 0.14; H_E = 0.41 ± 0.10) in the Xhosa ecotype. The Tankwa had an inbreeding coefficient (F_{IS}) of 0.15 ± 0.05, which was approximately 7.5 times greater than that for the Xhosa ecotype (-0.02 ± 0.05). Analysis of molecular variance indicated high levels of within-individual animal diversity relative to among individuals within populations and among populations variation across all population categories (**Table 5.3**).

Table 5.3 AMOVA analysis results using different data sets

Population group	Among populations	Among individuals within populations	Within individuals
All nine goat populations	6.39	1.92	91.69
Ecotype populations and commercial breeds	4.69	2.33	92.98
Ecotype populations	2.03	2.08	95.89
Commercial breeds	0.94	2.92	96.14

5.3.4 Population structure analysis

5.3.4.1 PCA-based clustering

The first principal component (PC1) explained 5.87% of the variation and separated the commercial breeds and ecotype populations. PC2 explained 3.16% of the variation and clearly separated the Tankwa from the other goat populations (Figure 5.2). The lower principal components resolved the divergence between (i) the Kalahari Red breed (PC3) and the ecotype populations (PC4 and PC5). PC3 indicated divergence between the Kalahari Red goats that were established from a population of indigenous goats from those derived from Boer goats. PC4 and PC5 separated the Zulu and Venda ecotypes from the Tswana and Xhosa ecotype populations. The Nguni ecotype, representing a population of unimproved goats kept by organized goat breeders, were concordantly located between village ecotypes. The Xhosa and Tswana goats clustered closer to the commercial breeds relative to the Zulu and Venda ecotype populations.

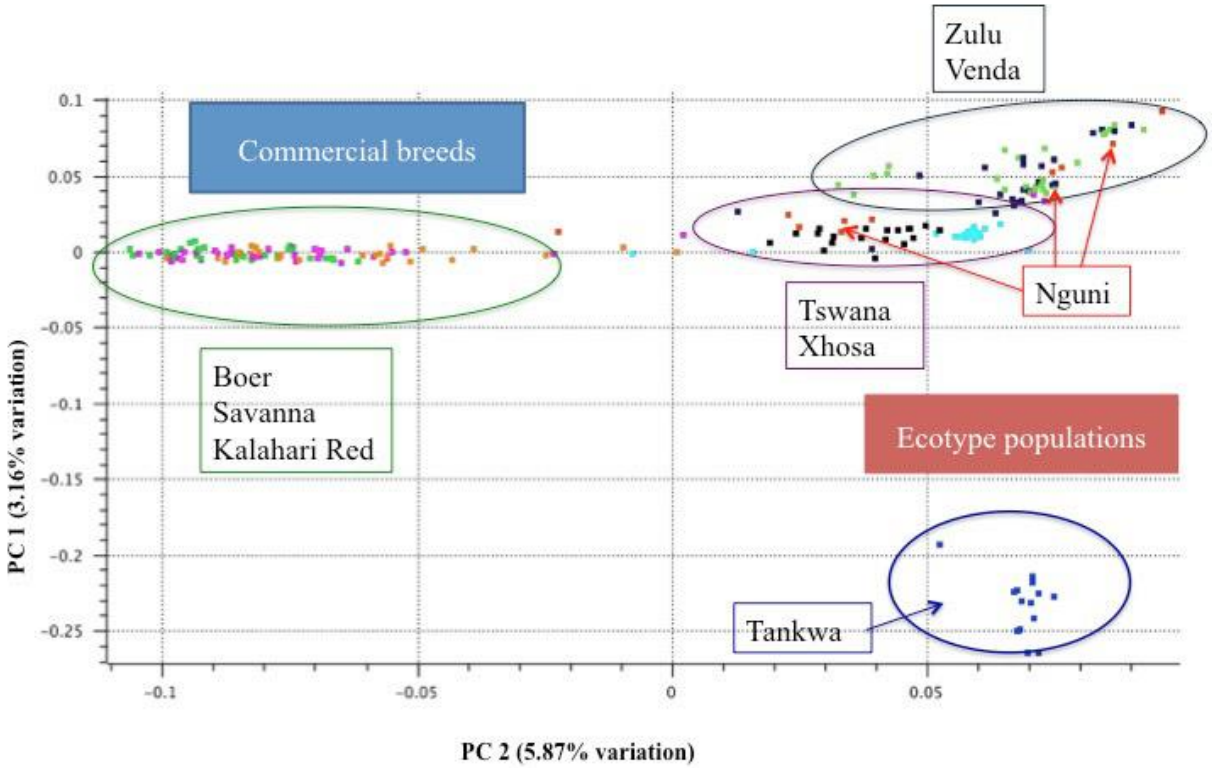


Figure 5.2 Principal components analyses plot for South African indigenous goats. PC1 and PC2 explain 5.87% and 3.16% of the total variance, respectively

5.3.4.2 Population clustering based on ADMIXTURE

ADMIXTURE-based clustering is presented in Figure 5.3. *K* = 2 clustered the commercial breeds (Boer, Kalahari Red and Savanna) separately from the Tankwa and the village goat ecotypes. The three clusters at *K* = 3 corresponded to the Tankwa, the commercial breeds and the village goats. The two Kalahari Red bloodlines separated at *K* = 4 (KR and KRI). *K* = 5 was the optimal population cluster based on cross-validation. Increasing K to 6 revealed a finer resolution of the commercial breeds where a cluster of a few Savanna goats emerged. *K* = 7 to *K* = 10 revealed the distinction between Xhosa and Tswana from Zulu, Venda and Nguni ecotype populations, as seen in PC4 and PC5. High genetic diversity and breed interrelatedness were observed in the clusters of the commercial breeds and the ecotype populations, which was contrary to that of the Tankwa from *K* = 3 to *K* = 10.

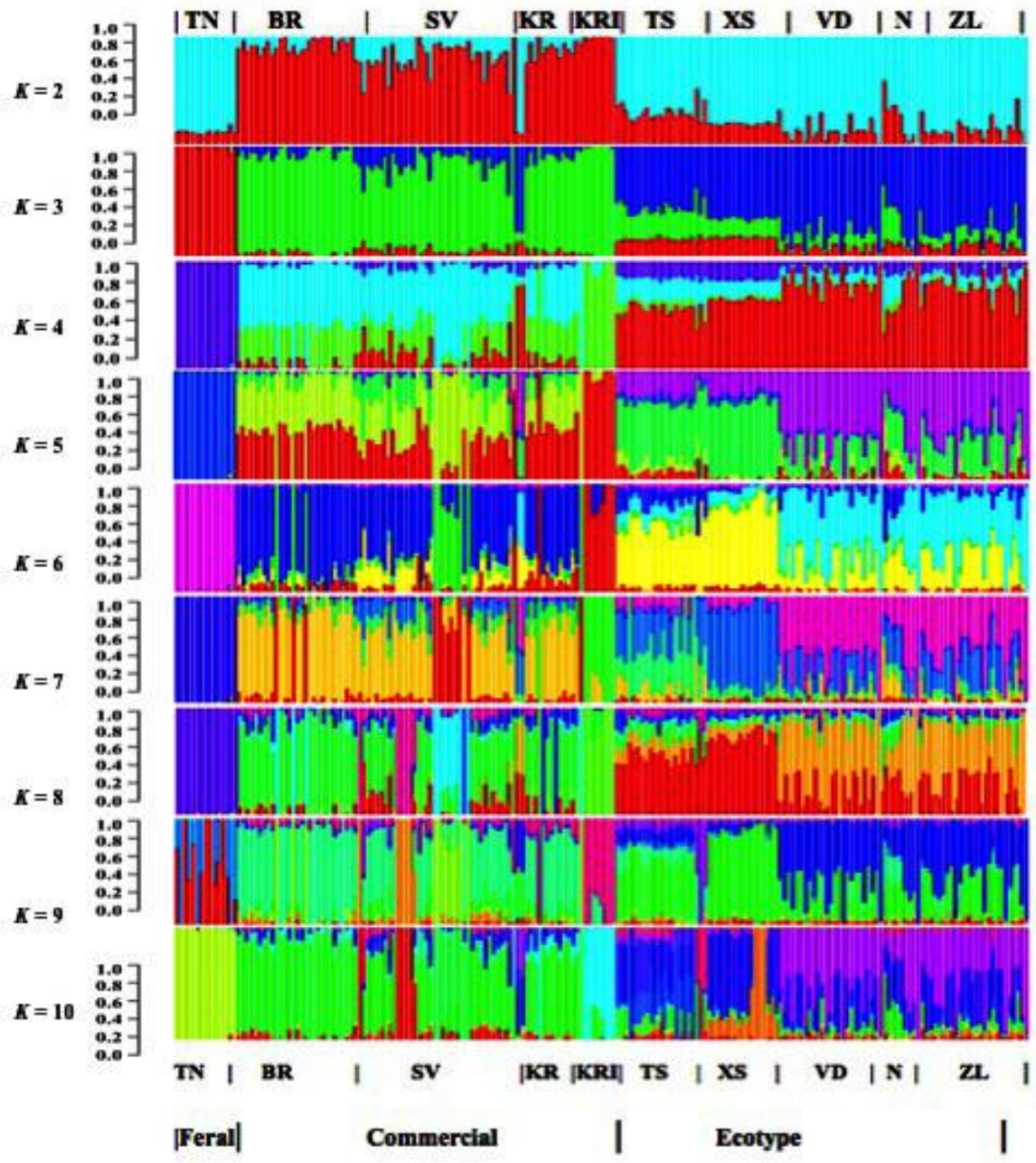


Figure 5.3 ADMIXTURE-based clustering of the South African goats breeds from *K* = 2 to *K* = 10. TN, Tankwa; BR, Boer; SV, Savanna; KR, Kalahari Red; KRI, Kalahari Red from Northern Cape (indigenous bloodline) ; TS, Tswana; XS, Xhosa; VD, Venda; N, Nguni (iMbuzi); ZL, Zulu

5.3.5 Genetic differentiation, outlier loci and candidate genes under selection

Pairwise population comparison using F_{ST} (Figure 5.4) indicated a clear distinction between Tankwa and all the domestic goats (F_{ST} = 0.108-0.163) and close relationships among the commercial breeds as well as among the ecotype populations. The level genetic differentiation (range F_{ST} = 0.051-0.092) was observed between commercial breeds and the ecotype populations, and among the commercial breeds and among the ecotype populations respectively.

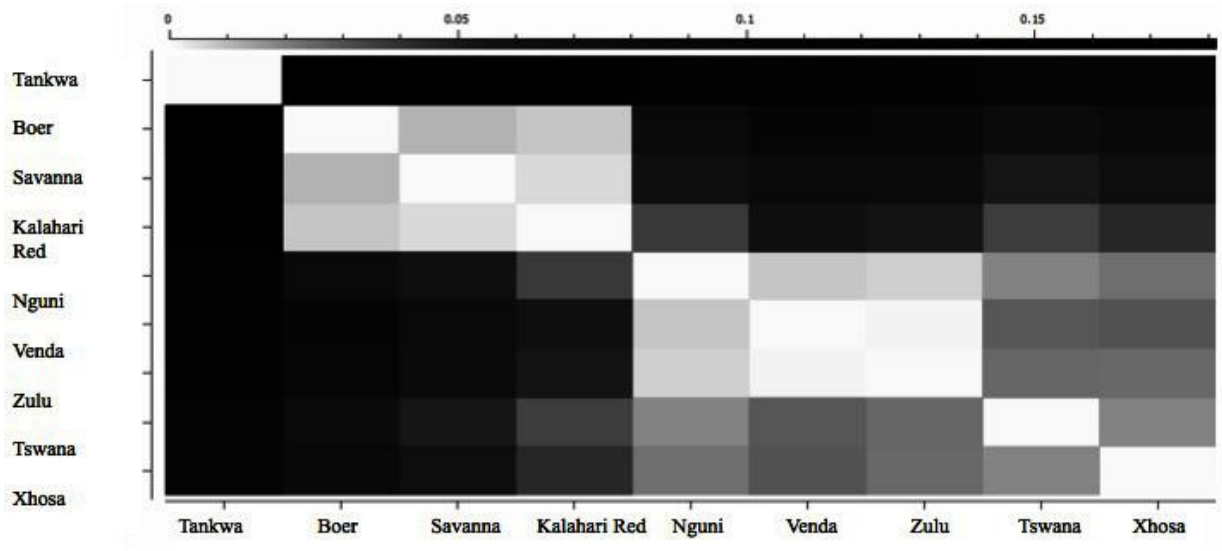


Figure 5.4 Heat map of pairwise F_{ST} for South African goat population

The mean genomic F_{ST} value across all SNPs was equal to 0.089 and values ranging from 0 to 0.81) according to Wright (1978) classification. The distribution of F_{ST} across all autosomes is shown in Figure 5.5. A total of 101 differentiated SNPs with $F_{ST} \geq 0.6$ were considered to be under selection (Figure 5.5). One SNP had an F_{ST} value of 0.81 located on chromosome 24, and 21 had F_{ST} values > 0.70. These SNPs were located on chromosomes 1, 2, 3, 7, 8, 9, 12, 13, 14, 15, 17, 18, 24, and 25. Seventy-nine SNPs had an F_{ST} between 0.60 and 0.69. Significantly selected SNPs were observed to lie within 38 genes (APPENDIX B). The loci with the greatest F_{ST} values (> 0.7) were located within the *FAM71E1*, *ADCY9*, *USP37*, *SFT2D2* genes, and two loci located on chromosome 7 were located within the *CHD1* gene (APPENDIX B).

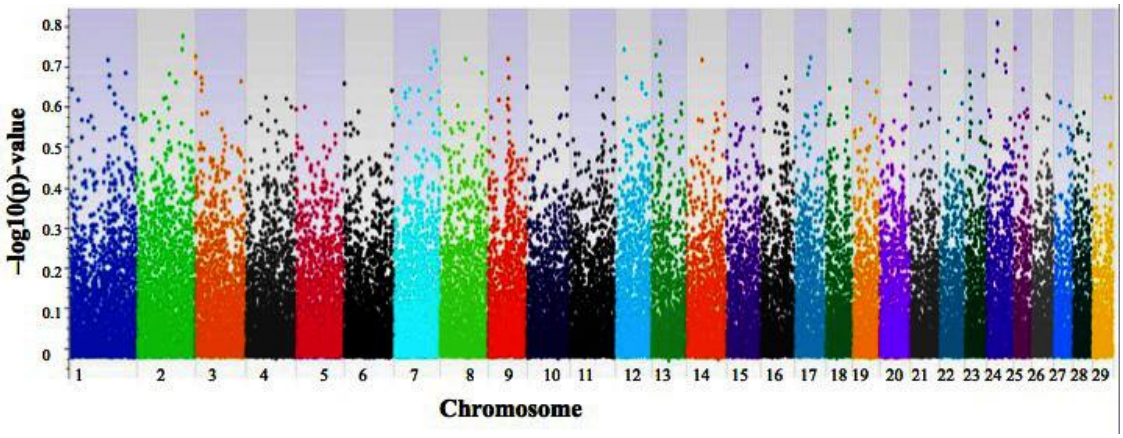


Figure 5.5 Distribution of F_{ST} across all 29 autosomes of Tankwa vs. commercial and domestic goats

5.3.6 Linkage disequilibrium variation

Information on the length of 29 autosomes and distribution of SNPs is presented in **Table 5.4**. The distribution of SNPs varied among the chromosomes but was consistent with the length of the chromosome, with chromosome 1 having the highest number of SNPs (n = 2991) and chromosome 25 having the lowest number (**Table 5.4**). The longest SNP interval of 245Kb was observed on chromosome 19. Linkage disequilibrium (r^2) averaged 0.07±0.11. The mean r^2 for the different autosomal chromosomes per population is summarized in Table 5.4. The lowest r^2 was observed in the village ecotypes and ranged from 0.09±0.12 in Zulu goats, 0.09±0.13 in the Xhosa and Venda goats, 0.11±0.14 in Tswana goats. The highest r^2 was observed on chromosome 6, whereas chromosomes 5, 23, 28 and 29 had the least r^2. Population by chromosome interactions were also observed. The Tankwa population had the highest average LD at chromosome 24 and the lowest r^2 on chromosome 29. Boer, Kalahari Red, Savanna, Tswana and Xhosa ecotypes had high r^2 on chromosome 6. The Nguni and Venda ecotypes had highest r^2 on chromosome 20, whereas r^2 was high on chromosome 27 of Zulu ecotypes. It was observed that breed, chromosomes and their interactions as well as SNP marker distance significantly influenced r^2 values (P < 0.0001; **Table 5.5**). High r^2 was observed in the Tankwa (0.25±0.26) followed by the commercial breed (Boer) and Nguni.

Table 5.4 Effect of chromosome on average r^2 values per and across populations

CHR	CHRL (Mb)	♯SNPs	Average r^2									
			Boer	Kalahari Red	Savanna	Tankwa	Nguni	Tswana	Venda	Xhosa	Zulu	All
1	155	2991	0.18±0.20	0.15±0.18	0.16±0.18	0.24±0.25	0.17±0.19	0.11±0.15	0.09±0.13	0.10±0.13	0.09±0.13	0.14±0.18
2	135.4	2571	0.18±0.20	0.15±0.18	0.15±0.17	0.25±0.27	0.15±0.18	0.11±0.14	0.09±0.12	0.09±0.12	0.09±0.12	0.14±0.18
3	116.8	2093	0.20±0.21	0.15±0.18	0.16±0.18	0.25±0.26	0.15±0.18	0.11±0.14	0.09±0.12	0.10±0.13	0.09±0.12	0.14±0.18
4	116	2181	0.17±0.19	0.14±0.16	0.14±0.16	0.25±0.28	0.16±0.18	0.11±0.14	0.10±0.13	0.11±0.14	0.09±0.13	0.14±0.18
5	111.1	2048	0.17±0.19	0.14±0.16	0.14±0.16	0.23±0.25	0.16±0.18	0.12±0.15	0.09±0.13	0.10±0.12	0.09±0.12	0.13±0.17
6	114.3	2128	0.23±0.25	0.20±0.22	0.17±0.19	0.24±0.26	0.17±0.20	0.12±0.15	0.09±0.14	0.10±0.14	0.09±0.14	0.16±0.19
7	106.5	1942	0.21±0.22	0.17±0.19	0.16±0.19	0.25±0.27	0.17±0.19	0.12±0.15	0.09±0.13	0.10±0.13	0.09±0.13	0.15±0.19
8	111	2094	0.21±0.22	0.18±0.20	0.17±0.20	0.23±0.25	0.17±0.19	0.11±0.13	0.09±0.13	0.10±0.13	0.09±0.12	0.15±0.19
9	90.3	1714	0.18±0.19	0.15±0.17	0.15±0.17	0.27±0.28	0.16±0.19	0.10±0.13	0.09±0.12	0.10±0.12	0.09±0.12	0.14±0.18
10	99.2	1926	0.16±0.19	0.13±0.16	0.14±0.16	0.24±0.26	0.15±0.18	0.09±0.13	0.08±0.12	0.09±0.12	0.08±0.12	0.14±0.17
11	105.3	1938	0.19±0.20	0.15±0.18	0.16±0.18	0.28±0.28	0.16±0.19	0.09±0.13	0.08±0.13	0.10±0.13	0.08±0.13	0.15±0.19
12	82.5	1521	0.21±0.22	0.17±0.20	0.18±0.20	0.28±0.30	0.16±0.19	0.11±0.15	0.09±0.14	0.11±0.15	0.09±0.13	0.15±0.20
13	80.6	1412	0.20±0.22	0.17±0.19	0.17±0.21	0.25±0.26	0.15±0.18	0.12±0.15	0.08±0.12	0.10±0.13	0.09±0.13	0.15±0.19
14	92.3	1697	0.19±0.21	0.18±0.20	0.16±0.18	0.23±0.25	0.15±0.18	0.10±0.13	0.08±0.12	0.11±0.13	0.09±0.12	0.14±0.18
15	79	1429	0.19±0.21	0.16±0.18	0.14±0.16	0.23±0.25	0.15±0.18	0.10±0.13	0.08±0.12	0.10±0.13	0.08±0.11	0.14±0.17

♯= number

Table 5.4 Effect of chromosome on average r^2 values per and across populations (continued)

CHR	CHRL (Mb)	♯SNPs	Average r^2									
			Boer	**Kalahari Red**	**Savanna**	**Tankwa**	**Nguni**	**Tswana**	**Venda**	**Xhosa**	**Zulu**	**All**
16	77.7	1457	0.16±0.18	0.14±0.17	0.13±0.16	0.27±0.28	0.16±0.18	0.11±0.14	0.08±0.12	0.10±0.13	0.09±0.12	0.14±0.18
17	71.9	1337	0.19±0.21	0.14±0.17	0.16±0.19	0.25±0.26	0.16±0.19	0.13±0.15	0.09±0.13	0.10±0.13	0.09±0.13	0.15±0.18
18	61.1	1132	0.18±0.21	0.16±0.19	0.15±0.18	0.26±0.27	0.16±0.20	0.11±0.15	0.09±0.14	0.11±0.14	0.09±0.14	0.15±0.19
19	62.1	1058	0.19±0.20	0.15±0.18	0.15±0.17	0.23±0.25	0.16±0.18	0.10±0.13	0.08±0.11	0.09±0.12	0.08±0.12	0.14±0.17
20	71.3	1305	0.18±0.20	0.17±0.18	0.18±0.19	0.26±0.28	0.18±0.20	0.11±0.14	0.11±0.14	0.10±0.13	0.09±0.13	0.15±0.19
21	66.8	1315	0.17±0.19	0.14±0.16	0.14±0.16	0.25±0.27	0.15±0.18	0.09±0.13	0.08±0.11	0.09±0.13	0.08±0.11	0.14±0.17
22	58	1055	0.19±0.21	0.13±0.16	0.16±0.18	0.23±0.25	0.15±0.18	0.11±0.14	0.09±0.13	0.09±0.13	0.08±0.13	0.14±0.18
23	49.4	960	0.16±0.18	0.13±0.15	0.12±0.15	0.27±0.27	0.15±0.17	0.09±0.12	0.07±0.10	0.09±0.11	0.08±0.11	0.13±0.17
24	61.8	1098	0.19±0.20	0.16±0.18	0.16±0.18	0.29±0.31	0.15±0.17	0.11±0.15	0.09±0.12	0.10±0.12	0.09±0.12	0.14±0.18
25	41.5	766	0.19±0.21	0.17±0.19	0.15±0.18	0.21±0.24	0.15±0.18	0.10±0.14	0.08±0.12	0.10±0.13	0.08±0.12	0.14±0.18
26	50.2	891	0.18±0.20	0.14±0.17	0.15±0.18	0.25±0.27	0.16±0.18	0.10±0.13	0.08±0.12	0.09±0.13	0.08±0.12	0.14±0.18
27	44.1	848	0.17±0.19	0.14±0.17	0.15±0.17	0.26±0.28	0.17±0.19	0.11±0.14	0.09±0.13	0.10±0.13	0.10±0.14	0.14±0.18
28	43.2	838	0.16±0.19	0.13±0.16	0.14±0.16	0.28±0.28	0.15±0.17	0.09±0.12	0.08±0.11	0.09±0.12	0.08±0.11	0.13±0.17
29	48.4	875	0.18±0.20	0.15±0.17	0.16±0.18	0.19±0.21	0.15±0.18	0.10±0.13	0.08±0.11	0.09±0.12	0.08±0.11	0.13±0.16
Overall		44660	0.18±0.20	0.16±0.18	0.15±0.18	0.25±0.26	0.16±0.18	0.11±0.14	0.09±0.13	0.09±0.13	0.09±0.12	0.07±0.11

♯= number

Table 5.5 Effects of breed, chromosome, SNP interval and interaction between breed and chromosome on measured r^2

Source	DF	Type III SS	Mean Square	F Value	*P* > F
Breed	8	8307.057870	1038.382234	34888.1	<.0001
Chromosome	28	198.141724	7.076490	237.76	<.0001
Distance	1	1477.642392	1477.642392	49646.6	<.0001
Breed*Chromosome	224	376.759358	1.681961	56.51	<.0001

5.3.7 Linkage disequilibrium decay

Linkage disequilibrium decay over increased SNP marker distances is illustrated in Figure 5.6. High r^2 (>0.4) was observed for SNP marker intervals within the minimum distance of <10Kb in all populations. r^2 rapidly decreased to an average of 0.2 across populations at 40Kb followed by a gradual decrease in r^2 below 0.2 at SNP marker intervals of 2000Kb observed for all breeds except the Tankwa (Figure 5.6). A more rapid decrease in r^2 was observed in the village ecotypes: An r^2 below 0.2 was observed at SNP marker intervals of 40Kb, and r^2 in the Boer was maintained above 0.2 up to SNP marker intervals of 1000Kb, whereas that of the Tankwa remained consistently higher than in other breeds and was above 0.2, even at 2000Kb SNP marker intervals.

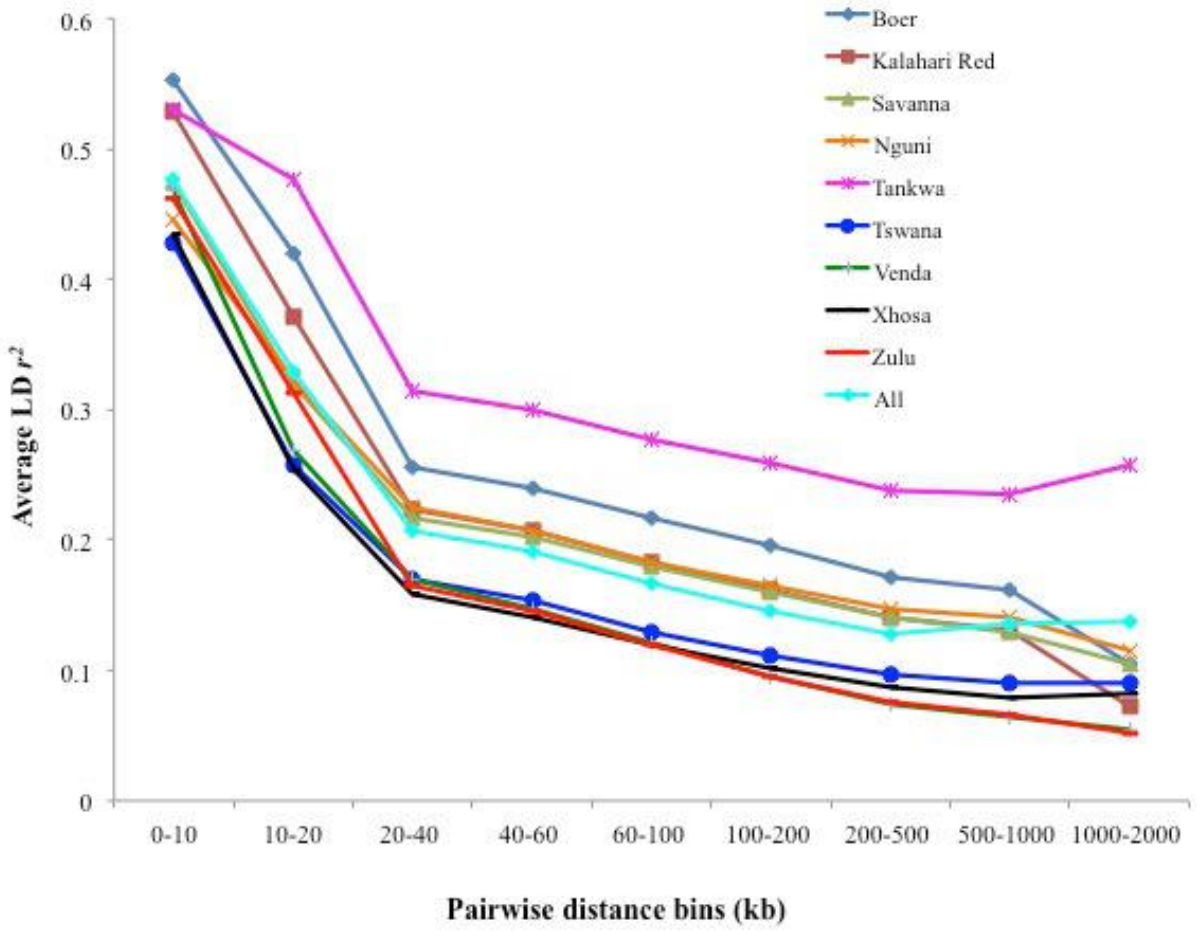


Figure 5.6 Decay of linkage disequilibrium over increased marker distances

5.3.8 Estimates of N_e

Estimated N_e (t) at t generations in the past are shown in the plots of Figure 5.7. Based on these results, a progressive decrease in N_e across generations was observed. The N_e indicated an effective population size of <150 for all the populations 13 generations ago. The major reduction in effective population size was observed 150 generations ago.

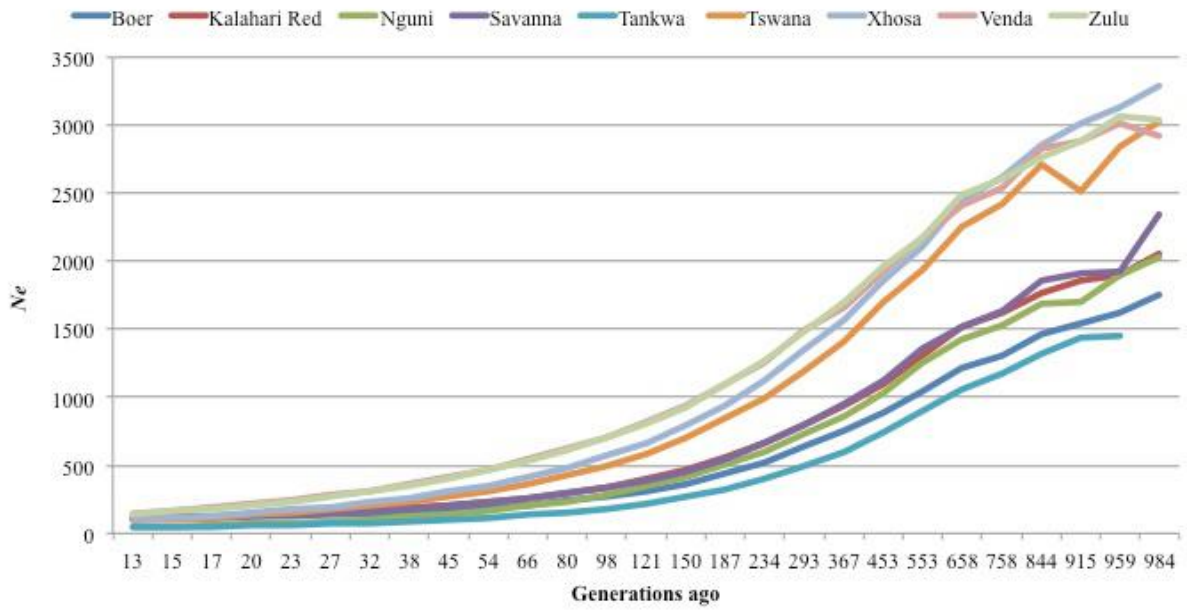


Figure 5.7 Effective population size over the past 984 generations

5.3.9 Correlation of gametic phase

Consistency of gametic phase between commercial meat-type breeds showed a higher correlation between Boer and Kalahari goats, as presented in Figure 5.8a. Correlation in gametic phase was >0.5 for marker pairs that were <10Kb in length and decreased to 0.5 for marker pairs within 20-40Kb, after which it fell to 0.25. A similar trend was observed for the other commercial breed pairs, although correlation of gametic phase fell close to 0, for Boer-Savanna and Savanna-Kalahari Red breed pairs, correlation of gametic phase fell beyond 60Kb. The persistence of gametic phase between the commercial and ecotypes goat populations are shown in Figure 5.8b. Highest correlation that persisted above 0.9 for all marker distances was observed between the Savanna commercial meat-type and Tswana village ecotypes. Among the ecotype populations, highest correlation was observed between the Zulu and Venda goats (Figure 5.8c). Consistency of linkage phase was also examined between Tankwa and the farm breeds, and the lowest correlation was observed between the Tankwa and Zulu as well as Venda goats (Figure 5.8d). Across all breed pairs, correlation of gametic phase took a sharp decline at the 20-40Kb marker distance.

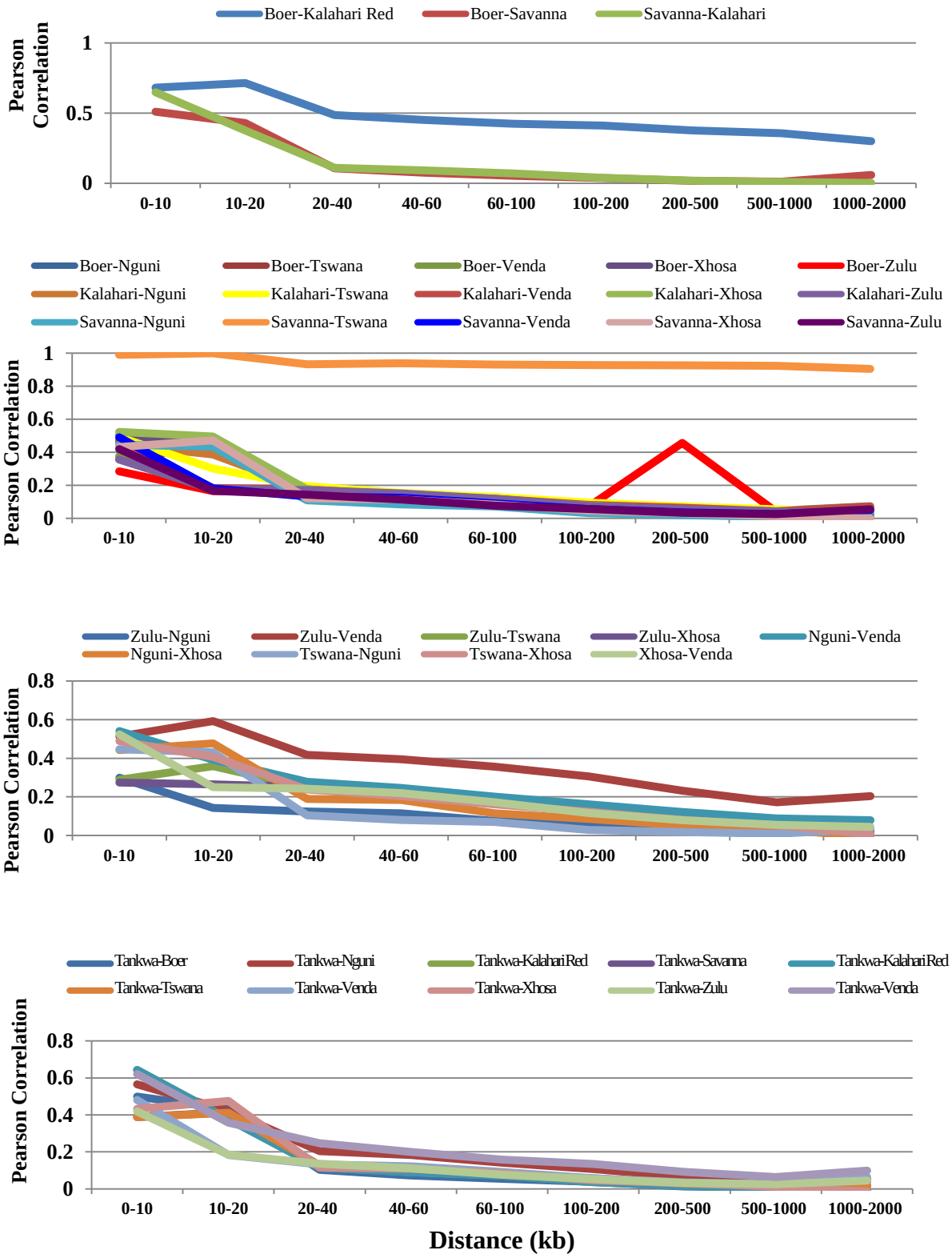


Figure 5.8 Pearson's correlations of signed r-values between all breed pairs (*n* = 36) at given distances. (a) Commercial meat-type, (b) commercial meat-type and ecotype populations (c) ecotype populations and (d) Tankwa and all domesticated breeds

5.4 Discussion

Previous studies suggest a valuable diversity in indigenous goat breeds that are claimed to be genetically diverse and unique and are adapted to particular eco-regions or production systems (Rege 1994; Morrison 2007). With the release of the Goat SNP50K panel in 2012 (Tosser-Klopp *et al.* 2012), several authors have applied this technology to assess genetic diversity, investigate loci under selection and evaluate genetic association between traits of economic importance (Kijas *et al.* 2013; Maroteau 2013; Brito *et al.* 2014; Palhière *et al.* 2014). Our study comprised a comprehensive analysis of South African indigenous goats using the genome-wide Illumina SNP50K BeadChip. The study demonstrated the utility of using the SNP50K BeadChip to study the diversity, population structure and breed relations of the local village ecotypes and feral goat populations.

The overall proportion of informative SNPs (87.09%) observed in the current study was comparable to the 43 759 (82.03%) observed for the South African Angora breed (Lashmar *et al.* 2015). A large proportion of SNPs was informative in the South African breeds, although only the Boer and Savanna goats were used in SNP discovery and development of the chip (Tosser-Klopp *et al.* 2014). The utility across non-discovery breeds could be attributed to a broad spectrum of breeds used when developing the chip, which represented diverse repertoire of dairy, meat and dual-purpose breeds as well as goats that not selected for any purpose/breeding objective. The observed utility of this chip in the South African population justifies the need to use representatives of global breed groups when generating genomic resources.

The level of polymorphic SNPs in this study was higher than has previously been reported for the Angora goat breed (Lashmar *et al.* 2015). However, the Angora goat study used a more stringent SNP call rate of 98% than the 95% call rate threshold used in this study. The differences in the SNP quality control thresholds could have influenced the resultant proportion of useful SNPs. Kijas *et al.* (2013) reported levels of polymorphism of 97% after applying a call rate threshold of 90%. The relatively lower number of informative SNPs in the Tankwa suggests its narrow genetic base and uniqueness from other South African breeds. High inbreeding due to a small breeding stock could have contributed to the low genetic diversity and relatively higher proportion of fixed alleles in the conservation of Tankwa population. The Tankwa may also have experienced founder effects and/or genetic drift when

fewer than 100 individuals were used to establish the conservation unit (pedigree data, Department of Agriculture-Northern Cape).

The higher genetic variability and low levels of inbreeding in the ecotype populations are attributed to the absence of strong artificial selection, the high gene flow and heterogeneous characteristics of communal smallholder production systems in South Africa. Broad and multiple breeding objectives result in effective population sizes that are much larger than for the commercial breeds that are raised as closed flocks with specific breeding goals. The level of among-population genetic variation observed in this study was lower than that reported in commercially developed breed populations including Angora (16.12%; Ozoje 2002; Pieters 2007). Genetic variation among the nine South African goat breeds was lower (6.39%) than the 7.49% reported for Italian goat breeds (Nicoloso *et al.* 2015).

Results from PCA and ADMIXTURE were in agreement, clustering populations according to the breed type and production system and emphasizing the unique genetic diversity in the Tankwa population. The results present the Tankwa breed as a unique genetic resource, which is in line with findings by Kotzé *et al.* (2014). Although Boer and Savanna were developed from unimproved indigenous goats, the PCA and ADMIXTURE results are generally consistent with breed history and separation into the different populations. The hypothesis that village ecotypes comprise several subpopulations reared in different geographic locations by different ethnic groups was supported at higher principal components and at ADMIXTURE $K = 5$. Although separated by geographic distances, Venda and Zulu ecotype populations have some genetic relatedness that may reflect the documented historic distribution of Nguni type goats and reflect similarities in production systems. Similar trends were seen in pairwise F_{ST} (Figure 5.4). Consistently across all analyses, the Tankwa is the most diverged of the goat breeds, possibly due to its unique genetic conformation and geographic isolation. Based on the genetic uniqueness of the Tankwa, analysis for outlier SNPs was conducted for Tankwa vs. the domesticated breeds from which over 100 loci within 39 genes were observed to be under selection. The genes *FAM71E1*, *ADCY9*, *USP37*, *SFT2D2*, and *CHD1* were found to be positively selected for.

The Tankwa and commercially developed meat-type breeds had higher average r^2 than did the non-descript village populations. However, the average r^2 for the South African Boer in this study was lower than the 0.28 observed for Boer goats in Canada (Brito *et al.* 2014). The

impact of low diversity and high inbreeding on r^2 present in the Tankwa goats and commercial meat-type goats has also been observed in similarly raised populations (Pritchard & Przeworski 2001). Whether low diversity is a threat to any of these goat populations has not been demonstrated; however, a general correlation between population fitness and genetic marker diversity has been demonstrated in sheep (Miller *et al.* 2014). The ecotype populations showed an r^2 that declined within the first 20-40Kb, whereas the r^2 for Tankwa commercial breeds had long-range r^2. The longer ranges of r^2 can be a result of intensive artificial selection in the improving commercial breeding populations and the reduction in effective population size, which is absent in village goat production systems. The r^2 also varied extensively between chromosomes, suggesting a variation in autosomal recombination rates due to the effects of genetic drift and selection within populations (Arias *et al.* 2009; Qanbari *et al.* 2010). Effective population sizes of the ecotype populations were slightly higher than for Tankwa and commercial breeds across generations, suggesting higher diversity and probable gene flow between village goat populations.

The persistence of LD phase is essential for effective genomic selection (Goddard *et al.* 2006) because it allows for determining the marker density to perform accurate multibreed genomic selection (Larmer *et al.* 2014). The persistence of gametic phase between Boer and Kalahari Red could be linked to breed relation and suggests the use of similar genomic tools across breed selection and in genomic evaluation programs. Brito *et al.* (2014) also observed the association between breed relatedness and high consistency of phase between the Australian and Canadian Boer goats. The high correlation between the Savanna and Tswana goats could be linked to their similarity in coat colors. The dominant color for the Tswana goats is white (Chapter 3), which is also characteristic of Savanna goats, suggesting that this correlation may be related to coat color. The low correlation among the ecotype populations was expected due to the high genetic within-breed diversity, heterogeneous nature of most village ecotypes as well as the lack of structured selection programs in communal production systems. Overall, the sudden decline in LD phase at low marker distance (2-40Kb) and the low correlation between breeds imply limited use of the goat SNP50K panel for genomic selection, particularly for across breed evaluations.

In conclusion, this study validated over 40 000 (80.09%) autosomal SNPs from the Illumina goat SNP50K panel that could be of use in genetic diversity studies of the South African goat populations. Results from this study presented valuable insights into the genetic diversity and

demographic history of South African indigenous goat breeds and ecotype populations in addition to highlighting the urgent need to implement crucial management and conservation strategies. Special consideration should be given to the Tankwa goat population, which appears to be the only remaining representative of a unique feral goat genetic resource. Tankwa also has the lowest genetic diversity; high levels of inbreeding and low effective population size and is the most geographically isolated population. Thus, initiatives for its sustainable conservation and utilization are imperative. The existence of multiple and fragmented goat breeds and populations, the majority of which are raised by communal farmers, hinder genetic improvement programs of South African goats. This study highlighted the limited use of the current SNP chip in across breed genomic selection programs with the exception of the Boer and Kalahari Red breeds.

CHAPTER 6: LANDSCAPE GENOMICS AND PATHWAY ANALYSIS TO UNDERSTAND GENETIC ADAPTATION OF SOUTH AFRICAN INDIGENOUS GOAT POPULATIONS

ABSTRACT

There is scant information regarding the response of South African indigenous goat populations to natural selection presented by the heterogeneous production systems and environment. In the previous Chapter, patterns of genetic differentiation among the breeds suggestive of local genetic adaptation were shown. The objectives of this study were to (i) describe the genetic structure of the South African indigenous goat populations across the geographical regions; (ii) evaluate the relative contribution of the environment in shaping the genetic variation; and (iii) screen loci under selection in geographically distant South African goat populations using a combination of the Illumina goat SNP50K data and GIS coordinates from various goat sampling locations. Population structure analysis showed that the isolated population of the Tankwa breed formed a separate group with a homogeneous genetic structure restricted to the Northern Cape province of South Africa. Village goat populations clustered into two genetic groups, one composed of goats from Limpopo and KwaZulu-Natal and the other composed of goat populations of the Eastern Cape and North West. The commercial breeds represented a genetic cluster spread across multiple sampling and geographic locations. The study assessed and quantified the relative importance of environmental factors (geographic location, altitude, annual and seasonal trend of temperature, and rainfall) on genetic variation of goat populations using Redundancy Analysis (RDA). Significantly more genetic variation was explained by climatic variables than geographic regions. Combined effects of climatic and geographic variables explained 14.34% (R^2_{adj} = 9%) of the total genetic variation. Correlation-based landscape genomic approaches (SAM and LFMM) were used to identify loci associated with environmental factors. SAM identified a total of 444 (0.92%) SNPs, while LFMM identified 377 (0.78%), and RDA detected 149 SNPs that were significantly associated with environmental variables. Significant markers were found within genes involved in diverse biological functions potentially important for environmental adaptation. Overall, the study suggested a strong role of environmental factors in shaping the genetic variation of South African indigenous goat populations. Loci observed to be significant and under selection may be responsible for the adaption of the goat populations to local production systems and environments.

Keywords: *Goats, production systems, landscape genomics, Spatial Analysis Method, Latent Factor Mixed Models, RDA, adaptive loci, environments.*

6.1 Introduction

Smallholder farmers raise the largest number of indigenous goats in South Africa, with a large portion kept in the arid zones of the country. According to the Joint Agricultural Weather Facility (1999) (agro-ecological zone classification) and the report on historic distribution of goat ecotype populations by Morrison (2007), majority of the Nguni ecotype goats occur in the arid and subtropical wet agro-ecological zones while the Eastern Cape Xhosa lobbed goats occupy the arid zones of the country. The Northern Cape Skilder goats occur in the desert-like agro-ecological zones. The Tankwa goats, which are a unique feral breed are found in the hot and drier desert-like agro-climatic zones of the Northern Cape. South African local goat populations have demonstrated adaptive genetic and phenotypic variation that is distributed across the geographical and environmental gradients. These indigenous goats have also demonstrated the ability to thrive and be productive in unfavorable environmental conditions characterized by high temperatures, humidity, and numerous pests and diseases.

Although South African indigenous goats are considered to be generally highly adapted to local environmental conditions, human interventions such as intensive management and selection under commercial goat farming systems can considerably reduce population genetic diversity and increase inbreeding thereby threatening their ability to cope with climate change. The Boer, Savannah and Kalahari Red goats of South Africa are such breeds that, although indigenous to South Africa, have been selected and reared under intensive commercial systems of production. Conversely, communal and feral goat populations are not buffered from the effects of adverse climatic and environmental conditions. Their geographic and environmental conditions impose natural selection pressures, which increase allele frequencies of selected genes involved in local adaptation.

Assessing the natural selection pressure and affected genetic loci will give insight into genetic mechanisms underlying local adaptation in marginal goat populations, which is crucial for guiding their genetic improvement strategies and conservation programs. South Africa has diverse goat breed populations raised in extreme and severe climates and heterogeneous production systems. Thus, it was hypothesized that the genetic diversity observed in local

goat populations is a reflection of the broad climatic and production systems under which they are raised. By imposing diverse selection pressures across geographic regions and environmental factors, the different production systems and host-pathogen interactions define the population genetic structure and distribution that determines the level of local adaptation. The environment plays a key role in shaping the genomic architecture of livestock populations and their continued evolution and consequently should be incorporated in evolutionary studies.

Landscape genomics is a field of study that combines population genetics, spatial statistics, and landscape ecology to decipher the geographic and environmental processes at play in population genetic structuring along the geographic gradients within production systems (Manel *et al.* 2010). It has proven a valuable tool in understanding the effects of the geographic and environmental factors and in identifying production landscape variables that influence the genetic structure of indigenous goat populations. Genetic variation that culminates in climatic adaptation in livestock species was suggested from research using phenotypic data (Moraa *et al.* 2015) and low density markers (Meutchieye *et al.* 2014). Previously, Pariset *et al.* (2009) found distinct patterns of genetic variation using twenty-seven SNP markers along environmental gradients and reported a correlation between 16 loci and the environmental parameters in North-East Mediterranean goat breeds. (Colli *et al.* 2014) also used AFLP markers and found associations with diurnal temperature range, frequency of precipitation, relative humidity and solar radiation in European and Western Asian Goat Breeds.

The great improvements in genomics and bioinformatics are facilitating higher resolution of livestock genomes and the identification of loci that are potentially of ecological significance (Manel *et al.* 2010). The availability of goat genome sequence data and the Illumina goat SNP50K panel are improvements providing comprehensive genome coverage and allowing finer mapping of loci under selection. Recently, a number of studies employing different analytical concepts have provided useful information on signals of selection on a genome-wide scale in indigenous cattle (Makina *et al.* 2015), and goats and sheep (Kim *et al.* 2016). This study sought to investigate the genetics of adaptation in South African goat populations using the Illumina SNP50K data and landscape genomics approaches. South African goat populations are exposed to different natural selection pressures from different production and environmental conditions. The study hypothesized that the divergence of different goat

populations is a reflection of the diverse selection pressures emanating from environmental factors. The aims of the study were therefore to: i) reveal the geographic population genetic structure; ii) assess and quantify the relative contribution of geographic and environmental factors to genetic variation using a combination of genetic and site-averaged environmental data iii) reveal evidence of associations between production environment and genetic variation; and iv) investigate the biological pathways effected by selection to ensure adaptation of indigenous goats to local environments. This study used three landscape genetics approaches namely (i) multivariate method (Redundancy Analysis; RDA), (ii) the spatial method (Spatial Analysis Method; SAM), and (iii) the latent factor mixed methods (LFMM) that accounts for the effect of population structure.

6.2 Materials and Methods

6.2.1 Geographic origin of goat populations and Genotyping

Goat populations were sampled from five provinces, Eastern Cape, KwaZulu-Natal, Limpopo, North West and Northern Cape that fall under three agro-ecological zones representing sub-tropical wet, arid and desert conditions. The subtropical wet agro-ecological zone consists of wet and humid eastern coastal regions of the Eastern Cape and KwaZulu-Natal provinces while dry conditions are experienced in the northwestern region (arid zone) of the Eastern Cape and KwaZulu-Natal and the Limpopo and the North West provinces. The Northern Cape is located in the desert zone of the country with a maximum summer temperature of 48°C. The same individuals were previously used in a 50K SNP genetic diversity study (Chapter 4, Figure 4.1). We excluded individuals without GPS coordinates (latitude and longitude), those with >5% missing data and related individuals (IBD > 0.45). The filtering criterion resulted in 194 goats representing South African indigenous goat populations from 19 sampling locations in the form of villages, commercial farmers and research stations. Breed representation included goats from locally developed meat-type breeds of the Boer (n = 33), Kalahari Red (n = 40) and Savanna (n = 31); non-descript village ecotype populations from smallholder farmers (n = 100) from Eastern Cape, KwaZulu-Natal, Limpopo, North West; Nguni breed from Kwazulu-Natal (n = 10); and a feral Tankwa (n = 25) goat population from the Carnarvon, Northern Cape. The random sampling represented an extensive coverage of major agro-ecological zones and diversity of the environmental conditions in the country (sampling locations and coordinates for each individual is provided in APPENDIX C. SNP subsets were

different for each analysis and have the quality control criterion used is discussed in detail for each method.

6.2.2 Climate data

Geographical coordinates (longitude and latitude) and altitude (m) were captured for each goat sampled. The Agricultural Research Council Institute for Soil Climate and Water provided climate data from 19 weather stations (Figure 6.1) over several years, and were averaged for 5 years (**Table 6.1**). Only data sets that had greater than 80% of the annual climatic data were utilized to ensure accurate trend results for each sampling location. Climate data included minimum and maximum trends of mean annual range in temperature (°C); mean annual range in total rainfall (mm) as well as seasonality trends of temperature (°C) and rainfall (mm) during summer (December-February) and winter (June-August) months. Means for the climatic variables of the sampled regions are given in **Table 6.1**.

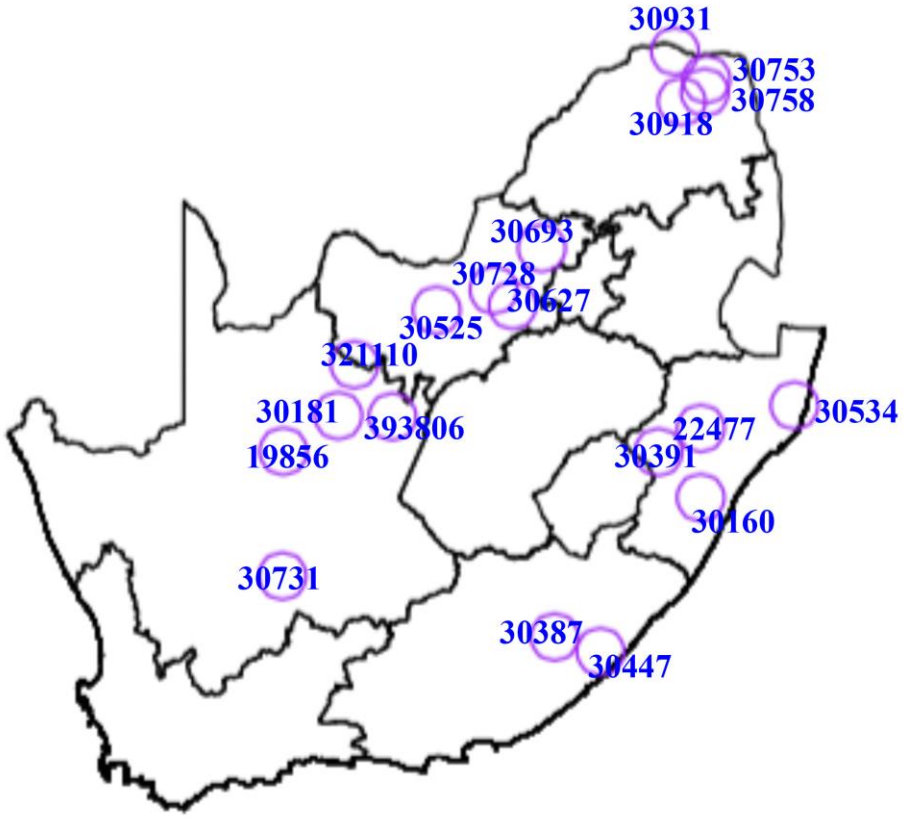


Figure 6.1 Map of South Africa with the 19 weather stations where climatic data was obtained

Table 6.1 Descriptions of climatic variables from weather stations

Province	Station	Climatic variables								
		Atmax	Atmin	Stmax	Stmin	Wtmax	Wtmin	AR	SR	WR
EC	30387	25.01	10.75	28.04	16.1	17.85	4.13	1.57	2.61	0.79
EC	30447	28.8	11.61	27.19	14.82	21.8	9.19	2.11	3.25	0.97
KZN	30160	28.8	11.61	27.19	14.82	21.8	9.19	2.11	3.25	0.97
KZN	30391	26.85	9.7	29.63	15.83	23.76	2.48	2.47	3.88	0.79
KZN	30534	28.8	11.61	27.19	14.82	21.8	9.19	2.11	3.25	0.97
KZN	22477	22.42	11.89	25.83	15.63	18.36	7.97	69.27	74.15	66.53
LMP	30931	32.6	15.14	35.52	20.93	29.03	7.92	1.06	3.19	0.1
LMP	30918	27.27	13.17	29.44	17.59	24.53	8.19	2.54	6.36	0.23
LMP	30758	27.18	14.54	29.04	18.72	23.72	9.1	2.35	5.33	0.24
LMP	30753	26.87	16.85	28.9	20.25	23.67	13.17	2.73	5.4	0.52
NW	30525	25.78	11.31	29.49	16.98	20.39	4.54	1.38	3.14	0.27
NW	30728	24.76	9.83	28.27	15.42	19.39	2.87	1.36	2.78	0.24
NW	30693	25.57	8.09	28.91	14.9	20.87	0.18	1.15	2.31	0.1
NW	30627	27.82	12.77	30.59	18.19	23.2	5.69	1.39	2.69	0.46
NC	30594	27.25	11.39	32.08	17.79	20.45	3.85	1.57	3.39	0.49
NC	30181	29.02	10.28	34.68	17.71	21.73	1.56	0.62	1.39	0.11
NC	30731	23.98	8.7	31.1	14.99	16.51	1.63	0.61	1.21	0.31
NC	393806	27.96	9.7	31.69	16.47	26.41	2.26	21.35	40.18	5.82
NC	321110	26.79	10.13	31.66	16.97	20.88	2.31	15.38	32.69	4.23

EC: Eastern Cape, KZN: KwaZulu-Natal, LMP: Limpopo, NW: North West, NC: Northern Cape, Atmax: annual maximum temperature, Atmin: annual minimum temperature, Stmax: summer maximum temperature, Stmin: summer minimum temperature, Wtmax: winter maximum temperature, Wtmin: winter minimum temperature, AR: annual rainfall, SR: summer rainfall, and WR: winter rainfall.

6.2.3 Population geographic structure

The filtering criterion resulted in 194 goats representing South African indigenous goat populations from 19 sampling locations in the form of villages, commercial farmers and research stations. The random sampling represented an extensive coverage of major agro-ecological zones and diversity of the environmental conditions in the country. The SNP panel

(n = 23 234) used had a call rate of 95%, MAF > 0.05, HWE P > 0.001 and r^2 > 0.2. Population structure computed using ADMIXTURE 1.21 (Alexander *et al.* 2009), from K = 2 to K = 10 independent runs, from which K = 5 was considered the optimal K value based on lowest CV error (Chapter 5), was used to compute cluster membership coefficients. The membership coefficients for each individual were then plotted as pie charts using *maps* function in R package (R Core Team 2013) onto the South African geographic map. In cases where numerous individuals were sampled in the same geographic location, pie charts were plotted near their original geographic location for ease of visualization.

6.2.4 Partitioning of genomic variation amongst geographic and climatic variables

A subset of 36128 SNPs with a genotyping call rate of 100% (including MAF 0.05, HWE P > 0.001) was used in the multivariate regression redundancy analysis (Van Den Wollenberg 1977; Legendre & Legendre 1998) to estimate the degree to which geographic coordinates, climate variables, and their combination explained the genomic variation in the indigenous goat populations. Genetic data (dependent matrix) and individual-specific climate and geographic variables (explanatory matrix) were compared in a partial RDA using 3 models implemented in *vegan* function in R package (R Core Team 2013). Model I was the full RDA model and it incorporated all geographic and climatic variables as explanatory variables. Two partial RDA models were also run: Model II, specified geographic variables (latitude and longitude) as a third 'conditioned' matrix and only considered variation explained by climate only, whilst Model III used geographic variables as explanatory variables only (Gugger 2012). To estimate the percentage of variance accounted for by each component and the direction of the effect, RDA scores were plotted for each individual and vectors depicting the direction of the predictor variables relative to the RDA axis.

6.2.5 Locus specific landscape genomics analysis to identify environment-associated SNPs

Landscape genomics approaches were used to determine SNPs significantly associated with the geographic variables (latitude, longitude and altitude), and climatic variables (Atmax, Atmin, Stmax, Stmin, Wtmax, Wtmin, AR, SR, and WR). For this analysis we excluded SNPs with >5% missing data, call rate of 95%, MAF > 0.05, HWE P > 0.001. The first landscape genomics analysis method used an individual-based Spatial Analysis Method (SAM), developed by Joost *et al.* (2007) and implemented in SAMβADA software program

[lasig.epfl.ch/sambada] to investigate the effects of the environment on locus-specific differentiation. The SAM method uses a logistic regression model whereby individuals were coded as either presence/absence for each given SNP allele and the association between the allele and the environmental parameters was measured across sampling sites. The model was considered significant when both the G and Wald tests were significant following a Bonferroni correction at 99% confidence level (Joost *et al.* 2007).

The second method used a non-spatial Latent Factor Mixed Model (Frichot *et al.* 2013) and was used to evaluate signals of environmental adaptation while controlling for false positives by accounting for population structure based on the ADMIXTURE-based population genetic structure at $K = 5$. An MCMC algorithm was used for each of the variables using 5000 sweeps for burn-in and 10 000 additional sweeps to compute LFMM parameters ($|z|$-scores) for all loci (Frichot *et al.* 2013). A false discovery rate (FDR) of 0.01 was used for detection of the outlier loci.

6.2.6 Outlier SNP detection

An outlier analysis of RDA results was done to determine SNPs that were strongly linked with multivariate environmental gradients. Outliers were identified as SNPs with the greatest squared scores (99%CI) along the first RDA axis of Model I using PROC UNIVARIATE analysis of Statistical Analysis System (SAS institute Inc. 2013).

6.2.7 Gene annotation and Pathway analysis

Significant SNPs were annotated to corresponding genes within the candidate gene region intervals (10,000 base pairs up and downstream of the significant SNP) were retrieved from the National Centre for Biotechnology Information (NCBI; www.ncbi.nlm.nih.gov) genome browser using the *Capra hircus* genome assembly ASM170441v1 (Accessed October 2016). Enriched functional annotation clusters were defined using a functional annotation tool implemented in Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa *et al.* 2004) database.

6.3 Results

6.3.1 Spatial Representation of Population Structure

Assuming five ancestral populations provided the lowest cross-validation standard error (0.62) indicating that this is the most likely number of ancestral breeds (Chapter 5) we used $K = 5$ to investigate spatial representation of population structure. The geographic distribution of the $K = 5$ population clusters across and along the geographical coordinates of South Africa is shown in Figure 6.2. The Tankwa goats represented as blue circles clustered as a small homogeneous genetic structure restricted to the Northern Cape province. The ecotypes were admixed to varying degrees, corresponding to two genetic substructures, one consisting of the Venda, Zulu and Nguni and the other of Tswana and Xhosa ecotype populations. Zulu, Venda and Nguni showed the highest Xhosa and Tswana genetic proportion, however, to a lesser degree, Zulu, Venda and Nguni genetic component was present in the Xhosa and Tswana ecotype population. The Zulu, Venda and Nguni genetic component was geographically distributed in Limpopo and KwaZulu-Natal provinces whilst the Tswana and Xhosa genetic component was predominantly found in North West and Eastern Cape provinces. Commercial breeds (Boer, Kalahari Red and Savanna) shared genetic components spread across geographic regions of Eastern Cape, KwaZulu-Natal, North West and parts of the Northern Cape provinces. In addition, few Savanna and Kalahari Red goats from the Northern Cape (-29.0313 latitude, 23.1826 longitude) showed low level of admixture and appeared to be the most homogenous individuals of these two breeds representing independent populations whilst the ones from other geographical locations are admixed (Figure 6.2). ADMIXTURE maps plotted for the Boer, Kalahari Red and Savanna are in APPENDIX D.

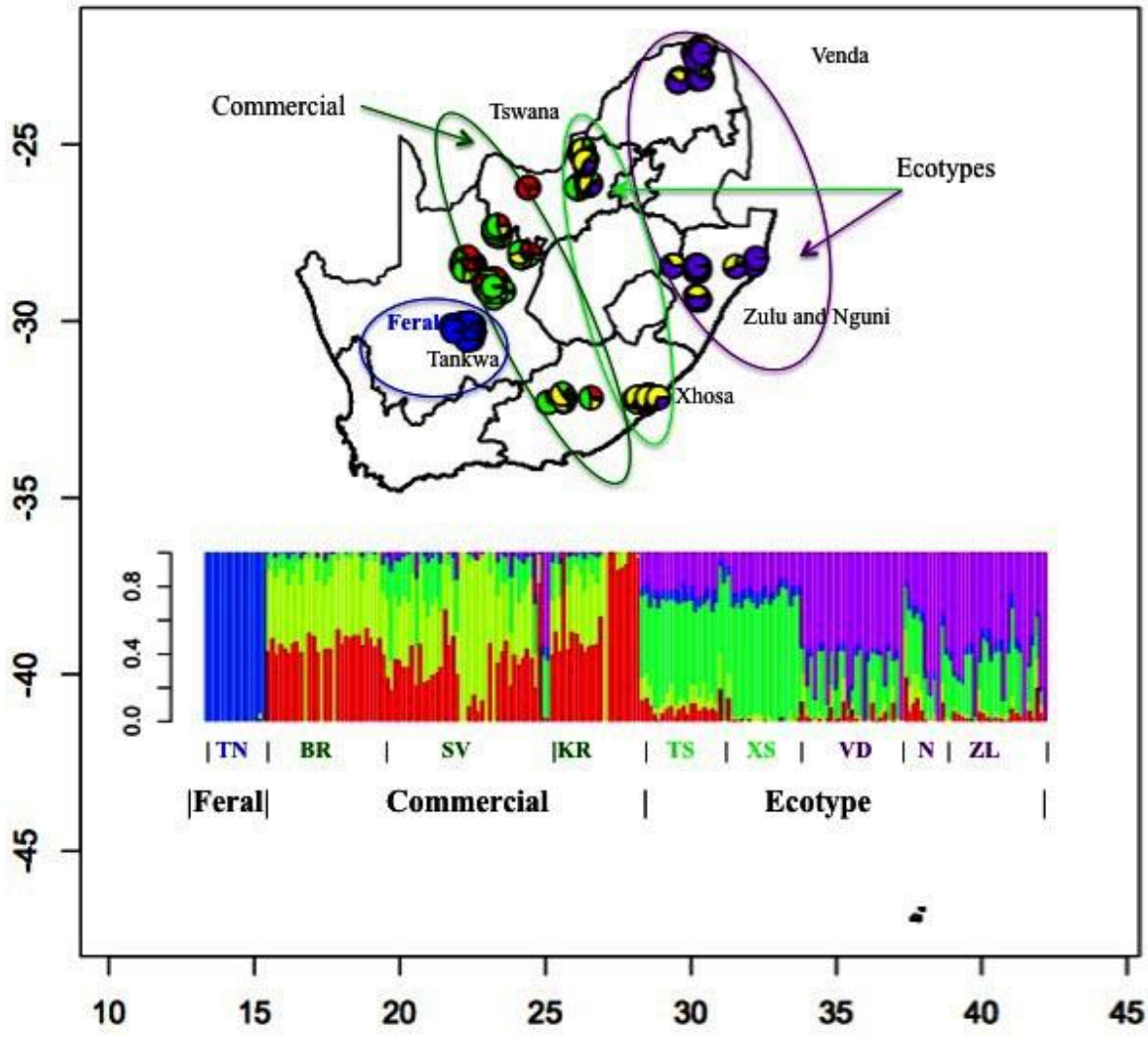


Figure 6.2 Map of South Africa. The ADMIXTURE plot represents average coefficients of membership resulting from the genetic structure analysis (best fit model, *K* = 5). Each colour represents a different gene pool. The included barplot represents each accession as a single vertical bar broken into *K* colour segments, with lengths proportional to the estimate probability of membership in each inferred cluster

6.3.2 Redundancy analysis models and the genomic variation partitioning

The contributions of the environmental and geographical variables to variation in genetic diversity are shown in **Table 6.1**. The Monte Carlo test permutation showed that the environmental and geographical variables explained a significant (p > 0.001) proportion of the genetic variation. Climate and geographic factors together (Model I) explained 14.34% (R^2_{adj} = 9%) of the total variation. Partitioning of variance analysis indicated that climate

alone (Model II) accounted for 9% (R^2_{adj} = 5%) of the variation. Geographic variables (Model III) on the other hand explained 1% (R^2_{adj} = 1%) of the total variance (**Table 6.2**).

Table 6.2 Redundancy analyses (RDA) of the contribution of both climate and geographic variables (Model I); climate alone (Model 1I), and geographic variables only (Model III) in genetic variation

Models	**Propotion**			R^2_{adj}	***P*-Value**
	Contrained	**Uncontrained**	**Conditioned**		
Model I (Climate and geographic)	0.14	0.86	-	0.09	0.001
Model II (Climate)	0.09	0.86	0.05	0.05	0.001
Model III (Geographic)	0.01	0.86	0.13	0.01	0.001

In Model I, variables strongly associated with the genetic variation were altitude, annual minimum temperature, winter minimum temperature, and longitude (Figure 6.3a). For Model II, partial redundancy analysis, annual maximum temperature, winter maximum temperature and altitude accounted for more variation than other variables when geographic variables were controlled (Figure 6.3b).

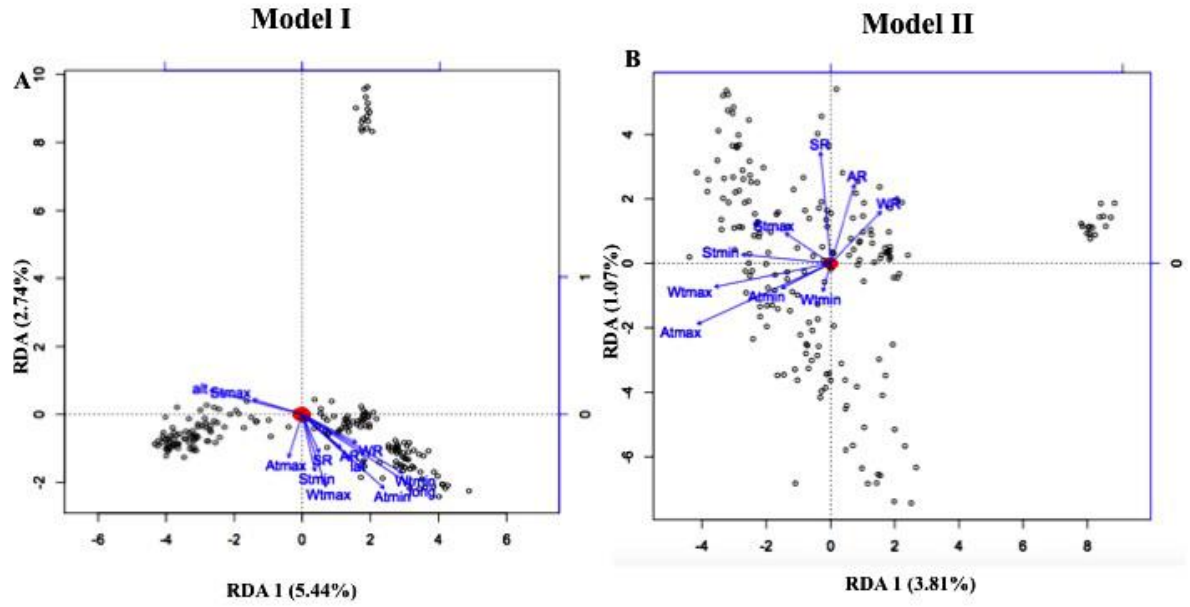


Figure 6.3 Redundancy analyses of contribution of both climate and geographic variables (Model I); climate alone (Model 1I) and geographic variables only (Model III) to genetic variation

6.3.3 Locus specific landscape genomic approach for SNPs associated with the environmental variables

SAM analysis detected a total of 444 SNPs (0.92%) showing significant association with one or more geographic and climatic variables using univariate models. The highest number (*n* = 422) of SNPs were associated with longitude while 160 SNPs were associated with winter minimum temperature (Wtmin). Only 15 SNPs were significantly associated with annual minimum temperature (Atmin) and 13 with altitude. No significant SNPs were observed for other variables. The significant SNPs and associated genes for the different climatic or geographic variables are presented in APPENDIX E.

The association of SNP markers with geographic and climatic variables was also analysed using LFMM analysis. The population structure-based association analysis revealed 377 (0.78%) SNPs at 1% FDR to be associated with eleven variables. The highest number of outlier SNPs were associated with summer minimum temperature (*n* = 54), followed by summer maximum temperature (*n* = 52), annual maximum temperature (*n* = 49), latitude (*n* = 46), winter rainfall (*n* = 41), annual rainfall (*n* = 39), summer rainfall (*n* = 39), annual

temperature (n = 36) and winter maximum rainfall (n = 17), while the lowest (n = 2) were associated with longitude and winter minimum rainfall. A complete summary of the LFMM analysis is presented in APPENDIX F.

Five (n = 5) outlier SNPs associated with different geographic and climatic variables were common between SAM and LFMM (**Table 6.3**). Among the SNPs commonly detected was snp59985-scaffold999-128817 (chromosome 6), snp31862-scaffold356-2790706 (chromosome 8), snp36146-scaffold431-10361406 (chromosome 12), snp9199-scaffold1335-368697 (chromosome 19) and snp14653-scaffold1591-11752 (chromosome 20).

Table 6.3 Summary of the single nucleotide polymorphism (SNP) detected by both Spatial Analysis Method (SAM) and Latent Factor Mixed Model and the associated environmental variables

SNP	Environmental variables		Chromosome
	SAM	LFMM	
snp59985-scaffold999-128817	Longitude	Atmax and Stmax	6
snp31862-scaffold356-2790706	Longitude	Stmax	8
snp36146-scaffold431-10361406	Atmax	Atmax, longitude and WR	12
snp9199-scaffold1335-368697	Longitude	Stmin	19
snp14653-scaffold1591-11752	Longitude and latitude	Longitude and latitude	20

Atmax: annual maximum temperature, Stmax: summer maximum temperature, Stmin summer minimum temperature, and WR: winter rainfall.

6.3.4 Outlier detection

The first RDA axis for Model I detected 149 outlier SNPs using a threshold of squared scores >0.14 (APPENDIX G). Of the hundred-forty nine detected SNPs, seven loci (snp28079-scaffold300-3453358, snp56493-scaffold89-902962, snp25341-scaffold261-1157571, snp49503-scaffold706-1271483, snp19366-scaffold1958-28303, snp12149-scaffold145-736257, snp45014-scaffold614-1443654) were also detected using the SAM approach for and were all associated with longitude (**Table 6.4**).

Table 6.4 Summary of the single nucleotide polymorphism (SNP) detected by both Spatial Analysis Method (SAM) and Redundancy Analysis (RDA)

SNP	SAM	RDA score[2]	Chromosome
snp25341-scaffold261-1157571	Longitude, Atmin	0.19	3
snp45014-scaffold614-1443654	Longitude	0.16	2
snp49503-scaffold706-1271483	Longitude	0.16	4
snp19366-scaffold1958-28303	Longitude	0.16	6
snp12149-scaffold145-736257	Longitude	0.15	1
snp56493-scaffold89-902962	Longitude	0.15	3
snp28079-scaffold300-3453358	Longitude	0.15	8

Atmin: annual minimum temperature.

6.3.5 Associated candidate genes and functional analysis

Within the candidate region intervals reported from SAM, LFMM and RDA, a total of 395, 220 and 97 genes were respectively identified (APPENDIX E-G). Two associated candidate genes (*GRID2* and *MARCH1*) were observed by all three methods, while 13 (*KCTD16*, *C8B*, *DGKB*, *DUSP16*, *C3H1orf87*, *SPATS2L*, *UHRF2*, *LDB2*, *NLGN1*, *VAV3*, *CNTNAP2*, *ST6GALNAC3*, and *NTNG1*) were common between RDA and SAM analyses. Seven associated genes (*PPM1H*, *DHX35*, *NSF*, *OCLN*, *CCSER1*, *LOC102187959*, and *SORCS2*) were common between SAM and LFMM, while only one was common for LFMM and RDA (*PIK3C2G*; Figure 6.4). The functional implications of the associated candidate genes were investigated using the KEGG pathways (http://www.kegg.jp; last accessed October 16, 2016) in goats. In the SAM analysis, the genes detected were involved in 183 pathways, while 75 and 87 pathways were observed for LFMM and RDA, respectively.

Amongst the functional associations detected, numerous pathways were considered particulalry relevant in adaptation of indigenous goats to their environment and production system. Among the genes significantly correlated with environmental variables detected by SAM, several notable examples included pathways involved in metabolic adaptations to poor feed, energy requirements and digestive efficiency, which included genes involved in metabolic pathways (*PLCB1*, *APIP*, *PLCH2*, *MTMR3*, *GOT2*, *DCXR*, *MTHFS*, *ABAT*, *XYLT1*, *PLCE1*, *PANK1*, *IDO2*, *ST6GALNAC3*, *NT5C1A*, *DGKB* and *GABBR2*), bile

secretion (*SCTR* and *ABCG2*), gastric acid secretion (*PLCB1*), pancreatic secretion (*SCTR*, *KCNMA1* and *PLCB1*), protein digestion and absorption (*COL6A6*, *PAG1* and *COL21A1*), and salivary secretion (*PLCB1*, *ADRA1D* and *KCNMA1*). One gene (*NSF*) was involved in the vasopressin-regulated water reabsorption pathway. Ten genes were found in pathways involved in response to diseases including African trypanosomiasis pathway (*PLCB1* and *IDO2*), Chagas disease pathway (of American trypanosomiasis; *PLCB1*, *GNA15*), Prion diseases pathway (*NCAM1*, *C8A* and *C8B*), tuberculosis pathway (*JAK2* and *ARHGEF12*), amoebasis (*C8A*, *C8B*, and *PLCB1*) and Toxoplasmosis (*NCAM1* and *BIRC2*). In addition, genes involved pathways responsible for thermoregulation included beta-transducin repeat containing E3 ubiquitin protein ligase (*BTRC*; circadian rhythm); and two involved in the circadian entrainment (*PLCB1*, *RYR3*).

Six significant genes (*PIK3C2G*, *TNK2*, *HYAL4*, *DNMT1*, *MTAP*, and *AK5*) were observed using LFMM and found to be regulating metabolic pathways. *MAPK10* gene on chromosome 6 influences the disease response pathways such as Chagas disease (American trypanosomiasis), Pertussis, Tuberculosis and Toxoplasmosis. Among the genes for environmental adaptation, three genes (*KCNJ3*, *CACNA1*, and *GRIA4*) were located in the circadian entrainment pathway. One gene (*NSF*) was found to influence the vasopressin-regulated water reabsorption, while *KCNMB2* was located in the pathway for vascular smooth muscle contraction. Further, the RDA analysis revealed six candidate genes (*CMAS*, *DGKB*, *PIK3C2G*, *CXADR*, *ST6GALNAC5*, *EPHX2* and *CMAS*) ranking within the top squared scores (>0.14) along the first RDA axis to be functionally involved in metabolic pathways. The gene *RAB5A* was located in the tuberculosis and the vasopressin-regulated water reabsorption pathways. The gene *ITPR2* was involved in numerous secretion pathways including those for the salivary, gastric acid and pancreatic secretions.

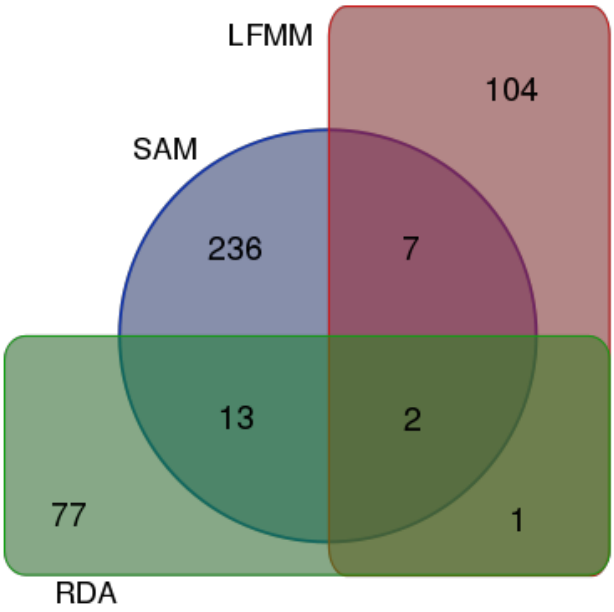


Figure 6.4 Venn diagram for genes detected among the three methods SAM, LFMM and RDA

6.4 Discussion

South Africa has considerable variation in climate and topography and is mainly classified as semi-arid. The adaptation of livestock to local environmental conditions is important for survival. This is especially true for goat populations that are raised under limited or no human intervention. Population structure typically results from a combination of demographic and adaptive factors (Ometto *et al.* 2015). Results from Chapter 5 demonstrated strong patterns of genetic differentiation of the South African goat populations. Analysis in Chapter 5 did not however, explore the actual causative factors responsible for structuring this genetic variation. This chapter therefore evaluated the contribution of the environmental and geographical factors on the genetic variation. It further correlated patterns of genetic variation with temperature and rainfall-related variables to specific loci to provide new insights on the adaptive potential of South African goat populations. An understanding of the population genetic structures and the associated geographic and environmental effects is important in guiding localized conservation and utilization of biodiversity. The study used landscape genomic approaches, which, although having growing interest in the study of population genetic structures, have not received enough recognition for their capacity to investigate the genetics of local adaptation in indigenous livestock species.

Results of geographical clustering detected high to moderate levels of genetic differentiation within the goat populations across the geographic regions. The Tankwa goat breed that is genetically distinct from the commercial and domestic village goat populations (Kotze *et al.* 2014) was compared to other goat populations restricted to a single geographic region of the Northern Cape province (Figure 6.2). This region is located in the second largest Nama-Karoo Biome (Low & Rebelo 1996) characterized by an arid climate with seasonal temperatures higher than 30°C and low rainfall (range of 100 mm to 500 mm; Desmet & Cowling 1999). The study also presented a genetic distribution and structure of village goat ecotypes that has not been reported before. Almost all village goat ecotypes were admixed which could be a result of the village goat production system that is characterized by communal breeding leading to high genetic diversity and gene flow amongst villages and populations in close proximity. The Tswana shared a genetic background with the Xhosa goats and the Zulu with Venda goats. These results suggested weak impact of latitude on the genetic structure of ecotype populations and are in line with the clustering based on morphological characteristics (Chapter 3). The detected clusters revealed geographically structured populations biased towards longitude. The Zulu and Venda goats fall in the 30° longitude whilst the Tswana and Xhosa cluster is spread between 26° and 28° longitude. Commercial breeds with different population genetic components showed a wide geographic spread. In spite of the complex admixture observed among the commercial breeds, as indicated by the admixture ($K = 5$), Kalahari red and Savanna from Northern Cape (-29°03’13”S; 23°18’26”E) have retained unique identities and are differentiated from individuals from other geographic areas. The Kalahari Red and Savanna goats from this sampling site had a low percentage of admixed individuals. The low percentage of admixture among the populations was presumably associated with the low gene flow from the neighboring regions. This is attributed to the geographical isolation, and the unique founder population for the Kalahari Red and Savanna goats. The Kalahari Red cluster with a small genetic spread restricted to the Northern Cape is said to have been developed from the local indigenous red goats while the Savanna goats were from the bloodline originally developed by Messers Cilliers and Sons (Personal communication with Mr. Kobus Botma, a farmer and breeder in Northern Cape). Majority of the Boer, and Savanna (from the geographical locations) also shared a large proportion of ancestry with the unique population of these Kalahari Red goats, which supports the documented breed development from indigenous populations (APPENDIX D). These findings suggest that these populations have not been separated long enough to afford genetic

differentiation. Few individuals of the Boer (located at -26°15’16’’S; 26°16’23” E) displayed a low level of admixture.

It is hypothesized that the harsh and heterogeneous environmental parameters have facilitated adaptive genetic variance that presents an integral component to the survival of natural populations, especially in the face of foreseen climate changes (McGaughran *et al.* 2014). Climatic variables such as temperature and rainfall have been shown to directly and indirectly affect livestock production. Evidence of adaptation to the different environmental regimes such as temperature, rainfall, and altitude has been suggested in goats (Lv *et al.* 2014; Kim *et al.* 2016; Song *et al.* 2016). These production landscape features are also important predictors of quality and quantity of pastures (Kirkman & de Faccio Carvalho 2003) ultimately influencing productivity of goats (Butswaat 1994) and of vector-borne diseases such as heartwater especially in low input and extensively farmers livestock. South Africa, for example is heartwater endemic, a tick-borne disease that is known to exhibit geographic restrictions because of the distribution of the disease vector (*Ambylomma hebraeum*). Although adults of this tick are present throughout the year, high incidences have been reported during wet summer months in some parts of the country (Spickett *et al.* 2011). Disease outbreak statistics from the South African National Department of Agriculture Disease database suggest that the disease burden is more pronounced during the summer periods. Wet seasons are also associated with high gastrointestinal parasites infection rates (Nwosu *et al.* 2007; Mbuh *et al.* 2008; Gwaze *et al.* 2009c; Mubi *et al.* 2012). Based on these known dynamics, this study further investigated the contribution of climatic and geographic variables to the explainable variation and the outlier SNPs associated with the annual and seasonal trends of temperature and rainfall in the different geographic regions.

Variance partitioning using the two partial RDA models (Model II and III) indicated a higher contribution of climate variables (9.0%) than the geographic variables (1.0%) to the total genetic variation. The positive association between climatic variables and genetic variation suggested the importance of rainfall, temperature and other climatic variables in shaping population genetic structures. To our knowledge, there are no other studies that have partitioned the genetic variance among goat populations or indigenous livestock populations into their environmental and geographic factors. As such, there is a dearth of data as well as studies available for comparison. In other non-livestock species, genetic variation observed was found to be highly associated with climatic variables than geographical variables and

examples are given for *Arabidopsis thaliana* (Lasky *et al.* 2012), *Pristionchus pacificus* (McGaughran *et al.* 2014) and *Hordeum vulgare* L. (Abebe *et al.* 2015). The RDA full model (Model I) showed that altitude, annual minimum temperature, winter minimum temperature, and longitude to have significant effects on genetic variation contrary to the partitioning of variance (Model II) whereby only annual maximum temperature, winter maximum temperature and altitude were the main contributors to the detected genetic variation. Annual temperature ranges were observed as the first significant explanatory variables in both models, followed by winter temperature ranges indicating the importance of temperature as an environmental selective pressure in these goats.

The second step in this analysis was to further investigate the genetic basis of local adaptation to temperature, rainfall, altitude, latitude and longitude in indigenous goat populations. Two correlation-based landscape association genomic approaches namely SAM (Joost *et al.* 2007) and LFMM (Frichot *et al.* 2013) and an outlier method using RDA axes (Hancock *et al.* 2012; Lasky *et al.* 2012) were used. The three association methods reported several significant loci associated with climatic variables. The association of climatic variables with SNP markers using SAM returned SNPs associated with longitude, annual minimum temperature and altitude as the important factors that affect selection pressure, while LFMM returned significant loci in relation to 11 of the 12 variables used. The SAM analysis revealed that the highest number of SNPs were associated with longitude. The population structure showed different genetic patterns of divergence among the goat populations is associated with longitude. In addition, RDA analysis (Model I) indicated that annual and winter temperatures, altitude and longitude were the most important variables suggest that adaptation to local climatic variables, and particularly temperature stress and geographic location (altitude and longitude), has shaped genomic variation. SAM detected a SNP (snp25341-scaffold261-1157571) significantly associated with longitude, winter minimum and annual minimum temperature variables, which had a high squared score (99% CI = 0.19) in the first axis for the RDA analysis. When population structure was considered in the LFMM analysis, warm temperatures (Stmin, Stmax, and Atmax) and rainfall (annual and seasonal ranges) had the highest number of SNPs associated, suggesting the differences among the geographic regions as an important selection pressure. Most of the loci detected using LFMM were associated with annual and seasonal temperature and rainfall variables followed by latitude, indicating the importance of these variables in driving local adaptation. Strong associations with rainfall variables were also observed, contrary to SAM, which did not return any associations with

rainfall variables. There was more common outlier loci observed between RDA and SAM than LFMM suggestive of the important role of environmental differences on genetic variation.

Goats have been found to be generally more adaptable of all domesticated livestock species with proven ability to thrive in low input production systems and environmental conditions particularly those prevailing in the agro-ecological regions across Africa (Yakubu *et al.* 2010a; Ojo *et al.* 2015). Indigenous goats have been observed to be adapted or tolerant to yearly and seasonal fluctuations of water and feed availability, and survive starvation periods during drought as well as extremely high temperatures and exposure to pests and diseases (Baker & Gray 2004; Daramola & Adeloye 2009). Identifying the genomic regions targeted by these natural selection pressures associated with goats in the tropics will help in identifying functionally important genes with a potential roles in local adaptation and optimal production particularly in the face of climate change induced selection pressures. Few studies have investigated signatures of selection in response to climatic variation in livestock species including goats (Lv *et al.* 2014; Kim *et al.* 2016; Song *et al.* 2016) differing to the current study by the various aspects (i.e. livestock species, analytic approach etc.). Identification of SNPs associated with climatic variables, altitude and geographic location enable identification of alleles associated with environmental stress thus could contribute to the understanding of the genetic basis of response to climatic change in indigenous goat populations. Kim *et al.* (2016) found eight candidate sweep regions, using the integrated Haplotype Score (*iHS*) approach, distributed across chromosomes 3, 6, 7, 11, 12, 14 and 17 in Barki goats found to be associated to response to hot arid environment. Similarly, outlier SNPs were identified spanning these chromosomes associated to low rainfall and high temperature ranges. Song *et al.* (2016) emphasized the role of altitude in creating and maintaining adaptive genetic differentiation in Tibetan cashmere goats inhabiting high altitudes. However, our study does not have representation of high altitude ranges known to affect oxygen as all altitudes were below 3000m.

Gene annotation and pathway analysis of associated genes reported metabolism and responses to heat, water scarcity and disease that allow indigenous goats to tolerate environmental pressure in their local production systems. However, adaptation is a complex trait (Scholtz *et al.* 2013; Lv *et al.* 2014; Kim *et al.* 2016) involving multiple genes acting together; therefore, it was not surprising that a large number of genes were identified. Some of the genes may

have major effects while other may have minor effects, complicating efforts aimed at identifying the genetic basis of such a trait. Numerous candidate genes involved in metabolism were identified in the present study including protein digestion and absorption, amino acid metabolism processes, secretion and taste transduction (*ASIC2*). An important attribute that goats possess for surviving and producing in semi-arid and arid areas is their ability to utilize low-grade roughage (Silanikove *et al.* 1993; Daramola & Adeloye 2009) due to an efficient system of digesting fiber to enable maximal food intake and utilization and high tolerance for bitter substances (Goatcher & Church 1970; Casey & Van Niekerk 1988; Silanikove *et al.* 1996; Silanikove 2000).

Given that South African goats occur in arid and semi-arid areas characterized by environments of high temperature, water-scarcity and frequent droughts, the overwhelming evidence for natural selection in our study is therefore not surprising. Indigenous goats also have the capacity to withstand prolonged periods of water deprivation (Silanikove *et al.* 1994) because of their ability to withstand and minimize water losses through urine and faeces (Daramola & Adeloye 2009), a mechanism to conserve water in times of heat and drought. The rumen, the salivary glands and the kidneys function in a coordinated way during dehydration and rapid rehydration because of a large ruminal volume, better capacity of the kidney to desalt the water absorbed from the gut and the maintenance of a salivary flow to the rumen (Silanikove 2000; Daramola & Adeloye 2009). In this study, genes (*NSF* and *RAB5A*) involved in vasopressin-regulated water re-absorption (*NSF* and *RAB5A*) and salivary secretion (*ITPR2*, *ADRAID*, *RYR3*, *PBL1*, and *KCNMA1*) were observed as under selection in these goat populations.

Heat stress is an important selection pressure in indigenous goats in response to environment stressors that mainly stem from high temperatures. Goats have been reported to be superior in managing heat stress due to their thermoregulatory response mechanisms. While coat color and patterns may indicate adaptability to harsh environments (Daramola & Adeloye 2009), the coat induced thermophysiological mechanisms to cope with heat stress are not well documented for South African indigenous goats. Several genes associated with the heat stress response pathways, including circadian entrainment and circadian rhythm, were observed to be under selection in this study. The *PLCB1* gene involved in numerous pathways was also reported by (Kim *et al.* 2016) for thermoregulation in hot arid environment.

Our findings also indicated that climate can also affect autoimmune regulation. Response to disease and disease pathogens is important in developing countries where populations are frequently exposed to various endemic diseases as well as disease outbreaks. Indigenous breeds usually display enhanced resistance to endemic diseases as compared to exotic ones reared in the same environment (Baker & Rege 1994). Evidence has accumulated over the last few years indicating that a number of indigenous goat breeds have natural resistance or tolerance to specific diseases (Malan 2000; Agyemang 2005; Ahmed & Othman 2006; Morrison 2007). The relative tolerance of the South African goats to prevalent heartwater in these populations was established in Chapter 4. Genes involved in response to diseases included *PCLB1*, *IDO2*, *GNA15*, *ARHGEF12*, *NCAM1*, *C8A*, *C8B*, *JAK2*, *BIRC2*, *RAB5A* and *MAPK10* and were found to be under selection in the study populations, suggestive of the importance of immune response and disease resistance/tolerance as a survival mechanism in the South African goat population.

Three statistical methods (SAM, LFMM, and RDA analysis) were used as a strategy to offer complementary comparisons between the different detection methods. A noteworthy observation was the differences between the association methods whereby few SNPs were common among the different methods. SAM and RDA analysis had 7 common outlier SNPs, whilst 5 SNPs were common between SAM and LFMM. However, among the genes detected, two (*GRID2* and *MARCH1*) have strong support as candidate genes after detection by the three approaches. A number of SNPs were detected by only one or two methods, which could be due to different outlier detection methods being differentially sensitive to data input and as such might fail to pick other loci in specific genomic regions. Redundancy analysis, for example, has a limitation of only making use of animals with a complete set of SNP data. As such, pathways such as circadian rhythm and circadian entrainment pathways could not therefore be picked by the redundancy method as some SNPs had low call rates (98%). Therefore, different outlier and association analyses would complement each other in determining loci under selection.

This study used a combination of multivariate statistics, landscape genomics and population genomics based analyses in order to increase accuracy and draw from the multidimensional effects of environmental factors in shaping the genetic variation in local indigenous goat populations. Using multiple methods of population and landscape analyses, the study provided evidence of genetic subdivision within the village ecotype populations and

demonstrated the presence of genetic subpopulations associated with specific geographic localities. Overall, the study provided insight into the primary selection pressures driving evolution and local adaptation of South African goats to important prevailing production conditions. The observed association between the environmental gradients and genetic differentiations confirm the hypothesis that the production environment presents natural selection pressures that shape the genetic divergence of goat populations. However, further experiments involving targeted gene sequencing and expression are necessary to confirm the causal mutations and the precise role of candidate genes in the process of local adaptation.

CHAPTER 7: GENERAL DISCUSSION, CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

7.1 General discussion

South Africa has a dual goat production industry that consists of a well-developed commercial sector and a predominantly subsistence farming in the communal areas. Majority of goats in South Africa are extensively managed in rural settings that exhibit considerable conformational diversity in coat color patterns and body sizes. In general, Eastern Cape, KwaZulu-Natal and Limpopo have the highest proportion of the country's small-scale farmers in the former homelands. Characteristic of small-scale farming is extensive production with low or limited management input. Livestock are forced to deal directly with environmental factors and goats are important adapters to such an environment. The adaptive potential and broad and heterogeneous distribution across the country make indigenous goats an interesting livestock species for studying the effects of natural selection pressures on adaptive genetic variation. Past attempts to improve goat production in South Africa focused on introduction of highly productive exotic breeds that require high level of management and inputs. These efforts were least successful thus creating a need to design proper breed improvement programs in order to enhance the utilization and conservation of the genetic diversity of the local goat populations.

Together with diverse agro-ecological and climatic features, South Africa experiences vast temperature and rainfall differences which leads to the hypothesis that these production landscapes might be the major explanation of the highly diversified goat populations distributed across the country. Landscape genomics, through modeling the multi-dimensional association between the environment and genetic variation can capture selective gradients that are combinations of multiple climate variables (Whittaker *et al.* 1973) and their effect on allele frequencies across numerous loci (Lasky *et al.* 2012). Genomic scans for adaptive genetic variation in goats have in the past been proven to be an effective approach to associate genetic variation with environmental factors using few SNP markers on specific genes (Pariset *et al.* 2009) and AFLP markers (Colli *et al.* 2014). However, advances in next generation sequencing technologies have resulted in high-throughput genotyping tools such as the Illumina Goat 50K SNP BeadChip with thousands of single nucleotide polymorphisms (SNPs) across the genome. Genome-wide SNP data can then be effectively used in

environmental association analyses to identify putatively adaptive genetic variation that displays exceptionally strong associations with the environment. This study was aimed at detecting the genetic signatures of local adaptation in South Africa indigenous goat populations using the landscape genomics approach. The study reports the first insight into the identification of adaptive loci by combining medium density SNP data of indigenous breed/populations with highly divergent climatic variables for different agro-ecological zones. The detection of these signatures of local adaptation will help in understanding the mechanisms of adaptation in the South African climate.

The patterns of morphological diversities and geographical distribution of these ecotype populations at community (village) level have not been fully investigated and remains unknown. This sampling strategy adopted in this study was designed to ensure representation of the different production landscapes in the country. However, limitations of a sampling scheme restricted to the specific districts, breeds, and production systems cannot be ignored. The sampling structure of the study considered populations based on geographic distribution, and targeted major goat producing provinces of the country (Eastern Cape, KwaZulu-Natal, Limpopo, Northern Cape and North West). These goat populations represented the extensively raised village goats, commercially developed breeds (Boer, Kalahari Red and Savanna) as well as the Tankwa feral goat from the Carnarvon Research station in the Northern Cape. Due to financial limitations, the sampling did not include dairy and mohair breeds that are exotic breeds raised predominantly by commercial farmers in South Africa. We speculate that by not including these breeds, their genetic relationship to village goats cannot be entangled in order to understand fully the effect of random crossbreeding with such breeds. The study however, targeted the major goat populations that include the village goat production sector comprising more than half of the goat population and the billion rand worth commercial goat meat sector. Majority of goats are known to be distributed in arid and semi-arid parts of the country, and the study purposively sampled from the major goat producing provinces of the Eastern Cape, Limpopo, KwaZulu-Natal, Northern Cape and North West.

The first experiment described the production landscape of South African indigenous goat populations by investigating the characteristics of the predominant village production systems using a questionnaire survey. Results from this study suggest the need for local adaptation for goat populations to survive and thrive under harsh conditions characteristic of these production systems. It was also observed that the number of breeding individuals per

household was low, with most households not owning bucks in KwaZulu-Natal and Limpopo. This highlights the importance for communities to prioritize saving breeding males, which contribute immensely towards achieving high genetic diversity, and reducing inbreeding levels. However, this depending on how many males are used for breeding and their pedigree. In addition, the study explored morphological variables of goat populations found in these communities. Phenotypic characteristics are important in breed identification and classification and are easily used by village farmers compared to genetic and other forms of characterization. The mode and level of inheritance of these traits has not been studied in South African goats but general conjecture links them to environment adaptation. Studies that model morphological characteristics to reflect history, ecological and human selection are important in understanding the impact of production factors on the phenotypic and genetic diversity of livestock populations. This study hypothesized that differences in the origin of SA goat populations coupled with ecological, geographical and management practices might have resulted in the diversity observed in the indigenous goat populations of South Africa. Geographically structured phenotypic variation is common in indigenous goats (Yakubu *et al.* 2010a) and has been considered key to local adaptation. The diversity patterns of qualitative traits and quantitative phenotypic characteristics in the non-descript village goat populations observed in the current study is evidence for the existence of genetic variability among the studied populations.

Linear body measurements variability in goats arises due to genotypic and environmental effects, and the magnitude of variability may differ under different management practices and environmental conditions. However, the step-wise discriminatory procedure showed that four measurements (CG, PB, CW and H) were significantly important implying that they could be more important in differentiating between the four non-descript village goat populations, commercial meat type breeds and Tankwa goats. Further, the canonical discriminant analysis suggested that the chest girth had the highest significant discriminative potential for the populations. Such differences between the types maybe related to environmental adaptation. The closeness between Venda and Zulu ecotypes compared to their Tswana and Xhosa counterparts might be as a result of body size, which may function as a guide to genetic and evolutionary relationships among ecotypes of South Africa. A significant morphological variation was also observed using the Principal Component Analysis based on the five morphometric variables. The principal component analysis and discriminatory analysis of body measurements showed high numbers of individuals of the Boer, Kalahari Red, Savanna

and Venda goats are not differentiated but are admixed. However, three overlapping clusters were observed: 1) the commercial meat type breeds (Boer, Kalahari Red, and Savanna); 2) the Zulu and Venda village ecotype populations, that inhabit KwaZulu-Natal and Limpopo provinces, respectively; and 3) the Tswana and Xhosa village ecotype populations of the North West and Eastern Cape, respectively and the Tankwa goat. Zulu, Tswana, Xhosa and Tankwa showed a low level of admixture in compared to the other population. The overlapping body measurements among individuals of different breeds/populations could also be a factor leading to misclassification of some animals. The present pattern of morphometric variation among the non-descript village goats indicated a geographical trend of population differentiation. We speculated that the village ecotype populations from KwaZulu-Natal and Limpopo belonged to the small-framed Nguni (iMbuzi) ecotype found in heartwater-endemic region thus the low level of differentiation and high level of misclassification between these populations. The high degree of correct classification of individuals in their respective genetic source breed/ecotype confirmed the high discriminating power of the quantitative characteristics especially chest girth in explaining morphological variability.

Ticks have specific environmental requirements for survival and reproduction, such as temperature, humidity, vegetation and host availability. The distribution and prevalence of vectors and thus vector-borne diseases is linked to environmental variables. The study hypothesized that there are significant regional differences in heartwater prevalence among the geographic regions. The heterogeneity in climate and habitat types affect the livestock population genetic structure especially when the disease is endemic to the region. The study used heartwater as a model disease due to distribution dynamics between endemic and non-endemic regions, seasonal role in infection dynamics such as rainfall, temperature and vegetation in the disease burden. Further, heartwater was reported to be one of the major disease challenge in goat production in the sampled communities. Therefore, the study investigated whether the variations in the seroprevalence among the study areas could be related to management differences in the different production systems, regions and the associated risk factors. The risk assessment was based on factors useful to predict disease incidence and prevalence such as geographical location, sex, age, tick infestation, breed/population, management system, and endemic status of the animal origin. However, the seroprevalence considered in our study were limited to detect the effects of zonal differences in rainfall, temperature, relative humidity and vegetation condition. Overall, risk of heartwater seroprevalence was linked to non-descript village goats under extensive smallholder

production system and infestation with ticks. These findings suggest an existence of distinct differential tick control management strategies across the production systems and the high seroprevalence highlights the animal health importance of heartwater in the smallholder systems, which often have limited resources to control the disease. It is also speculated that the overrepresentation of does and adult goats (Chapter 3) maintained in the sampled communities could also contribute to the higher seroprevalence in does and aldult goats. This result, together with those reported in Chapter 3 suggest that heartwater seroprevalence could be due to natural exposure and that goat populations in communal areas have evolved unique adaptation to thrive even under disease challenges. Therefore, differences in management practices (coupled with production constraints) and geographical regions are likely to shape the genetic diversity and adaptation of goats from diverse production systems and agro-ecological environments in which they are found. However, a proper determination of tick burden (numbers per animal) and identification of the tick species as well as would have provided additional important information.

As a follow up to Chapters 3 and 4, the third experiment genotyped 239 South African indigenous goats consisting of the 4 registered breeds (Boer, Kalahari Red, Savanna, and Veld Indigenous) and a feral Tankwa breed, from different geographic regions and production systems. These populations represent major goat breeds with the largest share in the meat and smallholder production systems targeting the major goat producing South African provinces. They were genotyped using the Illumina Goat 50K SNP BeadChip assay to investigate the level of their genetic diversity and population structure. The genetic diversity was higher in the non-descript village ecotype populations whilst, lower in the Tankwa goats due to differences in their population dynamics. The village ecotype populations are kept under low management systems with no controlled breeding, leading to high genetic diversity and gene flow from other areas all contributing to an admixed and highly diverse goat population. The population structure (admixture-based and principal component analysis) analysis revealed three major clusters assigned to their respective management systems. However, the pattern of clustering was slightly different compared to previous South African goat genetic studies (Kotze *et al.* 2004; Visser *et al.* 2004; Pieters 2007; Kotze *et al.* 2014) mainly because of the differences in sampling strategy incorporating breed representation (commercial, village and feral populations) and the number of genetic markers used. The analysis showed uniqueness of the Tankwa breed, with some typical features such as high level of differentiation from the domestic indigenous goat populations and a reduced intra-population genetic diversity (lowest

H_O = 0.35 ± 0.33; H_E = 0.33 ± 0.16). These factors have major implications on the conservation and management of Tankwa. Furthermore, the results are comparable to the estimates obtained using microsatellite data between Tankwa and Boer goats, which revealed high genetic differentiation for the relocated goats (F_{ST} = 0.162) and for original goats from the Tankwa Karoo National Park (F_{ST} = 0.235; Kotzé *et al.* 2014). Although variation in the ecotype populations was expected due to the extensive management system, it was not expected in the commercial breeds because they have been selected for specific traits. It is speculated that the high variability among individuals in the commercial breeds could have arisen by high levels of gene flow within the farms, and common breeding goals (i.e. hardiness). At a fine scale, the village ecotype populations could be further separated into two clusters, one comprising the Tswana and Xhosa and the other, Zulu and Venda. The pattern of genetic clustering slightly differed from the morphological clustering (Chapter 3). This may be related to varying sensitivity of data inputs, although the analysis was able to separate the commercial breeds from village populations. Although revealing fine-scale genetic structure, the village ecotype populations presented a certain level of genetic admixture, while the Tankwa population showed higher differentiation from the others.

The high levels of genetic differentiation (pairwise F_{ST} comparisons and principal component analysis) of Tankwa from other breeds and populations are suggestive of a unique genetic resource with unique alleles reflecting the presence of certain functional genes that may be possibly related to better adaptability to its agro-ecological zone and feral production system. In a recent study, using complete mtDNA, three clearly differentiated maternal lineages were observed in the South African goat population (Ncube 2016). Maternal lineages A, B and G were found in all the breeds analysed with different types predominating in each population. Using the *COX1* gene, Ncube (2016) suggested that the Tankwa and domestic counterparts had a common maternal founder population, which is supported by some animals sharing similar genetic components in the ADMIXTURE analysis. Ncube (2016) and this study further support the breed development of the Boer and other commercial meat type breeds from indigenous populations. Due to the low level of differentiation between the commercial breeds and village ecotypes, no outlier loci were reported. However, outlier loci were observed between the Tankwa and the rest of the domestic goat populations, which are potentially under selection in the Tankwa goat. In the Tankwa breed, SNPs with significantly higher F_{ST} values (> 0.7) were associated with *FAM71E1*, *ADCY9*, *USP37*, *SFT2D2* and *CHD1* genes suggesting the importance of the Karoo desert environment as a selection

pressure in the Tankwa. For an example, the Adenylate Cyclase 9 (*ADCY9*) is associated with numerous pathways such as circadian entrainment, vasopressin-regulated water reabsorption, salivary secretion pathways involved in responses to heat stress including in thermoregulation and water balance

The mean linkage disequilibrium (r^2) obtained between adjacent SNPs across all autosomes was lowest (ranging between 0.09±0.12 and 0.11±0.14) in village ecotypes, intermediate (0.15±0.18 to 0.18±0.20) for commercial breeds and Nguni (0.16±0.18), and highest in the Tankwa (0.25±0.26). Differences in the average r^2 and its decline in these populations can be explained by differences in effective population size (N_e) and breeding practices. The estimated N_e 13 generations ago was lowest for the Tankwa (41 individuals), followed by Nguni (46 individuals), Tswana (93 individuals), and highest (> 100) for the other breeds/populations. The decline in effective population size in the Tankwa could be due to a limited pool of the founder population at the Carnarvon research station. The decline in *Ne* in the commercial breeds is a reflection of the historical process of breed formation. In addition, the knowledge about breed- and chromosome-specific variation in LD is important in defining the optimal marker density in genome-wide association (GWAS) and genome selection (GS) studies. This could be used to design customized marker panels that with the desired LD levels across the genome. The estimated LD decreased rapidly with increasing distance between SNPs for the breed pairs. The useful linkage for GWAS and GS ($r^2 > 0.2$) had different patterns for each breed pair. Among the commercial breeds, linkage was found to span from 200Kb. The linkage for the village ecotypes was found to span from 10Kb and ~20Kb in the Tankwa breed and ~10Kb. Thus panels containing about 10 000 SNP markers are necessary for GWAS and GS studies between breeds of these pairs. The results of this analysis suggested higher between-breed genetic variation relative to within-breed diversity. This variation could be a valuable tool for genetic improvement and conservation.

Having described the production landscape (Chapter 3 and 4) and ascertained the genetic diversity and population structure of South African indigenous goat populations (Chapter 5), Chapter 6 investigated the role of various geographic and climate features in shaping patterns of genetic diversity. The adaptive potential of indigenous goat populations may have major economic and ecological consequences. Although they are considered to be highly adapted to climate, poor nutrition and other production challenges, uncontrolled inbreeding and unplanned crossbreeding may threaten the ability of these populations to cope with

environmental pressures. This study aimed to explicitly test the hypothesis that genetic diversity and population structure of South African indigenous goat populations is influenced by environmental variables that vary across geographic regions. To test this, landscape genomic analysis using the Redundancy analysis (RDA) statistical tools was used to investigate the contribution of geographic and climatic data to goat genetic variation. Research questions addressed included (i) whether there was any directionality in the population genetic structure along geographical locations, and (ii) what fraction of genetic variation was explained by the environmental and geographic variables?

The ADMIXTURE membership coefficients were overlaid on the South African geographic map, and a geographically-based population structure was revealed. Specifically, among the inferred population structure, the Tankwa breed genomic signature was associated with one geographic region of the Northern Cape with homogenous genetic backgrounds. The minimal level of admixture observed among the Tankwa could be associated with the low gene flow from other regions and between domestic and feral goat populations. This is attributed to their location (at a conservation unit on a research station) and the environmental landscape of the region (the Karoo desert). These findings suggest that this breed should be closely supervised and preservation strategies should be implemented to avoid further inbreeding. Results of the ecotype populations suggested a more complex substructure within these populations due to high admixture levels. The geographic distribution of the these components followed a longitudinal pattern where the Tswana and Xhosa goats were located between 26°- 28° E and the Zulu, Venda and Nguni goats located at 30° E. The population structure for the ecotypes revealed that they shared some genetic components reflecting common origin and similar rearing management. Commercial breeds shared genetic components and harbor different fractions of ancestries suggestive of a common origin of the populations and for some, a long period of genetic isolation. The variance models indicated that the variation contributed by climate variables, were higher than the variation introduced by geographic variables, thus suggesting an important influence of climate diversity in shaping genetic variation. We further used Spatial Analysis Methods (SAM), Latent Factor Mixed Models (LFMM) and the outlier detection method using first axis of RDA to investigate the association of SNP markers with climatic and geographic variables. The combined use of multiple landscape genome scan methods offered an advantageous strategy to obtain a more reliable identification of loci under selection. Altitude, annual minimum temperature, winter minimum temperature and longitude were shown to be the main contributors of the detected genetic variation for RDA Model I

and for the detected adaptive SNPs using SAM. When population structure was taken into consideration as in the LFMM, most significant SNPs were associated with summer and annual ranges of both temperature and rainfall, thus indicating the importance of these variables in determining local adaptation. It is therefore proposed that these variables are the selective gradients related to local adaptation of South African indigenous goats.

Despite the long history of the application of landscape genomic tools in goats, this is the first study, to our knowledge, to use genome-wide SNP data in goats. Previous studies used low-density markers and although important correlations with environmental factors were observed, this study is unique as it sampled different breeds representing different productions landscapes. The study captured seasonal trends of rainfall and temperature important in feed availability for grazing animals as well as prevalence of parasites and diseases. Alternative sequencing platforms to genotyping single nucleotide polymorphisms (SNPs), such as whole genome sequencing (WGS), whole genome re-sequencing and transcriptome sequencing (TS) can have important contributions to the study of adaptive genetic mechanism by giving more comprehensive arrays of regions under selection and should be considered in future studies.

7.2 Conclusion

The observed association of genetic differentiation with production system, and environmental gradients may suggest a role for management and natural selection in shaping the divergence in indigenous goat populations. The study highlights the need for the integration of environmental, geographical and genetic data in genome-wide scans for loci under natural selection as well as the promise of combining complementary approaches for the identification of functionally important genetic variation. The study identified climatic and geographic variables as important explanatory aspects of genetic variation. The detected candidate loci that showed selective response to important climatic variables and geographical location could be associated with local adaptation. In general, this study successfully demonstrated how landscape genomics contributes to uncover the loci under selection that potentially influence adaptation in indigenous goats in South Africa.

7.3 Study limitations and Challenges

The number of respondents and goat numbers in Chapter 3 was lower than the minimal threshold recommended for statistical analysis. According to Census (2011), human population density is 6.2, 10.2, 3.5, and 5.4 million for Eastern Cape, KwaZulu-Natal, North West and Limpopo, respectively. At 5% margin error and 95% CI the number of respondent should be 385 per province. However, this number of respondents was not achievable due to cost. KwaZulu-Natal and Limpopo had the highest number of goat farmers, whilst Eastern Cape and North West had the least number. Recruitment of households for interviews was biased towards those that were part of the NGO in KwaZulu-Natal, those registered as a goat farmer with the provincial Department of Agriculture extension service in North West, Eastern Cape and Limpopo. The reception was also different, with other farmers not willing to participate in the study for various reasons. In addition, goat population densities of goat farmers within the villages differed. Further, the presence of large variation is possibly the influenced by the number of households and the large variation in the number goats per the household or farm across the geographical regions resulted in variably larger standard deviation. Variation in the number of goats kept per household was large ranging from 2 to 62 in the North West, 1 to 60 in Limpopo, 5 to 200 in the Eastern Cape, and 4 to 200 in KwaZulu-Natal affecting the normal distribution curve thus led to the high standard deviation (SD) values observed.

A wide variety of data types are useful for a comprehensive landscape genomics analyses. Geographical, socio-economic and socio-demographic data, animal husbandry practices within production systems, animal populations, and types of prevalence of other diseases e.g. endo- and ecto-parasites, animal survival rates etc. and data on environmental and genetic parameters are such examples. The greatest challenge for this study was the complex sampling strategy whereby individual goat samples were collected from heterogeneous geographical landscapes and diverse environments in order to unravel the comparative adaptive fitness on the animal genetic resources. Collection of some data types such as vegetation, soil and genotyping data from other breeds was not feasible within the scope of the study. The study findings are therefore limited to predominantly geographic and environmental data and a limited set of data of production systems and disease challenges such as heartwater. Collecting the comprehensive data required for landscape genomics increases the cost of the research, which is sometimes limiting particularly in developing countries.

7.4 Future possibilities

Future attention should be given to spatio-temporal variation within the different geographical areas for instance by undertaking a comprehensive study including indigenous goat populations from the agro-ecological zones not covered by the present study such as the subtropical winter rain agro-ecological zone and possibly other African countries. There are many complementing approaches for evaluating disease prevalence and distribution that differ in terms of accuracy. Therefore, future studies should consider using combinations of seroprevalence with phylogenetic representation to provide a more advanced understanding of the potential historical relationships and the spatial genetic variation of the different animals from the geographic regions. In turn, this could be coupled with tick prevalence and taxonomy data of the different regions and further unravel distribution of tick-borne diseases within the distribution range of their vector.

It is also clear that future studies will require more in-depth analyses of environmental factors, epidemiology mapping of other diseases and not be limited to a single disease as in our case with heartwater. Diseases, among those reviewed by Mohlatlole *et al.* (2016), include gastrointestinal (GI) infections caused by nematodes like *Haemonchus contortus* and the diseases reported by the farmers outlined in Chapter 3.

It is also possible that some important genome regions might not have been identified in the present study due to inherent limitations of the SNP50K panel. Therefore, further studies could use higher density SNP arrays and/or whole genome and transcriptome sequences specific for South African goats to address these issues to some extent.

REFERENCES

Abebe T.D., Naz A.A. & J L. (2015) Landscape genomics reveal signatures of local adaptation in barley (*Hordeum vulgare* L.). *Frontiers in Plant Science* **6**, 1-17.

Adebayo J. & Chineke C. (2011) Evaluation of West African dwarf goat for some qualitative traits in Southwestern Nigeria. *African Journal of Agricultural Research* **6**, 6204-6207.

Adedeji T.A., Ozoje M.O., Peters S.O., Sanusi A., Ojedapo L.O. & Ige A.O. (2011) Coat pigmentation and Wattle genes effect on some haematological characteristics of heat stressed and extensively reared West African Dwarf goats. *World Journal of Life Sciences and Medical Research* **3**, 48-55.

Aganga A. & Mosimanyana N. (2001) Gender impact on sheep and goat production in Botswana. A case of Gaborone region. *Journal of Agriculture in the Tropics and Subtropics* **102**, 15-18.

Agyemang K. (2005) Trypanotolerant livestock in the context of trypanosomiasis intervention strategies. In: *PAAT Technical and Scientific*. FAO, Rome, Rome, Italy.

Ahmadu B., Lovelace C. & Mahan S. (2004) Some observations on the sero-prevalence of heartwater and tick infestation in Zambian goats: research communication. *Onderstepoort Journal of Veterinary Research* **71**, 161-164.

Ahmed S. & Othman O.E. (2006) A PCR-RFLP Methods for the analysis of Agyptian Goat MHC Class II DRB Gene. *Biotechnology* **5**, 58-61.

Ai H., Huang L. & Ren J. (2013) Genetic diversity, linkage disequilibrium and selection signatures in Chinese and Western pigs revealed by genome-wide SNP markers. *PloS One* **8**, e56001.

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

Allsopp M., Visser E., Du Plessis J., Vogel S. & Allsopp B. (1997) Different organisms associated with heartwater as shown by analysis of 16S ribosomal RNA gene sequences. *Veterinary Parasitology* **71**, 283-300.

Almeida A., Schwalbach L., De Waal H., Greyling J. & Cardoso L. (2006) The effect of supplementation on productive performance of Boer goat bucks fed winter veld hay. *Tropical Animal Health and Production* **38**, 443-449.

Anaeto M., Tayo G., Chioma G., Ajao A. & Peters T. (2009) Health and nutrition practices among smallholder sheep and goat farmers in Ogun State Nigeria. *Livestock Research for Rural Development* **21**, article 97.

Arias J.A., Keehan M., Fisher P., Coppieters W. & Spelman R. (2009) A high density linkage map of the bovine genome. *BMC Genetics* **10**, 18.

Awa D.N. (1997) Serological survey of heartwater relative to the distribution of the vector *Amblyomma variegatum* and other tick species in north Cameroon. *Veterinary Parasitology* **68**, 165-173.

Bakare A.G. & Chimonyo M. (2011) Seasonal variation in time spent foraging by indigenous goat genotypes in a semi-arid rangeland in South Africa. *Livestock Science* **135**, 251-256.

Baker R. & Gray G. (2004) Appropriate breeds and breeding schemes for sheep and goats in the tropics. *Worm control for small ruminants in tropical Asia*, 63-188.

Baker R.L. & Rege J.E.O. (1994) Genetic resistance to diseases and other stresses in improvement of ruminant livestock. In: *5th World Congress on Genetics Applied to Livestock Production*, pp. 405-12, University of Guelph, Ontario, Canada.

Bambou J., Vachiery N., Despois P., Giraud-Girard K., Arquet R., Pinarello V., Aprelon R., Barbier C., Gobardham J. & Mandonnet N. (2010) Assessment of genetic variability of resistance to heartwater in Creole goats. *Advances in Animal Biosciences* **1**, 408-409.

Barbato M., Orozco-terWengel P., Tapio M. & Bruford M.W. (2015) SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in genetics* **6**, 109.

Barry D. & Godke R. (2001) The Boer goat: The potential for cross breeding. *Louisiana State University, Baton Rouge, USA dnafrica@ worldonline. co. za*.

Bath G., Van Wyk J. & Pettey K. (2005) Control measures for some important and unusual goat diseases in southern Africa. *Small Ruminant Research* **60**, 127-140.

Bekker C.P., Vink D., Pereira C.M.L., Wapenaar W., Langa A. & Jongejan F. (2001) Heartwater (*Cowdria ruminantium* Infection) as a Cause of Postrestocking Mortality of Goats in Mozambique. *Clinical and Diagnostic Laboratory Immunology* **8**, 843-846.

Benhin J.K. (2006) Climate change and South African agriculture: Impacts and adaptation options. CEEPA discussion paper.

Bester J., Ramsay K.A. & Scholtz M.M. (2009) Goat farming in South Africa: findings of a national livestock survey. *Applied Animal Husbandry & Rural Development* **2**, 9-13.

Braker M., Udo H. & Webb E. (2002) Impacts of intervention objectives in goat production within subsistence farming systems in South Africa. *South African Journal of Animal Science* **32**, 185-91.

Brito L.F., Jafarikia M., Grossi D., Maignel L., Sargolzaei M. & Schenkel F. (2014) Characterization of linkage disequilibrium and consistency of gametic phase in Canadian goats. In: *Proceedings of 10th World Congress of Genetics Applied to Livestock Production* 17-22.

Bryson N., Tice G., Horak I., Stewart C. & du Plessis B. (2002) Ixodid ticks on indigenous goats owned by small-scale farmers in four communal grazing areas in South Africa. *Journal of the South African Veterinary Association* **73**, 26-30.

Butswaat, I.S. (1994) Study on seasonal variation in the reproductive status of sheep and goats in Bauchi. Ph.D. Thesis, Abubarkar Tafawa Belewa University.

Cam M., Olfaz M. & Soydan E. (2010) Possibilities of using morphometrics characteristics as a tool for body weight prediction in Turkish hair goats (Kilkeci). *Asian Journal of Animal and Veterinary Advances* **5**, 52-59.

Campbell Q. (2003) The origin and description of southern Africa's indigenous goats. *South African Journal of Animal Science* **33**, 18-22.

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

Carillier C., Larroque H. & Robert-Granié C. (2014) Comparison of joint versus purebred genomic evaluation in the French multi-breed dairy goat population. *Genetics Selection Evolution* **46**, 1.

Casey N. & Van Niekerk W. (1988) The Boer goat. I. Origin, adaptability, performance testing, reproduction and milk production. *Small Ruminant Research* **1**, 291-302.

Census (2011) Agricultural households. Statistics South Africa. http://www.statssa.gov.za/publications/Report-03-11-01/Report-03-11-012011.pdf.

Coetzee R. (1998) Socio-economic aspects of sustainable goat production. *Research and training strategies for goat production systems in South Africa. Department of Animal and Wildlife Sciences, University of Pretoria, Pretoria,* 14-7.

Colli L., Joost S., Negrini R., Nicoloso L., Crepaldi P., Ajmone-Marsan P. & Consortium E. (2014) Assessing the spatial dependence of adaptive loci in 43 European and Western Asian goat breeds using AFLP markers. *PloS One* **9**, e86668.

Coop G., Witonsky D., Di Rienzo A. & Pritchard J.K. (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**, 1411-1423.

Domestic Animal Genetic Resources Information System (DAGRIS). 2007. Domestic Animal Genetic Resources Information System (DAGRIS). (eds. S. Kemp, Y. Mamo, B. Asrat and T. Dessie). International Livestock Research Institute, Addis Ababa, Ethiopia. http://dagris.ilri.cgiar.org.

Department of Agriculture, Forestry and Fisheries (DAFF). 2012. Republic of South Africa DAFF Annual Report 2011/2012. Available at http://www.daff.gov.za/.

Daramola J. & Adeloye A. (2009) Physiological adaptation to the humid tropics with special reference to the West African Dwarf (WAD) goat. *Tropical Animal Health and Production* **41**, 1005-16.

De Villiers J., Gcumisa S., Gumede S., Thusi S., Dugmore T., Cole M., Du Toit J., Vatta A. & Stevens C. (2009) Estimation of live body weight from the heart girth measurement in KwaZulu-Natal goats. *Applied Animal Husbandry and Rural Development* **1**, 1-8.

De Villiers J. & Letty B. (2001) Information collected during the diagnostic phase in the Nkwezela Sub-ward, Hlanganani District. KZN Agri Report N. A/2001.

Delgado J.V., Barba C., Camacho M.E., Sereno F.T.P.S., Martinez A. & Vega-Pla J.L. (2001) Livestock characterisation in Spain. *AGRI* **29**, 7-18.

Dent M.C., Lynch S.D. & Schulze R.E. (1987) Mapping mean annual and other rainfall statistics over southern Africa. Water Research Commission REport, Pietermartzburg, South Africa.

Desmet P.G. & Cowling R.M. (1999) The climate of the karoo- as functional approach In: *The Karoo: ecological patterns and processes* (eds. by Dean WRJ & Milton SJ), Cambridge University Press, Cambridge 3-16.

Devendra C. & McLeroy G.B. (1982) *Goat and sheep production in the tropics*. Longman Group, UK 271.

De Waal D.T., Matthee O. & Jongejan F. (2000) Evaluation of the MAP1b ELISA for the diagnosis of heartwater in South Africa. *Annals of the New York Academy of Sciences* **916**, 622-627.

Dillon S., McEvoy R., Baldwin D.S., Rees G.N., Parsons Y. & Southerton S. (2014) Characterisation of adaptive genetic diversity in environmentally contrasted populations of Eucalyptus camaldulensis Dehnh.(River Red Gum). *PloS One* **9**, e103515.

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

Donkin E., Stewart C., MacGregor R., Els H. & Boyazoglu P. (1992) Resistance of indigenous and crossbred goats to heartwater (*Cowdria ruminantium*). In: Recent Advances in Goat Production, V International Conference on Goats, New Delhi, 1716-1719.

Dossa L.H., Wollny C. & Gauly M. (2007) Smallholders' perceptions of goat farming in southern Benin and opportunities for improvement. *Tropical Animal Health and Production* **39**, 49-57.

Du Plessis J. (1981) Application of the indirect fluorescent antibody test to the serology of heartwater. In: *Tick biology and control: proceedings of an International Conference January 27-29, 1981/edited by GB Whitehead and JD Gibson*. Grahamstown, South Africa: Tick Research Unit, Rhodes University.

Du Plessis J., Bezuidenhout J.D. & Ludemann C. (1984) The immunization of calves against heartwater: subsequent immunity both in the absence and presence of natural tick challenge. *Onderstepoort Journal of Veterinary Research* **51**, 193-196.

Du Plessis J., Jansen B. & Prozesky L. (1983) Heartwater in Angora goats. I. Immunity subsequent to artificial infection and treatment. *Onderstepoort Journal of Veterinary Research* **50**, 137-43.

Du Toit C.J.L., Van Niekerk W.A. & Meissner H.H. (2014) Direct greenhouse gas emissions of the South African small stock sectors. *South African Journal of Animal Science* **43**, 340-61.

Duchev Z., Distl O. & Groeneveld E. (2006) Early warning system for loss diversity in European livestock breeds. *Archiv Tierzucht* **49**, 512-531.

Ehret C. (1982) The first spread of food production to southern Africa. *The Archaeological and Linguistic Reconstruction of African History* **158**, 182.

Elphick R. (1985) Khoikhoi and the founding of white South Africa. Johannesburg: Ravan.

Epstein H. (1971) The origin of the domestic animals of Africa. Africana, New York.

Excoffier L., Laval G. & Schneider S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47.

Faburay B., Munstermann S., Geysen D., Bell-Sakyi L., Ceesay A., Bodaan C. & Jongejan F. (2005) Point seroprevalence survey of Ehrlichia ruminantium infection in small ruminants in The Gambia. *Clinical and Diagnostic Laboratory Immunology* **12**, 508-512.

Food and Agriculture Organization of the United Nations (FAO). 2012 Phenotypic characterization of animal genetic resources. FAO Animal Production and Health Guidelines No. 11. Rome.

Food and Agriculture Organization of the United Nations (FAOSTAT). 2013. FAOSTAT Database, FAO, Rome, Italy. http://faostat3.fao.org/.

Frichot E., Schoville S.D., Bouchard G. & François O. (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution* **30**, 1687-99.

Gbetibouo G. & Hassan R. (2005) Economic impact of climate change on major South African field crops: A Ricardian approach. *Global and Planetary Change*, 143-152.

Gebreyesus G., Haile A. & Dessie T. (2013) Breeding scheme based on community-based participatory analysis of local breeding practices, objectives and constraints for goats around Dire Dawa, Ethiopia. *Livestock Research for Rural Development* **25**, article 48.

Goatcher W.D. & Church D.C. (1970) Taste responses in ruminants. IV Reaction of pygmy goats, normal goats, sheep and cattle to acetic acid and quinine hydrochloride. *Journal of Animal Science* **31**, 373-382.

Goddard M., Hayes B., McPartlan H. & Chamberlain A. (2006) Can the same genetic markers be used in multiple breeds? In: *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, 2006.* 22-16. Instituto Prociência.

Gugger P.F. (2012) Redundancy analysis tutorial: Landscape genetics. *University of Minnesota.*< http://www. *tc. umn. edu/~ gugg0030/Redundancy% 20Analysis% 20for% 20Landscape* **2**.

Günther T. & Coop G. (2013) Robust identification of local adaptation from allele frequencies. *Genetics* **195**, 205-220.

Gwaze F.R., Chimonyo M. & Dzama K. (2009a) Variation in the functions of village goats in Zimbabwe and South Africa. *Tropical Animal Health and Production* **41**, 1381-1391.

Gwaze F.R., Chimonyo M. & Dzama K. (2009b) Communal goat production in Southern Africa: a review. *Tropical Animal Health and Production* **41**, 1157-1168.

Gwaze F.R., Chimonyo M. & Dzama K. (2009c) Prevalence and loads of gastrointestinal parasites of goats in the communal areas of the Eastern Cape Province of South Africa. *Small Ruminant Research* **84**, 132-4.

Hagan J.K., Apori S.O., Bosompem M., Ankobea G. & Mawuli A. (2012) Morphological characteristics of indigenous goats in the coastal savannah and forest eco-zones of Ghana. *Journal of Animal Science Advances* **2**, 813-821.

Hancock A.M., Brachi B., Faure N., Horton M.W., Jarymowycz L.B., Sperone F.G., Toomajian C., Roux F. & Bergelson J. (2012) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* **334**, 83-86.

Hassen H., Baum M., Rischkowsky B. & Tibbo M. (2012) Phenotypic characterization of Ethiopian indigenous goat populations. *African Journal of Biotechnology* **11**, 13838-46.

Hill W. & Robertson A. (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**, 226-231.

Holmgren M., Stapp P., Dickman C.R., Gracia C., Graham S., Gutiérrez J.R., Hice C., Jaksic F., Kelt D.A. & Letnic M. (2006) Extreme climatic events shape arid and semiarid ecosystems. *Frontiers in Ecology and the Environment* **4**, 87-95.

Huson H., Sonstegard T., Silverstein J., Woodward-Greene M., Masiga C., Muchadeyi F., Rees J., Sayre B., Elbetagy A. & Rothschild M. (2014) Genetic and phenotypic characterization of African goat populations to prioritize conservation and production efforts for small-holder farmers in Sub-Saharan africa. In: *Vancouver, Canada: 10th World Congress on Genetics Applied to Livestock Production*.

Ilatsia E.D., Roessler R., Kahi A.K., Piepho H.-P. & Zárate V. (2012) Production objectives and breeding goals of Sahiwal cattle keepers in Kenya and implications for a breeding programme. *Tropical Animal Health and Production* **44**, 519-530.

Iniguez L. (2005) Sheep and goats in West Asia and North Africa: an overview. *Characterization of Small Ruminant Breeds in West Asia and North Africa* **1**, 1-29.

Joost S. (2014) Biogeoinformatics of livestock genomic resources. In: *Livestock genomic resources in a changing world*.

Joost S., Bonin A., Bruford M.W., Després L., Conord C., Erhardt G. & Taberlet P. (2007) A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* **16**, 3955-3969.

Joost S., Kalbermatten M. & Bonin A. (2008) Spatial analysis method (SAM): a software tool combining molecular and environmental data to identify candidate loci for selection. *Molecular ecology resources* **8**, 957-60.

Kakono O., Hove T., Geysen D. & Mahan S. (2003) Detection of antibodies to the Ehrlichia ruminantium MAP1-B antigen in goat sera from three communal land areas of Zimbabwe by an indirect enzyme-linked immunosorbent assay. *Onderstepoort Journal of Veterinary Research* **70**, 243-249.

Kanehisa M., Goto S., Kawashima S., Okuno Y. & Hattori M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Research* **32**, D277–D280.

Khan R. & Usmani R. (2005) Characteristics of rural subsistence small holder livestock production system in mountainous areas of NWFP, Pakistan. *Pakistan Veterinary Journal* **25**, 115-120.

Khanyile K.S., Dzomba E.F. & Muchadeyi F.C. (2015) Population genetic structure, linkage disequilibrium and effective population size of conserved and extensively raised village chicken populations of Southern Africa. *Frontiers in Genetics* **6**, article 13.

Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Neto L.R.P., San Cristobal M., Servin B., McCulloch R., Whan V. & Gietzen K. (2012) Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biology* **10**, e1001258.

Kijas J.W., Ortiz J.S., McCulloch R., James A., Brice B., Swain B. & Tosser- Klopp G. (2013) Genetic diversity and investigation of polledness in divergent goat populations using 52 088 SNPs. *Animal genetics* **44**, 325-335.

Kim E., Elbeltagy A., Aboul-Naga A., Rischkowsky 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**, 255-264.

Kirkman K.P. & de Faccio Carvalho P. (2003) Management interventions to overcome seasonal quantity and quality deficits of natural rangeland forages. In: *VIIth International Rangelands Congress* (eds. by Allsopp N, Palmer AR, Milton SJ, Kirkman KP, Kerley GIH, Hurt CR & Brown CJ), pp. 1289-1297, Durban, South Africa.

Kosgey I.S. (2004) Breeding objectives and breeding strategies for small ruminants in the tropics. PhD thesis, Wageningen University, Wageningen.

Kosgey I.S., Baker R.L., Udo H.M.J. & Van Arendok J.A.M. (2006) Successes and failure of small ruminant breeding programmes in the tropics. A review. *Small Ruminant Research* **61**, 13-28.

Kotze A., Grobler J.P., Marle-Köster E.v., Jonker T. & Dalton D.L. (2014) The Tankwa Karoo National Park feral goat population: a unique genetic resource. *South African Journal of Animal Science* **44**, 43-48.

Kotze A., Swart H., Grobler J. & Nemaangani A. (2004) A genetic profile of the Kalahari Red goat breed from southern Africa. *South African Journal of Animal Science* **34**, 10-12.

Kristensen T.N., Hoffmann A.A., Pertoldi C. & Stronen A.V. (2015) What can livestock breeders learn from conservation genetics and vice versa? *Frontiers in genetics* **6**, 38.

Kristjanson P., Thornton P., Kruska R., Reid R., Henninger N., Williams T., Tarawali S., Niezen J. & Hiernaux P. (2001) Mapping livestock systems and changes to 2050: Implications for West Africa. In: *Proceedings from an international conference on "Sustainable crop-livestock production for improved livelihoods and natural resource management in West Africa*, 19-21.

Kunene N. & Fossey A. (2006) A survey on livestock production in some traditional areas of Northern Kwazulu Natal in South Africa. *Livestock Research for Rural Development* **18**, 30-33.

Larmer S., Sargolzaei M. & Schenkel F. (2014) Extent of linkage disequilibrium, consistency of gametic phase, and imputation accuracy within and across Canadian dairy breeds. *Journal of Dairy Science* **97**, 3128-3141.

Lashmar S., Visser C. & van Marle-Köster E. (2015) Validation of the 50k Illumina goat SNP chip in the South African Angora goat. *South African Journal of Animal Science* **45**, 56-59.

Lashmar S., Visser C. & van Marle-Köster E. (2016) SNP-based genetic diversity of South African commercial dairy and fiber goat breeds. *Small Ruminant Research* **136**, 65-71.

Lasky J.R., Des Marais D.L., McKay J.K., Richards J.S., Juenger T.S. & Keitt T.H. (2012) Characterizing genomic variation of Arabidopsis thaliana: the roles of geography and climate. *Molecular Ecology* **21**, 5512-29.

Lebbie S. (2004) Goats under household conditions. *Small Ruminant Research* **51**, 131-136.

Legendre P. & Legendre L. (1998) Numerical Ecology. Elsevier, New York.

Low A.B. & Rebelo A.G. (1996) Vegetation of South Africa, Lesotho and Swaziland. Pretoria. (ed. by Tourism DoEA), Pretoria, SA.

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

Lv F. H., Agha S., Kantanen J., Colli L., Stucki S., Kijas J.W., Joost S., Li M.H. & Marsan P.A. (2014) Adaptations to climate-mediated selective pressures in sheep. *Molecular Biology and Evolution*, msu264.

Mahan S., Semu S., Peter T. & Jongejan F. (1998) Evaluation of the MAP- 1B ELISA for Cowdriosis with Field Sera from Livestock in Zimbabwea. *Annals of the New York Academy of Sciences* **849**, 259-261.

Mahanjana A. & Cronje P. (2000) Factors affecting goat production in a communal farming system in the Eastern Cape region of South Africa. *South African Journal of Animal Science* **30**, 149-55.

Makina S.O., Muchadeyi F.C., Marle-Köster E., Taylor J.F., Makgahlela M.L. & Maiwashe A. (2015) Genome-wide scan for selection signatures in six cattle breeds in South Africa. *Genetics Selection Evolution* **47**, 92.

Malan S.W. (2000) The improved Boer goat. *Small Ruminant Research* **36**, 165-70.

Maleko D.D. & Koipapi M.L. (2015) Opportunities and constraints for overcoming dry season livestock feed shortages in communal semi-arid rangelands of Northern Tanzania: A case of Longido District. *Livestock Research for Rural Development* **27**, article 70.

Mamabolo M.J. (1999) Dietary, seasonal and environmental influences on the semen quality and fertility status of indigenous goats in Mpumalanga Province (South Africa). Msc. thesis, University of Pretoria.

Manel S., Joost S., Epperson B.K., Holderegger R., Storfer A., Rosenberg M.S., Scribner K.T., Bonin A. & Fortin M.J. (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology* **19**, 3760-3772.

Manton H. (2005) Dreams and Medicines: The Perspective of Xhosa Diviners and Novices in the Eastern Cape, South Africa. *Indo-Pacific Journal of Phenomenology* **5**, 1-22.

Maree C. & Casey N. (1993) Livestock production systems: Principles and practice. *Agri-Development Foundation, Brooklyn, Pretoria, South Africa*.

Maree C. & Plug I. (1993) Origin of Southern African livestock and their potential role in the industry. *Livestock Production Systems Eds. Maree, C. & Casey, NH., Agri-Development Foundation*, 5.13.

Maroteau C. (2013) QTL detection for traits of interest for the dairy goat industry. In: *64th Annual Meeting of the European Federation of Animal Science (EAAP), Nathes, FRANCE, Wageningen Academic Publishers*.

Marume U., Chimonyo M. & Dzama K. (2013) Some insights into the reproductive performance, selection criteria, production and market constraints in communal goat production systems of South Africa. *Scientific Research and Essays* **8**, 1448-1458.

Masika P. & Mafu J. (2004) Aspects of goat farming in the communal farming systems of the central Eastern Cape, South Africa. *Small Ruminant Research* **52**, 161-164.

Matheron G., Barre N., Camus E. & Gogue J. (1987) Genetic resistance of Guadeloupe native goats to heartwater. *Onderstepoort Journal of Veterinary Research* **54**, 337-340.

Mbole-Kariuki M., Sonstegard T., Orth A., Thumbi S., de C Bronsvoort B., Kiara H., Toye P., Conradie I., Jennings A. & Coetzer K. (2014) Genome-wide analysis reveals the ancient and recent admixture history of East African Shorthorn Zebu from Western Kenya. *Heredity* **113**, 297-305.

Mbuh J., Ndamukong K., Ntonifor N. & Nforlem G. (2008) Parasites of sheep and goats and their prevalence in Bokova, a rural area of Buea Sub Division, Cameroon. *Veterinary Parasitology* **156**, 350-352.

McCue M.E., Bannasch D.L., Petersen J.L., Gurr J., Bailey E., Binns M.M., Distl O., Guérin G., Hasegawa T. & Hill E.W. (2012) A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genetics* **8**, e1002451.

McGaughran A., Morgan K. & Sommer R.J. (2014) Environmental Variables Explain Genetic Structure in a Beetle-Associated Nematode. *PloS One* **9**, e87317.

Mdladla K., Dzomba E., Huson H. & Muchadeyi F. (2016a) Population genomic structure and linkage disequilibrium analysis of South African goat breeds using genome- wide SNP data. *Animal genetics* **47**, 471-482.

Mdladla K., Dzomba E.F. & Muchadeyi F.C. (2016b) Seroprevalence of Ehrlichia ruminantium antibodies and its associated risk factors in indigenous goats of South Africa. *Preventive Veterinary Medicine* **125**, 99-105.

Meutchieye F., Ema-Ngono P., Agaba M., Djikeng A. & Manjeli Y. (2014) Genetic diversity of Cameroon indigenous goat populations using microsatellites. *Livestock Research for Rural Development* **26**, 128.

Miller J., Malenfant R., David P., Davis C., Poissant J., Hogg J., Festa-Bianchet M. & Coltman D. (2014) Estimating genome-wide heterozygosity: effects of demographic history and marker type. *Heredity* **112**, 240-247.

Mohlatlole R.P., Dzomba E.F. & Muchadeyi F.C. (2015) Addressing production challenges in goat production systems of South Africa: The genomics approach. *Small Ruminant Research* **131**, 43-49.

Moraa G., Oyier P., Maina S., Makanda M., Ndiema E., Alakonya A., Ngeiywa K., Lichoti J. & Ommeh S. (2015) Assessment of phenotypic traits relevant for adaptation to hot environments

in indigenous chickens from four agro-climatic zones of Kenya. *Livestock Research for Rural Development* **27**, 10.

Morand-Fehr P., Boutonnet J., Devendra C., Dubeuf J., Haenlein G., Holst P., Mowlem L. & Capote J. (2004) Strategy for goat farming in the 21st century. *Small Ruminant Research* **51**, 175-183.

Morrison J.W. (2007) A Guide to the Identification of the Natural African Indigenous Veld Goats of Southern Africa. In: *Compiled for the Indigenous Veld Goat Breeders Society, Pretoria, South Africa*, Pretoria.

Mpofu N. (2002) Choice of genetic types for specific production environments and production systems.

Mrema M. & Rannobe S. (1996) Goat production in Botswana: Factors affecting production and marketing among small-scale farmers. In: *Small ruminant research and development in Africa. Proceedings of the third biennial conference of the African Small Ruminant Research Network, Kampala, Uganda*, pp. 105-9. International Livestock Research Institute (ILRI).

Mubi A.A., Bube M.M., Michika S.A. & Midau A. (2012) Infection pattern of Helminths in goats around Mubi metropolis, Adamawa State, Nigeria. *Current Trends in Technology and Science* **2**, 222-225.

Mukhebi A., Chamboko T., O'Callaghan C., Peter T., Kruska R., Medley G., Mahan S. & Perry B. (1999) An assessment of the economic impact of heartwater (Cowdria ruminantium infection) and its control in Zimbabwe. *Preventive veterinary medicine* **39**, 173-189.

Nampanzira D.K., Kabasa J.D., Nalule S.A., Nakalembe I. & Tabuti J.R.S. (2015) Characterization of the goat feeding system among rural small holder farmers in the semi-arid regions of Uganda. *SpringerPlus* **4**, 1.

National Department of Agriculture (NDA). (2012). A profile of the South African goat market value chain. Retrieved January 2016, 19, 1-33.

Ncube K. (2016) maternal lineages and diversity of the growth hormone gene of South African goat populations. Msc dissertation. University of South Africa.

Neitz W. (1967) The epidemiological pattern of viral, protophytal and protozoal zoonoses in relation to game preservation in South Africa. *Journal of the South African Veterinary Medical Association* **38**, 129-141.

Nicoloso L., Bomba L., Colli L., Negrini R., Milanesi M., Mazza R., Sechi T., Frattini S., Talenti A. & Coizet B. (2015) Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. *Genetics Selection Evolution* **47**, 1.

Norris D., Ngambi J., Benyi K. & Mbajiorgu C.A. (2011) Milk production of three exotic dairy goat genotypes in Limpopo province, South Africa. *Asian Journal of Animal and Veterinary Advances* **6**, 274-281.

Norval R. & Horak I. (2004) Vector Ticks. In: *Infectious diseases of livestock* (eds. by Coetzer JAW & Tustin RC), pp. 3-42. Oxford University Press,, Cape Town, South Africa.

Nwosu C.O., Madu P.P. & Richards W. (2007) Prevalence and seasonal changes in the population of gastrointestinal nematodes of small ruminants in the semi-arid zone of North-Eastern Nigeria. *Veterinary parasitology* **144**, 118-124.

Obexer-Ruff G., Sattler U., Martinez D., Maillard J.-C., Chartier C., Saitbekova N., Glowatzki M.-L. & Gaillard C. (2003) Association studies using random and" candidate" microsatellite loci in two infectious goat diseases. *Genetics Selection Evolution* **35**, 1.

Ojo O., Akpa G., Orunmuyi M. & Adeyinka I. (2015) Genetic Differentiation among Nigerian Indigenous Goat Populations. *Journal of Agricultural Science* **7**, 39-47.

Okpeku M., Yakubu A., Peters S., Ozoje M., Ikeobi C., Adebambo O. & Imumorin I. (2011) Application of multivariate principal component analysis to morphological characterization of indigenous goats in Southern Nigeria. *Acta Agriculturae Slovenica* **98**, 101-109.

Ometto L., Li M., Bresadola L., Barbaro E., Neteler M. & Varotto C. (2015) Demographic History, Population Structure, and Local Adaptation in Alpine Populations of Cardamine impatiens and Cardamine resedifolia. *PloS One* **10**, e0125199.

Ozoje M. (2002) Incidence and relative effects of qualitative traits in West African Dwarf goat. *Small Ruminant Research* **43**, 97-100.

Palhière I., Larroque H., Clement V., Tosser-Klopp G. & Rupp R. (2014) Genetic parameters and QTL detection for milking speed in dairy alpine and saanen goats. In: *10th World Congress on Genetics Applied to Livestock Production*, Vancouver, BC, Canada.

Palmer T. & Ainslie A. (2006) Country pasture/forage resource profiles. *Food and Agriculture Organization, South Africa*.

Pariset L., Joost S., Marsan P.A. & Valentini A. (2009) Landscape genomics and biased F ST approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genetics* **10**, 7.

Payne W.J.A. & Wilson R.T. (1999) An introduction to animal husbandry in the tropics. Blackwell science.

Peacock C., Devendra C., Ahuya C., Roets M., Hossain M. & Osafo E. (2005) Livestock and wealth creation: Improving the husbandry of animals kept by resource-poor people in developing countries. Nottingham University Press, UK.

Peter T., Perry B., O'Callaghan C., Medley G., Shumba W., Madzima W., Burridge M. & Mahan S. (1998) Distributions of the vectors of heartwater, Amblyomma hebraeum and Amblyomma variegatum (Acari: Ixodidae), in Zimbabwe. *Experimental & Applied Acarology* **22**, 725-740.

Petney T.N., Horak I.G. & Rechav Y. (1987) The ecology of the African vectors of heartwater, with particular reference to *Amblyomma hebraeum* and A*mblyomma variegatum*. *Onderstepoort Journal of Veterinary Research* **54**, 381-395.

Pieters A. (2007) Genetic characterization of commercial goat populations in South Africa. Msc thesis. University of Pretoria.

Pritchard J.K. & Przeworski M. (2001) Linkage disequilibrium in humans: models and data. *The American Journal of Human Genetics* **69**, 1-14.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., Maller J., Sklar P., De Bakker P.I. & Daly M.J. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559-575.

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

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, url: http://www.R-project.org/.

Randolph T., Schelling E., Grace D., Nicholson C.F., Leroy J., Cole D., Demment M., Omore A., Zinsstag J. & Ruel M. (2007) Role of livestock in human nutrition and health for poverty reduction in developing countries. *Journal of animal science* **85**, 2788-2800.

Rege J. (1994) Indigenous African small ruminants: a case for characterisation and improvement. In: *Small Ruminant Research and Development in Africa, Proceedings of the Second Biennial Conference of the African Small Ruminant Research Network*, 205-211.

Rege J. & Gibson J. (2003) Animal genetic resources and economic development: issues in relation to economic valuation. *Ecological economics* **45**, 319-330.

Rischkowsky B. & Pilling D. (2007) *The state of the world's animal genetic resources for food and agriculture*. Food & Agriculture Organisation, of the United Nations (FAO), Rome, Italy.

Robertshaw D. (2006) Mechanisms for the control of respiratory evaporative heat loss in panting animals. *Journal of Applied Physiology* **101**, 664-648.

SAS Institute Inc. (2013), SAS/ETS User's Guide, Version 13.1, Cary, NC: SAS Institute Inc.

Schaffer J., V. & Bolognia J. L. (2001) The melanocortin-1 receptor: red hair and beyond, Archives of Dermatology **137**, 1477-1485.

Schiere H. & Kater L. (2001) Mixed crop-livestock farming. A review of traditional technologies based on literature and field experience. FAO Animal Production and Health Paper.

Scholtz, M.M., Maiwashe, A., Neser, F.W.C., Theunissen, A., Olivier, W.J., Mokolobate, M.C. & Hendriks, J. (2013) Livestock breeding for sustainability to mitigate global warming, with the emphasis on developing countries. *South African Journal of Animal Science* **43**, 269-281.

Sebei P., McCrindle C. & Webb E. (2004a) An economic analysis of communal goat production. *Journal of the South African Veterinary Association* **75**, 19-23.

Sebei P., McCrindle C. & Webb E. (2004b) Factors influencing weaning percentages of indigenous goats on communal grazing. *South African Journal of Animal Science* **34**, 130-133.

Shabalala N. & Mosima B. (2002) Report on the survey of large and small scale agriculture/Statistics South Africa. In: *Statistics SA Library Cataloguing-in-Publication (CIP) Data Pretoria Statistics South Africa.*, South Africa.

Shongjia L., Xinagmo L., Gangy X. & Shenov C. (1992) Relationship between physical traits, litter size and milk yield in Saanen, Guanzhong and crossbred goats. In: *Proceedings of the International Conference on Goats*, 83.

Silanikove N. (2000) The physiological basis of adaptation in goats to harsh environments. *Small Ruminant Research* **35**, 181-193.

Silanikove N., Gilboa A., Perevolotsky A. & Nitsan Z. (1996) Goats fed tannin-containing leaves do not exhibit toxic syndromes. *Small Ruminant Research* **21**, 195-201.

Silanikove N., Nitsan Z. & Perevolotsky A. (1994) Effect of daily supplementation of polyethylene glycol on intake and digestion of tannin-containing leaves (*Ceratonia siliqua*) by sheep. *Journal of Agricultural and Food Chemistry* **42**, 2844-2847.

Silanikove N., Tagari H. & Shkolnik A. (1993) Comparison of rate passage fermentation rate and ef®ciency of digestion of high ®ber diet in desert black Bedouin goats as compared to Swiss Saanen goats. *Small Ruminant Research* **12**, 45-60.

Simela L. & Merkel R. (2008) The contribution of chevon from Africa to global meat production. *Meat Science* **80**, 101-109.

Sisson S. & Grossman J. (1953) The male urethra. *The anatomy of domestic animals. 4th edition. Philadelphia: WB Saunders*, 594-596.

Sivakumar M. (1992) Climate change and implications for agriculture in Niger. *Climatic Change* **20**, 297-312.

Song, S., Yao, N., Yang, M., Liu, X., Dong, K., Zhao, Q., Pu, Y., He, X., Guan, W., Yang, N. and Ma, Y. (2016). Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (Capra hircus). *BMC Genomics* **17**, 122.

Spickett A.M., Heyne I.H. & Williams R. (2011) Survey of the livestock ticks of the North West province, South Africa. *Onderstepoort Journal of Veterinary Research* **78**, 1-12.

Sponenberg D.P., Beranger J. & Martin A. (2014) An Introduction to Heritage Breeds. Storey Publishing, LLC.

Sved J. (1971) Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology* **2**, 125-141.

Swai E., Moshy W., Mtui P., Bwanga S., Machange G. & Sanka P. (2009) Serological survey of antibodies to Ehrlichia ruminantium in small ruminants in Tanzania. *Tropical Animal Health and Production* **41**, 959-967.

Swai E., Mtui P., Chang'a A. & Machange G. (2008) The prevalence of serum antibodies to Ehrlichia ruminantium infection in ranch cattle in Tanzania: a cross-sectional study. *Journal of the South African Veterinary Association* **79**, 71-75.

Szpiech Z.A., Jakobsson M. & Rosenberg N.A. (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**, 2498-2504.

Theiler G. (1948) Zoological survey of the Union of South Africa; tick survey. *Onderstepoort Journal of Veterinary Science and Animal Industry* **23**, 217-231.

Toro M.A. & Caballero A. (2005) Characterisation and conservation of genetic diversity in subdivided populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 1367–1378.

Toro M.A., Fernandez J. & Caballero A. (2009) Molecular characterization of breeds and its use in conservation. *Livestock Science* **120**, 174-95.

Tosser-Klopp G., Bardou P., Bouchez O., Cabau C., Crooijmans R., Dong Y., Donnadieu-Tonon C., Eggen A., Heuven H.C. & Jamli S. (2014) Design and characterization of a 52K SNP chip for goats. *PloS One* **9**, e86227.

Tosser-Klopp, G. (2012) Goat genome assembly, Availability of an international 50K SNP chip and RH panel: An update of the International Goat Genome Consortium projects. In Plant and Animal Genome XX Conference. Plant and Animal Genome. San Diego.

Traoré A., Tamboura H.H., Kabore A., Royo L.J., Fernandez I., Alvarez I., Sangare M., Bouchel D., Poivey J.P. & Francois D. (2008) Multivariate analyses on morphological traits of goats in Burkina Faso. *Archiv Tierzucht* **51**, 588-600.

Tsegaye D., Belay B. & Haile A. (2013) Morphological characterization of indigenous Hararghe highland goat breeds in their native environment, West Haraghe, Ethiopia. *American-Eurasian Journal of Scientific Research* 8, 72-79.

Turton J. (1999) Tick-borne diseases in ruminants. URL http://www.elsenburg.com/info/nda/tick-borne diseases.pdf.

Van Den Wollenberg A.L. (1977) Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika* **42**, 207-219.

Van Niekerk W. & Pimentel P. (2004) Goat production in the smallholder section in the Boane district in Southern Mozambique. *South African Journal of Animal Science* **34**, 123-125.

van Vuren D. & Hedrick P.W. (1989) Genetic conservation in feral populations of livestock. *Conservation Biology* **3**, 312-317.

Vatta A., Letty B., Van der Linde M., Van Wijk E., Hansen J. & Krecek R. (2001) Testing for clinical anaemia caused by Haemonchus spp. in goats farmed under resource-poor conditions in South Africa using an eye colour chart developed for sheep. *Veterinary Parasitology* **99**, 1-14.

Visser C., Hefer C., van Marle-Koster E. & Kotze A. (2004) Genetic variation of three commercial and three indigenous goat populations in South Africa. In: South African Journal of Animal Science: Conference on Goats 2004: Supplement 1, 24-27. Sabinet Online.

Visser C., van Marle-Köster E. & Friedrich H. (2011) Parentage verification of South African Angora goats, using microsatellite markers. *South African Journal of Animal Science* **41**, 250-255.

Walker J.B. & Olwage A. (1987) The tick vectors of Cowdria ruminantium (Ixodoidea, Ixodidae, genus Amblyomma) and their distribution. *Onderstepoort Journal of Veterinary Research* **54**, 353-379.

Webb E.C. & Mamabolo M. (2004) Production and reproduction characteristics of South African indigenous goats in communal farming systems. *South African Journal of Animal Science* **34**, 236-239.

Weir B.S. & Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 1358-1370.

Whittaker R.H, Levin SA, Root R.B (1973) Niche, Habitat, and Ecotope. *The American Naturalist* **107**, 321-338.

Wright S. (1978) Vol. 4: Variability within and among natural populations. *Evolution and the genetics of populations: a treatise in four volumes: University of Chicago Press*.

Yakubu A., Salako A. & Imumorin I. (2010a) Multivariate analysis of spatial patterns of morphological traits in West African Dwarf goats in three agro-ecological zones of Nigeria. *Journal of Applied Animal Research* **38**, 257-260.

Yakubu A., Salako A., Imumorin I., Ige A. & Akinyemi M. (2010b) Discriminant analysis of morphometric differentiation in the West African Dwarf and Red Sokoto goats. *South African Journal of Animal Science* **40**, 381-387.

Ziervogel G., New M., Archer van Garderen E., Midgley G., Taylor A., Hamann R., Stuart- Hill S., Myers J. & Warburton M. (2014) Climate change impacts and adaptation in South Africa. *Wiley Interdisciplinary Reviews: Climate Change* **5**, 605-620.

Zweygarth E., Josemans A.I. & Steyn H.C. (2008) Experimental use of the attenuated *Ehrlichia ruminantium* (Welgevonden) vaccine in Merino sheep and Angora goats. *Vaccine* **26**, G34-G9.

APPENDIX A: DISTRIBUTION OF SMALLHOLDER GOAT FARMERS HOUSEHOLDS USED FOR THE QUESTIONNAIRE SURVEY

Province	District	Village	Number
Eastern Cape	Amathole	Hobeni	**8**
		Lower Desi	**3**
KwaZulu-Natal	uMzinyathi	Coniliva	**11**
		Guqa-Mkhovini	**5**
		Mathengwenya	**9**
		Mbomeni	**6**
		Ncunjana	**4**
		Ngubo	**11**
Limpopo	Vhembe	Madimbo	**4**
		Malale	**11**
		Mpheni	**1**
		Muila-muumoni	**10**
		Mukovhabale	**2**
		Ribungwane	**2**
		Tshandama	**1**
		Tshikambe	**8**
		Tshimboni	**3**
		Tshivaloni	**2**
		Tshivhilwi	**5**
		Valdezia	**1**
		Vuvha	**5**
North West	Kenneth Kaunda	Appeldraai	**2**
		Boikhutsong	**1**
		Goedgenonden	**5**
	Bojanala	Toelane	**4**
		Pella	**6**
	Total		130

APPENDIX B: POTENTIAL CAUSAL GENES UNDERLYING OUTLIER SNPS (>0.6)

Marker	Chr	Position	F_{ST}	Gene
snp1758-scaffold1049-635356	24	25507539	0.81	-
snp26862-scaffold283-114619	18	55048519	0.79	*FAM71E1*
snp7402-scaffold127-1054244	2	107877611	0.77	-
snp39578-scaffold503-1794028	13	20650342	0.76	-
snp24690-scaffold2518-214293	25	3353128	0.75	*ADCY9*
snp11354-scaffold1410-162641	2	105828819	0.74	*USP37*
snp41062-scaffold531-955436	12	19263166	0.74	-
snp1767-scaffold1049-1053131	24	25089764	0.74	-
snp50263-scaffold718-182724	7	94039060	0.74	*CHD1*
snp50262-scaffold718-122726	7	94099058	0.73	*CHD1*
snp6115-scaffold1216-1446006	13	10521185	0.73	-
snp32071-scaffold361-508535	3	246193	0.73	*SFT2D2*
snp14461-scaffold1578-572824	17	37084927	0.72	-
snp54436-scaffold830-2333244	8	60619965	0.72	-
snp352-scaffold1008-2479850	9	47432421	0.72	-
snp26314-scaffold276-354652	1	88912946	0.72	-
snp4369-scaffold1137-1287981	7	99350420	0.71	-
snp51271-scaffold75-5070454	14	36054587	0.71	-
snp748-scaffold102-4832971	24	23726403	0.71	-
snp40867-scaffold526-814277	24	43767975	0.70	-
snp41661-scaffold542-781596	15	47060976	0.70	-
snp35541-scaffold428-1842450	17	32817364	0.70	-
snp20132-scaffold2-5195652	7	87178558	0.69	-
snp11266-scaffold1404-460405	22	12548224	0.69	*MOBP*
snp40899-scaffold526-2229065	24	45182763	0.69	-
snp49900-scaffold714-30038	23	11037749	0.68	-
snp32067-scaffold361-318023	3	436705	0.68	*LOC102187999*
snp23065-scaffold230-3508504	1	129682502	0.68	-
snp34166-scaffold401-329664	8	98008475	0.68	*SVEP1*

APPENDIX B: POTENTIAL CAUSAL GENES UNDERLYING OUTLIER SNPS (>0.6) (continued)

Marker	Chr	Position	F_{ST}	Gene
snp32075-scaffold361-731478	3	23250	0.68	-
snp16562-scaffold1747-177599	17	30729262	0.68	-
snp12725-scaffold1489-481443	2	76106617	0.68	*LOC102187631*
snp26358-scaffold276-2262816	1	90821110	0.68	-
snp10225-scaffold1368-727231	23	42588072	0.68	*TMEM14C*
snp46301-scaffold638-2380769	13	18666418	0.68	*NRP1*
snp3659-scaffold112-28499	16	58446275	0.67	*TDRD5*
snp50174-scaffold717-4415291	12	24800232	0.67	*NBEA*
snp362-scaffold1008-2966313	9	47918884	0.67	*RNGTT*
snp55019-scaffold841-1975437	3	13372807	0.67	-
snp24719-scaffold2523-63156	18	55714101	0.66	*SIGLECL1*
snp58715-scaffold958-497114	13	21189760	0.66	*C13H10orf113/ NEBL*
snp41790-scaffold543-3480826	3	106271692	0.66	-
snp40389-scaffold515-3901447	2	90388801	0.66	-
snp28455-scaffold303-4083255	19	33605591	0.66	*MYO15A*
snp49469-scaffold705-147060	20	70609585	0.66	-
snp54989-scaffold841-620945	3	14727299	0.66	-
snp42837-scaffold568-6566283	12	60742379	0.66	-
snp31747-scaffold354-655359	6	655359	0.66	-
snp59309-scaffold975-765057	23	11832943	0.65	*BTBD9*
snp6868-scaffold1250-177217	10	108202	0.65	-
snp26380-scaffold276-3178572	1	91736866	0.65	*NLGN1*
snp39946-scaffold51-2218519	21	43596524	0.65	-
snp42831-scaffold568-6344853	12	60963809	0.65	-
snp41915-scaffold546-2606320	18	8289876	0.64	-
snp43069-scaffold572-1135310	10	92120459	0.64	-
snp55855-scaffold868-569009	1	4013831	0.64	-
snp23705-scaffold239-2400740	25	21250719	0.64	-
snp12656-scaffold1485-31052	7	37229958	0.64	-
snp20760-scaffold204-3388900	11	76448163	0.64	-
snp24815-scaffold254-1143492	7	58771708	0.64	*GM2A*

APPENDIX B: POTENTIAL CAUSAL GENES UNDERLYING OUTLIER SNPS (>0.6) (continued)

Marker	Chr	Position	F_{ST}	Gene
snp54991-scaffold841-747978	3	14600266	0.64	*DCST1*
snp20320-scaffold201-155901	16	66545287	0.64	-
snp16465-scaffold1734-371724	6	110848916	0.64	*SORCS2*
snp21502-scaffold210-164871	7	105051869	0.64	-
snp3325-scaffold1101-1094567	19	54504863	0.64	*TEN1*
snp58713-scaffold958-432366	13	21125012	0.63	*NEBL*
snp56924-scaffold902-763358	7	27376463	0.63	-
snp14821-scaffold1599-948503	16	48216376	0.63	*PRKCZ*
snp39579-scaffold503-1833289	13	20689603	0.63	-
snp28549-scaffold305-1942611	12	71431604	0.63	-
snp46649-scaffold65-565888	1	103093925	0.63	-
snp58719-scaffold958-675479	13	21368125	0.63	-
snp58521-scaffold952-954794	20	60456485	0.63	-
snp47498-scaffold67-91973	26	35877445	0.63	-
snp18067-scaffold185-11090622	11	61968542	0.62	-
snp26180-scaffold274-631324	29	30882159	0.62	-
snp33493-scaffold393-503159	2	67202240	0.62	-
snp47792-scaffold673-76823	23	14759916	0.62	-
snp1024-scaffold1027-604298	26	40231957	0.62	*ATE1*
snp56927-scaffold902-874152	7	27487257	0.62	-
snp1563-scaffold10434-6585	29	43144521	0.62	*CORO1B*
snp4987-scaffold117-1975	4	45378364	0.62	-
snp14051-scaffold1559-114629	4	94884129	0.62	*EXOC4*
snp347-scaffold1008-2273537	9	47226108	0.62	*MDN1*
snp47015-scaffold658-62075	11	101594620	0.62	-
snp29011-scaffold312-3058404	2	62973288	0.62	-
snp20133-scaffold2-5244407	7	87129803	0.62	-
snp30544-scaffold338-37210	15	72147651	0.62	-
snp42852-scaffold57-388484	15	62874903	0.62	*TRIM44*
snp2674-scaffold1077-361080	1	19098932	0.62	-
snp40822-scaffold524-1110687	9	26153564	0.62	*REV3L*
snp24084-scaffold2457-275912	27	14865153	0.61	-

APPENDIX B: POTENTIAL CAUSAL GENES UNDERLYING OUTLIER SNPS (>0.6) (continued)

Marker	Chr	Position	F_{ST}	Gene
snp58667-scaffold956-623199	13	69599357	0.61	*PTPRT*
snp46706-scaffold65-3055804	1	105583841	0.61	*SCHIP1*
snp12875-scaffold1497-2511326	22	52346033	0.61	*PTH1R*
snp5975-scaffold121-1467260	17	59566212	0.61	-
snp4091-scaffold1129-824063	14	83761676	0.61	*TMEM55A*
snp44070-scaffold597-624744	16	42628638	0.60	*SLC45A1*
snp302-scaffold1008-297138	9	45249709	0.60	-
snp28139-scaffold300-6098892	8	40883974	0.60	-
snp51830-scaffold762-142511	27	37787018	0.60	*LRRC3B*

APPENDIX C: LOCATIONS (PROVINCE, GPS COORDINATES AND ALTITUDE) OF THE 194 INDIVIDUALS AND THE WEATHER STATION WHERE CLIMATIC DATA WAS COLLECTED

Sample Name	Breed/ population	Province	Original GPS coordinates		Displaced GPS coordinates		Altitude (m)	Nearest weather station
			Latitude	Longitude	Latitude	Longitude		
CRA24	Boer	Eastern Cape	-32.2772	25.621	-32.2772	25.621	893	30387
CRA08	Boer	Eastern Cape	-32.2991	25.0908	-32.2991	25.0908	805	30387
CRA13	Boer	Eastern Cape	-32.1598	25.621	-32.1598	25.621	1219	30387
CRA15	Boer	Eastern Cape	-32.1598	25.621	-32.1598	25.621	1219	30387
FOX06	Boer	KwaZulu-Natal	-29.4057	30.2254	-29.4057	30.2254	857	30160
FOX10	Boer	KwaZulu-Natal	-29.4057	30.2254	-29.4057	30.2254	857	30160
FOX17	Boer	KwaZulu-Natal	-29.4057	30.2254	-29.4057	30.2254	857	30160
FOX04	Boer	KwaZulu-Natal	-29.4057	30.2254	-29.4057	30.2254	857	30160
GD06	Boer	North West	-26.224	24.4022	-26.224	24.4022	1292	30525
GD27	Boer	North West	-26.224	24.4022	-26.224	24.4022	1292	30525
LIC01	Boer	North West	-26.1516	26.1623	-26.1516	26.1623	1512	30728
LIC09	Boer	North West	-26.1516	26.1623	-26.2585	26.1257	1512	30728
LIC17	Boer	North West	-26.1516	26.1623	-26.2191	26.0598	1512	30728
KOOP14	Boer	Northern Cape	-28.1214	24.241	-28.1214	24.241	1054	30594
KK04	Boer	Northern Cape	-28.5023	22.229	-28.5023	22.229	1076	30181
KOOP30	Boer	Northern Cape	-28.1214	24.241	-28.1214	24.241	1054	30594
KK22	Boer	Northern Cape	-28.5023	22.229	-28.3861	22.5441	1076	30181
KB02	Boer	Northern Cape	-29.0313	23.1826	-29.0313	23.1826	1271	321110
KB06	Boer	Northern Cape	-29.0313	23.1826	-29.1757	23.1813	1271	321110
KB15	Boer	Northern Cape	-29.0313	23.1826	-29.0989	23.3132	1271	321110
KB19	Boer	Northern Cape	-29.0313	23.1826	-29.0413	23.4011	1271	321110
KB20	Boer	Northern Cape	-29.0313	23.1826	-29.0605	23.1813	1271	321110
KB26	Boer	Northern Cape	-29.0313	23.1826	-28.926	23.2912	1271	321110

APPENDIX C: LOCATIONS (PROVINCE, GPS COORDINATES AND ALTITUDE) OF THE 194 INDIVIDUALS AND THE WEATHER STATION WHERE CLIMATIC DATA WAS COLLECTED

Sample Name	Breed/ population	Province	Original GPS coordinates		Displaced GPS coordinates		Altitude (m)	Nearest weather station
			Latitude	Longitude	Latitude	Longitude		
CRA09	Kalahari Red	Eastern Cape	-32.1598	25.621	-32.1598	25.621	1219	30387
CRA17	Kalahari Red	Eastern Cape	-32.1598	25.621	-32.1598	25.621	1219	30387
ROLL05	Kalahari Red	KwaZulu-Natal	-28.431	29.3738	-28.431	29.3738	1015	30391
ROLL04	Kalahari Red	KwaZulu-Natal	-28.431	29.3738	-28.434	29.3753	1015	30391
ROLL09	Kalahari Red	KwaZulu-Natal	-28.431	29.3738	-28.431	29.3738	1015	30391
KK32	Kalahari Red	Northern Cape	-28.5023	22.229	-28.4248	22.1267	1076	30181
KK41	Kalahari Red	Northern Cape	-28.5023	22.229	-28.3281	22.1486	1076	30181
KK47	Kalahari Red	Northern Cape	-28.5023	22.229	-28.3281	22.2805	1076	30181
KK51	Kalahari Red	Northern Cape	-28.5023	22.229	-28.4248	22.3683	1076	30181
KK30	Kalahari Red	Northern Cape	-28.5023	22.229	-28.4827	22.3683	1076	30181
KK34	Kalahari Red	Northern Cape	-28.5023	22.229	-28.5792	22.3683	1076	30181
KK35	Kalahari Red	Northern Cape	-28.5023	22.229	-28.4441	22.5441	1076	30181
KOOP11	Kalahari Red	Northern Cape	-28.1214	24.241	-28.1926	26.2136	1054	30594
KOOP19	Kalahari Red	Northern Cape	-28.1214	24.241	-28.0764	26.1696	1054	30594
KOOP10	Kalahari Red	Northern Cape	-28.1214	24.241	-28.1214	24.241	1054	30594
AB21	Kalahari Red	Northern Cape	-27.4633	23.4339	-27.4633	23.4339	1517	393806
AB05	Kalahari Red	Northern Cape	-27.4633	23.4339	-27.6101	23.423	1517	393806
AB26	Kalahari Red	Northern Cape	-27.4633	23.4339	-27.5711	23.2912	1517	393806
AB09	Kalahari Red	Northern Cape	-27.4633	23.4339	-27.4737	23.2253	1517	393806
AB16	Kalahari Red	Northern Cape	-27.4633	23.4339	-27.3567	23.2253	1517	393806
AB19	Kalahari Red	Northern Cape	-27.4633	23.4339	-27.2981	23.3571	1517	393806
KB40	Kalahari Red	Northern Cape	-29.0313	23.1826	-29.2332	23.2033	1271	321110

APPENDIX C: LOCATIONS (PROVINCE, GPS COORDINATES AND ALTITUDE) OF THE 194 INDIVIDUALS AND THE WEATHER STATION WHERE CLIMATIC DATA WAS COLLECTED

Sample Name	Breed/ population	Province	Original GPS coordinates		Displaced GPS coordinates		Altitude (m)	Nearest weather station
			Latitude	Longitude	Latitude	Longitude		
KB35	Kalahari Red	Northern Cape	-29.0313	23.1826	-28.926	23.3791	1271	321110
KB36	Kalahari Red	Northern Cape	-29.0313	23.1826	-29.1565	23.5571	1271	321110
KB31	Kalahari Red	Northern Cape	-29.0313	23.1826	-28.8105	23.1813	1271	321110
KB37	Kalahari Red	Northern Cape	-29.0313	23.1826	-28.8683	23.0275	1271	321110
KB32	Kalahari Red	Northern Cape	-29.0313	23.1826	-29.3482	23.2253	1271	321110
KB38	Kalahari Red	Northern Cape	-29.0313	23.1826	-29.1757	23.0715	1271	321110
KB33	Kalahari Red	Northern Cape	-29.0313	23.1826	-29.0305	23.1992	1271	321110
KB39	Kalahari Red	Northern Cape	-29.0313	23.1826	-29.0256	23.1899	1271	321110
ASH10	Nguni	KwaZulu-Natal	-29.404	30.2951	-29.404	30.2951	670	30160
CED02	Nguni	KwaZulu-Natal	-29.316	30.16	-29.316	30.16	1259	30160
CED06	Nguni	KwaZulu-Natal	-29.316	30.16	-29.3119	30.1637	1259	30160
CED07	Nguni	KwaZulu-Natal	-29.316	30.16	-29.3136	30.1641	1259	30160
CED10	Nguni	KwaZulu-Natal	-29.316	30.16	-29.3153	30.1637	1259	30160
HLU04	Nguni	KwaZulu-Natal	-28.4	32.153	-28.4	32.153	80	30534
HLU05	Nguni	KwaZulu-Natal	-28.4	32.153	-28.4	32.153	80	30534
HLU08	Nguni	KwaZulu-Natal	-28.4	32.153	-28.4	32.153	80	30534
MONZI03	Nguni	KwaZulu-Natal	-28.2221	32.245	-28.2221	32.245	28	30534
ZEN10	Nguni	KwaZulu-Natal	-28.4546	31.5338	-28.4546	31.5338	102	30534
CRA19	Savanna	Eastern Cape	-32.1598	25.621	-32.1598	25.621	1219	30387
CRA21	Savanna	Eastern Cape	-32.1598	25.621	-32.1734	26.5578	1219	30387
CRA02	Savanna	Eastern Cape	-32.1757	25.6527	-32.1757	25.6527	1501	30387
CRA07	Savanna	Eastern Cape	-32.1757	25.6527	-32.1757	25.6527	1501	30387
CRA12	Savanna	Eastern Cape	-32.1598	25.621	-32.1598	25.621	1219	30387

APPENDIX C: LOCATIONS (PROVINCE, GPS COORDINATES AND ALTITUDE) OF THE 194 INDIVIDUALS AND THE WEATHER STATION WHERE CLIMATIC DATA WAS COLLECTED

Sample Name	Breed/ population	Province	Original GPS coordinates		Displaced GPS coordinates		Altitude (m)	Nearest weather station
			Latitude	Longitude	Latitude	Longitude		
CRA14	Savanna	Eastern Cape	-32.1598	25.621	-32.1631	25.625	1219	30387
CRA25	Savanna	Eastern Cape	-32.1598	25.621	-32.1648	25.6135	1219	30387
KK31	Savanna	Northern Cape	-28.5023	22.229	-28.502	22.5002	1076	30181
KK_31	Savanna	Northern Cape	-28.5023	22.229	-28.2314	22.1926	1076	30181
KK43	Savanna	Northern Cape	-28.5023	22.229	-28.1926	22.3244	1076	30181
KK48	Savanna	Northern Cape	-28.5023	22.229	-28.5213	22.3464	1076	30181
KK28	Savanna	Northern Cape	-28.5023	22.229	-28.5599	22.3024	1076	30181
KK39	Savanna	Northern Cape	-28.5023	22.229	-28.5599	22.2145	1076	30181
KOOP37	Savanna	Northern Cape	-28.1214	24.241	-28.1345	26.1476	1054	30594
KOOP16	Savanna	Northern Cape	-28.1214	24.241	-28.1345	26.3674	1054	30594
KOOP23	Savanna	Northern Cape	-28.1214	24.241	-28.1151	26.4333	1054	30594
KOOP26	Savanna	Northern Cape	-28.1214	24.241	-28.1926	26.1476	1054	30594
KOOP24	Savanna	Northern Cape	-28.1214	24.241	-28.057	26.1257	1054	30594
KB46	Savanna	Northern Cape	-29.0313	23.1826	-28.926	23.2033	1271	321110
KB41	Savanna	Northern Cape	-29.0313	23.1826	-29.0413	23.423	1271	321110
KB47	Savanna	Northern Cape	-29.0313	23.1826	-29.0605	23.467	1271	321110
KB42	Savanna	Northern Cape	-29.0313	23.1826	-29.1181	23.423	1271	321110
KB48	Savanna	Northern Cape	-29.0313	23.1826	-29.1813	23.1813	1271	321110
KB43	Savanna	Northern Cape	-29.0313	23.1826	-28.8105	23.2693	1271	321110
KB49	Savanna	Northern Cape	-29.0313	23.1826	-29.0221	22.9177	1271	321110
KB50	Savanna	Northern Cape	-29.0313	23.1826	-29.1757	23.2912	1271	321110
KB45	Savanna	Northern Cape	-29.0313	23.1826	-29.039	23.1839	1271	321110

APPENDIX C: LOCATIONS (PROVINCE, GPS COORDINATES AND ALTITUDE) OF THE 194 INDIVIDUALS AND THE WEATHER STATION WHERE CLIMATIC DATA WAS COLLECTED

Sample Name	Breed/ population	Province	Original GPS coordinates		Displaced GPS coordinates		Altitude (m)	Nearest weather station
			Latitude	Longitude	Latitude	Longitude		
TWA30	Tankwa	Northern Cape	-30.221	22.1724	-30.2634	22.2805	1113	30731
TWA03	Tankwa	Northern Cape	-30.221	22.1724	-30.3392	22.1926	1113	30731
TWA05	Tankwa	Northern Cape	-30.221	22.1724	-30.3582	22.1267	1113	30731
TWA06	Tankwa	Northern Cape	-30.221	22.1724	-30.2634	22.0388	1113	30731
TWA07	Tankwa	Northern Cape	-30.221	22.1724	-30.1874	22.0607	1113	30731
TWA08	Tankwa	Northern Cape	-30.221	22.1724	-30.0734	22.1267	1113	30731
TWA19	Tankwa	Northern Cape	-30.221	22.1724	-30.3772	22.3464	1113	30731
TWA20	Tankwa	Northern Cape	-30.221	22.1724	-30.2064	22.3903	1113	30731
TWA21	Tankwa	Northern Cape	-30.221	22.1724	-30.5098	22.4782	1113	30731
TWA28	Tankwa	Northern Cape	-30.221	22.1724	-30.3013	22.5441	1113	30731
TWA29	Tankwa	Northern Cape	-30.221	22.1724	-30.1114	22.5222	1113	30731
TWA31	Tankwa	Northern Cape	-30.221	22.1724	-30.0544	22.3903	1113	30731
TWA32	Tankwa	Northern Cape	-30.221	22.1724	-30.5287	22.3024	1113	30731
TWA44	Tankwa	Northern Cape	-30.221	22.1724	-30.1304	21.7757	1113	30731
TWA46	Tankwa	Northern Cape	-30.221	22.1724	-30.2823	21.841	1113	30731
APPEL15	Tswana	North West	-26.136	26.52	-26.136	26.52	1502	30693
APPEL17	Tswana	North West	-26.136	26.52	-26.1599	26.4772	1502	30693
BOIK20	Tswana	North West	-26.1054	26.5442	-26.1054	26.5442	1523	30693
PELLA01	Tswana	North West	-25.285	26.2953	-25.285	26.2953	1145	30627
PELLA05	Tswana	North West	-25.2758	26.225	-25.2758	26.225	1129	30627
PELLA10	Tswana	North West	-25.249	26.2151	-25.249	26.2151	1043	30627
PELLA15	Tswana	North West	-25.257	26.2603	-25.257	26.2603	1056	30627
PELLA17	Tswana	North West	-25.257	26.2603	-25.3485	26.2603	1056	30627

APPENDIX C: LOCATIONS (PROVINCE, GPS COORDINATES AND ALTITUDE) OF THE 194 INDIVIDUALS AND THE WEATHER STATION WHERE CLIMATIC DATA WAS COLLECTED

Sample Name	Breed/ population	Province	Original GPS coordinates		Displaced GPS coordinates		Altitude (m)	Nearest weather station
			Latitude	Longitude	Latitude	Longitude		
PELLA08	Tswana	North West	-25.2196	26.2455	-25.2196	26.2455	1053	30627
PELLA14	Tswana	North West	-25.2758	26.225	-25.2758	26.225	1129	30627
TOE03	Tswana	North West	-25.2686	26.203	-28.2686	27.203	1058	30627
TOE05	Tswana	North West	-25.1801	26.2905	-25.1801	26.2905	1027	30627
TOE09	Tswana	North West	-25.1801	26.2905	-25.2691	26.3015	1027	30627
TOE13	Tswana	North West	-25.1499	26.2877	-25.1499	26.2877	1012	30627
TOE16	Tswana	North West	-25.1499	26.2861	-25.157	26.2914	1009	30627
TOE22	Tswana	North West	-25.4854	26.3497	-25.4854	26.3497	1517	30627
TOE01	Tswana	North West	-25.4854	26.3497	-25.5866	26.3674	1517	30627
TOE15	Tswana	North West	-25.4217	26.4144	-25.4217	26.4144	1360	30627
TOE19	Tswana	North West	-25.4854	26.3497	-25.4854	26.3497	1517	30627
MAL15	Venda	Limpopo	-25.4854	26.3497	-25.4817	26.3535	1517	30931
MAL03	Venda	Limpopo	-22.2405	30.3542	-22.2405	30.3542	417	30931
MADI01	Venda	Limpopo	-22.221	30.3512	-22.221	30.3512	400	30931
MADI06	Venda	Limpopo	-22.2714	30.3319	-22.2714	30.3319	437	30931
MADI07	Venda	Limpopo	-22.2714	30.3319	-22.4156	30.2346	437	30931
MAL02	Venda	Limpopo	-22.275	30.3329	-22.275	30.3329	437	30931
MAL06	Venda	Limpopo	-22.2359	30.3512	-22.2359	30.3512	409	30931
MAL11	Venda	Limpopo	-22.2359	30.3512	-22.3953	30.0808	409	30931
MPHE01	Venda	Limpopo	-23.143	29.5933	-23.143	29.5933	445	30918
MUKO04	Venda	Limpopo	-22.3247	30.334	-22.3247	30.334	502	30758
MUKO06	Venda	Limpopo	-22.3245	30.3357	-22.3245	30.3357	502	30758
MUUM01	Venda	Limpopo	-23.2265	29.5617	-23.2265	29.5617	1143	30918
MUUM07	Venda	Limpopo	-23.2212	29.5612	-23.2212	29.5612	1123	30918

APPENDIX C: LOCATIONS (PROVINCE, GPS COORDINATES AND ALTITUDE) OF THE 194 INDIVIDUALS AND THE WEATHER STATION WHERE CLIMATIC DATA WAS COLLECTED

Sample Name	Breed/ population	Province	Original GPS coordinates		Displaced GPS coordinates		Altitude (m)	Nearest weather station
			Latitude	Longitude	Latitude	Longitude		
MUUM12	Venda	Limpopo	-23.2211	29.5612	-23.2211	29.5612	1126	30918
RIBU03	Venda	Limpopo	-23.1412	30.1933	-23.1412	30.1933	532	30918
TSHI02	Venda	Limpopo	-23.1222	30.3749	-23.1222	30.3749	512	30753
TSHAND02	Venda	Limpopo	-22.3247	30.334	-22.3247	30.334	506	30931
TSHIKA11	Venda	Limpopo	-22.5143	30.383	-22.5143	30.383	491	30753
TSHIKA15	Venda	Limpopo	-23.1323	30.323	-23.1323	30.323	539	30753
TSHA04	Venda	Limpopo	-22.2932	30.4855	-22.2932	30.4855	362	30758
TSHI08	Venda	Limpopo	-22.5022	30.3752	-22.5022	30.3752	546	30753
TSHI11	Venda	Limpopo	-22.5022	30.3752	-22.578	30.3884	546	30753
VUV03	Venda	Limpopo	-23.1323	30.323	-23.1323	30.323	539	30931
VUV08	Venda	Limpopo	-22.5848	30.1655	-22.5848	30.1655	1273	30931
VUV11	Venda	Limpopo	-22.412	30.3656	-22.412	30.3656	584	30931
CRA22	Xhosa	Eastern Cape	-32.1598	25.621	-32.1598	25.621	1275	30387
CRA20	Xhosa	Eastern Cape	-32.1598	25.621	-32.2356	25.6423	1275	30387
CRA28	Xhosa	Eastern Cape	-32.1598	25.621	-32.0682	25.5764	1275	30387
HOB03	Xhosa	Eastern Cape	-32.1651	28.8692	-32.1651	28.8692	45	30447
HOB19	Xhosa	Eastern Cape	-32.1411	28.5214	-32.1411	28.5214	150	30447
HOB21	Xhosa	Eastern Cape	-32.1032	28.5354	-32.1032	28.5354	201	30447
HOB23	Xhosa	Eastern Cape	-32.1032	28.5354	-32.1798	28.5866	201	30447
HOB26	Xhosa	Eastern Cape	-32.1135	28.5624	-32.1135	28.5624	12	30447
HOB31	Xhosa	Eastern Cape	-32.1053	28.5453	-32.1053	28.5453	161	30447
HOB37	Xhosa	Eastern Cape	-32.1114	28.5444	-32.1114	28.5444	153	30447
HOB15	Xhosa	Eastern Cape	-32.2205	28.0559	-32.2205	28.0559	597	30447

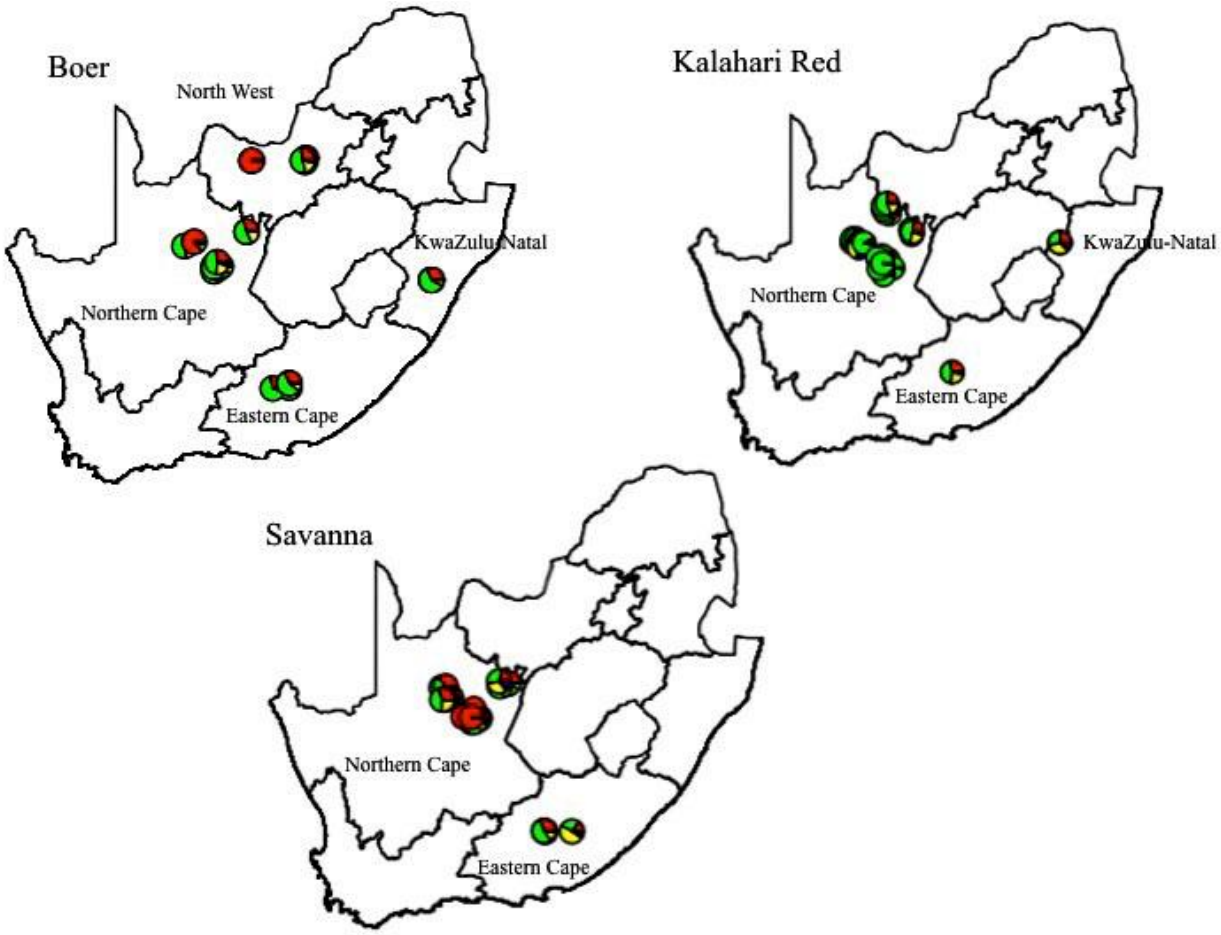
APPENDIX C: LOCATIONS (PROVINCE, GPS COORDINATES AND ALTITUDE) OF THE 194 INDIVIDUALS AND THE WEATHER STATION WHERE CLIMATIC DATA WAS COLLECTED

Sample Name	Breed/ population	Province	Original GPS coordinates		Displaced GPS coordinates		Altitude (m)	Nearest weather station
			Latitude	Longitude	Latitude	Longitude		
HOB33	Xhosa	Eastern Cape	-32.111	28.5516	-32.111	28.5516	130	30447
HOB40	Xhosa	Eastern Cape	-32.1636	28.2691	-32.1636	28.2691	498	30447
LOW04	Xhosa	Eastern Cape	-32.1733	28.2934	-32.1733	28.2934	440	30447
LOW05	Xhosa	Eastern Cape	-32.1733	28.2934	-32.1733	28.2934	440	30447
LOW07	Xhosa	Eastern Cape	-32.1733	28.2934	-32.3285	28.4548	440	30447
LOW08	Xhosa	Eastern Cape	-32.1733	28.2934	-32.2913	28.1691	440	30447
LOW10	Xhosa	Eastern Cape	-32.1733	28.2934	-32.1612	28.1472	440	30447
LOW12	Xhosa	Eastern Cape	-32.1733	28.2934	-32.1612	28.4987	440	30447
HOB01	Xhosa	Eastern Cape	-32.1651	28.8692	-32.1651	28.8692	45	30447
MAT06	Zulu	KwaZulu-Natal	-28.495	30.1858	-28.495	30.1858	1009	22477
CONI04	Zulu	KwaZulu-Natal	-28.4617	30.267	-28.4617	30.267	667	22477
CONI06	Zulu	KwaZulu-Natal	-28.4617	30.267	-28.502	30.2785	667	22477
CONI22	Zulu	KwaZulu-Natal	-28.446	30.1648	-28.446	30.1648	701	22477
CONI25	Zulu	KwaZulu-Natal	-28.4358	30.1614	-28.4358	30.1614	641	22477
GUQA01	Zulu	KwaZulu-Natal	-28.4951	30.1918	-28.4951	30.1918	1002	22477
GUQA04	Zulu	KwaZulu-Natal	-28.4943	30.1928	-28.4943	30.1928	1001	22477
GUQA08	Zulu	KwaZulu-Natal	-28.4946	30.1928	-28.5792	30.2785	1002	22477
GUQA10	Zulu	KwaZulu-Natal	-28.4953	30.1913	-28.4953	30.1913	1012	22477
IMB06	Zulu	KwaZulu-Natal	-29.3253	30.2415	-29.3253	30.2415	745	30160
IMB04	Zulu	KwaZulu-Natal	-29.3253	30.2415	-29.3291	302126	745	30160
MAT08	Zulu	KwaZulu-Natal	-28.4927	30.1831	-28.4927	30.1831	1012	22477
MAT11	Zulu	KwaZulu-Natal	-28.4921	30.1838	-28.4921	30.1838	1012	22477
MAT15	Zulu	KwaZulu-Natal	-28.495	30.1858	-28.495	30.1858	1009	22477
MAT18	Zulu	KwaZulu-Natal	-28.495	30.1858	-28.4634	30.2126	1009	22477

APPENDIX C: LOCATIONS (PROVINCE, GPS COORDINATES AND ALTITUDE) OF THE 194 INDIVIDUALS AND THE WEATHER STATION WHERE CLIMATIC DATA WAS COLLECTED

Sample Name	Breed/ population	Province	Original GPS coordinates		Displaced GPS coordinates		Altitude (m)	Nearest weather station
			Latitude	Longitude	Latitude	Longitude		
MAT20	Zulu	KwaZulu-Natal	-28.495	30.1858	-28.5792	30.1467	1009	22477
MBO04	Zulu	KwaZulu-Natal	-28.504	30.1815	-28.504	30.1815	1020	22477
MBO12	Zulu	KwaZulu-Natal	-28.5039	30.1817	-28.5039	30.1817	1072	22477
NCU09	Zulu	KwaZulu-Natal	-28.4726	30.142	-28.4726	30.142	919	22477
NCU12	Zulu	KwaZulu-Natal	-28.4726	30.142	-28.5599	30.2346	919	22477
NCU15	Zulu	KwaZulu-Natal	-28.4726	30.142	-28.4634	30.2785	919	22477
NGUBO06	Zulu	KwaZulu-Natal	-28.4358	30.2127	-28.4358	30.2127	679	22477
NGUBO10	Zulu	KwaZulu-Natal	-28.4356	30.2115	-28.4356	30.2115	706	22477
NGUBO19	Zulu	KwaZulu-Natal	-28.4344	30.212	-28.4344	30.212	682	22477
NGUBO21	Zulu	KwaZulu-Natal	-28.441	30.215	-28.441	30.215	656	22477

APPENDIX D: DISTRIBUTION OF ADMIXTURE FOR THE BOER, KALAHARI RED AND SAVANNA GOATS



Location of the South African commercial breed populations: Boer goats from North West, KwaZulu-Natal, Eastern Cape, and Northern Cape; Kalahari Red goats from Northern Cape, KwaZulu-Natal, and Eastern Cape; and Savanna goats from Northern Cape and Eastern Cape. Pie charts indicate the proportion of individuals with ancestry in up to five inferred ancestral clusters in each population.

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Altitude	3	snp6620-scaffold1236-866323	39469402	39468902-39469902	*ROR1*	receptor tyrosine kinase like orphan receptor 1	-
	3	snp25333-scaffold261-766882	32008750	32008250-32009250	*DAB1*	DAB1, reelin adaptor protein	-
	4	snp16362-scaffold1725-339300	106772042	106771542-106772542	*CNTNAP2*	contactin associated protein-like 2	Cell adhesion molecules (CAMs)
	5	snp2489-scaffold107-8760914	83787187	83786687-83787687	*GABBR2*	gamma-aminobutyric acid type B receptor subunit 2	2-Oxocarboxylic acid metabolism, Biosynthesis of amino acids, Metabolic pathways, Valine, leucine and isoleucine degradation, Valine, leucine and isoleucine biosynthesis, Cysteine and methionine metabolism, Pantothenate and CoA biosynthesis
	8	snp54499-scaffold830-4996258	63282979	63282479-63283479	*FUT8*	fucosyltransferase 8	N-Glycan biosynthesis, Metabolic pathways, Glycosaminoglycan biosynthesis - keratan sulfate, Transcriptional misregulation in cancer
	8	snp28013-scaffold300-596083	35381165	35380665-35381665	*PTPRD*	protein tyrosine phosphatase, receptor type D	-
	10	snp31070-scaffold343-2806214	25420824	25420324-25421324	*RYR3*	ryanodine receptor 3	Oxytocin signaling pathway, Calcium signaling pathway, Circadian entrainment, Salivary secretion
	10	snp16060-scaffold1690-257909	33415595	33415095-33416095	*FUT8*	fucosyltransferase 8	N-Glycan biosynthesis, Metabolic pathways, Glycosaminoglycan biosynthesis - keratan sulfate, Transcriptional misregulation in cancer

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Altitude	12	snp50210-scaffold717-5905175	26290116	26289616-26290616	-	-	-
	12	snp35965-scaffold431-2423076	48993330	48992830-48993830	-	-	-
	15	snp41697-scaffold542-2448188	48727568	48727068-48728068	*PDE2A*	phosphodiesterase 2A	Purine metabolism, cGMP-PKG signaling pathway, Olfactory transduction, Aldosterone synthesis and secretion
	15	snp41682-scaffold542-1835927	48115307	48114807-48115807	*GRAMD1B*	GRAM domain containing 1B	-
	15	snp41676-scaffold542-1367712	47647092	47646592-47647592	*SERGEF*	secretion regulating guanine nucleotide exchange factor	-
Annual minimum temperature	1	snp2702-scaffold1077-1551097	17908915	17908415-17909415	-	-	-
	1	snp26382-scaffold276-3252186	91810480	91809980-91810980	*NLGN1*	neuroligin 1	Cell adhesion molecules (CAMs)
	3	snp46893-scaffold654-1069751	33903498	33902998-33903998	*NTNG1*	netrin G1	Cell adhesion molecules (CAMs), Axon guidance
	3	snp37596-scaffold460-735979	30870653	30870153-30871153	*PROK1*	prokineticin 1	-
	3	snp37582-scaffold460-108115	31498517	31498017-31499017	*C8A*	complement C8 alpha chain	Amoebiasis, Prion diseases, Complement and coagulation cascades, Systemic lupus erythematosus

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Annual minimum temperature	3	snp25341-scaffold261-1157571	31618061	31617561-31618561	*C8B*	complement C8 beta chain	Amoebiasis, Prion diseases, Complement and coagulation cascades, Systemic lupus erythematosus
	4	snp49503-scaffold706-1271483	20477723	20477223-20478223	*DGKB*	diacylglycerol kinase beta	Glycerolipid metabolism, Glycerophospholipid metabolism, Metabolic pathways, Phosphatidylinositol signaling system, Phospholipase D signaling pathway
	4	snp16362-scaffold1725-339300	106772042	106771542-106772542	*CNTNAP2*	contactin associated protein-like 2	Cell adhesion molecules (CAMs)
	5	snp12969-scaffold15-95710	110174076	110173576-110174576	-	-	-
	9	snp3555-scaffold1110-68671	59402575	59402075-59403075	*SGK1*	serum/glucocorticoid regulated kinase 1	mTOR signaling pathway, Aldosterone-regulated sodium reabsorption, FoxO signaling pathway, PI3K-Akt signaling pathway
	10	snp43116-scaffold572-2955845	93940994	93940494-93941494	-	-	-
	12	snp35992-scaffold431-3698420	50165674	50165174-50166174	*RNF17*	ring finger protein 17	-
	14	snp34916-scaffold416-2141525	56247710	56247210-56248210	*SAMD12*	sterile alpha motif domain containing 12	-
	19	snp57768-scaffold931-563264	7530933	7530433-7531433	*LOC102185031*	musashi RNA binding protein 2	-
	29	snp14775-scaffold1596-57767	1067857	1067357-1068357	-	-	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	1	snp8057-scaffold1292-81777	24171311	24170811-24171811	-	-	-
	1	snp40542-scaffold519-1941005	78215896	78215396-78216396	*LOC102170736*	-	-
	1	snp36319-scaffold435-2778300	73496492	73495992-73496992	*FGF12*	fibroblast growth factor 12	MAPK signaling pathway, Ras signaling pathway, Rap1 signaling pathway, PI3K-Akt signaling pathway, Regulation of actin cytoskeleton
	1	snp29441-scaffold3180-133307	17625673	17625173-17626173	-	-	-
	1	snp27211-scaffold29-3020379	119057467	119056967-119057967	*AGTR1*	angiotensin II receptor type 1	Vascular smooth muscle contraction, Calcium signaling pathway, cGMP-PKG signaling pathway, Phospholipase D signaling pathway, Neuroactive ligand-receptor interaction, Adrenergic signaling in cardiomyocytes, Renin-angiotensin system
	1	snp2702-scaffold1077-1551097	17908915	17908415-17909415	-	-	-
	1	snp2700-scaffold1077-1462879	17997133	17996633-17997633	-	-	-
	1	snp26403-scaffold276-4147461	92705755	92705255-92706255	*SPATA16*	spermatogenesis associated 16	-
	1	snp26388-scaffold276-3471846	92030140	92029640-92030640	*NLGN1*	neuroligin 1	Cell adhesion molecules (CAMs)

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	1	snp26385-scaffold276-3367486	91925780	91925280-91926280	*NLGN1*	neuroligin 1	Cell adhesion molecules (CAMs)
	1	snp26382-scaffold276-3252186	91810480	91809980-91810980	*NLGN1*	neuroligin 1	Cell adhesion molecules (CAMs)
	1	snp23028-scaffold230-1726609	131464397	131463897-131464897	-	-	-
	1	snp21985-scaffold2163-176087	148657624	148657124-148658124	-	-	-
	1	snp14447-scaffold1577-466245	5049184	5048684-5049684	-	-	-
	1	snp13275-scaffold1510-59490	80111844	80111344-80112344	*MAP3K13*	mitogen-activated protein kinase 13	MAPK signaling pathway
	1	snp12149-scaffold145-736257	149143491	149142991-149143991	*COL6A6*	collagen type VI alpha 6 chain	ECM-receptor interaction, PI3K-Akt signaling pathway, Focal adhesion, Protein digestion and absorption
	1	snp11790-scaffold1437-451538	26920343	26919843-26920843	-	-	-
	2	snp8290-scaffold130-1346916	127183615	127183115-127184115	-	-	-
	2	snp8271-scaffold130-619798	126456497	126455997-126456997	*SEPN1*	selenoprotein N	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	2	snp7552-scaffold127-7361948	114185315	114184815-114185815	-	-	-
	2	snp6928-scaffold1252-766503	53795934	53795434-53796434	-	-	-
	2	snp53820-scaffold82-5619102	80868153	80867653-80868653	*LRP1B*	LDL receptor related protein 1B	-
	2	snp53804-scaffold82-4879979	81607276	81606776-81607776	-	-	
	2	snp53770-scaffold82-3369225	83118030	83117530-83118530	*LRP1B*	LDL receptor related protein 1B	-
	2	snp53692-scaffold82-52770	86434485	86433985-86434985	-	-	-
	2	snp51407-scaffold752-846826	57181104	57180604-57181604	-	-	-
	2	snp47703-scaffold670-2220263	125717904	125717404-125718404	-	-	-
	2	snp47681-scaffold670-1271342	124768983	124768483-124769483	*ZNF804A*	zinc finger protein 804A	-
	2	snp45109-scaffold614-5647948	51242109	51241609-51242609	*HECW2*	HECT, C2 and WW domain containing E3 ubiquitin protein ligase 2	-
	2	snp45073-scaffold614-4106093	49700254	49699754-49700754	*BOLL*	boule homolog, RNA binding protein	-
	2	snp45025-scaffold614-1981298	47575459	47574959-47575959	*ACTR2*	ARP2 actin related protein 2 homolog	-
	2	snp45024-scaffold614-1935094	47529255	47528755-47529755	*ORC4*	origin recognition complex subunit 4	Cell cycle
	2	snp45023-scaffold614-1891072	47485233	47484733-47485733	*ORC4*	origin recognition complex subunit 4	Cell cycle
	2	snp45014-scaffold614-1443654	47038315	47037315-47038315	*MBD5*	methyl-CpG binding domain protein 5	-
	2	snp45013-scaffold614-1411491	47005652	47005152-47006152	*SPATS2L*	spermatogenesis associated serine rich 2 like	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	2	snp40380-scaffold515-3538812	90026166	90025666-90026666	*LOC102173696*	KIAA2012 ortholog	-
	2	snp35737-scaffold430-565936	38447279	38446779-38447779	*GALNT5*	polypeptide N-acetylgalactosaminyltransferase 5	
	2	snp35269-scaffold422-2407176	12016946	12016446-12017446	-	-	-
	2	snp28963-scaffold312-1162210	64869482	64868982-64869982	*SCTR*	secretin receptor	
	2	snp20538-scaffold202-7447156	37223899	37223399-37224399	*PKP4*	plakophilin 4	-
	2	snp16987-scaffold1772-106087	103418198	103417698-103418698	-	-	-
	3	snp6620-scaffold1236-866323	39469402	39468902-39469902	*ROR1*	receptor tyrosine kinase like orphan receptor 1	-
	3	snp56493-scaffold89-902962	10304067	10303567-10304567	-	-	-
	3	snp47966-scaffold675-4312441	29546702	29546202-29547202	*USP24*	ubiquitin specific peptidase 24	-
	3	snp47961-scaffold675-4086912	29321173	29320673-29321673	*KCND3*	potassium voltage-gated channel subfamily D member 3	-
	3	snp47936-scaffold675-2937183	28171444	28170944-28171944	*GLIS1*	GLIS family zinc finger 1	-
	3	snp47924-scaffold675-2444945	27679206	27678706-27679706	*MAGI3*	membrane associated guanylate kinase, WW and PDZ domain containing 3	
	3	snp47878-scaffold675-365399	25599660	25599160-25600160	*VANGL1*	VANGL planar cell polarity protein 1	
	3	snp46920-scaffold654-2228032	35061779	35061279-35062279	*C3H1orf87*	chromosome 3 C1orf87 homolog	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	3	snp46893-scaffold654-1069751	33903498	33902998-33903998	*NTNG1*	netrin G1	-
	3	snp46892-scaffold654-1026090	33859837	33859337-33860337	*NTNG1*	netrin G1	-
	3	snp46885-scaffold654-729925	33563672	33563172-33564172	*VAV3*	vav guanine nucleotide exchange factor 3	-
	3	snp46873-scaffold654-255687	33089434	33088934-33089934	-	-	-
	3	snp43529-scaffold580-674688	30298510	30298010-30299010	-	-	-
	3	snp42150-scaffold55-4810954	61003556	61003056-61004056	*TTLL7*	tubulin tyrosine ligase like 7	-
	3	snp42148-scaffold55-4693053	60885655	60885155-60886155	-	-	-
	3	snp42143-scaffold55-4488683	60681285	60680785-60681785	-	-	-
	3	snp37600-scaffold460-923954	30682678	30682178-30683178	-	-	-
	3	snp37599-scaffold460-869294	30737338	30736838-30737838	-	-	-
	3	snp37596-scaffold460-735979	30870653	30870153-30871153	*PROK1*	prokineticin 1	-
	3	snp37587-scaffold460-329756	29853494	29852994-29853994	*STRIP1*	striatin interacting protein 1	-
	3	snp37582-scaffold460-108115	31498517	31498017-31499017	*C8A*	complement C8 alpha chain	-
	3	snp37302-scaffold455-537596	30337693	30337193-30338193	*KHDRBS2*	KH RNA binding domain containing, signal transduction associated 2	-
	3	snp25341-scaffold261-1157571	31618061	31617561-31618561	*C8B*	Complement C8 beta chain	-
	3	snp25340-scaffold261-1087491	31688141	31687641-31688641	*DAB1*	DAB1, reelin adaptor protein	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	3	snp25332-scaffold261-735786	32039846	32039346-32040346	*DAB1*	DAB1, reelin adaptor protein	-
	3	snp25327-scaffold261-524798	32250834	32250334-32251334	*WDR47*	WD repeat domain 47	-
	3	snp24972-scaffold257-48280	61724032	61723532-61724532	-	-	-
	3	snp1693-scaffold1047-2332604	52421122	52420622-52421622	-	-	-
	3	snp16924-scaffold1767-476739	11683867	11683367-11684367	-	-	-
	3	snp1691-scaffold1047-2269743	52483983	52483483-52484483	*ST6GALNAC3*	ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 3	-
	3	snp15835-scaffold1661-292564	113336914	113336414-113337414	*RAB17*	RAB17, member RAS oncogene family	-
	4	snp5880-scaffold1205-2568852	103025975	103025475-103026475	-	-	-
	4	snp5866-scaffold1205-1993299	103601528	103601028-103602028	-	-	-
	4	snp5857-scaffold1205-1567307	104027520	104027020-104028020	*NXPH1*	neurexophilin 1	-
	4	snp55964-scaffold87-1300451	23018211	23017711-23018711	*BZW2*	basic leucine zipper and W2 domains 2	-
	4	snp55951-scaffold87-764906	23553756	23553256-23554256	*CHCHD3*	coiled-coil-helix-coiled-coil-helix domain containing 3	-
	4	snp49503-scaffold706-1271483	20477723	20477223-20478223	*DGKB*	diacylglycerol kinase beta	-
	4	snp49500-scaffold706-1143916	20350156	20349656-20350656	*DGKB*	diacylglycerol kinase beta	-
	4	snp49210-scaffold701-2667997	45360621	45360121-45361121	*TNS3*	tensin 3	-
	4	snp44417-scaffold603-6413945	9709013	9708513-9709513	*CNTNAP2*	contactin associated protein-like 2	-
	4	snp44412-scaffold603-6178971	9943987	9943487-9944487	*CNTNAP2*	contactin associated protein-like 2	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	4	snp42217-scaffold552-702040	29001121	29000621-29001621	*GRM8*	glutamate metabotropic receptor 8	-
	4	snp27692-scaffold295-6146929	67415791	67415291-67416291	-	-	-
	4	snp27578-scaffold295-1171243	72391477	72390977-72391977	*TNS3*	tensin 3	-
	4	snp16362-scaffold1725-339300	106772042	106771542-106772542	*CNTNAP2*	contactin associated protein-like 2	-
	4	snp24707-scaffold2522-205328	115537623	115537123-115538123	-	-	-
	4	snp2239-scaffold1069-384905	9084076	9083576-9084576	*CNTNAP2*	contactin associated protein-like 2	-
	4	snp18348-scaffold1859-1248953	102222572	102222072-102223072	-	-	-
	5	snp54019-scaffold822-66684	6606867	6606367-6607367	*LOC108636125*	uncharacterized LOC108636125	-
	5	snp52276-scaffold775-827547	49290230	49289730-49290730	*PPM1H*	protein phosphatase, Mg2+/Mn2+ dependent 1H	-
	5	snp52273-scaffold775-705536	49168219	49167719-49168719	*PPM1H*	protein phosphatase, Mg2+/Mn2+ dependent 1H	-
	5	snp51082-scaffold740-2613701	86412830	86412330-86413330	*SLCO1C1*	solute carrier organic anion transporter family member 1C1	-
	5	snp47373-scaffold666-1167313	89574111	89573611-89574611	*DUSP16*	dual specificity phosphatase 16	-
	5	snp42484-scaffold562-1246992	58921345	58920845-58921845	*CCDC38*	coiled-coil domain containing 38	-
	5	snp38514-scaffold487-534533	639191	638691-639691	*TSPAN8*	tetraspanin 8	-
	5	snp33841-scaffold399-584264	2952043	2951543-2952543	-	-	-
	5	snp33833-scaffold399-259594	3276713	3276213-3277213	-	-	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	5	snp33542-scaffold394-770559	42491486	42490986-42491986	*PTPRB*	protein tyrosine phosphatase, receptor type B	-
	5	snp12969-scaffold15-95710	110174076	110173576-110174576	-	-	-
	6	snp926-scaffold1025-560411	94957295	94956795-94957795	-	-	-
	6	snp8927-scaffold1321-128489	114263833	114263333-114264333	*SORCS2*	sortilin related VPS10 domain containing receptor 2	-
	6	snp59985-scaffold999-128817	1438392	1437892-1438892	*MARCH-1*	membrane associated ring-CH-type finger 1	-
	6	snp5510-scaffold1193-22066	4909464	4908964-4909964	*PRDM5*	PR/SET domain 5	-
	6	snp46124-scaffold635-161069	6919777	6919277-6920277	*SEC24D*	SEC24 homolog D, COPII coat complex component	-
	6	snp46120-scaffold635-31540	6790248	6789748-6790748	*SYNPO2*	synaptopodin 2	-
	6	snp40151-scaffold511-5325929	57572037	57571537-57572537	*RBM47*	RNA binding motif protein 47	-
	6	snp31735-scaffold354-25231	25231	24731-25731	-	-	-
	6	snp31105-scaffold344-709956	8848750	8848250-8849250	-	-	-
	6	snp31096-scaffold344-349091	8487885	8487385-8488385	-	-	-
	6	snp26737-scaffold281-187160	34522041	34521541-34522541	*CCSER1*	coiled-coil serine rich protein 1	-
	6	snp12577-scaffold148-3245543	32084419	32083919-32084919	*GRID2*	glutamate ionotropic receptor delta type subunit 2	-
	6	snp23903-scaffold241-451166	47127534	47127034-47128034	-	-	-
	6	snp21377-scaffold2085-181875	9705253	9704753-9705753	-	-	-
	6	snp19366-scaffold1958-28303	8110392	8109892-8110892	-	-	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	6	snp14180-scaffold1567-155540	23571234	23570734-23571734	*LOC102181105*	alcohol dehydrogenase E chain	-
	6	snp12582-scaffold148-3413540	32252416	32251916-32252916	*GRID2*	glutamate ionotropic receptor delta type subunit 2	-
	6	snp12573-scaffold148-3041293	31880169	31879669-31880669	*GRID2*	glutamate ionotropic receptor delta type subunit 2	-
	6	snp11973-scaffold1440-206912	109051593	109051093-109052093	*LDB2*	LIM domain binding 2	-
	7	snp7370-scaffold1269-4675266	57271000	57270500-57271500	-	-	-
	7	snp7319-scaffold1269-2530432	55126166	55125666-55126666	*KCTD16*	potassium channel tetramerization domain containing 16	-
	7	snp59878-scaffold995-2691064	29621439	29620939-29621939	*PRR16*	proline rich 16	-
	7	snp59867-scaffold995-2183636	30128867	30128367-30129367	*WWC1*	WW and C2 domain containing 1	-
	7	snp59826-scaffold995-215671	32096832	32096332-32097332	-	-	-
	7	snp59822-scaffold995-11228	32301275	32300775-32301775	-	-	-
	7	snp52614-scaffold786-557000	26056006	26055506-26056506	-	-	-
	7	snp50006-scaffold716-1302734	74438604	74438104-74439104	-	-	-
	7	snp41464-scaffold54-493580	68458023	68457523-68458523	*ATP10B*	ATPase phospholipid transporting 10B (putative)	-
	7	snp30577-scaffold339-825223	42226561	42226061-42227061	*LOC108636345*	hepatitis A virus cellular receptor 1-like	-
	7	snp29797-scaffold323-1782765	66181579	66181079-66182079	*BTBD2*	BTB domain containing 2	-
	7	snp29767-scaffold323-473027	67491317	67490817-67491817	-	-	-
	7	snp29304-scaffold316-960149	19615271	19614771-19615771	-	-	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	7	snp29303-scaffold316-929807	19584929	19584429-19585429	*CDC42SE2*	CDC42 small effector 2	-
	7	snp24816-scaffold254-1177870	58771708	58771208-58772208	*SLC36A3*	proton-coupled amino acid transporter 3	-
	7	snp21537-scaffold210-1697586	103519154	103518654-103519654	-	-	-
	7	snp20069-scaffold2-2477286	89896924	89896424-89897424	*KIAA0825*	KIAA0825 ortholog	-
	8	snp59349-scaffold978-75724	27401137	27400637-27401637	*BNC2*	basonuclin 2	-
	8	snp54499-scaffold830-4996258	63282979	63282479-63283479	-	-	-
	8	snp54456-scaffold830-3227158	61513879	61513379-61514379	*ANP32B*	acidic nuclear phosphoprotein 32 family member B	-
	8	snp52645-scaffold788-39850	34257033	34256533-34257533	-	-	-
	8	snp42394-scaffold56-1217648	87933288	87932788-87933788	-	-	-
	8	snp34196-scaffold401-1557682	96780457	96779957-96780957	*EPB41L4B*	erythrocyte membrane protein band 4.1 like 4B	-
	8	snp32232-scaffold366-739084	84887498	84886998-84887998	-	-	-
	8	snp31898-scaffold356-4445540	102783778	102783278-102784278	-	-	-
	8	snp31870-scaffold356-3145635	101483873	101483373-101484373	*HSDL2*	hydroxysteroid dehydrogenase like 2	-
	8	snp31862-scaffold356-2790706	101128944	101128444-101129444	-	-	-
	8	snp28114-scaffold300-4928606	39713688	39713188-39714188	*LOC108636555*	uncharacterized LOC108636555	-
	8	snp28079-scaffold300-3453358	38238440	38237940-38238940	*JAK2*	Janus kinase 2	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	8	snp28078-scaffold300-3412212	38197294	38196794-38197794	*JAK2*	Janus kinase 2	-
	8	snp28077-scaffold300-3369813	38154895	38154395-38155395	*GLDC*	glycine decarboxylase	-
	8	snp28076-scaffold300-3335943	38121025	38120525-38121525	*GLDC*	glycine decarboxylase	-
	8	snp24577-scaffold25-698444	18426279	18425779-18426779	-	-	-
	8	snp24570-scaffold25-373025	18751698	18751198-18752198	-	-	-
	8	snp2075-scaffold1063-597215	105720942	105720442-105721442	*ASTN2*	astrotactin 2	-
	8	snp17201-scaffold1796-91323	11615229	11614729-11615729	-	-	-
	8	snp12913-scaffold1499-998881	67528652	67528152-67529152	*XPO7*	exportin 7	-
	8	snp12912-scaffold1499-953295	67574238	67573738-67574738	*NPM2*	nucleophosmin/nucleoplasmin 2	-
	8	snp12910-scaffold1499-866232	67661301	67660801-67661801	-	-	-
	9	snp50497-scaffold725-471271	65587263	65586763-65587763	-	-	-
	9	snp463-scaffold1011-1316509	19434896	19434396-19435396	*CEP85L*	centrosomal protein 85 like	-
	9	snp456-scaffold1011-1028523	19146910	19146410-19147410	*FAM184A*	family with sequence similarity 184 member A	-
	9	snp43757-scaffold588-1087718	14143970	14143470-14144470	*NKAIN2*	Na+/K+ transporting ATPase interacting 2	-
	9	snp43510-scaffold579-6303169	59516256	59515756-59516756	-	-	-
	9	snp43508-scaffold579-6223595	59595830	59595330-59596330	-	-	-
	9	snp43426-scaffold579-2901105	62918320	62917820-62918820	-	-	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	9	snp43420-scaffold579-2665013	63154412	63153912-63154912	*CCDC28A*	coiled-coil domain containing 28A	-
	9	snp43407-scaffold579-2018752	63800673	63800173-63801173	*REPS1*	RALBP1 associated Eps domain containing 1	-
	9	snp36861-scaffold447-3008093	80397341	80396841-80397841	*ZDHHC14*	zinc finger DHHC-type containing 14	-
	9	snp36854-scaffold447-2723696	80681738	80681238-80682238	*SNX9*	sorting nexin 9	-
	9	snp3556-scaffold1110-106176	59365070	59364570-59365570	-	-	-
	9	snp3555-scaffold1110-68671	59402575	59402075-59403075	*SGK1*	serum/glucocorticoid regulated kinase 1	-
	9	snp3554-scaffold1110-9525	59461721	59461221-59462221	*SGK1*	serum/glucocorticoid regulated kinase 1	-
	9	snp22963-scaffold2290-1020477	72855988	72855488-72856488	-	-	-
	9	snp16035-scaffold169-178962	8596722	8596222-8597222	-	-	-
	9	snp16034-scaffold169-145788	8629896	8629396-8630396	-	-	-
	10	snp56262-scaffold88-811488	62255902	62255402-62256402	*MDGA2*	MAM domain containing glycosylphosphatidylinositol anchor 2	-
	10	snp43087-scaffold572-1822323	92807472	92806972-92807972	-	-	-
	10	snp33153-scaffold388-1022075	37399646	37399146-37400146	-	-	-
	10	snp31485-scaffold349-1486287	39908107	39907607-39908607	-	-	-
	10	snp31472-scaffold349-911650	39333470	39332970-39333970	-	-	-
	10	snp2650-scaffold1076-3540567	88457789	88457289-88458289	*TSHR*	thyroid stimulating hormone receptor	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	10	snp16060-scaffold1690-257909	33415595	33415095-33416095	*C10H14orf105*	chromosome 3 C1orf87 homolog	-
	10	snp25145-scaffold259-3033563	56551430	56550930-56551930	*ATP8B4*	ATPase phospholipid transporting 8B4 (putative)	-
	10	snp10199-scaffold1366-531557	47110320	47109820-47110820	*UNC13C*	unc-13 homolog C	-
	11	snp6651-scaffold124-387371	52529271	52528771-52529771	-	-	-
	11	snp6057-scaffold1214-7768	4870338	4869838-4870838	*AFF3*	AF4/FMR2 family member 3	-
	11	snp52054-scaffold768-438427	55063660	55063160-55064160	*CTNNA2*	catenin alpha 2	-
	11	snp44104-scaffold598-297208	52030818	52030318-52031318	-	-	-
	11	snp39338-scaffold50-3397424	85214205	85213705-85214705	-	-	-
	11	snp33925-scaffold4-1818866	94161620	94161120-94162120	*NR6A1*	nuclear receptor subfamily 6 group A member 1	-
	11	snp33924-scaffold4-1773054	94115808	94115308-94116308	*NR6A1*	nuclear receptor subfamily 6 group A member 1	-
	11	snp33922-scaffold4-1671582	94014336	94013836-94014836	*CRB2*	crumbs 2, cell polarity complex component	-
	11	snp33896-scaffold4-444332	92787086	92786586-92787586	*STRBP*	spermatid perinuclear RNA binding protein	-
	11	snp33895-scaffold4-413151	92755905	92755405-92756405	*STRBP*	spermatid perinuclear RNA binding protein	-
	11	snp33894-scaffold4-349467	92692221	92691721-92692721	*RABGAP1*	RAB GTPase activating protein 1	-
	11	snp27512-scaffold293-1902861	10931323	10930823-10931823	*ALMS1*	ALMS1, centrosome and basal body associated protein	-
	11	snp24397-scaffold247-8994694	46347827	46347327-46348327	*IL1RN*	interleukin 1 receptor antagonist	-
	11	snp24293-scaffold247-4539140	41892273	41891773-41892773	-	-	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	11	snp20811-scaffold204-5757307	78816570	78816070-78817070	-	-	-
	11	snp20798-scaffold204-5252964	78312227	78311727-78312727	*PUM2*	pumilio RNA binding family member 2	-
	11	snp20182-scaffold20-912433	56414619	56414119-56415119	-	-	-
	11	snp18053-scaffold185-10538712	62520452	62519952-62520952	*ACTR2*	ARP2 actin related protein 2 homolog	-
	11	snp18052-scaffold185-10505563	62553601	62553101-62554101	-	-	-
	11	snp18044-scaffold185-10137014	62922150	62921650-62922650	-	-	-
	11	snp17989-scaffold185-7858813	65200351	65199851-65200851	-	-	-
	11	snp17970-scaffold185-7006873	66052291	66051791-66052791	*ARHGAP25*	Rho GTPase activating protein 25	-
	11	snp14766-scaffold1595-3825	79283834	79283334-79284334	-	-	-
	12	snp52092-scaffold77-674332	76474671	76474171-76475171	-	-	-
	12	snp50252-scaffold717-7820456	28205397	28204897-28205897	-	-	-
	12	snp50235-scaffold717-7093985	27478926	27478426-27479426	*BRCA2*	BRCA2, DNA repair associated	-
	12	snp50228-scaffold717-6783091	27168032	27167532-27168532	-	-	-
	12	snp50123-scaffold717-2243748	22628689	22628189-22629189	-	-	-
	12	snp50104-scaffold717-1520375	21905316	21904816-21905816	*LHFP*	lipoma HMGIC fusion partner	-
	12	snp36146-scaffold431-10361406	56828660	56828160-56829160	*LOC102187959*	WD repeat-containing protein 49-like	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	12	snp36140-scaffold431-10114671	56581925	56581425-56582425	*ALOX5AP*	arachidonate 5-lipoxygenase activating protein	-
	12	snp36110-scaffold431-8886228	55353482	55352982-55353982	*MTUS2*	microtubule associated tumor suppressor candidate 2	-
	12	snp36102-scaffold431-8591729	55058983	55058483-55059483	*FLT1*	fms related tyrosine kinase 1	-
	12	snp36100-scaffold431-8526387	54993641	54993141-54994141	*FLT1*	fms related tyrosine kinase 1	-
	12	snp35992-scaffold431-3698420	50165674	50165174-50166174	*RNF17*	ring finger protein 17	-
	12	snp35989-scaffold431-3552989	50020243	50019743-50020743	-	-	-
	12	snp35982-scaffold431-3206404	49673658	49673158-49674158	*UCHL3*	ubiquitin C-terminal hydrolase L3	-
	12	snp35967-scaffold431-2526076	48993330	48992830-48993830	-	-	-
	12	snp35965-scaffold431-2423076	48993330	48992830-48993830	-	-	-
	12	snp24107-scaffold246-144636	17915621	17915121-17916121	*CAB39L*	calcium binding protein 39 like	-
	13	snp7725-scaffold1278-1447019	67358729	67358229-67359229	*DHX35*	DEAH-box helicase 35	-
	13	snp5930-scaffold1209-402211	62943504	62943004-62944004	*CPNE1*	copine 1	-
	13	snp49071-scaffold7-4626335	27765659	27765159-27766159	*FRMD4A*	FERM domain containing 4A	-
	13	snp49066-scaffold7-4441201	27950793	27950293-27951293	*FRMD4A*	FERM domain containing 4A	-
	13	snp49005-scaffold7-1739012	30652982	30652482-30653482	*CUBN*	cubilin	-
	13	snp48983-scaffold7-757999	31633995	31633495-31634495	*CACNB2*	calcium voltage-gated channel auxiliary subunit beta 2	-
	13	snp48968-scaffold7-86243	32305751	32305251-32306251	*CACNB2*	calcium voltage-gated channel auxiliary subunit beta 2	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	13	snp48664-scaffold691-1156130	59355431	59354931-59355931	*TTLL9*	tubulin tyrosine ligase like 9	-
	13	snp39574-scaffold503-1642278	20498592	20498092-20499092	*PLXDC2*	plexin domain containing 2	-
	13	snp39569-scaffold503-1404542	20260856	20260356-20261356	*PLXDC2*	plexin domain containing 2	-
	13	snp32526-scaffold371-1443673	39230563	39230063-39231063	*CFAP61*	cilia and flagella associated protein 61	-
	13	snp27385-scaffold291-712138	642636	642136-643136	*PLCB1*	phospholipase C beta 1	-
	13	snp27383-scaffold291-642041	712733	712233-713233	*PLCB1*	phospholipase C beta 1	-
	13	snp23785-scaffold24-2232462	25323073	25322573-25323573	*MYO3A*	myosin IIIA	-
	13	snp23762-scaffold24-1242386	24332997	24332497-24333497	*ARHGAP21*	Rho GTPase activating protein 21	-
	13	snp23758-scaffold24-1057954	24148565	24148065-24149065	*KIAA1217*	KIAA1217 ortholog	-
	13	snp13054-scaffold150-3302650	49217446	49216946-49217946	*ADRA1D*	adrenoceptor alpha 1D	-
	13	snp13050-scaffold150-3127888	49042684	49042184-49043184	-	-	-
	14	snp58863-scaffold963-342613	34432285	34431785-34432785	*EIF3H*	eukaryotic translation initiation factor 3 subunit H	-
	14	snp52812-scaffold791-2229348	54880060	54879560-54880560	*PAG1*	phosphoprotein membrane anchor with glycosphingolipid microdomains 1	-
	14	snp51508-scaffold755-253634	60236242	60235742-60236742	-	-	-
	14	snp51207-scaffold75-2258261	38866780	38866280-38867280	*MRPS28*	mitochondrial ribosomal protein S28	-
	14	snp49570-scaffold708-265571	92042329	92041829-92042829	*SNTB1*	syntrophin beta 1	-
	14	snp42662-scaffold566-3573540	77002637	77002137-77003137	-	-	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	14	snp42661-scaffold566-3541454	76970551	76970051-76971051	-	-	-
	14	snp35377-scaffold425-795259	72578630	72578130-72579130	-	-	-
	14	snp34916-scaffold416-2141525	56247710	56247210-56248210	*SAMD12*	sterile alpha motif domain containing 12	-
	14	snp34916-scaffold416-2141525	56247710	56247210-56248210	*SAMD12*	sterile alpha motif domain containing 12	-
	14	snp23169-scaffold231-1260076	62271212	62270712-62271712	*CSMD3*	CUB and sushi domain-containing protein 3	-
	14	snp23156-scaffold231-637375	61648511	61648011-61649011	*CSMD3*	CUB and sushi domain-containing protein 3	-
	14	snp23145-scaffold231-83906	61095042	61094542-61095542	-	-	-
	14	snp1838-scaffold1052-332190	68809244	68808744-68809744	-	-	-
	14	snp13420-scaffold1518-983083	41365818	41365318-41366318	*MYBL1*	MYB proto-oncogene like 1	-
	14	snp13419-scaffold1518-947027	41401874	41401374-41402374	*MYBL1*	MYB proto-oncogene like 1	-
	14	snp13418-scaffold1518-891881	41457020	41456520-41457520	-	-	-
	14	snp12355-scaffold1466-277214	65483444	65482944-65483944	*NUDCD1*	NudC domain containing 1	-
	14	snp11096-scaffold14-1013966	70479832	70479332-70480332	*LRP12*	LDL receptor related protein 12	-
	15	snp85-scaffold100-265663	5337309	5336809-5337809	*BIRC2*	baculoviral IAP repeat containing 2	-
	15	snp54753-scaffold837-3247168	53654642	53654142-53655142	-	-	-
	15	snp42976-scaffold57-5665689	57597698	57597198-57598198	-	-	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	15	snp42948-scaffold57-4359751	58903636	58903136-58904136	*IMMP1L*	inner mitochondrial membrane peptidase subunit 1	-
	15	snp41697-scaffold542-2448188	48727568	48727068-48728068	*PDE2A*	phosphodiesterase 2A	-
	15	snp41682-scaffold542-1835927	48115307	48114807-48115807	*LOC102169978*	short transient receptor potential channel 2	-
	15	snp40425-scaffold516-971511	73156471	73155971-73156971	*C15H11orf49*	chromosome 15 C11orf49 homolog	-
	15	snp36763-scaffold445-3253592	31613500	31613000-31614000	-	-	-
	15	snp34632-scaffold41-1222140	17526001	17525501-17526501	*APIP*	APAF1 interacting protein	-
	15	snp32464-scaffold37-1900491	13881515	13881015-13882015	-	-	-
	15	snp32456-scaffold37-1565950	13546974	13546474-13547474	-	-	-
	15	snp30545-scaffold338-88317	72096544	72096044-72097044	-	-	-
	15	snp26596-scaffold277-2628547	18301481	18300981-18301981	-	-	-
	15	snp16007-scaffold1685-115772	21045899	21045399-21046399	-	-	-
	15	snp1136-scaffold103-5440	51432609	51432109-51433109	*ARHGEF12*	Rho guanine nucleotide exchange factor 12	-
	15	snp11206-scaffold1400-582048	64505726	64505226-64506226	*LOC102190533*	protein NPAT	-
	15	snp106-scaffold100-1313915	6385561	6385061-6386061	*ATG13*	autophagy related 13	-
	16	snp8633-scaffold131-2395896	56021781	56021281-56022281	*BRINP2*	BMP/retinoic acid inducible neural specific 2	-
	16	snp53577-scaffold815-1571965	5628345	5627845-5628845	*KCNT2*	potassium sodium-activated channel subfamily T member 2	-
	16	snp52590-scaffold785-614809	72027141	72026641-72027641	-	-	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	16	snp3727-scaffold112-2830032	61247808	61247308-61248308	*LOC102184834*	voltage-dependent R-type calcium channel subunit alpha-1E	-
	16	snp30134-scaffold331-518985	50526622	50526122-50527122	*LOC102174414*	chymotrypsin-like elastase family member 2A	-
	16	snp20913-scaffold205-2784009	44251791	44251291-44252291	*CAMTA1*	calmodulin binding transcription activator 1	-
	16	snp20859-scaffold205-345680	46690120	46689620-46690620	-	-	-
	16	snp14833-scaffold1599-1648545	48916418	48915918-48916918	*PLCH2*	phospholipase C eta 2	-
	16	snp14831-scaffold1599-1542713	48810586	48810086-48811086	*UBE2J2*	ubiquitin conjugating enzyme E2 J2	-
	16	snp14830-scaffold1599-1502937	48770810	48770310-48771310	*MMEL1*	membrane metalloendopeptidase like 1	-
	16	snp14828-scaffold1599-1382096	48649969	48649469-48650469	*LOC102189890*	ATPase family AAA domain-containing protein 3	-
	17	snp9926-scaffold1354-388856	68505280	68504780-68505780	*MTMR3*	myotubularin related protein 3	-
	17	snp6623-scaffold1238-69058	21288850	21288350-21289350	-	-	-
	17	snp5909-scaffold1206-678477	56393792	56393292-56394292	*SUDS3*	SDS3 homolog, SIN3A corepressor complex component	-
	17	snp53410-scaffold806-1045491	3062204	3061704-3062704	*DCHS2*	dachsous cadherin-related 2	-
	17	snp53404-scaffold806-807118	2823831	2823331-2824331	*PLRG1*	pleiotropic regulator 1	-
	17	snp44764-scaffold609-1510371	51120743	51120243-51121243	-	-	-
	17	snp43551-scaffold581-835766	34040957	34040457-34041457	*FSTL5*	follistatin like 5	-
	17	snp43537-scaffold581-227197	33432388	33431888-33432888	-	-	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	17	snp4007-scaffold1126-1730322	47901104	47900604-47901604	*SLC15A4*	solute carrier family 15 member 4	-
	17	snp3984-scaffold1126-701621	46872403	46871903-46872903	*TMEM132D*	transmembrane protein 132D	-
	17	snp35576-scaffold428-3304453	31355361	31354861-31355861	*RAPGEF2*	Rap guanine nucleotide exchange factor 2	-
	17	snp21117-scaffold207-433405	19897858	19897358-19898358	*NCOR2*	nuclear receptor corepressor 2	-
	17	snp19848-scaffold199-6393178	30162418	30161918-30162918	-	-	-
	17	snp16564-scaffold1747-257434	30649427	30648927-30649927	*RXFP1*	relaxin/insulin like family peptide receptor 1	-
	18	snp872-scaffold1023-930759	12456705	12456205-12457205	-	-	-
	18	snp40233-scaffold512-3318859	18458116	18457616-18458616	*NKD1*	naked cuticle homolog 1	-
	18	snp37912-scaffold467-1837460	30239750	30239250-30240250	-	-	-
	18	snp37896-scaffold467-1108407	29510697	29510197-29511197	-	-	-
	18	snp35658-scaffold43-2147682	22003155	22002655-22003655	-	-	-
	18	snp35658-scaffold43-2147682	22003155	22002655-22003655	-	-	-
	18	snp35656-scaffold43-2066150	22084687	22084187-22085187	-	-	-
	18	snp34670-scaffold410-763690	40720443	40719943-40720943	-	-	-
	18	snp14599-scaffold1589-2069848	26220784	26220284-26221284	*NUP93*	nucleoporin 93	-
	18	snp14590-scaffold1589-1713280	25864216	25863716-25864716	*GOT2*	glutamic-oxaloacetic transaminase 2	-
	19	snp9199-scaffold1335-368697	44656775	44656275-44657275	*NSF*	N-ethylmaleimide sensitive factor, vesicle fusing ATPase	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	19	snp57768-scaffold931-563264	7530933	7530433-7531433	*MSI2*	musashi RNA binding protein 2	-
	19	snp4870-scaffold1163-240403	16158606	16158106-16159106	*ASIC2*	acid sensing ion channel subunit 2	-
	19	snp4642-scaffold1149-1019159	39810131	39809631-39810631	*IKZF3*	IKAROS family zinc finger 3	-
	19	snp45201-scaffold616-880756	4319236	4318736-4319736	-	-	-
	19	snp3320-scaffold1101-846281	54256577	54256077-54257077	*SRSF2*	serine and arginine rich splicing factor 2	-
	19	snp3314-scaffold1101-583094	53993390	53992890-53993890	-	-	-
	19	snp32202-scaffold3643-393290	49794943	49794443-49795443	*DCXR*	L-xylulose reductase	-
	19	snp28441-scaffold303-3404356	32926692	32926192-32927192	*AKAP10*	A-kinase anchoring protein 10	-
	19	snp26095-scaffold2716-301565	50829289	50828789-50829789	*SLC38A10*	solute carrier family 38 member 10	-
	19	snp23097-scaffold2300-636007	51766960	51766460-51767460	*TBC1D16*	TBC1 domain family member 16	-
	19	snp1617-scaffold1046-887547	15030557	15030057-15031057	-	-	-
	20	snp58535-scaffold952-1559110	61060801	61060301-61061301	*CTNND2*	catenin delta 2	-
	20	snp58534-scaffold952-1528166	61029857	61029357-61030357	*CTNND2*	catenin delta 2	-
	20	snp57352-scaffold913-728668	7348854	7348354-7349354	-	-	-
	20	snp56495-scaffold89-972939	7348854	7348354-7349354	-	-	-
	20	snp52477-scaffold782-1503572	4804680	4804180-4805180	*CREBRF*	CREB3 regulatory factor	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	20	snp50994-scaffold74-20766	16141935	16141435-16142435	-	-	-
	20	snp45412-scaffold62-3233508	30443997	30443497-30444497	*FGF10*	fibroblast growth factor 10	-
	20	snp25966-scaffold269-90513	53619704	53619204-53620204	*CDH18*	cadherin 18	-
	20	snp14653-scaffold1591-11752	10250412	10249912-10250912	*OCLN*	177ccluding	-
	20	snp14320-scaffold1570-657183	57915104	57914604-57915604	*ANKH*	ANKH inorganic pyrophosphate transport regulator	-
	20	snp10071-scaffold1357-240695	6549046	6548546-6549546	-	-	-
	21	snp7155-scaffold1265-463334	17301758	17301258-17302258	*NTRK3*	neurotrophic receptor tyrosine kinase 3	-
	21	snp6514-scaffold1230-1489811	30150934	30150434-30151434	*PEAK1*	pseudopodium enriched atypical kinase 1	-
	21	snp50356-scaffold72-2158060	9924915	9924415-9925415	*LOC106503341*	uncharacterised LOC106503341	-
	21	snp46093-scaffold633-653652	24028150	24027650-24028650	-	-	-
	21	snp46088-scaffold633-451511	23826009	23825509-23826509	*MTHFS*	5-formyltetrahydrofolate cyclo-ligase	-
	21	snp39943-scaffold51-2086104	43464109	43463609-43464609	*RALGAPA1*	Ral GTPase activating protein catalytic alpha subunit 1	-
	21	snp38817-scaffold492-2285132	39092774	39092274-39093274	*STRN3*	striatin 3	-
	21	snp25956-scaffold2688-192534	65987348	65986848-65987848	*PPP2R5C*	protein phosphatase 2 regulatory subunit B'gamma	-
	21	snp21354-scaffold2084-1007300	63014390	63013890-63014890	-	-	-
	21	snp15688-scaffold165-4105460	50381469	50380969-50381969	-	-	-
	21	snp15609-scaffold165-793522	53693407	53692907-53693907	*TRIP11*	thyroid hormone receptor interactor 11	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	21	snp15594-scaffold165-138712	54348217	54347717-54348717	*LGMN*	legumain	-
	21	snp12283-scaffold1460-498799	58314366	58313866-58314866	*VRK1*	vaccinia related kinase 1	-
	22	snp56003-scaffold870-663942	56622506	56622006-56623006	*HMCES*	5-hydroxymethylcytosine (hmC) binding, ES cell-specific	-
	22	snp56001-scaffold870-565726	56524290	56523790-56524790	*MKRN2*	makorin ring finger protein 2	-
	22	snp15362-scaffold163-3429800	37611968	37611468-37612468	-	-	-
	23	snp49917-scaffold714-724493	10343294	10342794-10343794	*CPNE5*	copine 5	-
	23	snp46501-scaffold643-3449415	43936152	43935652-43936652	-	-	-
	23	snp46487-scaffold643-2935769	44449798	44449298-44450298	*COL21A1*	collagen type XXI alpha 1 chain	-
	23	snp39703-scaffold505-1556293	35190449	35189949-35190949	-	-	-
	23	snp39147-scaffold499-154126	44742332	44741832-44742832	-	-	-
	23	snp20298-scaffold2009-76962	48513436	48512936-48513936	*KHDRBS2*	KH RNA binding domain containing, signal transduction associated 2	-
	24	snp9698-scaffold1349-441591	5749814	5749314-5750314	-	-	-
	24	snp9696-scaffold1349-373329	5681552	5681052-5682052	-	-	-
	24	snp8182-scaffold13-555106	18338227	18337727-18338727	-	-	-
	24	snp7678-scaffold1277-988620	48734090	48733590-48734590	*DYM*	dymeclin	-
	24	snp663-scaffold102-1303046	20196478	20195978-20196978	*KIAA1328*	KIAA1328 ortholog	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	24	snp661-scaffold102-1223305	20116737	20116237-20117237	*KIAA1328*	KIAA1328 ortholog	-
	24	snp6563-scaffold1232-996159	7911275	7910775-7911775	*DOK6*	docking protein 6	-
	24	snp6559-scaffold1232-789738	8117696	8117196-8118196	-	-	-
	24	snp648-scaffold102-635295	19528727	19528227-19529227	-	-	-
	24	snp48317-scaffold685-528297	10234688	10234188-10235188	*CDH19*	cadherin 19	-
	24	snp46743-scaffold651-237398	9726770	9726270-9727270	-	-	-
	24	snp40916-scaffold526-2832048	45785746	45785246-45786246	*PSTPIP2*	proline-serine-threonine phosphatase interacting protein 2	-
	24	snp40884-scaffold526-1598069	44551767	44551267-44552267	*SETBP1*	SET binding protein 1	-
	24	snp34966-scaffold417-1322438	55302143	55301643-55302643	-	-	-
	24	snp34952-scaffold417-632509	54612214	54611714-54612714	*TCF4*	transcription factor 4	-
	24	snp34812-scaffold4144-157444	1251905	1251405-1252405	-	-	-
	24	snp27537-scaffold294-67724	7063423	7062923-7063923	-	-	-
	24	snp24999-scaffold2570-5047	909161	908661-909661	*ATP9B*	ATPase phospholipid transporting 9B (putative)	-
	24	snp20284-scaffold2008-357581	357581	357081-358081	-	-	-
	24	snp19090-scaffold1919-413014	53934222	53933722-53934722	-	-	-
	24	snp13162-scaffold1502-321603	47347699	47347199-47348199	-	-	-
	25	snp50794-scaffold734-2536448	26188006	26187506-26188506	*SULT1A1*	sulfotransferase 1A1	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	25	snp43039-scaffold570-1733961	7503657	7503157-7504157	*ABAT*	4-aminobutyrate aminotransferase	-
	25	snp38361-scaffold485-1646273	15771279	15770779-15771779	*XYLT1*	xylosyltransferase 1	-
	25	snp26909-scaffold2848-602248	38478038	38477538-38478538	-	-	-
	25	snp26742-scaffold281-373269	38478038	38477538-38478538	*ABCG2*	ATP binding cassette subfamily G member 2	-
	25	snp23682-scaffold239-1377449	22274010	22273510-22274510	-	-	-
	25	snp13492-scaffold1520-312869	18059682	18059182-18060182	-	-	-
	25	snp13490-scaffold1520-232171	17978984	17978484-17979484	*GP2*	glycoprotein 2	-
	25	snp12269-scaffold146-381647	28364139	28363639-28364639	*SBDS*	SBDS ribosome assembly guanine nucleotide exchange factor	-
	25	snp11659-scaffold143-1510888	33556248	33555748-33556748	*WBSCR22*	Williams-Beuren syndrome chromosome region 22	-
	25	snp11658-scaffold143-1470688	33516048	33515548-33516548	*BAZ1B*	bromodomain adjacent to zinc finger domain 1B	-
	25	snp11629-scaffold143-122235	32167595	32167095-32168095	-	-	-
	26	snp6239-scaffold122-811967	11805464	11804964-11805964	-	-	-
	26	snp47499-scaffold67-132575	35836843	35836343-35837343	*PLCE1*	phospholipase C epsilon 1	-
	26	snp41147-scaffold532-2467263	19263257	19262757-19263757	*DNMBP*	dynamin binding protein	-
	26	snp41126-scaffold532-1569400	20161120	20160620-20161620	*SEMA4G*	Semaphoring 4G	-
	26	snp41113-scaffold532-1051457	20679063	20678563-20679563	*BTRC*	beta-transducin repeat containing E3 ubiquitin protein ligase	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	26	snp41112-scaffold532-1001039	20729481	20728981-20729981	*DPCD*	deleted in primary ciliary dyskinesia homolog (mouse)	-
	26	snp30498-scaffold3363-112737	44763296	44762796-44763796	-	-	-
	26	snp19919-scaffold1992-2251393	37506837	37506337-37507337	-	-	-
	26	snp19915-scaffold1992-2074589	37683641	37683141-37684141	*EIF3A*	eukaryotic translation initiation factor 3 subunit A	-
	26	snp11757-scaffold1434-109091	4359713	4359213-4360213	*PCDH15*	protocadherin related 15	-
	26	snp11246-scaffold1403-97874	3342966	3342466-3343466	-	-	-
	26	snp1028-scaffold1027-781724	40054531	40054031-40055031	*PANK1*	pantothenate kinase 1	-
	27	snp7025-scaffold126-75677	31805712	31805212-31806212	*TENM3*	teneurin transmembrane protein 3	-
	27	snp59597-scaffold984-144768	1503838	1503338-1504338	*ZNF385D*	zinc finger protein 385D	-
	27	snp51865-scaffold762-1650776	39295283	39294783-39295783	-	-	-
	27	snp44435-scaffold604-404683	10876282	10875782-10876782	*LOC102186109*	disintegrin and metalloproteinase domain-containing protein 18-like	-
	27	snp30755-scaffold34-2405052	4053757	4053257-4054257	*THRB*	thyroid hormone receptor beta	-
	27	snp30708-scaffold34-595928	2244633	2244133-2245133	*CSMD1*	CUB and Sushi multiple domains 1	-
	27	snp26280-scaffold275-2529043	16999410	16998910-16999910	*NRG1*	neuregulin 1	-
	27	snp18611-scaffold1881-252963	33614189	33613689-33614689	*IDO2*	indoleamine 2,3-dioxygenase 2	-
	28	snp54839-scaffold838-3141592	14895606	14895106-14896106	*KAT6B*	lysine acetyltransferase 6B	-
	28	snp54811-scaffold838-2001811	13755825	13755325-13756325	*ANK3*	ankyrin 3	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Longitude	28	snp54782-scaffold838-778571	12532585	12532085-12533085	*KCNMA1*	potassium calcium-activated channel subfamily M alpha 1	-
	28	snp39803-scaffold509-2841876	29931763	29931263-29932263	*C28H10orf11*	chromosome 28 C10orf11 homolog	-
	28	snp12239-scaffold1457-237406	937357	936857-937857	-	-	-
	28	snp1044-scaffold1028-386572	4408220	4407720-4408720	-	-	-
	29	snp54098-scaffold825-43831	24064114	24063614-24064614	*NAV2*	neuron navigator 2	-
	29	snp17565-scaffold182-204098	44041920	44041420-44042420	*CPT1A*	carnitine palmitoyltransferase 1A	-
	29	snp16197-scaffold1703-381258	39523439	39522939-39523939	-	-	-
	29	snp16195-scaffold1703-234520	39376701	39376201-39377201	-	-	-
Winter minimum temperature	1	snp8057-scaffold1292-81777	24171311	24170811-24171811	-	-	-
	1	snp48942-scaffold699-17974	134122869	134122369-134123369	-	-	-
	1	snp29441-scaffold3180-133307	17625673	17625173-17626173	-	-	-
	1	snp2702-scaffold1077-1551097	17908915	17908415-17909415	-	-	-
	1	snp2700-scaffold1077-1462879	17997133	17996633-17997633	-	-	-
	1	snp26388-scaffold276-3471846	92030140	92029640-92030640	*NLGN1*	neuroligin 1	-
	1	snp26382-scaffold276-3252186	91810480	91809980-91810980	*NLGN1*	neuroligin 1	-
Variable	**CHX**	**Associated SNP**	**SNP**	**Gene region**	**Associated**	**Description**	**KEGG pathway**

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

			position		gene		
Winter minimum temperature	1	snp21985-scaffold2163-176087	148657624	148657124-148658124	-	-	-
	1	snp11790-scaffold1437-451538	26920343	26919843-26920843	-	-	-
	2	snp8271-scaffold130-619798	126456497	126455997-126456997	*SEPN1*	selenoprotein N	-
	2	snp45014-scaffold614-1443654	47037815	47037315-47038315	*SPATS2L*	spermatogenesis associated serine rich 2 like	-
	2	snp45013-scaffold614-1411491	47005652	47005152-47006152	*SPATS2L*	spermatogenesis associated serine rich 2 like	-
	2	snp28321-scaffold3010-324016	134070125	134069625-134070625	-	-	-
	2	snp17523-scaffold1811-460063	134561159	134560659-134561659	*PADI3*	peptidyl arginine deiminase 3	-
	3	snp6620-scaffold1236-866323	39469402	39468902-39469902	*ROR1*	receptor tyrosine kinase like orphan receptor 1	-
	3	snp47878-scaffold675-365399	25599660	25599160-25600160	*VANGL1*	VANGL planar cell polarity protein 1	-
	3	snp46920-scaffold654-2228032	35061779	35061279-35062279	-	-	-
	3	snp46893-scaffold654-1069751	33903498	33902998-33903998	*NTNG1*	netrin G1	-
	3	snp46892-scaffold654-1026090	33859837	33859337-33860337	*NTNG1*	netrin G1	-
	3	snp46885-scaffold654-729925	33563672	33563172-33564172	*VAV3*	vav guanine nucleotide exchange factor 3	-
	3	snp46873-scaffold654-255687	33089434	33088934-33089934	-	-	-
	3	snp42087-scaffold55-2111002	58303604	58303104-58304104	-	-	-
	3	snp37596-scaffold460-735979	30870653	30870153-30871153	*PROK1*	prokineticin 1	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Winter minimum temperature	3	snp37582-scaffold460-108115	31498517	31498017-31499017	*C8A*	complement C8 alpha chain	-
	3	snp25341-scaffold261-1157571	31618061	31617561-31618561	*C8B*	complement C8 beta chain	-
	3	snp25333-scaffold261-766882	32008750	32008250-32009250	*DAB1*	DAB1, reelin adaptor protein	-
	3	snp25327-scaffold261-524798	32250834	32250334-32251334	*WDR47*	WD repeat domain 47	-
	3	snp17502-scaffold181-510286	96132466	96131966-96132966	*LOC102175263*	vitamin D3 hydroxylase-associated protein	-
	3	snp1693-scaffold1047-2332604	52421122	52420622-52421622	-	-	-
	3	snp1691-scaffold1047-2269743	52483983	52483483-52484483	*ST6GALNAC3*	ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 3	-
	3	snp16222-scaffold1704-1001411	102628534	102628034-102629034	*NT5C1A*	5'-nucleotidase, cytosolic IA	-
	4	snp5880-scaffold1205-2568852	103025975	103025475-103026475	-	-	-
	4	snp5866-scaffold1205-1993299	103601528	103601028-103602028	-	-	-
	4	snp55951-scaffold87-764906	23553756	23553256-23554256	*CHCHD3*	coiled-coil-helix-coiled-coil-helix domain containing 3	-
	4	snp49503-scaffold706-1271483	20477723	20477223-20478223	*DGKB*	diacylglycerol kinase beta	-
	4	snp49500-scaffold706-1143916	20350156	20349656-20350656	*DGKB*	diacylglycerol kinase beta	-
	4	snp27578-scaffold295-1171243	72391477	72390977-72391977	*TNS3*	tensin 3	-
	4	snp16362-scaffold1725-339300	106772042	106771542-106772542	*CNTNAP2*	contactin associated protein-like 2	-
	4	snp24707-scaffold2522-205328	115537623	115537123-115538123	-	-	-
	4	snp2239-scaffold1069-384905	9084076	9083576-9084576	*CNTNAP2*	contactin associated protein-like 2	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Winter minimum temperature	5	snp52273-scaffold775-705536	49168219	49167719-49168719	*PPM1H*	protein phosphatase, Mg2+/Mn2+ dependent 1H	-
	5	snp47373-scaffold666-1167313	89574111	89573611-89574611	*DUSP16*	dual specificity phosphatase 16	-
	5	snp12969-scaffold15-95710	110174076	110173576-110174576	-	-	-
	6	snp926-scaffold1025-560411	94957295	94956795-94957795	-	-	-
	6	snp48555-scaffold690-432877	104631767	104631267-104632267	-	-	-
	6	snp31105-scaffold344-709956	8848750	8848250-8849250	-	-	-
	6	snp31096-scaffold344-349091	8487885	8487385-8488385	-	-	-
	6	snp26737-scaffold281-187160	34522041	34521541-34522541	*HERC6*	HECT and RLD domain containing E3 ubiquitin protein ligase family member 6	-
	6	snp12577-scaffold148-3245543	32084419	32083919-32084919	*GRID2*	glutamate ionotropic receptor delta type subunit 2	-
	6	snp14180-scaffold1567-155540	23571234	23570734-23571734	*LOC102181105*	alcohol dehydrogenase E chain	-
	7	snp7370-scaffold1269-4675266	57271000	57270500-57271500	-	-	-
	7	snp7319-scaffold1269-2530432	55126166	55125666-55126666	*JAKMIP2*	janus kinase and microtubule interacting protein 2	-
	7	snp7317-scaffold1269-2459871	55055605	55055105-55056105	*KCTD16*	potassium channel tetramerization domain containing 16	-
	7	snp50006-scaffold716-1302734	74438604	74438104-74439104	-	-	-
	7	snp30577-scaffold339-825223	42226561	42226061-42227061	*GNA15*	G protein subunit alpha 15	-
	7	snp29303-scaffold316-929807	19584929	19584429-19585429	*CDC42SE2*	CDC42 small effector 2	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Winter minimum temperature	7	snp21537-scaffold210-1697586	103519154	103518654-103519654	-	-	-
	7	snp20069-scaffold2-2477286	89896924	89896424-89897424	*KIAA0825*	KIAA0825 ortholog	-
	8	snp54499-scaffold830-4996258	63282979	63282479-63283479	*GABBR2*	gamma-aminobutyric acid type B receptor subunit 2	-
	8	snp42394-scaffold56-1217648	87933288	87932788-87933788	-	-	-
	8	snp31898-scaffold356-4445540	102783778	102783278-102784278	-	-	-
	8	snp31870-scaffold356-3145635	101483873	101483373-101484373	*HSDL2*	hydroxysteroid dehydrogenase like 2	-
	8	snp31862-scaffold356-2790706	101128944	101128444-101129444	-	-	-
	8	snp28114-scaffold300-4928606	39713688	39713188-39714188	*LOC108636555*	uncharacterised LOC108636555	-
	8	snp28079-scaffold300-3453358	38238440	38237940-38238940	*UHRF2*	ubiquitin like with PHD and ring finger domains 2	-
	8	snp1000-scaffold1026-533890	68958341	68957841-68958841	*XPO7*	exportin 7	-
	8	snp24570-scaffold25-373025	18751698	18751198-18752198	-	-	-
	8	snp12910-scaffold1499-866232	67661301	67660801-67661801	-	-	-
	9	snp43510-scaffold579-6303169	59516256	59515756-59516756	-	-	-
	9	snp3556-scaffold1110-106176	59365070	59364570-59365570	-	-	-
	9	snp3555-scaffold1110-68671	59402575	59402075-59403075	*SGK1*	serum/glucocorticoid regulated kinase 1	-
	10	snp43087-scaffold572-1822323	92807472	92806972-92807972	*PDE8B*	phosphodiesterase 8B	-
	10	snp16060-scaffold1690-257909	33415595	33415095-33416095	*C10H14orf105*	chromosome 3 C1orf87 homolog	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Winter minimum temperature	11	snp6057-scaffold1214-7768	4870338	4869838-4870838	*AFF3*	AF4/FMR2 family member 3	-
	11	snp52054-scaffold768-438427	55063660	55063160-55064160	*CTNNA2*	catenin alpha 2	-
	11	snp44104-scaffold598-297208	52030818	52030318-52031318	*CTNNA2*	catenin alpha 2	-
	11	snp33925-scaffold4-1818866	94161620	94161120-94162120	*NR6A1*	nuclear receptor subfamily 6 group A member 1	-
	11	snp33924-scaffold4-1773054	94115808	94115308-94116308	*NR6A1*	nuclear receptor subfamily 6 group A member 1	-
	11	snp33922-scaffold4-1671582	94014336	94013836-94014836	*CRB2*	crumbs 2, cell polarity complex component	-
	11	snp27512-scaffold293-1902861	10931323	10930823-10931823	*ALMS1*	ALMS1, centrosome and basal body associated protein	-
	11	snp18054-scaffold185-10574388	62484776	62484276-62485276	*ACTR2*	ARP2 actin related protein 2 homolog	-
	12	snp50252-scaffold717-7820456	28205397	28204897-28205897	-	-	-
	12	snp39124-scaffold498-1821142	13398326	13397826-13398826	*CLDN10*	claudin 10	-
	12	snp36146-scaffold431-10361406	56828660	56828160-56829160	*LOC102187959*	WD repeat-containing protein 49-like	-
	12	snp35992-scaffold431-3698420	50165674	50165174-50166174	*RNF17*	ring finger protein 17	-
	12	snp35967-scaffold431-2526076	48993330	48992830-48993830	-	-	-
	12	snp35965-scaffold431-2423076	48993330	48992830-48993830	-	-	-
	12	snp3156-scaffold1095-2895889	30201454	30200954-30201954	-	-	-
	13	snp5930-scaffold1209-402211	62943504	62943004-62944004	*CPNE1*	copine 1	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Winter minimum temperature	13	snp48664-scaffold691-1156130	59355431	59354931-59355931	*TTLL9*	tubulin tyrosine ligase like 9	-
	13	snp39574-scaffold503-1642278	20498592	20498092-20499092	*MALRD1*	MAM and LDL receptor class A domain containing 1	-
	13	snp39569-scaffold503-1404542	20260856	20260356-20261356	*MALRD1*	MAM and LDL receptor class A domain containing 1	-
	13	snp27385-scaffold291-712138	642636	642136-643136	*PLCB1*	phospholipase C beta 1	-
	13	snp27383-scaffold291-642041	712733	712233-713233	*PLCB1*	phospholipase C beta 1	-
	13	snp13050-scaffold150-3127888	49042684	49042184-49043184	-	-	-
	14	snp51547-scaffold755-2025740	58464136	58463636-58464636	*LOC102187607*	uncharacterized LOC102187607	-
	14	snp49570-scaffold708-265571	92042329	92041829-92042829	*SNTB1*	syntrophin beta 1	-
	14	snp34916-scaffold416-2141525	56247710	56247210-56248210	*SAMD12*	sterile alpha motif domain containing 12	-
	14	snp23145-scaffold231-83906	61095042	61094542-61095542	-	-	-
	14	snp13420-scaffold1518-983083	41365818	41365318-41366318	*MYBL1*	MYB proto-oncogene like 1	-
	14	snp13419-scaffold1518-947027	41401874	41401374-41402374	*MYBL1*	MYB proto-oncogene like 1	-
	15	snp85-scaffold100-265663	5337309	5336809-5337809	*BIRC2*	baculoviral IAP repeat containing 2	-
	15	snp42948-scaffold57-4359751	58903636	58903136-58904136	*NCAM1*	neural cell adhesion molecule 1	-
	15	snp41697-scaffold542-2448188	48727568	48727068-48728068	*PDE2A*	phosphodiesterase 2A	-
	15	snp41682-scaffold542-1835927	48115307	48114807-48115807	*GRAMD1B*	GRAM domain containing 1B	-
	15	snp41676-scaffold542-1367712	47647092	47646592-47647592	*SERGEF*	secretion regulating guanine nucleotide exchange factor	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Winter minimum temperature	15	snp40425-scaffold516-971511	73156471	73155971-73156971	*LOC108637685*	contactin-5-like	-
	15	snp34632-scaffold41-1222140	17526001	17525501-17526501	*APIP*	APAF1 interacting protein	-
	15	snp32464-scaffold37-1900491	13881515	13881015-13882015	*IDO2*	indoleamine 2,3-dioxygenase 2	-
	15	snp16007-scaffold1685-115772	21045899	21045399-21046399	-	-	-
	15	snp1136-scaffold103-5440	51432609	51432109-51433109	*ARHGEF12*	Rho guanine nucleotide exchange factor 12	-
	15	snp106-scaffold100-1313915	6385561	6385061-6386061	*ATG13*	autophagy related 13	-
	16	snp53577-scaffold815-1571965	5628345	5627845-5628845	*KCNT2*	potassium sodium-activated channel subfamily T member 2	-
	16	snp20913-scaffold205-2784009	44251791	44251291-44252291	*CAMTA1*	calmodulin binding transcription activator 1	-
	16	snp14833-scaffold1599-1648545	48916418	48915918-48916918	*PLCH2*	phospholipase C eta 2	-
	17	snp53410-scaffold806-1045491	3062204	3061704-3062704	*DCHS2*	dachsous cadherin-related 2	-
	17	snp53404-scaffold806-807118	2823831	2823331-2824331	*KREMEN1*	kringle containing transmembrane protein 1	-
	17	snp43537-scaffold581-227197	33432388	33431888-33432888	-	-	-
	17	snp3984-scaffold1126-701621	46872403	46871903-46872903	*TMEM132D*	transmembrane protein 132D	-
	18	snp37912-scaffold467-1837460	30239750	30239250-30240250	-	-	-
	18	snp37896-scaffold467-1108407	29510697	29510197-29511197	-	-	-
	18	snp35658-scaffold43-2147682	22003155	22002655-22003655	-	-	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Winter minimum temperature	19	snp45201-scaffold616-880756	4319236	4318736-4319736	-	-	-
	19	snp3320-scaffold1101-846281	54256577	54256077-54257077	*SRSF2*	serine and arginine rich splicing factor 2	-
	19	snp3314-scaffold1101-583094	53993390	53992890-53993890	-	-	-
	19	snp32202-scaffold3643-393290	49794943	49794443-49795443	*DCXR*	L-xylulose reductase	-
	20	snp56495-scaffold89-972939	7348854	7348354-7349354	-	-	-
	21	snp50356-scaffold72-2158060	9924915	9924415-9925415	*LOC106503341*	uncharacterized LOC106503341	-
	21	snp46088-scaffold633-451511	23826009	23825509-23826509	*ADAMTSL3*	ADAMTS like 3	-
	21	snp25956-scaffold2688-192534	65987348	65986848-65987848	*PPP2R5C*	protein phosphatase 2 regulatory subunit B'gamma	-
	21	snp15594-scaffold165-138712	54348217	54347717-54348717	*FKBP3*	FK506 binding protein 3	-
	21	snp12283-scaffold1460-498799	58314366	58313866-58314866	*VRK1*	vaccinia related kinase 1	-
	22	snp56001-scaffold870-565726	56524290	56523790-56524790	*MKRN2*	makorin ring finger protein 2	-
	22	snp15362-scaffold163-3429800	37611968	37611468-37612468	-	-	-
	23	snp49917-scaffold714-724493	10343294	10342794-10343794	*CPNE5*	copine 5	-
	23	snp39147-scaffold499-154126	44742332	44741832-44742832	-	-	-
	24	snp9696-scaffold1349-373329	5681552	5681052-5682052	-	-	-
	24	snp7678-scaffold1277-988620	48734090	48733590-48734590	*DYM*	dymeclin	-
	24	snp663-scaffold102-1303046	20196478	20195978-20196978	*KIAA1328*	KIAA1328 ortholog	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Winter minimum temperature	24	snp40884-scaffold526-1598069	44551767	44551267-44552267	*SETBP1*	SET binding protein 1	-
	24	snp27537-scaffold294-67724	7063423	7062923-7063923	-	-	-
	24	snp13162-scaffold1502-321603	47347699	47347199-47348199	-	-	-
	25	snp43039-scaffold570-1733961	7503657	7503157-7504157	*ABAT*	4-aminobutyrate aminotransferase	-
	25	snp13492-scaffold1520-312869	18059682	18059182-18060182	-	-	-
	25	snp13490-scaffold1520-232171	17978984	17978484-17979484	*GP2*	glycoprotein 2	-
	25	snp12269-scaffold146-381647	28364139	28363639-28364639	*SBDS*	SBDSribosome assembly guanine nucleotide exchange factor	-
	26	snp47499-scaffold67-132575	35836843	35836343-35837343	*PLCE1*	phospholipase C epsilon 1	-
	26	snp41113-scaffold532-1051457	20679063	20678563-20679563	*BTRC*		-
	26	snp41112-scaffold532-1001039	20729481	20728981-20729981	*DPCD*	deleted in primary ciliary dyskinesia homolog (mouse)	-
	26	snp19915-scaffold1992-2074589	37683641	37683141-37684141	*CPEB3*	cytoplasmic polyadenylation element binding protein 3	-
	26	snp10491-scaffold1374-333237	333237	332737-333737	-	-	-
	27	snp59605-scaffold984-422160	1226446	1225946-1226946	*ZNF385D*	zinc finger protein 385D	-
	27	snp59597-scaffold984-144768	1503838	1503338-1504338	*ZNF385D*	zinc finger protein 385D	-
	27	snp51865-scaffold762-1650776	39295283	39294783-39295783	-	-	-
	27	snp30708-scaffold34-595928	2244633	2244133-2245133	*CSMD1*	CUB and Sushi multiple domains 1	-

APPENDIX E: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM SAM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Description	KEGG pathway
Winter minimum temperature	27	snp18611-scaffold1881-252963	33614189	33613689-33614689	*IDO2*	indoleamine 2,3-dioxygenase 2	-
	28	snp54811-scaffold838-2001811	13755825	13755325-13756325	*ANK3*	ankyrin 3	-
	28	snp54782-scaffold838-778571	12532585	12532085-12533085	*KCNMA1*	potassium calcium-activated channel subfamily M alpha 1	-
	29	snp54949-scaffold840-662662	17573575	17573075-17574075	*INTS4*	integrator complex subunit 4	-
	29	snp54098-scaffold825-43831	24064114	24063614-24064614	*NAV2*	neuron navigator 2	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Altitude	1	snp25518-scaffold263-526750	108849810	108849310 - 108850310	*RSRC1*	arginine and serine rich coiled-coil 1	-
	1	snp19512-scaffold197-992686	16713835	16713335-16714335	-	-	-
	1	snp19509-scaffold197-876261	16597410	16596910-16597910	-	-	-
	1	snp19505-scaffold197-696238	16417387	16416887-16417887	-	-	-
	1	snp16300-scaffold1716-229833	12179283	12178783-12179783	-	-	-
	2	snp37745-scaffold464-3970814	98730933	98730433-98731433	*ERBB4*	erb-b2 receptor tyrosine kinase 4	ErbB signaling pathway, Calcium signaling pathway, Endocytosis, Proteoglycans in cancer
	3	snp56474-scaffold89-82651	11124378	11123878-11124878	*COL8A2*	collagen type VIII alpha 2 chain	-
	4	snp49513-scaffold706-1675437	20881677	20881177-20882177	*SLC13A4*	solute carrier family 13 member 4	-
	5	snp47363-scaffold666-695998	90045426	90044926-90045926	*PIK3C2G*	phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2 gamma	Phosphatidylinositol signaling system, Inositol phosphate metabolism, Metabolic pathways

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Altitude	6	snp53987-scaffold821-117611	39856451	39855951-39856951	-	-	-
	6	snp36242-scaffold433-2243946	81977976	81977476-81978476	*TMPRSS11F*	transmembrane protease serine 11F	-
	6	snp11736-scaffold1432-648951	71404857	71404357-71405357	*TMEM165*	transmembrane protein 165	-
	7	snp15568-scaffold1647-372458	8247303	8246803-8247803	*ADGRE3*	adhesion G protein-coupled receptor E3	-
	7	snp10058-scaffold1356-3300452	47378770	47378270-47379270	*GLRA1*	glycine receptor alpha 1	Neuroactive ligand-receptor interaction
	8	snp43685-scaffold585-2415660	76710664	76710164-76711164	*UBQLN1*	ubiquilin 1	Protein processing in endoplasmic reticulum
	8	snp34741-scaffold412-610778	89075630	89075130-89076130	*PLPPR1*	phospholipid phosphatase related 1	-
	11	snp45234-scaffold618-132771	17051401	17050901-17051901	-	-	-
	11	snp20795-scaffold204-5124667	78183930	78183430-78184430	-	-	-
	12	snp9213-scaffold1336-116675	74077591	74077091-74078091	*NALCN*	sodium leak channel, non-selective	-
	13	snp58723-scaffold958-851235	21543881	21543381-21544381	*MLLT10*	myeloid/lymphoid or mixed-lineage leukemia; translocated to, 10	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Altitude	14	snp59977-scaffold998-2033155	48536870	48536370-48537370	*SLCO5A1*	solute carrier organic anion transporter family member 5A1	-
	14	snp43167-scaffold574-1313646	22901159	22900659-22901659	*ZFPM2*	zinc finger protein, FOG family member 2	MicroRNAs in cancer
	14	snp1520-scaffold1041-64419	2633393	2632893-2633893	-	-	-
	15	snp41669-scaffold542-1084394	47363774	47363274-47364274	*USH1C*	USH1 protein network component harmonin	-
	16	snp55087-scaffold845-820562	17285944	17285444-17286444	-		
	16	snp29520-scaffold32-2027374	35750378	35749878-35750878	*F5*	coagulation factor V	-
	17	snp47302-scaffold663-3319308	6591307	6590807-6591807	*SH3D19*	SH3 domain containing 19	-
	17	snp21209-scaffold207-4324665	16006598	16006098-16007098	*ATXN2*	ataxin 2	-
	18	snp22661-scaffold225-1558202	33177967	33177467-33178467	-		
	18	snp17436-scaffold1804-554905	60232869	60232369-60233369	*LOC108638022*	zinc finger protein 160-like	-
	20	snp52561-scaffold782-5057480	1250772	1250272-1251272	-		-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Altitude	20	snp43224-scaffold575-1525888	14421702	14421202-14422202	*CWC27*	CWC27 spliceosome associated protein homolog	-
	20	snp14653-scaffold1591-11752	10250412	10249912-10250912	*OCLN*	occludin	-
	24	snp6562-scaffold1232-943770	7963664	7963164-7964164	-	-	-
	26	snp55130-scaffold847-759066	7324080	7323580-7324580	-	-	-
	26	snp19903-scaffold1992-1520152	38238078	38237578-38238578	-	-	-
Annual Rainfall	1	snp59392-scaffold979-935182	2750518	2750018-2751018	*TIAM1*	T-cell lymphoma invasion and metastasis 1	Ras signaling pathway, cAMP signaling pathway, Chemokine signaling pathway, Regulation of actin cytoskeleton, Proteoglycans in cancer, Rap1 signaling pathway
	1	snp48200-scaffold682-1736590	70362256	70361756-70362756	*TNK2*	tyrosine kinase non receptor 2	Glycerophospholipid metabolism, Metabolic pathways
	1	snp16297-scaffold1716-120264	12288852	12288352-12289352	-	-	-
	2	snp46036-scaffold631-2300868	91996682	91996182-91997182	-	-	-
	2	snp33487-scaffold393-259518	67445881	67445381-67446381	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Annual Rainfall	2	snp30944-scaffold341-2489356	94888037	94887537-94888537	*KCNJ3*	potassium voltage-gated channel subfamily J member 3	Morphine addiction, Retrograde endocannabinoid signaling, Circadian entrainment, Glutamatergic synapse, Cholinergic synapse, Serotonergic synapse, Dopaminergic synapse, Estrogen signaling pathway, Oxytocin signaling pathway
	3	snp30451-scaffold3351-51973	32781675	32781175-32782175	-	-	-
	4	snp49402-scaffold704-985754	32462367	32461867-32462867	-	-	-
	4	snp49392-scaffold704-546427	32023040	32022540-32023540	*HYAL4*	hyaluronoglucosaminidase 4	Glycosaminoglycan degradation, Metabolic pathways
	5	snp34488-scaffold405-2384232	11246408	11245908-11246908	-	-	-
	5	snp12974-scaffold15-337619	109932167	109931667-109932667	*CACNA1I*	calcium voltage-gated channel subunit alpha1 I	MAPK signaling pathway, Calcium signaling pathway, Circadian entrainment, Aldosterone synthesis and secretion
	7	snp52389-scaffold780-399610	11248048	11247548-11248548	*DNMT1*	DNA methyltransferase 1	Cysteine and methionine metabolism, Metabolic pathways, MicroRNAs in cancer
	8	snp31811-scaffold356-502396	98840634	98840134-98841134	*PTPN3*	protein tyrosine phosphatase, non-receptor type 3	-
	8	snp28089-scaffold300-3875901	38660983	38660483-38661483	*SPATA6L*	spermatogenesis associated 6 like	-
	8	snp12230-scaffold1455-682954	86564862	86564362-86565362	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Annual Rainfall	9	snp3890-scaffold1122-863454	22708701	22708201-22709201	-	-	-
	9	snp331-scaffold1008-1595503	46548074	46547574-46548574	-	-	-
	11	snp20714-scaffold204-1444199	74503462	74502962-74503962	*ITSN2*	intersectin 2	-
	12	snp9155-scaffold1332-724004	42641075	42640575-42641575	-	-	-
	12	snp48841-scaffold697-47047	47047	46547-47547	*LOC102177789*	multidrug resistance-associated protein 4-like	-
	12	snp39128-scaffold498-1948805	13525989	13525489-13526489	*LOC102187779*	multidrug resistance-associated protein 4-like	-
	13	snp25868-scaffold267-329710	36145313	36144813-36145813	-	-	-
	14	snp9811-scaffold1351-78102	84663940	84663440-84664440	-	-	-
	15	snp58553-scaffold953-464186	44912479	44911979-44912979	*INSC*	inscuteable homolog (Drosophila)	-
	15	snp30052-scaffold329-521629	1491228	1490728-1491728	*GRIA4*	glutamate ionotropic receptor AMPA type subunit 4	Amphetamine addiction, Circadian entrainment, cAMP signaling pathway, Neuroactive ligand-receptor interaction, Retrograde endocannabinoid signaling, Glutamatergic, Dopaminergic synapse
	16	snp19861-scaffold1991-80006	18491172	18490672-18491672	*ESRRG*	estrogen related receptor gamma	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Annual Rainfall	16	snp12333-scaffold1464-723952	23458488	23457988-23458988	-	-	-
	17	snp57753-scaffold930-409801	7715154	7714654-7715654	*TRPV4*	transient receptor potential cation channel subfamily V member 4	Inflammatory mediator regulation of TRP channels
	17	snp47302-scaffold663-3319308	6591307	6590807-6591807	*SH3D19*	SH3 domain containing 19	-
	18	snp34700-scaffold411-510192	44074130	44073630-44074630	-	-	-
	22	snp12841-scaffold1497-1032419	50867126	50866626-50867626	*USP4*	ubiquitin specific peptidase 4	MAPK signaling pathway
	23	snp50404-scaffold721-148784	4230514	4230014-4231014	*LOC102189642*	uncharacterized LOC102189642	-
	24	snp54921-scaffold84-1860777	3061259	3060759-3061759	-	-	-
	24	snp40902-scaffold526-2321342	45275040	45274540-45275540	-	-	-
	27	snp23614-scaffold2383-185069	37118899	37118399-37119399	-	-	-
	28	snp3399-scaffold1102-2245266	39856757	39856257-39857257	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Annual Rainfall	28	snp16590-scaffold175-3394	21802422	21801922-21802922	*CTNNA3*	catenin alpha 3	Adherens junction, Arrhythmogenic right ventricular cardiomyopathy (ARVC), Hippo signaling pathway, Tight junction, Leukocyte transendothelial migration, Bacterial invasion of epithelial cells, Pathways in cancer, Endometrial cancer
	29	snp40464-scaffold517-89595	2917738	2917238-2918238	-	-	-
	29	snp38574-scaffold489-921092	20422832	20422332-20423332	-	-	-
Annual maximum temperature	1	snp48527-scaffold689-2824259	87837359	87836859-87837859	*KCNMB2*	potassium calcium-activated channel subfamily M regulatory beta	Vascular smooth muscle contraction, cGMP-PKG signaling pathway, Insulin secretion
	1	snp23076-scaffold230-4010657	129180349	129179849-129180849	*RBP2*	retinol binding protein 2	Vitamin digestion and absorption
	1	snp14128-scaffold1564-604366	22355127	22354627-22355627	*ROBO2*	roundabout guidance receptor 2	Axon guidance
	3	snp24551-scaffold2498-99871	15881880	15881380-15882380	-	-	-
	3	snp17639-scaffold183-91099	6296876	6296376-6297376	*DDR2*	discoidin domain receptor tyrosine kinase 2	-
	3	snp16963-scaffold1771-266976	88444061	88443561-88444561	*SSBP3*	single stranded DNA binding protein 3	-
	4	snp49402-scaffold704-985754	32462367	32461867-32462867	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Annual maximum temperature	4	snp49392-scaffold704-546427	32023040	32022540-32023540	*HYAL4*	hyaluronoglucosaminidase 4	Glycosaminoglycan degradation, Metabolic pathways
	4	snp44315-scaffold603-2039906	14083052	14082552-14083552	*ICA1*	islet cell autoantigen 1	Type I diabetes mellitus, Nicotinate and nicotinamide metabolism
	4	snp36583-scaffold440-1838759	40853766	40853266-40854266	-	-	-
	4	snp34391-scaffold404-600825	2778708	2778208-2779208	*COBL*	cordon-bleu WH2 repeat protein	-
	5	snp34488-scaffold405-2384232	11246408	11245908-11246908	-	-	-
	6	snp59985-scaffold999-128817	1438392	1437892-1438892	*1-Mar*	membrane associated ring-CH-type finger 1	-
	7	snp52390-scaffold780-436708	11210950	11210450-11211450	*LOC102186858*	suppressor of SWI4 1 homolog	-
	7	snp52389-scaffold780-399610	11248048	11247548-11248548	*DNMT1*	DNA methyltransferase 1	Cysteine and methionine metabolism, Metabolic pathways, MicroRNAs in cancer
	8	snp12230-scaffold1455-682954	86564862	86564362-86565362	-	-	-
	10	snp18126-scaffold185-13659990	63195990	63195490-63196490	-	-	-
	11	snp45234-scaffold618-132771	17051401	17050901-17051901	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Annual maximum temperature	11	snp41014-scaffold53-63994	63994	63494-64494	*ZC3H6*	zinc finger CCCH-type containing 6	-
	12	snp9250-scaffold1336-1582602	75543518	75543018-75544018	-	-	-
	12	snp48841-scaffold697-47047	47047	46547-47547	*LOC102177789*	multidrug resistance-associated protein 4-like	-
	12	snp36053-scaffold431-6447095	52914349	52913849-52914849	*ATP8A2*	ATPase phospholipid transporting 8A2	-
	14	snp9811-scaffold1351-78102	84663940	84663440-84664440	*MMP16*	matrix metallopeptidase 16	MicroRNAs in cancer
	15	snp58553-scaffold953-464186	44912479	44911979-44912979	*INSC*	inscuteable homolog (Drosophila)	-
	15	snp42020-scaffold548-3470677	29651164	29650664-29651664	*PAAF1*	proteasomal ATPase associated factor 1	-
	16	snp29520-scaffold32-2027374	35750378	35749878-35750878	*F5*	coagulation factor V	-
	16	snp19861-scaffold1991-80006	18491172	18490672-18491672	*ESRRG*	estrogen related receptor gamma	-
	16	snp12333-scaffold1464-723952	23458488	23457988-23458988	-	-	-
	17	snp47302-scaffold663-3319308	6591307	6590807-6591807	*SH3D19*	SH3 domain containing 19	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Annual maximum temperature	17	snp41403-scaffold5388-198417	71099137	71098637-71099637	-	-	-
	18	snp41225-scaffold534-898045	50347088	50346588-50347588	*TEX101*	testis expressed 101	-
	18	snp34700-scaffold411-510192	44074130	44073630-44074630	-	-	-
	18	snp17782-scaffold1848-22570	56721950	56721450-56722450	-	-	-
	18	snp17436-scaffold1804-554905	60232869	60232369-60233369	*LOC108638022*	zinc finger protein 160-like	-
	18	snp14593-scaffold1589-1852440	26003376	26002876-26003876	-	-	-
	21	snp7648-scaffold1276-837684	64897809	64897309-64898309	*EIF5*	eukaryotic translation initiation factor 5	RNA transport
	21	snp38828-scaffold492-2711925	38665981	38665481-38666481	*FOXG1*	forkhead box G1	FoxO signaling pathway
	22	snp12841-scaffold1497-1032419	50867126	50866626-50867626	*USP4*	ubiquitin specific peptidase 4	-
	23	snp50404-scaffold721-148784	4230514	4230014-4231014	*LOC102189642*	uncharacterized LOC102189642	-
	23	snp17094-scaffold1786-19797	26236163	26235663-26236663	*LOC102171392*	uncharacterized LOC102171392	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Annual maximum temperature	24	snp27311-scaffold290-3060892	35595978	35595478-35596478	*COLEC12*	collectin subfamily member 12	Phagosome
	27	snp23614-scaffold2383-185069	37118899	37118399-37119399	-	-	-
	29	snp40464-scaffold517-89595	2917738	2917238-2918238	-	-	-
	29	snp38574-scaffold489-921092	20422832	20422332-20423332	-	-	-
	4	snp49392-scaffold704-546427	32023040	32022540-32023540	*HYAL4*	hyaluronoglucosaminidase 4	Glycosaminoglycan degradation, Metabolic pathways
	4	snp15282-scaffold1629-635153	105797490	105796990-105797990	-	-	-
	11	snp41014-scaffold53-63994	63994	63494-64494	*ZC3H6*	zinc finger CCCH-type containing 6	-
	12	snp36146-scaffold431-10361406	56828660	56828160-56829160	*LOC102187959*	WD repeat-containing protein 49-like	-
	12	snp36053-scaffold431-6447095	52914349	52913849-52914849	*ATP8A2*	ATPase phospholipid transporting 8A2	-
Latitude	1	snp19512-scaffold197-992686	16713835	16713335-16714335	-	-	-
	1	snp19505-scaffold197-696238	16417387	16416887-16417887	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Latitude	1	snp16300-scaffold1716-229833	12179283	12178783-12179783	-	-	-
	1	snp12141-scaffold145-385191	149494557	149494057-149495057	*DYRK1A*	dual specificity tyrosine phosphorylation regulated kinase 1A	-
	3	snp48292-scaffold684-4653	3581154	3580654-3581654	*LRRFIP1*	LRR binding FLII interacting protein 1	-
	3	snp22453-scaffold222-851829	37572717	37572217-37573217	-	-	-
	4	snp905-scaffold1024-50451	88004210	88003710-88004710	-	-	-
	4	snp49513-scaffold706-1675437	20881677	20881177-20882177	*SLC13A4*	solute carrier family 13 member 4	-
	4	snp49392-scaffold704-546427	32023040	32022540-32023540	*HYAL4*	hyaluronoglucosaminidase 4	Glycosaminoglycan degradation, Metabolic pathways
	4	snp38627-scaffold49-1724285	8298732	8298232-8299232	*SAMD9*	sterile alpha motif domain containing 9	-
	6	snp7623-scaffold1273-214583	38624908	38624408-38625408	*KCNIP4*	potassium voltage-gated channel interacting protein 4	-
	6	snp36242-scaffold433-2243946	81977976	81977476-81978476	*LOC102185449*	transmembrane protease serine 11F	-
	6	snp36198-scaffold433-398052	80132082	80131582-80132582	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Latitude	6	snp31695-scaffold352-251483	97979673	97979173-97980173	*MAPK10*	mitogen-activated protein kinase 10	MAPK signaling pathway, Salmonella pathway, Pertusis pathway, Toxoplasmosis, Tuberculosis, Chagas (American trypanosomiasis)
	6	snp16810-scaffold1760-380281	33737096	33736596-33737596	*CCSER1*	coiled-coil serine rich protein 1	-
	6	snp16466-scaffold1734-406221	110814419	110813919-110814919	*C1QTNF7*	C1q and tumor necrosis factor related protein 7	-
	6	snp16164-scaffold170-20288	5571293	5570793-5571793	-	-	-
	6	snp12510-scaffold148-213923	29052799	29052299-29053299	*GRID2*	glutamate ionotropic receptor delta type subunit 2	Neuroactive ligand-receptor interaction, Long-term depression
	6	snp11736-scaffold1432-648951	71404857	71404357-71405357	*TMEM165*	transmembrane protein 165	-
	7	snp4346-scaffold1137-309073	100329328	100328828-100329828	-	-	-
	8	snp21876-scaffold2147-257467	106584007	106583507-106584507	-	-	-
	11	snp33932-scaffold4-2112652	94455406	94454906-94455906	*DENND1A*	DENN domain containing 1A	-
	12	snp36053-scaffold431-6447095	52914349	52913849-52914849	*ATP8A2*	ATPase phospholipid transporting 8A2	-
	13	snp58723-scaffold958-851235	21543881	21543381-21544381	*NEBL*	nebulette	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Latitude	13	snp5375-scaffold1188-527212	65850534	65850034-65851034	*DHX35*	DEAH-box helicase 35	-
	14	snp59977-scaffold998-2033155	48536870	48536370-48537370	*SLCO5A1*	solute carrier organic anion transporter family member 5A1	-
	14	snp16529-scaffold174-102753	2164361	2163861-2164861	-	-	-
	15	snp57942-scaffold939-353937	66463081	66462581-66463581	*PIWIL4*	piwi like RNA-mediated gene silencing 4	Dorso-ventral axis formation
	15	snp34650-scaffold41-1991894	18295755	18295255-18296255	*LOC108637626*	uncharacterized LOC108637626	-
	16	snp29520-scaffold32-2027374	35750378	35749878-35750878	*F5*	coagulation factor V	-
	17	snp52367-scaffold779-1456203	560411	559911-560911	*LOC102176353*	Ig lambda chain V-I region BL2-like	-
	17	snp19778-scaffold199-3354327	27123567	27123067-27124067	-	-	-
	17	snp14463-scaffold1578-643250	37014501	37014001-37015001	-	-	-
	18	snp34700-scaffold411-510192	44074130	44073630-44074630	-	-	-
	18	snp22661-scaffold225-1558202	33177967	33177467-33178467	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Latitude	18	snp17436-scaffold1804-554905	60232869	60232369-60233369	*LOC108638022*	zinc finger protein 160-like	-
	19	snp9199-scaffold1335-368697	44656775	44656275-44657275	*NSF*	N-ethylmaleimide sensitive factor, vesicle fusing ATPase	Synaptic vesicle cycle, GABAergic synapse, Vasopressin-regulated water reabsorption, Glycolysis / Gluconeogenesis
	19	snp41382-scaffold537-4632196	60263552	60263052-60264052	*ABCA9*	ATP-binding cassette sub-family A member 9	ABC transporters
	19	snp3316-scaffold1101-672669	54082965	54082465-54083465	-	-	-
	20	snp52561-scaffold782-5057480	1250772	1250272-1251272	-	-	-
	21	snp12295-scaffold1460-979199	57833966	57833466-57834466	*PRIMA1*	proline rich membrane anchor 1	-
	24	snp27311-scaffold290-3060892	35595978	35595478-35596478	*COLEC12*	collectin subfamily member 12	Phagosome
	26	snp55130-scaffold847-759066	7324080	7323580-7324580	-	-	-
	26	snp19903-scaffold1992-1520152	38238078	38237578-38238578	-	-	-
	26	snp11047-scaffold1397-327144	45203276	45202776-45203776	*DOCK1*	dedicator of cytokinesis 1	Fc gamma R-mediated phagocytosis
	27	snp51869-scaffold762-1819501	39464008	39463508-39464508	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Longitude	4	snp15282-scaffold1629-635153	105797490	105796990 - 105797990	-	-	-
	12	snp36146-scaffold431-10361406	56828660	56828160-56829160	*LOC102187959*	WD repeat-containing protein 49-like	-
Summer rainfall	1	snp59392-scaffold979-935182	2750518	2750018-2751018	*TIAM1*	T-cell lymphoma invasion and metastasis 1	Ras signaling pathway, cAMP signaling pathway, Chemokine signaling pathway, Regulation of actin cytoskeleton, Proteoglycans in cancer, Rap1 signaling pathway
	1	snp48200-scaffold682-1736590	70362256	70361756-70362756	*TNK2*	tyrosine kinase non receptor 2	Glycerophospholipid metabolism, Metabolic pathways
	2	snp46036-scaffold631-2300868	91996682	91996182-91997182	-	-	-
	4	snp49402-scaffold704-985754	32462367	32461867-32462867	-	-	-
	4	snp49392-scaffold704-546427	32023040	32022540-32023540	*HYAL4*	hyaluronoglucosaminidase 4	Glycosaminoglycan degradation, Metabolic pathways
	4	snp38627-scaffold49-1724285	8298732	8298232-8299232	*SAMD9*	-	-
	7	snp5704-scaffold12-796210	83494847	83494347-83495347	*LMNB1*	lamin B1	Apoptosis
	7	snp52389-scaffold780-399610	11248048	11247548-11248548	*DNMT1*	-	-
	8	snp31811-scaffold356-502396	98840634	98840134-98841134	*PTPN3*	protein tyrosine phosphatase, non-receptor type 3	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Summer rainfall	9	snp3890-scaffold1122-863454	22708701	22708201-22709201	-	-	-
	9	snp331-scaffold1008-1595503	46548074	46547574-46548574	-	-	-
	11	snp33932-scaffold4-2112652	94455406	94454906-94455906	*DENND1A*	DENN domain containing 1A	-
	12	snp9155-scaffold1332-724004	42641075	42640575-42641575	-	-	-
	12	snp48841-scaffold697-47047	47047	46547-47547	*LOC102177789*	multidrug resistance-associated protein 4-like	-
	12	snp39128-scaffold498-1948805	13525989	13525489-13526489	*LOC102187779*	multidrug resistance-associated protein 4-like	-
	13	snp25868-scaffold267-329710	36145313	36144813-36145813	-	-	-
	14	snp9811-scaffold1351-78102	84663940	84663440-84664440	*MMP16*	matrix metallopeptidase 16	-
	14	snp16529-scaffold174-102753	2164361	2163861-2164861	-	-	-
	15	snp58553-scaffold953-464186	44912479	44911979-44912979	*INSC*	inscuteable homolog (Drosophila)	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Summer rainfall	15	snp30052-scaffold329-521629	1491228	1490728-1491728	*GRIA4*	glutamate ionotropic receptor AMPA type subunit 4	Amphetamine addiction, Circadian entrainment, cAMP signaling pathway, Neuroactive ligand-receptor interaction, Retrograde endocannabinoid signaling, Glutamatergic synapse, Dopaminergic synapse
	16	snp19861-scaffold1991-80006	18491172	18490672-18491672	*ESRRG*	estrogen related receptor gamma	-
	16	snp12333-scaffold1464-723952	23458488	23457988-23458988	-	-	-
	17	snp57753-scaffold930-409801	7715154	7714654-7715654	*TRPV4*	transient receptor potential cation channel subfamily V member 4	Inflammatory mediator regulation of TRP channels
	17	snp47302-scaffold663-3319308	6591307	6590807-6591807	*SH3D19*	SH3 domain containing 19	-
	18	snp34700-scaffold411-510192	44074130	44073630-44074630	-	-	-
	18	snp33207-scaffold389-1327069	45545637	45545137-45546137	*WDR62*	WD repeat domain 62	-
	18	snp17782-scaffold1848-22570	56721950	56721450-56722450	-	-	-
	19	snp6963-scaffold1256-306323	20515332	20514832-20515832	*EFCAB5*	EF-hand calcium binding domain 5	-
	22	snp12841-scaffold1497-1032419	50867126	50866626-50867626	*USP4*	ubiquitin specific peptidase 4	-
	23	snp50404-scaffold721-148784	4230514	4230014-4231014	*LOC102189642*	uncharacterized LOC102189642	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Summer rainfall	23	snp25263-scaffold2590-415388	26671447	26670947-26671947	*LOC102188814*	-	-
	24	snp54921-scaffold84-1860777	3061259	3060759-3061759	-	-	-
	24	snp40902-scaffold526-2321342	45275040	45274540-45275540	-	-	-
	27	snp23614-scaffold2383-185069	37118899	37118399-37119399	-	-	-
	28	snp3399-scaffold1102-2245266	39856757	39856257-39857257	-	-	-
	28	snp16590-scaffold175-3394	21802422	21801922-21802922	*CTNNA3*	catenin alpha 3	Adherens junction, Arrhythmogenic right ventricular cardiomyopathy (ARVC), Hippo signaling pathway, Tight junction, Leukocyte transendothelial migration, Bacterial invasion of epithelial cells, Pathways in cancer, Endometrial cancer
	29	snp50434-scaffold722-192613	37199185	37198685-37199685	-	-	-
	29	snp40464-scaffold517-89595	2917738	2917238-2918238	-	-	-
	29	snp38574-scaffold489-921092	20422832	20422332-20423332	-	-	-
Summer maximum temperature	1	snp37544-scaffold46-1948609	151371167	151370667-151371667	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Summer maximum temperature	1	snp16300-scaffold1716-229833	12179283	12178783-12179783	-	-	-
	3	snp6319-scaffold1223-258363	12153807	12153307-12154307	*LOC102182460*	CD48 antigen	-
	3	snp48292-scaffold684-4653	3581154	3580654-3581654	*LRRFIP1*	LRR binding FLII interacting protein 1	-
	3	snp16203-scaffold1704-149140	101776263	101775763-101776763	-	-	-
	4	snp800-scaffold1021-442704	85792546	85792046-85793046	-	-	-
	4	snp49513-scaffold706-1675437	20881677	20881177-20882177	*SLC13A4*	solute carrier family 13 member 4	-
	4	snp49392-scaffold704-546427	32023040	32022540-32023540	*HYAL4*	hyaluronoglucosaminidase 4	Glycosaminoglycan degradation, Metabolic pathways
	5	snp5576-scaffold1195-1778115	39436115	39435615-39436615	*CNTN1*	contactin 1	Cell adhesion molecules (CAMs)
	5	snp49863-scaffold712-250701	49964487	49963987-49964987	*PPM1H*	protein phosphatase, Mg2+/Mn2+ dependent 1H	-
	5	snp47333-scaffold665-1313341	34131356	34130856-34131856	-	-	-
	5	snp33421-scaffold392-2865428	53855350	53854850-53855850	*LRIG3*	leucine rich repeats and immunoglobulin like domains 3	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Summer maximum temperature	6	snp59985-scaffold999-128817	1438392	1437892-1438892	*MARCH1*	membrane associated ring-CH-type finger 1	-
	6	snp23913-scaffold241-889775	46688925	46688425-46689425	*STIM2*	stromal interaction molecule 2	Calcium signaling pathway
	6	snp16810-scaffold1760-380281	33737096	33736596-33737596	*CCSER1*	coiled-coil serine rich protein 1	-
	6	snp16466-scaffold1734-406221	110814419	110813919-110814919	*C1QTNF7*	C1q and tumor necrosis factor related protein 7	-
	6	snp12510-scaffold148-213923	29052799	29052299-29053299	*GRID2*	glutamate ionotropic receptor delta type subunit 2	Neuroactive ligand-receptor interaction, Long-term depression
	7	snp56946-scaffold902-1662793	28275898	28275398-28276398	-	-	-
	8	snp55267-scaffold851-819441	1276495	1275995-1276995	-	-	-
	8	snp34597-scaffold409-991445	57484244	57483744-57484744	-	-	-
	8	snp31862-scaffold356-2790706	101128944	101128444-101129444	-	-	-
	8	snp19316-scaffold1950-134982	85746827	85746327-85747327	*ROR2*	receptor tyrosine kinase like orphan receptor 2	-
	8	snp12230-scaffold1455-682954	86564862	86564362-86565362	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Summer maximum temperature	9	snp13751-scaffold1533-178329	20366390	20365890-20366890	*DCBLD1*	discoidin, CUB and LCCL domain containing 1	-
	10	snp48046-scaffold68-1551899	50060186	50059686-50060686	*TCF12*	transcription factor 12	-
	11	snp33932-scaffold4-2112652	94455406	94454906-94455906	*DENND1A*	DENN domain containing 1A	-
	11	snp17842-scaffold185-1498399	71560765	71560265-71561265	*RBKS*	ribokinase	Pentose phosphate pathway
	12	snp9250-scaffold1336-1582602	75543518	75543018-75544018	*LOC102181510*	protein-lysine methyltransferase METTL21E	-
	12	snp42817-scaffold568-5730781	61577881	61577381-61578381	*CCNA1*	cyclin A1	Cell cycle, Viral carcinogenesis, AMPK signaling pathway, Progesterone-mediated oocyte maturation, Hepatitis B, Epstein-Barr virus infection
	12	snp39094-scaffold498-544295	12121479	12120979-12121979	-	-	-
	13	snp58684-scaffold956-1283634	68938922	68938422-68939422	*PTPRT*	protein tyrosine phosphatase, receptor type T	-
	13	snp5375-scaffold1188-527212	65850534	65850034-65851034	*DHX35*	DEAH-box helicase 35	-
	16	snp8975-scaffold1327-71381	21739030	21738530-21739530	-	-	-
	16	snp19861-scaffold1991-80006	18491172	18490672-18491672	*ESRRG*	estrogen related receptor gamma	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Summer maximum temperature	16	snp12333-scaffold1464-723952	23458488	23457988-23458988	-	-	
	17	snp47302-scaffold663-3319308	6591307	6590807-6591807	*SH3D19*	SH3 domain containing 19	-
	17	snp41403-scaffold5388-198417	71099137	71098637-71099637	-	-	-
	17	snp14463-scaffold1578-643250	37014501	37014001-37015001	-	-	-
	18	snp17436-scaffold1804-554905	60232869	60232369-60233369	*LOC108638022*	zinc finger protein 160-like	-
	19	snp55453-scaffold859-1193527	9287823	9287323-9288323	-	-	-
	19	snp37159-scaffold452-66379	18655606	18655106-18656106	*LOC102189615*	galectin-9	-
	19	snp3316-scaffold1101-672669	54082965	54082465-54083465	-	-	-
	20	snp37272-scaffold454-1102616	40701775	40701275-40702275	-	-	-
	21	snp7648-scaffold1276-837684	64897809	64897309-64898309	*EIF5*	eukaryotic translation initiation factor 5	-
	21	snp23803-scaffold240-762888	13991914	13991414-13992414	*SV2B*	synaptic vesicle glycoprotein 2B	ECM-receptor interaction

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Summer maximum temperature	21	snp12295-scaffold1460-979199	57833966	57833466-57834466	*PRIMA1*	proline rich membrane anchor 1	-
	23	snp29279-scaffold3157-14408	1707323	1706823-1707823	-	-	-
	24	snp752-scaffold102-4983326	23876758	23876258-23877258	*ASXL3*	additional sex combs like 3, transcriptional regulator	
	24	snp636-scaffold102-112408	19005840	19005340-19006340	-	-	-
	24	snp27311-scaffold290-3060892	35595978	35595478-35596478	*COLEC12*	collectin subfamily member 12	Phagosome
	26	snp55130-scaffold847-759066	7324080	7323580-7324580	-	-	-
	29	snp38574-scaffold489-921092	20422832	20422332-20423332	-	-	-
Summer minimum temperature	1	snp19505-scaffold197-696238	16417387	16416887-16417887	-	-	-
	1	snp16300-scaffold1716-229833	12179283	12178783-12179783	-	-	-
	1	snp12141-scaffold145-385191	149494557	149494057-149495057	*DYRK1A*	dual specificity tyrosine phosphorylation regulated kinase 1A	-
	2	snp37655-scaffold464-129557	102572190	102571690-102572690	*KCNH7*	potassium voltage-gated channel subfamily H member 7	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Summer minimum temperature	3	snp48292-scaffold684-4653	3581154	3580654-3581654	*LRRFIP1*	LRR binding FLII interacting protein 1	-
	3	snp47079-scaffold659-1230265	22017615	22017115-22018115	*WARS2*	tryptophanyl tRNA synthetase 2, mitochondrial	Aminoacyl-tRNA biosynthesis
	3	snp22453-scaffold222-851829	37572717	37572217-37573217	-	-	-
	3	snp17049-scaffold1777-185533	7769340	7768840-7769840	*SAG*	S-antigen visual arrestin	Leukocyte transendothelial migration, Hepatitis B, Phototransduction, Toxoplasmosis, MicroRNAs in cancer
	4	snp905-scaffold1024-50451	88004210	88003710-88004710	-	-	-
	4	snp53930-scaffold820-710141	66404911	66404411-66405411	*FOXP2*	forkhead box P2	-
	4	snp49513-scaffold706-1675437	20881677	20881177-20882177	*SLC13A4*	solute carrier family 13 member 4	-
	4	snp49392-scaffold704-546427	32023040	32022540-32023540	*HYAL4*	hyaluronoglucosaminidase 4	Glycosaminoglycan degradation, Metabolic pathways
	5	snp49863-scaffold712-250701	49964487	49963987-49964987	*FAM19A2*	family with sequence similarity 19 member A2, C-C motif chemokine like	-
	5	snp34158-scaffold4005-211391	71440016	71439516-71440516	*TOM1*	target of myb1 membrane trafficking protein	-
	6	snp9894-scaffold1352-2257871	92138914	92138414-92139414	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Summer minimum temperature	6	snp7623-scaffold1273-214583	38624908	38624408-38625408	*KCNIP4*	potassium voltage-gated channel interacting protein 4	-
	6	snp56633-scaffold8959-56291	113873325	113872825-113873825	*SORCS2*	sortilin related VPS10 domain containing receptor 2	-
	6	snp36198-scaffold433-398052	80132082	80131582-80132582	-	-	-
	6	snp31695-scaffold352-251483	97979673	97979173-97980173	*MAPK10*	mitogen-activated protein kinase 10	MAPK signaling pathway, Salmonella pathway, Pertusis pathway, Toxoplasmosis, Tuberculosis, Chagas (American trypanosomiasis)
	6	snp16810-scaffold1760-380281	33737096	33736596-33737596	*CCSER1*	coiled-coil serine rich protein 1	-
	6	snp16466-scaffold1734-406221	110814419	110813919-110814919	*C1QTNF7*	C1q and tumor necrosis factor related protein 7	-
	6	snp12510-scaffold148-213923	29052799	29052299-29053299	*GRID2*	glutamate ionotropic receptor delta type subunit 2	Neuroactive ligand-receptor interaction, Long-term depression
	6	snp11736-scaffold1432-648951	71404857	71404357-71405357	*TMEM165*	transmembrane protein 165	-
	7	snp4346-scaffold1137-309073	100329328	100328828-100329828	-	-	-
	8	snp51324-scaffold750-924334	79618015	79617515-79618515	*C8H9orf3*	chromosome 8 C9orf3 homolog	-
	8	snp51314-scaffold750-520800	80021549	80021049-80022049	*FANCC*	Fanconi anemia complementation group C	Fanconi anemia pathway
	8	snp34597-scaffold409-991445	57484244	57483744-57484744	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Summer minimum temperature	8	snp29219-scaffold314-1734599	21735860	21735360-21736360	*MTAP*	methylthioadenosine phosphorylase	Cysteine and methionine metabolism, Metabolic pathways
	9	snp40829-scaffold524-1397409	26440286	26439786-26440786	-	-	-
	11	snp33932-scaffold4-2112652	94455406	94454906-94455906	*SCAI*	suppressor of cancer cell invasion	-
	11	snp24226-scaffold247-1539518	38892651	38892151-38893151	-	-	-
	11	snp20197-scaffold20-1655564	57157750	57157250-57158250	-	-	-
	12	snp9250-scaffold1336-1582602	75543518	75543018-75544018	*LOC102181510*	protein-lysine methyltransferase METTL21E	-
	12	snp42759-scaffold568-3191823	64116839	64116339-64117339	-	-	-
	12	snp36053-scaffold431-6447095	52914349	52913849-52914849	*ATP8A2*	ATPase phospholipid transporting 8A2	-
	13	snp5375-scaffold1188-527212	65850534	65850034-65851034	-	-	-
	13	snp25868-scaffold267-329710	36145313	36144813-36145813	*DHX35*	DEAH-box helicase 35	-
	14	snp9263-scaffold1337-272572	88145570	88145070-88146070	*LOC102185933*	solute carrier family 25 member 36-like	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Summer minimum temperature	14	snp59977-scaffold998-2033155	48536870	48536370-48537370	*SLCO5A1*	solute carrier organic anion transporter family member 5A1	-
	16	snp29520-scaffold32-2027374	35750378	35749878-35750878	*F5*	coagulation factor V	-
	17	snp52367-scaffold779-1456203	560411	559911-560911	*LOC102176353*	Ig lambda chain V-I region BL2-like	-
	18	snp34700-scaffold411-510192	44074130	44073630-44074630	-	-	-
	18	snp22661-scaffold225-1558202	33177967	33177467-33178467	-	-	-
	18	snp17436-scaffold1804-554905	60232869	60232369-60233369	*LOC108638022*	zinc finger protein 160-like	-
	19	snp9199-scaffold1335-368697	44656775	44656275-44657275	*NSF*	N-ethylmaleimide sensitive factor, vesicle fusing ATPase	Synaptic vesicle cycle, GABAergic synapse, Vasopressin-regulated water reabsorption, Glycolysis / Gluconeogenesis
	19	snp37159-scaffold452-66379	18655606	18655106-18656106	*LOC102189615*	-	-
	20	snp52561-scaffold782-5057480	1250772	1250272-1251272	*LOC102183265*	-	-
	21	snp23803-scaffold240-762888	13991914	13991414-13992414	*SV2B*	synaptic vesicle glycoprotein 2B	ECM-receptor interaction
	23	snp51777-scaffold761-2827156	18740444	18739944-18740944	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Winter rainfall	24	snp3499-scaffold1109-248	12815624	12815124-12816124	-	-	-
	24	snp27311-scaffold290-3060892	35595978	35595478-35596478	*COLEC12*	collectin subfamily member 12	Phagosome
	26	snp55130-scaffold847-759066	7324080	7323580-7324580	-	-	-
	26	snp19903-scaffold1992-1520152	38238078	38237578-38238578	-	-	-
	27	snp51869-scaffold762-1819501	39464008	39463508-39464508	-	-	-
	1	snp59392-scaffold979-935182	2750518	2750018-2751018	*TIAM1*	T-cell lymphoma invasion and metastasis 1	-
	1	snp48527-scaffold689-2824259	87837359	87836859-87837859	*KCNMB2*	potassium calcium-activated channel subfamily M regulatory beta	Vascular smooth muscle contraction, cGMP-PKG signaling pathway, Insulin secretion
	1	snp16297-scaffold1716-120264	12288852	12288352-12289352	-	-	-
	2	snp46036-scaffold631-2300868	91996682	91996182-91997182	-	-	-
	2	snp33487-scaffold393-259518	67445881	67445381-67446381	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Winter rainfall	2	snp30944-scaffold341-2489356	94888037	94887537-94888537	*KCNJ3*	potassium voltage-gated channel subfamily J member 3	Morphine addiction, Retrograde endocannabinoid signaling, Circadian entrainment, Glutamatergic synapse, Cholinergic synapse, Serotonergic synapse, Dopaminergic synapse, Estrogen signaling pathway, Oxytocin signaling pathway
	3	snp30451-scaffold3351-51973	32781675	32781175-32782175	-	-	-
	3	snp1656-scaffold1047-697575	54056151	54055651-54056651	*AK5*	adenylate kinase 5	Purine metabolism, Metabolic pathways
	4	snp49402-scaffold704-985754	32462367	32461867-32462867	-	-	-
	4	snp49392-scaffold704-546427	32023040	32022540-32023540	*HYAL4*	hyaluronoglucosaminidase 4	Glycosaminoglycan degradation, Metabolic pathways
	5	snp56988-scaffold903-448348	91189871	91189371-91190371	-	-	-
	5	snp34488-scaffold405-2384232	11246408	11245908-11246908	-	-	-
	5	snp12974-scaffold15-337619	109932167	109931667-109932667	*CACNA1I*	calcium voltage-gated channel subunit alpha1 I	MAPK signaling pathway, Calcium signaling pathway, Circadian entrainment Aldosterone synthesis and secretion
	6	snp7623-scaffold1273-214583	38624908	38624408-38625408	*KCNIP4*	potassium voltage-gated channel interacting protein 4	-
	7	snp52389-scaffold780-399610	11248048	11247548-11248548	*DNMT1*	DNA methyltransferase 1	Cysteine and methionine metabolism, Metabolic pathways, MicroRNAs in cancer
	7	snp39971-scaffold510-501631	6868888	6868388-6869388	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Winter rainfall	8	snp28089-scaffold300-3875901	38660983	38660483-38661483	*SPATA6L*	spermatogenesis associated 6 like	-
	8	snp12230-scaffold1455-682954	86564862	86564362-86565362	-	-	-
	9	snp3890-scaffold1122-863454	22708701	22708201-22709201	-	-	-
	9	snp331-scaffold1008-1595503	46548074	46547574-46548574	-	-	-
	10	snp49332-scaffold703-4982668	12867868	12867368-12868368	*NRXN3*	neurexin 3	Cell adhesion molecules
	11	snp20714-scaffold204-1444199	74503462	74502962-74503962	*ITSN2*	intersectin 2	-
	12	snp9155-scaffold1332-724004	42641075	42640575-42641575	-	-	-
	12	snp48841-scaffold697-47047	47047	46547-47547	*LOC102177789*	multidrug resistance-associated protein 4-like	-
	12	snp39128-scaffold498-1948805	13525989	13525489-13526489	*LOC102187779*	multidrug resistance-associated protein 4-like	-
	13	snp25868-scaffold267-329710	36145313	36144813-36145813	-	-	-
	14	snp9811-scaffold1351-78102	84663940	84663440-84664440	*MMP16*	matrix metallopeptidase 16	MicroRNAs in cancer

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Winter rainfall	15	snp58553-scaffold953-464186	44912479	44911979-44912979	*INSC*	inscuteable homolog (Drosophila)	-
	15	snp30052-scaffold329-521629	1491228	1490728-1491728	*GRIA4*	glutamate ionotropic receptor AMPA type subunit 4	Amphetamine addiction, Circadian entrainment, cAMP signaling pathway, Neuroactive ligand-receptor interaction, Retrograde endocannabinoid signaling, Glutamatergic, Dopaminergic synapse
	16	snp19861-scaffold1991-80006	18491172	18490672-18491672	*ESRRG*	estrogen related receptor gamma	-
	16	snp12333-scaffold1464-723952	23458488	23457988-23458988	-	-	-
	17	snp47302-scaffold663-3319308	6591307	6590807-6591807	*SH3D19*	SH3 domain containing 19	-
	18	snp34700-scaffold411-510192	44074130	44073630-44074630	-	-	-
	23	snp50404-scaffold721-148784	4230514	4230014-4231014	*BMP5*	bone morphogenetic protein 5	TGF-beta signaling pathway, Hippo signaling pathway
	24	snp54921-scaffold84-1860777	3061259	3060759-3061759	-	-	-
	24	snp40902-scaffold526-2321342	45275040	45274540-45275540	-	-	-
	27	snp23614-scaffold2383-185069	37118899	37118399-37119399	-	-	-
	28	snp3399-scaffold1102-2245266	39856757	39856257-39857257	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Winter rainfall	28	snp16590-scaffold175-3394	21802422	21801922-21802922	*CTNNA3*	catenin alpha 3	Adherens junction, Arrhythmogenic right ventricular cardiomyopathy (ARVC), Hippo signaling pathway, Tight junction, Leukocyte transendothelial migration, Bacterial invasion of epithelial cells, Pathways in cancer, Endometrial cancer
	29	snp40464-scaffold517-89595	2917738	2917238-2918238	-	-	-
	29	snp38574-scaffold489-921092	20422832	20422332-20423332	-	-	-
Winter maximum rainfall	3	snp17639-scaffold183-91099	6296876	6296376-6297376	*DDR2*	discoidin domain receptor tyrosine kinase 2	-
	3	snp16963-scaffold1771-266976	88444061	88443561-88444561	*SSBP3*	single stranded DNA binding protein 3	-
	4	snp49402-scaffold704-985754	32462367	32461867-32462867	-	-	-
	4	snp49392-scaffold704-546427	32023040	32022540-32023540	*HYAL4*	hyaluronoglucosaminidase 4	Glycosaminoglycan degradation, Metabolic pathways
	4	snp36583-scaffold440-1838759	40853766	40853266-40854266	-	-	-
	6	snp7623-scaffold1273-214583	38624908	38624408-38625408	*KCNIP4*	potassium voltage-gated channel interacting protein 4	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
	10	snp18126-scaffold185-13659990	63195990	63195490-63196490	-	-	-
	11	snp45234-scaffold618-132771	17051401	17050901-17051901	-	-	-
Winter maximum rainfall	12	snp9250-scaffold1336-1582602	75543518	75543018-75544018	*LOC102181510*	protein-lysine methyltransferase METTL21E	-
	12	snp36053-scaffold431-6447095	52914349	52913849-52914849	*ATP8A2*	ATPase phospholipid transporting 8A2	-
	13	snp7747-scaffold1278-2333650	66472098	66471598-66472598	*TTI1*	TELO2 interacting protein 1	mTOR signaling pathway
	15	snp58553-scaffold953-464186	44912479	44911979-44912979	*INSC*	inscuteable homolog (Drosophila)	-
	15	snp34650-scaffold41-1991894	18295755	18295255-18296255	*LOC108637626*	uncharacterized LOC108637626	-
	18	snp34700-scaffold411-510192	44074130	44073630-44074630	-	-	-
	18	snp17436-scaffold1804-554905	60232869	60232369-60233369	*LOC108638022*	zinc finger protein 160-like	-
	24	snp27311-scaffold290-3060892	35595978	35595478-35596478	*COLEC12*	collectin subfamily member 12	Phagosome
	29	snp38574-scaffold489-921092	20422832	20422332-20423332	-	-	-

APPENDIX F: SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM LFMM ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

Variable	CHX	Associated SNP	SNP position	Gene region	Associated gene	Gene description	KEGG pathway
Winter minimum rainfall	11	snp41014-scaffold53-63994	63994	63494-64494	*ZC3H6*	zinc finger CCCH-type containing 6	-
	12	snp36146-scaffold431-10361406	56828660	56828160-56829160	*LOC102187959*	WD repeat-containing protein 49-like	-

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
3	snp46886-scaffold654-758578	33592325	33542325-33642325	*VAV3*	vav guanine nucleotide exchange factor 3	B cell receptor signaling pathway, Leukocyte transendothelial migration, cAMP signaling pathway, Chemokine signaling pathway, Focal adhesion, Natural killer cell mediated cytotoxicity, T cell receptor signaling pathway, Fc epsilon RI signaling pathway, Fc gamma R-mediated phagocytosis, Regulation of actin cytoskeleton	0.2
3	snp25341-scaffold261-1157571	31618061	31568061-31668061	*C8B*	Complement C8 beta chain	Systemic lupus erythematosus, Complement and coagulation cascades, Prion diseases, Amoebiasis	0.19
4	snp5865-scaffold1205-1942450	103652377	103602377-103702377	-	-	-	0.19
6	snp11969-scaffold1440-30837	109227668	109177668-109277668	*LDB2*	LIM domain binding 2		0.19
29	snp14776-scaffold1596-98969	1164748	1114748-1214748	-	-	-	0.19
1	snp3229-scaffold1097-405939	37664341	37614341-37714341	-	-	-	0.18

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
1	snp37563-scaffold46-2727355	150592421	150542421-150642421	-	-	-	0.18
5	snp51075-scaffold740-2304048	86722483	86672483-86772483	*CMAS*	cytidine monophosphate N-acetylneuraminic acid synthetase	Amino sugar and nucleotide sugar metabolism, Metabolic pathways	0.18
7	snp5700-scaffold12-617113	83673944	83623944-83723944	*MARCH3*	membrane associated ring-CH-type finger 3	-	0.18
7	snp50009-scaffold716-1437133	74573003	74523003-74623003	-	-	-	0.18
8	snp54504-scaffold830-5185904	63472625	63422625-63522625	*ANKS6*	ankyrin repeat and sterile alpha motif domain containing 6	-	0.18
8	snp19563-scaffold1972-84578	64829974	64779974-64879974	*INVS*	inversin	Wnt signaling pathway	0.18
1	snp45640-scaffold627-834807	40536709	40486709-40586709	*EPHA6*	EPH receptor A6	Axon guidance	0.17
2	snp6926-scaffold1252-687266	53875171	53825171-53925171	-	-	-	0.17
2	snp29021-scaffold312-3475375	62556317	62506317-62606317	*ERG*	ERG, ETS transcription factor	-	0.17
3	snp46895-scaffold654-1152065	33985812	33935812-34035812	*NTNG1*	netrin G1	Cell adhesion molecules (CAMs), Axon guidance,	0.17
3	snp55801-scaffold865-890958	3375550	3325550-3425550	*UBE2F*	ubiquitin conjugating enzyme E2 F (putative)	Ubiquitin mediated proteolysis	0.17

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
3	snp43603-scaffold582-546044	36059562	36009562-36109562	-	-	-	0.17
3	snp46915-scaffold654-2025384	34859131	34809131-34909131	-	-	-	0.17
3	snp52744-scaffold790-1270662	111843331	111793331-111893331	*TOMM40L*	translocase of outer mitochondrial membrane 40 like	Amyotrophic lateral sclerosis (ALS)	0.17
4	snp2031-scaffold106-461271	99293666	99243666-99343666	*TRIM24*	tripartite motif containing 24	-	0.17
5	snp43882-scaffold593-525511	92986513	92936513-93036513	*PTPRO*	protein tyrosine phosphatase, receptor type O	-	0.17
5	snp28629-scaffold308-102299	93685434	93635434-93735434	*C5H12orf60*	chromosome 5 C12orf60 homolog	-	0.17
6	snp12588-scaffold148-3749188	32588064	32538064-32638064	*GRID2*	glutamate ionotropic receptor delta type subunit 2	Neuroactive ligand-receptor interaction, Long-term depression	0.17
6	snp9282-scaffold1338-12300	98692861	98642861-98742861	*PTPN13*	protein tyrosine phosphatase, non-receptor type 13	Apoptosis	0.17
7	snp59827-scaffold995-258454	32054049	32004049-32104049	-	-	-	0.17

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
7	snp7371-scaffold1269-4720182	57315916	57265916-57365916	*PPARGC1B*	PPARG coactivator 1 beta	Insulin resistance	0.17
7	snp8836-scaffold1317-742522	32548906	32498906-32598906	-	-	-	0.17
8	snp28057-scaffold300-2523826	37308908	37258908-37358908	*IL33*	interleukin 33	Influenza A, Cytosolic DNA-sensing pathway	0.17
1	snp46388-scaffold640-654371	24321451	24271451-24371451	-	-	-	0.16
1	snp12147-scaffold145-640752	149238996	149188996-149288996	*TTC3*	tetratricopeptide repeat domain 3	RNA degradation	0.16
1	snp26383-scaffold276-3296025	91854319	91804319-91904319	*NAALADL2*	N-acetylated alpha-linked acidic dipeptidase like 2	-	0.16
1	snp14220-scaffold1568-1401817	50038931	49988931-50088931	*ALCAM*	activated leukocyte cell adhesion molecule	Cell adhesion molecules (CAMs)	0.16
1	snp40505-scaffold519-327438	76602329	76552329-76652329	-	-	-	0.16
2	snp28292-scaffold301-2721451	28918563	28868563-28968563	-	-	-	0.16
2	snp45014-scaffold614-1443654	47037815	46987815-47087815	*SPATS2L*	spermatogenesis associated serine rich 2 like	-	0.16

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
2	snp37764-scaffold464-4814885	97886862	97836862-97936862	-	-	-	0.16
2	snp53802-scaffold82-4806953	81680302	81630302-81730302	-	-	-	0.16
2	snp35885-scaffold430-6814791	44696134	44646134-44746134	-	-	-	0.16
2	snp26998-scaffold287-1034787	116999916	116949916-117049916	*DNER*	delta/notch like EGF repeat containing	-	0.16
3	snp46921-scaffold654-2263991	35097738	35047738-35147738	*C3H1orf87*	chromosome 3 C1orf87 homolog	-	0.16
3	snp29613-scaffold321-1836730	48647322	48597322-48697322	*RPAP2*	RNA polymerase II associated protein 2	-	0.16
4	snp49503-scaffold706-1271483	20477723	20427723-20527723	*DGKB*	diacylglycerol kinase beta	Glycerolipid metabolism, Glycerophospholipid metabolism, Metabolic pathways, Phosphatidylinositol signaling system, Phospholipase D signaling pathway, Choline metabolism in cancer	0.16
4	snp24708-scaffold2522-236009	115506942	115456942-115556942	-	-	-	0.16
4	snp49504-scaffold706-1312647	20518887	20468887-20568887	*DGKB*	diacylglycerol kinase beta	Glycerolipid metabolism, Glycerophospholipid metabolism, Metabolic pathways, Phosphatidylinositol signaling system, Phospholipase D signaling pathway, Choline metabolism in cancer	0.16
4	snp44414-scaffold603-6273000	9849958	9799958-9899958	*CNTNAP2*	contactin associated protein-like 2	Cell adhesion molecules (CAMs)	0.16

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
5	snp47364-scaffold666-740937	90000487	89950487-90050487	*PIK3C2G*	phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2 gamma	Phosphatidylinositol signaling system, Inositol phosphate metabolism, Metabolic pathways	0.16
5	snp48928-scaffold698-2590548	103597059	103547059-103647059	*ANO2*	anoctamin 2	Olfactory transduction	0.16
5	snp36650-scaffold443-872633	16406791	16356791-16456791	-	-	-	0.16
5	snp48875-scaffold698-181195	101187706	101137706-101237706	*LOC102185252*	antigen WC1.1-like	-	0.16
6	snp12578-scaffold148-3277158	32116034	32066034-32166034	*GRID2*	glutamate ionotropic receptor delta type subunit 2	Neuroactive ligand-receptor interaction, Long-term depression	0.16
6	snp48554-scaffold690-382655	104681989	104631989-104731989	-	-	-	0.16
6	snp59988-scaffold999-243335	1552910	1502910-1602910	*MARCH1*	membrane associated ring-CH-type finger 1	-	0.16
6	snp19366-scaffold1958-28303	8110392	8060392-8160392	-	-	-	0.16

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
6	snp58130-scaffold94-7008249	69906314	69856314-69956314	*AASDH*	aminoadipate-semialdehyde dehydrogenase	Pantothenate and CoA biosynthesis	0.16
6	snp4391-scaffold1139-684249	105748992	105698992-105798992	*RAB28*	RAB28, member RAS oncogene family	-	0.16
7	snp1788-scaffold105-472853	36135770	36085770-36185770	*LOC102186589*	hyaluronan mediated motility receptor pseudogene	-	0.16
7	snp30578-scaffold339-857546	42258884	42208884-42308884	*LOC106502364*	uncharacterized LOC106502364	-	0.16
7	snp29806-scaffold323-2204390	65759954	65709954-65809954	-	-	-	0.16
8	snp54428-scaffold830-1973402	60260123	60210123-60310123	-	-	-	0.16
8	snp28115-scaffold300-5002508	39787590	39737590-39837590	*SLC1A1*	solute carrier family 1 member 1	Glutamatergic synapse, Protein digestion and absorption	0.16
8	snp58892-scaffold964-453108	64292189	64242189-64342189	*LOC106502417*	uncharacterized LOC106502417	-	0.16
8	snp44201-scaffold600-974110	15809098	15759098-15859098	-	-	-	0.16
8	snp28017-scaffold300-779469	35564551	35514551-35614551	-	-	-	0.16

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
28	snp12234-scaffold1457-7594	707545	657545-757545	*TAF5L*	TATA-box binding protein associated factor 5 like	Basal transcription factors, Herpes simplex infection	0.16
28	snp1053-scaffold1028-730214	4064578	4014578-4114578	*FAM35A*	family with sequence similarity 35 member A	-	0.16
28	snp4322-scaffold1134-921448	36894154	36844154-36944154	*LYST*	lysosomal trafficking regulator	-	0.16
28	snp26088-scaffold2713-44065	1112793	1062793-1162793	-	-	-	0.16
29	snp4251-scaffold1132-235670	10527021	10477021-10577021	*DLG2*	discs large MAGUK scaffold protein 2	Hippo signaling pathway	0.16
29	snp59593-scaffold983-980092	5093861	5043861-5143861	*FOLH1B*	folate hydrolase 1B	Alanine, aspartate and glutamate metabolism, Metabolic pathways, Vitamin digestion and absorption	0.16
1	snp2699-scaffold1077-1406515	18053497	18003497-18103497	*CXADR*	coxsackie virus and adenovirus receptor	Viral myocarditis	0.15
1	snp6737-scaffold1244-1457031	145326929	145276929-145376929	-	-	-	0.15
1	snp48325-scaffold686-119085	47303473	47253473-47353473	-	-	-	0.15

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
1	snp47762-scaffold672-822164	81534263	81484263-81584263	*ABCC5*	ATP binding cassette subfamily C member 5	ABC transporters	0.15
1	snp6730-scaffold1244-1127970	145655990	145605990-145705990	-	-	-	0.15
1	snp12149-scaffold145-736257	149143491	149093491-149193491	*RIPPLY3*	ripply transcriptional repressor 3	-	0.15
1	snp53422-scaffold809-29101	154639837	154589837-154689837	*RAB5A*	RAB5A, member RAS oncogene family	Endocytosis, Tuberculosis, Ras signaling pathway, Phagosome, Vasopressin-regulated water reabsorption, Amyotrophic lateral sclerosis (ALS), Amoebiasis	0.15
2	snp51419-scaffold752-1380042	56647888	56597888-56697888	-	-	-	0.15
2	snp45055-scaffold614-3398825	48992986	48942986-49042986	-	-	-	0.15
2	snp45015-scaffold614-1480545	47074706	47024706-47124706	*SPATS2L*	spermatogenesis associated serine rich 2 like	-	0.15
2	snp3643-scaffold1113-1019430	118487400	118437400-118537400	*CCDC141*	coiled-coil domain containing 141	-	0.15
2	snp32326-scaffold367-967699	9107165	9057165-9157165	*CEP85*	centrosomal protein 85	-	0.15
2	snp20487-scaffold202-5324352	35101095	35051095-35151095	-	-	-	0.15

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
2	snp45006-scaffold614-1153619	46747780	46697780-46797780	*EPC2*	enhancer of polycomb homolog 2		0.15
3	snp46898-scaffold654-1299029	34132776	34082776-34182776	*FGGY*	FGGYcarbohydrate kinase domain containing	-	0.15
3	snp5258-scaffold1181-308803	95400318	95350318-95450318	*SPAG17*	sperm associated antigen 17		0.15
3	snp40796-scaffold522-1686534	2441361	2391361-2491361	*HDAC4*	histone deacetylase 4	Epstein-Barr virus infection, MicroRNAs in cancer	0.15
3	snp18872-scaffold190-674518	55429401	55379401-55479401	*ADGRL4*	adhesion G protein-coupled receptor L4	-	0.15
3	snp56493-scaffold89-902962	10304067	10254067-10354067	*ZFYVE9*	zinc finger FYVE-type containing 9	Endocytosis, TGF-beta signaling pathway	0.15
3	snp47902-scaffold675-1515949	26750210	26700210-26800210	*TRIM33*	tripartite motif containing 33	-	0.15
3	snp1671-scaffold1047-1340790	53412936	53362936-53462936	*ST6GALNAC5*	ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 5	Glycosphingolipid biosynthesis - ganglio series, Metabolic pathways	0.15
3	snp56482-scaffold89-467312	10739717	10689717-10789717	-	-	-	0.15

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
3	snp46900-scaffold654-1368681	34202428	34152428-34252428	*FGGY*	FGGYcarbohydrate kinase domain containing	-	0.15
3	snp56492-scaffold89-853164	10353865	10303865-10403865	*ZMYM4*	zinc finger MYM-type containing 4	-	0.15
3	snp47903-scaffold675-1570544	26804805	26754805-26854805	*CC2D1B*	coiled-coil and C2 domain containing 1B	-	0.15
3	snp11944-scaffold144-4727052	66927020	66877020-66977020	*TNNI3K*	serine/threonine-protein kinase TNNI3K	-	0.15
3	snp41714-scaffold543-98568	102889434	102839434-102939434	*MACF1*	microtubule-actin cross-linking factor 1	-	0.15
4	snp46568-scaffold646-172942	96126083	96076083-96176083	*AGBL3*	ATP/GTP binding protein like 3	-	0.15
4	snp20988-scaffold2056-332954	113908043	113858043-113958043	-	-	-	0.15
4	snp5883-scaffold1205-2692488	102902339	102852339-102952339	-	-	-	0.15
4	snp24710-scaffold2522-302031	115440920	115390920-115490920	*NCAPG2*	non-SMC condensin II complex subunit G2	-	0.15
5	snp56978-scaffold903-6977	90748500	90698500-90798500	-			0.15

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
5	snp47374-scaffold666-1211775	89529649	89479649-89579649	*DUSP16*	dual specificity phosphatase 16	MAPK signaling pathway	0.15
5	snp12968-scaffold15-33807	110234896	110184896-110284896	*GRAP2*	GRB2-related adaptor protein 2	T cell receptor signaling pathway	0.15
5	snp54018-scaffold822-23515	6563698	6513698-6613698	-	-	-	0.15
5	snp2335-scaffold107-1957186	76983459	76933459-77033459	*DENND5B*	DENN domain containing 5B	-	0.15
5	snp14303-scaffold157-2558865	17128954	17078954-17178954	-	-	-	0.15
6	snp27055-scaffold289-32356	54089115	54039115-54139115	*OTOP1*	otopetrin 1	-	0.15
6	snp4379-scaffold1139-124328	105189071	105139071-105239071	-	-	-	0.15
6	snp927-scaffold1025-601285	94998169	94948169-95048169	*COPS4*	COP9 signalosome subunit 4	-	0.15
6	snp18698-scaffold189-729955	4157344	4107344-4207344	*LOC108636255*	endogenous retrovirus group K member 5 Gag polyprotein-like	-	0.15
6	snp42239-scaffold553-359292	19666966	19616966-19716966	*GSTCD*	glutathione S-transferase C-terminal domain containing	-	0.15
7	snp29823-scaffold323-2910326	65054018	65004018-65104018	*CYFIP2*	-	RNA transport, Regulation of actin cytoskeleton	0.15

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
7	snp2565-scaffold1075-600374	19596		-	-	-	0.15
7	snp1792-scaffold105-660423	35948200	35898200-35998200	*COL23A1*	collagen type XXIII alpha 1 chain	-	0.15
7	snp24828-scaffold254-1667537	59295753	59245753-59345753	*NRG2*	neuregulin 2	ErbB signaling pathway, EGFR tyrosine kinase inhibitor resistance	0.15
8	snp44611-scaffold606-3594931	51692420	51642420-51742420	-	-	-	0.15
8	snp28003-scaffold300-204469	34989551	34939551-35039551	*LOC108636636*	-	-	0.15
8	snp44224-scaffold600-1873310	16708298	16658298-16758298	*MOB3B*	MOB kinase activator 3B	-	0.15
8	snp28079-scaffold300-3453358	38238440	38188440-38288440	*UHRF2*	ubiquitin like with PHD and ring finger domains 2	-	0.15
8	snp28131-scaffold300-5755525	40540607	40490607-40590607	*GLIS3*	GLIS family zinc finger 3	-	0.15
28	snp3352-scaffold1102-351080	37962571	37912571-38012571	*GRID1*	glutamate ionotropic receptor delta type subunit 1	Neuroactive ligand-receptor interaction	0.15

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
28	snp16641-scaffold175-2273073	24072101	24022101-24122101	*LOC102191766*	core histone macro-H2A.2	-	0.15
29	snp19435-scaffold1961-601262	15033747	14983747-15083747	-	-	-	0.15
1	snp36273-scaffold435-778554	75496238	75446238-75546238	-	-	-	0.14
1	snp46380-scaffold640-385838	24589984	24539984-24639984	-	-	-	0.14
1	snp26409-scaffold276-4440247	92998541	92948541-93048541	*NLGN1*	neuroligin 1	Cell adhesion molecules (CAMs)	0.14
1	snp394-scaffold1009-1164506	112960571	112910571-113010571	-	-	-	0.14
1	snp40507-scaffold519-390745	76665636	76615636-76715636	-	-	-	0.14
2	snp50992-scaffold739-4324077	74381812	74331812-74431812	*ZRANB3*	zinc finger RANBP2-type containing 3	-	0.14
3	snp54993-scaffold841-846994	14501250	14451250-14551250	*BMP8B*	bone morphogenetic protein 8b	-	0.14
3	snp43596-scaffold582-238964	35752482	35702482-35802482	-	-	-	0.14

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
3	snp1690-scaffold1047-2226391	52527335	52477335-52577335	*ST6GALNAC3*	ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 3	Glycosphingolipid biosynthesis - ganglio series, Metabolic pathways	0.14
3	snp28604-scaffold3068-290611	100745676	100695676-100795676	-	-	-	0.14
4	snp18345-scaffold1859-1115241	102088860	102038860-102138860	*SSBP1*	single stranded DNA binding protein 1	DNA replication, Mismatch repair, Homologous recombination	0.14
5	snp54020-scaffold822-115839	6656022	6606022-6706022	-	-	-	0.14
5	snp5538-scaffold1195-78873	41135357	41085357-41185357	-	-	-	0.14
5	snp2439-scaffold107-6624674	81650947	81600947-81700947	*ITPR2*	inositol 1,4,5-trisphosphate receptor type 2	Serotonergic synapse, Vascular smooth muscle contraction, Calcium signaling pathway, cGMP-PKG signaling pathway, cGMP-PKG signaling pathway, Phosphatidylinositol signaling system, Oocyte meiosis, Apoptosis, Gap junction, Platelet activation, Long-term potentiation, Retrograde endocannabinoid signaling, Glutamatergic synapse, Cholinergic synapse, Dopaminergic synapse, Long-term depression, Inflammatory mediator regulation of T RP channels, GnRH signaling pathway, Estrogen signaling pathway, Thyroid hormone synthesis, Oxytocin signaling pathway, Salivary secretion, Gastric acid secretion, Pancreatic secretion Alzheimer's disease	0.14
6	snp27121-scaffold289-2919323	56976082	56926082-57026082	*HIP2*	ubiquitin conjugating enzyme E2 K	Ubiquitin mediated proteolysis, Insulin signaling pathway	0.14

APPENDIX G: OUTLIER SNPs WITH STRONGEST ASSOCIATION TO CLIMATIC AND GEOGRAPHIC VARIABLES FROM RDA ANALYSIS, INCLUDING FUNCTIONAL ANNOTATIONS FOR EACH CANDIDATE GENE (continued)

CHX	Associated SNP	Genomic position	Genomic region	Associated genes	Gene description	KEGG pathways	RDA score[2]
6	snp4233-scaffold1131-229217	3116474	3066474-3166474	-	-	-	0.14
6	snp28932-scaffold3114-9539	108375129	108325129-108425129	-	-	-	0.14
7	snp1808-scaffold105-1334035	35274588	35224588-35324588	*PDLIM7*	PDZ and LIM domain 7	-	0.14
7	snp7320-scaffold1269-2563532	55159266	55109266-55209266	*KCTD16*	potassium channel tetramerization domain containing 16	-	0.14
8	snp10569-scaffold1376-1806887	72794174	72744174-72844174	*EPHX2*	epoxide hydrolase 2	Metabolic pathways, Arachidonic acid metabolism, Peroxisome	0.14
8	snp22008-scaffold2164-632415	28916836	28866836-28966836	*TTC39B*	tetratricopeptide repeat domain 39B	-	0.14
8	snp44635-scaffold606-4665686	52763175	52713175-52813175	*PRUNE2*	prune homolog 2	-	0.14
27	snp44439-scaffold604-543600	11015199	10965199-11065199	*LOC102168418*	disintegrin and metalloproteinase domain-containing protein 2	-	0.14
29	snp19431-scaffold1961-424641	14857126	14807126-14907126	-	-	-	0.14
29	snp11241-scaffold1402-432416	46158519	46108519-46208519	*LOC102176955*	chitinase domain-containing protein 1	-	0.14