Genome-wide characterization of South African pig breeds

by

Nompilo Lucia Hlongwane

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

July 2022

PREFACE

The research contained in this thesis was completed by the candidate while based in the Discipline of Genetics, School of Life Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa and the Agricultural Research Council-Biotechnology Platform. The research was financially supported by University of Kwa-Zulu Natal, Department of Science and Innovation-National Research Foundation (DSI-NRF Free Standing, Innovation and Scarce Skills Masters and Doctoral Scholarships) and the Agricultural Research Council-Biotechnology Platform.

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:

Dr Farai Catherine Muchadeyi

Date: 29 July 2022

Dr Edgar Farai Dzomba

Date: 29 July 2022

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, Nompilo Lucia Hlongwane, 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: Nompilo Lucia Hlongwane

Date: 29 July 2022

DECLARATION 2: PUBLICATIONS

My role in each paper and presentation is indicated.

Chapter 3

1. **Hlongwane, N.L.,** Dzomba, E.F., Hadebe, K., Soma, P., and Muchadeyi, F.C. 2017. Genome-wide SNP diversity of the South African domestic and wild pig populations. International Society for Animal Genetics Conference, 16th to 21st July, 2017, Dublin, Ireland. Presented by NL Hlongwane.

2. **Hlongwane, N.L.,** Hadebe, K., Soma, P., Dzomba E.F., and Muchadeyi, F.C. 2019. Detection of gentic variation in South African genetic resources using SMP60K panel. University of Kwa-Zulu Natal, School of Life Sciences, Postgraduate Research Day, 22nd May, 2019, Durban, Kwa-Zulu Natal. Presented by NL Hlongwane.

3. **Hlongwane, N.L.,** Hadebe, K., Soma P., Dzomba E.F., and Muchadeyi, F.C. 2020. Genome-wide assessment of genetic variation and population distinctiveness of the pig family in South Africa. Frontiers in Genetics 11:34. https://doi.org/10.3389/fgene.2020.00344.

Chapter 4

1. **Hlongwane, N.L.,** Dzomba, E.F., Hadebe, K., Soma, P., and Muchadeyi, F.C. 2019. Linkage disequilibrium and haplotype block partitioning in village, commercial and indigenous pigs of South Africa. 37th International Society for Animal Genetics Conference, 7th to 12th July, 2019, Lleida, Spain. Presented by K Hadebe.

Chapter 3 and 4

1. **Hlongwane, N.L.**, Hadebe, K., Dzomba, E.F., Soma, P., van der Nest, M.A., Pierneef, R., and Muchadeyi, F.C. 2019. Genetic diversity, linkage disequilibrium analysis using genomic data in South African pigs. 5th Professionl Development Program Postgraduate Conference, Agricultural Research Council-Vegetable and Ornamental Plants Institute, Roodeplaat, South Africa, 27th to 29th August, 2019. Presented by NL Hlongwane.

Signed: Nompilo Lucia Hlongwane

Date: 29 July 2022

ABSTRACT

South African pig production is dominated by the commercial sector with a small portion of the smallholder farmers. With an expected increase in pork consumption, commercial breeds such the Large White, South African Landrace and Duroc comes with proven growth potential. On the other hand, indigenous breeds like Kolbroek and Windsnyer have low growth potential, although they are well adapted to the hardy and harsh conditions of South Africa. Village pigs kept by communal farmers have potential to provide food security and assists in poverty alleviation. The free-roaming Warthogs, Bush Pigs and Wild Boars that are found in the villages and in national parks and zoos could be a contribution to the pig genetic resource of South Africa. However, the possibility of interactions of domestic and wild pigs is there and could be a concern to the pork industry, with potential challenges of indiscriminate crossbreeding and introgression. In addition, the zoos are home to some exotic and wild populations such as the Vietnamese Potbelly and Warthogs.

Genetic diversity, history and relationships of pigs found in South Africa is poorly understood. Pigs are currently understudied in South Africa resulting in a poor understanding of (i) the genetic diversity and how it is distributed within and among breeds and populations, (ii) the effect of the current production systems on the genetic diversity and threat to populations that has a potential to cause inbreeding and extinction and (iii) the interactions within and between domestic and wild pig populations. This lack of research and information is to the detriment of the species and the pig sector and makes it difficult to implement successful genetic improvement programs further predisposing the unimproved and indigenous pig populations to genetic loss and risk of extinction. This study aimed to close the information gap through genomic characterisation of the pig populations that are within the borders of South Africa. The study sampled broadly from domestic pig populations consisting of villages (Mopani, Capricorn, Alfred Nzo and O.R. Tambo), commercial (Large White, South African Landrace and Duroc), indigenous (Kolbroek and Windsnyer) and Vietnamese Potbelly representing Asian ancestry and wild pig populations consisting of Wild Boars, Warthogs and Bush Pigs. The pigs were sampled from diverse production systems consisting of intensive, semi-intensive and free-range/scavenging as well as pigs in national parks and zoos. The Illumina Porcine SNP60K was used to genotype the pigs generating 62 163 genome-wide SNPs that were analysed to infer of a number of genomic parameters that included (i) within and between population genetic diversity and population structure, (ii) linkage disequilibrium, haplotype

blocks and effective population size, (iii) population history, gene flow and introgression and finally (iv) genomic signatures of selection within and between populations.

Genetic diversity levels ranged from 0.204 ± 0.151 for Warthog to 0.371 ± 0.126 for Capricorn village pigs. Clustering of South African populations was comparable to the genetic cluster using the worldwide genotypes from a deposited database. There were four major clusters consisting of (i) Duroc, (ii) Vietnamese Potbelly, (iii) Bush Pigs and Warthogs and (iv) a large cluster of commercial Large White and South African Landrace, indigenous Kolbroek and Windsnyer and village pigs of Mopani, Capricorn, Alfred Nzo and, O.R. Tambo (villages) as well as Wild Boar populations. The PCA clustering which included the global populations also displayed four clusters. Bush Pigs and Warthogs clustered together while the Durocs formed a separate cluster. The Vietnamese Potbelly clustered with Chinese populations. The South African villages, worldwide village dataset, Wild Boars, South African indigenous and all the Large Whites and Landraces constituted one major cluster. Admixture based population structure had $K = 10$ as the optimal which made up of eight distinct clusters of the Warthogs and Bush Pigs; Wild Boars; Vietnamese Potbelly; Kolbroek; Windsnyer; Large White; South African Landrace and Duroc, as well as admixed clusters of Alfred Nzo together with O.R. Tambo (Eastern Cape Province) and Mopani and Capricorn (Limpopo Province). Clustering of populations also depicted the production systems the pigs are reared under. Generally, the effective population sizes of all pig populations have been on the decrease since 12 to 22 generations ago. The effective population sizes of indigenous and commercial breeds were lower compared to that of village populations. Overall, linkage disequilibrium and haplotype blocks results reflected the differences between breeds and production systems. Average LD ranged from 0.18±0.22 for Mopani to 0.40±0.36 for Kolbroek. Most populations (Alfred Nzo, Mopani, South African Landrace, Duroc and Windsnyer) had a high LD on chromosome 14. The highest mean correlation was observed at 1 kb between Duroc and Large White at r = 0.79±0.73. A total of 23 969 haplotypes were observed across populations with Duroc having the highest number of haplotype blocks of 3 402 while Alfred Nzo had the least at 1 350. Duroc pigs had the most unique haplotype blocks 2 513 and associated QTLs 562. Mopani shared the most haplotype blocks with 10 other populations while Mopani and Capricorn shared the most (963) haplotypes. Three haplotype blocks were shared across populations and the corresponding genes were associated with carcass quality and growth, immune response, fertility, and milk production. There was a total of 910 genomic regions with high LD (LD > 0.80) and most QTLs were concentrated on chromosome 7 at 236.

Analysis of population history, gene flow and introgression demonstrated that breeds with little to no admixture at K = 10, for Windsnyer, Kolbroek, Large White, South African Landrace, Duroc, Vietnamese Potbelly and Wild Boars had membership co-efficient ranging from 92 to 100 %. Village pigs of Alfred Nzo dominated cluster 3 with membership co-efficient of 85%. Phylogenetic analysis identified four clusters including cluster 1 (Vietnamese Potbelly, Warthogs and Bush Pig), cluster 2 (Alfred Nzo and O.R. Tambo), cluster 3 (South African Landrace and Large White) and cluster 4 (Wild Boar, Duroc and Kolbroek). Migration events were set at 10, M = 1 reflected gene flow from Wild Boar to Mopani while f_3 and f_4 tests confirmed admixtures between populations. Evolutionary relationships revealed ancestries of village populations from Alfred Nzo (21% WIN, 22% DUR, 37% ORT, 20% SAL); O.R. Tambo (16% WIN, 72% ALN, 6% LWT, 6% SAL); Mopani (16% WIN, 26% DUR, 22% LWT, 18% SAL, 18% ALN) and Capricorn (22% WIN, 23% SAL, 24% ALN, 18% LWT, 14% DUR). CAP_ORT had the highest distribution of shared IBD segments of 21 190 with the longest shared segments going up to 25.11 Mb shared between Mopani and O.R. Tambo. The number of genes shared ranged from 13 751 for ALN_ORT to 13 888 for ALN_MOP. Meat quality gene *DECR1* (chromosome 4) was identified in all populations pairs whereas feed efficiency (*DNAJC15*) and fertility (*EPSTI1*) genes were observed on chromosome 11. This analysis demonstrated the effects of domestication, migration, genetic drift and downstream production system driven selection on the pig populations of South Africa. Furthermore, with the analysis of genomic signatures of selection, the study revealed genes associated with traits of economic traits across the different pig populations. Candidates' genes *DECR1*, *DLX1*, *BRPF1*, *CLPTM1*, *FANCD2*, *SEC13*, *FHL3*, *FSTL5*, *CEP135*, *EXOC1*, *FOXO1*, *ASTN2*, *MYO18B*, *PLXNA1*, *DNAH2*, *HECTD2*, *TMEM39B*, *TXLNA*, *CSMD2*, *COL16A1*, *SCARA3*, *ZFAND3* and *PTPRD* that are associated with meat and carcass quality traits were observed to be under selection in these pig populations. The observation of genes associated with meat and growth as well as immune and adaptation traits is suggestive of the potential indigenous breeds to contribute to pork meat production.

EXTENDED ABSTRACT

Pigs are important in agriculture as they produce animal-based protein for human consumption. Knowledge of genetic diversity, levels of inbreeding, effective population size and linkage disequilibrium can assist in defining better pig genetic improvement programs culminating in sustainable use of pig genetic resources. The history and pattern of introgression of populations have implications on the maintenance and utilization of genetic diversity and together with analysis of selection of signatures can reveal genes associated with the phenotypic traits as a result of either natural or artificial selection. Pig populations are poorly characterized in South Africa. Hence studies aimed at evaluating genetic distinctiveness and pig breed diversity will contribute to developing a rational plan for population conservation programs among other applications.

A total of 234 pigs representing 91 non-descript village pigs from Mopani (n = 27), Capricorn (n = 25) in Limpopo and O.R. Tambo (n = 22) and Alfred Nzo (n = 17) in Eastern Cape provinces; 60 commercial pigs from Duroc (n = 20), Large White (n = 20) and South African Landrace (n = 20); 40 indigenous pigs from Kolbroek (n = 20) and Windsnyer (n = 20); 38 wild pigs from Wild Boar (n = 4), Warthog (n = 31) and Bush Pig (n = 3) as well as 5 Asian pigs of the Vietnamese Potbelly were genotyped using the Porcine SNP60K Bead Chip (62 134 SNPs). Quality control removed animals with call rates less than 85%, SNPs with minor allele frequency < 0.02 and Hardy Weinberg. Equilibrium of $P < 0.0001$. The quality-controlled genotypes were used to characterize the genomic architecture of South African pig populations. The first analysis investigated the genetic diversity and population structure of South African populations and compared it to the worldwide pig populations. The worldwide pig populations consisted of 389 genotypes from 24 countries and included villages, indigenous, commercial and wild pigs. Cluster analysis resulted in four distinct clusters aligned to some extent to genetic similarities, geographic location and production systems. Village populations showed highly admixed individuals presumably as a result of indiscriminate crossbreeding in the communal production systems. Population pairwise F_{ST} analysis showed genetic differentiation ($P \leq 0.05$) between the village, commercial and wild populations. A per marker per population pairwise F_{ST} analysis revealed SNPs associated with QTLs for traits such as meat quality, cytoskeletal and muscle development, glucose metabolism processes and growth factors between both domestic populations as well as between wild and domestic breeds.

The second study described the genome-wide extent of LD and haplotype block structure within and between South African pig populations from different genetic and production backgrounds. The overall mean LD for villages pigs was smaller (0.19±0.24) compared to commercial (0.32±0.33) and indigenous (0.34±0.33) while the persistence phase between the commercial and indigenous pairs was highly correlated at distance of 0-1 kb at 0.74±0.03. The Kolbroek also had the greatest number of haplotype blocks, while the conserved population of Kolbroek and Duroc populations showed the highest levels of haplotype diversity. The village populations of O.R. Tambo, Capricorn and Mopani shared more haplotypes than the other populations. *DECR1* gene related to meat quality and growth rate traits was detected within the shared haplotypes, revealing the adaptive traits in the populations. High LD regions revealed significant QTLs associated with meat and carcass quality and fertility traits such as QTLs for IMF content, marbling, feed intake, body weight, litter size, teat number, and behavioural, traits amongst others.

The inconsistent and unreliable information available in relation to the history of pigs in South Africa prompted this study to investigate the population history, ancestral relationships, breed interactions, and gene flow between populations Village populations showed the most cluster variation while the phylogenetic construction confirmed the Capricorn population as being the most introgressed by the commercial pigs and other village populations. Vietnamese Potbelly appeared to be a common ancestor for all the studied populations. Alfred Nzo appeared to be an ecotype as it was less admixed compared to other village populations and was a common ancestor to all other village populations. The gene flow of Wild Boar and Vietnamese Potbelly confirmed Europe and Asia as the presumed domestication centres. Long shared IBD segments dominated in the village pigs signifying inheritance from a recent common ancestor. Significantly ($P \leq 0.01$) enriched GO terms of shared IBD segments were associated with meat quality, feed efficiency, muscle growth and fertility.

The final analysis of this study investigated signals of selection within and between populations using three different approaches of (i) Integrated Haplotype Score (*iHS*) that was used to investigate recent positive and past selection within a population; (ii) *XP-EHH* that identified long-range haplotypes between pairs of populations with selected SNPs under selection in one population while not being selected in another and (iii) *HapFLK* method that detected signatures of selection based on differences of haplotype frequencies between populations on a multiple population genotyping data. Findings from this signature of selection analysis

highlighted genomic regions, some harbouring QTLs of traits of interests within and between wild and domestic pigs. Several candidate genes related to growth, exterior traits, reproduction and meat and carcass quality traits were identified. The large number of genetic variants detected in this study offers an opportunity to further explore the genetic diversity underlying different phenotypes in domestic and wild pigs.

Overall, this study was able to provide a baseline understanding of South African porcine diversity and identify high LD regions and shared haplotype blocks for the application and development of genomic tools for pig studies. Additionally, the study established the different populations' genetic history inferred on breeding patterns of South African pig populations and increased our knowledge regarding the gene flow and introgression of our pig genetic resource. Traits of economic importance affecting phenotypic characteristics and related to differences in adaptation to specific environments and productive systems were identified. Importantly, the uniqueness and importance of village pigs as a national and global genetic resource was validated.

Keywords: Pigs, genetic diversity, haplotype blocks, introgression, signatures, Porcine SNP60K

ACKNOWLEDGMENTS

KoNgwane kaNgwadi, Ozikhali ezingalingani nezamaxhegwana. Ngiswele imilomo ezinkulungwane namazwi anembayo ngidllulisa ukubonga.

It is with immense gratitude that I acknowledge the support and assistance of my supervisor Dr Farai Muchadeyi. Her mentorship, guidance and professional support throughout this journey has opened doors for me. Thank you for your understanding, giving me opportunities and allowing me space to grow as a scientist. This would have not been possible without the academic support, endless patience and positive attitude of my co-supervisor Dr Edgar Dzomba.

To Dr Khanyi Hadebe, Dadewethu, I am forever grateful for the friendship, strength and above all, for being my light on my darkest days. Like I always say to you, ubizelwe kulento. My sincere appreciation goes to Dr Magriet van Der Nest. You have always given me moral support and believed in my capabilities. I would have not been able to complete this thesis without your generosity by providing your laptop when technology failed me. I am grateful to both of you for taking interest in my study and for consistence support. Most importantly, you kept pushing me through hard times.

I am indebted to the Agricultural Research Council Biotechnology Platform, Agricultural Research Council Professional Programme and National Research Foundation Scarce Skill Scholarship for providing financial support towards this study.

I would like to thank Dr Olga Makina for leaving me in the capable hands of Dr Muchadeyi and Dr Dzomba. Thanks also goes to Dr Pranisha Soma for her encouragement. I am incredibly thankful to Dr Nokuthula Mchunu, for taking care of my son and providing stability during times I spent away from home. You are a blessing in my life.

I am grateful to the bioinformatics team from Biotechnology Platform: Dr Rian Pierneef, Annie Chan and Dr Quentin Santana. I have learnt so much over the years and my bioinformatics skills is better because of your contributions.

Animal Genomics team is acknowledged for their assistance during sampling: Dr Petunia, Dr Keabetswe, Prescilla, Njabulo and Michael. A sincere appreciation to the agriculture provincial

departments' officials from Limpopo and Eastern Cape, SANParks and Pilanesberg National Parks, Bontebok Ridge Reserve, Johannesburg Zoo and Mockford Farm. A big thank you goes to all the farmers in the villages for giving access to get samples from their animals.

I would like to thank my fellow postgraduate students Dr Simon, Njabulo, Dr Petunia, Dr Bongo Prescilla, Nonhle, Mary and Madeline for walking this journey with me as students. I give thanks to Khulekani Khanyile and Alister Ngobeni for assistance with processing of samples. I also thank Lerato Mpye for all the time spent in her office trying to source samples and making sure my paperwork was always in order. Her importance on administrative issues at BTP cannot be ignored.

Thanks to my friends; MaKhumalo, Cate, Nozipho, Jabu, Onke, Lemo, Thina, Sindy, Maemu, Refilwe and Nokuthula. Your support has been amazing. Please forgive me if I forgot your name. Mnganami, Samke, your untimely death forever remains painful in my heart. Thanks for all the constructive contribution.

To Prof Lucky Nedambale and Mr Tlou Chokoe, you might have not been part of this project, but I shall never forget your positive contributions in my academics.

Special thanks go to my parents, Mr Thamsanqa (Sandile) Hlongwane and Mrs Zandile Hlongwane. I would have not reached this level without your incredible support. You have always been there for me, and Sisonke and all your sacrifices are highly appreciated. My son, Sisonke Hlongwane, my love for you have always encouraged me. You have been the most patient and understanding son during my absent days. My extended family: Dumakude, Magwaza and Gcabashe. Thank you for the unconditional support and keeping me motivated. My niece, Thenjiwe Hlatshwayo, your kind words and actions kept me sane in the most challenging times.

This is dedicated to the late Mr Michael Dumakude (grandfather), Ms Thokani Dumakude (Mam'ncane) and Mr Jabulani Hlongwane (Uncle) for shaping my life. May your souls continue resting in peace.

Last but not least, I thank the Almighty Lord for granting me the strength, wisdom and protection during the course of my study.

TABLE OF CONTENTS

Page

PREFACE ii
DECLARATION 1: PLAGIARISM iii
ABSTRACT v
EXTENDED ABSTRACT viii
ACKNOWLEDGMENTS xi
TABLE OF CONTENTS xiii
LIST OF TABLES xviii
LIST OF FIGURES xx
ABBREVIATIONS 1
CHAPTER 1: INTRODUCTION 3
1.1 Rationale for the research (nature and scope) 3
1.2 Justification 9
1.3 Aims 10
1.4 Objectives 11
1.5 Animal populations and SNP genotypes 11
1.6 Outline of dissertation structure 12
CHAPTER 2: LITERATURE REVIEW 14
2.1 Introduction 14
2.2 Domestication of pigs 16
2.3 The South African pig industry 17
2.4 Pig production systems in South Africa 20
2.5 South African pig breeds 22
2.5.1 Commercial breeds 22
2.5.1.1 Large White 23
2.5.1.2 South African Landrace 23
2.5.1.3 Duroc 24
2.5.2 Indigenous breeds 25
2.5.2.1 Kolbroek 25
2.5.2.2 Windsnyer 26

2.5.3 Wild pigs 26
2.5.3.1 European Wild Boar 26
2.5.3.2 Warthog and Bush Pig 27
2.5.4 Village populations 29
2.6 Crossbreeding commercial and indigenous breeds 29
2.7 Introgression and hybridization of pigs 30
2.8 Genetic studies of South African pig breeds 31
2.9 Genomic platform tools and techniques 32
2.10 Factors influencing genetic diversity in pigs 32
2.10.1 Natural and artificial selection pressures 32
2.10.2 Migration (Gene flow) 34
2.10.3 Geographical isolation and fragmentation 34
2.11 Population genomic parameters 35
2.11.1 Linkage Disequilibrium 35
2.11.2 Effective population size 37
2.12 Identity by Descent (IBD) 37
2.13 Selection of signatures 38
2.13.1 Fixation index (F_{ST}) 40
2.13.2 Haplotype based methods 40
2.14 Conclusion 41
CHAPTER 3: GENOME-WIDE ASSESSMENT OF POPLATION VARIATION AND POPULATION DISTINCTIVENESS OF THE PIG FAMILY IN SOUTH AFRICA 42
3.1 Abstract 42
3.2 Introduction 43
3.3 Materials and methods 45
3.3.1 Breed/populations sampled 45
3.3.2 Genotyping and quality control 46
3.3.3 Genetic diversity within population 48
3.3.4 Population differentiation and structure 49
3.4 Results 50
3.4.1 Genotype and quality control 50
3.4.2 Genetic diversity across populations 51
3.4.3 Effective population size 51
3.4.4 AMOVA 52

3.4.5 Population structure 53
3.4.6 Population differentiation 56
3.4.7 Per marker pairwise F_{ST}, SNP annotation and association with Porcine QTL 57
3.5 Discussion 60
3.6 Conclusion 67
CHAPTER 4: THE EXTENT OF LINKAGE DISEQUILIBRIUM AND HAPLOTYPE BLOCK STRUCTURE OF COMMERCIAL, INDIGENOUS AND VILLAGE PIGS 69
4.1 Abstract 69
4.2 Introduction 70
4.3 Materials and methods 72
4.3.1 Genotyping and quality control 72
4.3.2 Linkage disequilibrium (LD) analysis 73
4.3.3 Persistence phase 74
4.3.4 Haplotype block partitioning and diversity 74
4.3.5 Unique and shared haplotype blocks 74
4.3.6 Annotation of high LD regions 75
4.4 Results 75
4.4.1 Sample and SNP quality control 75
4.4.2 Linkage disequilibrium (LD) 75
4.4.3 Persistence phase 77
4.4.4 LD haplotype block partitioning and haplotype diversity 79
4.4.5 Analysis and annotation of unique and shared haplotype blocks 81
4.4.6 Annotation of high LD regions 83
4.5 Discussion 84
4.6 Conclusion 90
CHAPTER 5: GENOMIC ANALYSES REVEAL POPULATION HISTORY AND INTERACTIONS OF SOUTH AFRICAN PIG POPULATIONS 91
5.1 Abstract 91
5.2 Introduction 91
5.3 Materials and Methods 94
5.3.1 Animal samples, genotyping and quality control 94
5.3.2 Population structure and ancestry estimation 95
5.3.3 Phylogenetic analysis 95
5.3.4 TreeMix, f_3 and f_4 analysis 95

5.3.5 Evolutionary relationships and tracing admixture using PCAdmix 96
5.3.6 Pairwise IBD detection 97
5.3.7 GO-enrichment analysis 97
5.4 Results 97
5.4.1 Sample and SNP quality control 97
5.4.2 Population structure and ancestry estimation 98
5.4.3 Phylogenetic analysis 99
5.4.4 TreeMix, f_3 and f_4 analysis 99
5.4.5 Evolutionary relationships and tracing admixture using PCAdmix 101
5.4.6 Pairwise IBD detection 104
5.4.7 GO-enrichment analysis 105
5.5 Discussion 107
5.5 Conclusion 113
CHAPTER 6: IDENTIFICATION OF SIGNATURES OF SELECTION OF AND AFFECTED QTLS IN PIGS THROUGH GENOMIC SCANS USING DIFFERENT METHODS 115
6.1 Abstract 115
6.2 Introduction 115
6.3 Materials and Methods 118
6.3.1 Samples, data and quality control 118
6.3.2 Detection of signatures using iHS 119
6.3.3 Detection of signatures using XP-EHH 119
6.3.4 Detection of signatures using HapFLK 120
6.3.5 Annotation and function analyses of identified genomic regions 120
6.4 Results 121
6.4.1 Detection of signatures within a population using iHS 121
6.4.2 Detection of selection of signatures between populations using XP-EHH 126
6.4.3 Detection of selection of signatures between populations using HapFLK 132
6.4.4 Genes identified using different signature of selection methods 137
6.4.5 Functional enrichment analysis 137
6.5 Discussion 138
6.5 Conclusion 141
CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH 143

7.1 Introduction 143
7.2 Aims and objectives 145
7.3 General discussion 146
7.4 Conclusions 150
7.5 Study challenges 150
7.6 Final comments and recommendations 151
REFERENCES 153
APPENDIX A: SUPPLEMENTARY FILES 191

LIST OF TABLES

Table **Page**

Table 2.1 Systems of pig production 22

Table 3. 1 Summary of the sampled South African pig populations 46

Table 3. 2 Summary of the genetic diversity measures across South African pig populations 51

Table 3.3: Pairwise genetic differentiation (F_{ST} values) between 10 pig populations 57

Table 3.4 Most significant SNPs detected with F_{ST} analysis and the associated genes 58

Table 4.1: Mean and standard deviation of r^2 for chromosome one to eighteen in South African pig breeds and populations 76

Table 4.2 The effects of population, chromosome and SNP marker interval and population X chromosome interaction on linkage disequilibrium (r^2) 77

Table 4.3 Descriptive summary of haplotype block distribution for the nine populations used in this study 80

Table 4.4 Haplotype frequencies and percentages for all haplotypes in each window in all the pig populations 81

Table 4.5 Functions associated with haplotype blocks that were shared by all the nine populations 83

Table 4.6 QTLs associated with high LD regions 84

Table 5.1 *Frappe* based estimates of proportions for South African pig populations 98

Table 5.2 Significant f_3 statistics for pig populations. Three population tests were performed on all possible combinations of populations. A negative Z-score of 3 or more correlating to a *p-value* of 0.001 101

Table 5.3 Significant f_4 statistics for pig populations. Four population tests were performed on all possible combinations of populations. A negative Z-score of 3 or more corresponds to a *p-value* of 0.001 102

Table 5.4 Chromosome genome sharing for population pairs 105

Table 5.5 The top enriched (*P < 0.01*) Gene Ontology (GO) terms relevant to the pairwise for biological process 106

Table 6.1 Within population list of genomic regions under selection within and candidates' genes detected using the *iHS* method 121

Table 6.2 Selected regions and candidates genes detected between pairs of populations using the *XP-EHH* method 126

Table 6.3 Regions under selection detected by *HapFLK* methods in South African pigs (*P-value* < 0.10) 133

Table 6.4 Common genomic regions and genes under signatures of selection detected using *iHS*, *XP-EHH* and *HapFLK* methods 137

LIST OF FIGURES

Figure **Page**

Figure 3.1 Map showing geographic locations of the 13 pig populations in the present study 47

Figure 3.2 Average effective population size plotted against generation in the past 52

Figure 3.3 Principal Component Analysis based population clustering 54

Figure 3.4 Principal Component Analysis based population clustering including Burgos *et al.* (2013) genotypes (22 430) 55

Figure 3.5 ADMIXTURE based clustering $K = 2 – K = 10$. Each individual is represented by a single column divided into K coloured segments, where *K* is the number of clusters assumed with lengths proportional to each of the K inferred cluster. 56

Figure 4.1 LD decay represent the average r^2 according to marker distances for the 18 chromosomes of the nine South African pig breeds/populations 77

Figure 4.2 Pearson correlation between breed pairs ($n = 36$) at different physical distances. (a) Eastern Cape (EC) and Limpopo (LIM), b Villages (VIL) and Commercial (COM), (c) Villages and Indigenous (IND), (d) Commercial (COM) and Indigenous (IND)..... 79

Figure 4.3 Distribution of unique haplotype blocks for pig populations 82

Figure 4.4 Percentage of unique SNPs of haplotype blocks for the studied populations 82

Figure 5.1 Phylogenetic tree based on the South African pig populations. The tree was constructed by the neighbour-joining method using TreeMix program without migration edges ($M = 0$) population 99

Figure 5.2 Maximum likelihood tree inferred from 13 populations with migration edges based on population pairwise allelic covariance inferred with TreeMix (Pickrell & Pritchards, 2012). Coloured lines reflect the weights of migration according to the colour shades of lineage obtained from the contributed population..... 100

Figure 5.3: a) Proportions of ancestry from assuming six ancestral populations in village individual populations b) Genome admixture simplified using PCADMIX for village individuals' haplotypes based on 6 ancestral populations 103

Figure 5.4 Distribution of short and long segments that are identical by descent (IBD) for pairs of individuals from populations 104

Figure 6.1a Manhattan plot of the genome-wide distribution of selection of signatures detected by *iHS* across the 18 chromosomes (Villages) 124

Figure 6.1b Manhattan plot of the genome-wide distribution of selection of signatures detected by *iHS* across the 18 chromosomes (Commercial) 124

Figure 6.1c Manhattan plot of the genome-wide distribution of selection of signatures detected by *iHS* across the 18 chromosomes (Indigenous and Vietnamese Potbelly) 125

Figure 6.1d Manhattan plot of the genome-wide distribution of selection of signatures detected by *iHS* across the 18 chromosomes (Wild pigs) 125

Figure 6.2a Manhattan plot of the genome-wide distribution of selection of signatures detected by *XP-EHH* across the 18 chromosomes (Duroc group) 129

Figure 6.2b Manhattan plot of the genome-wide distribution of selection of signatures detected by *XP-EHH* across the 18 chromosomes (Warthog group) 130

Figure 6.2c Manhattan plot of the genome-wide distribution of selection of signatures detected by *XP-EHH* across the 18 chromosomes (Vietnamese Potbelly group) 131

Figure 6.3 Manhattan plot for signature of selection of South African pig populations detected by *HapFLK* across 18 chromosomes 132

ABBREVIATIONS

AD - Average Daily Gain

AI - Artificial Insemination

AMOVA - Analysis of molecular variance

ARC -Agricultural Research Council

ASF - African Swine Fever

BFAP - Bureau for Food and Agricultural Policy

BLUP - Best Linear Unbiased Prediction

cM - Centi morgans

CSF - Classical Swine Fever

DAFF - Department of Agriculture Forestry and Fisheries

DALLRD - Department of Agriculture, Land Affairs and Rural Development

DMI - Dry matter intake

EBVs - Estimated Breeding Values

EHH - Extended Haplotype Homozygosity

EM - Expectation Maximization

FAO - Food and Agricultural Organisation

FCR - Feed Conversion Ratio

FDR - False discovery rate

FMD - Foot and mouth disease

F_{ST} - Fixation index

GDP - Gross domestic product

GWAS - Genome-wide association studies

HWE - Hardy-Weinberg equilibrium

IBD - Identity by descent

iHS - Integrated haplotype score

IMF - Intramuscular fat

KZN - KwaZulu-Natal

LD- Linkage disequilibrium

MAF - Minor allele frequency

ML - Maximum likelihood

NAFU - National African Farmers Union

NGS - Next generation sequencing

NJ - Neighbour-joining

PBS - Pig Breeders Society of South Africa

PCA - Principal component analysis

PEDv - Porcine Epidemic Diarrhoea virus

PIC - Pig Improvement Company

PSE - Pale, soft flesh and exudative

QC - Quality control

QTL - Quantitative trait locus

SA - South Africa

SADC - South African Development Community

SAPPO - South African Pork Producers

SNP - Single Nucleotide Polymorphism

SSR - Simple Sequence Repeat

XP-EHH - Cross population extended haplotype homozygosity

CHAPTER 1: INTRODUCTION

1.1 Rationale for the research (nature and scope)

South African pigs (*Sus scrofa domestica*) provide a less expensive alternative source of animal protein compared to ruminants' species such as cattle, sheep and goats (Weinberger & Qaim, 2009; Antwi & Seahlodi, 2011; Dietze, 2011). Pigs also play a role towards poverty alleviation, food security and employment opportunities in the rural areas. In June 2021, the Bureau for Food and Agricultural Policy (BFAP) in conjunction with The South African Pork Producers' Organisation (SAPPO) reported the number of pigs in the country to be 2 800 million (BFAP, 2021). Pigs are distributed in all the nine provinces of South Africa with Limpopo having the largest proportion at 25% (DALRRD, 2021). The Eastern Cape and Northern Cape have the lowest proportion of pigs at less than a percent each. Pork production provides around 2.45% GDP of primary agricultural sector (BFAP, 2021). Twenty-seven million pigs have been slaughtered in the past decade with an average annual production of over three million pigs being slaughtered in 2021 and Gauteng province having the most slaughtered pigs at 126 659 (SAPPO, 2022).

There are three categories of pig farmers in South Africa consisting of the large-commercial, smallholder communal and emerging commercial farmers. Commercial farmers are associated with intensive systems of production, while smallholder with semi-extensive, backyard and free-range farming. These sectors differ by the breeds they keep, infrastructure, facilities and management (Mokoele *et al*., 2014). Smallholder farmers tend to keep less than 50 pigs (Munzhelele *et al*., 2017). The total number of pigs owned by the smallholder farmers in rural areas were estimated at 1 400 million (50%) (BFAP, 2021). There are 4 000 producers in the commercial sector with production units having anything between 600-5 000 sow units (DALRRD, 2021).

Commercial breeds such as the Large White, South African Landrace and Duroc are raised under the intensive production systems (Swart *et al*., 2010). The Large White was given recognition as breed in England in 1884 while the South African Landrace has its origins in Denmark (Briggs, 1983; Treacy, 1976; Swart *et al*., 2010). Duroc pigs were developed in the USA with the first import to South Africa from Canada taking place in 1980 (Visser *et al*., 1993; Rothschild & Ruvinsky, 1998; Swart *et al*., 2010). Commercial breeds are provided with commercial feed and supplement diets such as vitamins, probiotics and Omega 3 (Hossain *et*

al., 2011). In a previous study, pigs reared under intensive systems had better health care, hygiene, biosecurity and disease outbreaks were well managed (Mokoele, 2015). Well-conditioned pigs like commercial breeds tend to produce large number of litters when established management conditions are provided (Swart *et al*., 2010). Economic traits such as growth, reproductive performance and carcass and meat quality are important in the intensive production system (Munzhelele *et al*., 2017). Even though the commercial breeds are high performance, they are not as hardy and resilient as indigenous breeds. South African commercial pig genetics are imported to be able to meet the customers' needs (Swart *et al*., 2010). This often leads to mating of large number of sows with few elite boars using artificial insemination (AI) to increase productivity, thus resulting in reduced genetic diversity of founding populations.

Non-descript village pigs are reared by smallholder farmers in the rural areas under limited resources (Madzimure *et al*., 2013). The cost of commercial feed is high for the rural farmers and as such farmers mostly provide kitchen waste and other low-quality feed to their pigs (Madzimure *et al*., 2013). Pigs reared under the smallholder farmers are characterised by long farrowing intervals and small litter sizes (Manchidi, 2009). Lack of infrastructure as well as uncharacterised and unselected genetics contributes to the low productivity in the village production (Madzimure *et al*., 2013). The village pigs are characterized by small and narrow frame size with long legs, short forehead and elongated snout (Plug, 2001). On the other hand, village pigs also have good mothering abilities, high fertility, heat tolerance, hardiness, foraging ability, tolerance to endemic diseases and parasites, lean meat and adaption to low input management levels (Plug, 2001; Mutua *et al*., 2010). While high production breeds may not compete with low-input breeds, indigenous livestock harbour a large amount of the genetic variation (Tapio *et al*., 2006; Olivier, 2009).There is however very little information on the genetic background of village pigs. Other studies have suggested that factors such as migration, imports, and admixture levels have contributed to village pigs having more genetic diversity than commercial breeds (Mujibi *et al*., 2018). Their existence in harsh and marginalised environments has presumably contributed to unique adaptive genetic merits (Munzhelele *et al*., 2017). However, any valuable adaptation traits in village pigs are being threatened by the constant genetic erosion through indiscriminate crossbreeding with commercial breeds (Mokoele *et al.*, 2014. This is caused by the farmer's intention to improve their livestock for faster growth and access to high value markets (Mokoele *et al*., 2014).

In addition to commercial and predominantly exotic breeds and the non-descript village pigs, South Africa is home to indigenous Kolbroek and Windsnyer pigs. Although the history of the Windsnyer and Kolbroek pigs is not fully known, Moyo *et al*. (2018) reported the Windsnyer to be related to the Wild Boar (*Sus scrofa*) and its ancestry being for Europe. Windsnyers are indigenous to Southern African countries such as Zimbabwe and Mozambique and resemble the local Mukota pigs (Phogole, 2017; van Tonder, 2019). Ramsay *et al*. (1994) speculated the Kolbroek to be of Chinese origin. Swart *et al*. (2010) described the Kolbroek and Windsnyer as being hardy, able to survive under harsh conditions and suitable to be raised in rural areas. These breeds are also able to survive on little or no inputs (Masenya *et al*., 2011). The indigenous Kolbroek and Windsnyer are also tolerant and/or resilient to disease and parasites (Mohlatlole *et al*., 2013; Ncobela, 2017). According to Madzimure *et al*. (2012) and Moyo *et al*. (2018) the indigenous Windsnyer has longer black hair and thinner epidermis for increased heat tolerance. Halimani *et al*. (2010), Madzimure *et al*. (2012) and Mokoele *et al*. (2014) noticed that the indigenous Kolbroek and Windsnyer are prevalent in Limpopo and Eastern Cape provinces of South Africa. The Agricultural Research Council, South Africa also retains a conserved population of Kolbroek and Windsnyer populations. A shift from farming indigenous pigs to commercial breeds has been observed in KZN (Gcumisa *et al*., 2016), Mpumalanga (Munzhelele *et al*., 2017) and Limpopo (Mokoele *et al*., 2014).

Although not contributing to the agricultural sector, the Warthogs, Bush Pigs and conservation pigs like the Vietnamese Potbelly and other breeds constitute and contribute to the South African pig population. The Vietnamese Potbelly has its ancestry in South China (Reimer *et al*., 2018). Glodek and Oldigs (1981) and FAO (2005) reported the original Vietnamese Potbelly being crossbred with European breeds. Their abandonment in Spain has also led to interbreeding with wild populations (Delibes-Mateos & Delibes, 2013). Currently in South Africa, the Vietnamesse Potbelly is kept as captured and conserved populations at the Johannesburg Zoo. The common Warthog (*Phacochoerus africanus*) is widely distributed and is free roaming in South African national parks with incidents of escaping happening now and then (Swanepoel *et al*., 2010). The earliest Warthog fossils date back to 3.2 million years ago in Africa (Harris, 2013a). Gongora *et al*. (2011) reported the Giant Forest Hog (*Hylochoerus meinertzhageni*) to have diverged from the common Warthog 7 to 14 million years ago while also sharing an ancestor with the Desert Warthog (*Phacochoerus aethiopicus*) until 4 to 8 million years ago. Characterizing wild pigs is important to understand disease susceptibility and tolerance which may have an impact on pig production and therefore pose a threat to food

security. The Warthogs are hosts and reservoir to ticks causing African Swine Fever (ASF) Classical Swine Fever (CSF) and Foot and mouth diseases (FMD) and can transmits to and between domestic, Bush Pigs and Wild Boars resulting in great economic impact (Dixon *et al*., 2019). Bush Pigs (*Potamochoerus larvatus*) are extensively indigenous and distributed in Western Central, Eastern and Southern Africa. They have been reported in all the provinces of South Africa with the exception of Northern Cape (Venter *et al*., 2016). Bushmeat is popular in rural Africa and Bush Pigs might be declining or facing extinction if there are no monitoring strategies in place (Lindsey *et al*., 2013). There is also a potential occurrence of transmission of diseases to humans and domestic pigs. In an experimental study by Gers *et al*. (2011), the Bush Pigs presented clinical symptoms of Classical Swine Fever (CSF) similar to domestic pigs. Venter *et al*. (2016) also reported on interbreeding between the Bush Pigs and Wild Boars in South-Eastern KZN. This suggests the importance of understanding their genetic diversity and possibilities of interbreeding and genetic introgression. Wang *et al*. (2021) noted the negative effects of interbreeding and introgression in chickens of genomic differences and functions introduced by introgression not yet known, negative effects of interbreeding and introgression is found, whereas for pigs, the admixture values of introgressed populations is unusually high compared to other indigenous Chinese breeds (Yang *et al*., 2003; Wang *et al*., 2021). Outbreeding depression and genetic adaptation to captivity are some of the adverse effects of interbreeding that were reported by Frankham and Ralls (1998).

Overall, the South African pig population is diverse consisting of many breeds and populations raised under different production systems. This genetic pool needs to be protected for future food security (FAO, 2015). The predicted climate change resulting in increased temperatures in South Africa will negatively affect pig production particularly the exotic breeds. Natural and artificial disasters are a risk to genetic diversity and often communities and governments are not prepared to respond and mitigate the effects (Rischkowsky & Pilling, 2007). Disease outbreaks and indiscriminate breeding threaten the existence of indigenous breeds (Halimani *et al*., 2010). The outbreak of African Swine Fever and Classical Swine Fever (CSF) in the Western and Eastern Cape provinces was found to be predominant in the free-range production systems (NAFU, 2007; Penrith *et al*., 2011). Many pigs were lost and restocking after culling could not have replaced the lost biodiversity as there were no records of the animals that were kept (Halimani *et al*., 2010). While non-descript and free-range pigs are found in the villages, indiscriminate crossbreeding and the use of commercial pigs to improve the indigenous breeds are prevalent (Matabane *et al*., 2015). Chances of wild and domestic pigs mixing are high in

the free-range production system with possibilities of interbreeding and introgression that will give rise to dilution and loss of diversity in local pig populations (Penrith *et al*., 2011). This dilution leads to the loss of adaptability traits needed to survive in the harsh and marginalised rural areas. Genetic loss by dilution and eradication of the indigenous genetic pool has also been reported (Rege & Gibson, 2003; Halimani *et al*., 2010). Not only does crossbreeding contributes to the non-descript pigs often seen in the villages. It is also a threat to the existence and preservation of indigenous genotypes.

Declining diversity amongst breeds also mean the size of a population is getting smaller. Inbreeding is therefore likely to increase without any formal breeding strategies in place. A loss of the genetic diversity and increased risk of extinction is a major concern for many livestock species including pigs (Frankham, 2003). The loss of genetic diversity in a breed could cause reduction of survival, reproduction efficiency and ability to adapt to environmental changes (Frankham & Ralls, 1998). FAO (2000) indicated that a number of adapted breeds are at the brink of extinction. These breeds facing extinction are in developing countries (Rege & Gibson, 2003). With Africa being home to 49 indigenous pigs, 5% are endangered and 49% have an unknown status. The number of breeds declining means conservation is needed. If nothing is done, important genetic diversity may be lost of which the impact on food security for smallholder farmers and their communities will be damaging. It is therefore important that population diversity be characterised and maintained. Preservation of genetic variation and population distinctiveness is important to allow sustained use and genetic improvement of the current pig genetic resources.

The possibility of domestic and wild pigs interbreeding cannot be ignored. Hybridization is known to have the prospect to generate diversity that is needed to allow populations to respond to abrupt environmental changes (Baskett & Gomulkiewics, 2011). Although there are advantageous effects of genetic introgression including heterosis, improved growth rates and larger litter size, wild pigs can be challenging to control (Goedbloed *et al*., 2013). Venter *et al.* (2016) noticed that a feral population of Wild Boars are mating with local Bush Pigs in South-Eastern KwaZulu-Natal. Jori and Bastos (2009) suggested that a hybrid of domestic and Bush Pigs could become asymptomatic carrier animals that can be able to maintain and disseminate African Swine Fever in domestic pigs. Any evidence of introgression of wild pigs' genes into the domestic pig population would affect their genetic integrity and possibly cause outbreeding depression. Further and deep characterisation of the pig population of South Africa is therefore

essential to unveil the genomic architecture and its implications on sustainability and productivity of pig populations.

Lack of information is a major threat to optimal use and conservation of South African pig resources. There is very limited information available on South African pigs' characteristics, genetic attributes and their fitness for different environment (Ramsay *et al*., 2000). Studies have been published in Southern Africa on pig production (Mashatise *et al*., 2005), disease and parasite tolerance (Zanga *et al*., 2003), genetics (Chimonyo & Dzama, 2007) and diversity (Ramírez *et al*., 2009; Swart *et al*., 2010).

Genetic characterization of indigenous and local genetic resources is a pre-requisite particularly in the face of such challenges and threats to biodiversity that are prevalent in South Africa and other countries globally. Genetic characterisation would uncover the genetic variation in domestic and wild pigs and allow sustainable utilization of pig breeds. The South African pig population needs to be characterised to facilitate optimal utilisation and conservation and to prepare the sector in case of future disease outbreaks.

Historical molecular markers had limitations in characterizing the South African pig population. Genetic parameters such as levels of diversity, gene flow and interaction of both wild and domestic populations, effective population size and genetic selection driven by both natural and human interactions remains unknown. The availability of new and advanced genomic tools can change population and conservation genetic approaches to support the livestock sector (Taberlet *et al*., 2008; Joost *et al*., 2011).

Modern genomic tools can be applied to shed more light on the genetic background and origins of the South African pig population, to identify wild ancestral species, and infer on geographic zones of origin, evolutionary bottlenecks and expansions, hybridization or introgression between related species as well as investigate any geographical expansions and migrations. Based on the interconnectedness of the extant pig populations, it is important to find out if there is any gene flow occurring not only between populations but between diverse species. Pig breeds in Africa are poorly characterised at genetic level (Amills *et al*., 2012). Hence studies aimed at in-depth evaluation of genetic uniqueness and pig breed diversity will assist in developing a national and global plan for pig biodiversity and conservation programs.

1.2 Justification

According to FAO (2021), a decrease of pig consumption per capita/year from 5.72 in 2008 to 3.94 kg in 2013 was observed in sub-Saharan Africa. This comes with an estimated annual increase in pork consumption of 155% from 2000 to 2030 in the sub-Saharan region (FAO, 2011). There is a gap between pig production and estimated consumption of pork meat. Therefore, action needs to be taken to meet the future consumption levels. The declining number of pigs in South Africa, from 1.52 million in 2016 to 1.36 million in 2020 (Statista, 2022), is concerning. 25 percent of pigs are free ranging, with increased chance of contact between domestic and wild pigs. Anderson *et al*. (1998) reported interactions between domestic and Bush Pigs particularly in the communal areas. Warthogs have also been noted to have scattered further than their introductory sites (Nyafu *et al*., 2009). This existence of many South African pig populations including wild and domestic pigs that are interconnected brings opportunities for increased functional biodiversity as well as challenges of introgression-based dilution of genetic diversity.

There is therefore need for investigation of the genetic structure of the pig population of South Africa as well as gene flow and introgression mediated admixture in connected populations. There have been no studies conducted to date to investigate hybridization and introgression that might have an effect on the genetic make-up of the local population. In the process of evolution, hybridization and introgression are essential as they lead to the formation of unique genotypes and phenotypes (Bosse *et al*., 2014). This contributes to a significant source of genetic diversity and adaptation. On the other hand, according to Martinez *et al*. (2018), hunting, habitat loss and hybridization are a threat to diversity and conservation of pigs and other species.

The current genetic structure of pig breeds in South Africa is thought to be complex as pig populations are differentiated according to their regions of origin. Knowledge of the pig breeds in the country is crucial to respond to the agricultural, socio-economic and conservation challenges. Genetics (Chimonyo & Dzama, 2007) and diversity (Ramírez *et al*., 2009; Swart *et al*., 2010; Halimani *et al*., 2012) studies have been conducted in Southern Africa on pigs. However, these studies were also based on the less informative microsatellite markers and limited to diversity levels. The holistic representations of commercial, indigenous, villages, wild and Asian genotypes that we have in the country has not really been investigated.

The developments in high throughput sequencing technology have generated thousands of SNPs presenting opportunities for genome-wide association studies, genetic mapping, genomics-assisted selection and genetic diversity analysis (Duran *et al*., 2009). The availability of genetic resources has been a major limitation in the past when assessing the genomic architecture of pigs. SNPs are fast becoming the marker of choice for various applications in evolutionary studies, genomics, and conservation genetics as they have prospective for genotyping efficiency, cost-effective genotyping techniques, genome-wide coverage and data quality (Sachidanandam *et al*., 2001; Morin *et al*., 2004; Chen *et al*., 2007; Osei-Amponsah *et al*., 2017; Mujibi *et al*., 2018). The Porcine SNP60K bead chip was designed and first used by Ramos *et al*. (2009) and Groenen *et al*. (2012) in Duroc, Landrace, Pietrain and Large White breeds. In 2013, Burgoz-Paz *et al*. used genotypes generated from the Porcine SNP60K from 14 countries to show the ancestry of American pig populations. Osei-Amponsah *et al*. (2017) characterized the Ghananian Ashanti Dwarf pigs using the bead chip to understand the effects of admixture on economically importance traits. Mujibi *et al*. (2018) used the Porcine SNP60K to investigate the genetic diversity, and breed composition of village pigs. The study by Mujibi *et al.* (2018) additionally assessed genomic signatures of selection including samples from Warthogs, Bush Pigs and Wild Boars.

This study intends to provide in depth baseline data on the genomic architecture of South African pig resources inclusive of domestic (commercial, indigenous and non-descript village pigs), wild and exotic pigs kept in conservation units such as national parks and zoos. The analysis and data generated in this study will assist with a better understanding of the local pig genetic resource as well as facilitate future conservation and breed improvement strategies particularly for non-descript village pig populations.

1.3 Aims

To characterize the genomic architecture of South African pig populations using the Porcine SNP60K genome-wide SNP data.

1.4 Objectives

The specific objectives of this study were to:

- Determine genetic diversity of South African pigs relative to global populations of pigs consisting of villages and out-group pigs from South America, Europe, United States, and China amongst other regions.

- Assess the extent of genome-wide extent of LD and haplotype diversity (i) across several South African pig populations and (ii) between South African pig populations from different genetic and production backgrounds.

- Investigate the population histories, genetic migration, introgression and other genetic interactions amongst South African pigs using genome-wide SNP data.

- Identify and characterize genomic regions harbouring signatures of selection in South African commercial, villages, indigenous, wild and exotic pig populations using the SNP60K Porcine array.

1.5 Animal populations and SNP genotypes

This study used samples from various geographic locations. Populations sampled were a representation of non-descript village, commercial, indigenous, wild and Asian pig populations. The non-descript village pigs were sampled from the Eastern Cape and Limpopo provinces. In the Eastern Cape, hair samples were collected from pigs in various villages in Alfred Nzo (ALN) and O.R. Tambo (ORT) districts while hair and ear tissue samples were collected from Mopani (MOP) and Capricorn (CAP) pigs in Limpopo. Village pigs have no breed standards, are non-descript populations and are therefore termed as ecotypes based on their geographical locations. Production system varied from semi-intensive, backyard and free ranging in the villages. From Mockford commercial farm, which is located in Polokwane, twenty samples were collected from each commercial breed raised in the intensive production system. These were Large White (LWT), South African Landrace (SAL) and Duroc (DUR). Twenty ear tissues per breed was collected from the conserved indigenous Kolbroek (KOL) and Windsnyer (WIN) populations. These pigs are reared intensively at the Agricultural Research Council (ARC) - Animal Production Institute, which is based in Irene, Pretoria. Five samples from the

Vietnamese Potbelly (VIT) kept in conservation at the Johannesburg Zoo were collected as a representative of the Asian genetic pool in this study. A total of thirty-eight Warthog (WAT) hair samples were collected from the South African Kruger National Park biobank, Pilanesberg National Park and the Eastern Cape wilderness while three Bush Pigs (BSP) hair samples were donated by farmers around Magaliesburg area. From Western Cape, four Wild Boar (WBO) samples were collected from a private farmer who specialises with Wild Boar meat. In addition to our own samples, 389 genotypes deposited on the database from Burgoz-Paz *et al*. (2013) were also used. These were genotypes of pigs from 24 countries representing village, wild/feral, indigenous and commercial pigs. The presumed pig domestication centres of Europe and Asia were represented in this dataset.

1.6 Outline of dissertation structure

This thesis contains seven chapters, four of which are experimental chapters. Each experimental chapter has an abstract/summary, introduction, materials and methods, results, discussion, conclusion and references sections. The seven chapters are as follows:

Chapter 1 contains the general introduction to the thesis, study rationale, justification, aims and objectives.

Chapter 2 presents the literature review of the study. It outlines the domestication process of pigs and different pig breeds dominating the pig industry in South Africa. The production systems and genomic technologies and methods available for their characterisation are also discussed in this chapter.

Chapter 3 is the first experimental chapter that reports on the genetic diversity and structure of the pig populations found in South Africa and in relation to global pig populations.

Chapter 4 is the second experimental chapter that investigated and reported the extent of linkage disequilibrium, prevalence and distribution of haplotype blocks within and between the commercial, indigenous and village populations. In this study the Warthogs, Vietnamese Potbelly, Bush Pigs and Wild Boar samples were not used due to low numbers of polymorphic alleles for downstream analysis.

Chapter 5 investigated and reported the gene flow, shared ancestry and relationships of South African pig populations in relations to worldwide pig populations.

Chapter 6 is the last experimental chapter that identified potential genomic regions under selection and revealing genes linked to Quantitative Trait Locus (QTLs) (of economic importance using three different genomic signatures of selection methods.

Chapter 7 is the general discussion chapter that integrates findings from the 4 experimental chapters, draws conclusion and reflects on the findings of the study. Recommendations and future research prospects are also provided in this chapter.

In order to avoid repetition, all references were consolidated and put at the end of the thesis.

1.7 Ethical statement

Ethical approval for sample collection for this study was attained from the Agricultural Research Council-Irene Animal Ethics Committee (APIEC16/028). Permission was granted by the Department of Agriculture, Forestry and Fisheries to conduct investigation in terms of Section 20 of the Animal Diseases ACT of 1984 (ACT No. 35 of 1985). This was needed as domestic and wild pigs' samples were sampled and therefore precaution was needed to prevent the spread of African Swine Fever and Foot and Mouth Diseases (12/11/1/1).

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

Domestic pigs (*Sus scrofa domesticus*) that are part of the artiodactyl mammals (*i.e.*, hoofed animals) belong to the family Suidae (Frantz *et al.*, 2016). This family originated about 20 to 30 million years ago and consists of 17 classes that are clustered into five genera (Darwin, 1868). Domestic pigs are the result of separate domestication events that occurred in Europe and Asia, with Wild Boars (*Sus scrofa*) the most likely ancestors (Smith *et al.*, 2014). Previous studies showed that ancient introgression occurred between the European and Asian pig clades (Ramírez *et al.*, 2009; Amills *et al.*, 2010).

Modern domestic pig breeds were introduced and are farmed in many countries, including South Africa. According to Visser (2014), the South African commercial pig industry dates back to 1652 in the Cape of Good Hope. Small-scale pig production in previously disadvantaged communities is also commonly practiced in South Africa. In these communities, pigs help provide food security (Mashatise *et al.*, 2005; Mtileni *et al.*, 2006; Rischkowsky & Pilling, 2007). This is because pigs provide less expensive source of animal protein as well as a source of fat (Anti & Seahlodi, 2011; Dietze, 2011; Weinberg & Qaim, 2009). Pig farming in these communities also help with poverty alleviation, as it provides additional income, act as additional investments, as well as settling other obligations such as serving as dowry (Chimonyo *et al.*, 2005; Ajala *et al.*, 2007; Kamuribo *et al.*, 2011; Nissen *et al.*, 2011; Madzimure *et al.*, 2013).

The commercial pig industry mainly farms with exotic breeds (*e.g.*, Large White and Duroc) that were introduced into South Africa. Large White and Duroc are bred for intensive production systems (Chimonyo *et al.*, 2005; Halimani *et al.*, 2020). Commercial animals therefore represent useful models to study phenotypic traits that are under selection, which allow the identification of causative alleles that determine phenotypic differences (Wong *et al.*, 2004). Such information is essential for swift and precise genetic improvements of selected breeds for diverse environments, and moreover to facilitate rapid adaptation to possible adaptation in breeding goals (Notter 1999; Bijima *et al.*, 2002).

The commercial breeds are bred for intensive production systems and are not suitable for the extensive environment in South African (Chimonyo *et al*., 2005). In contrast, small-scale pig farmers mainly use South African indigenous breeds, such as the Windsnyer breed that were introduced to Southern African countries in 1600-1700s by the Chinese and European traders (Nicholas, 1999; Ramsey *et al*., 2000; Robinson & Penrith, 2009; Smith *et al*., 2014). These South African indigenous breeds are regarded as being hardy, since they are more tolerant to the harsh South African conditions and more resilient to inadequate resources and support (Lekule & Kyvasgaard, 2003, Mutua *et al*., 2010). However, previous studies found evidence of gene flow between South African indigenous populations along populations of other African countries, as well as with Eastern and European breeds (Mujibi *et al*., 2018).

Wild pigs such as Warthogs, Wild Boars and Bush Pigs are known to be potential carriers of African Swine Fever (Thomas & Kolbe, 1942). This poses a health risk for transmissions to domestic pigs via direct contacts, feeding on infected meat products or from shared water sources. Pentrith *et al*. (2013) associated the increase of ASF outbreaks with the increase in pig numbers the countries affected by ASF. In South Africa, outbreaks usually occur in places where pigs are kept in poorly fenced areas or allowed to roam freely and interact with Warthogs. Under experimental conditions undertaken by Anderson *et al*. (1998), Bush Pigs infection rates of ASF to domestic pigs was low compared to that of Warthogs, whereas transmission between domestic pigs is more efficient and poses a safety threat to humans because ASF virus is able to survive in pork meat at a wide range of temperatures and pH (Pentrith *et al*., 2019). The interbreeding of Wild Boars and Bush Pigs in KwaZulu-Natal is a concern as this can lead to genetic contamination (Venter *et al*., 2016).

The aim of this chapter is to provide a review on the South African pig industry, with a focus on the different production systems and breeds utilized by commercial and small-scale farmers. This chapter also reviews on the available genomic tools that can be used to study pigs for characterisation, improved production and conservation of existing breeds and populations. The review looks at availability of pig reference genome and other resources made available from genome sequencing initiatives that can offer understanding into the evolution, genetic complexity and variations of pig breeds and are necessary for understanding the origins, spread and management of pig breeds.

2.2 Domestication of pigs

Animal domestication, including that of pigs, has played a critical part in the expansion of human civilization by providing a more reliable food source, transportation, clothing and protection. Eurasia Wild Boar (*Sus scrofa*) is believed to be the origins of domestic pigs, with the deviation of the two main clades (namely the Asian and European clades) of domestication that is thought to have happened about 58 000 years ago (Kim *et al*., 2002; Larson *et al*., 2005).

Domestication of pigs resulted in genetic and phenotypic changes that occurred over generations that resulted in domestic pigs differing from their wild relatives (Price, 2002). Domestication of Wild Boar gave rise to intense phenotypic variations in domestic pigs which includes changes a wide range of phenotypic traits, such as behaviour, body composition, reproduction and coat colour (Rubin *et al*., 2012). Domestication together with a long history of migrations have generated a variety of breeds via selection and adaptation (Groenveld *et al*., 2010). These breeds also differ greatly in terms of reproduction abilities, production, colour, shape and size and overall, pig breeds display a massive quantity of phenotypic diversity (Groenveld *et al*., 2010). For any species, it is important to understand the domestication and breed formation events as they would have downstream implications on other genetic forces and parameters like increase in inbreeding, reduction in effective population size and fixation of certain genomic regions (Lindblad-Toh *et al*., 2005).

The history of African pig breeds remains generally unidentified and questionable as they are poorly studied and documented (Blench & MacDonald, 2000; Amills *et al*., 2012). According to literature, European and Asian populations had a major contribution to the current gene pool of pigs in Africa (Amills *et al*., 2012). Asian pigs were introduced into Africa, as the Chinese explored the East coast of Africa five centuries ago and had ancient commercial relationships with African countries (Levathes, 1994). However, it has also been observed that pig breeds occurring in Western Africa do not exhibit far Eastern genotypes, signifying that they are a product from the admixture of indigenous and exotic populations alongside the European ancestry (Ramírez *et al*., 2009). South African pigs are proclaimed to have reached Sub-Saharan Africa by chance through the Nile corridor and overtime spread to the West-Central Africa (Blench & MacDonald, 2000).

In South Africa, the Wild Boars, Warthogs and Bush Pigs are free-roaming and scavenging wild animals. Wild Boars live in scrub and forest areas. Warthogs can be found roaming in the game reserves as well as in villages across the country. According to Cumming (1999), Southern Africa has an estimated population of 250 000 Warthogs and about 5 000 reside in the Kruger National Park (Swanepoel *et al*., 2016). Warthogs are not in any threat of declining as have an annual growth rate of up to 45% (Swanepoel *et al*., 2016). The Warthogs practise a polyandrous mating system (Swanepoel *et al*., 2016), whereby females are able to mate with more than one male and males only mates with one female. Humans have been implicated in relocating the wild pigs and allowing them to move freely among agricultural lands which then increases contacts with domestic pigs. Wild pigs are popular for trophy and recreational hunting (Hoffman *et al*., 2017). The wild pigs are consumed as game meat. Although wild pigs are hardy, they are a reservoir for diseases, which is why they remain a concern for pig farming. The Warthogs resides in the burrows (Martinez *et al*., 2018) while Bush Pigs prefer a woodland, savannah and dense vegetation landscape (Jori & Bastos, 2009; Venter *et al*., 2016). Bush Pigs are mainly concentrated on the eastern areas of South Africa in KwaZulu-Natal, Eastern Cape and Western Cape provinces. However, there are recent recordings of Bush Pigs in the Free State province (Venter *et al*., 2016) as well as in Limpopo, North-West and Gauteng. It is estimated that there are 10 000 mature individuals of Bush Pigs in the Southern region. According to Venter *et al*. (2016), there is possible migration of Bush Pigs from Mozambique to Northern Limpopo and North-Eastern KwaZulu-Natal. Warthogs and Bush Pigs have natural immunity and show no signs to African Swine Fever (ASF) virus (Zhang *et al*., 2020; Feng *et al*., 2021). Domestic pigs and Wild Boars are on the other hand, are susceptible to acute diseases (Gabriel *et al*., 2011) and a previous study demonstrated disease transmission between domestic pigs and Wild Boars (Martinez *et al*., 2018). Fatality of ASF is 100% resulting in production and financial losses in the livestock industry (Feng *et al*., 2021).

2.3 The South African pig industry

There is an ever-increasing need for food production worldwide, with a 76% increase of animal protein demand projected by 2050. It was suggested that increasing the production represent an important bridge to overcome the shortfall (Alexandrotos *et al*., 2012; Happer & Wellesley, 2019). A huge role is played by the South African pig industry that dates back with the introduction of pigs to the Cape of Good Hope in 1652 by Jan Van Riebeeck (Visser, 2014). South African pig production is divided into commercial and smallholder communal sectors.

SAPPO (South African Pork Producers) is the commercial pig farmers' organization and serves the entire pork value chain. The emerging farmers' portfolio from SAPPO coordinates and manages emerging farmers' projects in the country. The main interests of SAPPO includes animal health, training, and development, liaison with stakeholders including government and abattoir owners, research, promotion of pork and pork products. SAPPO represents 700 producers from the formal sector whereas there are 208 312 households from the informal sector (SAPPO, 2022; BFAP, 2021).

The Pig Breeders Society (PBS) of South Africa formed in 1919 ensures compliance in terms of breed standards by the breeders. Performance records of the pedigrees of purebred boars and sows' registration is kept by PBS. PBS is affiliated to the South African Book and Livestock Improvement Association. Selection of sows and boars is based on the BLUP (Best Linear Unbiased Prediction) estimated breeding values (EBVs). There are 18 members under PBS. Standards are set for the Large White, South African Landrace, Chester White, Duroc, Hampshire, Large Black, Piétrain and Kolbroek. However, smallholder farmers mainly found in villages do not have access to modern production and lack genetic resources (Halimani *et al*., 2020). There are no formal selection breeding programs. Halimani *et al*. (2012) noted the willingness of smallholder farmers to participate in in situ conservation programs in order to improve pig genetic resources.

The pig industry contributed 34% of the gross agricultural production from 2018 - 2020. According to BFAP/SAPPO (2021), 8% of pork meat is consumed against 22% beef and 4% mutton. In the last decade, the South African pork industry has experienced an average growth of 9% per annum. South Africa roughly produces 250 000 tonnes of pork meat per annum (DALRRD, 2020). Only about 10% of pork is imported to the country. According to BFAP (2021), half of the small pork producers sell their meat to abattoirs and informal market. The primal cuts of pork meat sold by abattoirs goes to processors (27%), wholesalers (26%) while 26% goes to retailers, butcheries and the hospitality sector (BFAP/SAPPO, 2021).

Kriel (2020) reported the advancement of technology to have accelerated genetic improvement for South African pork producers. This has enabled improvement in genomic and breeding selection decisions. Quality purebred genetics from Large White, Landrace and Duroc are being exported to SADC and countries like Nigeria by reputable pig breeders in conjunction with

Alliance Genetics South Africa breeders' company (Taaibosch Piggery, 2021). TOPIGS SA and PIC (Pig Improvement Company) are other international companies supplying 73% of breeding stock to the South African market (Krüger *et al*., 2017). There is a demand for superior semen selected from top genetic pool of maternal and terminal lines. Visser *et al*. (2014) noticed the use of AI (Artificial Insemination) for mating in the pig industry to be around 75%. Canada and USA pig semen is allowed to be imported to SA under strict conditions (Krüger *et al*., 2017). The Department of Agriculture, Forestry and Fisheries (2014), halted pig semen imports to SA due to Porcine Epidemic Diarrhoea virus (PEDv) breakouts. South African producers are beginning to obtain 30 to 32 weaners/sows/year versus the average of 25 weaners/sows/year (Kriel, 2020, Reynders, 2021). Selection strategies has also contributed to improved growth rates, feed conversion ratio and a decrease in mortality rates (Kriel, 2020). According to Krüger *et al*. (2017), fertility traits (number born alive (NBA) and 21-day litter weight (21DLWT)), growth (average daily gain (ADG) and feed conversion ratio (FCR)) and production traits (back fat thickness (BF)) are the most important production traits for Large White, South African Landrace and Duroc. Feed is one of the biggest expenses in pig production, accounting for 80% of production costs (Magazines, 2012). It therefore important to feed the pig correctly as it has an impact on the growth rate. FCR has decreased from 3,8:1 to 3,1:1 in the last decade (Kriel, 2020). Inadequate quality and quantity of feed is critical constraint faced by smallholder farmers (Halimani *et al*., 2012).

The pig industry has not been without challenges. A devastating blow affected the industry by a listeria outbreak in 2018 on the market for the selling processed meat. Live pig imports have been banned to curb the devastation to the pig industry due to diseases and pathogens (Kriel, 2020). The COVID-19 lockdown period saw temporary closure of restaurants and fast-food retailers (BFAP/SAPPO, 2021). Nearly 40% of the pork industry in the country was affected and pork prices were pushed further down. Informal cross-border sales between South Africa and Mozambique were affected during lockdown (BFAP/SAPPO, 2021). The July 2021 unrest in South Africa led to loss of 6 000 tonnes of pork ribs due to theft in Durban (Pig Progress, 2021). Pig farms, slaughterhouses, semen deliveries were also not spared during the unrest. There was also a shortage of feed as the major highways were closed thus affecting 20 000 pigs who were at finishing stages (Pig Progress, 2021).

2.4 Pig production systems in South Africa

There is an ever-increasing need for food production worldwide, with a 76% increase of animal protein demand projected by 2050. It was suggested that increasing the production represent an important bridge to overcome the shortfall (Alexandrotos *et al*., 2012; Happer & Wellesley, 2019). South Africa's pig agricultural sector consists of different production systems, of intensive commercial, semi-intensive commercial as well as free-range/scavenging production systems.

Intensive production system is in the well-developed large-scale, requires high inputs and is usually associated with commercial breeds like the Large White, Duroc, South African Landrace, Pietrain and Hampshire (Hossain *et al*., 2011). In terms of the occupancy in the market, the Large White is the mostly used breed at 60% then Landrace (30%) and Duroc (5.5%) while more breeds (Kolbroek, Chester White and Pietrain) make up 4.5% in the market (Kirsten *et al*., 2009; BFAP, 2013). Commercial breeds are renowned for their growth potential although this great growth comes at an expense of high productions inputs (Hoffman *et al*., 2005). Commercial breeds outperform indigenous breeds in size, prolificacy, litter sizes and carcass confirmation (Visser, 2014; Munzhelele *et al*., 2017). Indigenous Windsnyer and Kolbroek breeds have been declining in the commercial sector because abattoirs and auctioneers prefer commercial breeds (Cupido, 2020). Commercial farms are estimated to be 4 000 in the country (DALRRD, 2021). From the 100 000 sows owned by the commercial farmers, the majority at 15% are found in the Western Cape (Meissner *et al*., 2013; Visser, 2014). Commercial farmers have 50 up to 250 sows on the unit (Mokoele *et al.*, 2014). According to Pig Breeders Society of South Africa Members (2022), South Africa has eighteen pig stud breeders for pig breeding stock. International companies, TOPIGS and PIC supplies 73% of the breeding stock (Kirsten *et al*., 2009; Buchnan & Stadler, 2011). An estimated 70-75% of mating is done via artificial insemination in the commercial sector (Kruger *et al*., 2017). In general, artificial insemination allows superior boar semen to be used for breeding, with disadvantages of reliance on a narrow genetic base (Whittemore, 2006a).

In the intensive production system, proper and adequate housing is provided to protect pigs from harsh weather conditions (Matabane *et al*., 2015). This also offers protection against diseases and biosecurity thus increasing production whilst minimising incidences of diseases (Madec *et al*., 2010). In terms of controlling diseases and treating individual animals, the

intensive system is more efficient compared to other production systems, as biosecurity measures are easier to apply. Feed contributes about 70% on the running expenses for pig production (Mokoele *et al.*, 2014). Pigs under intensive systems cannot fend for themselves and thus require a dietary intake of balanced diets to get all the required nutrients.

Free-range/scavenging production systems are mainly connected to the smallholder and subsistence farmers (Pienaar & Traub, 2015; Thamaga-Chita & Morojele, 2017). These farmers are important for food security in rural areas even though their pigs are considered as being unproductive and farming methods as backward (Kirsten & van Zyl, 1998). There are about 10 000 sows in this sector (DALRRD, 2020). Smallholder farmers stock up to 50 sows per unit (Mokoele *et al*., 2014). Limpopo and Eastern Cape provinces have the most smallholder farmers in the country (Madzimure *et al*., 2014). Most smallholder and subsistence farmers are dominated by men in Gauteng, Mpumalanga, KwaZulu Natal and Mpumalanga whereas in the Eastern Cape, women are dominant in pig farming (Madzimure *et al*., 2013; Mokoele *et al.*, 2014; Matabane *et al*., 2015; Gcumisa *et al*., 2016; Munzhelele *et al*., 2017). The involvement of women in the farming sector must be encouraged as it can contribute to gender redress and improved livelihoods of the marginalised.

Pig production plays various important roles for smallholder and subsistence farmers. According to Mokoele *et al.* (2014), farmers in Limpopo use pig farming as substitute to investing for the future. Pig farming also provides animal protein and income used to support families (Gcumisa *et al*., 2016; Meissner *et al*., 2013). Pigs also produce manure, can be exchanged as dowry (Meissner *et al*., 2013) and are offered as gifts (Madec *et al*., 2010). Mokoele *et al*. (2015) and Matabane *et al*. (2015) recommended formal training for pig farmers in order to alleviate poverty and increase productivity.

Table 2.1 summarizes the system of pig production in South Africa in terms of housing, ownership, feeding and breeding. Matabane *et al*. (2015) described semi-intensive housing to consist of fences and roof while the farmer has no control of the temperature in the housing structures. Pigs are free to move anyhow with no housing in free-range/scavenging systems, while increases the threat to diseases like the African Swine Fever (ASF) and Classical Swine Fever (CSF) to domestic pigs (Cupido, 2020). Surprisingly, pigs kept under extensive production are free from the majority of infectious diseases like respiratory syndrome, porcine

reproductive and porcine pleuropneumonia, enzootic pneumonia, swine dysentery, porcine proliferative enteropathy that are common with pigs kept in confinement (Saade *et al*., 2020). Contacts between wild pigs and domestic pigs is also observed under extensive production systems.

Table 2.1 Systems of pig production

Production system	Characteristics			
	Housing	**Ownership**	**Feeding**	**Breeding**
Free-range/scavenging	None	Often communal	None	Uncontrolled
Semi-intensive	Semi-permanent construction from local materials	Individual smallholders	Household waste	Uncontrolled or use of local boars
Intensive	Controlled housing	Urban-based entrepreneurs and businessman	Agro-industrial by-products	Only selected boars used for stud

2.5 South African pig breeds

Out of the 541 pig breeds globally (Rischkowsky & Pilling, 2007), only six dominate the commercial pork sector (Weka *et al*., 2021). These are the Piétrain, Landrace, Large White, Hampshire, Duroc and Berkshire and (Rothschild & Ruvinsky, 1998). It is presumed that the European settlers of the 1600s brought the founding breeds of the developed commercial pigs found in Southern Africa (Krige, 1950; Blench & MacDonald, 2000; Swart *et al*., 2010) starting with Jan van Riebeeck who brought some pigs to the Cape of Good Hope around this time (Naude & Visser, 1994).

2.5.1 Commercial breeds

The Duroc, Large White and South African Landrace are established breeds used in the commercial industry (Kem, 1993; Ramsey *et al*., 2000). Landrace and Large White breeds are some of the pig breeds that are fairly distributed in many countries worldwide (DADIS, 2006) because of their outstanding reputation in the production of white meat, bacon fat and preserved products and are found to be a mix of European-Asian origin (Groenveld *et al*., 2009).

2.5.1.1 Large White

The Large White pig breed was first developed in the late 1700s in Yorkshire, England and is one of the commonly utilised breeds in commercial pig production (Taylor & Roese, 2005). Large White pigs can be identified through its short snout with erect ears. Although it has pink skin and white hair like the Landrace, it is a very strong breed (Kemm, 1993). It is thought to be one of the popular breeds as it has shown to have better performance when compared to other breeds. Large Whites were first recognized as a distinct breed in 1884 (Briggs, 1983). During 1958 and 1959, there was a big change in pig breeding in South Africa when Landrace pigs were first introduced (S.A Stud Book Association, 1971). Large White pigs are known for their fast growth rate, desirable feed to bodyweight conversion ratio and optimal lean meat percentage and as well as substantial milk production, large litters and exceptionally mothering characteristics (Bosse *et al.*, 2014). The breed is believed to be a cross with certain traits possibly the outcome of selected Chinese haplotypes (Bosse *et al.*, 2014).

Large White pigs are a large-framed breed with long middle and light shoulders. Their legs tend to be longer than other breeds and are associated with poor ham development. The head is moderately long. Its coat and skin colour make it vulnerable to sunburn. Large White sows are well-known for their large litter size, high milk yield and good mothering instincts (Taylor & Roese, 2005). The current breeding objectives are driven by meat consumer's demand for small amounts of fat and high quantity of lean meat and good quality of bacon and pork (Taylor & Roese, 2005). The males can reach up to 155 kg while the females reach about 117 kg (Taylor & Roese, 2005). The Large White are associated with high milk production, large litters, lean carcasses and fast growth rates. The main shortcoming is that the breed is more suited to intensive productions systems with commercial feed rations, and do not do well in other production systems.

2.5.1.2 South African Landrace

The Landrace breed rate as the second most important breed in the country with its small black spots that are intermittently found on the back of some pigs. It was developed in Denmark by crossing a Large White with a Danish native pig and was improved over the years and exported to various countries in 1949 (Taylor & Roese, 2005; Treacy, 1976). The Landrace breed is widely used throughout the world because of good mothering ability and for being docile.

Twenty-four, Danish Landrace were exported to the United States of America in the early 1930s while four boars and eight gilts were exported to England in 1949. In South Africa, today the Landrace breed is the second-best defined breed (Visser *et al*., 1993). As one of the preferred breeds, it is utilised for production of pork and bacon equally. Even though the Landrace has similar production rates to those of the Large White, it is longer, has a larger meat percentage and tend to produce PSE (pale, soft flesh and exudative) carcass. PSE leads to pale meat and is associated with low quality pork caused by stress and rapid decline in pH in pigs resulting in economic losses (Trevison & Brum, 2020).

Landraces have white skin with floppy ears, long snout, fine hair and slightly dished face (Smith *et al*., 2014). It also has a desirable character of giving birth to large litters, high growth rates, high milk production and high proportion of lean carcass weight (Hoffman *et al*., 2005b). As a commercial breed, it performs well under intensive production system. The Landrace has disadvantages of health difficulties such as splay legs, nervous disorders like porcine stress syndrome and leg weakness (Swart *at al.*, 2010). The majority of breeds today are known to contain Landrace and Large White genes (Taylor & Roese, 2005). The Landrace is also popular among non-commercial farmers who slaughter mainly for domestic consumption.

2.5.1.3 Duroc

There are different theories about the development of the Duroc breed. The Duroc breed is assumed to have been brought by Christopher Columbus to America on his second voyage with Hernando the Spanish explorer also introducing them to South Africa in the 1500's (Dohner, 1951; Martínez *et al*., 2012). Another theory is that the Duroc is one of several pig strains to have been developed around 1800's in New England while it is also presumed to have been imported from the Guinea coast of Africa at the time of slave trade (Dohner, 1951). The foundation that forms today's Duroc is thought to have come in early 1812 from Eastern America in New York and New Jersey.

The Duroc is a large framed, late maturing type, medium length with dropping ears and slight dish of the face (Martínez *et al*., 2012). Farmers who like to keep their pigs outdoors would choose the Duroc, as it can withstand various environments. Durocs are common in crossbreeding programs, as they are known for producing large litters and improving other

breeds (Esteve *et al*., 2012). The skin has a reddish-brown colour though wide variations may occur (Smith *et al*., 2014). It has high ratio of intramuscular marbling to carcass fat, higher pH and above average feed conversion and daily gains. It is known to have a rich mouth-watering flavour with superior tenderness and juiciness because of its bright reddish pink colour that is highly desirable (Hoffman *et al*., 2005b).

2.5.2 Indigenous breeds

Indigenous breeds are resilient, able to live on very little and have greater immunity (Munzhelele *et al*., 2017). Mortality is higher, and growth is slower in the indigenous populations (Madec *et al*., 2010; Munzhelele *et al*., 2017) and mostly attributed to the management systems. The Kolbroek and Windsnyer are considered as indigenous and are common in the villages (Kem, 1993; Ramsey *et al*., 2000). In the rural areas, indigenous breeds such as Kolbroek and Windsnyer are a source of food and income (Madzimure *et al*., 2013).

Halimani *et al.* (2012) found that farmers were prepared to farm with indigenous pigs and conserve them as they are genetically resistance to diseases, can survive and produce on low inputs costs and have good adaptability. In rural areas, there are no formal breeding programs and therefore sows and boars are housed or roam freely together. Indigenous pigs still thrive on low quality feed, and this is helpful to rural farmers who have little resources (Madec *et al*., 2020). Farmers in the rural areas cannot afford commercial feeds and thus very few are supplemented with commercial feed or concentrates. The majority of the pigs are fed swill, grains and/or kitchen leftovers (Ocampo *et al*., 2005; Nsoso *et al.*, 2006; Ajala *et al*., 2007; Kumaresan *et al*., 2009; Mtileni *et al*., 2009; Kagira *et al*., 2010).

2.5.2.1 Kolbroek

Kolbroek is generally accepted as an indigenous breed although there are several theories about its origin (Ramsay, 1998). It is thought that when a shipwrecked on the Cape coast in 1778, the Kolbrook was on board (Swart *et al*., 2010). On the other hand, the theory is that the Portuguese introduced this breed in the 15th century, and it appeared to be of Chinese origin (Swart *et al*., 2010). Although it is smaller in size equated to the Landrace and Large White breeds, it is ideally suitable for free-range and smallholder systems. Its lighter spots and patterns with a belly that almost touches the ground are typical morphological attributes used to best identify

it (Swart *et al*., 2010). Mason and Maule (1960) labelled the Kolbroek as a short, fat, short snout pig, resembling the Chinese lard pig, whilst referring to the Windsnyer (Wind-Cutter) as long-nosed, razor-back pig. Local breeds such as the Kolbroek and Windsnyer are marginalized in benefit of high-output breeds and therefore their genetic variation might be threatened. Herrero-Medrano *et al*. (2014) considered the Kolbroek as a source of biodiversity that represented a genomic stock that will possibly be essential for future adaptation needs.

2.5.2.2 Windsnyer

Windsnyer indigenous pigs are known to be dominant in Zimbabwe and parts of Mozambique and Zambia (Holness, 1973, 1991). The genotypes of the Windsnyer breed are distributed in Northern Zimbabwe, Mozambique and Eastern part of South Africa (Halimani *et al*., 2010; Bovula, 2017). Windsnyer pigs are renowned for their long nose, narrow back and razorback (Smith *et al*., 2014). They are a variety of reddish-brown, black, black and white or speckled. Their young have stripes like the Bush Pigs. These pigs have larger heads and longer legs with excellent mothering qualities. Mature females weigh between 40-120 kg (Bovula *et al*. 2017). As indigenous breeds, they are hardy, resistance to diseases and parasites and heat tolerance (Ramsay *et al*., 1994; Chimonyo *et al*., 2005). The Windsnyer pigs survive well outside and scavenge for food with little management, which makes them suitable for resource poor rural farmers (Bovula, 2017). This pig breed can utilise poor quality food and is associated higher back fat thickness.

2.5.3 Wild pigs

2.5.3.1 European Wild Boar

The European Wild Boar is considered indigenous to the Northern hemisphere and has spread to several parts of the world, including South Africa (Choi *et al*., 2014; Khalilzadeh *et al*., 2016). Its physical characteristics include large head, long legs and long snout, narrow and straight with small erect ears (Mapston, 2007). This breed is dark grey to black or brown and even whitish with thick and course double coat of fur. It also has excellent maternal skills, large litter size and heavy milk production (Tack, 2018). Boars can weigh up to 270 kg and sows up to 120 kg (Tack, 2018). They are known to create an immense amount of damage and are a

huge problem to farmers. They are also carriers of ASF like other wild pigs (Thomas & Kolbe, 1942). However, some farmers breed them and sell their meat to restaurants.

The Wild Boar is the descendant of the domestic pig thus the two-share strong genetic similarities and are able to hybridize (Scandura *et al*., 2011). Therefore, genetic introgression of the domestic pig could bring new genetic variants into the Wild Boar and possess danger to its genetic integrity. Albarella *et al*. (2007) noted that after the transformation of pigs from the wild to entirely tamed domestic animals, the interactions between human and pigs grew into much solid, nevertheless more diverse and multifaceted. Bosse *et al*. (2012) explained the latest admixture of Asian and European breeds that happened at the time of the industrial revolution when Asian pigs were introduced in Europe to improve the local pigs. Burgos-Paz *et al*. (2013) also found a direct link to the Asian introgression of pig in the modern villages of America, as a result of several independent events. The introgression of Asian pigs into European populations was anticipated to have improved the overall genetic variation in European pigs in contrast to their wild equivalent. Higher variation and fewer IBD tracts could have been introduced together with distinct haplotypes through hybridization (Bosso *et al*., 2012). Furthermore, higher nucleotide diversity was expected and observed (Giuffra *et al*., 2000; SanCristobal *et al*., 2006; Megens *et al*., 2008).

2.5.3.2 Warthog and Bush Pig

The Warthogs are wild pigs found throughout sub-Saharan Africa in open savannah, grasslands areas around water holes and marshy zones (Feng *et al*., 2021). Its name comes from wart-like protrusions on the face with four for boars and two for sows (Rognes, 2011). They are black or brown in colour with curls running to the bottom of the spine to central of the back. Warthogs have large, flat head with four sharp tusks (Feng *et al*., 2021). The coat is sparse and has no subcutaneous fat with a long tail with a tuft of hair. They use burrows each night for protection against predators, for thermoregulation and giving birth (Rognes, 2011). The boar plays no part in rearing piglets and leaves the sounder indefinitely at 15 months but is only able to attract a sow at 4 years. The reproductive rates of Warthogs is high at 3,4 hoglets each year (Somers & Penzhorn, 1992).

Bush Pigs are found in the forests, woodlands, long grass and reed beds of East and Southern Africa. (Botha, 1989; Jori & Bastos, 2009). Bush pigs are similar to domestic pigs and are recognized by a blunt muscular nose, sharp ears and small eyes. Their shades range from reddish brown to dark brown and gets darker as they age with this providing camouflage. Its face and ears are slightly in colour and have coarser hair and are larger in size. The head of the Bush Pigs is long with hard flattened snout. Bush Pigs are a nuisance to farmers and hunted extensively. Its meat is considered leaner than the traditional pork. Bush Pigs run with their tails down while the Warthog runs with the tail held vertically. They wallow in the mud to regulate temperature and protection themselves against insects and are mainly active at night. Boars assist with the rearing of the piglets. There is greater chance of contact between Bush Pigs and domestic pigs than of Warthogs and domestic pigs in the communal areas (Venter *et al*., 2016). This can increase chances of CSF and ASF outbreaks as previously seen in the Eastern and Western Cape provinces (NAFU, 2007; Penrith *et al*., 2011). The Department of Agriculture in South Africa has already defined zones of ASF in Mpumalanga, KwaZulu-Natal, Eastern Cape and North-West provinces. Blomstrom *et al*. (2012) considered Bush Pigs to be carriers of porcine parvovirus while wild pigs in general from time to time are infected by *trichinelloasi*s, bovine tuberculosis and/or foot and mouth diseases (Jori *et al*., 2016).

2.5.3.3 Captured and conserved populations: The Vietnamese Potbelly

The original Vietnamese Potbelly is short with wrinkled face, black in colour with the prominent potbelly (Blaney, 2004). They are sometimes black with white spots in colour or grey because of being crossbred (Simianer & Kohn, 2010.). At some point, they were popular as pets in North America (imported mid 1980s), Europe and many regions in the world as they are considered friendly but there has recently been increasing incidents of abandonment (Delibes & Delibes, 2013). Vietnamese Potbelly can grow to between 113 to 227 kg, which makes them unsustainable to keep as pets. This has led to unknown free roaming Vietnamese Potbelly populations living in the wild. Delibes and Delibes (2013) believe they are interbreeding with the Wild Boar populations in Spain. Concern is growing on the genetic erosion of the Spanish Wild Boar breeds as a result of the Vietnamese Potbelly invasion. These breeds are highly adaptive and can survive the climatical conditions of Europe.

Vietnamese Potbelly are known to have high fertility but tend to reach maturity early (Blaney, 2004). With an average lifespan of between 12-18 years, Vietnamese Potbelly have been

recorded to live up to 21 years in conserved sites (Corapi *et al*., 2011). The Vietnamese Potbelly has been used to develop breeds such as the Göttingen Minipig from Germany from a population kept at Wilhelma Zoo in Stuttgart, Germany (Simianer & Kohn, 2010). There is no clear history as when the Vietnamese Potbelly was first imported to South Africa, but a small population is conserved at the Johannesburg Zoo. As these are a small population size, this gives rise to genetic drift and inbreeding.

2.5.4 Village populations

Pigs in the villages are a source of income and meat for home consumption (Gcumisa *et al*., 2016). Even though they represent a valuable source of genetic pool (Subalina *et al*., 2010), there are no defined breeding programs for village pig farmers. Piglets in the villages mostly die before weaning as they are being squashed by sows and exposed to unfavourable weather conditions (Ajala *et al*., 2007; Madzimure *et al*., 2017). These pigs represent unknown source of lineages. There is no selection taking place in village pigs. These populations are also heterogenous genetic pool. Parameters such as growth patterns, phenotypic variations and reproduction traits have not been evaluated in South African village pigs. Undertaking research inclusive of village pigs is important as it can contribute to a better understanding of village pigs, their management and genetic resource strategies. This is despite thriving under harsh environmental conditions. Village populations differ from the studied commercial and indigenous breeds which are already standardised according to the specific breeds. Therefore, pigs found in villages classified ecotypes based on geographic locations. This study was interested in diverse geographical regions hence pigs were sampled from villages in Limpopo and Eastern Cape provinces. Therefore, in this thesis, each village population will be referred to by their geographical context.

2.6 Crossbreeding commercial and indigenous breeds

Due to the low production performance, some farmers in rural areas abandon the indigenous pigs in favour of exotic and commercial breeds (Halimani *et al*., 2012; Lekule & Kyvasgaard, 2003). Provinces such as KwaZulu-Natal (Gcumisa *et al*., 2016), Limpopo (Mokoele *et al*., 2014) and Mpumalanga (Muzhelele, 2015) have seen a move from indigenous to commercial breeds for smallholder farming. Smallholder farmers also favour Large White, Durocs or their crosses (Madec *et al*., 2010; Munzhelele *et al*., 2017). Nevertheless, KwaZulu-Natal, Eastern

Cape and Limpopo still keep indigenous populations (Madzimure *et al*., 2012; Gcumisa *et al*., 2016; Mokoele *et al*., 2014).

Other farmers in rural areas have started crossbreeding indigenous breeds with exotic breeds (Halimani *et al*., 2012; Lekule & Kyvasgaard, 2003). For example, farmers have crossed indigenous populations such as Kolbroek alongside commercial breeds such as Large White and Duroc to combine hardiness and production benefits (Visser, 2014). These crosses have also led to better carcass quality, feed conversion ratio, growth rates and better grading results (Visser, 2014). However, crossbreeding requires planning and well-maintained record keeping being able to preserve heterosis. Heterosis tend to decline overtime as a result of backcrossing with parental breeds (Sorensen *et al*., 2008). Crossbreeding can also induce outbreeding depression which might prompt a decrease in production and reproduction traits (Kirkpatrick, 2017.). Genetic erosion can be caused by indiscriminate crossbreeding (Berthouly-Salazar *et al*., 2012). This leads to the disappearance of local genetic diversity and therefore increasing chances of extinction.

2.7 Introgression and hybridization of pigs

Harrison (1990, 1993) defined natural hybridization as the interbreeding of animals from two or more divergent populations resulting individuals distinguishable from the original breeds on the basis of one or more of the inherited features. Although occurring at a very low frequency, hybridization with domestic pigs under ordinary environments is a regular habit through the freeing of confined pigs and it signifies a serious threat for populations (Scandura *et al.*, 2011). Levy *et al*. (1996) noted many studies to have verified the effect of genetic diminishing by introgression/hybridization thus resulting in extinction. Diminishing in fitness, endurance, population capability, including interruption of indigenous adaptations and life-historical attributes through the establishment of abnormal gene complexes are some of the effects of introgression (Rhymer & Simberloff; 1996; Allendorf *et al*., 2001; McGinnity *et al*., 2003). Anderson and Hubricht (1938) together with Anderson (1949) described introgression as the incorporation of genes out of one species to the genetic pool of subsequent divergent species. The ecological background whereupon hybridization happens, the distant that resultant individuals disperse, and the nature of natural selection are the factors that will influence rate and effects of introgression (Harrison & Larson, 2014).

According to Frankham and Ralls (1998), genetic introgression via a domestic or hybrid origin results in loss of genetic adaptation and possibly outbreeding depression (Iacolina *et al*., 2018). However possible advantageous effects of introgression including heterosis, greater litter size and fast growth rates (Goedbloed *et al*., 2013). Though relevant for conservation and disease risk management applications, the extent of introgression of pigs from Europe, Far East and Wild Boar is unknown. Genome-wide comparisons can be used to investigate developmental evolution or divergence events and confirm hybridization between divergent ancestries (Patterson *et al*., 2006; Nadeau *et al*., 2013; Prüfer *et al*., 2014).

2.8 Genetic studies of South African pig breeds

The study of the genetic variety and population structure in South African indigenous and commercial pig breeds are relevant from social or historical perspective, as well as for pig breeding and improvement. High resolution and inexpensive tools to assess genetic population structures and factors influencing them have been provided by molecular techniques particularly of SNPs (Ramos *et al*., 2009). The accessibility of the entire genomic sequence data has increased the tools currently available for population and genetic studies for pigs and other livestock species.

A good reference genome is important for more comprehensive analysis of within and between species genetic diversity. Inbreeding has long been acknowledged as a main reason of decline in fitness in both wild and domesticated populations (Bosse *et al*., 2012). As a result of inbreeding, Herrero-Medrano *et al*. (2014) noted that due to introgression, coupled with recent inbreeding, English pigs displayed ROH associated with detrimental mutations. Consanguineous mating causes the inherited of haplotypes identical by descent (IBD), causing extensive homozygous stretches through the genome of the progenies (Bosse *et al*., 2012). An understanding of the genomic distributions of IBD alleles will assist in unveiling the effects of inbreeding, introgression and other forces of evolution on populations. The pig genome has relatively heterogenous recombination and could be used as a model to study demographic influences on genetic variation (Bosse *et al*., 2012).

2.9 Genomic platform tools and techniques

Genetic diversity presents opportunities for scientists to provide solutions of meeting the growing demand for food. Useful tools such as microsatellites or Single Sequence Repeats (SSR) has been extensively used in genetic studies in South Africa for cattle (van der Westhuisen *et al*., 2020), goats (Visser *et al*., 2011), chicken (Mtileni *et al*., 2011) and sheep (Qwabe *et al*., 2013). The abundant robust variation of SSR have facilitated their use in population genetics, phylogenetic inference and genome mapping for over 20 years (FAO, 2011). However, technology has evolved and in recent years SNPs have replaced SSRs. High density arrays made it possible to analyse economically important traits which was restricted by insufficient number of genetic markers (Ramos *et al.*, 2009). SNPs sequenced by the Swine Genome Sequencing Consortium (Schook *et al*., 2005) were used to build a porcine reference genome. In the end, 549K SNPs constituted the panel that was subsampled, optimised and validated into the 64 232 Illumina Porcine SNP60K chip. The success of the development of the high throughput bead chip has resulted in over 260 genome-wide studies identifying 4 143 meat and carcass QTLs over the years. Over the years, genotyping using the Porcine SNP60K has become accessible and enabled mapping of phenotypic effects such as fatness (Ponsuksili *et al*., 2011) and reproductive (Uimari *et al*., 2011) traits. GeneSeek GGP Porcine HD Genomic Profiler *version 1* containing 68 516 SNPs, an improved version of the Illumina Porcine SNP60K chip has also been used to genotype European domestic and Wild Boars by Munoz *et al*. (2019.) This array has 42 372 common SNPs with the Illumina Porcine SNP60K. The progress and decreasing costs of ultra-high throughput next generation sequencing (NGS) is resulting in increased use of the technology.

2.10 Factors influencing genetic diversity in pigs

2.10.1 Natural and artificial selection pressures

The domestication (artificial selection) process 10 000 years ago happened as result of human interference (Larson *et al*, 2010). Following domestication of Wild Boars (*Sus scrofa*), there have been some phenotypic changes in domestic pigs for a various trait, including behaviour, growth, coat colour and reproduction. (Rubin *et al*., 2012) suggesting that artificial selection was the principal driver that shaped resultant domesticated pig breeds. Bosse *et al*. (2015) noted early pig farmers in Europe importing pigs from Asia to improve traits for commercial interests. The improvement of pigs can be done by both artificial selection and natural selection. Semen

from genetically superior boars can be used to improve traits of interests via AI whereas though natural selection, beneficial and advantageous traits are inherited over generations (Hill, 2004).

Both artificial and natural selection have played a role shaping modern pigs. Natural selection contributes to adaptation of species to their environment. Amaral *et al*. (2011) and Rubin *et al*. (2010) have noted the coat colour and size of pigs have been as a result of artificial selection during domestication whereas disease resistance in Wild Boars is influenced by survival and fitness via natural selection. Windsnyer, Kolbroek and the indigenous Mangalitza fatty-type pigs from Hungary all have curly hair phenotype beneficial for thermal insulation as effects of natural selection (Schachler *et al*., 2020). In South Africa, commercial breeds such as Large White, Duroc and S.A. Landrace are artificially selected for frugally important attributes like carcass traits and growth (Visser, 2014). Village and indigenous pigs such as Kolbroek and Windsnyer are predominantly under natural selection that leads to characteristics favourable to adaption to local environments such as harsh climate and hardiness (Madzimure *et al*., 2013). Natural and/ or artificial selection both leave trails across the genome known as signatures of selection that defines breeds (Bertolini *et al*., 2018). Both artificial and natural selection can also lead to lower diversity at selected genomic regions leading to fragmented populations (Amills *et al*., 2010).

Currently, information on genetic variation and its association with phenotypes in South African pig populations are limited. However, the study by Osei-Amponsah *et al*. (2017) of the indigenous Ashanti Dwarf revealed signatures influenced by European and Asian genetics. The black coat colour and patterns displayed by the Ashanti Dwarf is as a result of the *MC1R* gene. The mutation on this gene was found to be carried by haplotypes from Large Black (European) and Black Meishan (Asian) pigs. Using the homozygosity based *iHS*, Mujibi *et al*. (2018) noted chromosomes 1, 2, 3, 7, 9, 14, 15 and 18 to be regions under selection in Busia populations while chromosomes 1, 6, 9 and 12 were observed in Homobay populations. Genes such as *AMHR2*, *NPFF* and *SP1* on chromosome 5 were shared by Busia, Homobay and Wild pig populations. As a follow up to extensive research done on growth including meat and carcass quality of pigs in indigenous, commercial and Warthogs pigs in South Africa by Hoffman *et al*. (2005, 2007); Dube *et al.* (2011) and Kanengoni *et al*. (2014) amongst others, this research hope to find genes and QTLs linked with these traits.

2.10.2 Migration (Gene flow)

Processes such as migration, genetic drift, selection and mutation have an effect on the genetic structure of populations. Movement of genes in and out of the population, known as migration or gene flow, may be a driving force to evolution of populations (Gao *et al.*, 2020; Andrews *et al.*, 2010). A review by Whitlock and McCauley (1999) observed that movement of genes associated with adaptation and adaptive traits is determined by the rate of gene flow. Alleles within a population have a high probability to be identical by descent (IBD) if populations are geographically isolated overtime with no gene flow from other populations (Ciofi *et al.*, 1999; Clegg & Phillimore, 2010). Gene flow contribute to rates of phenotypic divergence and genetic variation overtime. Gene flow can either stimulate genetic diversity through new alleles introduced from other population or cause genetic similarity between populations as populations will end up acting as one continuous genetic population (Clegg & Phillimore, 2010). According to Ciofi *et al.* (1999), movement through migration has an effect on gene flow even though migration is not always a guarantee to gene flow. This is because of the need to exchange genes between migrating individuals and local populations. Human interference when introducing unrelated individuals into populations, is one type of gene flow that can assist in increasing genetic diversity. However, Zhan (2016) and Gao *et al.* (2020) also noticed that gene flow can also contribute to the decreasing genetic differentiation.

Artificial insemination dominates in the South African intensive production system as a source for superior genetics for profitability (Kruger *et al.*, 2017). This means there is continuous gene flow from European genetics into the South African population. Artificial insemination program for smallholder farmers was initiated in Gauteng in order to provide farmers with superior genetics (Thivhilaheli *et al.*, 2020). Conserved populations found at research stations, Zoos and National parks are fragmented with restricted gene flow. This contributes to small population size, disappearance of genetic variation, widening genetic drift and inbreeding (Andrews *et al.*, 2010). Village populations are free roaming which means there is gene flow between them and any other breed or species that they interact with. Increase in gene flow may lead to increased genetic diversity in populations and erosion of genetic differentiation.

2.10.3 Geographical isolation and fragmentation

Domestication of pigs has had an effect on the genetic diversity within and between populations. According to Marchelli and Gallo (2001), geographic and environmental dynamics

further influences the genetic diversity patterns. Selection pressures causes populations to drift or diverge apart. Keyghobadi (2007) described fragmentation as breaking of a population into smaller isolated patches. Thus, genetic processes are more likely to influence population fragmentation and isolation as a result of high genetic differentiation and selection techniques (Provan & Maggs, 2016; March *et al*., 2016). This fragmentation of populations can lead to a population separating and becoming homogenous and ultimately an ecotype. The negative effect on fragmented populations is the reduction of population sizes that increases the potential of genetic drift, inbreeding and chances of extinction (Ellstrand & Elam, 1993; Tóth *et al*., 2019). In South Africa, Swart *et al*. (2010), found the Kolbroek to be of a narrow genetic base caused by the small effective population size as these populations are kept at a research station as nucleus herds where they are reared under intensive production system. Therefore, conservation of such genetically distinct populations is important. The Vietnamese Potbelly population also has no gene flow as it is kept isolated at the Johannesburg Zoo. For conservation purposes, population size is important as small populations are at risk of going extinct. Genetic erosion and loss of genetic variation remains a challenge in implementing conservation and development strategies thus population size needs regular monitoring to avoid fragmentation of populations.

2.11 Population genomic parameters

2.11.1 Linkage Disequilibrium

Linkage disequilibrium (LD) measures the degree of non-random linkage of alleles at different points in a population and could be employed to assess mutation, selection, breeding history and genetic drift, and to depict causal mutations and QTLs in populations (Pritchard & Przeworski, 2001). The reduction in genotyping costs have created a platform to gather genotypic data with higher marker density genome-wide and allow assessment of whole genome LD profiles of populations. In pigs, the Porcine SNP60K and GeneSeek GGP Porcine HD have been used to investigate LD in both commercial and indigenous type breeds (Amaral *et al*., 2008; Badke *et al*., 2012; Ai *et al*., 2013; Veroneze *et al*., 2013; Munoz *et al*., 2019). An average r^2 of above 0.3 was observed by Veroneze *et al*. (2013) for Duroc and Landrace breeds whereas Badke *et al*. (2012) detected an average r^2 of 0.36 for Landrace and the highest for Duroc at 0.46. In pigs, LD extends for hundreds of kilobases, and studies have shown that

30,000 - 50,000 SNPs would be necessary for the whole genome association analysis (Du *et al*., 2007; McKay *et al*., 2007; Solberg *et al*., 2008).

One of the simplest LD measures is D_{ij} (Lewontin, 1960) which has the disadvantage of being dependent on allele frequencies and having no direct interpretation for the downstream application (Slatkin, 2008). D' and r^2 are alternative measures of LD with values ranging from 0 (no LD) to 1(high LD) (Slatkin, 2008). A $D' = 1$ means that the least allele per loci is entirely linked with another allele at a different locus. D' of 1 is often maintained at 1 until a recombinant or mutation event disruptions the original haplotype (Slatkin, 2008). Nonetheless, $D'<1$ does not have a logical explanation thereby the D' value from many studies is limited (Du *et al*., 2007). Hence, it becomes difficult to compare across diverse marker pairs and studies. Another LD estimate, r^2 that measures the correlation between two alleles at two loci (Ardlie *et al*., 2002) is more preferred as it is less affected by allele frequencies (Slatkin, 2008).

The consistency of selection over time, its duration, intensity and direction have an effect of on LD. Bulmer (1971) revealed that selection diminished genetic diversity for subsequent generations while generating undesirable LD among loci. Previous studies observed variation of LD amongst populations and genetic regions (Ardlie *et al*., 2002). In the Northern European pig populations, the extent of LD varied from 10-30 kb to hundreds of kilobases whereas for the Northern African populations the extent of LD is considerably lower (Ardlie *et al*., 2002). Studying LD allows for the assessment of effective population size of populations especially when pedigree information is unavailable (Hayes *et al*., 2003). Karimi *et al*. (2020) found effective population size to be 116 at 5 generation ago by using LD to estimate population size in American mink. LD has been applied to analyse signatures of selection identifying genes associated with growth and development (*PRDM1*); reproduction (*SLCO4C1*) and response to nutrient (*BCKDHB*) in European pig breeds (Munoz *et al*., 2019). To date, no LD studies have been done in South African pig populations. Khanyile *et al*. (2015); Makina *et al*. (2015) and Mdladla *et al*. (2016) have used high density array to estimate LD in goats, chickens and cattle respectively. Average LD observed in goats ranged from 0.09 for village populations to 0.25 for Tankwa breed (Mdladla *et al*., 2016). In chickens, LD was low in villages and high in conserved populations (Khanyile *et al*., 2015).

2.11.2 Effective population size

Crow and Kimura (1970) and Wright (1931) described effective size population (N_e) as the size of an idealised population that can contribute to the rate of inbreeding (ΔF). Fluctuations in breeding sex ratio, sample of population size and variance in reproductive success can influence N_e (Barbato *et al*., 2015*)*. Inbreeding is a major challenge when managing indigenous populations, as the effective population size of these populations is inclined to be smaller, thereby increasing the threat of fitness declining and extinction (Herrero-Medrano *et al*., 2013). It is presumed that choosing animals for breeding by choosing the lowest average co-ancestry maximizes genetic diversity in the proceeding generations (Lacy, 1995; Lindren *et al*., 1996). Risk status in livestock can be used to manage conservation programs as its acts as early warning systems. Effective population size (N_e) is the recommended estimate for assessing threats in livestock (FAO, 1992; Gandini *et al*., 2004), and is estimated based on the number of the sows and boars breeding animals in a population. Knowledge of N_e and its trends can be used to monitor and manage the loss of genetic variation within populations using methods that restrict the rate of inbreeding (Groenveld *et al*., 2009). Anderson (1991) suggested calculating the source of the gene origins for each likely candidate for breeding and then choosing animals with the highest effective number of founders. Monitoring changes in effective population size might facilitate the prediction of the demographic history of populations and provides for a better understanding of the risk of populations to inbreeding and extinction. Long-term viability of the populations can also be inferred using *Ne* that can be used to predict the possibility of kinship and genetic drift (Bittles, 2010). Effective population studies done using SNP data found effective population size of commercial Finnish Landrace and Finnish Yorkshire to be at 91 and 61, respectively (Uimari & Tapio, 2010). These were in line with the recommended effective population of 50 (FAO, 2000). Chinese indigenous breeds had higher effective population size compared to commercial breeds (Wang *et al*., 2021). This is reflective of the different selection programs.

2.12 Identity by Descent (IBD)

The advancements in molecular genetics have allowed scientists to develop strategies to predict genomic similarities inherited prior generations. Common ancestral relationships can be identified through segments coinherited from parents or grandparents dating generations ago (Palamara *et al*., 2012). High density SNPs has made it possible to estimate probabilities of a

pair of related individuals along the chromosome sharing either 0, 1 or 2 alleles by descent. Various authors (Gusev *et al*., 2008; Browning & Browning, 2012; Thompson, 2013) described identical by descent (IBD) as two alleles at a chromosomal region descending from a common ancestor. Additionally, this is more likely to be primary regions which are continually dispersed throughout the genome. The IBD concept which is now widely used in genomic mapping studies was first introduced by Malecot (1941). Downstream applications such as recombination rates (Zhou *et al*., 2020), mutation rates (Palamara *et al*., 2017), recent selection detection (Han & Abney, 2013) and estimation of kinship (Browning & Browning, 2011) can be derived based on the IBD theory. In livestock, IBD assists with locating dormant traits and fine mapping QTL. Charlier (1996) was the first to report on IBD in livestock for bovine hereditary syndactyly on chromosome 15. IBD can assist in detecting QTLs from pairs of populations sharing genomic regions due to inheritance as well as identify regions undergoing selection.

The IBD segments inherited at different periods can be used to trace recent or ancient shared ancestry. Broman and Weber (1999) and Browning and Browning (2012) observed that recent shared ancestry that occurred within the past 25 generations tend to be longer (> 3 cM) whereas older common ancestry is associated with short segments (1-3 cM). The Porcine SNP60K high-density SNP panel has the power to detect both longer and shorter IBD segments (Browning & Browning 2012). Both Browning and Browning (2012) and Palamara *et al*. (2012) noted long segments of about 8.3 cM and 6.25 cM for fifth and seventh level cousins respectively. Thompson (2013) also observed short segments emanating from high linkage disequilibrium (LD) regions. Additionally, more IBD segments tend to be shared when effective population size is smaller (Zhou *et al*., 2020). Zhou *et al*. (2020) noticed short segments to be associated with recent effective population size while IBD segments in general are useful with migration rates issues. Szpiech *et al*., (2013) observed recent shared segments to be tens of kilo bases in size while ancient IBD segments were several mega-bases long. IBD analysis in South African populations will enable fine scaling population structure and search for shared haplotypes in order to identify QTLs.

2.13 Selection of signatures

A new standard for connecting genotype to phenotype has been offered by population genomics in the form of selection signature analysis (Qanbari & Simianer, 2014). Selection mapping

methods are being applied progressively to be able to detect genomic regions associated with phenotypic diversity in domesticated livestock (Wiener & Wilkinson, 2011). Qanbari and Simianer (2014) defined selection signatures as regions of the genome that harbour useful important sequence variations and thus have been or are concealed by either natural or artificial selection, leaving special patterns of DNA. Domestication involved strong selection for specific phenotypes and could have in the process left signals of selection that should be evident in the genome (Sabeti *et al*., 2007). Nielsen *et al*. (2005) recognized that there would be modifications in the basic genetic content of the population as soon as selection is applied on a population. The genes or genomic regions targeted by domestication and selective breeding are anticipated to demonstrate signatures of selection (Chen *et al*., 2007). Since pigs have been through recent and intensive selection directed at phenotypes to advance performance in agriculture and resistance to diseases, they offer the opportunity to identify genes or genomic regions encoding quantitative trait loci (QTLs). Domestic pigs are subjected to artificial selection pressures of fertility and growth with reduction in fitness traits (Maselli *et al*., 2014; Fulgione *et al*., 2016). The interbreeding of the domestic and wild populations might have also contributed to the camouflage coat colour making them more visible to predators and hunters (Fajardo *et al*., 2018). In South African pigs, signatures of selection should be evident in regions targeted by natural and artificial selection.

Whole genome scans are used to detect genomic regions subjected to selection that can be associated with phenotypes using association mapping and gene pathways analysis. Genome tests for revealing signature of selection in natural populations are recommended when gene function or phenotype of interest is unknown, (Schlotterer, 2003). In a study by Li *et al*. (2014), selective sweep analysis detected signatures of selection in genomic regions that harbour genes underlying economic and adaptive traits such as disease resistance fertility, pork yield and body length. According to Oleksyke *et al*. (2010), genomic regions can display reduced polymorphism due to other processes like population bottlenecks, or reduced founders' effects that make it challenging to differentiate signatures of selection from other evolutionary pressures. Different selection approaches such as F_{ST} (Weir & Cockerham, 1984; Akey *et al*., 2002; Gianola *et al*., 2010), *iHS* (Voight *et al*., 2006; Sabetti *et al*. 2007)), *XP-EHH* (Sabetti *et al*. (2007)) and *HapFLK* (Fariello *et al*., 2013) have been suggested and can be used to investigate signatures.

2.13.1 Fixation index (F_{ST})

Lewontin and Krakauer (1973) were the first to describe F_{ST} statistics which is based on population differentiation. The method was further developed and used by Weir and Cockerham (1984), Akey *et al*. (2002) and Gianola *et al*. (2010). The F_{ST} detects genomic regions between two populations which are differentially fixed (Gianola *et al*., 2010). It uses differences in allele frequencies to assume genetic differentiation (Wright, 1951). While using this method, genes responsible for adaptation to high altitude fertility and body length have been observed in pigs by Ai *et al*. (2013) and Li *et al*. (2014). Both recent to ancient selective sweeps can be evaluated using F_{ST} (Guo *et al*., 2021). However, Ma *et al*. (2015) noticed that the F_{ST} method has a high amount of false positive when compared to other methods *i.e., XP-EHH* method.

2.13.2 Haplotype based methods

Chen *et al*. (2010) found haplotype-based methods to provide consistent results compared to single locus technique. Similar and rare haplotypes are clustered together and by using this method, false positives are more controlled than with single SNP based methods. The focus with haplotype-based methods is on the entire haplotype block versus each SNP (Hamazaki & Iwata, 2020) which increases accuracy in identification of complex variants.

The *iHS* technique is based on the Extended Haplotype Homozygosity (*EHH*) method by Sabetti *et al*. (2007). This is an LD based technique which targets extended homozygous regions with high frequency haplotypes (Voight *et al*., 2006). The extent of *EHH* around a SNP with inherited allele is measured against the derived allele. This method detects recent or ongoing positive selection signatures within a population (Voight *et al*., 2006). Zhao *et al*. (2015) noticed this method to be less sensitive to ascertainment bias. Chen *et al*. (2018) and Wu *et al*. (2020) used this method to investigate signatures of selection in pigs and reported 175 genomic regions associated with fat deposition in muscle (Chen *et al*., 2018), meat quality, adaptability, body size and appetite (Wu *et al*., 2020). Gouveia *et al*. (2014) reviewed signatures on the use of *iHS* method to detect signatures in livestock.

Ongoing selection that are still segregating between populations are detected using the *XP-EHH* method. Sabeti *et al*. (2007) observed this method to be efficient in identifying selection signatures from different regions by comparing two populations. *XP-EHH* is a haplotype-based method that measures LD which traces selection. This method is able to test for genome-wide

variance amongst populations (Sabeti *et al*., 2007). Ma *et al*. (2015) was able to detect signals of selection associated with ear morphology, coat colour and fertility in Chinese indigenous and Landrace and Yorkshire commercial breeds.

The *HapFLK* method is used to identify recent selection signatures. *HapFLK* detects genetic differentiation due to selection between populations using multiple population genomic data (Fariello *et al*., 2013). It takes the amount of genetic drift into consideration and estimate inbreeding co-efficient from the meta population. However, *HapFLK* is not efficient in revealing ancient signals. Molotsi *et al*. (2018) utilized *HapFLK* technique to identify signatures in South African sheep populations. The challenges noted by Molotsi *et al*. (2018) was detecting specific selection targets caused by various genes and genomic regions. The *HapFLK* method has been applied in various species including pigs (Vahedi *et al*., 2021), cattle (Maiorano *et al*., 2022), sheep (Molotsi *et al*., 2018), chicken (Bihan-Duval *et al*., 2018), goats (Onzima *et al*., 2018) and horses (Avila *et al*., 2018).

2.14 Conclusion

South Africa has a rich diversity of pig breeds, which consist of commercial, indigenous, village and wild pigs. The different pig production systems and breeds present a unique genetic resource that offers prospects for improvement of pig production to meet present and upcoming needs. Knowledge about these breeds, their production systems, the genetic diversity and population structure and their evolution is vital for comprehensive characterisation, optimal utilisation and management of the local breeds. This information is also particularly valuable for the conservation of village and indigenous pig populations in rural areas as it is currently difficult to apply genomic strategies on populations lacking genomic characterization information. The Illumina Porcine SNP60K can be applied to provide accurate and comprehensive information on the South African pig populations.

CHAPTER 3: GENOME-WIDE ASSESSMENT OF POPLATION VARIATION AND POPULATION DISTINCTIVENESS OF THE PIG FAMILY IN SOUTH AFRICA

Published in Frontiers in Genetics[1]

3.1 Abstract

Genetic diversity is of great importance and a prerequisite for genetic improvement and conservation programs in pigs and other livestock populations. The present study provides a genome-wide analysis of the genetic variability and population structure of pig populations from different production systems in South Africa relative to global populations. A total of 234 pigs sampled in South Africa and consisting of village (n = 91), commercial (n = 60), indigenous (n = 40), Asian (n = 5) and wild (n = 38) populations were genotyped using Porcine SNP60K bead chip. In addition, 389 genotypes representing village and commercial pigs from America, Europe, and Asia were accessed from a previous study and used to compare population clustering and relationships of South African pigs with global populations. Moderate heterozygosity levels, ranging from 0.204 for Warthogs to 0.371 for village pigs sampled from Capricorn municipality in Eastern Cape province of South Africa were observed. Principal Component Analysis of the South African pigs resulted in four distinct clusters of (i) Duroc; (ii) Vietnamese Potbelly; (iii) Bush Pig and Warthog and (iv) a cluster with the rest of the commercial (SA Large White and Landrace), village, Wild Boar and indigenous breeds of Kolbroek and Windsnyer. The clustering demonstrated alignment with genetic similarities, geographic location and production systems. The PCA with the global populations also resulted in four clusters that where populated with (i) all the village populations, Wild Boars, SA indigenous and the large white and landraces; (ii) Durocs (iii) Chinese and Vietnamese Potbelly pigs and (iv) Warthog and Bush Pig. K = 10 (The number of population units) was the most probable ADMIXTURE based clustering, which grouped animals according to their populations with the exception of the village pigs that showed presence of admixture. AMOVA reported 19.92% to 98.62% of the genetic variation to be within populations. Sub structuring was observed between South African commercial populations as well as between Indigenous and commercial breeds. Population pairwise F_{ST} analysis showed genetic differentiation ($P \leq$

[1] Hlongwane NL, Hadebe K, Soma P, Dzomba EF, Muchadeyi FC. 2020. Genome-wide assessment of genetic variation and population distinctiveness of the pig family in South Africa. Frontiers in Genetics 11: 344. doi: 10.3389/fgene.2020.00344.

0.05) between the village, commercial and wild populations. A per marker per population pairwise F_{ST} analysis revealed SNPs associated with QTLs for traits such as meat quality, cytoskeletal and muscle development, glucose metabolism processes and growth factors between both domestic populations as well as between wild and domestic breeds. Overall, the study provided a baseline understanding of porcine diversity and an important foundation for porcine genomics of South African populations.

Keywords: *pigs, diversity, population structure, genetic characterization, SNP60K*

3.2 Introduction

Pigs were domesticated over 5,000 years ago, leading to the gradual and cumulative development of modern pig breeds with very distinctive phenotypes and production abilities (Zeder *et al*., 2006; Rothschild & Ruvinsky, 1998). Domesticated pig (*Sus scrofa domesticus*) originated from the *Sus scrofa*, which is commonly known as the Wild Boar belonging to the Suidae family (Jones, 1998). This family includes species of wild pigs such as *Phacochoerus africanus* (Common Warthog), *Potamochoerus larvatus* (Bush Pig) and *Hylochoerus meinertzhageni* (Giant Forest Hog) some that are indigenous to Africa (Jones, 1998). The Wild Boars are widely distributed covering areas such as Europe, Asia, and North Africa and were introduced as game species in all other continents including Africa (Jones, 1998; Scandura *et al*., 2011).

Pig breeds worldwide are either of well-defined ancestry or in certain instances crossbreds from populations of diverse origins (Amills *et al*., 2010). South African pig production consists of a commercial intensive sector with defined breeds and an extensive sector that is mainly associated with small- scale farmers in the rural areas. Village production system is characterized by non-descript populations raised under extensive low-input management. Commercial breeds such as the Large White, Landrace and Duroc have worldwide distribution in modern commercial farming systems including South Africa and are widely used (Amills *et al*., 2010). Indigenous breeds classified under *Sus indica* such as Kolbroek and Windsnyer are geographically restricted to Southern Africa (Nicholas, 1999). The Kolbroek, which is of Chinese origin, is speculated to have pigs that ended up in the hands of South African farmers when a sailing shipwrecked at the Cape Hangklip (Ramsay *et al*., 1994). Although the origin of the Windsnyer is unknown, there are observed similarities to Chinese breeds (Nicholas, 1999) thereby suggesting that it is of Chinese origin. Regardless of their origins and domestication

routes, pig breeds in South Africa have become closed genetic pools restricted to specific farming systems and moulded by artificial selection and possibly genetic drift (Amills *et al*., 2010). In addition to these domesticated breeds are the Warthog, Bush Pig and Red River Hog wild pigs that are native to Africa and are found roaming in forests or in the zoos (Porter, 1993). The common Warthog (*Phacochoerus africanus*) which was first discovered at Cape Verde, Senegal is one of the three species found in Africa. The Cape Warthog (*Phacochoerus aethiopicus*) is now extinct due to the rinderpest epizootic of the 1860s (Pallas, 1766; Gmelin, 1788; D'Huart & Grubb, 2003). Another Warthog (*Phacochoerus delamerei*) species was described in Somalia and later renamed *Phacochoerus aethiopicus eelamerei* as it is similar to the Cape Warthog (Lönnberg, 1908, 1912; Roosenvelt & Heller, 1915). Muwanika *et al*. (2003) studied the phylogeography of the common Warthog in Africa and found three clades representing West, South and East African Warthogs. There is not enough evidence to support the origin of the Bush Pig, which was assumed to have originated from Asia (White & Harris, 1977). There are recordings of the Bush Pig in the Swellendam and Outeniqualand in the Western Cape provinces of South Africa (Rookmaaker, 1989). Hybrids between the domestic and Bush pigs have been recorded with the introduction of Bush Pigs to South Africa going back as far as 1 400 years ago (Linnaeus, 1758; Mujibi *et al*., 2018). The existence of hybrids is a concern, as they could become asymptomatic carriers of diseases such African swine fever (Jori & Bastos, 2009).

Indigenous breeds are often geographically restricted and harbour unique genetic variants that may provide future breeds with the flexibility to change in response to product market preferences and production environments. While low-input and indigenous breeds may not compete with exotic breeds in terms of production performance, they are considered hosts to unique genetic diversity that should be protected as sources of variation. Local pigs are important because of their hardiness and ability to survive in extreme conditions (Taverner & Dunkin, 1996; Zadik, 2005). Most indigenous breeds are, however, threatened by small and fragmented flock sizes, which predispose them to lose genetic diversity as a result of genetic drift and indiscriminate crossbreeding with exotic germplasm that can lead to genetic erosion and the eradication of the local genetic pool. Globally, 35% of pig breeds are classified as at risk or already extinct (FAO, 2009) demonstrating the threat to local biodiversity.

Genomics have emerged as an effective tool for assessing diversity within and amongst populations. Swart *et al*. (2010) observed low differentiation among pig populations in

Southern Africa using microsatellites. Heterozygosity levels ranged from 0.531 to 0.692 for commercial and indigenous breeds. The availability of the Porcine SNP60K bead chip has opened new avenues of examining genetic diversity (Ramos *et al*., 2009) at a genome-wide scale relative to that using microsatellite and other low-coverage markers. Mujibi *et al*. (2018) observed close clustering of Warthogs and Bush Pigs using the Porcine SNP60K bead chip. The Porcine SNP60K bead chip has been used to infer on population structure and selection signatures in Chinese and European pig populations (Ai *et al*., 2013). Using this SNP panel in South African pig populations will provide comprehensive information on the genomic architecture of local, exotic and wild pig populations, which will guide future management and conservation. The objective of the present study was to provide a large-scale analysis of the genetic diversity and structure of South African local pig populations using the Porcine SNP60K bead chip. The study investigated diversity of South African pigs relative to global populations of 389 pigs consisting of villages and out-group pigs from South America, Europe, United States and China amongst other countries.

3.3 Materials and methods

3.3.1 Breed/populations sampled

South African specimens were collected from a total of 234 samples from different production systems, representing village, intensively farmed populations in conservation units and free ranging populations. Village and non-descript pig populations were sampled from Alfred Nzo (ALN; *n* = 17) and Oliver Reginald Tambo (ORT; *n* = 22) districts in Eastern Cape province and Mopani (MOP; *n* = 27) and Capricorn (CAP; *n* = 25) districts in Limpopo province. Commercial pig breeds of Large White (LWT; *n* = 20), South African Landrace (SAL; *n* = 20) and Duroc (DUR; *n* = 20) were sampled from commercial farmers in Limpopo province. Indigenous populations Kolbroek (KOL; *n* = 20.) and Windsnyer (WIN; *n* = 20) were sampled from the Agricultural Research Council - Animal Production Institute in Pretoria, South Africa (Table 3.1). Vietnamese Potbelly breed (VIT; *n* = 5) was sampled from the Johannesburg Zoo and represents a breed that is endangered in Vietnam, its country of origin but has been raised in a conservation zoo in South Africa. European Wild Boar (*n* = 4), Warthogs (*n* = 31), and Bush Pigs (*n* = 3) were sampled as representatives of the wild pig populations. The European Wild Boar and Bush Pigs were sampled from the surrounding villages in the North-West whilst the Warthog samples were collected from geographically separated National Parks from North

-West (*n* = 4), Eastern Cape (*n* = 3), and Limpopo (*n* = 24). The distribution of the sampled individuals is illustrated in Figure 3.1. Ear tissue samples were collected using the tissue sampling applicator gun while pliers ere used to collect the hair samples according to standard procedures and ethical approval from ARC-Irene Animal Ethics Committee (APIEC16/028).

Table 3. 1 Summary of the sampled South African pig populations

Category	Population	Code	N
Village	Mopani	MOP	27
Village	Capricorn	CAP	25
Village	Oliver Reginald Tambo	ORT	22
Village	Alfred Nzo	ALN	17
Commercial	Large White	LWT	20
Commercial	SA Landrace	SAL	20
Commercial	Duroc	DUR	20
Indigenous	Kolbroek	KOL	20
Indigenous	Windsnyer Type	WIN	20
Asian	Vietnamese Potbelly	VIT	5
Wild	Wild Boar	WBO	4
Wild	Warthog	WAT	31
Wild	Bush Pig	BSP	3

3.3.2 Genotyping and quality control

DNA was extracted at the Agricultural Research Council- Biotechnology Platform from the ear tissue and hair samples using a commercially available Perkin Elmer Genomic DNA kit according to the manufacturer's protocol. DNA concentration was quantified using the Qubit 2.0 Fluorometer. Gel electrophoresis (5%) was used to assess the quality and integrity of the DNA.

Figure 3.1 Map showing geographic locations of the 13 pig populations in the present study

All 234 animals were genotyped using Porcine SNP60K *version* 2 genotyping bead chip (Illumina, United States) containing 62 163 SNPs with an average gap of 43.4 kb. Genotyping was done using the standard Infinium assay at the ARC- Biotechnology Platform in South Africa. GenomeStudio *version* 2.0 (Illumina, United States) was used to process the genotype data, including raw data normalization, clustering and genotype calling. A final custom report was created to be able to generate a Plink Ped (Pedigree file) and Map (SNP panel file) for use in downstream analysis.

Golden Helix SNP Variation Suite (SVS) *version* 8.5 was used to update the SNPs marker file (Golden Helix Inc., 2016) based on the pig genome assembly (*Sus scrofa version* 10.2). Markers

were then filtered to exclude SNPs located on the sex chromosomes. From this data set, Minor allele frequency (MAF) and deviation from Hardy-Weinberg equilibrium (HWE) were estimated per population for the 10 populations that excluded BSP, VIT, and WBO, which were left out due to small sample sizes. Additional quality control (QC) was also performed per population to remove SNPs with less than 85% call rate, MAF < 0.02 and HWE < 0.0001. The resultant filtered dataset was used to calculate observed (H_O), and expected (H_E) heterozygosities, inbreeding (F_{IS}) and effective population size (N_e).

Quality control was then performed overall population to remove SNPs with less than 85% call rate, MAF < 0.02 and HWE < 0.0001 and generate a dataset used for analysis of molecular variance (AMOVA) and F_{ST} analysis. Using this dataset, further QC filtered for SNPs in high LD (r^2 = 0.2) and closely related individual [identity by descent (IBD) ≥ 0.45] to produce a filtered dataset used for population structure analysis using ADMIXTURE and Principal Component Analysis (PCA).

3.3.3 Genetic diversity within population

The MAF, H_E and H_O were calculated as measures of within population genetic variation using PLINK 1.07 (Purcell *et al*., 2007). In addition, inbreeding coefficient (F_{IS}) was calculated on Golden Helix SNP Variation Suite (SVS) *version* 8.5 (Golden Helix Inc., 2016). Effective population size (N_e) trends across generations were estimated based on a relationship between r^2 (expected LD), *Ne* and C (recombination rate). SNeP software (*v* 1.1) tool was used based on the following formula suggested by Corbin *et al*. (2012) using the equation:

$$N_{T\ (t)} = \frac{1}{\left(4f(C_t)\right)} \frac{1}{E\left[r^2_{adj} \middle| C_t\right]} - \propto$$

Where:

$N_{T(t)}$: Effective population size estimated t generations ago

C_t:Recombination rate t generations ago

r_{2adj}: Linkage disequilibrium estimation adjusted for sampling biasness

α: a constant.

The recombination rate by using the following formula proposed by Sved (1971):

$$f(c) = c\left[\frac{\left(1-\frac{c}{2}\right)}{(1-2)^{2}}\right]$$

The Bush Pig, Vietnamese Potbelly and Wild Boar were excluded from the diversity within population analysis due to their small sample sizes. The few available samples were sampled from zoos and game reserves in the country where only few animals are often rescued and kept in conservation.

3.3.4 Population differentiation and structure

Analysis of Molecular Variance (AMOVA) was used to determine the genetic variance within populations (F_{IS}), among populations within group (F_{SC}) and among groups (F_{CT}) using ARLEQUIN v3.5 (Excoffier *et al*., 2005). The populations were categorized into villages, commercial, indigenous and wild populations and consisted of animals sampled in South Africa as well global populations from Burgos-Paz *et al*. (2013) which consisted of 389 genotypes of villages and out-group pigs from 24 countries of America (United States), South America (Mexico, Cuba, Guadeloupe, Guatemala, Costa Rica, Columbia, Ecuador, Peru, Brazil, Bolivia, Paraguay, Argentina, and Uruguay), Europe (Spain, Portugal, Italy, Poland, Hungary, Tunisia, Denmark, Holland, United Kingdom) and China. Variance components were also estimated for groups consisting of different categories, *i.e*., village and indigenous; indigenous and commercial; South African village and global villages; South African commercial and global commercial *etc*.

Principal Component Analysis (PCA) using SVS *version* 8.5 (Golden Helix Inc., 2016) and the eigenvector method was used to determine population clustering. ADMIXTURE *version* 1.20 (Alexander & Lange, 2011) was used to detect the most likely clusters (K) for the population. ADMIXTURE was run from $K = 2$ to $K = 15$. The number of potential genetic clusters (K) was tested from 1-15 to reassign each sample to its population of origin. The optimum K-value was that with the lowest cross- validation error value. Initially, all the 13 populations sampled from South Africa were included in the population structure analysis. After this, the South African data set was merged to Porcine SNP60K genotype data from Burgos-Paz *et al*. (2013) described above.

Population pairwise F_{ST} values were estimated according to the formula of Weir and Cockerham (1984) implemented in the Golden Helix SNP Variation Suite (SVS) *version* 8.5 (Golden Helix Inc., 2016). Based on population pairwise F_{ST} values, PCA and ADMIXTURE based clustering, F_{ST} analysis per marker was estimated between pairs of highly differentiated populations of the village populations, indigenous populations and commercial breeds as well as amongst highly differentiated commercial breeds and wild populations. To lessen the discord, an F_{ST} averaged smooth value was used to identify genomic regions differentiating pairs of populations. Manhattan plots of per marker F_{ST} values between pairs of populations were plotted against chromosomal coordinates using the porcine assembly (*Sus scrofa* 10.2). Highly differentiating SNPs ($F_{ST} \geq 0.8$) were sub-sampled, and genes associated with these SNPs searched using genome browse including their associations with known QTLs in the pig genome based on the *Sus scrofa* 10.2 on Ensemble.

3.4 Results

3.4.1 Genotype and quality control

The percentage of polymorphic and number of SNPs (N_{SNP}) remaining after QC per population and overall is presented in Table 3.2. Two hundred and eleven individuals with a genotyping rate of 85% remained after QC. Windsnyer pigs had the highest percentage of informative markers (95%) after QC, whilst Warthog had the lowest at 82%. About 31 705 SNPs were removed leaving 30 458 polymorphic SNPs of the loci distributed over 18 autosomal chromosomes, which were used for AMOVA and F_{ST} analysis. After LD and IBD pruning, 23 345 SNPs and 176 individuals were used for the population structure analysis.

Table 3. 2 Summary of the genetic diversity measures across South African pig populations

Pop	N	%SNP	MAF ± SD	N_{SNP}	H_O ± SD	H_E ± SD	F_{IS} ± SD	*P-value*
MOP	27	92	0.262±0.149	52 925	0.299±0.129	0.369±0.131	0.198±0.134	0.495
CAP	24	94	0.264±0.147	54 078	0.332±0.140	0.371±0.126	0.117±0.155	0.582
ORT	22	93	0.259±0.153	52 238	0.315±0.145	0.370±0.130	0.163±0.113	0.553
ALN	15	94	0.238±0.157	53 580	0.336±0.160	0.359±0.134	0.056±0.168	0.695
LWT	18	93	0.227±0.161	49 773	0.358±0.177	0.348±0.144	-0.023±0.009	0.721
SAL	19	94	0.221±0.162	49 191	0.372±0.186	0.345±0.144	-0.052±0.085	0.704
DUR	19	94	0.177±0.168	40 632	0.359±0.182	0.337±0.147	-0.067±0.153	0.764
KOL	20	94	0.173±0.167	39 560	0.364±0.182	0.339±0.144	-0.051±0.087	0.727
WIN	19	95	0.220±0.164	47 402	0.385±0.171	0.360±0.134	-0.056±0.158	0.733
WAT	28	82	0.076±0.109	3 967	0.188±0.155	0.204±0.151	0.0771±0.251	0.710

%SNP used to calculate MAF analysis; N_{SNP}, the number of SNPs in the subset 62,163 SNP; H_O, observed heterozygosity; H_E, expected heterozygosity; SD, standard deviation; F_{IS}, inbreeding co-efficient; MAF, minor allele frequency, $P < 0.05$.

3.4.2 Genetic diversity across populations

Genetic diversity parameters among the 10 populations are summarized in Table 3.2. Warthog pigs had the lowest H_O (0.188 ± 0.155) and Windsnyer the highest (0.385 ± 0.171). Expected heterozygosity values ranged from 0.204 ± 0.151 from Warthog to 0.371 ± 0.126 for Capricorn. The highest inbreeding coefficient (F_{IS}) was for Warthog at 0.398 ± 0.475 while the Duroc had the lowest and slightly negative value of −0.067 ± 0.153. F_{IS} values were positive for all village populations as well as Warthog suggesting some level of inbreeding within these populations. MAF was the highest in village population from Capricorn (0.264 ± 0.147) and the least in Warthog pigs (0.076 ± 0.109).

3.4.3 Effective population size

Figure 3.2 shows trends in effective population size across all of the studied populations. The Warthog was excluded in this analysis because the number of polymorphic SNPs was not enough to generate results. Effective population size values are presented in Supplementary

Table 3.1. There was a general decline in N_e across all the populations across generations. The indigenous and commercial populations had lower effective population size compared to the village populations. The Kolbroek had the lowest effective population size 12 generations prior.

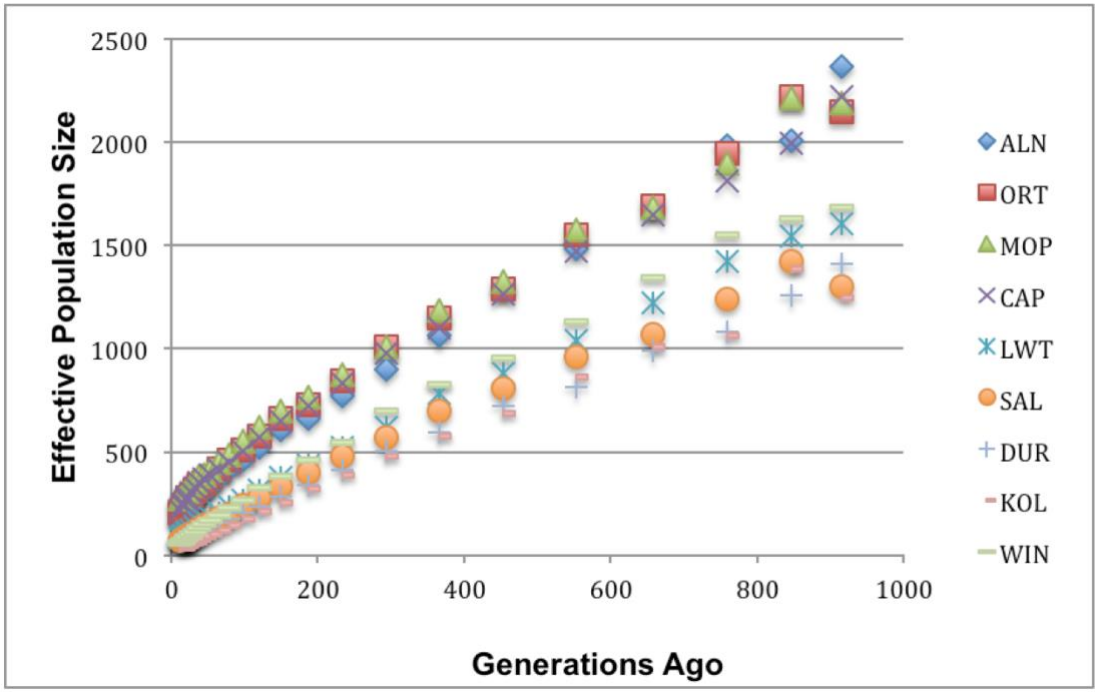


Figure 3.2 Average effective population size plotted against generation in the past

3.4.4 AMOVA

Genetic differentiation between populations is presented in Supplementary Table 3.2. The major proportion of the genetic variance was attributed to variation within South African populations with F_{IS} values ranging from 76.41 to 98.62%. Diversity within populations (F_{IS}) in village populations from this study and those from Burgos-Paz *et al.* (2013) was 35.52% while variation among groups (F_{CT}) was 62.35%. Diversity of South African commercial pigs were 76.41% within populations, 18.17% among populations within group and 5.42% among groups. When including the commercial breeds from Burgos-Paz *et al.* (2013), the diversity parameters changed to F_{IS} = 30.97%, F_{SC} = 8.31% and F_{CT} = 60.72%. High F_{CT} (> 60%) were observed in the category consisting of South African indigenous and Chinese indigenous (F_{CT}

= 70.08%) as well as that consisting of the South African Wild Boar and the worldwide Wild Boar (F_{CT} = 73.58%).

3.4.5 Population structure

Principal component one (PC1) and principal component two (PC2) explained approximately 30.7% and 11.8% of the total variation, respectively. The PCA of South African breeds yielded four main genetic clusters (Figure 3.3). The Duroc clearly separated from the Large White and South African Landrace that clustered together with the Wild Boar and village populations. The Warthog and the Bush Pig clustered together as a third cluster whilst the fourth cluster consisted of Vietnamese Potbelly sampled from the zoo. The PCA analysis using South African samples and those from Burgos-Paz *et al.* (2013) demonstrated the same clustering with all the village pigs grouping together with the Large White and Landraces separated from clusters of (i) Warthog and Bush Pig, (ii) Chinese and Vietnamese Potbelly breed and (iii) Duroc (Figure 3.4).

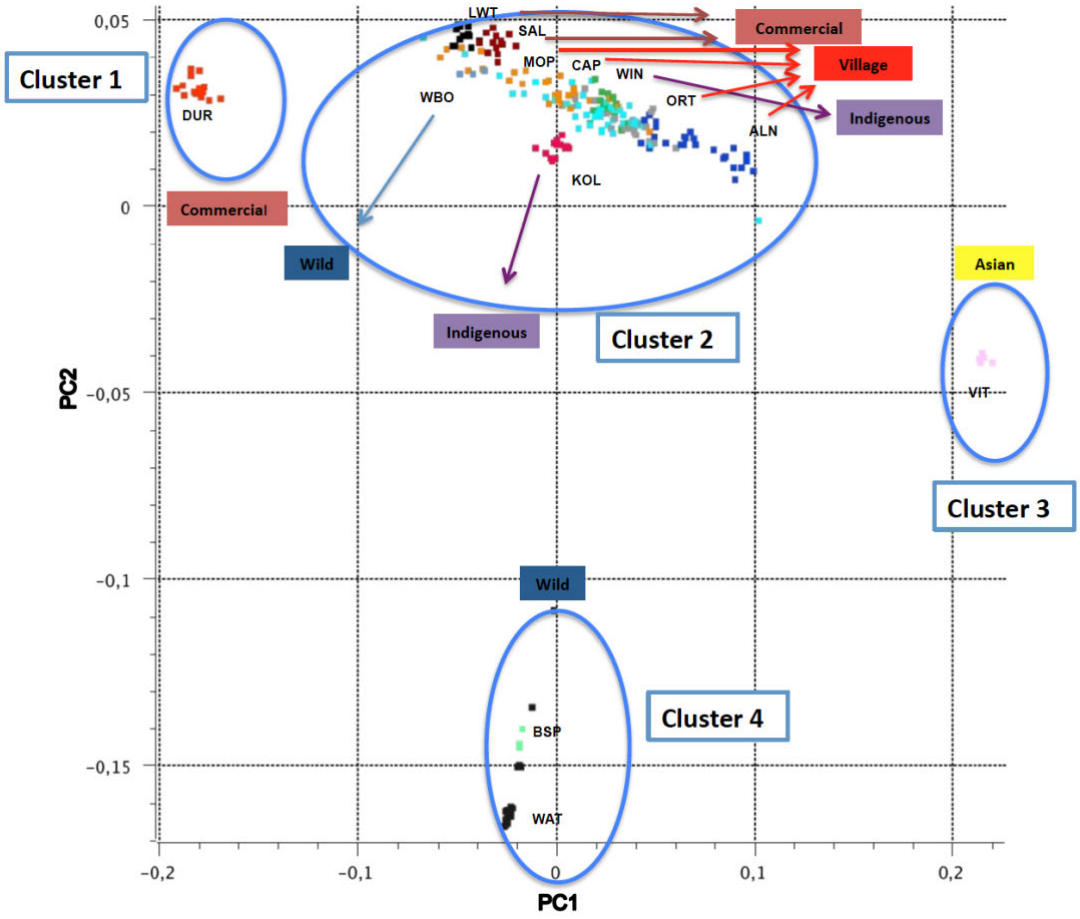


Figure 3.3 Principal Component Analysis based population clustering

1541
1542

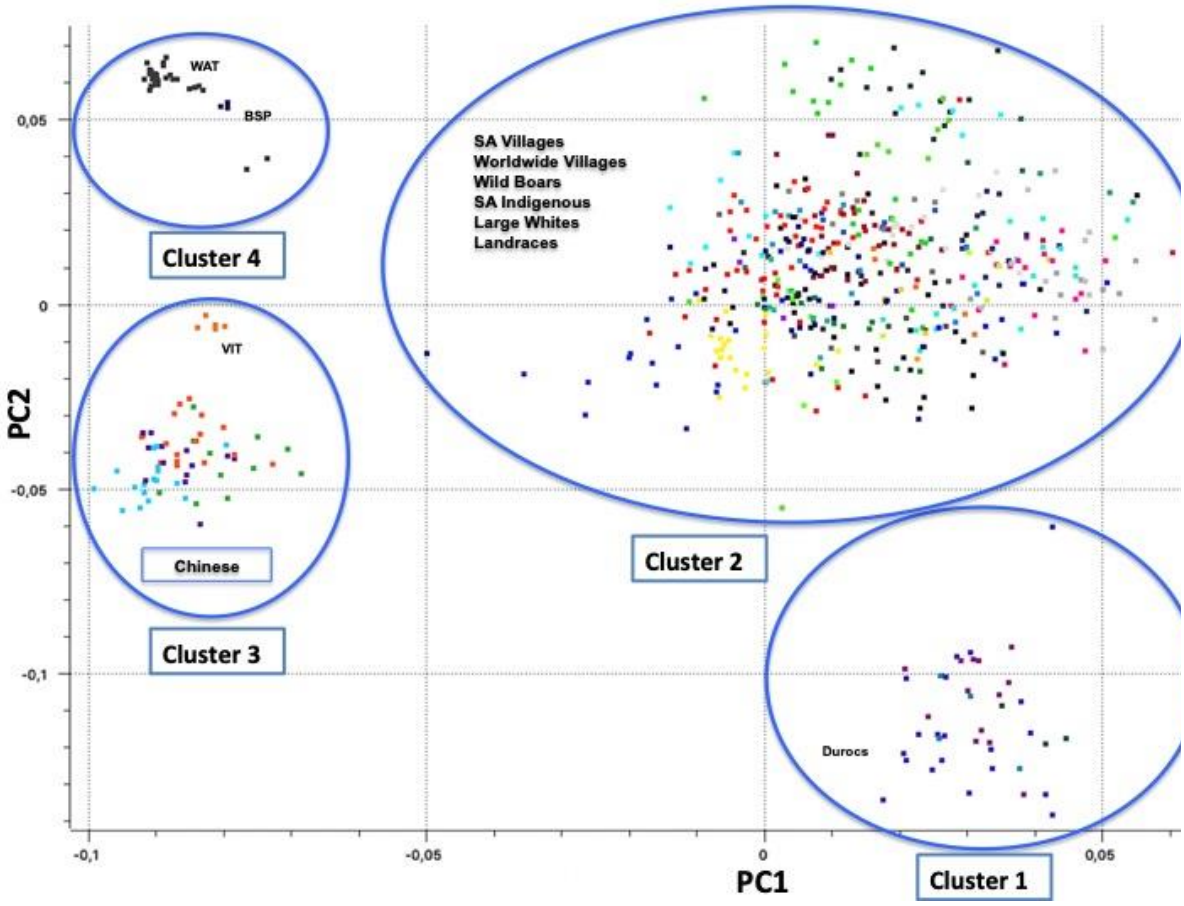


Figure 3.4 Principal Component Analysis based population clustering including Burgo-Paz *et al*. (2013) genotypes (22 430)

Genetic structure of the South African breeds was further investigated using ADMIXTURE. The results presented in Figure 3.5 show the Warthog and Bush Pigs populations clustering together and clearly separated from the rest of the other populations at $K = 2$. Duroc separated from the rest of the populations at $K = 3$ followed by Vietnamese Potbelly at $K = 4$. $K = 4$ clustered animals in the same way observed with PCA based clustering. Beyond $K = 8$, the genetic clusters of the commercial, indigenous, Asian and wild breeds are maintained whilst the added K is distributed within the village populations. $K = 10$ which was the optimal K (Supplementary Figure 2.1) with lowest CV (0.551) resulted in the eight distinct genetic clusters of commercial, indigenous, Asian and wild breeds plus highly admixed clusters consisting of all village pig populations from Limpopo and Eastern Cape provinces of South Africa.

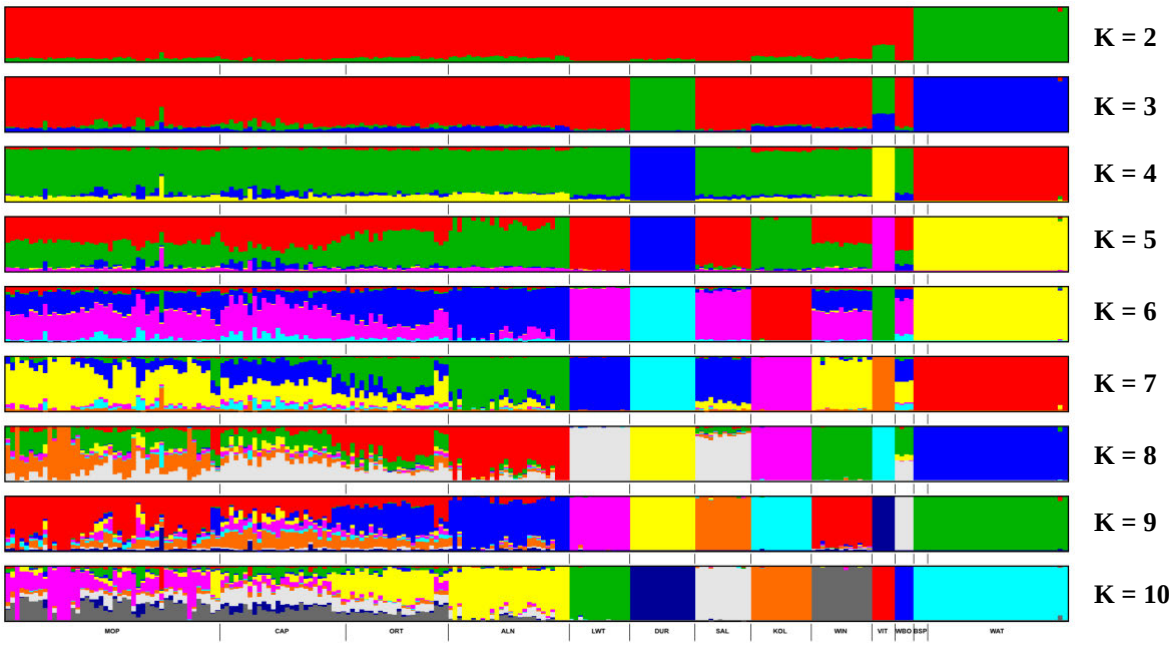


Figure 3.5 ADMIXTURE based clustering *K* = 2 – *K* = 10. Each individual is represented by a single column divided into K coloured segments, where *K* is the number of clusters assumed with lengths proportional to each of the K inferred cluster.

3.4.6 Population differentiation

Population pairwise F_{ST} values are shown in Table 3.3. Low F_{ST} were observed between village populations with values ranging from 0.022–0.060 ($P < 0.05$) within South Africa and in global populations. The highest differentiation was found between Warthog and Duroc at F_{ST} = 0.481. Warthog and Kolbroek pigs showed the high differentiation at 0.468. All other populations had F_{ST} values above 0.282. The extent of differentiation between Warthog and all the other populations was high ranging from 0.312 (Warthog and Creole from Columbia) to 0.589 (Warthog and Vietnamese Potbelly). Highest F_{ST} observed was between Vietnamese Potbelly and Bush Pigs populations at 0.700 (Supplementary Table 3.3).

Table 3.3: Pairwise genetic differentiation (F_{ST} values) between 10 pig populations

	MOP	CAP	ORT	ALN	LWT	DUR	SAL	KOL	WIN	WAT
MOP										
CAP	0.022									
ORT	0.031	0.026								
ALN	0.059	0.060	0.040							
LWT	0.091	0.073	0.096	0.130						
DUR	0.134	0.126	0.143	0.174	0.183					
SAL	0.094	0.073	0.099	0.132	0.120	0.194				
KOL	0.120	0.116	0.129	0.162	0.189	0.237	0.194			
WIN	0.061	0.064	0.077	0.106	0.143	0.189	0.144	0.173		
WAT	0.282	0.306	0.314	0.350	0.433	0.481	0.435	0.468	0.410	

3.4.7 Per marker pairwise F_{ST}, SNP annotation and association with Porcine QTL

Per population, per marker pairwise F_{ST} values were computed for highly differentiated populations and are illustrated in Table 3.3, Supplementary Figure 2.2. SNPs. High F_{ST} values (≥ 0.8) where considered breed differentiating and the associated SNPs were functionally annotated for genes within a 1 MB region. Fixed SNPs (F_{ST} = 1.0) where observed on chromosome 9 between Duroc and Warthog, on chromosome 12 between Kolbroek and Warthog and on chromosome 18 between Windsnyer and Warthog. For all the pairwise comparisons, 281 SNPs ($F_{ST} \geq 0.8$) were detected (Supplementary Figure 2.2) with only 123 candidate genes within 1 MB of those SNPs. Pairwise comparison of village pigs from Alfred Nzo, South Africa and Warthog yielded genes related to acute heat stress (*RPL18*) and inflammatory response (*IL17B* and *ARHGAP23*) as illustrated in Table 4 and Supplementary Figure 2.2. Gene *ADGRB3* was in close proximity of SNPs rs81353971, rs81353988, rs81353991, rs81297001, and rs81333295 that were of significant between Duroc and Warthog. Inflammatory response genes such as *ARHGAP23* were associated with the significant SNPs observed between Kolbroek, Large White and Windsnyer populations. For reproduction traits, genes *CD28, TCP11L2, TLK1, ATPB2, GPR137C, ZNF609, ARHGAP22, EPSTI1, GPR63, TCTE3, PTP4A2, ZSCAN20, CLU*, and *CACNA2D3* were observed within 14 significant SNPs

Table 3.4 Most significant SNPs detected with F_{ST} analysis and the associated genes

Pop	SNP	Chr	Position	Gene	Function
ALN and WAT	rs81355030	1	84,376,735	*RPL18*	Acute heat stress (Newton *et al*., 2012)
	rs81367521	2	50,546,025	*IL17B*	Embryonic development, tissue regeneration and inflammation (Bie *et al*., 2017)
	rs81285672	12	23,638,629	*ARHGAP23*	Inflammatory response (Liu, 2015)
DUR and WAT	rs81353971	1	49,024,494	*ADGRB3*	Growth traits (Emrani *et al*., 2017)
	rs81353988	1	49,350,539	*ADGRB3*	
	rs81353991	1	49,392,902	*ADGRB3*	
	rs81297001	1	49,458,254	*ADGRB3*	
	rs81333295	1	49,592,586	*ADGRB3*	
	rs80946298	13	33,531,504	*DOCK3*	Induces axonal growth (Kimura *et al*., 2016)
	rs81444796	13	33,481,604	*DOCK3*	
	rs81478683	13	34,024,632	*IQCF3*	Conjunctival UV to auto fluorescence (Yazar *et al*., 2015)
	rs81478482	13	34,117,528	*ACY1*	Amino acid and heat shock protein (Martínez-Montermayor *et al*., 2008)
	rs81454214	15	107,134,695	*CD28*	Endometrial gene expression (Gu *et al*., 2014)
KOL and WAT	rs81341610	3	4,508,681	*LOC102160627*	Uncharacterized
	rs80993200	4	234,605	*ARHGAP39*	Milk production related and mastitis resistance (Wang *et al*., 2015)
	rs80851822	5	13,913,761	*POLR3B*	Residual feed intake (Gondret *et al*., 2017)
	rs80873063	5	13,940,475	*TCP11L2*	Regulate in small atretic follicles for healthy follicles (Hatzirodos *et al*., 2014a)
	rs80999600	5	66,998,856	*TSPAN9*	ADG (Fontanesi *et al*., 2014)
	rs80929588	5	67,092,749	*TSPAN9*	
	rs80883075	5	67,132,255	*TEAD4*	Regulation in organ size control and cell proliferation (Frankenberg *et al*., 2016)
	rs81385003	5	67,292,728	*ITFG2*	Disease resistance (Moioli *et al*., 2016)
	rs81285672	12	23,638,629	*ARHGAP23*	Inflammatory response (Liu, 2016)
	rs81325261	12	44,771,203	*FOXN1*	Regulation of hair follicles development (Song *et al*., 2017)
	rs80801871	13	33,170,033	*DOCK3*	Induces axonal growth (Kimura *et al*., 2016
	rs80802886	13	33,202,454	*DOCK3*	
	rs81444784	13	33,306,071	*DOCK3*	
	rs81444796	13	33,481,604	*DOCK3*	
	rs80946298	13	33,531,504	*DOCK3*	
	rs81478683	13	34,024,632	*IQCF3*	Conjunctival UV to auto fluorescence (Yazar *et al*., 2015)
	rs335091311	15	148,461	*STAM2*	Residual feed intake (Gondret *et al*., 2017)
	rs80852223	15	77,232,829	*TLK1*	Decrease expression in the endometrium (Gray *et al*., 2006)
	rs80999734	15	77,318,065	*TLK1*	
	rs81453662	15	78,190,260	*DLX1*	Muscling and meat availability (Li *et al*., 2010)
LWT and WAT	rs81349766	1	182,224,202	*GPR137C*	Litter size (Sosa-Madrid *et al*., 2018)
	rs81296498	1	182,722,677	*DDHD1*	Lipid metabolism (Parker Gaddis *et al*., 2018)
	rs81349773	1	182,756,343	*DDHD1*	
	rs332395415	1	246,195,557	*ABCA1*	Mediates the transport of excess cholesterol (Schwartz *et al*., 2000)
	rs321979518	1	246,199,966	*ABCA1*	
	rs81383185	5	21,606,108	*RNF41*	Lipid rafts immune signalling (McGraw & List, 2017)
	rs808220161	5	21,745,636	*STAT2*	Milk production (Salehi *et al*., 2005)
	rs80894897	5	21,727,701	*PAN2*	Fat yield (Suchocki *et al*., 2016)
	rs80940129	5	21,970,939	*BAZ2A*	Nutrition related (Cornelis & Hu, 2013)
	rs3252229936	5	22,338,939	*MYO1A*	Coat colour and pigmentation (Gutiérrez-Gil *et al*., 2007)
	rs81285672	12	23638,629	*ARHGAP23*	Inflammatory response (Liu, 2015)
	rs80854565	14	89,185,576	*ARHGAP22*	Fertility (Browett *et al*., 2018)
	rs80833618	14	89,227,581	*ARHGAP22*	
	rs80957034	14	89,255,703	*ARHGAP22*	
	rs80962102	14	89,309,115	*ARHGAP22*	
SAL and LWT	rs81395957	6	51,328,753	*NECTIN2*	Cell recognition and adhesion (Wang *et al*., 2010)
	rs81395929	6	51,427,663	*CLPTM1*	Marbling score (Lim *et al*., 2013)
WIN and WAT	rs13811252	4	65,339	*ZNF609*	Fertility (Hatzirodos *et al*., 2014b)
	rs81285672	12	23,638,629	*ARHGAP23*	Inflammatory response (Liu, 2015)
	rs81325261	12	44,771,203	*FOXN1*	Regulation of hair follicle development (Song *et al*. 2017)
	rs331955329	13	66,004,327	*MTMR14*	Reduced with age accelerates skeletal muscle aging (Romero-Suarez *et al*., 2010)
	rs80971430	13	66,026,240	*BRPF1*	Intramuscular fatty acid (Puig-Olíveras *et al*., 2016)
	rs80945527	13	66,104,857	*ARPC4*	Mastitis resistance (Grossi *et al*., 2014)
	rs80885182	13	66,270,725	*FANCD2*	Muscle weight (Lionikas *et al*., 2010)
	rs45430493	13	66,515,894	*SEC13*	Muscle weight (Lionikas *et al*., 2012)
	rs81248260	13	66,583,753	*ATPB2*	Heat stress on reproductive performance (Dash *et al*., 2016)
	rs1446451	13	66,668,301	*ATPB2*	
	rs81446497	13	66,691,206	*ATPB2*	
	rs81446475	13	66,725,741	*ATPB2*	
	rs1446484	13	66,777,686	*ATPB2*	
	rs81478601	13	66,795,578	*ATPB2*	
IND and DUR	rs80866460	4	106,698,412	*PTPN22*	Immune response (Lamsyah *et al*., 2009)
	3rs81413279	9	79,010,742	*NXPH1*	DMI (Olivieri *et al*., 2016)
	rs81413279	9	79,010,742	*ABCB5*	Immune function (Lee *et al*., 2017)

Table 3.4 Most significant SNPs detected with F_{ST} analysis and the associated genes

Pop	SNP	Chr	Position	Gene	Function
	rs81306790	6	89,661,963	*PHC2*	Mastitis (Chen *et al*., 2015)
	rs80854994	4	106,719,032	*PTPN22*	Immune response (Lamsyah *et al*., 2009
	rs80854994	4	106,719,032	*BCL2L15*	Mastitis (Chen *et al*., 2015)
Villages and DUR	rs81282695	6	94,442,844	*POU3F1*	Neurobehavioral functioning (Eusebi *et al*., 2018)
	rs81282695	6	94,442,844	*FHL3*	Carcass traits (Zou *et al*., 2004 ; 2007)
Villages and KOL	rs81430450	11	24,063,007	*DNAJC15*	Feeding efficiency (Reyer *et al*., 2017a)
	rs81430450	11	24,063,007	*EPSTI1*	Fertility traits (Gaddis *et al*., 2016), fat deposition (Zhang *et al*., 2018)
SAL&LWT and IND	rs81232179	8	51,070,662	*FSTL5*	Meat quality (Gaddis *et al*., 2016), fat deposition (Zhang *et al*., 2018)
	rs5431508	8	69,912,174	*CXCL8*	Pig disease (Wang *et al*., 2019)
	rs81400554	8	55,181,102	*CEP135*	Intramuscular fat (Hamill *et al*., 2012); milk production (Rui *et al*., 2019)
	rs81400554	8	55,181,102	*EXOC1*	Marbling score (Wu *et al*., 2016)
	rs81400740	8	63,119,376	*EPHA5*	Feeding efficiency (Reyer *et al*., 2017b)
	rs81400500	8	52,213,568	*NPY5R*	Feed efficiency and fat deposition (Chen *et al*., 2018)
	rs81400500	8	52,213,568	*NPY1R*	Feed efficiency and fat deposition (Chen *et al*., 2018)
	rs81302014	8	69,950,857	*RASSF6*	Body conformation (Fang & Pausch, 2019)
	rs80904678	11	15,274,089	*FOXO1*	Meat quality and carcass traits (Ropka-Molik *et al*., 2018)
	rs81400500	8	52,213,568	*SLC7A11*	Feed efficiency (Vigors *et al*., 2016)
	rs81300083	9	78,940,661	*NXPH1*	DMI (Olivieri *et al*., 2016)
IND and VIT	rs81350922	1	257,096,974	ASTN2	Carcass weight in cattle (Júnior *et al*., 2016)
	rs80970078	14	43,524,181	MYO18B	Meat quality and carcass traits (Ropka-Molik *et al*., 2018)
	DRGA0006738	6	117,857,953	NOL4	Fatness (Li *et al*., 2011)
	rs80860919	1	64,018,444	GPR63	Fertility traits (Moran *et al*., 2017)
	rs80921694	13	73,023,057	PLXNA1	Meat quality (Martínez-Montes *et al*., 2016)
	rs1327396	12	53,063,765	DNAH2	Intramuscular fat (Lou *et al*., 2012), carcass weight (Kang *et al*., 2013)
Villages and WBO	rs81244815	2	50,167,007	*SWAP70*	Disease resistance (Ma *et al*., 2011; Zhang *et al*., 2018)
	rs81244815	2	50,167,007	*SBF2*	Fertility (Zhang *et al*., 2014), immune function (Ibeagha-Awemu *et al*., 2016)
	rs81401075	8	73,841,435	*FRAS1*	Sow reproductive traits (Fischer *et al*., 2015), feed efficiency (Messad *et al*., 2019)
	rs81401075	8	73,841,435	*NPY2R*	Obesity (Siddiq *et al*., 2007; Hunt *et al*., 2011)
Villages and VIT	INRA0003181	1	95,198,598	*SLC14A2*	Conformation traits (Le *et al*., 2017)
	rs81332040	6	45,777,816	*ZBF382*	Conformation traits (Le *et al*., 2017
	INRA0045852	14	103,086,988	*HECTD2*	Fat and meat quality traits ((Piórkowska *et al*., 2018)
	rs80980839	4	93,722,493	*RHBG*	Ammonia transporter (Xiang *et al*., 2016)
	rs80971176	5	49,876,132	*SOX5*	Ear morphology (Edea *et al*., 2017)
WBO and DUR	rs80837120	1	565,627	*TCTE3*	Involved in spermatogenesis (Du *et al*., 2016)
	rs80988392	1	213,780,848	*PTPRD*	Meat quality (Raschetti *et al*., 2013)
	rs80790807	4	106,750,789	*PTPN22*	Immune response (Lamsyah *et al*., 2009)
	rs80790807	4	106,750,789	*BCL2L15*	Mastitis (Chen *et al*., 2015)
	rs80855522	4	110,552,282	*GNAI3*	Heat tolerance (Berihulay *et al*., 2019)
	rs81389936	6	88,264,983	*COL16A1*	Carcass and meat quality traits (Choi *et al*., 2012)
	rs81389959	6	88,334,239	*PTP4A2*	Reproductive traits (Verardo *et al*., 201), intramuscular fat (Martínez-Montes *et al*., 2016)
	rs81317489	6	89,640,457	*ZSCAN20*	Scrotal circumference (Sweett *et al*., 2018)
	rs81317489	6	89,640,457	*CSMD2*	Meat pH trait (Dong *et al*., 2014), Body weight (Yoshida *et al*., 2017)
	rs81390106	6	88,751,010	*TMEMM39B*	Intramuscular fat (Cesar et al., 2018)
	rs81390106	6	88,751,010	*TXLNA*	Meat quality (Ropka-Molik *et al*., 2018)
	rs81390106	6	88,751,010	*HDAC1*	Altitude (Ban *et al*., 2015)
	rs81390106	6	88,751,010	*MARCKSL1*	Feed intake (Lindholm-Perry *et al*., 2016)
	rs80975991	7	33,481,446	*ZFAND3*	Growth and carcass quality traits (Li & Kim, 2015)
	rs81330369	9	7,449,894	*FCHDS2*	Milk production traits (Kemper *et al*., 2015)
	rs80894853	9	78,663,586	*NXPH1*	DMI (Olivieri *et al*., 2016)
	rs343528814	13	36,608,977	*CACNA2D3*	Reproductive traits (Smith *et al*., 2019), body width in gilts and sows (Rothschild, 1998), body weight traits (Borowska *et al*., 2017), altitude (Zhang *et al*., 2014)
	rs81478390	13	53,707,241	*RYBP*	Body conformation traits-body weight, body length, body height and chest circumference (Zhou *et al*., 2016)
	rs80911350	14	11,345,116	*SCARA3*	Meat quality traits (Tizioto *et al*., 2015)
	rs80911350	14	11,345,116	*CLU*	Fertility (Kumar *et al*., 2015), intramuscular fat (de Jager *et al*., 2015)

on chromosomes 1, 2, 5, 6, 11, 14 and 15. Genes that had association with meat traits such as *DLX1*, *BRPF1*, *CLPTM1*, *FANCD2*, *SEC13*, *FHL3*, *FSTL5*, *CEP135*, *EXOC1*, *FOXO1*, *ASTN2*, *MYO18B*, *PLXNA1*, *DNAH2*, *HECTD2*, *TMEM39B*, *TXLNA*, *CSMD2*, *COL16A1*, *SCARA3*, *ZFAND3* and *PTPRD* were also reported. Comparison with indigenous pigs showed

genes that were associated with mastitis resistance (*ARHGAP39*, *ARPC4*, *PHC2* and *BCL2L15*) and hair follicle development (*FOXN1*). A total of eight SNPs associated with growth traits (*ADGRB3*, *TSPAN* and *ZFAND3*) were detected. *PTPN3* gene associated with immune response was observed between indigenous and Wild Boar. Wild Boar and Duroc comparison resulted in genes associated with adaptation (*HDAC1* and *GNAI3*).

3.5 Discussion

The Porcine SNP60K bead chip was developed in 2009 (Ramos *et al*., 2009) and has been used to analyse genetic diversity and population structure in several pig populations (Ai *et al*., 2013; Burgos-Paz *et al*., 2013; Yang *et al*., 2017; Mujibi *et al*., 2018). This is the first report using the Porcine SNP60K bead chip to explore diversity of domestic and wild pig populations covering the commercial, village, wild and conserved pigs farmed and reared in Africa. Pigs are possibly known to have reached Sub- Saharan Africa through the Nile corridor and later dispersed to the West-Central Africa (Blench & MacDonald, 2000). There are 541 pig breeds worldwide (Rischkowsky & Pilling, 2007) but the dominating commercial breeds in the pork industry are the Large White, Landrace, Duroc, Hampshire, Berkshire and Piétrain (Rothschild & Ruvinsky, 1998). The source of the improved breeds found in Southern Africa is believed to be the European settlers in 1600s (Krige, 1950; Blench & MacDonald, 2000; Swart *et al*., 2010). This was when Jan van Riebeeck brought some pigs to the Cape of Good Hope (Naude & Visser, 1994). The Large White, South African Landrace and the Duroc are the breeds mostly found and used in the commercial sector while the Kolbroek and Windsnyer are considered as indigenous and are mostly found in rural areas (Kem, 1993; Ramsay *et al*., 2000). The Vietnamese Potbelly, Bush Pigs and Wild Boar populations constitute a small component of the genetic pool of pigs in the country often restricted to the game reserves and zoos.

The Porcine SNP60K bead chip was designed using genomic resources from Western pig genomes (Ramos *et al*., 2009) and hence the number of SNPs after QC for the commercial population was higher. The village populations had a higher number of polymorphic SNPs and moderate-high MAF compared to that of commercial pigs. Non-descript livestock populations including pigs are often observed to be highly diverse probably due to open mating systems and gene flow between populations. In South Africa similar observations of highly diverse and polymorphic populations were observed in village chicken populations (Khanyile *et al*., 2015), cattle (Makina *et al*., 2014), and village goats (Mdladla *et al*., 2016). The Warthog and other

indigenous pigs were observed to be the least polymorphic and diverse which could be attributed to ascertainment bias as the Kolbroek, Windsnyer, Vietnamese Potbelly, Warthog and Bush Pigs were not used in the development of the Porcine SNP60K bead chip. Overall, the Porcine SNP panel showed moderate MAF for the village, commercial and indigenous purebred pig populations such as the Windsnyer implying utility of the chip in the prevalent farmed pig populations of South Africa.

A study conducted by Swart *et al.* (2010) using microsatellite markers in various Southern African pig breeds revealed higher levels of diversity within population than was observed in this study for the same breeds. High heterozygosity levels (0.61-0.75) were also reported by Halimani *et al.* (2012). In contrast to Swart *et al.* (2010) the Large White had the lowest diversity (*Ho* = 0.358) compared to the South African Landrace (*Ho* = 0.372) and other breeds of the Duroc and Kolbroek. It must be noted that these previous studies used microsatellite markers that are highly polymorphic markers and cannot be compared to SNPs that are biallelic in nature. High gene diversity is therefore expected in microsatellites markers. However, results on genetic diversity from this study were comparable to other studies that used the Porcine SNP60K bead chip in Chinese and Western pig populations (Ai *et al.*, 2013).

The heterozygosity values for the indigenous pigs were relatively similar to those of the commercial pigs. A lower diversity was expected for the commercial pigs as they are under selection while the indigenous pigs are known to be rich reservoirs of distinct alleles, coupled with presence of gene flow (Amills *et al.*, 2012). However, the indigenous pig populations are also of very small flock sizes and often fragmented and restricted to specific farming communities and conservation units hence diversity was low. Small and fragmented populations and the possibility of natural selection due to disease and unfavourable climatic conditions could explain the genetic diversity observed in the village populations. The high inbreeding levels observed in the Warthog populations might have been promoted by its family structuring where pigs are organized into fragmented breeding and social units. Somers *et al.* (1995) noted that a group of Warthogs consist of about 40% of adults with changes seasonally. The number of mature individuals is estimated to be between 2000 and 5000 in the Kruger National Park (Ferreira *et al.*, 2013). The geographical separation of the three national parks from which the warthogs were sampled, could have created small and fragmented subpopulations leading to escalated F_{IS} values due to Wahlund effect. As expected, we found that the village pig populations of South Africa had high inbreeding values compared with other

populations. The negative F_{IS} values for commercial and indigenous populations are reflective of their intensive production environment as individuals are outbred to avoid mating to close relatives.

The low levels of effective population size (N_e) in the recent 12-22 generations for both commercial and indigenous populations are of concern. More so in the indigenous breeds since low levels of genetic diversity are likely to diminish overtime and increase the risk of extinction. The effective population of the Kolbroek of 34 at 12 generations ago is even lower than the minimum threshold N_e of 50 set by the FAO (2000). Franklin (1980) recommended a N_e of at least more than 500 while Willi *et al*. (2006) suggested N_e of more than 1,000 to maintain the evolutionary potential of any population. The genetic diversity of these populations will likely continue to be negatively impacted by the small number of founders and them being farmed in fragmented populations. Small effective population size of the Kolbroek might be due to pigs being raised in a research facility with limited boars and sows. Large White, Duroc and South African Landrace are commercial pigs that have undergone strong selection for meat and carcass traits thus resulting in small effective population sizes. Long-term sustainability of the populations might be compromised due to the small population size as it increases the effects of genetic drift and reduction in fitness traits (Frankham *et al*., 1995).

The high F_{IS} values observed within populations across breeds are similar to previous studies (SanCristobal *et al*., 2006; Swart *et al*., 2010; Gama *et al*., 2013; Edea *et al*., 2014). An overall AMOVA F_{IS} value of 93.95% was comparable to Halimani *et al*. (2012) value of 92.90% in indigenous pigs of Southern Africa. Diversity amongst South African populations that ranged from F_{CT} = 0.92 (village pigs) to F_{CT} = 5.42 (Commercial populations) might be due to gene flow between different populations within a sub-population. Moderate diversity within population (*i.e*., F_{IS} ranging from 19.92 in the category consisting of South African Wild Boar and worldwide Wild Boar to F_{IS} = 35.52 in the categories consisting of South African villages and Worldwide villages) relative to elevated F_{CT} in the same categories implies a higher genetic variation distributed among groups from different geographic locations. This genetic variation observed amongst groups of the South African and Burgos- Paz *et al*. (2013) pig populations (*i.e*., F_{CT} = 62.35–73.58) is higher than the variation reported amongst Angora goats from South Africa, France and Argentina using 50K SNP bead chip (Visser *et al*., 2016), which could be explained by limited exchange of breeding animals across geographic boundaries in the studied pig populations. The amongst population within groups diversity values ranging from F_{CT} =

0.46 for South African villages to F_{SC} = 18.17 for South African commercial demonstrates evidence of population sub-structure and genetic differentiation between the well-defined commercial and indigenous breeds relative to non-descript village populations that are characterized by weak population boundaries.

The PCA demonstrates the impact of domestication and geographic history on the clustering of populations. European populations as represented by Wild Boar, South African Landrace, and Large White, clustered together as expected. Considering the history that the Wild Boar is an ancestor to the domestic pigs of today, some gene flow may have remained from the Wild Boar in the domestic pigs (Giuffra *et al.*, 2000). The clustering of the Wild Boars reflects a European ancestry of those populations within that cluster. The slight difference between the Wild Boar and domestic populations might have been due to geographic isolation and artificial selection. Geographic structures were evident amongst most of the pig populations that were aligned to production systems and their founder effects. The clustering of the Windsnyer and the village populations could be due to gene flow between indigenous breeds and village populations. Limpopo populations had a closer proximity to Large White and South African Landrace, and farmers in this region are more likely to buy pigs from commercial herds. The Large White and South African Landrace are also closer together as these are both European breeds. It was interesting that generally the village populations were closer to the Windsnyer and Kolbroek as these are both indigenous breeds in South Africa. Although not much is known about our indigenous breeds, different theories suggest that the Kolbroek might have far Eastern alleles while the Windsnyer is known to be dominant in other parts of Southern Africa like Mozambique, Zambia and Zimbabwe (Holness, 1973, 1991). The village populations and other Large Whites and Landraces from the global data set clustered together with the South African village, commercial and indigenous pigs demonstrating genetic similarities that could be aligned to founder effects and similarities in production systems.

The clustering of Duroc away from other commercial populations (Large White and South African Landrace) was expected. The Duroc breed was created in the United States with pigs of several ancestries, including African pigs (Porter, 1993). Studies conducted by Visser and Kotze (1996) and Swart *et al*. (2010) using the microsatellite markers on the Large White, South African Landrace and Duroc also reported similar results. The Large White and South African Landrace were more genetically similar when compared to the Duroc. The inclusion of global populations did not alter this clustering.

The distance of Vietnamese Potbelly population from the rest of the domestic pigs is clear evidence of independent domestication that took place between the European and Asian subspecies of the Wild Boar (Giuffra *et al.*, 2000). The PCA including pigs genotyped from all over the world clearly shows the geographical effect of the populations as the Vietnamese Potbelly clustered in close proximity to the Chinese population.

ADMIXTURE $K = 2$ presented the first level of ancestry of the Suidae family representing *Phacochoerus africanus* (Warthog) and *Potamochoerus larvatus* (Bush Pig) versus *Sus scrofa* (domesticated pigs including the Wild Boar) species. The presence of the Wild Boar genomic signature in the domestic pigs from $K = 2$ to $K = 7$ is not surprising. It is well documented that the domestic pigs diverged from each other and originated from the ancestral Wild Boars around 8,000-10,000 years ago (Giuffra *et al.*, 2000; Laval *et al.*, 2000; Larson *et al.*, 2005). The Asian and European ancestral Wild Boars also originated from different subspecies thus the Vietnamese Potbelly diverged early ($K = 2$) from the rest of the domestic pig population. The results for the village populations showed high levels of admixture and weak between population sub- structuring. As opposed to pigs from the commercial sector that practices the intensive production systems, pigs in the villages are farmed under semi-intensive of free-range production systems, which might explain the admixture observed in this study. There is considerable indiscriminate crossbreeding that is taking place in village populations (Rege & Gibson, 2003). European and Asian pigs were used to improve the South African pig breeds, but the actual contribution is unknown. Although phenotypically distinct from each other, the Bush Pigs and warthogs clustered together which is suggestive of either common founder effect or selection pressures in the natural environments.

According to Wright (1978), F_{ST} estimation with values of less than 0.05 represents low differentiation while values between 0.05 and 0.15 represent a moderate genetic differentiation and those between 0.15 and 0.25 and beyond reflect highly differentiated populations. The low levels of genetic differentiation of the village populations from this study is consistent to pairwise F_{ST} values of Halimani *et al.* (2012) of village populations from Zimbabwe and South Africa. Most pig farmers from the villages practice free ranging or semi-controlled farming where there is continuous gene flow between populations within villages thereby explaining the low levels of population sub-structuring observed. Moderate F_{ST} values implies closer relationship between the South African Landrace and Large White and agrees with their

breeding history, whereby the Landrace was developed from crossing the Large White from England and a Denmark indigenous. Greater genetic differentiation between the Warthog and the other pig populations (F_{ST} = 0.36–0.53) might be attributed to the (i) pressures of natural selection (ii) the separate histories of domestic and wild populations and (iii) the unique population dynamics of Warthogs that are known to live in clans of adult females, males and their offspring while maintaining minimal contacts with other clans (Cumming, 1975; Somers *et al*., 1994). In South Africa, Warthog populations are restricted to nature reserves thus creating a physical barrier and huge genetic differentiation between them and other pig populations. This will be in contrast to the greater interaction between village, commercial and indigenous populations. Low F_{ST} values between the villages in South African and village populations from South America from Burgos-Paz *et al*. (2013) study, might be an indication that either common founder populations or similarities in production systems leading to common selection pressures. Ramírez *et al*. (2009) demonstrated that the African and South American pigs were derived from Europe and Far Eastern pigs. The very high genetic differentiation between the Vietnamese Potbelly and Bush Pig agrees with the PCA and Admixture clustering.

Per marker pairwise F_{ST} were estimated between pairs highly differentiated populations which were from villages, commercial, indigenous, Asian and wild populations. From the pairwise F_{ST}, Warthog was found to be genetically different from the rest of the populations. The per marker pairwise F_{ST} analysis used a threshold of 0.8 and above to plot Manhattan graphs of the Warthog against the rest of the populations. From the SNPs showing a threshold of $F_{ST} \geq 0.8$, we looked at candidate genes and QTLs that can be associated with those SNPs to infer on traits that might have genetically differentiated the Warthog from Alfred Nzo, Duroc, Kolbroek, Large White, South African Landrace and Windsnyer populations.

Majority of the SNPs that were above the threshold between the Warthog and the rest of the populations were from chromosomes 1, 4, 5, 12, 13 and 15. Chromosomes 2 (Warthog vs. Alfred Nzo), 3 (Warthog vs. Kolbroek), 6 (Warthog vs. South African Landrace) and 14 (Warthog vs. Large White) seemed to be less common. Chromosome 1 with a total number of 12 SNPs was associated with reproduction and growth traits while the indigenous populations of Kolbroek and Windsnyer were differentiated on chromosome 4 that was also linked to reproduction and growth traits.

Warthog vs. Alfred Nzo had three SNPS ($F_{ST} \geq 0.8$) that are associated with reproduction (*RPL18, IL17B)* and growth (*IL17B, ARHGAP23*) characteristics. It is known that good nutrition is vital to be able to maximize growth performance. Genes *IL17B* and *ARHGAP23* are linked to inflammatory response (Liu, 2015; Bie *et al.*, 2017) and the gastrointestinal tract where they play a role in the digestion and absorption of the nutrients. Inflammatory responses lead to reduction of feed intake, which in turn affects the growth of the animal (Liu, 2015). Selection on genes associated with inflammation in the populations of Warthog vs. Alfred Nzo might be an effect of the different diets these populations scavenge on. Medzhitov (2008) noted the inflammation response to be a protective mechanism from the stress and harmful environment.

Growth linked genes *ADGRB3*, and *ACY1* were dominant in differentiating Warthog vs. Duroc populations with an overall total of 10 SNPs. Emrani *et al.* (2017) associated *ADGRB3* to body weight traits in the broiler chickens. The association of *ADGRB3* gene to Duroc rather than Large White or South African Landrace breeds might be linked to the higher percentage of IMF in Duroc compared to the other two commercial breeds (De Vries *et al.*, 2000). Mature males of Warthog can also reach up to 100 kg and possesses good meat and carcass qualities (Hoffman & Sales, 2007).

A total number of 20 significant SNPs ($F_{ST} \geq 0.8$) were linked to the Warthog vs. Kolbroek populations. Growth traits were associated with five of the SNPs between Warthog vs. Kolbroek. Indigenous Kolbroek are reported to be smaller in size when compared to commercial breeds such as Large White (Chimonyo *et al.*, 2005). Kutwana *et al.* (2015) reported no significant difference ($P > 0.05$) between the Kolbroek and Large White populations that had higher fat percentages when compared to the other commercial breeds (Nicholas, 1999).

Chromosome 13 was also highly notable with significant SNPs differentiating Warthog vs. Kolbroek and Warthog vs. Windsnyer. Only two SNPs appeared for Warthog vs. South African Landrace and were on chromosome 6. The Warthog vs. Windsnyer had a total of fourteen SNPs differentiating them. The identification of *BRPF1* gene in the Warthog vs. Windsnyer populations is an important observation as this gene is associated with the intramuscular fat (IMF). When it comes to the value and taste of the pork meat, IMF is an important characteristic because meat that is high in IMF tends to be juicy and tender (Eikelenboom *et al.*, 1996; de

Koning *et al*., 1999). The gene *ATPB2* associated with six significant SNPs is linked to heat stress and reproductive performance (Dash *et al*., 2016). Heat stress might result in poor reproduction for both sows and boars. Pigs cannot sweat and this makes them sensitive to high environmental temperatures making and of concern particularly to commercial pig farmers (Ross *et al*., 2015).

Genes linked to immune response and mastitis were observed in Indigenous vs. Duroc comparisons. *PTPN22* gene on chromosome 4 has a regulatory effect on T- and B- cell activation in immune response (Lamsyah *et al*., 2009). *PTPN22* plays a role in susceptibility to tuberculosis. Pigs are generally natural hosts of mycobacterial infections (de Lisle, 1994). Porcine TB has been reported in South Africa where infections are commonly via infected cattle fecal matter fed to piglets as well as interactions with wild pigs (Muwonge *et al*., 2012). *NXPH1* gene is associated with DMI (dry matter intake) in cattle (Olivieri *et al*., 2016). Both P*TPN* and *NXPH1* genes were fixed in the Duroc implying natural selection of the Duroc when compared to both indigenous and Wild Boars. Breeds in the commercial sector are mainly selected for growth, carcass and meat quality traits. The indigenous and village population on the other hand has not been systematically selected for such traits.

The *NPY5R* located on chromosome 8, was associated with feed efficiency and fat deposition. This gene was also reported in Jinhua and Rongchang pigs that belong to Chinese breeds (Chen *et al*., 2018). Fat deposition genes observed in Indigenous vs. Vietnamese Potbelly, Villages vs. Kolbroek and South African Landrace with Large White vs. Indigenous are evidence in agreement with suggestions that Kolbroek and other indigenous pigs tend to carry their weight in their bellies and backs (Hoffman *et al*., 2005). Hoffman *et al*. (2005) also reported breed type and diet to have an influence on the composition of the meat. This study therefore presented a diverse genomic architecture of South African pigs with differentiating selection pressures for meat and carcass quality traits in the different pigs raised in diverse production systems.

3.6 Conclusion

Overall, the study demonstrated the utility of the Porcine SNP60K bead chip in elucidating genetic diversity and population genomic structure of South African pig populations relative to other global populations. Village pigs demonstrated distinctiveness from other domestic and commercial populations within South Africa and when compared to global populations. The

study provided baseline knowledge with regards to the genetic diversity of the domestic and wild pig populations of South Africa, which is a prerequisite for population/breed characterization, utilization and conservation. A more in-depth analysis of patterns of genetic variations is required to get more insight into factors shaping genetic diversity of these populations.

CHAPTER 4: THE EXTENT OF LINKAGE DISEQUILIBRIUM AND HAPLOTYPE BLOCK STRUCTURE OF COMMERCIAL, INDIGENOUS AND VILLAGE PIGS

4.1 Abstract

Genomic tools for selection and breeding provides opportunities towards conservation and improvement of genetic resources. The aim of this study was to describe the genome-wide extent of linkage disequilibrium (LD) and haplotype block structure in South African commercial, indigenous and village pigs. For this purpose, the Porcine SNP60K array (62 134 SNPs) genotypic data from a total of 191 pigs representing indigenous (n = 40), commercial (n = 60) and village pigs from Limpopo and Eastern Cape provinces (n = 91) was used. The overall mean LD for village pigs was lower (0.19±0.24) compared to the means of the commercial (0.32±0.33) and indigenous (0.34±0.33). Extent of LD differed amongst chromosomes with stronger LD on chromosomes under selection harbouring QTLs. LD decay was the highest for the commercial Duroc ranging from 0.73±0.75 at 0-1 kb to 0.16 at 4000-5000 kb whereas Mopani village populations had the lowest LD at 0.55±0.53 at 0-1 kb to 0.09±0.06 at 3000-4000kb. The persistence of phase between the commercial and indigenous pairs was highly correlated with a Pearson correlation co-efficient of 0.74±0.73 at distance of 0-1 kb. On average, more haplotype blocks were observed in indigenous (4 766) and commercial (3 075) populations whereas the village pig populations shared the least number (2 145) of haplotype blocks. The indigenous Kolbroek population also had the greatest number of haplotype blocks, while the Kolbroek and Duroc populations showed the highest levels of haplotype diversity. Analysis of the shared and unique haplotype blocks also allowed the identification of chromosomal regions under selection. The villages shared more haplotype blocks compared to the other populations in this study. The *DECR1* gene related to meat quality and growth rate traits was detected within the haplotype blocks shared by all the populations, reflecting on the relevance of production traits in the local pig populations. Other observed genes were associated with immune response (*LENG9*), fertility (*TARS*) and milk production (*ADAMTS12*). Overall, results of this study revealed similarities and differences between breeds and production systems, thereby increasing our understanding of the extent of LD across South African pig populations, and identification of genomic regions harbouring shared and unique haplotype blocks that could find application in the development of genomic tools for pig studies.

Keywords: *Pig populations, production system, LD decay, persistence of phase, high-density SNP genotypes, unique and shared blocks*

4.2 Introduction

Efforts have been made in genomics characterisation, selection and improvement of livestock species including pigs (Amills *et al*., 2020). Population studies using whole genome Single Nucleotides Polymorphism (SNPs) have been used to enhance pig breeding, as it provides opportunities towards conservation and improvement of genetic resources (Amaral *et al*., 2011). The level and extend of linkage disequilibrium (LD) provide understanding into the past and evolution of breeds or populations and affects the utility of available SNP panels (Slatkin, 2008). LD can assist to understand the history of the population and breed development (Slatkin, 2008). Factors such as small effective population size, selection, migration and more genetic events are known to affect LD (Lander & Schork, 1994).

Whole genome SNP data have been used to haplotype pigs as a beneficial framework tool for animal genomic selection programs (Gabriel *et al*., 2002; Andreia *et al*., 2008; Uimari & Taio, 2011; Veroneze *et al*., 2013; Grossi *et al*, 2017; Jasielczuk *et al*., 2020). Haplotypes are the result of inbreeding and are derived from one ancestry to the next generation as components called haplotype blocks (Gabriel *et al*., 2002; Amaral *et al*., 2008). Haplotype blocks are segments of the genome that display low recombination rates, strong LD and inadequate haplotype diversity (Luikart *et al*., 2003; Phillips *et al*., 2003). Haplotype blocks are influenced by population bottlenecks, mutations, population admixture, chromosomal recombination and selection and have been used in animal breeding (Guryev *et al*., 2006).

The popularity and reduced cost of Single Nucleotides Polymorphism (SNPs) genotyping platforms have made it possible to exploit mapping linkage for various traits in pig breeds (Yang *et al*., 2017; Muñoz *et al*., 2019; Lee *et al*., 2020) and other livestock species (O'Brien *et al*., 2014; Khanyile *et al*., 2015; Makina *et al*., 2015; Mdladla *et al*., 2016). The identification of quantitative trait loci (QTL) associated with important production traits can aid to increase the accuracy in choosing genetically superior livestock to meet the growing demand of pork products (Archibald *et al*., 1995). Haplotype blocks have been used to find significant variants in genome-wide association studies (GWAS), as well as in predicting the genomic breeding values of animals in genomic selection schemes (Meuwissen *et al*., 2001; Calus *et al*., 2008;

Cuyabano *et al*., 2015; Chen *et al*., 2018). In contrast to traditional quantitative trait loci (QTL), GWAS that depends on markers close to the contributing loci being strong in LD have a much higher resolution and makes it possible to identify genes associated with different traits in a single population (Bush & Moore, 2012; Otyama *et al*., 2019). When two alleles occur together on the same haplotype more than expected, this can lead to positive LD (Barnes, 2007; Calabrese, 2019). This is because correlation exist across genetic variation of different loci thus one marker can provide information of the second genetic marker.

The Porcine SNP60K is an important genomic tool with a genome-wide coverage of SNPs to assist in revealing haplotypes that may be linked with selected alleles and improve genomic predictions (Veroneze *et al*., 2013; Grossi *et al*, 2017; Jasielczuk *et al*., 2020). Previous studies have demonstrated the extent of LD in indigenous and village populations are comparable to that of cattle and larger compared to human populations (Uimari & Taio, 2011; Ai *et al*., 2013; Khanyile *et al*., 2015; Makina *et al*., 2015; Mdladla *et al*., 2016). The high LD in commercial pigs was attributed to small effective population sizes and strong selection for traits of economic importance (Huang & Ren 2013). Whilst some data is available on the extent of LD in commercial pig populations, there are very few studies conducted on indigenous and village-based pigs. Although haplotype blocks and LD has been widely studied in pigs (Uimari & Taio, 2011; Veroneze *et al*., 2013; Grossi *et al*, 2017; Jasielczuk *et al*., 2020), very little is known for South African pig populations.

In South Africa, pigs are raised under different productions systems namely, commercial intensive system, small-scale farmers, as well as free backyard management systems characterized by lack of proper housing (Matabane *et al*., 2015; Mokoele *et al*., 2014; Molotsi *et al*., 2021). The commercial farming system is associated with well characterised international breeds like the Large White, Duroc and Landrace, good management practices and set breeding strategies. The small scale and backyard production systems that utilize non-descript, indigenous and village pigs are characterized by low to medium production inputs (Madzimure *et al*., 2012; Phogole, 2017). However, the indigenous and non-descript village pigs in South Africa are considered well adapted to harsh climate conditions and diseases (Madzimure *et al*., 2012) and therefore of importance to low-input production systems and for coping with the effects of environmental changes. In Chapter 3, genetic similarities were demonstrated between the village populations, Wild Boars, SA indigenous and the Large White and Landrace pigs that

clustered together separated from the Durocs, Chinese and Vietnamese Potbelly pigs as well as the Warthog and Bush Pig. Sub structuring was noted between SA commercial populations as well as between indigenous and commercial breeds (Chapter 3).

The aims of this study were to assess the extent of genome-wide LD and haplotype diversity haploblock structure (i) across several South African pig populations and (ii) between South African pig populations with different genetic and production backgrounds. This study further analysed the South African pig populations by focusing on the village, indigenous and commercial pigs and investigating the genome-wide linkage disequilibrium, haploblock structure and persistence of gamete phase to avail information that can be used to guide future genomics assisted breeding and improvement of local pigs. The analysis of LD structure would shed information on the (i) utility of the SNP60K panel to support mapping of causal mutations through GWAS or QTL mapping and (ii) the genetic relationship between populations and its implications on across breed/population selection programs. Furthermore, by screening and analysing haplotypes, this study sets to provide information that will assist in dissecting the origin and ancestral relationship of South African pigs, thereby providing insights on the genomic organization of South African pig populations for comparative studies with other populations such as European and Asian pigs.

4.3 Materials and methods

4.3.1 Genotyping and quality control

A total of 191 pigs were genotyped with the Illumina Porcine SNP60K (llumina Inc., San Diego, United States). The genotyped data represented pigs from commercial (n = 60) and indigenous (n = 40) populations and villages (n = 91). A full description of the geographic locations of these populations is provided in the materials and methods section of Chapter 3. The Large White (LWT, n = 20), South African Landrace (SAL, n = 20) and the Duroc (DUR, n = 20) represented commercial breeds, while the Kolbroek (KOL, n = 20) and Windsnyer (WIN, n = 20) represented indigenous breeds. The villages were represented by Limpopo province sampling in Capricorn (MOP, n = 25) and Mopani (MOP, n = 27) districts and Eastern Cape province sampling in Alfred Nzo (ALN, n = 17) and O.R. Tambo (ORT, n = 22) districts.

SNP data QC was done using Golden Helix SNP Variation Suite (SVS) *version* 8.5 (Golden Helix Inc., 2016) to remove within a population, individual animals with a call rate less than 85%, SNPs with greater than 10% missing genotypes, minor allele frequencies (MAF) less than 0.02, and with significant deviation from Hardy-Weinberg Equilibrium ($P < 0.0001$). The same quality control parameters as above was used for across populations' analysis. SNPs that were not assigned to chromosomes and those on sex chromosomes of the pig reference genome (*Sus scrofa* 11.2) were excluded from further downstream analysis.

4.3.2 Linkage disequilibrium (LD) analysis

The correlation coefficient (r^2) was used to calculate pairwise LD within each population using PLINK 1.07 (Purcell *et al*., 2007). Linkage disequilibrium defined by PLINK as '--r2 --ld-window-kb 4000 --ld-window-r2 0' was used for estimating LD for each population. Extent of LD was calculated for each population using mean and standard deviation of r^2 for chromosome 1 to 18. Inter-SNP distances (kb) were classified using R Studio 3.6.0 (Team, 2020) into the following bins, 0-1; 1-10; 10-20; 20-40; 40-60; 60-100; 100-200; 200-500; 500-1000; 1000-2000; 2000-3000; 3000-4000 kb to calculate LD decay. The average pairwise LD (r^2) was used for each bin of inter-SNP distance. The following equation was used to predict the extent of decline of the LD between the populations Sved (1971), Heifetz *et al*. (2005) and Amaral *et al*. (2008):

$$LD_{ijk} = \frac{1}{1 + 4\beta_{jk}\, d_{ijk}} + e_{ijk}\,,$$

where LD_{ijk} is the observed LD for marker pair i of population j in genomic region k, d_{ijk} is the distance in base pairs for marker pair i of population j in genomic region k, β_{jk} is the coefficient that describes the decline of LD with distance for population j in genomic region k and e_{ijk} is the random residual.

The effect of population, chromosome, SNP interval and the interaction between population and chromosome on LD was investigated using GLM Procedure in Statistical Analysis System (SAS, 2013) using the equation: $LD_{ijk} = \mu + d_i + \text{population}_j \times BTA_k + e_{ijk}$ where LD_{ijk} (Sved, 1971; Heifetz *et al*., 2005) was the observed LD over population mean μ; marker distance d_i for-marker pair *i;* of population *j* and on *BTA* chromosome *k*.

4.3.3 Persistence phase

For these analyses, only 24 678 SNPs common across the nine populations were used for each marker pair with a measure of r^2. The aim of this analyses was to measure the genetic relationship between population pairs. The signed *r* value was calculated by taking the square root of the *r* value which is not affected by the sample size and minor allele frequency unlike D' and allocating the correct sign established from the calculated *D*' value (Sargolzaei *et al*., 2008). PLINK 1.9 (Chang *et al*., 2015) was used to calculate the r^2 using '--r2 --ld-window-kb 5000 --ld-window-r2 0' for population pair marker distance from 0-5000 kb. The r^2 values of population 1 and population 2 were then categorised into the following bins: 0-1; 1-10; 10-20; 20-40; 40-60; 60-100; 100-200; 200-500; 500-1000; 1000-2000; 2000-3000; 3000-4000; 4000-5000 kb using R Studio *version* 3.6.0 (Team, 2020).

4.3.4 Haplotype block partitioning and diversity

Genotypes were phased per chromosome using BEAGLE 5.1 (Browning & Browning, 2007) using 100 iterations parameters. After phasing, PLINK 1.9 (Chang *et al*., 2015) was used to perform haplotype blocks analysis using --hap plink.blocks function for the 18 pairs of autosomal chromosomes for SNPs within 200 kb. Haplotype blocks were estimated using Expectation Maximization (EM) algorithm approach (Qin *et al*., 2002). Haplotypes for this study was reflected as the number of haplotypes found within a block. Crawford and Nickerson (2005) defined a haplotype block as two or more closely linked alleles that segregated together. Haplotypes are inherited as a single units called haplotype blocks from one generation to the next (Gabriel *et al*., 2002). The frequency of haplotype blocks was then estimated using the --hap plink.blocks --hap-freq function in order to determine the number of times blocks occurred. Blocks were grouped according to frequencies of < 0.1, 0.1-0.25, 0.25-0.5, 0.5-0.75, 0.75-1.0 were estimated for each population in RStudio.

4.3.5 Unique and shared haplotype blocks

Unique haplotype blocks were defined as the block regions targeted for each population while shared haplotype blocks were characterized as the overlapping block regions shared by two or more populations (Oyelami *et al*., 2020). Ouput file from --hap plink.blocks was used to screen

for haplotype blocks unique and shared between the populations using UpSetR (Conway *et al*., 2017) package in RStudio. Ggplot (Wickham, 2016) in RStudio was used to visualise the results. RStudio was further employed to sort the number of unique or shared haplotype blocks. To annotate the unique and shared haplotype blocks, Ensemble Biomart was used to mine for genes and Pig QTL database (https://www.animalgenome.org/cgi-bin/QTLdb/SS/index) linked to *Sus scrofa* 11.2 to identify QTLs.

4.3.6 Annotation of high LD regions

High LD SNPs ($r^2 > 0.8$), unique and shared haplotype blocks were used to identify QTLs and their traits ontology. The study hypothesised that important and complex traits could be harboured in regions with the highest LD and in regions of unique and shared haploblocks that have undergone natural or artificial selection. Ensemble Biomart and Pig QTL database was used to map the high LD regions for each chromosome (Ashburner *et al*., 2000). QTLs were based on exterior, health, meat and carcass quality, production and reproduction traits.

4.4 Results

4.4.1 Sample and SNP quality control

After applying the filtering criteria, a total number of 52 925 (MOP), 54 078 (CAP), 52 238 (ORT), 53 580 (ALN), 49 773 (LWT), 49 191 (SAL), 40 632 (DUR), 39 560 (KOL) and 47 402 (WIN) SNPs were retained for each population for LD and haplotype block analysis. After pruning, 183 (MOP 27; CAP 24; ORT 22; ALN 15; LWT 18; SAL 19; DUR 19; KOL 20; WIN 19) individuals were used for LD analysis. A total of 27 740 polymorphic SNPs common across the nine populations were retained for persistence of gamete phase analysis.

4.4.2 Linkage disequilibrium (LD)

Table 4.1 shows the average extent of linkage disequilibrium per and overall population across the 18 chromosomes for the nine South African pig breeds/populations. The mean LD ranged from 0.18±0.22 (MOP) to 0.40±0.36 (KOL). KOL and DUR had the highest overall mean LD at 0.40±0.36 and 0.38±0.36, respectively. Chromosome 14 had the highest LD for five of the populations (ALN, MOP, SAL, DUR, WIN) followed by chromosome 1 (MOP, LWT, KOL) whereas chromosome 18 (MOP), had the least at 0.13±0.19 in the village populations.

Table 4.1: Mean and standard deviation of r^2 for chromosome one to eighteen in South African pig breeds and populations

Chr	Size	Villages				Commercial			Indigenous		Overall
		ALN	ORT	MOP	CAP	LWT	SAL	DUR	KOL	WIN	Mean
Chr 1	315,321,322	0.24±0.27	0.22±0.25	0.22±0.25	0.22±0.25	0.36±0.34	0.32±0.33	0.41±0.38	0.48±0.41	0.32±0.33	**0.31±0.09**
Chr 2	162,569,375	0.24±0.26	0.21±0.25	0.19±0.23	0.19±0.23	0.31±0.31	0.34±0.33	0.41±0.37	0.40±0.37	0.28±0.29	**0.29±0.08**
Chr 3	44,787,322	0.22±0.25	0.18±0.22	0.18±0.22	0.24±0.23	0.27±0.30	0.30±0.31	0.34±0.35	0.42±0.37	0.28±0.30	**0.27±0.08**
Chr 4	143,465,943	0.24±0.27	0.19±0.24	0.19±0.22	0.19±0.23	0.31±0.32	0.30±0.31	0.37±0.36	0.35±0.33	0.28±0.30	**0.26±0.07**
Chr 5	111,506,441	0.22±0.25	0.19±0.23	0.18±0.22	0.18±0.22	0.28±0.30	0.26±0.29	0.35±0.35	0.38±0.35	0.27±0.29	**0.29±0.09**
Chr 6	157,765,593	0.22±0.25	0.20±0.25	0.19±0.23	0.20±0.25	0.32±0.32	0.34±0.34	0.43±0.39	0.37±0.36	0.30±0.32	**0.26±0.09**
Chr 7	134,764,511	0.21±0.25	0.20±0.24	0.18±0.21	0.18±0.22	0.28±0.30	0.28±0.31	0.39±0.37	0.37±0.36	0.27±0.29	**0.27±0.08**
Chr 8	148,491,826	0.23±0.27	0.20±0.25	0.19±0.23	0.20±0.22	0.29±0.31	0.28±0.31	0.37±0.35	0.40±0.36	0.30±0.31	**0.28±0.08**
Chr 9	153,670,197	0.22±0.26	0.28±0.22	0.17±0.22	0.18±0.22	0.30±0.32	0.29±0.30	0.37±0.36	0.41±0.36	0.27±0.29	**0.23±0.07**
Chr 10	79,102,373	0.19±0.23	0.16±0.21	0.15±0.20	0.15±0.20	0.26±0.29	0.25±0.29	0.32±0.32	0.35±0.35	0.25±0.27	**0.25±0.09**
Chr 11	87,690,581	0.20±0.24	0.16±0.22	0.16±0.22	0.17±0.21	0.25±0.29	0.30±0.32	0.34±0.35	0.42±0.38	0.27±0.29	**0.25±0.08**
Chr 12	63,588,571	0.21±0.26	0.18±0.23	0.16±0.20	0.17±0.24	0.25±0.27	0.25±0.28	0.35±0.35	0.41±0.36	0.25±0.28	**0.30±0.08**
Chr 13	218,635,234	0.25±0.29	0.22±0.25	0.21±0.25	0.21±0.24	0.34±0.33	0.32±0.33	0.41±0.37	0.40±0.37	0.31±0.31	**0.32±0.08**
Chr 14	153,851,969	0.27±0.29	0.23±0.27	0.22±0.25	0.23±0.26	0.33±0.33	0.36±0.34	0.47±0.38	0.40±0.36	0.34±0.34	**0.28±0.09**
Chr 15	157,681,621	0.24±0.26	0.20±0.24	0.19±0.23	0.19±0.23	0.31±0.33	0.28±0.30	0.41±0.37	0.43±0.38	0.30±0.31	**0.26±0.08**
Chr 16	86,898,991	0.21±0.25	0.18±0.22	0.17±0.21	0.17±0.21	0.27±0.29	0.30±0.30	0.35±0.35	0.39±0.36	0.29±0.30	**0.26±0.09**
Chr 17	69,701,581	0.20±0.24	0.18±0.23	0.16±0.21	0.17±0.21	0.28±0.29	0.29±0.30	0.37±0.35	0.42±0.37	0.27±0.29	**0.26±0.11**
Chr 18	61,220,071	0.21±0.24	0.15±0.20	0.13±0.19	0.15±0.20	0.24±0.28	0.27±0.30	0.37±0.34	0.44±0.36	0.30±0.31	**0.27±0.08**
Overall Mean		**0.22±0.26**	**0.20±0.23**	**0.18±0.22**	**0.19±0.23**	**0.29±0.31**	**0.30±0.31**	**0.38±0.36**	**0.40±0.36**	**0.29±0.30**	**0.27±0.01**

There was a significance difference ($P < 0.0001$) on the LD amongst populations, chromosome and SNP intervals as well as the interaction between chromosome X population (Table 4.2). Figure 4.1 shows the decline in r^2 for each population at increasing marker distances. DUR population had the highest average values of r^2 = 0.73±0.75 at 0-1 kb that decreased to r^2 = 0.29±0.28 at 3000-4000 kb. For the KOL population, LD declined from r^2 = 0.70±0.74 (0-1 kb) to r^2 = 0.58±0.55 (1-10 kb) but remained constant from 500-2000 kb (0.38±0.42-0.39±0.45) and then showed a sharp decrease until 4000 kb. Village populations MOP and CAP had the lowest LD of r^2 = 0.07±0.09 at 4000-5000 kb that declined to r^2 = 0.10±0.12 at 3000-4000 kb.

All the village populations displayed a similar LD decay pattern whereby LD decayed from r^2 = 0.56±0.59 at 0-1 kb to r^2 = 0.37±0.40 at 1-10 kb.

Table 4.2 The effects of population, chromosome and SNP marker interval and population X chromosome interaction on linkage disequilibrium (r^2)

Factor	DF	SS	MS	*F*-value	*P*-value
Population	8	8120.43	1015.05	14683.3	***
Chromosome	17	740.50	43.56	630.10	***
SNP Interval	136	700.22	5.15	74.48	***
Population X Chromosome	1	1314.16	1314.16	19010.1	***

DF = Degrees of Freedom; SS = Sums of Squares MS = Mean Square error $^{***}P < 0.0001$

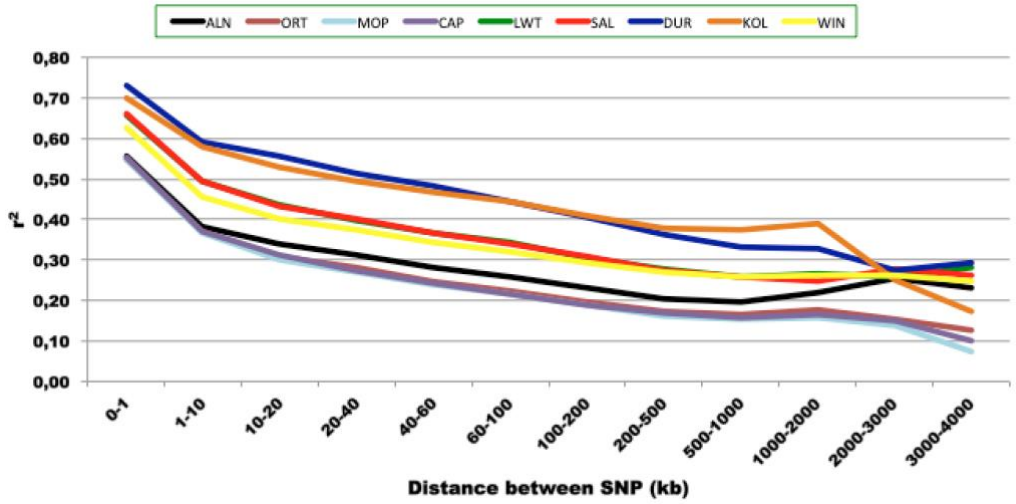
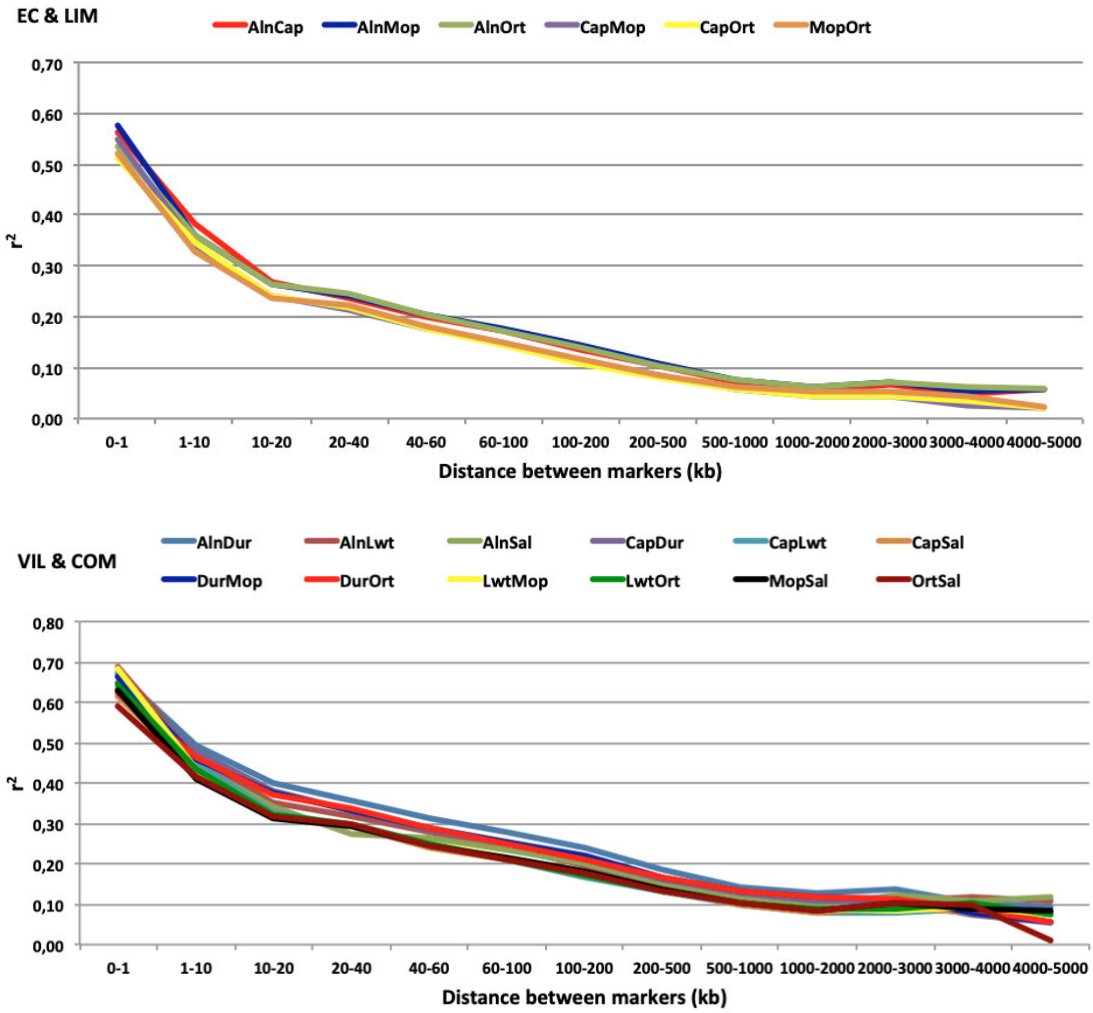


Figure 4.1 LD decay represent the average r^2 according to marker distances for the 18 chromosomes of the nine South African pig breeds/populations

4.4.3 Persistence phase

The persistence phase was estimated across the nine population pairs using Pearson correlation coefficient *(r^2) at* different marker distances (Figure 4.2). Correlations were higher at shorter marker distances for all the population pairs. DUR and LWT (r = 0.79±0.73) followed by KOL and LWT (r = 0.78±0.74) and DUR and KOL *(r* = 0.76±0.79) had the highest mean correlation at distances less than 1 kb compared to other populations. At the same marker distance bin, a smaller value ($r \leq 0.53$±0.59) was noticed between ORT and three other village pig populations

(CAP, MOP and ALN). There was a decrease in correlations at distances of 4000-5000 kb to *r* = 0.16±0.18 for CAP and MOP as well as CP and ORT. Persistence phase almost fell to zero amongst the villages as well as between the villages and commercial at higher marker distances. The persistence dropped dramatically > 5000 kb to *r* = 0.01±0.02 between the village and commercial type pigs of ORT and SAL respectively. Overall, commercial and indigenous populations had the highest correlation with the village populations. Trends between marker distance bins of 0-20 kb were similar for all the studied populations as the Pearson correlation decreased below *r* = 0.40±0.45.



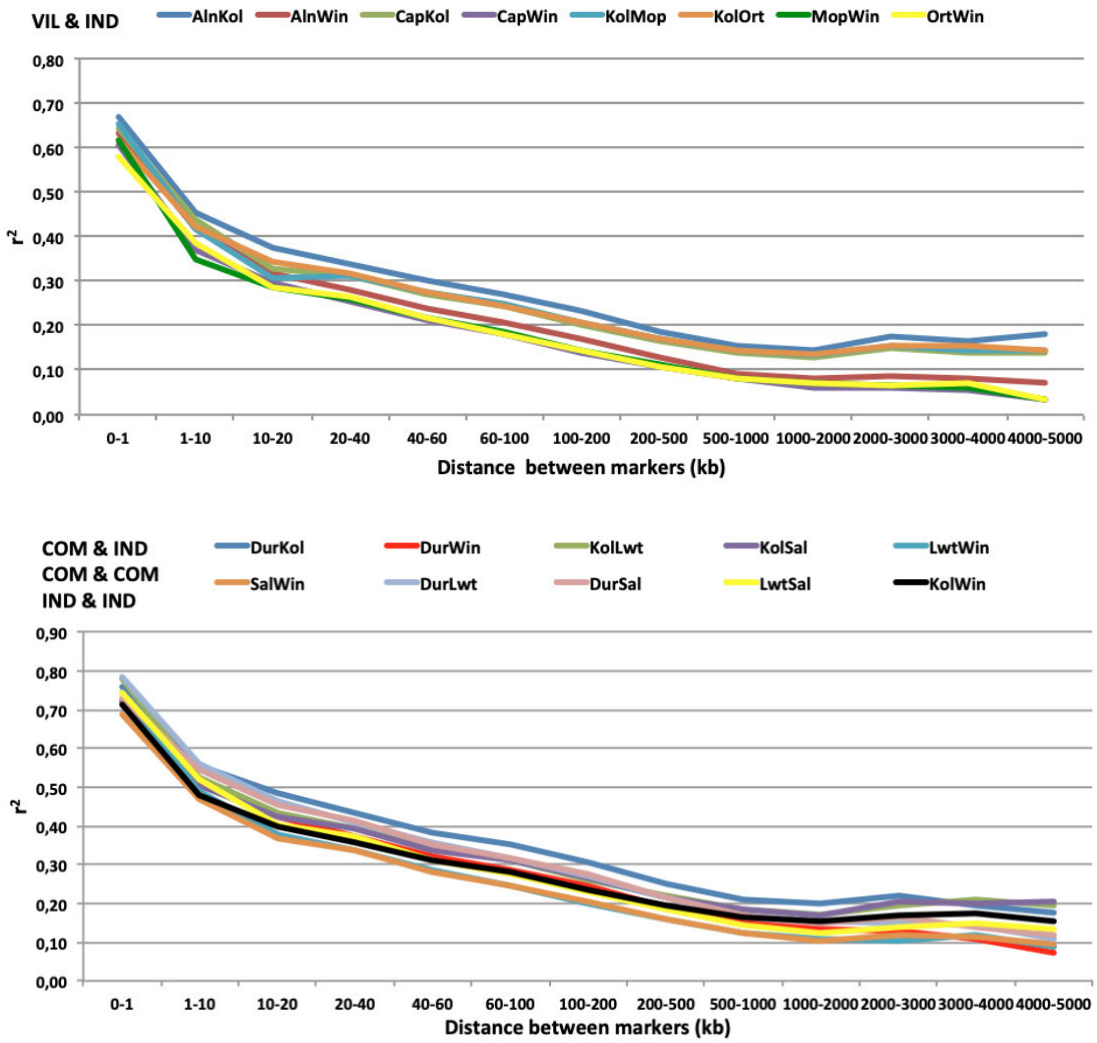


Figure 4.2 Pearson correlation between breed pairs (*n* = 36) at different physical distances. (a) Eastern Cape (EC) and Limpopo (LIM), b Villages (VIL) and Commercial (COM), (c) Villages and Indigenous (IND), (d) Commercial (COM) and Indigenous (IND)

4.4.4 LD haplotype block partitioning and haplotype diversity

In total, 23 969-haplotypes were observed across populations. DUR had the highest number of haplotype blocks at 3 402 whilst ALN had the least number (*n* = 1 350) (Table 4.3). Overall, village populations had the least number of haplotypes. Supplementary Figure S1 shows the number of haplotype blocks for each chromosome. Chromosome 1 had the most haplotype blocks (Supplementary Figure S1). From chromosome 1, MOP population had the highest haplotype blocks at 423. ALN had the lowest haplotype blocks for all the chromosomes.

Longest average block length 103.17±6.7Mb was reported in KOL pigs followed by DUR (102.08±9.67Mb) with the shortest block in CAP pigs at 63.14±10.84Mb (Table 4.3). DUR had the longest the average block length of 117.10±15.3Mb on chromosome 7 (Supplementary Figure S2). The shortest average block length was observed on chromosome 18 in MOP (37.27±9.67).

Table 4.3 Descriptive summary of haplotype block distribution for the nine populations used in this study

Haplotype block data description	ALN	ORT	MOP	CAP	LWT	SAL	DUR	KOL	WIN
No. of haplotype blocks	1 350	2 185	2 728	2 315	2 730	3 084	3 402	3 356	2 819
Av. block length (Mb)	78.48±8.15	67.56±10.25	63.27±11.23	63.14±10.84	85.82±12.06	88.14±10.53	102.08±9.67	103.17±6.71	82.02±7.14
% of the genome captured by the blocks	4.02	5.63	6.79	5.67	9.10	10.24	13.15	12.95	8.73
Total No. of SNPs in the blocks	4 403	6 595	8 127	7 003	9 502	10 726	12 285	11 741	9 121
Av. No. of SNPs per block	3	3	3	3	3	3	4	3	3
Sum of block length (Mb)	108±3.87	152±5.71	183±7.51	153±5.86	246±8.62	276±9.18	355±9.24	350±9.35	236±6.45

Haplotype blocks covered 13.15% of the total genome length of the DUR population while only 4.02% of the genome of ALN pigs was covered by haplotype blocks. Generally, the commercial and indigenous populations had the highest genome coverage followed by the village populations. Detailed information on haplotype blocks, average block length, average number of SNPs and sum of block length are presented in Supplementary Figures S2, S3 and S4. Overall, DUR had the highest total number of SNPs in the blocks at 12 285 SNPs. Average number of SNPs per block was three for all populations except for DUR (n = 3.94) at chromosome 14. DUR had the highest sum of block length, at 355 Mb while ALN had the least at 108 Mb. However, the KOL had the highest of block length with chromosome 1 at 45 Mb (Supplementary Figure S3). Chromosome 1 which is the largest autosome (315.32 Mb) had the longest number of haplotype blocks and highest sum of block length and average block length across all the studied populations. Most haplotype blocks were between 100-250 kb (Supplementary Figure S4).

Table 4.4 shows the blocks of different frequencies occurring the most at ranges from 0.25 to 0.50 (25-50%) for all the population except for MOP, with most frequent haplotypes occurred at < 0.1 (Table 4.4). For all the populations the least frequent haplotypes occurred at a frequency between 0.75 and 1.0. Haplotypes H901, H1430, H1605 and H1969 occurred at the highest haplotype frequency at 0.907 in MOP pigs. The most frequent haplotypes for the other eight populations are represented in [Supplementary Table 4.1].

Table 4.4 Haplotype frequencies and percentages for all haplotypes in each window in all the pig populations

Frequency (%)	ALN	ORT	MOP	CAP	LWT	SAL	DUR	KOL	WIN
< 0.1	273 (20.22)	564 (25.79)	850 (31.15)	665 (28.72)	570 (20.88)	690 (22.36)	489 (14.35)	470 (13.99)	394 (13.99)
0.1-0.25	162 (12.02)	259 (11.87)	423 (15.51)	369 (15.95)	326 (11.94)	348 (11.28)	509 (14.96)	455 (13.55)	419 (14.88)
0.25-0.50	522 (38.63)	727 (33.26)	779 (28.55)	671 (28.99)	992 (36.35)	1 057 (34.29)	1 222 (35.93)	1 244 (37.06)	1 057 (37.50)
0.50-0.75	365 (27.05)	539 (24.67)	561 (20.55)	521(22.48)	765 (28.06)	865 (28.06)	1008 (29.67)	1 036 (30.87)	844 (29.93)
0.75-1.0	28 (2.08)	96 (4.41)	115 (4.24)	89 (3.86)	77 (2.77)	124 (4.01)	174 (5.09)	151 (4.53)	105 (3.70)
No. of haplotype blocks	**1 350**	**2 185**	**2 728**	**2 315**	**2 730**	**3 084**	**3 402**	**3 356**	**2 819**

4.4.5 Analysis and annotation of unique and shared haplotype blocks

Commercial and indigenous populations had the highest number of unique haplotype blocks with DUR pigs being highest at 2 513 (Figure 4.3, Supplementary Table 4.2). Overall, thirty-eight (38%) percentage of unique haplotype blocks were detected. Most of the unique haplotype blocks were characterised by 2 and 3 SNPs and observed in the ORT and ALN population (Figure 4.4). A total of 2 477 genes (562 QTLs) represented the highest number of genes for DUR while the least at 552 genes was represented by ALN (471 QTLs) (Supplementary Table 4.2).

MOP shared the most haplotype blocks (*n* =10) with other populations compared to ALN sharing only 4. MOP and CAP (*n* = 963) shared the most haplotypes followed by ORT and CAP (*n* = 772) then ORT and MOP (*n* =761) sharing the third highest number of blocks (Supplementary Table 4.3). The village populations of ORT, CAP and MOP shared 479 haplotypes amongst them while the commercial populations of SAL and LWT shared 523 haplotypes. The indigenous KOL and WIN shared a total of 242 haplotype blocks. A total of 773 genes (Supplementary Table 4.4) were detected in MOP_CAP with QTLs ranging from teat numbers to marbling. *VAV1* (meat QTLs: ham weight and subcutaneous fat area) in chromosome 2 and *ADAMTS12* (QTLs: daily feed intake, drip loss, carcass weight, *etc*) in chromosome 16 were some of the genes observed in ALN_LWT populations sharing the least number of haplotype blocks at 157.

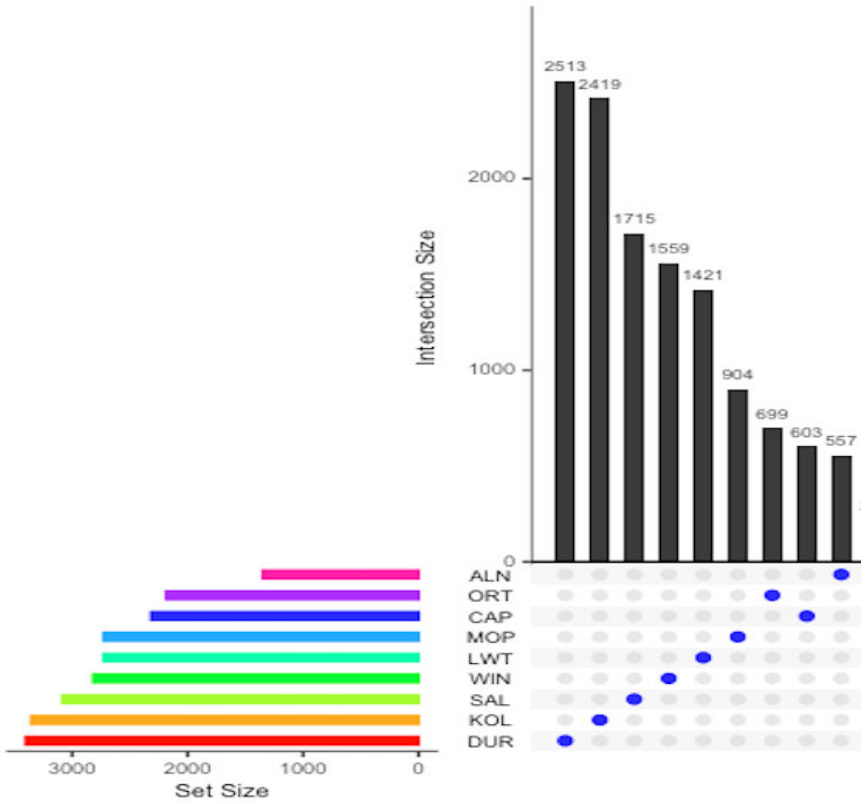


Figure 4.3 Distribution of unique haplotype blocks for pig populations

2165

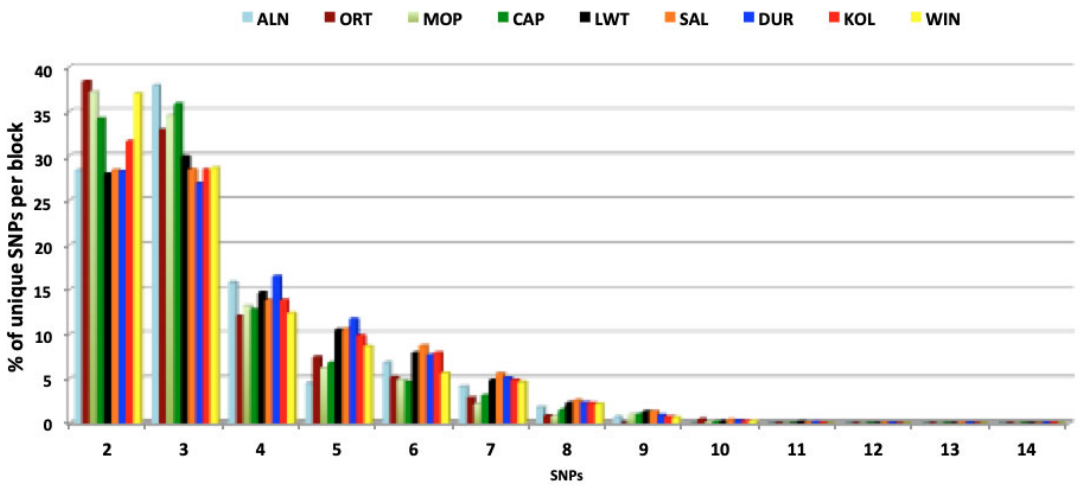


Figure 4.4 Percentage of unique SNPs of haplotype blocks for the studied populations

Only three haplotype blocks consisting of 3, 4 and 7 SNPs were shared by all nine populations while the rest were unique to each population (Figure 4.3). Four loci were detected on the shared haplotype blocks on chromosomes 4, 6 and 16 (Table 4.5). The genes associated with the shared regions were *DECR1*, *LENG9*, *TARS* and *ADAMTS12*. These genes were linked to carcass and growth quality (*DECR1*), immune response (*LENG9*), fertility (*TARS*) and milk production (*ADAMTS12*).

Table 4.5 Functions associated with haplotype blocks that were shared by all the nine populations

Chr	Location	Size (kb)	Gene	Function
4	51,208,996 - 51,268,855	56.86	*DECR1*	Carcass, Growth quality
6	53,368,946 - 53,422,895	53.95	*LENG9*	Immune response
16	20,397,715 - 203,977,914	199.36	*TARS, ADAMTS12*	Fertility, Milk production

4.4.6 Annotation of high LD regions

Across the nine populations, a total of 910 genomic regions had LD > 0.80 (Supplementary Table 4.5). Interestingly, most of the QTLs detected were found on Chromosome 7 (236) while Chromosomes 17 had 81 QTLs (Table 6, Supplementary Table 4.6). Most of the QTLs found were related to meat and carcass quality at 53.02 %QTLs. Carcass weight, skin thickness, spleen weight, shear force, saturated fatty acid content, pH, marbling, backfat, IMF content, lean meat percentage and loin weight were some of the traits related to meat and carcass quality. On average, meat and carcass quality constituted 53.02% of the observed QTLs followed by reproduction at 14.24% of the QTLs (Table 4.6). QTLs for litter traits were prevalent on chromosomes 1, 7, 9 and 13.

Table 4.6 QTLs associated with high LD regions

Chr	Location	Size (kb)	No. of QTLs	% QTLs detected				
				Exterior	Health	Meat & Carcass	Production	Reproduction
1	538,161 - 304,789,220	304.251	154	7.67	11.66	54.29	12.58	13.80
2	6,552,956 - 156,980,351	150.427	186	3.33	12.68	70.48	6.65	6.86
3	18,531,817 - 140,269,499	121.737	160	5.26	17.76	50.99	12.50	13.49
4	95,368 - 143,449,789	143.354	150	5.56	5.82	58.73	20.37	9.52
5	2,593,525 - 108,813,391	106.219	103	18.30	19.61	41.83	7.84	12.42
6	9,929,286 - 156,282,160	146.352	207	5.31	8.61	69.41	8.97	7.69
7	38,455,603 - 109,349,209	70.893	236	7.70	9.30	66.00	8.63	8.37
8	1,901,674 - 148,081,003	146.179	143	7.80	17.02	45.39	12.06	17.73
9	8,780,883 - 138,222,849	137.35	135	7.00	8.23	53.91	13.99	16.87
10	54,839 - 77,335,839	77.281	93	11.43	10.00	54.29	12.14	12.14
11	8,789,418 - 85,498,495	76.709	96	7.33	20.00	50.00	9.33	13.33
12	95,953 - 61,530,557	61.434	110	10.33	16.85	44.57	10.87	17.39
13	20,373,820 - 218,211,270	197.837	125	11.52	11.52	41.56	9.47	25.93
14	6,529,806 - 152,480,709	145.950	143	7.36	10.85	51.94	11.24	18.60
15	29,324,766 - 135,199,210	105.874	100	8.70	7.07	53.26	8.15	22.83
16	19,775,350 - 86,808,222	67.032	110	8.52	16.48	53.41	7.39	14.20
17	26,302,529 - 60,460,437	34.157	81	4.41	29.41	41.91	10.29	13.97
18	8,457,527 - 47,045,445	38.587	81	6.35	16.67	52.38	13.49	11.11
			Average	**7.99**	**13.86**	**53.02**	**10.89**	**14.24**

4.5 Discussion

LD analyses were important in identifying closely linked genetic markers connected to village, indigenous and commercial populations. Closely linked SNPs tend to be in high LD and only one SNP between the two is necessary to be used as a marker for genomic selection (Song *et al*., 2019). Chromosomal regions that are high in LD display patterns showing block-like structures called haplotype blocks. These haplotype blocks are important as they show evidence of historical recombination, shared evolutionary history and sometime selection (Templeton *et al*., 2005).

Average LD per population, ranged from 0.18±0.22 for Mopani to 0.40±0.36 for Kolbroek. The strong LD in Kolbroek, Duroc and South African Landrace could reflect recently formed mutations that did not allow recombination to break the haplotypes, or reduced diversity in the populations. The low genetic diversity levels reported in Chapter 3 supports the high LD levels observed in Duroc and Kolbroek. The low genetic diversity is consistence with the low recombination rates which will translate to high LD (Frankham, 2012). Populations raised in the intensive production system had a higher LD compared to the village populations. Mdladla *et al*. (2016) also observed low LD levels in village goats' populations that are raised extensively. The high LD levels in indigenous and commercial populations is also consistent with the low effective population size (Chapter 3). Under commercial production systems and in the case of indigenous Kolbroek and Windsnyer pigs, the breeding population is small and under intense artificial selection pressures compared to the more diverse village populations that are not artificially selected for given production traits. Makina *et al*. (2015) also demonstrated LD to be high in population with small sample size. Low rates of gene flow can have also contributed to the extensive r^2 values of indigenous and commercial populations (Pritchard & Przeworski, 2001). Village populations are highly admixed due high gene flow within and amongst populations coupled with indiscriminate crossbreeding as was demonstrated by the ADMIXTURE and Principal Component Analysis (PCA) based genetic structure reported in Chapter 3. The Kolbroek and Windsnyer pigs used in this study on the other hand are closed populations sampled from conserved flocks from the Agricultural Research Council. These populations are raised in the intensive production system like the commercial breeds with a small breeding pool. Similarly, Khanyile *et al*. (2015) observed that LD in conserved chicken populations was high compared to the admixed village chicken populations.

O'Brien *et al*. (2014) advocated for useful LD threshold ≥ 0.3 for genomic selection with Du *et al*. (2007) supporting this r^2 for QTL mapping in pigs. Ardlie *et al*. (2002) also observed an LD of 0.30 to be strong and valuable for QTL mapping. The r^2 values for indigenous and commercial populations shows that Porcine SNP60K can be useful to associate with economic important traits. The low LD levels for all village populations being less than 0.3 highlights the limitations of the current SNP array for genomic applications in these populations. A SNP chip with adequate marker density will be recommended for village pig populations to support effective genomic programs in the future. Gurgul *et al*. (2014) cautioned on the direct

comparison on the extent of LD between studies. This is because LD may be influenced by factors such as by sample size, marker type, density and distribution of markers and the strictness of SNP filtering. This then make it challenging to compare the LD observed in this study with other pig populations in other studies.

Several authors (Khanyile *et al*., 2015; Mdladla *et al*., 2016; Makina *et al*., 2015; Lashmar *et al*., 2016) have studied the extent to which LD decays in South African livestock. With increased marker distance, LD decayed to very low values of 0.07±0.10 at distances of 3000-4000 kb in the South African village pig populations. The value of r^2 peaked at 0.56±0.59 (Alfred Nzo) for the village populations at the distance of 0-1 kb and decreased to 0.19±0.23 at 500-1000 kb distances. Similar patterns of LD decay displayed by the village populations agrees with common genetic background and crossbreeding in village pigs as discussed in Chapter 3. The LD decay observed in village populations in this study was similar to that observed by Khanyile *et al*. (2015) in conserved and village crossbred chicken populations. LD decayed faster in village populations with a sharp decay observed within the 0-10 kb bin. There are several factors that could explain the faster and deeper LD decay observed in village populations. Firstly, it has been suggested that unlike commercial and indigenous pigs, village pigs originate from larger ancestral populations characterised by high genetic diversity levels of 0.359-0.370 (Chapter 3). Secondly, these populations are exposed to uncontrolled mating systems and indiscriminate crossbreeding that exposed the genomes to high recombination rates that break LD. Regions with low LD have a higher recombination rate. Grossi *et al*. (2017) also observed fast decaying values in crossbred Canadian pigs. Such similarities might be attributed to the outbred nature of both village and crossbred and most admixed populations, which will result in animals being less related to each other and thereby accumulating less haplotypes (Toosi *et al*., 2020). Overall, the low LD and faster and deeper LD decay observed in village pig populations suggests for a denser SNP array for these populations.

There is a clear distinction in the decay of LD between all the populations versus Kolbroek and South African Landrace populations. This is because commercial and the indigenous intensively raised Kolbroek and Windsnyer exhibit lower levels genetic diversity (Chapter 3). These populations are genetically and sometimes geographically restricted at research stations, in the case of Kolbroek and Windsnyer, with no outside gene flow and low effective population size. The LD observed in the commercial and indigenous populations could be because of history of the breed formation and current breeding practices of commercial pigs that make

intensive use of artificial insemination of a limited genetic base of proven sires (Slatkin *et al*., 2008). The Large White, South African Landrace and Duroc populations are raised under intensive environments subjected to artificial insemination (Krüger *et al*., 2017) with limited genetic pool. Hoglund *et al*. (2014) also noticed an increase of LD due to use of AI and selection intensity in Nordic Holstein, Red and Jersey cattle. Artificial selection has led to close genetic pools in the intensive sector that is mainly practised by commercial farmers (Amills *et al*., 2010).

In this study, certain chromosomes had a higher LD value when compared to others. A significant ($P < 0.0001$) was observed in this study. Chromosome 1 which is the largest chromosome for *Sus scrofa* species showed higher LD values followed by chromosome 14 and 13. The village populations had their highest r^2 on chromosome 14. Uimari *et al.* (2011) also observed chromosomes 1, 13 and 14 to have the highest LD values in Finnish pig breed whereas Chromosome 1 harboured *ESR1* gene associated with reproduction while chromosome 13 was linked to ovulation rate QTL. Espigolan *et al*. (2013) observed no relationship between chromosome size and LD that might have been affected by the medium density SNP panels used. Arias *et al*. (2009) observed that as the length of the chromosomes increased the recombination rates decrease. The population and chromosome effect were also supported by Nsengimana *et al*. (2004) who pointed to intensive selection for traits influenced by QTLs on given chromosomes. Lozada-Soto *et al.* (2021) showed there is a variation in recombination rates between Large White and Landrace breeds. In this study, across chromosomes, the commercial and indigenous populations had higher LD compared to the villages.

Persistence of phase can be used to assess relatedness between populations within a species (Goddard, 2006). The high correlation between Duroc and Large White ($r = 0.79$ at 0-1kb) was not anticipated as these two breeds grouped under different PCA clusters shown in Chapter 3 indicative of their dissimilar genetic backgrounds. However, this persistence of gametic phase between Duroc and Large White could be linked to both populations being reared under the intensive production system under selection for similar traits (production efficiency traits). Badke *et al*. (2012) and Veroneze *et al*. (2014) also observed high correlation of 0.87 (0-1 kb) between Duroc and Large White. The higher correlation at 0-1 kb found between villages and commercial ($r > 0.62$) compared with correlation amongst villages ($r > 0.51$) was expected as $K = 10$ on ADMIXTURE and PCA clustering reported in Chapter 3 showed village pigs sharing a certain component of the commercial breeds genetics. The low correlation amongst village

pigs suggests for a more breed specific genomic tools and denser bead chip to support across population genetic improvement initiatives.

The average haplotype block length observed in this study is less than that reported in a study on beef cattle (Mokry *et al*., 2014). The short haplotype block length observed in village pigs could be due to crossbreeding with commercial breeds that is taking place in these village populations. Oyelami *et al*. (2019) also observed smaller haplotype block sizes in admixed populations that were crossbred with western breeds. The large number of haplotype blocks and haplotypes found in the DUR and KOL populations also corresponds with the high LD values observed in those populations. For almost all the populations, majority of the haplotypes were observed at a frequency ranging from 0.25 to 0.50 in agreement with a study conducted by Khanyile *et al*. (2015) using 60K SNP panel in chicken populations. Luikart and Allendorf (1996) noted the variations in frequencies of haplotypes to be acquired within a couple generations if the genomic flow remains sturdy. The crossbreeding taking place in the village populations could be contributing to low frequency haplotypes. Low frequency haplotypes may also be due to recombination events such as fitness mechanism of animals' thus breaking haplotypes. For example, to remove a haplotype susceptible to a disease takes many generations by using boars that are not carriers of the concerned haplotype (Chen *et al*., 2020). Oyelami *et al*. (2020) also observed low LD coupled with low haplotype frequency of 0.465 and 0.467 in admixed pig populations of Huaibei and Hongdenglong from Jiangsu province in China respectively. Longer haplotypes at high frequencies assist in maintaining within population LD (Lai, 2012). The observation that Mopani pigs had most haplotypes occurring at < 0.1, is therefore disturbing as this means that it will be difficult to detect positive selection in these populations.

Source of distinctive phenotypic characteristics can be revealed through unique and shared haplotype blocks. Unique haplotype blocks can be of assistance for future conservation programs. Khanyile *et al*. (2015) indicated that unique haplotypes arise from independent development and genomic structuring of populations. Duroc harboured the highest number of 2 513 (74%) unique haplotype blocks associated with 2 384 genes. Various QTLs ranging from reproduction, health and meat and carcass were observed within these unique haplotype blocks indicative of selection driven accumulation of these haplotype block variants. Mopani pigs shared the highest number of haplotype blocks with Capricorn (963) in line with the ADMIXTURE results reported in Chapter 3. These results suggest that Mopani pigs might be

from several ancestries that are shared with other populations or experienced introgression with or from other populations. Zhang *et al*. (2016) reported severe genetic erosion on Chinese indigenous pigs that shared more haplotype blocks with commercial pigs. Genetic erosion is therefore a possibility with Mopani pigs that shared most of their haplotype blocks. In Chapter 3, results of ADMIXTURE $K = 10$ analysis implied that Alfred Nzo was the least admixed and unique village population, which explains the least number of haplotypes blocks this population shared with other breeds. The highest numbers of haplotype blocks shared between populations were from the villages (MOP_CAP 963; ORT_CAP 772; ORT_MOP 761). Generally, village pigs were the highly admixed suggestive of shared ancestry. The sharing of 523 haplotype blocks between South African Landrace Large and White support their breeding history of Large White contributing to the genetic material of the South African Landrace (Porter, 1993; Chapter 3). The QTLs and genes harboured by the shared haplotype blocks reveals selection pressures towards superior meat quality and growth rate. *DECR1* gene, detected in shared haplotype blocks in this study has an association to meat and growth traits and is located in chromosome 4 which harbours 135 QTLs. Clop *et al*. (2003) and Muñoz *et al*. (2019) also found an association of *DECR1* gene to meat quality and carcass traits. The percentage of meat and carcass quality QTLs was well represented on chromosome 2, 6 and 7 at 70.48%, 69.41% and 66% respectively. Several studies have also reported a considerable QTLs for meat and carcass quality in pigs (Choi *et al*., 2010; Choi *et al*, 2011; Stratz *et al*., 2013; Uimari & Sironen, 2014; Stratz *et al*., 2018; Oyelami *et al*., 2020). In this study, chromosome 7 harboured more QTLs at 236 while the least QTLs were found in chromosomes 17 and 18. Nsengimana *et al*. (2004) also noted more QTLs on chromosome 7 for five European and synthetic pig populations aligned with the selection pressures associated with this chromosome. The high percentage of high LD regions related to meat and carcass quality traits observed in this study implies that pig farmers are prioritising selecting animals based on these economic traits.

It was also encouraging to observe QTLs associated with adaptation and fertility linked to the haplotype blocks observed in this study. Chromosome 6 was linked to immune response (*LENG9*). In our previous study, immune response genes genomic regions were also observed in Chapter 3 and highly differentiated populations of indigenous, Duroc and Wild Boar.

As expected, chromosome regions harbouring QTL had stronger LD and long haplotype. Overall, results of this study highlight the importance of haplotype block-based methods in identify SNPs associated with important phenotypic traits. Different selection pressures for the

different populations might have led to high LD between pairwise causative/associated SNPs and accumulation of haplotype blocks particularly in regions harbouring specific QTLs associated with the traits under selection (Ardlie *et al*., 2002).

4.6 Conclusion

High LD was observed at short distances in all the South African pig populations. Commercial and indigenous breeds showed the highest levels of LD compared to the village populations. The higher LD in the Kolbroek and Duroc breeds are suggestive of the effects of genetic drift, and recent bottlenecks events in these populations. Genomic selection and assisted improvement could be limited by the low LD and that persisted up to very short markers distances and the low persistence in gametic phase. The unique haplotype blocks observed in this study, are suggestive of genetically different populations emanating from either different genetic background or evolutionary forces. The QTLs associated with high LD regions, unique and shared haplotype blocks revealed economic and adaptive traits that could be useful for future improvement and conservation decisions of these pig populations.

CHAPTER 5: GENOMIC ANALYSES REVEAL POPULATION HISTORY AND INTERACTIONS OF SOUTH AFRICAN PIG POPULATIONS

5.1 Abstract

The well-defined breeds, non-descript breeds and wild populations are found across the geographic spread of South Africa, with phenotypic evidence of geneflow amongst the populations. However, genomic evidence linking the commercial, indigenous, village and wild populations is lacking and constraining efforts to optimise utilisation and conservation of these indigenous genetic resources. Illumina Porcine SNP60K genotyping data from 234 pigs was used to trace the gene flow and introgression between indigenous, commercial, village and wild pig populations from different production systems. *Frappe* results showed Clusters 1 (Windsnyer) and 3 (Alfred Nzo) depicting the highest genetic variations in the village populations. Phylogenetic analyses presented the Capricorn population as being the most introgressed by commercial and village populations. PCAdmix showed the domination of Alfred Nzo as an ancestral population for other village pig populations. Migration of Wild Boar to O. R. Tambo (M = 1) cements its place as the founder of domestic pigs. Treemix results of the Vietnamese Potbelly supports the belief of Asia being one of the major domestication centres. Long shared IBD (> 1cM) segments prevailed in the genomes of the South African pigs suggesting inheritance from recent common ancestry. Significantly enriched GO terms for the shared IBD segments were associated with meat quality, feed efficiency, muscle growth and fertility. Overall, the study demonstrated contributions of domestication, admixture, selection and adaptation in shaping the phenotype and genomic variations that exists in the South African pig populations. This study also showed that the history and lineage of the pig populations globally is complicated by the common ancestries between domestic and wild populations that occurred through various admixture events.

Keywords: *Breed history, introgression, migration, segment detection, IBD, pigs*

5.2 Introduction

There is patchy and unreliable data and information on the history of pig populations in South Africa and other regions of Africa (Blench & MacDonald, 2000; Amills *et al*., 2013). Blench and MacDonald (2000) suggested pigs to have spread to South Africa via the Nile Corridor. The indigenous pigs found in sub-Saharan Africa have phenotypic and genetic similarities with ancient Egyptian black pigs supporting the suggested spread of pigs from North to Southern

Africa (Blench & McDonald, 2000). Medrano *et al*. (2013) found a link between pigs from North Africa and the Iberian Peninsula. The existence of the non-indigenous Wild Boar in Africa implies a human-aided introduction and spread of wild pigs into the African continent. According to Oliver (1993), the first entry of the Wild Boar from Europe in the continent was through North Africa, after which they scattered naturally all over the continent (Blench & MacDonald, 2000). The commercial pig breeds (Large White, Duroc and Landrace) that are thriving in the commercial industry in South Africa are reported to have arrived from Europe, Oceania and America with genetic and phenotypic distinctiveness (Holness, 1974; Ramírez *et al*., 2009; Burgos-Paz *et al*., 2013). According to Maule and Maule, (1960), the introduction of pigs in Eastern and Southern Africa is thought to have happened around the 19th century with Indian traders bringing Asian and Mediterranean breeds. Several authors have also traced the origins of African pigs to India and Southeast Asia (Noce *et al*., 2015; Osei-Amponsah *et al*., 2017).

Domestic pigs spread from sub-Saharan African to South Africa via the Nile Corridor (Blench & MacDonald, 2000). Intensive production during the last 200 years coupled with crossbreeding between the various domestic pig populations has resulted in moderate to high genetic diversity in South African porcine populations (Scandura *et al*., 2011; Frantz *et al*., 2013; Chapter 3) that include pure indigenous (Kolbroek, Windsnyer) and commercial (Large White, Landrace, Duroc) pig breeds (Chapter 3). According to government statistics, there are 1.3 million pigs in South Africa largely concentrated in Limpopo (329K) with Eastern Cape (79K) and Northern Cape (16K) having the least numbers (DALRRD, 2021). The pig industry is mainly dominated by exotic breeds of the commercial sector (Swart *et al*., 2010). Wild pig populations like the Warthogs, Wild Boars and Bush Pigs are found free ranging around game reserves or raised by farmers for meat production throughout the country (Porter, 1993). There has been an elevated suspicion of hybridization of Bush Pigs with domestic pigs, as well as interbreeding of Bush Pigs with Wild Boars in South-Eastern KwaZulu-Natal province of South Africa (Vercammen *et al*., 1993; Kingdom, 2003; Venter *et al*., 2016). However, early records of Warthog and Bush Pig species are treated with caution, as early literature seem to have confused the two (Skead, 2007). Both Bush Pig and Wild Boars remain the asymptomatic carriers of African Swine Fever disseminating the virus to domestic pigs (Jori & Bastos, 2009). According to White (2011) European breeds still have some phenotypic similarities to the Wild Boars, their wild ancestor.

In South Africa, commercial farmers keep international breeds of Large White, Landrace and Duroc, while smallholder village-based farmers keep indigenous breeds such as the Kolbroek and Windsnyer to crosses and non-descript in villages (Chapter 3). Additionally, Kolbroek and Windsnyer breeds are also kept as conserved populations. The modern Kolbroek is a synthetic breed developed at the ARC in 1996 from the Great White, Tamworth, Sandveld and Windsnyer, and genetics with the indigenous Kolbroek being from East and European pig genetics (Coleman, 2018). Coleman also noted the Windsnyer to be an ancient popular breed from Zimbabwe. The gene flow of exotic pigs into village pigs is evident and driven by commercial farmers selling the replacement stock directly to farmers in the local communities or regions for breeding (Madzimure *et al.*, 2012; Matabane *et al.*, 2015; Gcumisa *et al.*, 2016). Mokoele *et al.* (2014) noted that the emerging small-scale farmers in Limpopo farmed with Large White (28,29%), Landrace (10,53%) and Duroc (7,24%).

Amills *et al.* (2021) stressed the need for comprehensive studies to quantify ancestry and genetic relations of geographically connected and dispersed African pigs in order to fill in the information gaps regarding their origin, and onward evolution. Genetic relatedness provides an understanding of how populations are connected. The Porcine SNP60K is an important genomic tool with a genome-wide coverage of SNPs to assist in increasing our understanding of the population genetic history, and relationships of pig populations (Moyo *et al.*, 2018; Reimer *et al.*, 2018; Amills, 2021). Ovine SNP50K genome-wide SNPs have found application in detection of haplotypes linked to introgression in livestock species (Amaral *et al.*, 2008; Herrero-Medrano *et al.*, 2012). Determining the lineage per chromosome segment in amalgamated individuals is crucial for understanding the population ancestral history including admixture events, their timing thereof and evolutionary forces that influences their prevalence and distribution in populations (Stam, 1980; Pool & Nielsen, 2009; Pugach *et al.*, 2011), as well as population growth after these admixture events (Chapman & Thompson, 2002). The ability to identify genomic segments between individuals that are identical-by-descent (IBD) can help trace recent or ancient shared genetic ancestry between individuals and populations (Browning & Browning, 2012; Taylor *et al.*, 2020). IBD segments are useful for estimating relatedness between individuals and provides for a wide range of analyses including tracing population history and gene mapping of traits of interest such as those predisposing populations to diseases (Belbin *et al.*, 2017; Han *et al.*, 2017; Ramstetter *et al.*, 2018). IBD segments can be inherited at different times. The shorter (< 1 cM) IBD segments are a result of older shared ancestry while recent (within the past 25 generations) common ancestry is associated with long genomic

segments shaped by migration and other historical events (Broman *et al*., 1999). Browning and Browning (2012) also noted the importance of high frequency of shared segments to signal ancient relationships between populations.

A limited number of studies have been performed to increase our understanding of the diversity and genetic structure of South African pig populations (Swart *et al*., 2010). For example, worldwide genotypes have demonstrated admixture between village and commercial breeds in American pigs (Burgoz-Paz *et al*., 2013; Chapter 3). Similar studies in other countries have suggested that the existence of wild populations alongside domestic pigs particularly in the villages has resulted in gene flow and introgression between wild and domestic pigs (Burgoz-Paz *et al*., 20103; Frantz *et al*., 2013; Mujibi *et al*., 2018). Gene flow and introgression evolutionary forces have not been investigated in the wild, village, indigenous and commercial pigs of South Africa and other neighbouring countries. The overall aim of this study was therefore to examine population histories, migration, introgression and other genetic interactions amongst South African pigs using genome-wide SNP data. Using the Porcine SNP60K genotypes generated in the previous Chapter 3, this Chapter undertook analyses that included *frappe* for estimating individual ancestry proportion, Neighbor-Joining tree for phylogenetic analysis, Treemix to infer on gene flow between populations and PCAdmix to infer on population structure and ancestry. Coupled to that, the study, screened and analysed for the frequency and distribution of IBD segments within and amongst South African pig populations to assess recent or ancient shared segments.

5.3 Materials and Methods

5.3.1 Animal samples, genotyping and quality control

A total of 234 genotypes were used in this study, which included pig populations from the villages (n = 91), commercial (n = 60), indigenous (n = 40), Vietnamese Potbelly (n = 5) and wild (n = 38) as described in Chapter 1 and 3. The village pigs were represented by ALN (Alfred Nzo, n = 17), O.R. Tambo (ORT, n = 22), CAP (Capricorn, n = 25), MOP (Mopani, n = 27) while LWT (Large White, n = 20), SAL (South African Landrace, n = 20), DUR (Duroc, n = 20) represented the commercial, KOL (Kolbroek, n = 20), WIN (Windsnyer, n = 20) the indigenous pigs and WBO (Wild Boar, n = 4), WAT (Warthog, n = 31) and BSP (Bush Pigs, n = 3) the wild pigs. For genotyping quality control, SNPs were filtered according to the following

criteria: SNP call rate ≥ 85%, minor allele frequency ≥ 0.02, markers that significantly deviated from Hardy-Weinberg Equilibrium (*P < 0.0001*). Quality control was performed in Golden Helix SNP Variation Suite (SVS) *version* 8.8.1 (Golden Helix Inc., 2016). SNPs without chromosome number and/or position information or on sex chromosomes of the *Sus scrofa* genome (assembly *version* 11.1) were removed. In addition, for Treemix and PCAdmix analysis was pruning for individuals with an IBD ≥ 0.45 to remove one of the closely related individuals and LD pruning at $r^2 = 0.2$ to remove one of the pair of closely associated SNPs.

5.3.2 Population structure and ancestry estimation

Frappe was used to estimate individual admixture by using a genotype and parameter file with individual genotypes (Tang *et al*., 2005). ADMIXTURE *version* 1.20 (Alexander *et al*., 2011) was used to detect the most likely optimum *K*-value. This is because *frappe* does not infer on the probable number of clusters (*K*). The maximum likelihood (ML) algorithm was then implemented in the *frappe* software (http://smstaging.stanford.edu/tanglab/software/frappe.html) to estimate the proportional ancestries for each population using unphased genotype data (Tang *et al*., 2005). *Frappe* uses a defined number of ancestries (*K*-values) and in this case, *K*-value of 10 detected by ADMIXTURE was used. Maximum iteration of EM to run was set to 10 000 with no less four independent runs per populations.

5.3.3 Phylogenetic analysis

Neighbour-joining (NJ) relationship tree was built based from the Treemix output .treeout file. A bootstrap consensus tree was obtained by using 1 000 replicates. Itol web-based software was used to visualize the results (Letunic & Bork, 2021).

5.3.4 TreeMix, f_3 and f_4 analysis

Treemix package *version* 1.13 (Pickrell & Pritchard, 2012) was used to estimate Maximum likelihood (ML) tree of the 13 populations with migration edges and to fit admixture gene flow events between branches of population pairs. Blocks of 500 SNPs (-k) were grouped together to account for linkage disequilibrium and ensure independency of blocks. To build the ML tree, 0-15 migration (-m) events were used. Ten independent replicates were run in order to confirm

consistency of migration events. R script plotting_func.R and ggplot2 package was used to visualize the graph.

In addition, using Treemix *version* 1.13, the f_3 and f_4 statistics were calculated to verify for the admixture in the history of tested populations (Reich *et al*., 2009; Patterson *et al*., 2012). These f_3 and f_4 statistics measure allele frequency correlations between populations. The *f*-statistics work on population level, which examine phylogenetic relationships by estimating the genetic drift as variance in allele frequencies along tree branches that are shared amongst populations. In the f_3 test (A; B, C), population A is tested as a mixture of B and C, and this was run with all possible triplets from the thirteen populations. For the f_4 test (A, B; C, D) with frequency difference between A and B being tested against populations C and D.

5.3.5 Evolutionary relationships and tracing admixture using PCAdmix

PCAdmix uses a technique that estimates local lineage through principal components analysis (PCA) using phased haplotypes. PCAdmix requires pre-defined ancestries and identifies ancestry patterns that were inherited on each chromosome in admixed individuals thus providing the most likely ancestry classification for everyone. PCA (https://www.sites.google.com/site/pcadmix/home) was used to assign variants that are more informative about ancestry at greater weights (Brisbin *et al*., 2012). Principal components were then used to determine ancestry scores for each window. This was done using the command line -anc anc.tx -adm admixed.txt -map file mapfile.txt -rho genetic_map.txt -w 40 -wseg 20 -r2 0.8. Default parameters were used for PCAdmix analysis except for the threshold for window size of the haplotypes that was set at 40 SNPs. Highly related alleles from different populations were pruned using linkage disequilibrium threshold of $r^2 > 0.8$ to remove highly linked SNPs. The ancestral (anc) populations for each admixed (adm) village were as follows: ALN (WIN, DUR, ORT, SAL); ORT (WIN, ALN, LWT, SAL); MOP (WIN, DUR, LWT, SAL, ALN) and CAP (WIN, SAL, ALN, LWT, DUR). Phasing was done using BEAGLE 5.1 (Browning & Browning, 2007) separately for each breed and per chromosome using the following parameters: 30 iterations of the phasing algorithm on a chromosomal region of 5 Mb and sample haplotype pairs for each individual per iteration.

5.3.6 Pairwise IBD detection

Germline 1.5.3 (http://www.gusevlab.org/projects/germline/) software (Gusev *et al*., 2008) was used to estimate the amount of shared common ancestor segments using the command line germline -input (ped file) (map file) -output with additional parameters (min_m =1; err_hom=1; err_het=2) and map file for marker positions. For our analysis, the length of 0-1 cM (short segments) was used to define ancient, shared segments while length of >1 cM (long segments) was considered recent ancestry segments. A count of the number of segments for each pair was also done. Histograms were created for the total number of segments shared between pairs of individuals within each population. A segment was detected as a set when at least two consecutive SNPs had a probability of more than 50% of sharing of at least one IBD (Han & Abney, 2013).

5.3.7 GO-enrichment analysis

SNPs generated from the pairwise IBD analysis was used for enrichment analysis. Ensemble BioMart (*Sus scrofa* 11.1) gene database was used to annotate genes in the candidate regions (release 106). Web-based DAVID Bioinformatics Resources *version* 6.8 server was then employed by uploading gene IDs for functional and pathway enrichment analysis (Mi *et al*., 2017). Revigo (Reduce + Visualize Gene Ontology), a web-based tool (http://revigo.irb.hr) was used to summarise and visualise GO (Gene Ontology) biological process terms in all populations by removing redundant GO terms by simple clustering algorithm method (Supek *et al*., 2011). A significant *P-value* ($P < 0.01$) was based on EASE score: a modified Fisher's exact test.

5.4 Results

5.4.1 Sample and SNP quality control

One hundred and seventy-six individuals (with 85% call rate) and 27 740 SNPs remained for downstream analysis. This quality-controlled data set was used to analyse for number of shared IBD for each population pair. Further QC of LD ($r^2 = 0.2$) and IBD (≥ 0.45) resulted in 23 345 SNPs and 176 individuals used for *Frappe*, Treemix and PCAdmix analysis.

5.4.2 Population structure and ancestry estimation

Based on results from Chapter 3, the ADMIXTURE analysis showed the lowest cross validation error was at K = 10, displaying the presence of ten genetic clusters. Table 5.1 shows the summarized proportional contribution of the assumed ancestral populations using the ten ancestral populations. Pig populations VIT, WBO and DUR scored the highest ancestries of 100% (1.00±0.00) at clusters 6, 7 and 9 respectively. Clusters 1 and 3 was constituted by pigs from village populations that had the most variations. Cluster 1 was dominated by WIN at 97%, (0.97±0.08) and had substantial contribution from MOP (43%; 0.43±0.11), CAP (28%; 0.28±0.10) and ORT (12%; 0.12±0.03). The village population ALN was dominant in cluster 3 at 85% (0.85±0.13) followed by ORT (57%; 0.57±0.10), MOP (16%; 0.16±0.05) and CAP (14%; 0.14±0.08). BSP and WAT shared cluster 4 at 87% (0.87±0.04) and 92% (0.92±0.25), respectively. LWT led cluster 5 with 98% (0.98±0.05) and had MOP and CAP villages contributing at 15% and 10% respectively. KOL population was represented in cluster 10 at 99% (0.99±0.03). Supplementary Table 5.1 have detailed results on individual cluster contributions for each population.

Table 5.1 *Frappe* based estimates of proportions for South African pig populations

	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7	Cluster8	Cluster9	Cluster10
ALN	0.04±0.05	0.01±0.02	**0.85±0.13**	0.01±0.01	0.02±0.03	0.01±0.01	0.02±0.02	0.02±0.02	0.01±0.01	0.01±0.01
BSP	0.00±0.00	0.00±0.00	0.00±0.00	**0.87±0.04**	0.00±0.00	0.00±0.00	0.00±0.00	0.13±0.04	0.00±0.00	0.00±0.00
CAP	**0.28±0.10**	**0.23±0.08**	**0.14±0.08**	0.00±0.01	0.15±0.06	0.01±0.03	0.03±0.02	0.00±0.01	0.08±0.07	0.07±0.02
DUR	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	**1.00±0.00**	0.00±0.00
KOL	0.01±0.02	0.00±0.00	0.00±0.01	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.01	0.00±0.00	0.99±0.03
LWT	0.00±0.00	0.01±0.03	0.00±0.00	0.00±0.00	**0.98±0.05**	0.00±0.01	0.00±0.01	0.00±0.00	0.00±0.00	0.00±0.00
MOP	**0.43±0.11**	**0.13±0.05**	**0.16**±0.05	0.00±0.00	**0.10**±0.07	0.02±0.01	0.03±0.03	0.03±0.02	0.05±0.07	0.05±0.02
ORT	**0.12±0.03**	0.09±0.05	**0.57±0.10**	0.01±0.01	0.06±0.04	0.01±0.01	0.04±0.02	0.02±0.01	0.03±0.02	0.05±0.03
SAL	0.01±0.01	**0.92±0.16**	0.00±0.00	0.00±0.00	0.05±0.14	0.00±0.01	0.01±0.02	0.00±0.00	0.00±0.00	0.01±0.01
VIT	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	**1.00±0.00**	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
WAT	0.00±0.00	0.00±0.00	0.00±0.00	**0.92±0.25**	0.00±0.00	0.00±0.00	0.00±0.00	0.08±0.25	0.00±0.00	0.00±0.00
WBO	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	**1.00±0.00**	0.00±0.00	0.00±0.00	0.00±0.00
WIN	**0.97±0.08**	0.00±0.01	0.00±0.02	0.00±0.00	0.00±0.01	0.00±0.01	0.00±0.01	0.00±0.00	0.00±0.01	0.01±0.04

5.4.3 Phylogenetic analysis

The result from the phylogenetic tree shows WIN as outgroup from the rest of the population (Figure 1). A total of 4 clusters were identified. These included cluster 1 (VIT, WAT and BSP), cluster 2 (ALN and ORT), cluster 3 (SAL and LWT) and cluster 4 (WBO, DUR, KOL). WAT and BSP share an ancestry with VIT while ALN and ORT are genetically closed. There is also sharing of a descendant by WBO, DUR and KOL.

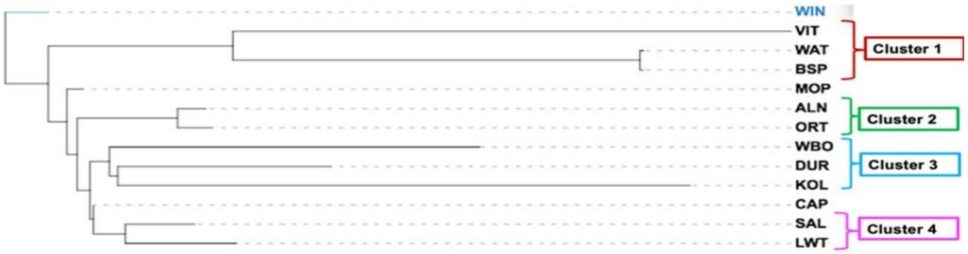


Figure 5.1 Phylogenetic tree based on the South African pig populations. The tree was constructed by the neighbour-joining method using TreeMix program without migration edges (*M* = 0) population

5.4.4 TreeMix, f_3 and f_4 analysis

Migration events were set at 10, explaining marginally higher 99.68% (Supplementary Table 5.2) variance for the relatedness between the populations. Migration edges showed introgression between populations and was coloured according to the percentage received from the contributing population (Figure 5.2). At *M* = 1, high migration gene flow from WBO to ORT populations was observed. Gene flow from WBO to WAT and BSP was also observed at *M* = 3 while *M* = 2 connected WAT and BSP to Kolbroek and DUR populations. VIT was a

donor population to MOP (*M* = 5), CAP (*M* = 6), WIN (*M* = 7), ALN (*M* = 9) and LWT with SAL (*M* = 10). Migration from the VIT together with BSP and WAT populations to the KOL were also indicated at *M* = 4. There was gene flow of SAL to DUR at *M* = 8.

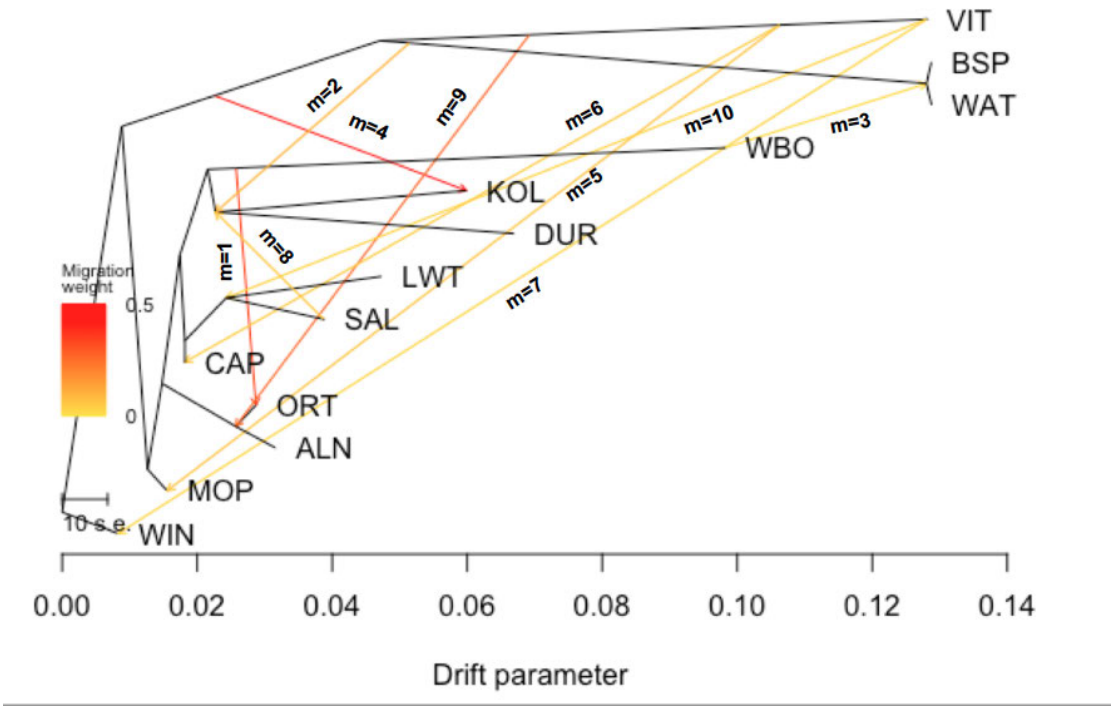


Figure 5.2 Maximum likelihood tree inferred from 13 populations with migration edges based on population pairwise allelic covariance inferred with TreeMix (Pickrell & Pritchards, 2012). Coloured lines reflect the weights of migration according to the colour shades of lineage obtained from the contributed population

Three population f_3 tests was utilized to further highlight admixtures between populations. Table 5.2 illustrates all populations with significantly negative f_3 tests. The significant Z-Scores supports VIT introgression into the village populations (Figure 5.2 and Table 5.2). We also observed significant contributions of commercial and indigenous populations to the CAP.

Table 5.2 Significant f_3 statistics for pig populations. Three population tests were performed on all possible combinations of populations. A negative Z-score of 3 or more correlating to a *p-value* of 0.001

Pop A	Pop B	Pop C	f_3	Standard Error	Z-Score
ORT	VIT	WBO	-0.01336	0.00241	-5.55662
CAP	VIT	WBO	-0.01224	0.00261	-4.68002
CAP	WIN	LWT	-0.00269	0.00063	-4.28068
MOP	VIT	WBO	-0.01152	0.00278	-4.14655
CAP	SAL	WIN	-0.00245	0.00070	-3.51137
CAP	MOP	SAL	-0.00121	0.00039	-3.10585
ALN	VIT	WBO	-0.00834	0.00313	-2.66884
CAP	BSP	LWT	-0.00190	0.00080	-2.36179
CAP	LWT	WAT	-0.00192	0.00082	-2.34383
CAP	LWT	VIT	-0.00376	0.00161	-2.33765
CAP	SAL	WAT	-0.00216	0.00095	-2.26729
CAP	SAL	BSP	-0.00202	0.00095	-2.12177
CAP	ALN	SAL	-0.00163	0.00077	-2.11290
CAP	SAL	VIT	-0.00294	0.00151	-1.94619
CAP	VIT	DUR	-0.00216	0.00177	-1.21598
CAP	KOL	LWT	-0.00068	0.00063	-1.09183
CAP	KOL	SAL	-0.00073	0.00073	-1.00609

The admixture nature of the village populations was further supported by the more sensitive f_4 statistics (Table 5.3) which showed top 20 significant negative and positive Z-Scores. These results depict common ancestries between the commercial, indigenous, villages and wild populations

5.4.5 Evolutionary relationships and tracing admixture using PCAdmix

The ancestral population for each village was based on the $K = 10$ ADMIXTURE results with eight distinct genetic clusters as reported in Chapter 3. All the populations were group according to their clusters except for the villages which showed admixture. The proportional ancestry for each population based on *frappe* results are provided as a summary in Table 5.1 and Supplementary Table 5.1 shows ancestry at individual level Figure 5.3a shows ALN ancestry was well represented in all the village populations. WIN had the most common ancestry

Table 5.3 Significant f_4 statistics for pig populations. Four population tests were performed on all possible combinations of populations. A negative Z-score of 3 or more corresponds to a *p-value* of 0.001

Pop A	Pop B	Pop C	Pop D	f_4	Standard Error	Z-Score
CAP	WAT	BSP	LWT	-0.13858	0.00210	-65.98860
MOP	WAT	BSP	LWT	-0.13315	0.00204	-65.19580
SAL	WAT	BSP	LWT	-0.14650	0.00227	-64.57750
CAP	WAT	BSP	DUR	-0.13288	0.00210	-63.31030
SAL	WAT	BSP	DUR	-0.13538	0.00214	-63.20800
MOP	WAT	BSP	DUR	-0.12833	0.00218	-58.84180
CAP	WAT	BSP	WBO	-0.13528	0.00231	-58.59150
BSP	LWT	DUR	WAT	-0.13426	0.00230	-58.44650
BSP	DUR	LWT	WAT	-0.13429	0.00231	-58.19430
WIN	WAT	BSP	LWT	-0.12971	0.00224	-58.00350
ORT	WAT	BSP	LWT	-0.13010	0.00226	-57.54270
WIN	WAT	BSP	DUR	-0.12613	0.00222	-56.70890
ALN	WAT	BSP	LWT	-0.12500	0.00225	-55.49650
SAL	WAT	BSP	WBO	-0.13789	0.00255	-53.97980
ORT	WAT	BSP	DUR	-0.12562	0.00236	-53.31170
MOP	WAT	BSP	WBO	-0.13224	0.00249	-53.19570
KOL	WAT	BSP	LWT	-0.12681	0.00242	-52.35520
WIN	WAT	BSP	WBO	-0.13047	0.00258	-50.57600
BSP	DUR	WBO	WAT	-0.13597	0.00270	-50.32510
BSP	WBO	DUR	WAT	-0.13591	0.00270	-50.29800
MOP	WAT	SAL	BSP	0.13236	0.00183	72.32330
MOP	BSP	SAL	WAT	0.13230	0.00187	70.88290
SAL	BSP	CAP	WAT	0.13882	0.00203	68.28920
SAL	WAT	CAP	BSP	0.13868	0.00206	67.47790
CAP	BSP	LWT	WAT	0.13856	0.00210	66.11790
MOP	WAT	CAP	BSP	0.13142	0.00200	65.79200
MOP	BSP	CAP	WAT	0.13149	0.00200	65.66910

dispersed in the entire windows. Ancestries from the indigenous WIN and commercial SAL populations were present in all the four villages. The ancestries for the village populations were as follows: Alfred Nzo (21% WIN, 22% DUR, 37% ORT, 20% SAL); O.R. Tambo (16% WIN,

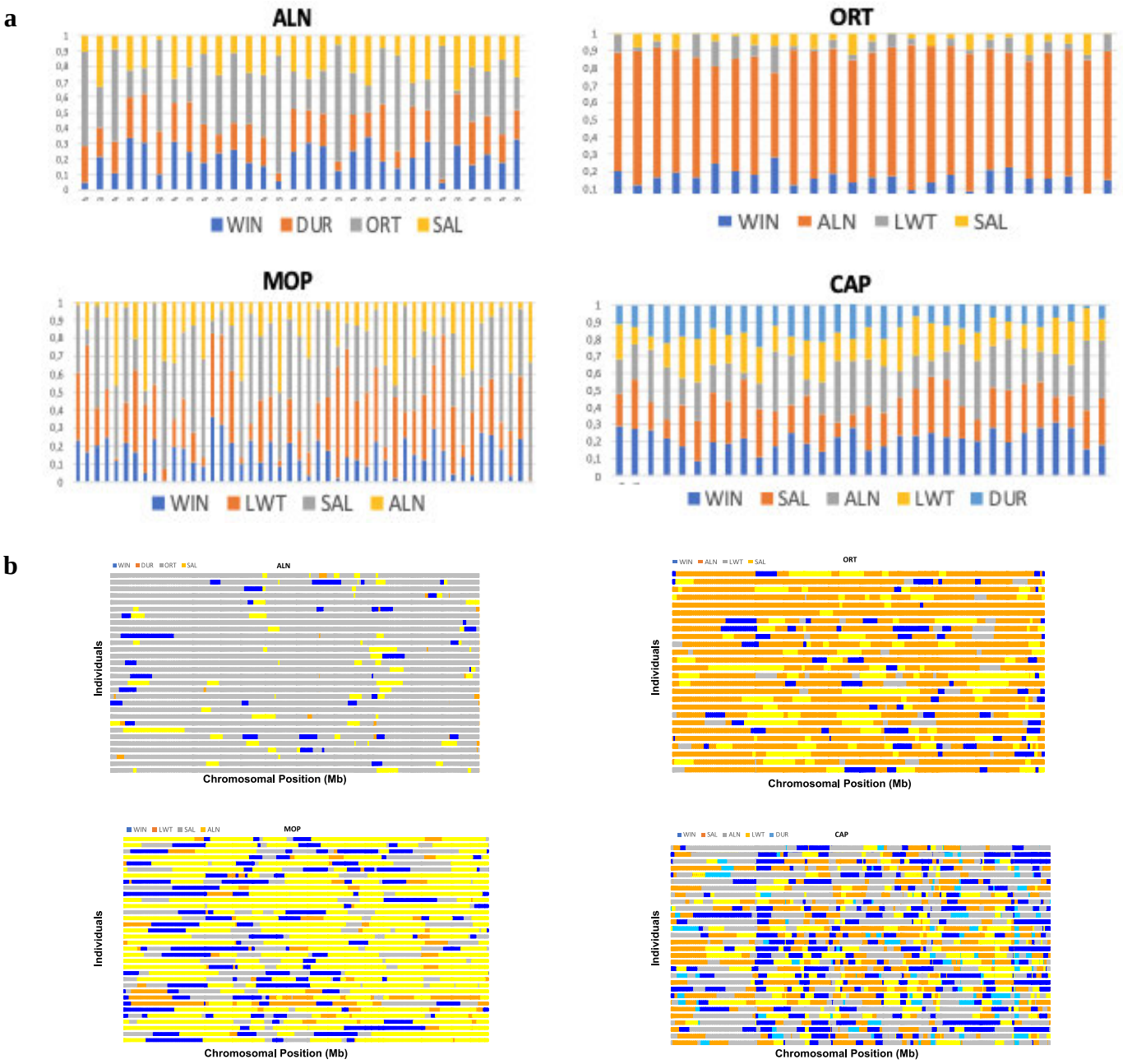
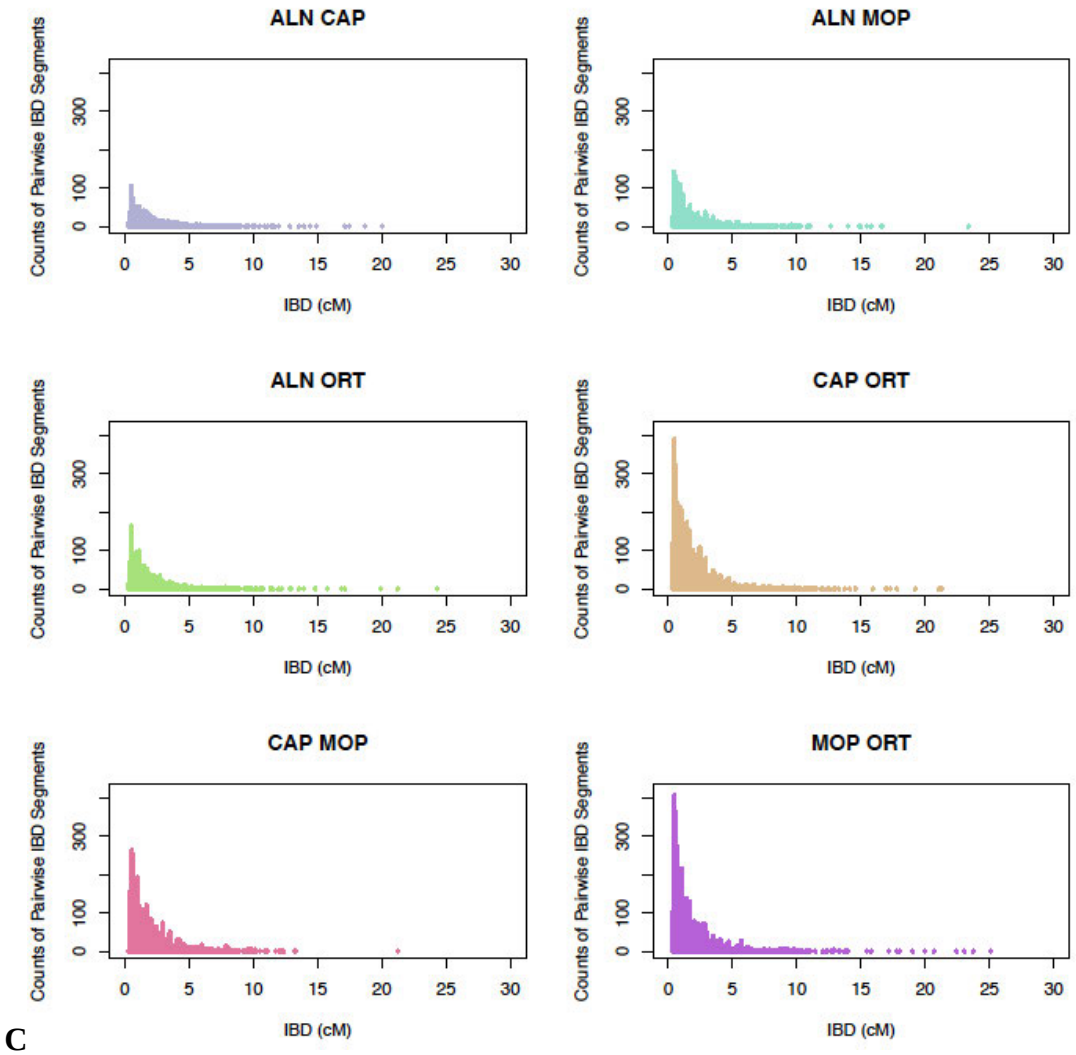


Figure 5.3: a) Proportions of ancestry from assuming six ancestral populations in village individual populations b) Genome admixture simplified using PCADMIX for village individuals' haplotypes based on 6 ancestral populations

2660

5.4.6 Pairwise IBD detection

Figure 5.4 shows the distribution shared IBD segments between pairs of populations. MOP_ORT populations had 407 shared IBD segments as well as the longest shared IBD segments of up to 25.11 cM. The lowest number of shared IBD segments were from ALN_CAP at 7 638. CAP_ORT individuals shared the most IBD segments (21 190) followed by MOP_ORT (20 486) as illustrated in Table 5.4. Most segments were found in chromosome 1 while most of the segments' distribution were short (< 1 cM). ALN_CAP, CAP_MOP and



C

Figure 5.4 Distribution of short and long segments that are identical by descent (IBD) for pairs of individuals from populations

MOP_ORT IBD segments were predominantly distributed on chromosome 7 whereas ALN_MOP, ALN_ORT and CAP_ORT had the greatest shared segments on chromosomes 2, 11 and 4 respectively. Chromosome 18 had the least shared IBD segments across all populations. The number of genes within shared IBD segments ranged from 13 751 (ALN_ORT) to 13 888 (ALN_MOP) showing no major differences (Table 5.4). On chromosome 6, genes such as *POU3F1, FHL3* and *ZNF382* were observed in all population pairs. Meat quality gene *DECR1* was also observed in chromosome 4 for all population pairs. Feed efficiency (*DNAJC15*) and fertility (*EPSTI1*) genes were observed on chromosome 11.

Table 5.4 Chromosome genome sharing for population pairs

Populations	Short Segments	Long Segments	No Genes
ALN_CAP	3 242	4 396	13 751
ALN_MOP	3 287	4 362	13 888
ALN_ORT	3 414	4 921	13 876
CAP_MOP	6 873	9 368	13 885
CAP_ORT	8 503	12 687	13 879
MOP_ORT	8 707	11 779	13 878

5.4.7 GO-enrichment analysis

Supplementary Tables 5.3-5.8 shows the detailed analyses performed on DAVID for functional enrichment ($P < 0.01$) for the GO of the pairwise village populations. The total number of the significant biological GO terms were 75 (ALN_CAP, CAP_MOP and MOP_ORT); 77 (ALN_MOP) and 72 (ALN_ORT and 74 (CAP_ORT). 46 unique GO terms were observed from ALN_CAP, 44 (ALN_ORT) and 47 (ALN_MOP, CAP_MOP, CAP_ORT, MOP_ORT) (Table 5.5). Among the highly significant biological process identified for all population pairs, were those related to protein phosphorylation (GO:0006468), Transcription DNA-template (GO:0006351), intracellular protein transport (GO:0006886), cell differentiation (GO:0030154). These GO terms were associated with *P-values* ranging from 1.16E-03 to 3.71E-07 (Table 5.5). Table 5.5 is based on the *P-values* representing the proportion of specified GO term within the whole Uniprot protein annotation database calculated by REVIGO (Supek *et al*., 2011) as seen on the detailed Supplementary Table 5.9. The GO, transcription, DNA-

template had the most genes at 221. Intracellular signal transduction (GO:0035556) was only relevant for all populations with 157 genes related to it.

Table 5.5 The top enriched (*P < 0.01*) Gene Ontology (GO) terms relevant to the pairwise for biological process

GO term	Description	Genes	*P-Value*
ALN_CAP GO identified (46)			
GO:0006468	Protein phosphorylation	67	6.16E-04
GO:0035556	Intracellular signal transduction	157	4.01E-03
GO:0006351	Transcription, DNA-templated	219	9.49E-04
GO:0006886	Intracellular protein transport	119	3.08E-04
GO:0030154	Cell differentiation	97	1.48E-03
ALN_MOP GO identified (47)			
GO:0006468	Protein phosphorylation	67	9.21E-04
GO:0006351	Transcription, DNA-templated	221	8.46E-04
GO:0006886	Intracellular protein transport	120	2.75E-04
GO:0030154	Cell differentiation	98	1.20E-03
ALN_ORT GO identified (44)			
GO:0006468	Protein phosphorylation	67	8.97E-04
GO:0006351	Transcription, DNA-templated	221	8.01E-04
GO:0006886	Intracellular protein transport	120	2.64E-04
GO:0030154	Cell differentiation	98	1.16E-03
CAP_MOP GO identified (47)			
GO:0006468	Protein phosphorylation	67	9.36E-04
GO:0006351	Transcription, DNA-templated	221	8.74E-04
GO:0006886	Intracellular protein transport	120	2.82E-04
GO:0030154	Cell differentiation	98	1.22E-03
CAP_ORT GO identified (47)			
GO:0006468	Protein phosphorylation	67	9.12E-04
GO:0006351	Transcription, DNA-templated	221	8.28E-04
GO:0006886	Intracellular protein transport	120	2.70E-04
GO:0030154	Cell differentiation	98	1.18E-03

Table 5.5 The top enriched (*P < 0.01*) Gene Ontology (GO) terms relevant to the pairwise for biological process

MOP_ORT GO identified (47)			
GO:0006468	Protein phosphorylation	67	9.12E-04
GO:0006351	Transcription, DNA-templated	164	3.71E-07
GO:0006886	Intracellular protein transport	120	2.70E-04
GO:0030154	Cell differentiation	98	1.18E-03

5.5 Discussion

The history of South African pig populations dates back to the arrival of the first explorer and settlers from Europe that along with them brought pig breeds from around the world (Swart *et al*., 2010). In this study, the ancestry and history of South African pig populations were constructed using a population structure approach similar to that reported in previous studies in livestock species (Wang *et al*., 2011; Hulsegge *et al*., 2019; Liang *et al*., 2019). The study hypothesised that South African pig populations share common ancestry, and that Large White and South African Landrace commercial breeds are descents of European ancestry.

Based on the results from Chapter 3, the genomic composition of South African pig population was influenced by the genetic background and the production system. In line with this, the indigenous breeds of Windsnyer and Kolbroek, the commercial breeds of South African Landrace, Duroc and Large White as well as the Vietnamese Potbelly and Wild Boar, were all assigned to single genomic clusters demonstrating that they were pure breeds with little to no admixture at $K = 10$. *Frappe* results additionally supported this by giving estimations ranging from 92% to 100% and assigning these populations to their own genomic clusters. Swart *et al*. (2010) also reported individual genetic clusters for South African Landrace, Large White, Duroc and Kolbroek populations based on microsatellite analysis. The village populations on the other hand were assigned to several clusters that implied crossbreeding amongst themselves, as well as with commercial and indigenous breeds. Admixture of village pigs with commercial breeds was also reported in pigs from Ghana (Osei-Amponsah *et al.*, 2017) and Kenya (Mujibi *et al*., 2018). The analysis pursued in this study were meant to further investigate the genetic

connectedness of the pig population by interrogating the underlying ancestry as well as the evolutionary forces behind the admixture in South African pig populations.

At M = 2, the commercial South African Duroc shared ancestry with the indigenous breed, Kolbroek. The shared ancestry of Duroc and Kolbroek can be explained by the fact that the Iberian pigs share an ancestry with American breeds (Lee *et al*. 2020). Duroc is originally an American pig breed. Such shared ancestry across different continents were also demonstrated in other countries (Vaughan 1950; Alves *et al*. 2009; Ravidatti *et al*. 2021). For example, the Duroc-Jersey has ancestry contribution of the Iberian pigs from Portugal and Spain (Vaughan 1950; Alves *et al*., 2009) and the American Criolli (Creole) which is a descendant of Iberian ancestry was used in the development of the Duroc in the 1800's (Ravidatti *et al*., 2021). Chen *et al*. (2017) revealed that the Duroc and Large White were crossed to form a White Duroc.

The gene flow at M = 2 of Warthog and Bush Pig to Kolbroek and Duroc demonstrates the genetic history of these Kolbroek and Duroc pig breeds. In Chapter 4, Kolbroek and Duroc shared considerable haplotype blocks levels further suggesting that there are similar selection pressures that these populations share probably for traits they have in common. The persistence phase between these two populations was also high at r = 0.76±0.79 at distances less than 1 kb, showing the genetic relatedness. The genetic connections between Kolbroek and Duroc implies that the two breeds could share genomic tools for selection and improvement such that SNPs designed for Duroc could also be used as a tool for selection for Kolbroek pigs. Paper review by Misztal *et al*. (2022) suggested that across breed genomic evaluations could be implemented in closely related breeds.

The Large White also contributed the genetic pool of the South African Landrace. While there was a gene flow of the South African Landrace to Duroc at M = 8, Large White breed seemed to have shaped both the South African Landrace and Duroc. Both the South African Landrace and Duroc indicated a strong linkage disequilibrium (LD) of 0.36±0.34 and 0.47±0.38 on chromosome 14 (Chapter 4). This demonstrates that these populations which are reared in an intensive environment are selected on the same important phenotypic traits. Chapter 4 observed a high correlation of 0.73 at 0-1 kb and a total of 252 shared haplotype blocks between SAL and Duroc. Additionally, a moderate F_{ST} value of 0.194 demonstrated their closer genetic relationship (Chapter 3).

Gene flow was observed between Wild Boar to Warthog and Bush Pigs at $M = 3$. Wild Boars (*Sus scrofa*) are known to be founding members of the domestic pigs (*Sus scrofa domestica*). However, Asia and Europe are domestication centres for pigs (Yu *et al*., 2013). Three major clusters of the Asian Vietnamese Potbelly, Warthog and Bush Pig and Wild Boar clustering with the rest of the domestic pigs were found in Chapter 3. According to Lee *et al*. (2020), Wild Boar, Warthog and Bush Pig share SLA-6 haplotypes whereas Kim *et al*. (2002) noticed the genetic distance of 0.0168 between European and Asian breed from which they inferred that European Wild Boars are ancestors of both the Asian and European pigs. Wild Boar, Warthog and Bush Pig share a common ancestor as they are all from the *Suidae* family. *Suidae* diversified into multiple species from their common ancestor in Southeast Asia, (Gongora *et al*., 2011). These species further diverged to other lineages. Based on dental and cranial records, Thenius (1970) and Cook (1978) suggested that the Wild Boar was related to the Bush Pig, though this was overturned by Gongora *et al*. (2011). Interestingly, Venter *et al*. (2016) reported on Bush Pigs interbreeding with Wild Boars in South-Eastern KwaZulu-Natal. This is not farfetched with these populations being in an environment where there is no controlled breeding. In 1942, Thomas and Kolbe, related on Bush Pig and domestic pigs inter-breeding while Notter *et al*. (2013) have reported fertile progenies of crosses between the domestic and the Bush Pigs.

Worldwide PCA (Chapter 3), showed clustering using Burgos-Paz *et al*. (2013) genotypes showed the Vietnamese Potbelly sharing a cluster with the Chinese populations. The sub clustering of Vietnamese Potbelly with Warthog and Bush Pig populations might be explained by some introgression during their breeding history, and this was also evidence in $K = 2$ and $K = 3$ in Chapter 3. On both the phylogenetic tree and Treemix, the Vietnamese Potbelly and Bush Pigs are clustering together. Migration edges at $M = 4$ supports the assumption of White and Harris (1977) that the Bush Pigs originate from Asia. According to Sawyer and Camp (2020), the Vietnamese Potbelly was developed in the 1960s as dwarf pigs. As with the Vietnamese Potbelly used in the study, such breeds are now kept in zoos for conservation (Bosse *et al*., 2015). Based on the proportions of ancestry and the migration gene flow results of this study, it is evident that the pig populations are as a result of extensive crossbreeding in South Africa with influence from both European and Asian pigs.

In this study, a gene flow from Vietnamese Potbelly to Large White and South African Landrace ($M = 10$) was observed. Pham *et al*. (2014) reported on European introgression into Vietnam pigs while Reimer *et al*. (2018) observed the Vietnamese Potbelly pig to carry the South

Chinese haplotype and a crossbreeding history with European breeds (Glodek & Oldigs, 1981; FAO, 2005). South-Eastern Asian Wild Boar pigs of which Vietnam is geographically located, spread to Europe approximately 1 million ages ago (Giuffra *et al*., 2000; Groenen *et al*., 2012). This might explain the significance of the Vietnamese Potbelly gene flow to other domestic pigs in this study. The abandonment of the Vietnamese Potbelly pigs as pets in Spain has also led to them breeding with wild and/or feral pig populations (Delibes-Mateos & Delibes, 2013).

The local African pig, commonly known as the Kolbroek has its ancestries from the Iberian pig (Inter Africa Bureau for Animal Resources, 2015). This includes the Mukota, Somo, Busia, Bakosi, Ashanti, Bush Pigsand West African Dwarf pigs from Zimbabwe, Mali, Kenya, Gabon, Ghana, Togo and Nigeria respectively, all originating from the Iberian amid the third to seventh centuries (Bester & Kusel, 1998). In Herrero-Medrano *et al*. (2012) study, the Spanish Iberian and Berkshire pig breeds clustered together. In a study by Visser (2012), Kolbroek and Berkshire also shared phenotypic similarities, further showing a genetic contribution of the Berkshire to Kolbroek. Additionally, the White Chinese Hog and Black Siamese pigs were included in the establishment of the Berkshire (Tonder, 2019). This shows a pattern of introgression of European and Asian ancestries to the indigenous Kolbroek. The Berkshire is an English breed from Berkshire which was the epicentre for pig breeding in the 18th and 19th century (Tonder, 2019). The Iberian and Berkshire breeds had a large influence on the development of other breeds (Herrero-Medrano *et al*., 2012). Moreover, other African pigs have an Iberian ancestry like the Kolbroek and Mukota (Bester & Kusel, 1998; Plug & Badenhorst, 2001). Due to its phenotypic structure, Windsnyer is said to look like Mukota breed of Zimbabwe and have been portrayed as being closely related to each other by Holness (1991). Previous studies have reported South-East Asian influence on Mukota pigs (Olalde *et al*., 2015; Adeola *et al*., 2017) further explaining the gene flow of Vietnamese Potbelly to Windsnyer (*M* = 7) observed in this study. This was further supported by Adeola *et al*. (2017) who reported on Mukota similarities to a Chinese lard pig.

Despite the geographical distance barrier of about 1 300 km apart for Eastern Cape (O.R. Tambo and Alfred Nzo) and Limpopo (Mopani and Capricorn), results from this study suggest shared genetic materials amongst these populations. These populations shared ancestries with the Windsnyer, South African Landrace and Duroc. Large White breed was a popular common ancestry in O.R. Tambo, Capricorn and Mopani. The high appearance of the South African Landrace and Large White in the village pigs confirms first-hand information from Limpopo

farmers of sourcing their breeding from the nearby commercial farms. Farmers might also be selling and exchanging pigs between them. In studies conducted by Madzimure *et al*., 2012; Matabane *et al*., 2015 and Gcumisa *et al*., 2016, village pig farmers either sold or exchanged their pigs with the neighbours or other villages. Windsnyer and South African Landrace were common ancestries in all the village populations. Windsnyer population was well represented in the ancestries of all the village populations and reveals percentages ranging from 16% to 22%. Mujibi *et al*. (2018) also observed the admixture in Kenyan village pigs with Landrace, Large White and Wild Boar populations. This was also clearly indicated by the gene flow of the Wild Boar to the O.R. Tambo village population at $M = 1$ and further evidenced by Wild Boar and the O.R. Tambo village pigs being in the same PCA clusters of South African populations alone and that with international populations (Chapter 3). Wild Boars are descents stocks of the modern pigs while surveys conducted by Matabane *et al*. (2015) and Gcumisa *et al*. (2016) has demonstrated the appetite of commercial breeds in the villages. Capricorn displayed the highest level of admixture of all the village populations with more than 10% of Windsnyer (28%), South African Landrace (23%), Alfred Nzo (14%) and Large White (15%) while Kolbroek was 7%. Introgression of commercial breeds into the villages may be a result of historic improvement strategies with high performing breeds and indiscriminate breeding by farmers. This was confirmed first-hand by the Limpopo and Eastern Cape farmers included in this study, as they verified to use commercial breeds to improve their animals (unpublished). Furthermore, previous studies in Eastern Cape (Madzimure *et al*., 2012), KwaZulu-Natal (Gcumisa *et al*., 2016), Mpumalanga (Munzhelele *et al*., 2017) and Limpopo (Phogole, 2017) also noted that most of their respondents in the villages kept pig crosses. The 5^{th}, 6^{th} and 9^{th} migration events indicated gene flow from Vietnamese Potbelly to Mopani, Capricorn and O.R. Tambo pigs, respectively.

Studies by (Madzimure *et al*., 2012; Matabane *et al*., 2015; Gcumisa *et al*., 2016; Munzhelele *et al*., 2017; Phogole, 2017) showed that village farmers have different preference in breeds in order to improve their livestock. Although admixture occurred between all of the village populations, it appears that Alfred Nzo pigs were less admixed with other populations when compared to the other village populations. Chapter 3 also indicated that the Alfred Nzo pigs represent a unique distinguished population. As this population represent a distinct genetic resource, this population therefore has the potential to be classified as specific breed. This shows distinct and unique genetic resources with a potential of classifying this population as a breed.

This study further examined whether individual pigs had a shared IBD of two alleles or more originating from a common ancestor. Shared regions could be due to gene flow and shared ancestries (Browning & Browning, 2012). Previous results (Chapter 4) showed the village populations sharing the most haplotype blocks amongst themselves and therefore linked ancestry. The sharing of shorter IBD segments which are associated with ancient ancestry for CAP_ORT and MOP_ORT compared to other pairs suggests that these populations are distantly related. Short segments result from recombination events that interrupts long genomic segments predating continental migrations (Broman *et al*., 1999). Therefore, the low number of short segments in the study compared to long IBD segments (> 1 cM) supports the common ancient ancestry of the South African pig populations. Short segments are associated with high inbreeding levels (Kardos *et al*., 2017) and in Chapter 3, inbreeding levels in villages being low to moderate (0.056-0.198). The unequal distribution of segments seen in the populations were also noted in the study by Bosse *et al*. (2012) and presumed to have been caused by the recombination landscape of the pig family that is highly heterogenous.

Shared IBD segments are an integral component to the study of genetic diversity, effective population size, inbreeding and selection processes. In this Chapter, genomic regions harbouring shared IBD segment were annotated. Genes observed in the IBD segments were also observed in previous Chapters 3 and 4. The *DECR1* gene on chromosome 4 that was initially detected within the shared haplotype blocks (Chapter 4) is related to meat quality and growth rate traits. Furthermore, genes observed by per marker pairwise F_{ST} analysis were also found in this study. These were *DNAJC15* related to feed efficiency which was found in Villages vs. Kolbroek populations. Gene associated with fertility traits (*SBF*) on chromosome 4 was observed in this study as well as F_{ST} for Villages and Wild Boar (Chapter 3) and on high LD regions (Chapter 4). Shared IBD regions also revealed genes such as S*WAP70* (chromosome 2) associated with disease resistance that were initially reported in Chapter 3 as those within highly differentiated regions (F_{ST} > 0.8.) between Villages and Wild Boars. Genes harboured within shared IBD regions in village populations were associated with important QTL regions. The observation of the similar genes and QTL regions using different methods reported in Chapters 3, 4 and this Chapter 5 confirms (i) the evolutionary history of the South African pig population and (ii) the important traits that under either natural or artificial selection in the pig sector.

GO pathway analysis revealed that they were significantly enriched in GO terms with *p-value* < 0.01 including protein phosphorylation (GO:0006468), intracellular signal (GO:0035556), transduction, transcription, DNA-templated (GO:0006351), intracellular protein transport (GO:0006886) and cell differentiation (GO:0030154) for all population pairs. A total of 67 genes observed in GO:0006468 were associated with high fat high fibre diet (Gondret *et al*., 2016) and feed intake (Liu *et al*., 2016; Messad *et al*., 2019). GO:0035556 (Gondret *et al*., 2016) and GO:0006886 (Krombeen *et al*., 2017; Dobrzyn *et al*., 2019) were shown to have been associated with fertility. GO:0006351 was observed to be associated with muscle growth and meat quality traits (Ovilo *et al*., 2014; Vigors *et al*., 2019). In addition, 120 genes significantly enriched for high fat content and high fibre diets (Gondret *et al*., 2016) and disease susceptibility (Zhu *et al*., 2019; Jiao *et al*., 2021) were observed in GO:0006886. In the previous Chapter 3, SNPs that differentiated village pigs were associated with disease resistance, an adaptive mechanism to allow pigs to perform in harsh disease infested environments.

Ayuso *et al*., 2015; Zhao *et al*., 2019 observed the GO:0030154 which are associated with deposition of fat and is significant to pigs as it contributes to traits of economic important such as meat quality, reproductive performance, muscle growth and fattening efficiency (Zhao *et al*., 2019). In the GO0030154 gene ontology, *EHF* gene was also observed. This gene is associated with scrotal hernia which is a congenital defect in pigs caused by environmental and genetic factors (Du *et al*., 2009). Whilst Zhao *et al*. (2019) noted that improving sanitation and hygiene conditions could lead to reduced chance of getting this condition in village pigs, the presence of the *EHF* gene within shared IBD segments signifies the role of genetics in the occurrence of scrotal hernia in pigs. Scrotal hernia has a negative effect on pigs on growth and increases the chance of morbidity (Nowacka-Woszuk, 2021). Pig farmers in the villages lack resources and the possibility of selecting against this genetic disorder might provide resource limited farmers with an alternative solution. Currently, there are also no breeding programs to facilitate genetic management of scrotal hernia defects in herds. *FOXN2* gene was also part of the GO:0030154 term and is associated with meat quality traits in pigs (Lee *et al*., 2014).

5.5 Conclusion

South African pig populations tend to be highly introgressed predominantly by European and Asian ancestries. Shared ancestry and introgression with wild pigs was also observed. Further examination in future requires more samples including those of Wild Boars and Bush Pigs, that will help give a clearer picture on the introgression of the wild pig populations in domestic

populations. Crossbreeding has been the basis in the formation of the modern pig populations in search of desirable traits from different parts of the world. Indiscriminate crossbreeding is evident in village pig populations leading to admixed populations. The village population category, presented as a unique and distinct population feeding into the genomic architecture of other breeds and populations thus implicating the effects of natural selection on IBD patterns. The short segments distribution in villages are in line with the previously observed inbreeding levels, low and short-spanned LD and low haplotype blocks prevalence and diversity. Overall, the study demonstrated the role of domestication, admixture, selection and adaptation in shaping the phenotype and genomic structure that exists in the South African pig populations.

CHAPTER 6: IDENTIFICATION OF SIGNATURES OF SELECTION OF AND AFFECTED QTLS IN PIGS THROUGH GENOMIC SCANS USING DIFFERENT METHODS

6.1 Abstract

South Africa has diverse pig populations ranging from intensively raised and artificially selected commercial breeds to naturally select indigenous and village pigs that are reared under low input extensive production systems. Natural and artificial selection results in phenotypic variations such as reproduction, growth, body composition, disease resistance and coat colour prevalent in pig populations. This study focused on investigating genomic regions affected by different selection pressures in the South African pigs sampled from diverse genetic backgrounds and production systems. To uncover signature of selection, 234 pigs from the villages, commercial, indigenous, Asian and wild population were genotyped using the Porcine SNP60K bead chip. Three different methods were used to identify signatures of selection. We implemented the *iHS, XP-EHH* and *HapFLK* techniques to detect potential regions associated with different QTLs. *iHS* revealed population specific signatures. In the Alfred Nzo and Mopani villages *TRIM44* gene on chromosome 2 linked to spinal curvature was identified. Intensive raised Large White, Kolbroek and Windsnyer demonstrated a strong link to meat and carcass quality traits. Chromosome 2 of the Wild Boar found *F2* gene linked to several QTLs. Importantly, village populations showed more genomic regions using the *iHS* within each population. Our findings revealed candidates' genes related to the animals' phenotypic variation using different methods.

Keywords: *Genetic signatures, iHS, XP-EHH, HapFLK, pigs, gene enrichment analyses*

6.2 Introduction

In January 2020, the world population of pigs was estimated at 677.6 million (Statista, 2021). Pigs represents an important livestock species worldwide and are important for livelihoods, food security and economic growth. In developing countries, pigs are reared under harsh environments while providing proteins for resource limited households (Lunney, 2007). Besides providing proteins for humans, pigs are also used as model animals to research on human diseases (Lunney, 2007).

Among the wild hog species (also referred to as wild pigs) are the Warthog (*Phacochoreus africanus*), pig-deer (*Babyrousa babyrussa*) and the pygmy hog (*Porcula salvania*), only Wild Boar (*Sus scrofa*) has been domesticated (Giuffra *et al*., 2000; Li *et al*., 2012). Changes in the phenotypic characteristics between domestic and wild pigs are highly noticeable and were driven by natural and artificial selection (Gurgul *et al*., 2018). Independent domestication events from local Wild Boar (*Sus scrofa*) in Europe and Asia, gave rise to European and East Asian pigs (Larson *et al*., 2005; Wu *et al*., 2007). As a result of strong artificial selection, there is a considerable genetic distance between European and Asian domestic pigs (Giuffra *et al*., 2000; Larson *et al*., 2005; Wu *et al*. 2007). While commercial lines of European pigs are characterised by extended body length and lean growth, East Asian domestic pigs have good fat deposition and high reproductive performance (Zhang, 2021; van Laere *et al*., 2003; Rubin *et al*. 2012). In the absence of a reproductive barrier between East Asian and European domestic pigs, hybridization between East Asian and European and later American pig breeds has been successfully used to increase pig production (Zhang, 1986; Guo *et al*., 2009; Li *et al*. 2009b). Previous studies clearly demonstrated a hybrid origin of the European Large White breed with Asian pigs (Watanabe *et al*., 1986). Hybridization of domesticated pigs with Wild Boars on European farms has also been used to increase reproduction and genetic diversity in inbred pig lines (Scandura *et al*., 2011). Village and smallholder pigs that are farmed predominantly under free-range production system, allows for gene flow, intogression and hybridization to occur with wild pigs (*e.g*., Warthogs, Wild Boars and Bush Pigs) (Nyafu *et al*., 2009; Amar *et al*., 2021). Although hybridization between pigs and wild pigs can increase production, these events may also have a negative impact on pig production. (Molotsi *et al*., 2021). It is suggested, for example that the outbreak of Classical Swine Fever (CSF) is related to wild and domestic pigs mixing in free-range production system (NAFU, 2007; Penrith *et al*., 2011).

Commercial pig breeds from Europe and America were introduced in South Africa for commercial farming in the 1600s by European settlers (Krige, 1988; Blench & MacDonald, 2000; Swart *et al*., 2010). These commercial pig breeds of Large White, Landrace and Duroc, that are commercially farmed in South Africa are known for their high performance (*e.g*., litter size, high growth rate, meat and carcass quality) (Visser, 2004; Visser *et al*., 2014; Munzhelele, 2015). In addition to European pig breeds, indigenous South African pig breeds are also used by commercial farmers (Mokoele *et al*., 2014; Gcumisa *et al*., 2016; Munzhelele *et al*., 2017; Molotsi *et al*., 2021). These indigenous breeds have developed an adaptation to the harsh South African environmental conditions. For example, the indigenous Windsnyer has longer black

hair and thinner epidermis for increased heat tolerance that will shield them against extreme climatic conditions (Madzimure *et al.*, 2012; Moyo *et al.*, 2018).

South Africa has 4 000 producers in the commercial sector and 1 400 million smallholder farmers (BFAP, 2021). While commercial pig farmer's practise controlled breeding and intense artificial selection for key production traits, the landscape of the small-scale and village farming is characterised by poorly organised and indiscriminate crossbreeding (Madzimure *et al.*, 2012). Pig farmers in the rural are increasingly shifting towards the use of commercial exotic breeds and established indigenous breeds of the Kolbroek and Windsnyer (Mokoele *et al.*, 2014; Gcumisa *et al.*, 2016; Munzhelele *et al.*, 2017; Molotsi *et al.*, 2021). The Kolbroek and Windsnyer are hardy and adapted to survive under harsh local conditions (Halimani *et al.*, 2010; Swart *et al.*, 2010; Madzimure *et al.*, 2012; Mokoele *et al.*, 2014), while crossbreeding is aimed at improving the performance and production of populations (Esfandyari *et al.*, 2015). The indigenous breeds (*e.g.*, Kolbroek and Windsnyer) and some non-descript village populations display high tolerance and/or resistance to disease and parasites (Mohlatlole *et al.*, 2013; Ncobela, 2017). The positive characteristics of indigenous and local populations (*e.g.*, heat tolerance and disease resistance) are valuable and need to be characterised and conserved as they are also important to the livelihood of subsistence and small-scale farmers.

Adaptation (natural selection), domestication (artificial selection), as well as breed development creates signatures of selection in genomic regions of populations (Raudsepp *et al.*, 2019). Signatures of selection have been identified in pig populations associated with important traits, such as adaptation to high altitudes (Ai *et al.*, 2013), muscle growth (Wang *et al.*, 2014) and body size (Rubin *et al.*, 2012). Genomic sequences of domestic and wild pigs have been observed to be predominantly similar, except in regions under strong selection pressure (Rubin *et al.* 2012). Various authors have reported on selection in domestic pigs for resistance and tolerance (Guy *et al.*, 2012) and productivity (Hermesch *et al.*, 2015). Fulgione *et al.* (2016) welcomed the idea of applying artificial selection for Wild Boars in order to increase fitness.

Information of signatures of selection is valuable and can be used in management strategies to improve production and adaptability. To date there has not been any study that investigated signatures of selection in South African pig populations. The analysis done in this study in Chapters 3, 4 and 5, however point towards a population that has been shaped by complex evolutionary forces including domestication, continuous interactions between domestic and

wild populations. This includes natural and artificial selection under the different production systems as a result to adapt and survive under the prevailing climatic conditions, low input production systems and diseases. The aim of this Chapter was therefore to apply high resolution in-depth analysis to identify and characterize genomic regions that display signatures of selection in the South African commercial, villages, indigenous, wild and Vietnamese Potbelly pig populations using the SNP60K Porcine genotypes generated and used in previous chapters. Three different population genetic methods of detecting signatures of selections were used. To identify recent signatures of selection, (i) the integrated haplotype score (*iHS*) that identifies positive selection signatures within population (Voight *et al*., 2006) and (ii) the cross-population extended haplotype homozygosity (*XP-EHH*) that calculates differences between two populations (Sabeti *et al*., 2006) were used. The last method, *HapFLK* method (Fariello *et al*., 2013) that accounts for migration and bottlenecks in populations and populations being structured by haplotypes (Bonhomme *et al*., 2010) was used to identify selection signatures across multi-populations. The identified signatures of selection were annotated in order to facilitate inferences on genes and associated genetic mechanisms linked to traits of economic importance in the South African and global context.

6.3 Materials and Methods

6.3.1 Samples, data and quality control

Porcine SNP60K genotypes from 234 animals from the villages, commercial, indigenous, wild and an Asian pig breed as described in Chapter 3 were used. This included 60 pigs from commercial farms represented by the Large White (LWT), South African Landrace (SAL) and Duroc (DUR) breeds, as well as 40 indigenous pigs represented by Kolbroek (KOL) and Windsnyer (WIN). It also included 91 pigs obtained from villages in the Alfred Nzo (ALN), O.R. Tambo (ORT), Capricorn (CAP) and Mopani (MOP) districts. In addition, 5 Vietnamese Potbelly pigs (VIT) from the Johannesburg Zoo and 38 wild pigs represented by Warthog (WAT), Wild Boar (WBO) and Bush Pig (BSP) were collected from various game reserves.

Markers with a call rate lower than 85% and not physically mapped to the *Sus scrofa* 11.2 genome assembly were discarded using the Golden Helix SNP Variation Suite (SVS) *version* 8.8.1. Markers with a minor allele frequency (MAF) lower than 2% and those that deviated from Hardy-Weinberg Equilibrium ($P < 0.0001$) were also excluded. After QC, 211 individuals

and 27 740 SNPs remained for analysis. A further 318 SNPs were identified to have the same position and therefore filtered out for duplication. BEAGLE (*version* 5.1) was used to phase the autosomal genome on 30 iterations of the phasing algorithm on a chromosomal region of 5 Mb and sample haplotype pairs for each individual per iteration (Browning & Browning, 2007 for the data set used for *XP-EHH*, *iHS* and *HapFLK* analysis.

6.3.2 Detection of signatures using iHS

iHS was used screen for non-overlapping regions within a population under positive selection. Plink 1.9 was used to exclude duplicate SNPs and recode all genotypes using --allele1234 script. Plink format (map and ped files) were converted to fastPHASE format using --recode fastphase script. This generated a fastphase.inp file which was used in the fastPHASE software. The fastPHASE 1.4.8 (Scheet & Stephans, 2006) software was used to estimate missing genotypes and unobserved haplotypes from unphased data for each chromosome. This then created an input file fastphase_hapguess_switch.out file which was used to calculate *iHS*. Once phasing was done, *iHS* was calculated on individual sites for possible signatures using rehh (Gauter & Vitalls, 2012) software in the R environment. The absolute unstandardized *iHS* (un*iHS*) is identified as log ratio iHH^A ancestral to derived iHH^D allele for each SNP (Voight *et al*., 2006). A cut off of $-\log_{10}$ (*p-value*) = 3 (*p value* ≤ 0.0001) was set as the threshold for considering the *iHS* score to be significant under selection with at least five SNPs $\leq$ 100 kb. Manhattan plots were generated in the R package qqman.

6.3.3 Detection of signatures using XP-EHH

Selective sweeps between populations were detected using the *XP-EHH*. *XP-EHH* pairwise makes it possible to find selected regions using genetic distance between adjacent SNPs based on the of *EHH* model (Sabeti *et al*., 2002). Cross population *EHH* (*XP-EHH*) statistic, which is similar to *Rsb* and compares one population to the other using haplotypes (Sabeti *et al*., 2007). The difference being that *Rsb* is based on *iES* instead of *inES*. Sabeti *et al*. (2007) defined *XP-EHH* as standardised (un*XP-EHH*) identified as mean (un*XP-EHH*) and standard deviation (un*XP-EHH*) for each given SNP. The argument was set to, pop1 identified as p_{XP-EHH}^{right} relative to pop2 as p_{XP-EHH}^{left}.to find regions associated with each population. *XP-EHH* used the same phased file as *iHS* and therefore *iHS* was firstly calculated for each population using rehh package (Gauter & Vitalls, 2012) on R. The significance p-value ($-\log_{10}$ [2ϕ -| *iHS* |]) = 3 (p ≤

0.001) was set as threshold for considering the *XP-EHH* score to be significantly under selection with at least five SNPs ≤ 100 kb. Manhattan plots were generated in the R package qqman.

6.3.4 Detection of signatures using HapFLK

To reveal genetic differentiations in genomic regions subjected to selection from multiple populations, *HapFLK* method was employed. This test accounts for haplotype structure of the population whilst it undertakes polymorphic SNPs in ancestral populations. Reynolds distances was calculated according to *HapFLK* 1.3.0 software (https://forge-dga.jouy.inra.fr/projects/hapflk) and then converted to a kinship matrix with HapFLK package on the RStudio. FastPHASE cross validation procedure was used to determine haplotype diversity (Scheet & Stephens, 2006). A total of 20 clusters with 30 maximization iterations on per chromosome basis was used to calculate *HapFLK* statistic. A standard normal distribution was calculated at each SNP using *p-values*. Selected regions were identified using P-value < 0.10 (Storey & Tibshirani, 2013). For this study, the indigenous breed Kolbroek, was used as outgroup.

6.3.5 Annotation and function analyses of identified genomic regions

The candidate regions identified using the three different methods (*i.e.*, *iHS*, *XP-EHH*, and *HapFLK*) were annotated for genes and QTLs and functional pathways. For this purpose, the BioMart on the Ensembl gene database website was used to annotate genes at particular genome coordinates for all selected regions (release 89). Candidate regions search, and identification was performed within 1 Mb to the left and right of statistically significant SNPs. The current pig genome *Sus scrofa* 11.1 assembly was used to extract gene symbols. Web-based Panther was employed for functional and pathway enrichment analysis. False Discovery Rate (FDR) < 0.10 was used to assess the significance of enriched pathways. Pig QTL (Release 46) database was used to align candidate genes to available QTL.

6.4 Results

6.4.1 Detection of signatures within a population using iHS

After quality check, a total of 27 422 SNPs were retained for further analysis. The *iHS* method used to detect positive selection within a population, identified potential genomic regions in all 13 populations included in this study (Table 6.1). However, the number of regions differed greatly between the different populations, ranging from 87 for CAP to 4 for BSP (Supplementary Table 6.1). The number of regions that displayed evidence of signatures of selection differed according to the management strategy used. Most selection regions were identified among the village pigs (ALN 40; ORT 71; CAP 87; MOP 68), followed by commercial pigs (LWT 34; SAL 31; DUR 12) while the indigenous pig population (KOL 17; WIN 22) had the fewest regions (Supplementary Table 6.1). For the wild pigs, WBO had the most regions at 32, followed by WAT at 10 while BSP had only 4 regions.

Table 6.1 Within population list of genomic regions under selection within and candidates' genes detected using the *iHS* method

Pop	Chr	Start Length	End Length	Gene name	QTLs
ALN	2	25,247,505	25,378,565	TRIM44	Spinal Curvature
	7	95,982,739	96,254,303	DPF3	Teat number
	8	31,921,587	32,301,975	APBB2	Stearic acid content
	13	29,753,308	29,775,371	*PTH1R*	Front leg conformation, Hind leg conformation, Hip structure
ORT	1	162,364,055	162,737,492	*NEDD4L*	Intramuscular fat content
	2	87,678,811	87,866,856	*DMGDH*	Litter weight piglets born alive
	8	80,290,715	80,669,404	*NR3C2*	Loin muscle, Teat number
	12	53,979,215	54,055,377	*PIK3R5*	Intramuscular fat content
	12	27,957,732	28,617,585	*CA10*	Hemaglobin
	12	44,617,499	44,660,271	*VTN*	Body depth, Drip loss, Hind leg conformation, Hip structure, pH 24 hr post-mortem (loin), ph 45 minutes post-mortem
MOP	4	71,045,333	71,318,030	NKAIN3	Body weight (birth)
	8	31,921,587	32,301,975	APBB2	Stearic acid content
	16	27,537,885	27,631,130	SELENOP	Meat colour a
	16	32,142,034	32,148,335	PELO	Obesity index, Teat number
	16	32,336,292	32,437,103	ITGA2	Body weight (5 weeks)
CAP	1	183,915,339	184,140,122	*SAMD4A*	Intramuscular fat content
	12	27,416,337	27,436,160	*NME1*	Conductivity 24 hours post-mortem (loin), Cooking loss, Loin muscle depth, Loin weight
	12	27,957,732	28,617,585	*CA10*	Hemoglobin
	14	101,123,381	101,254,725	*LIPA*	Average daily gain, Front leg weight, HDL/LDL ratio, Litter size, Loin muscle area, Monounsaturated fatty acid content, Oleic acid content, Skin thickness, Sperm concentration, Teat number
	14	103,992,037	104,106,576	*IDE*	Abdominal fat percentage, Age at puberty, Age at slaughter, Average backfat thickness, Average daily gain, Backfat at first rib, Backfat at last rib, Backfat between 3rd and 4th last ribs,

Table 6.1 Within population list of genomic regions under selection within and candidates' genes detected using the *iHS* method

Pop	Chr	Start Length	End Length	Gene name	QTLs
					Body depth, Body height, body length, Body weight (slaughter), Body weight (weaning), Body width, CD4-negative CD8-positive leukocyte percentage, CD4-positive CD8-negative leukocyte percentage, CD4-positive/CD8-positive leukocyte ratio, Carcass length, Carcass weight (hot), Conductivity 24 hours post-mortem (loin), Cooking loss, Daily feed intake, Days to 100 kg, Dressing percentage, Ear area, Fat androstenone level, Fat percentage in carcass, Fat-cuts percentage, Feed conversion ratio, Feed intake, Femur length, Front leg conformation, Front leg weight, Gait score (overall), Ham weight, Hemoglobin, Hind foot size, Hind leg conformation, Hip bone length, Hip structure, Humerus length, Intramuscular fat content, Lean meat percentage, Linoleic acid content, Litter size, Loin muscle area, Loin muscle depth, Loin weight, Lymphocyte number, Lymphocyte percentage, Marbling, Maternal infanticide, Mean corpuscular hemoglobin content, Mean corpuscular volume, Meat colour L*, Meat colour a*, Meat colour score, Meat colour-L, Meat to fat ratio, Monounsaturated fatty acid to polyunsaturated fatty acid ratio, Muscle moisture percentage, Number of visits to feeder per day, Number weaned, Oleic acid content, Oleic acid stearic acid ratio, PH for longissmus dorsi, PRRSV antibody titer, Palmitic acid content, Palmitoleic acid content, Polyunsaturated fatty acid content, Red blood cell count, Red cell distribution width, Rib shape, Salmonella shedding status, Shoulder weight, Side fat thickness, Sperm concentration, Sperm per ejaculate, Stearic acid content, Teat number, Teat number (maximum per side), Tibia length, Total number born alive, Triglyceride level, Ulna length, Vaccenic acid content, White blood cell number, Androstenone laboratory, Body mass index, cis-11-Eicosenoic acid content, Indole laboratory, Skatole laboratory
	14	105,036,770	105,044,765	*RBP4*	Litter size, Total number born alive
	14	12,437,515	12,563,544	*EXTL3*	Fat androstenone level
	8	76,482,022	76,699,351	*FBXW7*	Body mass index
	8	31,921,587	32,301,975	*APBB2*	Stearic acid content
LWT	2	108,255,930	108,536,881	*PAM*	Intramuscular fat content, Loin percentage, Maternal infanticide, Teat number
	7	26,860,140	26,990,017	*LRRC1*	Meat colour L
	9	48,120,802	48,286,278	*TECTA*	Backfat between 3rd and 4th last ribs
	13	177,365,717	179,013,542	*ROBO2*	Feed efficiency, Linolenic acid content
	14	69,200,215	70,938,204	*CTNNA3*	Teat number
DUR	6	137,595,524	138,010,444	*SLC44A5*	Diameter of muscle fibers
	7	27,389,874	27,895,570	*KHDRBS2*	Loin muscle area, Loin muscle depth, Teat number
KOL	4	61,628,299	61,716,546	*JPH1*	Intramuscular fat content
	14	79,352,396	80,106,258	*KCNMA1*	Meat colour b
WIN	5	18,718,158	18,723,843	*TARBP2*	Backfat between 3rd and 4th last ribs
	8	127,731,903	128,953,331	*CCSER1*	Backfat between 3rd and 4th last ribs, Intramuscular fat content
VIT	2	79,766,349	80,141,293	*COL23A1*	Front foot size, Hip structure
	4	79,687,359	79,847,281	*PRKDC*	Feed conversion ratio
	6	79,849,687	79,958,271	*HSPG2*	Days to 113 kg, Marbling

Table 6.1 Within population list of genomic regions under selection within and candidates' genes detected using the *iHS* method

Pop	Chr	Start Length	End Length	Gene name	QTLs
	14	79,352,396	80,106,258	*KCNMA1*	Meat colour b*
WAT	18	31,027,031	31,125,465	*MDFIC*	Fat androstenone level
WBO	2	15,791,451	15,819,138	*F2*	Age at slaughter, Average backfat thickness, Average daily gain, Backfat at last rib, Backfat at rump, Backfat thickness between 3rd and 4th rib, Body weight (end of test), Body weight (weaning), CD4-negative CD8-positive leukocyte percentage, CD4-positive CD8-positive leukocyte percentage, Calcium to phosphorus ratio, Carcass weight (hot), Cis-11-Elcosenoic acid to arachidic acid ratio, Feed conversion ratio, Ham weight, Lean meat percentage, Litter size, Loin and neck weight, Loin muscle area, Mean platelet volume, Number weaned, Platelet distribution width, Salmonella shedding status, Shoulder weight, Teat number, Total number born alive, pH 24 hr postmortem (ham), pH 24 hr postmortem (loin)
	5	4,769,801	48,49,334	*SHISAL1*	Intramuscular fat content
	5	6,997,444	7,014,422	*POLR3H*	Fat androstenone level
	13	66,316,436	66,452,917	*GHRL*	Age at slaughter, Average daily gain, Daily feed intake, Days to 100 kg, Feed intake, Loin weight, Marbling
	14	21,754,463	22,121,051	*SPOCK3*	Fat androstenone level

The regions displaying significant selection were distributed on different chromosomes, harbouring genes associated with important traits (Supplementary Tables 6.2). The strongest selection signal ($-\log_{10}[2\phi - |iHS|]$) = 24.09) was observed on chromosome 13 (Figure 6.1) in WBO overlapping the *ZBTB20* gene associated with chromosome 13 (145.36 Mbp). For various traits categories, this study identified QTLs that were common in the 13 populations. The QTL for number of teats was the most common and detected on several chromosomes, including chromosome 7 (ALN, DUR), chromosome 8 (ORT), chromosome 16 (MOP), chromosome 14 (CAP), chromosome 2 (LWT, WBO) chromosome 14 (LWT). Other regions identified using *iHS* are those associated with IMF content and detected in the ORT (chr 1, 12) CAP (chr 1, 14), LWT (chr 2), KOL (chr 4), WIN (chr 8) and WBO (chr 5) populations.

The regions that displayed significant evidence of selection harboured genes associated with important traits. For example, genes such as *LIPA* (101.13 Mbp), *IDE* (103.99 Mbp) and *RBF4* (105.04 Mbp) on chromosome 4 in the CAP population were associated with litter size.

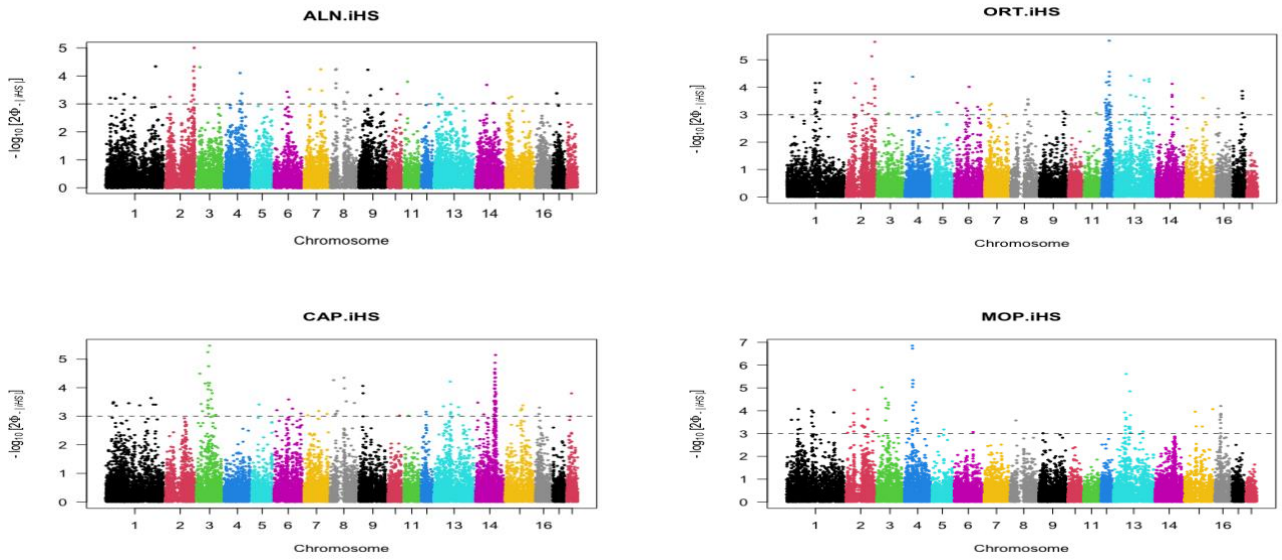


Figure 6.1a Manhattan plot of the genome-wide distribution of selection of signatures detected by *iHS* across the 18 chromosomes (Villages)

3149

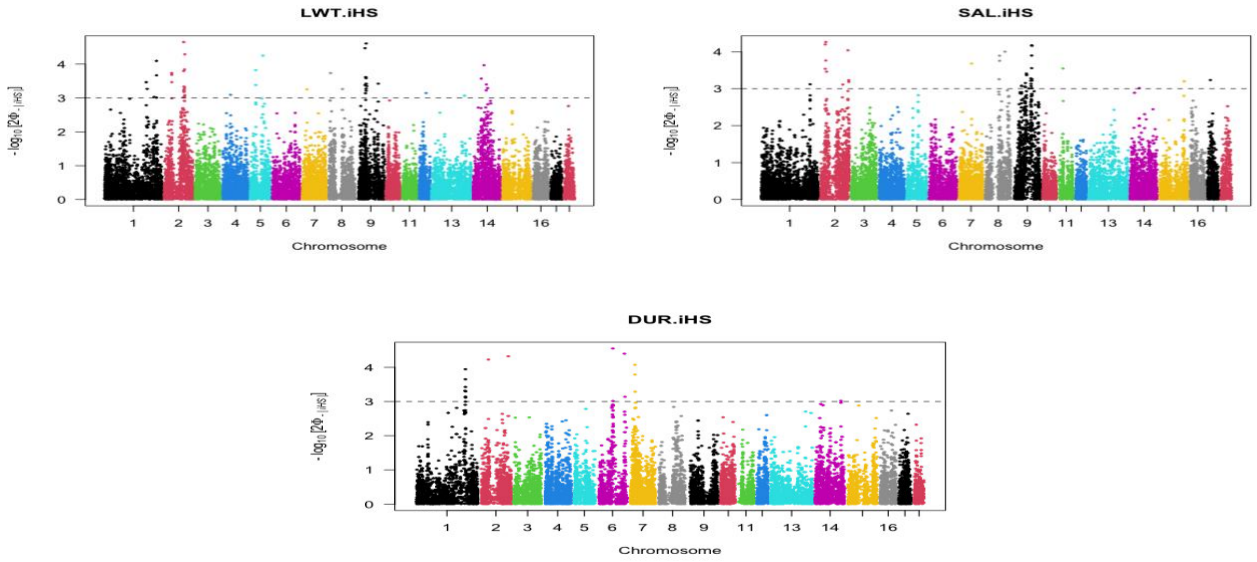


Figure 6.1b Manhattan plot of the genome-wide distribution of selection of signatures detected by *iHS* across the 18 chromosomes (Commercial)

3150

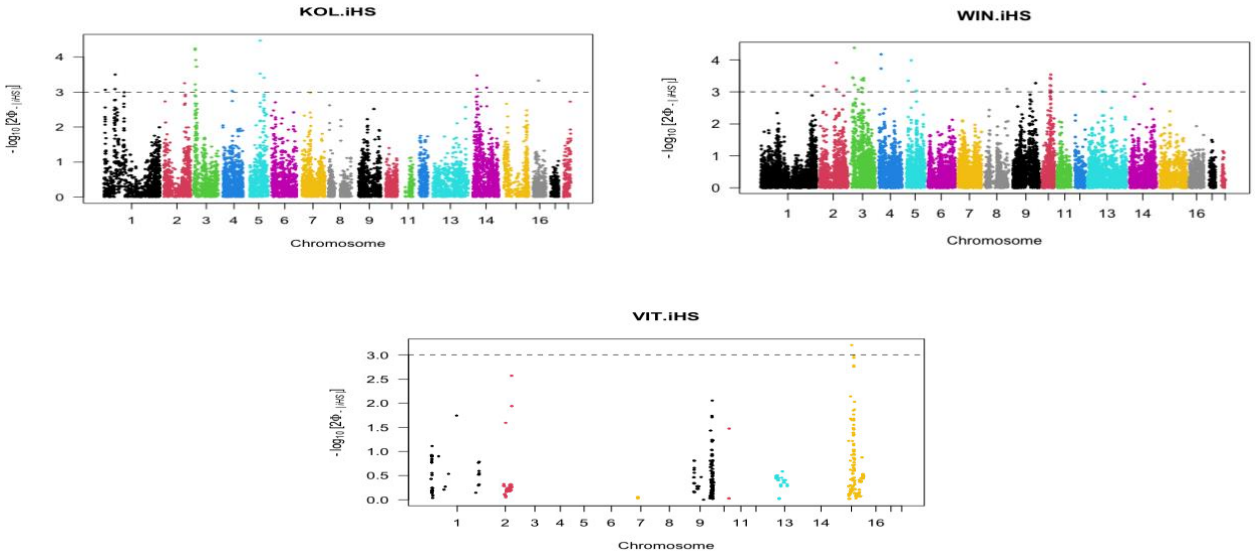


Figure 6.1c Manhattan plot of the genome-wide distribution of selection of signatures detected by *iHS* across the 18 chromosomes (Indigenous and Vietnamese Potbelly)

3151

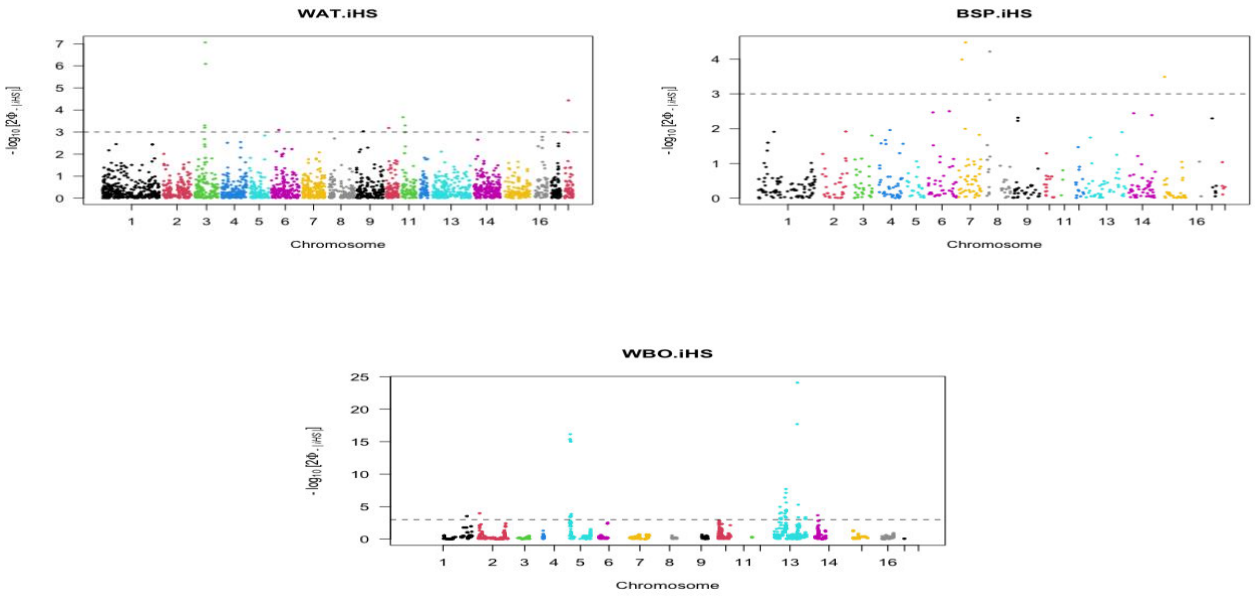


Figure 6.1d Manhattan plot of the genome-wide distribution of selection of signatures detected by *iHS* across the 18 chromosomes (Wild pigs)

3152

The region on chromosome 14 identified in VIT and KOL populations harbour the *KCNMA1* gene that is associated with meat colour, while the region on chromosome 8 (31.92 Mbp) detected in ALN, MOP and CAP has the *APBB2* gene associated with stearic acid content. *IDE* on chromosome 14 (103.99 Mbp) and *F2* on chromosome 2 (15.79 Mbp) were detected in CAP and WBO and are associated with several QTLs associated with feed conversion ratio, age at slaughter, average backfat thickness and average daily gain amongst others.

6.4.2 Detection of selection of signatures between populations using XP-EHH

Several regions that displayed significant evidence of selection were identified using *XP-EHH* to identify genomic regions under selection between pairs of populations (Supplementary Table 6.3). The number of regions differed between the different populations paired, with VIT_WBO having the least number of regions (1 region) and WAT_EC had the most (61 regions). The populations paired with WAT showed the most diverse regions, while Vietnamese Potbelly paired with other populations presented the smallest number of regions.

Table 6.2 Selected regions and candidates genes detected between pairs of populations using the *XP-EHH* method

Pop	Chr	Start Length	End Length	Gene name	QTLs
DUR_EC	1	166,173,135	166,310,972	*ITGA11*	Obesity index, Teat number
	2	113,774,412	114,206,930	*FER*	Abdominal circumference, Average backfat thickness, Average daily gain, Backfat at last lumbar, Biceps brachii weight, Body height, Body weight (3 weeks), Carcass weight (hot), Double-bond index
	6	80,649,143	80,843,567	*EPHB2*	Litter weight total
DUR_LIM	1	166,173,135	166,310,972	*ITGA11*	Obesity index, Teat number
	6	80,649,143	80,843,567	*EPHB2*	Litter weight total
DUR_IND	18	40,820,644	41,409,087	*PDE1C*	Backfat at rump
	18	42,030,510	42,046,184	*GHRHR*	Backfat at last rump, Carcass length, Fat-cuts percentage
DUR_LWT	2	113,774,412	114,206,930	*FER*	Abdominal circumference, Arachidonic acid content, Aspartate aminotransferase activity, Average backfat thickness, Average daily gain, Backfat at last lumbar, Backfat at mid-back, Backfat at rump, Backfat at tenth rib, Backfat between 3rd and 4th last ribs, Backfat between 6th and 7th ribs, Backfat between 3rd and 4th rib, Biceps brachii weight, Blood pH, Body height, Body length, Body weight (150 days), Body weight (birth), Body weight (weaning), Body weight (3 weeks), Body width, Cannon bone circumference, Carcass length, Chest circumference, Cross-sectional area of muscle fibers, Days to 100 kg, Days to 110 kg, Dihomo-gamma-linolenic acid content, Carcass weight (hot), Double-bond index, Drip loss, Ear size, Eicosadienoic acid content, Estimated carcass lean content, Fat androstenone lean content, Fat weight (total), Feed conversion ratio, Front feet

Table 6.2 Selected regions and candidates genes detected between pairs of populations using the *XP-EHH* method

Pop	Chr	Start Length	End Length	Gene name	QTLs
					conformation, Front leg conformation, Hematocritic, Hemoglobin, Hind feet conformation, Humerus length, Immunoglobulin G level, Intramuscular fat content, Iris pigmentation, Linoleic acid content, Litter size, Lymphocyte percentage, Mean corpuscular haemoglobin content, Mean corpuscular volume, Mean platelet volume, Meat colour L*, Meat colour a*, Monounsaturated fatty acid content, Monounsaturated fatty acid to polyunsaturated fatty acid ratio, Monounsaturated fatty acid to saturated fatty acid ratio, Monounsaturated fatty acid to saturated fatty acid ratio, Myristic acid content, Number of muscle fibers per unit area, Number weaned, Oleic acid to stearic acid ratio, PRRSV antibody titer, Palmitic acid content, Palmitoleic acid content, Percentage type IIb fibers, Peroxidability index, Platelet count, Polyunsaturated fatty acid content, Polyunsaturated fatty acid to saturated fatty acid ratio, Pseudorabies virus antibody titer, Red bood cell count, Red cell distribution width, Relative area of type IIb fibers, Residual glycogen, Rib shape, Seminiferous tubule diameter, Shoulder subcutaneous, fat thickness, Sperm motility, Sperm progressive motility, Stearic acid content, Teat number, Tibia length, Total muscle fiber number, Total number born alive, Ulna length, Vaccenic acid content, Vertebra number, White blood cell number, Backfat above muscle dorsi, Body mass index, Body weight (10 weeks) CIS-11-Eicosenoic acid content, Leaf fat percentage, pH 24 hr postmortem (ham)
	7	27,389,874	27,895,570	*KHDRBS2*	Loin muscle area, Loin muscle depth, Teat number
	7	30,708,332	30,724,045	*SNRPC*	Loin muscle area, Loin muscle depth
	7	30,731,461	30,802,735	*UHRF1BP1*	Femur length, Hip bone length, Humerus length, Tibia length, Ulna length
	7	30,812,920	30,995,361	*ANKS1A*	Femur length, Hip bone length, Humerus length, Tibia length, Ulna length, Galt score (front), Loin muscle area, Loin muscle depth
	7	31,586,555	31,603,617	*ARMC12*	Facial morphology
	7	31,722,990	31,792,904	*SLC26A8*	Facial morphology, Femur length, Humerus length, Tibia length, Ulna length
	14	111,834,168	111,914,304	*PAX2*	Monounsaturated fatty acid to saturated fatty acid ratio, Oleic acid to stearic acid ratio, Palmitoleic acid to palmitic acid ratio, Stearic acid content
DUR_VIT	2	72,326,714	72,391,648	*VAV1*	Average daily gain, Backfat between 3rd and 4th last rib, Birth weight variability, Body weight (end of test), Conductivity 45 minutes post-mortem, Fat androstenone level, Intramuscular fat content, Time in feeder per day, pH 24 hr postmortem (ham), pH 45 minutes postmortem
DUR_WBO	1	193,722,164	193,906,565	*ESR2*	Front leg conformation, Gait score (overall), Hind leg conformation, Litter size, Maternal infanticide, Plasma droplet rate, Semen volume, Sperm concentration, Sperm motility, Total number born alive
WAT_EC	1	254,683,885	254,703,225	*AMBP*	Conductivity 24 hours post-mortem (loin), pH 24 hr postmortem (ham), pH 24 hr post-mortem (loin), pH 45 minutes postmortem
	3	39,957,269	39,957,727	*NPW*	Lean meat percentage
	8	37,530,815	37,809,761	*CORIN*	Platelet count

Table 6.2 Selected regions and candidates genes detected between pairs of populations using the *XP-EHH* method

Pop	Chr	Start Length	End Length	Gene name	QTLs
	8	37,797,875	37,875,540	*NFXL1*	Mean corpuscular hemoglobin content, Mean corpuscular volume
	8	47,473,787	47,601,359	*RXFP1*	Red blood cell count
	8	71,520,275	71,554,766	*PPEF2*	Platelet distribution width
	8	71,573,783	71,603,338	*NAAA*	Platelet distribution width
	8	72,543,620	72,679,173	*SEPTIN11*	Teat number
	8	73,502,664	73,958,083	*FRAS1*	Teat number
WAT_LIM	1	254,683,885	254,703,225	*AMBP*	Conductivity 24 hours post-mortem (loin), pH 24 hr postmortem (ham), pH 24 hr post-mortem (loin), pH 45 minutes postmortem
	8	71,520,275	71,554,766	*PPEF2*	Platelet distribution width
	8	71,573,783	71,603,338	*NAAA*	Platelet distribution width
	14	113,414,264	113,429,936	*PSD*	Intramuscular fat content, Oleic acid to stearic acid ratio
	14	113,464,174	113,478,699	*MFSD13A*	Oleic acid content, Oleic acid to stearic acid ratio, Stearic acid content
	14	113,480,197	113,498,740	*ACTR1A*	Oleic acid content, Stearic acid content
	9	45,400,464	45,435,916	*TMPRSS4*	Backfat between 3rd and 4th last ribs
WAT_LWT	4	105,804,586	105,845,725	*CSDE1*	Intramuscular fat content
	4	105,868,897	105,893,771	*AMPD1*	Juiciness score, Overall impression, sensory panel, Tenderness score
WAT_WBO	7	89,120,822	89,168,270	*MAX*	Meat colour b*, Teat number, maximum per side
VIT_EC	18	29,895,878	29,936,233	*TES*	Average daily gain, Backfat between 3rd and 4th last rib, Birth weight variability, Body weight (end of test), Conductivity 45 minutes post-mortem, Fat androstenone level, Intramuscular fat content, , Time in feeder per day, pH 24 hr postmortem (ham), pH 45 minutes postmortem, Teat number
VIT_LIM	2	72,326,714	72,391,648	*VAV1*	Average daily gain, Backfat between 3rd and 4th last rib, Birth weight variability, Body weight (end of test), Conductivity 45 minutes post-mortem, Fat androstenone level, Intramuscular fat content, Time in feeder per day, pH 24 hr postmortem (ham), pH 45 minutes postmortem
	18	29,895,878	29,936,233	*TES*	Average daily gain, Backfat between 3rd and 4th last rib, Birth weight variability, Body weight (end of test), Conductivity 45 minutes post-mortem, Fat androstenone level, Intramuscular fat content, Time in feeder per day, pH 24 hr postmortem (ham), pH 45 minutes postmortem, Teat number
VIT_IND	18	29,895,878	29,936,233	*TES*	Average daily gain, Backfat between 3rd and 4th last rib, Birth weight variability, Body weight (end of test), Conductivity 45 minutes post-mortem, Fat androstenone level, Intramuscular fat content, , Time in feeder per day, pH 24 hr postmortem (ham), pH 45 minutes postmortem, Teat number
	18	31,027,031	31,125,465	*MDFIC*	Fat androstenone level
VIT_WBO	1	193,722,164	193,906,565	*ESR2*	Front leg conformation, Gait score (overall), Hind leg conformation, Litter size, Maternal infanticide, Plasma droplet rate, Semen volume, Sperm concentration, Sperm motility, Total number born alive

Several QTLs and genes were on the genomic regions identified using the *XP-EHH* method (Table 6.2, Supplementary Table 6.4). The strongest signal (*XP-EHH* = 6.91) was observed for

VIT_LIM on chromosome 9 though it was not associated with any QTL (Figure 6.2a,b,c) Another region on chromosome 2 (113.82 Mbp) that was identified in the DUR_EC and DUR_LWT populations harbour the *FER* gene associated with important production traits including meat and carcass quality traits. The regions on chromosomes 1 and 6 respectively for DUR paired with the village populations (EC and LIM) harbour the *ITGA11* (166.17 Mbp) and *EPHB2* (80.64 Mbp) genes associated with reproduction.

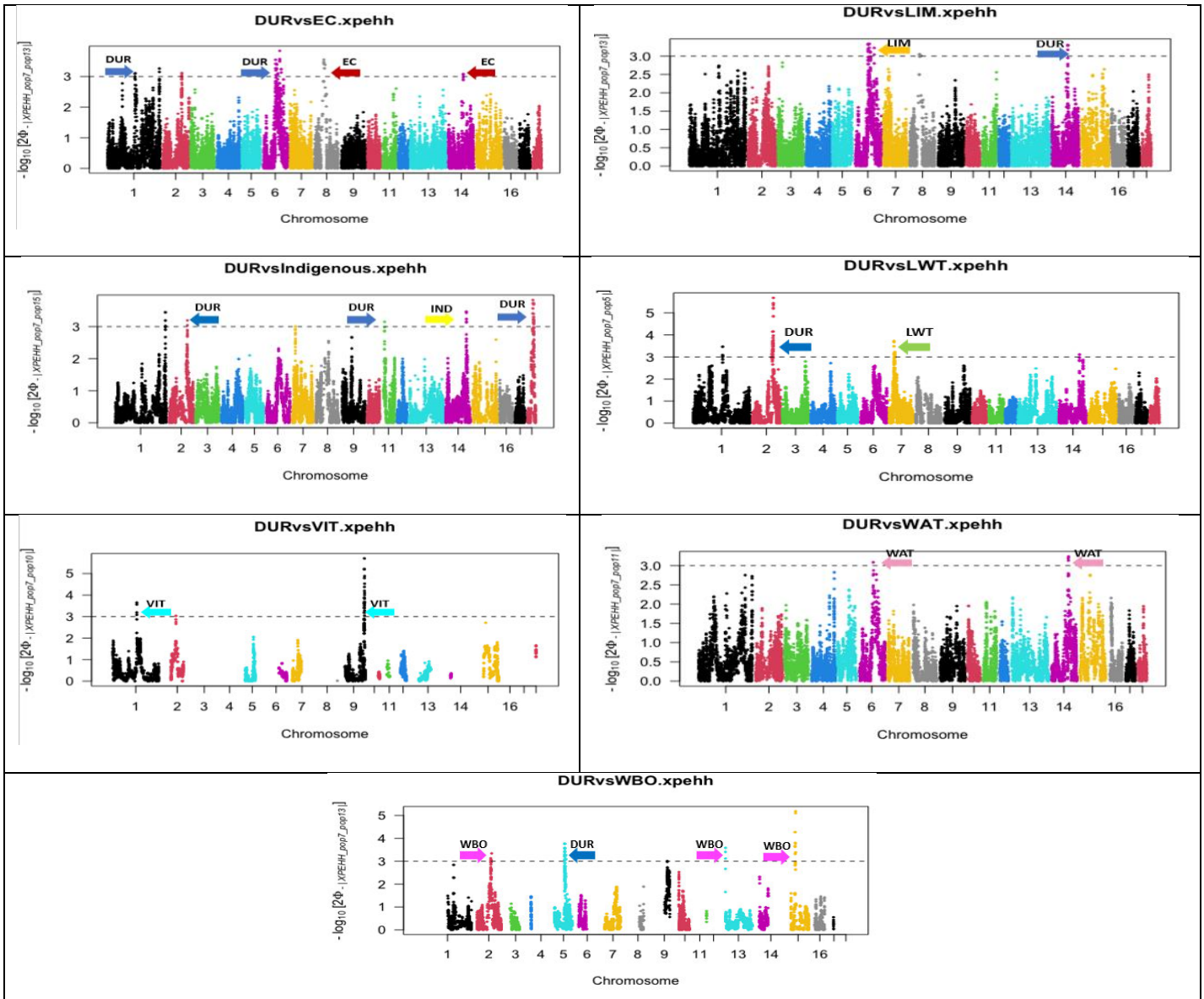


Figure 6.2a Manhattan plot of the genome-wide distribution of selection of signatures detected by *XP-EHH* across the 18 chromosomes (Duroc group)

3179

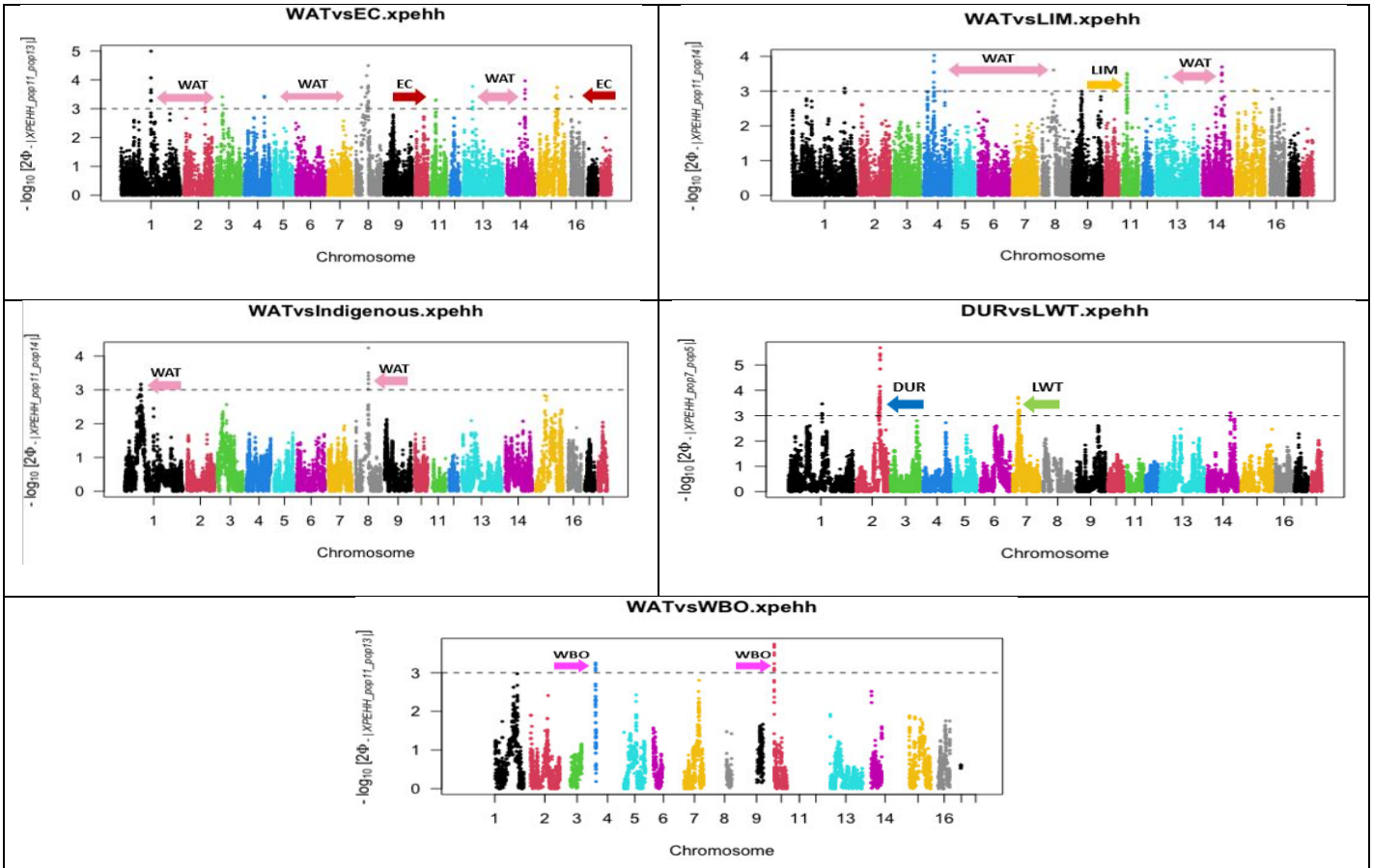


Figure 6.2b Manhattan plot of the genome-wide distribution of selection of signatures detected by *XP-EHH* across the 18 chromosomes (Warthog group)

3180
3181

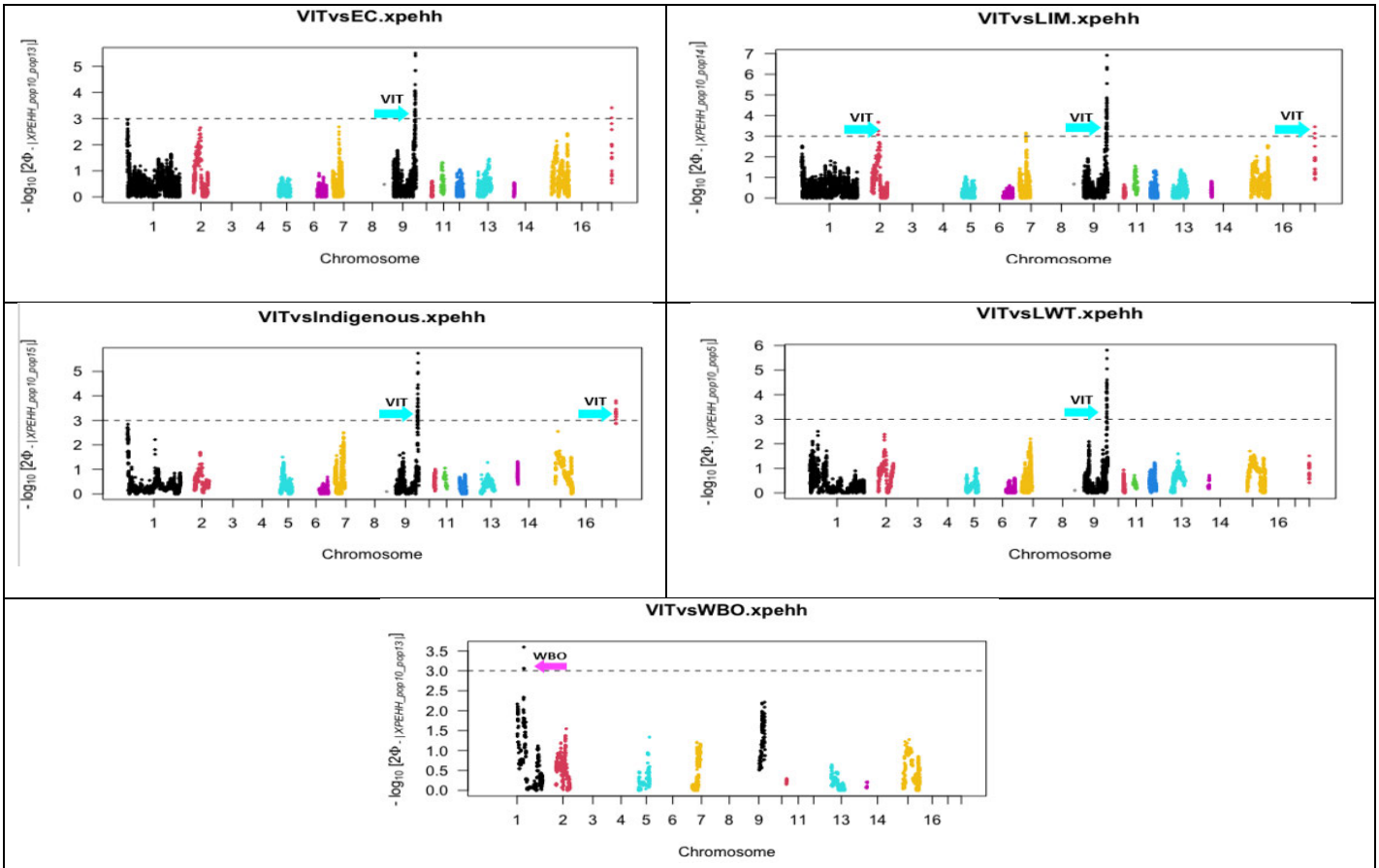


Figure 6.2c Manhattan plot of the genome-wide distribution of selection of signatures detected by *XP-EHH* across the 18 chromosomes (Vietnamese Potbelly group)

The *ESR2* gene on chromosome 1 (193.82 Mbp) detected in the DUR_WBO and VIT_WBO was associated with exterior and reproduction traits such as front leg conformation, gait score, hind leg conformation, litter size, maternal infanticide, plasma droplet rate, semen volume, sperm concentration, sperm motility and total number born alive. Results of the WAT_LIM and WAT_EC population pairwise analysis identified a region on chromosome 1 that harbour the *AMBP* (254.68 Mbp) gene was associated with meat quality traits, as well as *NAAA* (71.57 Mbp) and *PPEF2* (71.52Mbp) related to platelet distribution width on chromosome 8.

6.4.3 Detection of selection of signatures between populations using HapFLK

Across all populations, regions displaying significant evidence of selection were identified on chromosomes 5 and 6 using KOL as outgroup (Figure 6.3). A total of 5 924 segments were found as signatures of selection and 1 179 genes allocated (Supplementary Table 6.5). A total of 110 genes were linked to QTLs (Table 6.3). Genes on chromosomes 2 (*CAT*) and 14 (*IDE*) had the most diverse QTLs. These genes were associated with various QTLs such as age at slaughter, meat to fat ratio, body weight, teat number and litter size to mention the few. Signatures associated with IMF content were revealed on chromosomes 1 (*NEDD4L, CORO2B, PAQR5, CDH20*), 2 (*CAT*), 3 (*CYRIA1*), 4 (*CSDE1, JPH1, DPT, KIRREL1*), 6 (*CDH2, PIK3C3*), 11 (*GPC6*) and 14 (*IDE, ABLM1*). QTL for number of teats is related to *GRM1, ANKS6, CAT, FGF1, PRMT8, GNAL, IPA, IDE* and *CHST15* genes on chromosomes 1, 2, 5, 6 and 14 respectively.

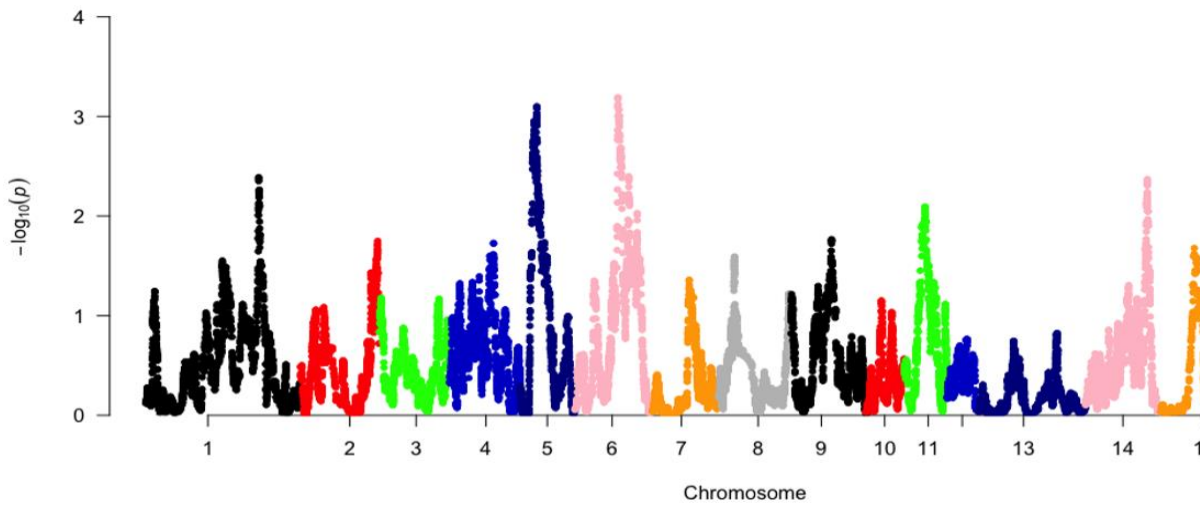


Figure 6.3 Manhattan plot for signature of selection of South African pig populations detected by *HapFLK* across 18 chromosomes

Table 6.3 Regions under selection detected by *HapFLK* methods in South African pigs (*P-value* < 0.10)

Chr	Start Length	End Length	Gene name	QTLs
1	162,144,471	162,367,879	*ALPK2*	Average daily gain, Lean meat percentage
1	162,364,055	162,737,492	*NEDD4L*	Intramuscular fat
1	19,173,791	19,572,769	*GRM1*	Teat number
1	163,237,100	163,271,164	*CILP*	Body length, Front leg conformation
1	166,173,135	166,310,972	*ITGA11*	Obesity index, Teat number
1	166,465,953	166,608,367	*CORO2B*	Average daily gain, Intramuscular fat content. Obesity index
1	167,154,627	167,234,807	*PAQR5*	Intramuscular fat
1	193,722,164	193,906,565	*ESR2*	Front leg conformation, Gait score (overall), Hind leg conformation, Litter size, Maternal infanticide, Plasma droplet rate, Semen volume, Sperm concentration, Sperm motility, Total number born alive
1	209,725,636	209,745,705	*TYRP1*	Gestation length
1	238,197,666	238,220,524	*POLR1E*	Fat androstenone level
1	147,568,673	147,685,288	*MBP*	Conductivity 24 hours post-mortem (loin), pH 24 hr post-mortem (ham), pH 24 hr post-mortem (loin), pH 45 minutes postmortem
1	240,522,837	240,584,918	*ANKS6*	Teat number
1	156,295,431	156,435,823	*CDH7*	Spinal curvature
1	14,217,036	14,493,363	*ESR1*	Litter size, Non-return rate, Number weaned, Plasma droplet rate, Sperm motility, Total number born alive, Androstenone laboratory, Indole laboratory
1	159,817,557	160,024,988	*CDH20*	Average daily gain, Intramuscular fat content. Obesity index, Lean meat percentage
2	26,095,038	26,115,464	*APIP*	Front feet conformation, Front leg conformation, Hind leg conformation, Rib shape
2	26,487,653	26,581,452	*CAT*	Abdominal circumference, Age at slaughter, Average backfat thickness, Average daily gain, Average glycogen, Average glycolytic potential, Average lactate, Backfat at last rib, Backfat at rump, Backfat between 3rd and 4th last ribs, Biceps brachii weight, Body weight, Body weight (end of test), Carcass length, Conductivity 24 hours post-mortem (ham), Conductivity 24 hours post-mortem (loin), Cooking loss, Days to 100 kg, Drip loss, Estimated carcass lean content, Fat androstenone level, Fat percentage in carcass, Fat skatole level, Fat-cuts percentage, Feed conversion ratio, Gadoleic acid content, Ham weight, Hind foot size, Intramuscular fat content, Lean meat percentage, Linoleic acid content, Litter size, Loin and neck weight, Loin muscle area, Meat colour L*, Meat colour a*, Meat colour b*, Meat colour score, Meat colour-L, Meat colour-a, Meat to fat ratio, Monounsaturated fatty acid to polyunsaturated fatty acid ratio, Muscle cathepsin B activity, Number of muscle fibers per unit area, Oleic acid content, PH for Longissmus dorsi, Polyunsaturated fatty acid content, Residual glycogen, Shear force, Shoulder subcutaneous fat thickness, Stearic acid content, Teat number, Thawing loss, Thoracic vertebra number. Time in feeder per day, Total muscle fiber number, Total number born alive, pH 24 hr postmortem (ham), pH 24 hr post-mortem (loin), pH 45 minutes postmortem, pH 48 hr post-mortem (loin)
2	27,145,774	27,184,735	*FBXO3*	Days to 100 kg
2	32,734,691	32,856,322	*LIN7C*	Front leg conformation
2	144,133,390	144,241,323	*FGF1*	Body weight, Functional teat number, Teat number

Table 6.3 Regions under selection detected by *HapFLK* methods in South African pigs (*P-value* < 0.10)

Chr	Start Length	End Length	Gene name	QTLs
2	144,822,937	144,956,095	*NR3C1*	Carcass length, Conductivity 24 hours post-mortem (loin), Conductivity 45 minutes post-mortem, Drip loss, Front leg conformation, Gait score (Overall), Loin muscle area, Meat colour OPTO, Meat colour-L, Meat colour-a, Meat colour-b, Leaf fat percentage, pH 24 hr post-mortem (loin)
3	6,220,939	6342,637	*ARPC1B*	Fat androstenone level
3	49,466,903	49,521,875	*TGFBRAP1*	Meat colour a*
3	120,968,060	121,069,242	*CYRIA*	Intramuscular fat content
4	20,143,864	20,569,129	*SAMD12*	Body mass index
4	37,090,935	37,820,419	*VPS13B*	Backfat at rump
4	105,804,586	105,845,725	*CSDE1*	Intramuscular fat content
4	46,730,461	46,767,427	*DECR1*	Cholesterol level, LDL cholesterol
4	61,628,299	61,716,546	*JPH1*	Intramuscular fat content
4	111,824,896	112,202,825	*VAV3*	Average daily gain
4	117,496,908	117,511,226	*VCAM1*	Backfat thickness between 3rd and 4th rib
4	117,635,861	117,805,254	*CDC14A*	Cannon bone circumference
4	118,176,241	118,268,835	*AGL*	Average backfat thickness, Average daily gain, Backfat thickness between 3rd and 4th rib, Body weight (3 weeks), Body weight (birth), Carcass weight (hot), Sperm motility, Sperm progressive motility, Androstenone motility, Body weight (10 weeks)
4	69,616,035	69,806,496	*CYP7B1*	pH for longissmus dorsi
4	70,961,465	71,021,839	*GGH*	Fat androstenone level
4	71,045,333	71,318,030	*NKAIN3*	Body weight (birth)
4	79,687,359	79,847,281	*PRKDC*	Feed conversion ratio
4	82,287,730	82,319,367	*DPT*	Intramuscular fat content
4	84,961,867	85,017,820	*TMCO1*	Residual glycogen
4	92,176,745	92,290,610	*KIRREL1*	Intramuscular fat content
4	88,591,280	88,818,372	*ATF6*	Loin muscle area
5	44,381,009	44,558,744	*FAR2*	Feed conversion ratio
5	56,817,606	57,005,114	*EPS8*	Fat androstenone level
5	64,519,186	65,002,098	*VWF*	Litter size
5	66,665,263	66,838,922	*PRMT8*	Teat number
5	28,300,950	28,605,120	*SRGAP1*	Ear area
5	29,695,839	29,863,599	*MSRB3*	Ear area
5	97,092,883	97,148,601	*SLC6A15*	Time in feeder per day
5	33,858,572	34,033,939	*CCT2*	Feed conversion ratio
5	34,067,992	34,218,513	*MYRFL*	Feed conversion ratio
5	34,660,029	34,794,321	*PTPRB*	Feed conversion ratio
5	36,274,364	36,658,703	*TRHDE*	Feed conversion ratio
6	97,342,474	97,429,364	*GNAL*	Age at puberty, Arachidic acid content, Average backfat thickness, Average daily gain, Backfat at last lumbar, Backfat at last rib, Backfat at rump, Backfat at tenth rib, Body weight (16 days), Carcass weight (hot), ear area, Feed conversion ratio, Lean meat percentage, Loin muscle area, Loin muscle depth, Oleic acid content, Oleic acid to stearic acid ratio, PH for longissmus dorsi, Stearic acid content, Teat number, Vertebra number, Androstenone laboratory
6	108,115,886	108,227,748	*CABLES1*	Average daily gain, Backfat at rump
6	108,548,837	108,805,693	*LAMA3*	Average daily gain
6	75,696,484	75,755,892	*PADI2*	Fat androstenone level
6	112,396,721	112,628,432	*CDH2*	Average daily gain, Intramuscular fat content, Lean meat percentage, Obesity index

Table 6.3 Regions under selection detected by *HapFLK* methods in South African pigs (*P-value* < 0.10)

Chr	Start Length	End Length	Gene name	QTLs
6	117,631,227	118,334,105	*NOL4*	Average backfat thickness
6	79,849,687	79,958,271	*HSPG2*	Days to 113 kg, Marbling
6	125,890,708	126,043,161	*PIK3C3*	Average backfat thickness, Average daily gain, Intramuscular fat content, Loin muscle area
6	80,649,143	80,843,567	*EPHB2*	Litter weight (total)
6	137,595,524	138,010,444	*SLC44A5*	Diameter of muscle fibers
6	43,933,759	44,069,330	*GPI*	Average backfat thickness, Body weight (5 weeks), Intramuscular fat content, Osteochondrosis score
6	46,442,857	46,470,075	*ZNF570*	Lean meat percentage
7	78,49,1053	78,530,184	*TEP1*	Sperm concentration, Sperm per ejaculate
7	87,567,379	87,766,911	*SV2B*	Age at puberty
8	39,715,534	40,145,887	*SCFD2*	Red blood cell count
8	41,809,116	41,856,339	*KDR*	Red blood cell count
8	29,370,481	29,593,274	*TBC1D1*	Age at puberty
8	31,921,587	32,301,975	*APBB2*	Stearic acid content
8	37,530,815	37,809,761	*CORIN*	Platelet count
8	37,797,875	37,875,540	*NFXL1*	Mean corpuscular hemoglobin content, Mean corpuscular volume
9	67,726,154	67,829,220	*C4BPA*	Carcass weight (hot), Cooking loss, Meat colour L*, Meat colour a*, Muscle fat content, Water holding capacity, Muscle protein percentage, pH 24 hr post-mortem (loin), pH 45 minutes postmortem
9	72,385,263	72,444,616	*PEX1*	Fat androstenone level
9	86,511,369	86,555,943	*AHR*	Age at puberty, Litter size, Total number born alive
9	45,400,464	45,435,916	*TMPRSS4*	Backfat between 3rd and 4th last ribs
9	46,404,999	46,512,834	*CBL*	Average daily gain, CD-4positive CD-8positive leukocyte percentage, Fat androstenone level
9	48,120,802	48,286,278	*TECTA*	Backfat between 3rd and 4th last ribs
10	27,588,582	27,623,077	*FBP1*	Front feet conformation, Front leg conformation
10	32,755,991	32,892,869	*UBAP2*	Arachidic acid content
10	50,540,556	50,680,154	*ARHGAP21*	Mean corpuscular hemoglobin content
10	52,862,835	53,060,740	*DNAJC1*	Dihomo-gamma-linolenic acid content
11	49,164,141	49,432,858	*MYCBP2*	Iris pigmentation
11	62,438,652	63,560,903	*GPC6*	Intramuscular fat content
12	24,987,347	25,022,616	*TTLL6*	Red cell distribution width
14	75,518,528	75,721,502	*MCU*	Palmitoleic acid to palmitic acid ratio
14	79,352,396	80,106,258	*KCNMA1*	Meat colour b*
14	87,644,240	87,722,245	*OPN4*	Fat androstenone level
14	87,742,459	87,891,793	*BMPR1A*	Thoracic vertebra number
14	89,353,367	89,496,349	*WDFY4*	Numbers of muscle fibers per unit area
14	90,338,503	90,480,986	*PARG*	Backfat at tenth rib, Body width, Front feet conformation, Hind feet conformation, Rib shape
14	101,123,381	10,1254,725	*LIPA*	Average daily gain, Front leg weight, HDL/LDL ratio, Litter size, Loin muscle area, Monounsaturated fatty acid content, Oleic acid content, Skin thickness, Sperm concentration, Teat number
14	103,992,037	104,106,576	*IDE*	Abdominal fat percentage, Age at puberty, Age at slaughter, Average backfat thickness, Average daily gain, Backfat at first rib, Backfat at last rib, Backfat between 3rd and 4th last ribs, Body depth, Body height, body length, Body weight (slaughter), Body weight (weaning), Body width, CD4-negative CD8-positive leukocyte percentage, CD4-positive CD8-negative leukocyte percentage, CD4-positive/CD8-positive

Table 6.3 Regions under selection detected by *HapFLK* methods in South African pigs (*P-value* < 0.10)

Chr	Start Length	End Length	Gene name	QTLs
				leukocyte ratio, Carcass length, Carcass weight (hot), Conductivity 24 hours post-mortem (loin), Cooking loss, Daily feed intake, Days to 100 kg, Dressing percentage, Ear area, Fat androstenone level, Fat percentage in carcass, Fat-cuts percentage, Feed conversion ratio, Feed intake, Femur length, Front leg conformation, Front leg weight, Gait score (overall), Ham weight, Haemoglobin, Hind foot size, Hind leg conformation, Hip bone length, Hip structure, Humerus length, Intramuscular fat content, Lean meat percentage, Linoleic acid content, Litter size, Loin muscle area, Loin muscle depth, Loin weight, Lymphocyte number, Lymphocyte percentage, Marbling, Maternal infanticide, Mean corpuscular hemoglobin content, Mean corpuscular volume, Meat colour L*, Meat colour a*, Meat colour score, Meat colour-L, Meat to fat ratio, Monounsaturated fatty acid to polyunsaturated fatty acid ratio, Muscle moisture percentage, Number of visits to feeder per day, Number weaned, Oleic acid content, Oleic acid stearic acid ratio, PH for longissmus dorsi, PRRSV antibody titer, Palmitic acid content, Palmitoleic acid content, Polyunsaturated fatty acid content, Red blood cell count, Red cell distribution width, Rib shape, Salmonella shedding status, Shoulder weight, Side fat thickness, Sperm concentration, Sperm per ejaculate, Stearic acid content, Teat number, Teat number (maximum per side), Tibia length, Total number born alive, Triglyceride level, Ulna length, Vaccenic acid content, White blood cell number, Androstenone laboratory, Body mass index, cis-11-Eicosenoic acid content, Indole laboratory, Skatole laboratory
14	47,946,396	48,040,822	*LIMK2*	Fat androstenone level, Melanoma susceptibility
14	122,777,130	122,828,051	*ACSL5*	Fat androstenone level
14	123,343,694	123,546,417	*TCF7L2*	Carcass weight (hot), Number of visits to feeder per day
14	133,460,167	133,544,018	*CHST15*	Teat number
14	74,734,189	74,887,763	*PSAP*	CD4-negative CD8-positive leukocyte percentage, CD4-positive CD8-negative leukocyte percentage, CD4-positive/CD8-positive leukocyte ratio
14	124,760,398	125,010,892	*ABLIM1*	Fat androstenone level, Intramuscular fat content
15	100,868,469	101,211,242	*ANKRD44*	Skin thickness
15	101,623,818	101,957,516	*PLCL1*	Skin thickness
15	118,335,925	118,452,296	*XRCC5*	Average backfat thickness, Conductivity 24 hours post-mortem (loin), Cooking loss, Fat weight (total), Lean meat percentage, Loin muscle area, Loin muscle depth, Loin weight, PH for longissmusdorsi, Subcutanous fat area, pH 24 hr postmortem (ham), pH 24 hr post-mortem (ham), pH 24 hr post-mortem (loin)
15	79,935,994	80,025,850	*SP3*	Cooking loss, Meat colour b*, Shear force, Thawing loss
16	32,130,235	32,304,592	*ITGA1*	Obesity index, Teat number
16	32,336,292	32,437,103	*ITGA2*	Body weight (5 weeks)
16	33,126,494	33,565,542	*ARL15*	Loin muscle area
18	34,006,688	34,906,268	*IMMP2L*	Age at puberty
18	40,820,644	41,409,087	*PDE1C*	Backfat at rump
18	42,030,510	42,046,184	*GHRHR*	Backfat at last rib, Carcass length, Fat-cuts percentage
18	51,387,836	51,802,945	*HECW1*	Teat number

3206

6.4.4 Genes identified using different signature of selection methods

Twenty-seven candidate genes were identified using the *iHS*, *XP-EHH* and *HapFLK* methods. These included 16 on chromosome 8. A total of 8 (*EPB41L3, METTL4, EPHA8, LYPLA2, FUCA1, PNRC2, SRSF10, MYOM3*), 6 (*EPHA8, LYPLA2, FUCA1, PNRC2, SRSF10*), 4 (*METTL4, SLC16A12, PANK1, PCGF5*), 2 (*NECAP1, KCNJ3*), 2 (*PANK1, PCGF5*) and 5 (*EIF3H, OLFM4, SLC16A12, PANK1, PCGF5*) candidates genes were identified and common for DUR_EC, DUR_LIM, DUR_WAT, DUR_WBO, WAT_EC and WAT_LIM populations (Figure 6.4).

The following genes revealed in *HapFLK* method were also common for *iHS* method. These were *NEDDFL* (ORT), *ITGA11* (EC and LIM), *SLC44A5* (DUR), *APBB2* (ALN, MOP, CAP), *TECTA* (LWT), *IDE* (CAP) and *ITGA2* (MOP). Regions associated with *ESR2* (VIT_WBO), *EPHB2* (DUR_EC and DUR_LIM), *CORIN* (WAT_EC), *TMPRSS4* (WAT_LIM), *PDE1C* (DUR_IND) and *GHRHR* (DUR_IND) were observed using both *HapFLK* and *XP-EHH* methods.

6.4.5 Functional enrichment analysis

The genes located in genomic regions were accessed for functional enrichment using David and Panther 17.0 (Mi *et al.*, 2020). DUR_EC (*EPB41L3, METTL4, EPHA8, LYPLA2, FUCA1, PNRC2, SRSF10, MYOM3*) and DUR_LIM (*EPHA8, LYPLA2, FUCA1, PNRC2, SRSF10*) had the most overlapping genes detected using the three methods.

Table 6.4 Common genomic regions and genes under signatures of selection detected using *iHS, XP-EHH* and *HapFLK* methods

Populations	Chromosomes	No. Genes	Genes
DURvsEC	6	8	*EPB41L3, METTL4, EPHA8, LYPLA2, FUCA1, PNRC2, SRSF10, MYOM3*
DURvsLIM	6	6	*EPHA8, LYPLA2, FUCA1, PNRC2, SRSF10*
DURvsWAT	6, 14	4	*METTL4, SLC16A12, PANK1, PCGF5*
DURvsWBO	5, 15	2	*NECAP1, KCNJ3*
WATvsEC	14	2	*PANK1, PCGF5*
WATvsLIM	4, 11. 14	5	*EIF3H, OLFM4, SLC16A12, PANK1, PCGF5*

Most of the genes for DUR_EC and DUR_LIM were involved in nucleus, cytoplasm, RNA binding and mRNA splicing GO terms (Table 6.4; Supplementary Table 6.6). Most of these GO terms (Supplementary Table 6.6) were shared amongst the studied populations. However, DUR_WBO reported diverse compared to other populations and this included regulation of ion transmembrane transport, clathrin vesicle coat, voltage-gated potassium channel activity, plasma membrane, vesicle-mediated transport, ligand-gated ion channel activity and potassium transmembrane transport were. Important pathways were found to be related to Dopamine receptor mediated signalling pathway (DUR_EC) and Coenzyme A biosynthesis (DUR_WAT, WAT_EC and WAT_LIM). A total of four pathways (P0026, P0027, P05731, P0043) were associated with genomic regions under selection between the DUR_WBO (Supplementary Table 6.7).

6.5 Discussion

To date, this is the first study identifying signatures of selection footprints in South African villages, commercial, indigenous, wild and Vietnamese Potbelly pigs. The *iHS* statistical analyses, widely used in pigs (by Chen *et al*., 2018, Mujibiet *et al*., 2018; Wu *et al*., 2020) allowed for the identification of regions displaying signatures of selection within populations while the *XP-EHH* method allowed for the identification of signatures of selection based on long ranges of linkage disequilibrium between two populations (Vatsiou *et al*., 2016). The *HapFLK* method allowed for the identification of selection of signatures across all the nine populations. Identification of signatures of selection is important, as this information increase our knowledge about the genomic regions and evolutionary processes that shape populations and affects important phenotypic traits (Gouveia *et al*., 2014). For example, genomic regions identified in this study were associated with meat and fertility traits in Large White, and Duroc commercial breeds. Regions under natural and artificial selection that has resulted in the distinct imprints in these populations increase our knowledge about the genes and pathways that influence production and adaptation.

In this study, several regions displayed significant evidence of positive selection and included genes and QTLs associated with meat and carcass quality. In Chapter 3, meat and carcass QTL were highlighted as traits in genomic regions with high F_{ST} values and differentiating breeds. Chromosome 4 showed the most signatures linked to mostly meat and carcass quality traits. As

was expected, this included regions associated with meat and carcass quality among the Large White and Duroc breeds, as commercial pigs are selected for meat production (Rubin *et al*., 2012; Gouveia *et al*., 2014; Gurgul *et al*., 2018). Genes in these regions associated with meat and carcass quality included *CORIN* on chromosome 8, *TMPRRSS4* on chromosome 9, *SLC44A5* on chromosome 6, *APBB2* on chromosome 8, *TECTA* on chromosome 9, *LIPA* and *IDE* on chromosome 14 and *ITGA2* on chromosome 16. *DECR1* gene on chromosome 4 that is associated with cholesterol levels amongst other meat quality and growth traits was also detected within shared haplotype blocks in Chapter 4.

Among the indigenous Kolbroek and Windsnyer breeds also included the *JPH1* gene observed on chromosome 4 that has previously been linked to meat and carcass quality in pigs (Duarte *et al*., 2016; Qin *et al*. 2020). This shows that indigenous breeds can also be identified with traits despite their slow growth rate. Furthermore, Hoffman *et al*. (2005) observed that meat from Kolbroek pigs can be processed into bacon, ham, and chops. Identification of QTL associated with meat colour in this study is important, as it implies on the possibility to select for higher growth rates improved meat taste and meat colour.

In this study, several genomic regions were identified to be associated with fatness that is an important economic trait in pigs (Kogelman *et al*., 2013). The QTLs associated with IMF content identified in this study in indigenous breeds (e.g., Kolbroek), that are associated with excess fat deposition when fed improved diets (Wang *et al*., 2013; Reardon *et al*., 2010), presents an opportunity to genetically improve on meat quality in these breeds. A study by Jung *et al*. (2015) and Ren *et al*. (2017) observed that consumers preferred lean pork with high IMF content. However, pig breeds vary when it comes to fat tissue deposition with heritability levels being around 0.5 (Switonski *et al*., 2010; Stachowiak *et al*., 2016). While commercial breeds (*e.g*., Large White, Duroc and Landrace) have low levels of fat tissue, European breeds (*e.g*., Iberian and Mangalica pigs) and Chinese breeds are predisposed to accumulate excess amounts of adipose tissue (Lopes-Bote, 1998; Egerszegi *et al*., 2003; Stachowiak *et al*., 2016). Obesity index and IMF are related meat and carcass quality that can be used as adaptive tools to select animals (Stachowiak *et al*., 2016). Hoffman *et al.* (2005) has reported consumers' preference towards meat with higher lean percentage. The *ITGA11* gene on chromosome 1 and the *NPY2R* gene on chromosome that were identified in both this Chapter and the previous Chapter 3 are associated with obesity index that determines fat deposition in pigs and other animals (Stachowiak *et al*., 2016)). The *EPHB2* gene driven by Duroc signatures is also within QTLs

that play a role in obesity and reproductive traits (teat numbers). Regions harbouring *F2* and *SHISAL1* genes associated with IMF were identified among the Wild Boar populations. Wild Boar have low IMF and it is categorised under game meat that has high protein and iron and is considered healthier than ordinary pork or beef meat (Ludwiczak *et al*., 2020; Niewiadomska *et al*., 2020; Ciobanu *et al*., 2022).

QTLs associated with loin muscle area and loin muscle depth were identified among the commercial breeds in this study, likely because commercial breeds such as Large White, Durocs and South African Landrace are selected for high lean growth and have higher growth rate (Nevrkla *et al*., 2021; Zhang *et al.*, 2019). QTLs associated with growth rate, meat and carcass quality were identified among the Vietnamese Potbelly, including the region on chromosome 18 associated with average daily gain, fat androstenone level and IMF content.

Genomic regions displaying signatures of selection were associated with important reproduction traits (*e.g*., litter size and total number born alive from a sow, semen volume, sperm concentration, sperm motility, *etc*.) that has impact on pig reproduction. For example, the genomic regions showing significant signatures of selection on chromosome 1 harbour the *ESR2* gene that are associated with fertility traits (*e.g*., litter size, semen volume, sperm concentration, sperm motility and total number born alive). These traits identified in this study demonstrated genomic regions influencing the ability and performance of boars and sows to reproduce and production. Litter size is important since pigs differ greatly in litter size. For example, Wild Boar sows average 6.6 litters (Fonseca *et al*., 2011) per year versus an average of 14 to 15.3 litters per sow in Large White breeds (Dall'Olio *et al*., 2018; Krupa *et al*., 2020), while indigenous breeds such as Kolbroek averages 8-10 piglets (Halimani *et al*., 2012). Nowadays, the pig industry in Europe has been yielding 18-20 litters per sow (Oliviero *et al*., 2019). The high litter number has a negative implication on the physiological tolerance for both sows and litters. The good mothering ability and hardiness of sows ensure high survival rates for the litters. Commercial breeds have an advantage being raised in the intensive production system.

Several genomic regions contain QTLs associated with number of teats on chromosomes 1, 2, 5, 6, 14, 16 and 18. The number of teats is an important trait as it ensures that piglets have adequate access to milk from the sow. The number of teats can have effects on the weaning weight of a piglet and lesser number of teats in a sow reduces piglets survival rate (Lopes *et al*.,

2014). In commercial breeds such as Large White and Duroc, a sow can have as many as 19 teats (Duijvesteijn *et al*., 2014). Makhanya (2018) reported the number of teats to average 10 in the indigenous Kolbroek pigs. Various studies have shown that the number of teats is an essential morphological and reproductive trait that has been under selections for many generations in the pig industry (Duijvesteijn *et al*., 2015; Verado *et al*., 2015).

More signatures under selection were detected among the village pig populations compared to the other populations. Van Hossou *et al*. (2021) also reported a higher number of selection signatures on admixed West African cattle populations in Benin. This might be caused by admixture as a result of being crossbred with commercial pigs to improve productive economic traits such as reproduction, growth and carcass traits amongst others. Higher levels of genetic diversity were observed among the village pig populations in comparison to the other populations in Chapter 3. Village pigs also displayed the largest number of shared haplotype blocks in Chapter 4. The identification of genes such as *APBB2* (chromosome 8) in Alfred Nzo, *NEDD4L* (chromosome 1), *NR3C* (chromosome 8), *PIK3R5* (chromosome 12), *VTN* (chromosome 12) in O.R. Tambo, *APBB2* (chromosome 1), *SELENOP* (chromosome 16) in Mopani, *SAMD4A* (chromosome 1), *APBB2* (chromosome 8), *LIPA*, *IDE* (chromosome 14) in Capricorn population that are related to meat and conformation shows that these admixed village populations have undergone some divergent development towards the same production traits targeted in commercial breeds.

6.5 Conclusion

This study identified a number of regions showing signatures of selection as well as candidates' genes and functional pathways with potential effect on phenotypic traits, especial IMF content. We were able to observe signatures related to reproduction, production, health and meat and carcass QTLs. Village populations showed more regions compared to commercial breeds using the *iHS* method. The *XP-EHH* method indicated the villages, commercial, indigenous, Asian Vietnamese Potbelly and wild populations to harbour QTLs associated with a number of economic importance. Meat and carcass QTLs were prevalent in all the populations, showing the potential of village and indigenous populations' ability to be managed and improved for such traits. The chip may not be dense enough in order to fully understand the signatures between domestic and wild, further research is required. Genetic resource from village is

important for research as they are not influenced by selection when compared to commercial breeds.

CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

7.1 Introduction

Pigs found in South Africa (SA) contributes to the economic welfare of the country and are an important source of protein, as well contribute to household income, food security and poverty alleviation. The average pork products consumption in South Africa from 2018 to 2020 stood at 22% compared to 21% in the world while pork meat consumption is estimated to grow from 250 000 tonnes in 2017 to about 300 000 tonnes by 2029 (BFAP, 2021). However, the future of South African pigs is worrying a decline in pig numbers was noted from 1.52 million in 2016 to 1.36 million in 2020 (Statista, 2022) with a possibility of further decreases in the recent years. It is therefore vital to find ways to increase pig production in South Africa. The demand of higher growth animals has led to the introduction of international commercial breeds of the Large White, South African Landrace and Duroc in the villages. The genetic pool of indigenous breeds is on the other hand degenerating. In order to have proper develop conservation strategies for long-term sustainable use, characterization of pig populations especially indigenous, village and wild pigs where knowledge is still limited is important.

There have been very limited studies on the genetic diversity and population structures of South African pigs. Few previous studies were undertaken and these focused on few breeds and restricted to less comprehensive genetic markers that would not allow an in-depth analysis of the genetic diversity and genomic architecture (Swart *et al*., 2010; Halimani *et al*., 2012). The characterization of South African pig populations by Swart *et al*. (2010) and Halimani *et al*. (2012) over 10 years ago used outdated microsatellite markers, limited populations of 22 households in the OR Tambo and Alfred Nzo districts (Halimani *et al*., 2012) while Swart *et al*. (2010) characterised Kolbroek, Large White, South Africa Landrace and Duroc breeds. The study by Swart *et al*. (2010) observed the genetic differentiation, structure and diversity levels. Halimani *et al*. (2012) analysed the heterozygosity, inbreeding and genetic differentiation levels.

The Food and Agricultural Organisation, World Watch Lists (FAO, 2000) highlighted the dangers that face indigenous and local breeds and emphasised the need for identification and characterisation of breeds and populations to facilitate their conservation. According to the FAO, majority of species and breeds are lost because nothing is known about them. Failure to

characterise breeds and populations therefore predispose indigenous animal genetic resources to risk of extinction (FAO, 2000).

A range of pig production systems and breeds are utilized in South Africa. It makes more sense for rural farmers to keep indigenous breeds, such as Kolbroek and Windsnyer that are able to survive with low quality feed, known to be hardy and tolerant/resistant to disease resistance and environmental stress. While the commercial breeds play an important role of feeding urban populations and contributing to around 2.45% towards the GDP for the country (DAFF, 2019). Pigs kept under free ranging/backyard can survive and produce optimally under low-inputs system where they move freely and scavenge for their own food. Warthog, Wild and Bush Pigs are prevalent in the villages and National Parks of South Africa. Wild Boars and Warthogs also play an important role of as a source of protein globally (Sales & Kotrba, 2013; Swanepoel *et al*., 2016). Furthermore, Bush Pigs hunted as bush meat in African countries (Pearce, 2005; Ebewore *et al*., 2015). In general, wild pigs contributes to long-term improvement of biodiversity. South African National parks and zoos are home to abandoned and rescued animals including exotic and wild pigs such as the Vietnamese Potty Belly.

Overall, the South African pig population is a complex one consisting of multiple breeds and species distributed across wide and contrasting geographical spread and raised under diverse production systems facing different natural and artificial selection pressures. Disease control remains a major challenge in the pork industry. Indigenous and village pigs might have contacts with wild animals such as Warthogs, Wild Boars and Bush Pigs as they either roam freely or escape from neighbouring game reserves and therefore increasing chances of disease transmission. Diseases such as Classical Swine Fever (CSF), African Swine Fever (ASF) and Foot and mouth diseases (FMD) have an ability to cause major economic loses for the farmers and economy at large (Dixon *et al*., 2019). Production of indigenous and village pigs might be challenged by unknown parasites and worms. Intensive has its own challenges as pigs kept together can spread infectious diseases easily amongst the droves because they share feed and water.

The continuous interaction of breeds and populations including domestic and wild pigs presents both opportunities of increased genetic diversity and hybrid vigour and challenges of sharing of diseases, genetic erosion, and dilution particularly of the indigenous and village populations that are raised in extensive production systems. Again, the limited information on extant gene

flow, crossbreed and introgression patterns makes it difficult to assess the impact of such interactions.

This study took advantage of the availability of the Porcine SNP60K bead chip as a valuable genomic resource to study pig populations from different genetic and production backgrounds as confirmed by previous studies (Burgos-Paz *et al*., 2013). Genotyping pigs on the Porcine SNP60K array allows one to conduct comprehensive analysis from background genetic diversity and population structure (Burgos-Paz *et al*., 2013; Ai *et al*., 2013; Chen *et al*., 2017; Osei-Amponsah *et al*., 2017; Mujibi *et al*., 2018); LD and haplotype block patters (Badke *et al*., 2012; Ai *et al*., 2013; Veroneze *et al*., 2013); gene flow and introgression patterns (Bosse *et al*., 2015; Mujibi *et al*., 2018; Chen *et al*., 2020) and signature of selection analysis (Bosse *et al*., 2012; Badke *et al*., 2012; Ai *et al*., 2013; Burgos-Paz *et al*., 2013; Borowska *et al*., 2017; Chen *et al*., 2018). All these and other analysis are necessary to achieve an in-depth understanding of the local pig gene pool including making inferences of forces of evolution at play and their impact on the productivity and sustainability of populations. The Porcine SNP60K bead chip is a universal tool that can allow studies to (i) share genotypes across studies and undertake global analysis and (ii) with certain considerations compare results between populations analysed from different studies. Such opportunities were limited with previous markers that were predominantly variable between studies.

7.2 Aims and objectives

The overall aim of the research was to undertake a comprehensive characterisation of genetic diversity and genomic architecture of South African pig gene pool, with the ultimate goal of providing baseline information towards their inventory, optimal usage and conservation. To achieve this the Illumina Porcine SNP60K bead chip was used for extensive genome characterization of variety of pig breeds distributed across the total landscape of South Africa and included commercial, indigenous, non-descript village, wild populations and captured populations in zoos and national parks. In addition, and to allow interpretation of findings in the global context, the study made use of publicly available Porcine SNP60K genotypes from other global populations for joint analysis. The overall aim of the study was achieved through the following specific objectives:

- Investigate the genetic diversity and population genetic structure of SA pig populations in comparison with global populations (Chapter 3).

- Evaluate and describe linkage disequilibrium and haplotype block structure in village, indigenous and commercial pigs, and to scan for existence of genes in shared haplotype blocks that may have effect on traits of economic importance (Chapter 4).

- Investigate the evolutionary history of SA pigs using genome-wide SNP data. This analysis tried and disentangled the demographic history looking at genetic ancestries, introgression and gene flow and assisted in tracing the migration events between indigenous, commercial, village and wild pig populations from different production systems (Chapter 5).

- Screen for genomic regions showing signatures of selection as well as candidate genes and functional pathways associated with potential effect on phenotypic traits of importance in the different production environments (Chapter 6).

7.3 General discussion

Chapter 3 revealed the bias towards the design of the bead chip with higher number of polymorphic SNPs remaining after quality control for commercial breeds. It is because only breeds such as the Large White, Duroc and Landrace were used in the development towards the chip. Ascertainment bias was reported in other studies in pigs (Ramos *et al*., 2009) and different South Africa studies on different species of cattle (Makina *et al*., 2015), chickens (Khanyile *et a*l., 2015), goats (Mdladla *et al*., 2016) and sheep (Sandenbergh *et al*., 2016). Breeds that were not used in the discovery of SNPs and development of SNP panel retained lesser SNPs for downstream analysis. Such an observation calls for (i) continued and additional studies on local breeds for SNP discovery and (ii) inclusion of local breeds in future upgrades or new developments of SNP panels. Examples of such initiatives have been observed in goats on IGGC (Tosser-Klopp *et al*., 2014) and AdaptMap (Stella *et al*., 2018); and 1000 Bull Genomes Project on cattle (Hayes & Daetwyler, 2019).

Despite the expected ascertainment bias, a considerable number of SNPs were retained for downstream analysis and were used for further investigations. The results in this analysis revealed genetic diversity and population genetic diversity that were driven by the genetic background (different breed categories) and production environments. Breeds that had a common founder populations' *e.g.,* Large White and South African Landrace commercial

breeds as well as free roaming Warthogs and Bush Pigs clustered together as well as breeds from similar production systems *e.g*., village populations from Mopani, Capricorn, Alfred Nzo and O.R. Tambo. Increasing risks of genetic drifts and inbreeding were observed for commercial breeds that demonstrated moderate heterozygosity indices and relatively elevated inbreeding co-efficient. Commercial breeds raised in the intensive system and mainly selected for traits of higher lean meat causing small effective population sizes. In contrast, that village pigs represent an important genetic reservoir which might be because of mating system and gene flow amongst populations. Population effective size was higher in the villages compared to indigenous and commercial breeds. Population structure also demonstrated high levels of admixture in village populations with the exception of Alfred Nzo pigs, which might have continuously experienced population bottlenecks and geographical restriction during the ASF outbreaks that are prevalent in Alfred Nzo, and O.R. Tambo districts of the Eastern Cape province. These outbreaks have occurred in 2006 and as recent as May 2020 with farmers having to cull all their animals. This therefore means farmers have to restock again, which then causes bottlenecks and fragmentation and prevented gene flow. This study also suggests that genetic improvement programs be set up for Alfred Nzo population as there is a potential in this ecotype for benefits of rural farmers.

The observed genetic diversity and population genetic structures observed in Chapter 3 were confirmed in Chapter 4, that focused on the LD patterns and haplotype blocks structure of commercial, indigenous and village populations. The intensively raised Duroc and Kolbroek had the highest LD levels compared to other populations, demonstrating the effect of production systems on LD and supports the low genetic diversity and effective population size levels displayed in Chapter 3 for these breeds. Village pigs consistently demonstrated high diversity as reflected by the low LD levels and high frequency of short haplotype blocks whereas the high LD levels in commercial and indigenous reflected on the role of artificial selection, low genetic diversity and narrow genetic base and limited gene flow between commercial, indigenous, and other breeds and populations. LD decay patterns reveals low and short-range LD particularly in the village pigs but also in all SA pigs as a whole, with implications raised on the utility of the SNP60K bead chip for genomic applications such as genomic selection and QTL mapping. Previous studies have advocated for an LD of > 0.3 (O'Brien *et al*., 2014) for meaningful genomic predictions and QTL and GWAS mapping studies (O'Brien *et al*., 2014). A denser SNP panel is always recommended when populations are characterised by low and short LD (Ai *et al*., 2013).

Persistence of phase was able to provide relatedness across populations. This revealed a potential of using SNP60K data from reference populations for genomic evaluation and improvement on selection programs. As LD decay, increased marker distance yielded lower *r* values. Correlation between commercial breeds was the highest, indicating high degree of similarity and possibilities of shared genomic tools and across breed genomic evaluations. This confirms closer relationships of domestic pigs based on the development of breeds around the world. In Sheep (Brito *et al*., 2017) demonstrated the possibility of across breed genomic evaluations in closely related breeds that could benefit from these shared reference populations because of low population sizes. The higher LD in the Kolbroek and Duroc breeds suggests effect of genetic drift, recent bottlenecks events and selection pressures in the populations. Production system seems to be a factor in haplotype blocks as they were less in village pigs compared to commercial and indigenous breeds. The unique haplotype blocks shown were able to give clarity on the independent development and genomic structuring of populations. The high number of shared haplotype blocks in villages demonstrated their admixture history and high levels of introgression with other populations.

Chapter 5 further investigated the genetic relatedness and interactions of breeds and populations. The indigenous breeds of Windsnyer and Kolbroek, the commercial breeds of South African Landrace, Large White and Duroc, as well as the Vietnamese Potbelly and Wild Boar, all assigned to single genomic clusters by giving membership coefficients estimates ranging from 92 % to 100 %. The village populations on the other hand assigned to several clusters that implied crossbreeding and admixture within these non-descript village pigs. Gene flow and shared IBD analyses were able to uncover genetic history of shared ancestries between Mopani, Capricorn and O.R. Tambo and ongoing interactions in Capricorn which populations have continuous gene flow between populations. The major concern of continuous gene flow resulting in introgression is the resultant dilution and genetic erosion often witnessed in local indigenous breeds and populations. FAO (2007a; 2009) and other reports (Madzimure *et al*., 2012) highlights the need to manage indiscriminate crossbreeding or mating of indigenous genetic resources and commercial exotic breeds as this is a major risk to their unique biodiversity that will subsequently be lost through dilution or erosion (Berthouly-Salazar *et al*., 2012). In combination, the results from Chapter 3 of high admixture and weak population boundaries, Chapter 4 (low and short range LD as well as low number and short haplotype blocks) and Chapter 5 (mixed ancestry, gene flow and introgression) highlights the major

concerns around non-descript village pigs and other livestock species (village chickens - Khanyile *et al.,* 2015; village goats - Mdladla *et al.,* 2016) that their existence and sustainable use is threatened by replacement by other breeds and populations in most cases from commercial and exotic breeds. In the case of South African pigs, their existence as a unique and distinct genetic resource is further threatened by the co-existence and interactions with wild populations.

An additional analysis in Chapter 5 was the annotation of genomic regions within shared IBD segments revealed genes and genetic mechanisms also observed in Chapters 3 and 4 associated with growth, meat and carcass traits, reproduction and immune response as an indication of traits that are under either natural or artificial selection in these populations. Chapter 6 went on further to do an in-depth analysis of signatures of selection using three different but complementary methods of *iHS*, *XP-EHH* and *HapFLK*. The analyses were able to provide insights on the signatures harboured by the different pig breeds and populations. The *iHS* method showed that they were more genomic regions harbouring signatures of selections within the village populations. Genomic regions on chromosomes 1, 7, 8, 16 and 14 emphasis influences traits of economic (growth, meat and carcass quality, and reproduction) and adaptive (disease resistance, immune response) importance in the village pig populations that are predominantly under natural selection. *XP-EHH* harboured QTLs associated with economic importance such as meat and carcass, diseases, adaptability, reproductive, growth, morphological and growth in the different pairs of populations compared. Of worth to mention is IMF content found in genomic regions 113,774,412-114,206,930:2 (DUR_LWT), 72,326,714-72,391,648:2 (DUR_VIT, VIT_LIM), 113,414,264-113,429,936:14 (WAT_LIM), 105,804,586-105,845,725:4 (WAT_LWT), 29,895,878-29,936,233:18 (VIT_EC, VIT_LIM, VIT_IND). The *iHS* detected genes such as *F2* (chromosome 2), *SHISAL1* (chromosome 5) and *GHRL* (chromosome 13). These genes are associated with growth (average daily gain, feed conversion ratio, body weight at weaning and end of test), fertility (litter size) and meat and carcass traits (backfat, ham weight, lean meat, IMF content and marbling). This suggests introgression and interbreeding between Wild Boars and domestic pigs as these traits are associated with artificial selection used by the farming industry. Results from the *XP-EHH* method between Warthog and Villages revealed signatures mostly from coming from the Warthog such as *AMBP* (chromosome 1), *NPW* (chromosome 3) *PSD* (chromosome 14) and *TMRSS* (chromosome 9). Genes linked to meat and carcass traits and suggests, Warthogs are losing their natural adaptive traits. Genetic erosion also displayed in the DUR_WBO signatures

of *ESR2* gene on chromosome 1 associated with fertility QTLs such as litter size, total number born alive, semen volume, sperm concentration and motility. Findings in this study even shows that village pig populations have been under selection for meat and carcass traits, which denounces the perception of indigenous and village pigs having no ability to provide the consumers with meat of desired yield and quality. The strong signals towards meat and carcass quality traits points to both natural and artificial selection pressures. Overall, the result of this study demonstrates the role that selection provides in shaping the genomic landscape of domestic and wild pig populations and enhances the understanding of the phenotypic and genotyping traits of relevance in these populations.

7.4 Conclusions

(i) Chapter 3 showed that the genetic background and the production system influenced the genomic architecture of South African pig

(ii) Chapter 4 displayed the use of Porcine SNP60K as a genomic tool for future breeding and conservation strategies in pig populations of South Africa

(iii) Chapter 5 demonstrated the domestication events, traders, ancestries, and migration having an effect on the development South African pig population.

(iv) Chapter 6 revealed important candidates' genes and QTLs within and between populations

7.5 Study challenges

The greatest challenge in this study was getting adequate samples from Vietnamese Potbelly, Wild Boars and Bush Pigs. The numbers of these populations are presumably low in South Africa and the animals are currently found in a few areas of National Parks and Zoos. Sampling in the National Parks was also a challenge as the animals will be scattered within a spread-out area with no information on what animals are where.

As a result of the small sample size, certain analysis could not be done, which overall limited the findings from the study. An increase in sample size of these populations from several geographic locations will allow more in-depth analysis of these breeds and provide a more holistic picture.

Illumina Porcine SNP60K bead chip first published in 2009 has been widely applied and found utility in most breeds and populations. High ascertainment bias was observed particularly in the village and more so in the wild populations calling for a more representative SNP panel.

LD was also low in these breeds also calling for a denser SNP panel. It would have been best to use a higher resolution Porcine *version* 3.0 chip which contains ~80K SNPs to perform molecular studies. Unfortunately, it only became available after this study had commenced. A higher density 650K SNP chip from Affymetrix will also be soon available providing possibility of even more in-depth analysis than provided in this study. Cost was a major challenge in this study and as a result not as many animals were genotyped even when the samples were available. SNP genotyping is also becoming less expensive thus providing a tool to probe genetic variation across several pig species and breeds and allowing more in-depth sampling within a breed or population to enable more and comprehensive analysis. The use of SNPs to determine genetic variability and for routine assessments of trends in genetic diversity particularly for threatened breeds and populations can be encouraged and be feasible with reduced cost of genotyping.

7.6 Final comments and recommendations

Based on the findings of this study within the acknowledged limitations, this study makes the following recommendations:

(i) Proper management of indigenous breeds and non-descript village pigs to ensure maintenance and conservation of genetic diversity. It is recommended that indigenous stock is moved from research facilities and possibly be distributed in villages to conserve their genetic pool and improve production. Lack of breeding programs and uncontrolled crossbreeding with commercial breeds were confirmed in the village populations. This poses as a major threat to the nation's genetic resources as it contributes to genetic erosion. Awareness drives should be established to emphasise and educate rural communities of the unique traits and potential of increase of performance from indigenous breeds such as Kolbroek and Windsnyer. Furthermore, this study has demonstrated the potential of increased genetic diversity that can be achieved by village populations by implementing breeding and genetic programs. This will also improve production by maximizing heterozygosity.

(ii) Proper management of wild populations to ensure maintenance and conservation of genetic diversity as well as optimal usage of their genetic potential. The study has additionally provided a baseline of wild pigs' populations that can be utilised for spread of pig diseases for protection and mutually beneficiation of domestic and wildlife industries.

(iii) Commercial breeds contribute immensely to the South African pork industry because of their fast growth rates feed to meat conversion ratio thus contributing to the country's GDP. Therefore, results from this study can be used to improve response to selection to be able to maintain acceptable genetic levels.

(iv) Capture populations like the Vietnamese Potbelly have spread worldwide and crossed with other pig breeds to widen the genetic diversity. Their small and easy to handle friendly nature has also made them be used as pets as well as in scientific research. Proper management of the Vietnamese Potbelly is important, as it has been valuable in development of other pig breeds.

Overall, this study provided comprehensive baseline information on the South African pig population. The information can be used to design and implement strategies to manage both domestic and wild pig populations and implement optimal usage, improvement, and conservation programs. Further studies are recommended to build up from this study pursing other genomic approaches such as (i) gene expression profiling to further understand the functional attributes of the highlighted genomic regions (ii) epi-genomic approaches to build an understanding of both the genomic and environmental influences on phenotypes and (iv) microbiome and other second genome profiling approaches that can shed more light on the genetic and environmental complexities influencing the production and performance of these pig populations. In addition, findings from this study need to be shared with the broader stakeholder to guide ongoing and new breed improvement initiatives.

REFERENCES

Adeola AC, Oluwole OO, Oladele BM, et al. 2017. Analysis of the genetic variation in mitochondrial DNA, Y-chromosome sequences, and MC1R sheds light on the ancestry of Nigerian indigenous pigs. Genetics Selection Evolution 49: 52.

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.

Ajala, M. K., A. O. K. Adesehinwa and A. K. Mohammed, 2007. Characteristics of smallholder pig production in southern Kaduna area of Kaduna State, Nigeria. American-Eurasian Journal Agriculture and Environmental Science 2: 182-188.

Akey JM, Zhang G, Khang K, et al. 2002. Interrogating a high-density SNP map for signatures of natural selection. Genome Research 12: 1805-1814.

Albarella U, Manconi F, Vigne JD, et al. 2007. The ethnoarchaeology of traditional pig husbandry in Sardinia and Corsica. In Albarella U, Dobney K, Ervynck A (Eds.) Pigs and Humans: 10,000 years of interaction, 285-307. Oxford, Oxford University Press.

Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12: 246.

Alexandratos N, Bruinsma J. 2012. World agriculture towards 2030/2050: The 2012 Revision; ESA Working Paper No. 12-03. FAO: Rome, Italy.

Allendorf FW, Leary RF, Spruell P, et al. 2001. The problems with hybrids: Setting conservation guidelines. Trends in Ecology and Evolution 16: 613-622.

Alves E, Fernandez AI, Fernandez-Rodriguez A, et al. 2009. Identification of mitochondrial markers for genetic traceability of European Wild Boars and Iberian and Duroc pigs. Animals 3: 1216-23.

Amar S, De Boni L, de Voux A, et al. 2021. An outbreak of African swine fever in small-scale pigs, Gauteng, South Africa, July 2020. International Journal of Infectious Diseases 110: S44-S49.

Amaral AJ, Megens HJ, Crooijmans RP, et al. 2008. Linkage disequilibrium decay and haplotype block structure in the pig. Genetics 179: 569-579.

Amaral AJ, Ferretti L, Megens HJ, et al. 2011. Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. PLoS One, 6(4): e14782.

Amills M, Clop A, Ramírez O, et al. 2010. Origin and diversity of pig breeds. Encyclopedia of Life Sciences. Chichester: John Wiley & Sons.

Amills M, Ramírez O, Galman-Omitogun O, et al. 2012. Domestic pigs in Africa. African Archaeological Review 30: 73-82.

Anderson DE. 1991. Genetic study of eye cancer in cattle. Journal of Heredity 82: 21-6.

Anderson E. 1949. Introgressive hybridization. (New York: John Wiley and Sons).

3722 Anderson E, Hubricht L. 1938. Hybridization in Tradescantia. III. The evidence for introgressive
3723 hybridization. American Journal of Botany 25: 396-402.

3724 Anderson EC, Hutchings GH, Mukarati N, et al. 1998. African swine fever virus infection of the bushpig
3725 (Potamochoerus porcus) and its significance in the epidemiology of the disease. Veterinary
3726 Microbiology 62: 1-15.

3727 Andrews KR, Karczmarski L, Au WWL, et al. 2010 Rolling stones and stable homes: social structure,
3728 habitat diversity and population genetics of the Hawaiian spinner dolphin (*Stenella longirostris*).
3729 Molecular Ecology 19: 732-748.

3730 Antwi M, Seahlodi P. 2011. Marketing constraints facing emerging small-scale pig farmers in Gauteng
3731 province, South Africa Journal of Human Ecology 36: 37-42.

3732 Archibal AL, Haley CS, Brown JF, et al. 1995. The PiGMaP consortium linkage map of the pig (Sus
3733 Scrofa). Mammalian Genome, 6, 157-175.

3734 Ardlie KG, Kruglyak L, Seielstad M. 2002. Patterns of linkage disequilibrium in the human genome.
3735 Nature Reviews Genetics 3: 299-309.

3736 Arias JA, Keehan M, Fisher P, et al. 2009. A high density linkage map of the bovine genome. BMC
3737 Genetics 10: 18.

3738 Ashburner M, Ball CA, Blake JA, et al. 2000. Gene ontology: tool for the unification of biology. Nature
3739 Genetics 25: 25-29.

3740 AU-IBAR. 2015. Local African Pig. African Union Inter-African Bureau for Animal Resources.
3741 Available at http://www.au-ibar.org/2012-10-01-13-08-42/features/series/know-
3742 youranimalsseries/local-african-pig. [Accessed: 20 August 2019].

3743 Avila F, Mickelson JR, Schaefer RJ, et al. 2018. Genome-wide signatures of selection reveal genes
3744 associated with performance in American quarter horse subpopulations. Frontiers in Genetics 9: 249.

3745 Ayuso M, Fernández A, Núñez Y, et al. 2015. Comparative analysis of muscle transcriptome between
3746 pig genotypes identifies genes and regulatory mechanisms associated to growth, fatness and
3747 metabolism. PLoS ONE 10: e0145162.

3748 Badke YM, Bates RO, Ernst CW, et al. 2012. Estimation of linkage disequilibrium in four US pig breeds.
3749 BMC Genomics 13: 24.

3750 Ban DM, Zhang B, Wang ZX, et al 2015. Differential gene expression of epigenetic modifying enzymes
3751 between Tibet pig and Yorkshire in high and low altitudes. Genetics and Molecular Research 14:
3752 3274-3280.

3753 Barbato M, Orozcoter Wengel P, Tapio M, et al. 2015. SNeP: A tool to estimate trends in recent effective
3754 population size trajectories using genome-wide SNP data. Frontiers in Genetics 6: 109.

3755 Baskett ML, Gomulkiewicz R. 2011. Introgressive hybridization as a mechanism for species rescue.
3756 Theoretical Ecology 4: 223-239.

3757 Belbin GM, Odgis J, Sorokin EP, et al. 2017.Genetic identification of a common collagen disease in
3758 Puerto Ricans via identity-by-descent mapping in a health system. eLife 6: e25060.

Berihulay H, Abied A, He X, et al. 2019. Adaptation mechanisms of small ruminants to environmental heat stress. Animals 9:75.

Berthouly-Salazar C, Thévenon S, Van TN, et al. 2012. Uncontrolled admixture and loss of genetic diversity in a local Vietnamese pig breed. Ecology and Evolution 2: 962-75.

Bertolini F, Servin B, Talenti A, et al. 2018. Signatures of selection and environmental adaptation across the goat genome post-domestication. Genetics Selection Evolution 50: 57.

Bester J, Kusel US. 1998. Early domesticated animals in South Africa. Proceedings of the 4th Global Conference on Conservation of Domestic Animal Genetic Resources, Lallitpur, Nepal, 17-21 August 1998.

BFAP - Bureau for Food and Agricultural Policy (BFAP). 2020. Baseline Outlook for South African. Agriculture: 2020-2029. BFAP. Pretoria. Available at https://www.bfap.co.za/wpcontent/uploads/2020/08/BFAP-Baseline-2020.pdf. [Accessed: 15 July 2022].

BFAP - Bureau for Food and Agricultural Policy (BFAP). 2013. Evaluating the South African pork value chain. October. Available at: http://www.bfap.co.za/wpcontent/uploads/reports/BFAP%20Pork%20value%20chain%202013_Final.pdf. [Accessed: 28 November 2022].

BFAP - Bureau for Food and Agricultural Policy (BFAP). 2021. Quantification of the South African Pork value chain: Phase 1. Available at https://www.bfap.co.za/wp-content/uploads/2021/07/Pork-Value-Chain-Study.pdf [Accessed: 09 December 2022].

BFAP/SAPPO. 2021. Summary of the BFAP research report: Explaining consumption and procurement preferences. Published 05 October 2021. Available from: https://sappo.org/lets-talk-indepth/summary-of-the-bfap-research-report-explaining-consumption-and-procurement-preferences/. [Accessed: 14 December 2022].

Bie Q, Jin C, Zhang B, et al. 2017. IL-17B: A new area of study in the IL-17 family. Molecular Immunology 90: 50-56.

Bihan-Duval EL, Hennequet-Antier C, Berri C, et al. 2018. Identification of genomic regions and candidate genes for chicken meat ultimate pH by combined detection of selection signatures and QTL. BMC Genomics 19: 294.

Bijma P, Mulder HA, Ellen ED. 2002. Sustainable livestock genetic improvement. Proceedings of the Association for the Advancement of Animal Breeding and Genetics 17: 235-239.

Bittles AH. 2010. Consanguinity, genetic drift, and genetic diseases in populations with reduced numbers of founders. M. Speicher et al. (Eds.). Vogel and Motulsky's Human Genetics: Problems and Approaches.

Blair MW, Cortés AJ, Farmer AD, et al. 2018. Uneven recombination rate and linkage disequilibrium across a reference SNP map for common bean (Phaseolus vulgaris L.). PloS ONE 13: e0189597.

Blench RM, MacDonald KC. (Eds). 2000. The origins and development of African livestock. Archaeology, genetics, linguistics and ethnography. UCL Press: Taylor and Francis Group.

3797 Blaney J. 2004. History of the US potbellied pig. 6th Annual Pet Pig Symposium Ames, June 4-6, 1A.

3798 Blomstrom AL, Stahl K, Masembe C, et al. 2012. Viral metagenomic analysis of bushpigs
3799 (Potamochoerus larvatus) in Uganda identifies novel variants of Porcine parvovirus 4 and Torque
3800 teno sus virus 1 and 2. Virology Journal 9:192.

3801 Bonhomme M, Chevalet C, Servin B, et al. 2010. Detecting selection in population trees: the Lewontin
3802 and Krakauer test extended. Genetics 186: 241-62.

3803 Borowska A. Reyer H, Wimmers K, et al. 2017. Detection of pig genome regions determining
3804 production traits using an information theory approach. Livestock Science 205: 31-35.

3805 Bosse M, Megens HJ, Madsen O, et al. 2012. Regions of homozygosity in the porcine genome:
3806 consequence of demography and the recombination landscape. PLoS Genetics 8: e1003100.

3807 Bosse M, Megens HJ, Frantz LA, et al. 2015. Genomic analysis reveals selection for Asian genes in
3808 European pigs following human-mediated introgression. Nature Communications 5: 4392.

3809 Bosse M, Megens HJ, Madsen O, et al. 2014. Untangling the hybrid nature of modern pig genomes: a
3810 mosaic derived from biogeographically distinct and highly divergent Sus scrofa populations.
3811 Molecular Ecology 23: 4089-4102.

3812 Botha SA. 1989. Feral pigs in the Western Cape Province: failure of a potentially invasive species. South
3813 African Forage Journal 151: 17-25.

3814 Bovula N. 2017. Response in fertility of Windsnyer boars to increasing α-tocopherolsupplementation.
3815 Agriculture (Animal Science) dissertation, University of KwaZulu-Natal, South Africa.

3816 Briggs HM. 1983. International pig breed Encyclopedia. Elanco Animal Health, Indianapolis, USA.

3817 Brisbin A, Bryc K, Byrnes J, et al. 2012. PCAdmix: principal components-based assignment of ancestry
3818 along each chromosome in individuals with admixed ancestry from two or more populations. Human
3819 Biolology 84: 343-364.

3820 Brito LF, McEwan JC, Miller SP, et al. 2017. Genetic diversity of a New Zealand multi-breed sheep
3821 population and composite breeds' history revealed by a high-density SNP chip. BMC Genetics 18:
3822 25.

3823 Broman KW, Weber JL. 1999. Long homozygous chromosomal segments in reference families from
3824 the centre d'Etude du polymorphisme humain. American Journal of Human Genetics, 65: 1493-500.

3825 Browett S, McHugo G, Richardson IW, et al. 2018. Genomic characterisation of the indigenous Irish
3826 Kerry cattle breed. Frontiers in Genetics 9:51.

3827 Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference
3828 for whole-genome association studies by use of localized haplotype clustering. American Journal of
3829 Human Genetics 81: 1084-1097.

3830 Browning S, Browning B. 2011. Haplotype phasing: existing methods and new developments. Nature
3831 Reviews Genetics 12: 703-714.

3832 Browning SR, Browning BL. 2012. Identity by descent between distant relatives: detection and
3833 applications. Annual Review of Genetics 46: 617-633.

Buchanan DS, Stalder K. 2011. Breeds of pigs. M. F. Rothschild & A. Ruvinsky (Eds.). The Genetics of the Pig (2nd ed.). Oxfordshire, UK: CAB International.

Bulmer MG. 1971. The effect of selection on genetic variability. The American Naturalist 105: 201-211.

Burgos-Paz W, Souza C, Megens H. et al. 2013. Porcine colonization of the Americas: a 60k SNP story. Heredity 110: 321-330.

Bush WS, Moore JH. 2012. Chapter 11: Genome-Wide association studies. PLoS Computational Biology 8: e1002822.

Calabrese B. 2019. Linkage disequilibrium. Encyclopedia of Bioinformatics and Computational Biology 1: 763-765.

Calus MPL, Meuwissen THE, de Roos APW, et al. 2008 Accuracy of genomic selection using different methods to define haplotypes. Genetics 178: 553-561.

Cesar AS, Regitano LC, Reecy JM, et al. 2018. Identification of putative regulatory regions and transcription factors associated with intramuscular fat content traits. BMC Genomics 19: 499.

Chang CC, Chow CC, Tellier LC, et al. 2015. Second-generation PLINK: Rising to the challenge of Larger and Richer Datasets. Gigascience 4: 7.

Charlier C, Farnir F, Berzi P, et al. 1996. Identity-by-descent mapping of recessive traits in livestock: Application to map the bovine syndactyly locus to chromosome 15. Genome Research 6: 580-589.

Chen H, Huang M, Yang B, et al. 2020. Introgression of Eastern Chinese and Southern Chinese haplotypes contributes to the improvement of fertility and immunity in European modern pigs, GigaScience 9.

Chen K, Baxter T, Muir WM, et al. 2007 Genetic resources, genome mapping and evolutionary genomics of the pig (Sus scrofa). International Journal of Biological Sciences 3: 153-165.

Chen M, Su G, Fu J. et al. 2017. Population admixture in Chinese and European Sus scrofa. Scientific Reports 7: 13178.

Chen M, Wang J, Wang Y, et al. 2018. Genome-wide detection of selection signatures in Chinese indigenous Laiwu pigs revealed candidate genes regulating fat deposition in muscle. BMC Genetics 19: 31.

Chen W, Abecasis GR. 2007. Family-based association tests for genomewide association scans. The American Journal of Human Genetics 81: 913-926.

Chen X, Cheng Z, Zhang S, et al. 2015. Combining genome-wide association studies and differential gene expression data analyses identifies candidate genes affecting mastitis caused by two different pathogens in the dairy cow. Open Journal of Animal Science 5: 358-393.

Chen Y, Li X, Li J. 2010.A novel approach for haplotype-based association analysis using family data. BMC Bioinformatics 11: S45.

Chimonyo M, Bhebhe E, Dzama K, et al. 2005. Improving smallholder pig production for food security and livelihoods of the poor in Southern Africa. Proceedings of African Crop Sciences Conference 7: 569-573.

Chimonyo M, Dzama K. 2007. Estimation of genetic parameters for growth performance and carcass traits in Mukota pigs. Animals 2: 317-323.

Choi I, Bates RO, Raney NE, et al. 2012. Evaluation of QTL for carcass merit and meat quality traits in a US commercial Duroc population. Meat Science 92: 132-138.

Choi I, Steibel JP, Bates RO, et al. 2010. Application of alternative models to identify QTL for growth traits in an F2Duroc x Pietrain pig resource population. BMC Genetics 11: 97.

Choi I, Steibel JP, Bates RO, et al. 2011. Identification of carcass and meat quality QTL in an F2 Duroc x Pietrain pig resource population using different least-squares analysis models. Frontiers in Genetics 2: 18.

Choi SK, Lee JE, Kim YJ, et al. 2014. Genetic structure of wild boar (Sus scrofa) populations from East Asia based on microsatellite loci analyses. BMC Genetics 15: 85.

Ciobanu MM, Postolache AN, Lipsa FD, et al. 2022. Meat fatty acid composition of wild boars hunted in Romania in relationship to gender and age-class. Animals 12: 810.

Ciofi C, Beaumont MA, Swingland IR, et al. 1999. Genetic divergence and units for conservation in the Komodo dragon Varanus komodoensis. Proceedings of the Royal Society London B 266: 2269-2274.

Clark AG, Glanowski S, Nielsen R, et al. 2003. Inferring non-neutral evolution from human-chimp-mouse orthologous gene trios. Science 12: 1960-1963.

Clegg SM, Phillimore AB. 2010. The influence of gene flow and drift on genetic and phenotypic divergence in two species of Zosterops in Vanuatu. Philosophical Transactions of the Royal Society B 365: 1077-1092.

Clop A, Ovilo C, Perez-Enciso M, et al. 2003. Detection of QTL affecting fatty acid composition in the pig. Mammalian Genome 14: 650-656.

Cook H. 1978. Suid evolution and correlation of African horminoid localities: an alternative taxonomy. Science: 201: 460-463.

Coleman A. 2018. The indigenous Kolbroek pig - a South African success story. Available at: https://www.farmersweekly.co.za/animals/pigs/indigenous-kolbroek-pig-south-african-success-story/. [Accessed: 14 September 2021].

Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: An R package for the visualization of intersecting sets and their properties. Bioinformatics 33: 2938-2940.

Corapi WV, Rodrigues A, Lawhorn DB. 2011. Mucinous Adenocarcinoma and T-Cell Lymphoma in the Small Intestine of 2 Vietnamese Potbellied Pigs (Sus scrofa). Veterinary Pathology 48: 1004-1007.

Corbin LJ, Liu AY, Bishop SC, et al. 2012. Estimation of historical effective population size using linkage disequilibrium with marker data. Journal of Animal Breeding and Genetics 129: 257-270.

Cornelis MC, Hu FB. 2013. Systems epidemiology: a new direction in nutrition and metabolic disease research. Current Nutrition Reports 2: 225-235.

Crow JF, Kimura M. 1970. An Introduction to Population Genetics Theory. Harper and Row: New York, USA.

Cumming DHM. 1975. A field of study of the ecology and behaviour of Warthog Museum memoir no. 7. Trustees of the National Museums and Monuments of Rhodesia Salisbury.

Cumming DHM. 1999. Study on the development of transboundary natural resource management areas in Southern Africa - Environmental Context. Natural Resources, Land Use, and Conservation. Biodiversity Support Program. Washington DC, USA.

Cupido M. 2005. An insight into the livelihoods of small-scale pig farmers in the Western Cape, South Africa. MSc in Animal Sciences thesis, Stellenbosch University, South Africa.

Cuyabano BC, Su G, Lund MS. 2015. Selection of haplotype variables from a high density marker map for genomic prediction. Genetics Selection Evolution 47: 61.

D'Huart J, Grubb P. 2003. Distribution of the common Warthog (Phacochoerus africanus) and the desert Warthog (Phacochoerus aethiopicus) in the Horn of Africa. African Journal of Ecology 39: 156-169.

DAFF. 2019. A profile of the South African pork market value chain, Department of Agriculture, Forestry and Fisheries. Available at https://www.dalrrd.gov.za/doaDev/sideMenu/Marketing/Annual%20Publications/Pork%20Market%20Value%20Chain%20Profile%202019.pdf. [Accessed: 20 July 2020].

Dall'Olioa S, Ribania A, Moscatelli G, et al. 2018. Teat number parameters in Italian Large White pigs: Phenotypic analysis and association with vertnin (VRTN) gene allele variants. Livestock Science 210: 68-72.

DALRRD. 2020. South Africa yearbook 2019/20. Available at: https://www.gcis.gov.za/sites/default/files/docs/resourcecentre/yearbook/Yearbook2020-3-Agriculture.pdf [Accessed: 07 December 2022].

DALRRD-Department of Agriculture, Land Reform and Rural Development. 2021. Newsletter: National Livestock Statistics. February, 2021. Available at: https://www.dalrrd.gov.za/Home/Crop-Estimates/Statistical-Information/Livestock. [Accessed: 14 September 2021].

Darwin C. 1868. The variation of animals and plants under domestication. The American Naturalist 2: 208-209.

Dash S, Chakravarty AK, Singh A, et al. 2016. Effect of heat stress on reproductive performances of dairy cattle and buffaloes: A review. Veterinary World 9: 235-244.

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

Delibes-Mateos M, Delibes A. 2013. Pets becoming established in the wild: free-living Vietnamese Potbellied pigs in Spain. Animal Biodiversity and Conservation 36: 209-215.

de Jager N, Hudson NJ, Reverter A, et al. 2013. Gene expression phenotypes for lipid metabolism and intramuscular fat in skeletal muscle of cattle. Journal of Animal Science 91: 1112-1128.

3944 de Koning DJ, Janss LL, Rattubk P, et al. 1999. Detection of quantitative traits loci for backfat thickness
3945 and intramuscular fat content in pig. Genetics 152: 1679-1690.

3946 de Lisle GW. 1994. Mycobacterial infections in pigs. Surveillance 21: 23-25.

3947 de Vries AG, Faucitano L, Sosnicki A, et al. 2000. The use of gene technology for optimal development
3948 of pork meat quality. Food Chemistry 69: 397-405.

3949 Dietze K. 2011. Pigs for Prosperity. Diversification Booklet Number 15. Food and Agriculture
3950 Organization of the United Nations, Rome, Italy.

3951 Dixon LK, Sun H, Roberts H. 2019. African Swine Fever. Antiviral Research 165: 34-41.

3952 Dobrzyn K, Szeszko K, Kiezun M, et al. 2019. In vitro effect of orexin A on the transcriptomic profile
3953 of the endometrium during early pregnancy in pigs. Animal Reproduction Science 200: 31-42.

3954 Dohner JV. 1951. The encyclopedia of historic and endangered livestock and poultry breeds. Yale
3955 University Press, Exhibits Department, New Haven, CT.

3956 Dong Q, Liu H, Li X, et al. 2014. A genome-wide association study of five meat quality traits in
3957 Yorkshire pigs. Frontiers of Agricultural Science and Engineering 1: 137-143.

3958 Du FX, Clutter AC, Lohuis MM. 2007. Characterizing linkage disequilibrium in pig populations.
3959 International Journal of Biological Sciences 3: 166-178.

3960 Du Y, Li M, Chen J, et al. 2016. Promoter targeted bisulfite sequencing reveals DNA methylation
3961 profiles associated with low sperm motility in asthenozoospermia. Human Reproduction 31: 24-33.

3962 Du FX, Clutter AC, Lohuis MM. 2007. Characterizing linkage disequilibrium in pig populations.
3963 International Journal of Biological Sciences 3: 166-178.

3964 Du ZQ, Zhao X, Vukasinovic N, et al. 2009. Association and haplotype analyses of positional candidate
3965 genes in five genomic regions linked to scrotal hernia in commercial pig lines. PLoS ONE 4: e4837.

3966 Duarte DAS. Fortes MRS., Duarte MS, et al 2017. Genome-wide association studies, meta-analyses and
3967 derived gene network for meat quality and carcass traits in pigs. Animal Production Science 58:
3968 1100-1108.

3969 Dube B, Mulugeta SD, van der Westhuizen, et al. 2011. Non-genetic factors affecting growth
3970 performance and carcass characteristics of two South African pig breeds. South African Journal of
3971 Animal Science 41 161-175.

3972 Duijvesteijn N, Veltmaat JM, Knol EF, et al. 2014. High-resolution association mapping of number of
3973 teats in pigs reveals regions controlling vertebral development. BMC Genomics 15: 542.

3974 Duran C, Appleby N, Edwards D, Batley J. 2009. Molecular genetic markers: Discovery, application,
3975 data storage and visualisation. Current Bioinformatics 4: 16-27.

3976 Ebewore SO, Ovharhe OJ, Emaziye PO. 2015. Acceptability of Bush Meat as a source of animal protein
3977 in delta state, Nigeria: implication for extension services. Journal of Northeast Agricultural
3978 University 22: 67-78.

3979 Edea Z, Bhuiyan MSA, Dessie T, et al. 2014. Genome-wide genetic diversity, population structure and
3980 admixture analysis in African and Asian cattle breeds. Animals 9: 218-226.

3981 Edea Z, Hong JK, Jung JH, et al. 2017. Detecting selection signatures between Duroc and Duroc
3982 synthetic pig populations using high-density SNP chip. Animal Genetics 48: 473-477.

3983 Egerszegi I, Ratky J, Solti L, et al. 2003. Mangalica - An indigenous swine breed from Hungary.
3984 Archives Animal Breeding 46: 245-256.

3985 Eikelenboom G, Hoving-Bolink AH, Van Der Wal PG. 1996. The eating quality of pork: The influence
3986 of intramuscular fat. Fleischwirtschaft 76: 517-518.

3987 Ellstrand NC, Elam DR 1993. Population genetic consequences of small population size: implications
3988 for plant conservation. Annual Review of Ecology and Systematics 24: 217-242.

3989 Emrani H, Torshizi RV, Masoudi AA, et al. 2017. Identification of new loci for body weight traits in F2
3990 chicken population using genome-wide association study. Livestock Science 206: 125-131.

3991 Esfandyari H, Sørensen AC, Bijma P. 2015. Maximizing crossbred performance through purebred
3992 genomic selection. Genetics Selection Evolution 47: 16.

3993 Esteve A, Kofler R, Himmelbauer H, et al. 2012. Sequencing the Iberian pig genomes. 7th International
3994 Symposium on the Mediterranean Pig 101: 21-24.

3995 Eusebi PG, Cortés O, Carleos C, et al. 2018. Detection of selection signatures for agonistic behaviour
3996 in cattle. Journal of Animal Breeding and Genetics 135: 170-177.

3997 Excoffier L, Laval G, Schneider S. 2005. ARLEQUIN ver. 3.0: an integrated software package for
3998 population genetics data analysis. Evolutionary Bioinformatics. Online 1: 47-50.

3999 Fan B, Onteru SK, Plastow GS, et al. 2009. Detailed characterization of the porcine MC4R gene in
4000 relation to fatness and growth. Animal Genetics 40: 401-419.

4001 Fang ZH, Pausch H. 2019. Multi-trait meta-analyses reveal 25 quantitative trait loci for economically
4002 important traits in Brown Swiss cattle. BMC Genomics 20: 695.

4003 FAO. 1992. The role of ruminant livestock in food security in developing countries. FAO Committee
4004 on World Food Security, 17th Session, 23 to 27 March 1992, Rome, Italy.

4005 FAO. 2000. Secondary guidelines of farm animal genetic resources management plans. Management of
4006 small populations at risk. Rome: Food and Agriculture Organization of the United Nations.

4007 FAO. 2000. World Watch List for Domestic Animal Diversity. Rome, Italy.

4008 FAO. 2005. Gene flow in animal genetic resources. A study on status, impact and trends. Rome, Italy.

4009 FAO. 2015. Climate change and food security: Risks and responses. Rome, Italy.

4010 FAO 2009. Status and trends report on animal genetic resources-2008. Rome: Food and Agriculture
4011 Organization of the United Nations. Rome, Italy.

4012 FAO. 2011. World livestock: Livestock and food supply. Rome, Italy.

4013 Fariello MI, Boitard S, Naya H, et al. 2013Detecting signatures of selection through haplotype
4014 differentiation among hierarchically structured populations. Genetics 193: 929-41.

4015 Feng W, Zhou L, Zhao P, et al. 2021. Genome Assemblies of the Warthog and Kenyan Domestic Pig
4016 Provide Insights into Suidae Evolution and Candidate Genes for African Swine Fever Tolerance.
4017 BioRxiv. doi: https://doi.org/10.1101/2021.12.17.473133.

4018 Ferreira S, Gaylard A, Greaver C, et al. 2013. Summary report: Animal abundances in Parks 2012/2013.
4019 SANPARKS RESEARCH REPORT 2013. Skukuza: SANParks.

4020 Fischer D, Laiho A, Gyenesei A, et al. 2015. Identification of reproduction-related gene polymorphisms
4021 using whole transcriptome sequencing in the Large White pig population. G3 (Bethesda) 5: 1351-
4022 1360.

4023 Fonseca C, da Silva AA, Alves J. et al. 2011. Reproductive performance of Wild Boar females in
4024 Portugal European Journal of Wildlife Research 57: 363-371.

4025 Fontanesi L. Schiavo G, Galimberti G, et al. 2014. A genomewide association study for average daily
4026 gain in Italian Large White pigs. Journal of Animal Science 92: 1385-1394.

4027 Frankenberg SR, De Barros FRO, Rossant J, et al. 2016. The mammalian blastocyst. Journal of
4028 Developmental Biology 5: 210-232.

4029 Franklin IR. 1980. Evolutionary change in small populations. In M Soulé and B Wilcox (Eds). In
4030 Conservation Biology: An Evolutionary-Ecological Perspective. Sunderland: Massachusettsr. pp.
4031 135-149.

4032 Frankham R. 2003. Genetics and conservation biology. Comptes Rendus Biologies 326: 22-29.

4033 Frankham R. 2012. How closely does genetic diversity in finite populations conform to predictions of
4034 neutral theory? Large deficits in regions of low recombination. Heredity 108: 167-78.

4035 Frankham R, Ballou JD, Briscoe DA. 1995. Conservation genetics. Annual Review of Genetics 29: 311-
4036 327.

4037 Frankham R, Ralls K. 1998. Conservation biology: Inbreeding leads to extinction. Nature 392: 441-442.

4038 Frantz LA, Schraiber JG, Madsen O, et al. 2013. Genome sequencing reveals fine scale diversification
4039 and reticulation history during speciation in Sus. Genome Biology 14: R107.

4040 Frantz L, Meijaard E, Gongora J, et al. 2016. The Evolution of Suidae. Annual Review of Animal
4041 Biosciences 4: 61-85.

4042 Fu W, Lee WR, Abasht B. 2016. Detection of genomic signatures of recent selection in commercial
4043 broiler chickens. BMC Genetics 17: 122.

4044 Fulgione D, Rippa D, Buglione M, et al. 2016. Unexpected but welcome. Artificially selected traits may
4045 increase fitness in Wild Boar. Evolutionary Applications 17: 769-76.

4046 Gabriel C, Blome S, Malogolovkin AS, et al. 2011. Characterization of African Swine Fever virus
4047 Caucasus isolate in European Wild Boars. Emerging Infectious Diseases 17: 2342-2345.

4048 Gabriel SB. 2002. The structure of haplotype blocks in the human Genome Science 296: 2225-2229.

4049 Gaddis KP, Null DJ, Cole JB. 2016. Explorations in genome-wide association studies and network
4050 analyses with dairy cattle fertility traits. Journal of Dairy Science 99: 6420-6435.

4051 Gama LT, Martínez AM, Carolino I, et al. 2013. Genetic structure, relationships and admixture with
4052 wild relatives in native pig breeds from Iberia and its island. Genetics Selection Evolution 45: 18.

4053 Gandini GC, Ollivier L, Danell B, et al. 2004. Criteria to assess the degree of endangerment of livestock
4054 breeds in Europe. Livestock Production Science 91: 173-182.

4055 Gao F, Chen C, Li B, et al. 2020. The gene flow direction of geographically distinct phytophthora
4056 infestans populations in china corresponds with the route of seed potato exchange. Fronters in
4057 Microbiology 11: 1077.

4058 Gautier M, Vitalis R. 2012. Rehh: An R package to detect footprints of selection in genome-wide SNP
4059 data from haplotype structure. Bioinformatics 28: 1176-1177.

4060 Gers S, Vosloo W, Drew T, et al. 2011. Experimental infection of common warthogs (Phacochoerus
4061 africanus) and bushpigs (Potamochoerus larvatus) with classical swine fever virus II: A comparative
4062 histopathological study. Transboundary and Emerging Diseases 58: 135-144.

4063 Gcumisa ST, Oguttu JW, Masafu MM. 2016. Pig farming in rural South Africa: A case study of
4064 uThukela District in KwaZulu-Natal. Indian Journal of Animal Research 50: 614-620.

4065 Gianola D, Simianer H, Qanbari S. 2010. A two-step method for detecting selection signatures using
4066 genetic markers. Genetics Research 92: 141-55.

4067 Giuffra E, Kijas JM, Amarger V, et al. 2000. The origin of the domestic pig: independent domestication
4068 and subsequent introgression. Genetics 154: 1785-1791.

4069 Glodek P, Oldigs B. (Eds). 1981. Das Göttinger Miniaturschwein. Parey: Berlin and Hamburg. pp. 9-
4070 17.

4071 Glover KA, Solberg MF, McGinnity P, et al. 2017. Half a century of genetic interaction between farmed
4072 and wild Atlantic salmon: Status of knowledge and unanswered questions. Fish and Fisheries 18:
4073 890-927.

4074 Gmelin JF. 1788. Systema naturae per regna tria naturae secundum classes, ordines, genera, species cum
4075 caracteribus, differentiis, synonmis. Locis 1: 3021-3909. Leipzig: CRC Press.

4076 Goddard ME, Hayes B, McPartlan H, et al. 2006. Can the same genetic markers be used in multiple
4077 breeds? 8th World Congress on Genetics Applied to Livestock Production 8: 22-14.

4078 Goedbloed DJ, Megens HJ, Van Hooft P, et al. 2013. Genome-wide single nucleotide polymorphism
4079 analysis reveals recent genetic introgression from domestic pigs into Northwest European Wild Boar
4080 populations. Molecular Ecology 22: 856-866.

4081 Gondret F, Vincent A, Houée-Bigot M, et al. 2016. Molecular alterations induced by a high-fat high-
4082 fiber diet in porcine adipose tissues: variations according to the anatomical fat location BMC
4083 Genomics 17: 120.

4084 Gondret F, Vincent A, Houée-Bigot M, et al. 2017. A transcriptome multi-tissue analysis identifies
4085 biological pathways and genes associated with variations in feed efficiency of growing pigs. BMC
4086 Genomics 18: 244.

4087 Gongora J, Cuddahee R, do Nascimento F, et al. 2011. Rethinking the evolution of extant sub-Saharan
4088 African suids (Suidae, Artiodactyla). Zoologica Scripta 40: 327-335.

4089 Gouveia JJD, Silva MVG, Paiva SR, et al, 2014. Identification of selection signatures in livestock
4090 species. Genetics and Molecular Biology 37: 330-342.

4091 Gray CA, Abbey CA, Beremand PD, et al. 2006. Identification of endometrial genes regulated by early
4092 pregnancy, progesterone, and interferon tau in the ovine uterus. Biology of Reproduction 74: 383-
4093 394.

4094 Groenen MA, Archibald AL, Uenishi H, et al. 2012. Analyses of pig genomes provide insight into
4095 porcine demography and evolution. Nature 491: 393-398.

4096 Groeneveld E, Westhuizen BV, Maiwashe A, et al. 2009. POPREP: A generic report for population
4097 management. Genetics and Molecular Research 8: 1158-78.

4098 Groeneveld LF, Lenstra JA, Eding H. 2010. Genetic diversity in farm animals – a review. Animal
4099 Genetics 41: 6-31.

4100 Grossi DA, Jafarikia M, Brito LF, et al. 2017. Genetic diversity, extent of linkage disequilibrium and
4101 persistence of gametic phase in Canadian pigs. BMC Genetics 18: 6-19.

4102 Grossi DA, Abo-Ismail MK, Koeck A, et al. 2014. Genome-wide association analyses for mastitis in
4103 Canadians Holsteins. In: Proceedings of the 10th World Congress of Genetics Applied to Livestock
4104 Production: 524. Vancouver, Canada.

4105 Gu T, Zhu M, Schroyen M, et al. 2014. Endometrial gene expression profiling in pregnant Meishan and
4106 Yorkshire pigs on day 12 of gestation. BMC Genomics 15: 156.

4107 Guo T, Zhao H, Yuan C, et al. 2021. Selective sweeps uncovering the genetic basis of horn and
4108 adaptability traits on fine-wool sheep in China. Frontiers in Genetics 12: 604235.

4109 Guo Y, Mao H, Ren J, et al. 2009. A linkage map of the porcine genome from a large-scale White Duroc
4110 x Erhualian resource population and evaluation of factors affecting recombination rates. Animal
4111 Genetics 40: 47-52.

4112 Gurgul A, Jasielczuk I, Ropka-Molik K, et al. 2018. A genome-wide detection of selection signatures
4113 in conserved and commercial pig breeds maintained in Poland. BMC Genetics 19: 95.

4114 Gurgul A, Semik E, Pawlina K, et al. 2014. The application of genome-wide SNP genotyping methods
4115 in studies on livestock genomes. Journal of Applied Genetics 55: 197-20.

4116 Gusev A, Lowe JK, Stoffel M, et al. 2008. Whole population, genome-wide mapping of hidden
4117 relatedness. Genome Research 19: 318-26.

4118 Gutiérrez-Gil B, Wiener P, Williams JL. 2007. Genetic effects on coat colour in cattle: dilution of
4119 eumelanin and phaeomelanin pigments in an F2 backcross Charolais x Holstein population. BMC
4120 Genomics 8: 56.

4121 Guy SZ, Thomson PC, Hermesch S. 2012. Selection of pigs for improved coping with health and
4122 environmental challenges: breeding for resistance or tolerance? Frontiers in Genetics 2012 3: 281.

4123 Halimani TE, Muchadeyi FC, Chimonyo M, et al. 2010. Pig genetic resource conservation: The
4124 Southern African perspective. Ecology Economy 69: 944-951.

Halimani TE, Muchadeyi FC, Chimonyo M, et al. 2012. Some insights into the phenotypic and genetic diversity of indigenous pigs in Southern Africa. South African Journal of Animal Science 42: 507-510.

Halimani, TE, Mapiye, O, Marandure, T, et al. 2020. Domestic Free-Range Pig Genetic Resources in Southern Africa: Progress and Prospects. Diversity, 12, 68.

Hamazaki K, Iwata H. 2020. RAINBOW: Haplotype-based genome-wide association study using a novel SNP-set method. PLoS Computational Biology 16: e1007663.

Hamill RM, McBryan J, McGee C, et al. 2012. Functional analysis of muscle gene expression profiles associated with tenderness and intramuscular fat content in pork. Meat Science 92: 440-450.

Han E, Carbonetto P, Curtis RE, et al. 2017. Clustering of 770,000 genomes reveals post-colonial population structure of North America. Nature Communications 8: 14238.

Han L, Abney M. 2013. Using identity by descent estimation with dense genotype data to detect positive selection. European Journal of Human Genetics 21: 205-211.

Happer C, Wellesley L. 2019. Meat consumption, behaviour and the media environment: A focus group analysis across four countries. Food Security 11: 123-139.

Harris J. 2013a. Subfamily Suinae-pigs, hogs. In: Kingdon J, Hoffman M. Mammals of Africa. London: Bloosmbury.

Harrison RG. 1990. Hybrid zones: windows on evolutionary process. Futuyma D, Antonovics J. (Eds). Oxford surveys in evolutionary biology. Vol. 7. New York: Oxford University Press. p. 69-128.

Harrison RG. (Ed). 1993. Hybrids and hybrid zones: historical perspective. Hybrid zones and the evolutionary process. New York: Oxford University Press. pp.3-12.

Harrison RG, Larson EL. 2014. Hybridization, introgression, and the nature of species boundaries. Journal of Heredity 105: 795-809.

Hatzirodos N, Hummitzsch K, Irving-Rodgers HF, et al. 2014a. Transcriptome profiling of granulosa cells from bovine ovarian follicles during atresia. BMC Genomics 15: 40.

Hatzirodos N, Irving-Rodgers HF, Hummitsch K, et al. 2014b. Transcriptome profiling of granulosa cells of bovine ovarian follicles during growth from small to large antral sizes. BMC Genomics 15: 24.

Hayes BJ, Daetwyler HD. 2019. Map simple and complex traits in cattle: Applications and outcomes. Annual Review of Animal Biosciences 7: 89-102.

Hayes BJ, Visscher PM, McPartlan HC, et al. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Research 13: 635-643.

Heifetz EM, Fulton JE, Sullivan NO, et al. 2005. Extent and consistency across generations of linkage disequilibrium in commercial layer chicken breeding populations. Genetics 171: 1173-1181.

Hermesch S, Li L, Doeschl AB. 2015. Selection for productivity and robustness traits in pigs. Animal Production Science 55: 1437-1447.

4161 Herrero-Medrano HJ, Megens RP, Crooijmans JM, et al. 2012. Farm-by-farm analysis of microsatellite,
4162 mtDNA and SNP genotype data reveals inbreeding and crossbreeding as threats to the survival of a
4163 native Spanish pig breed. Animal Genetics 44: 259-266.

4164 Herrero-Medrano JM, Megens H, Groenen MAM, et al. 2013. Conservation genomic analysis of
4165 domestic and wild pig populations from the Iberian Peninsula. BMC Genetics 14: 1-13.

4166 Herrero-Medrano JM, Megens HJ, Groenen M, et al. 2014. Whole-genome sequence analysis reveals
4167 differences in population management and selection of European low-input pig breeds. BMC
4168 Genomics 15: 601.

4169 Hill W. 2004. Influences of genetic variance in phenotypic variability on response to artificial selection.
4170 Proceedings of the British Society of Animal Science 2004: 30-30.

4171 Hoffman LC, Sales J. 2007. Physical and chemical quality characteristics of Warthog (Phacochoerus
4172 africanus) meat. Livestock Research for Rural Development 19: 153.

4173 Hoffman LC, Styger E, Muller M, et al. 2005. Sensory, physical and chemical quality characteristics of
4174 bacon derived from South African indigenous and commercial pig breeds. South African Society for
4175 Animal Science 6: 38-48.

4176 Hoffman LC, Styger WF, Brand TS, et al. 2005. The growth, carcass yield, physical and chemical
4177 characteristic of two South African indigenous pig breeds. South African Journal of Animal Science
4178 6: 25-35.

4179 Hoffman LC, Swanepoel M, Leslie AJ. 2017. African game meat and the safety pertaining to free-
4180 ranging wildlife: example of a wild suid in South Africa. Game meat hygiene: Food safety and
4181 security. Paulsen P, Bauer A, Smulders FJM. (Eds). Wageningen, Netherlands: Wageningen
4182 Academic Publishers. pp. 17-50.

4183 Holness DH. 1973. The role of indigenous pigs as source of protein in Africa: A review. Rhodesian
4184 Journal of Agricultural Research 73: 59-63.

4185 Holness DH. 1991. The Tropical Agriculturalist. London, UK:. Macmillan Education Ltd.

4186 Hossain ME, Chakma S, Khatun MM, et al. 2011. Production systems of swine in the rural areas of
4187 Rangamati and Khagrachari districts of Bangladesh. Bangladesh Journal of Animal Science 40: 28-
4188 33.

4189 Huang DW, Sherman BT, Zheng X, et al. 2009. Extracting biological meaning from large gene lists
4190 with DAVID. Current Protocols in Bioinformatics. 27: 13-11.

4191 Hulseggea I, Schoona M, Windig J, et al. 2019. Development of a genetic tool for determining breed
4192 purity of cattle. Livestock Science 223: 60-67.

4193 Hunt SC, Hasstedt SJ, Xin Y, et al. 2011) Polymorphisms in the NPY2R Gene Show Significant
4194 Associations with BMI that are Additive to FTO, MC4R, and NPFFR2 Gene Effects. Obesity 19:
4195 2241-2247.

4196 Iacolina L, Pertoldi C, Amills M, et al. 2018. Hotspots of recent hybridization between pigs and wild
4197 boars in Europe. Scientific Reports 8: 17372.

4198 Ibeagha-Awemu EM, Peters SO, Akwanji KA, et al. 2016. High density genome-wide genotyping-by-
4199 sequencing and association identifies common and low frequency SNPs, and novel candidate genes
4200 influencing cow milk traits. Scientific Reports 6: 31109.

4201 Jasielczuk I, Gurgul A, Szmatoła T, et al. 2020. Comparison of linkage disequilibrium, effective
4202 population size and haplotype blocks in Polish Landrace and Polish native pig populations. Livestock
4203 Science 231: 103887.

4204 Jiao H, Zhao Y, Zhou Z, et al. 2021. Identifying circular RNAs in HepG2 expressing genotype IV swine
4205 hepatitis E virus ORF3 via whole genome sequencing. Cell Transplantation 30: 1-13.

4206 Johnson M. Whalen A, Ros-Freixedes R, et al. 2021. Genetic variation in recombination rate in the pig.
4207 Genetics Selection Evolution 53: 54.

4208 Jones GF. 1998. Genetic aspects of domestication, common breeds and their origin. MF Rothschild,
4209 Ruvinsky A. (Eds). The Genetics of the Pig. Wallingford: CABI Publishing.

4210 Joost S, Colli L, Bonin A, et al., The GLOBALDIV Consortium. 2011. Promoting collaboration between
4211 livestock and wildlife conservation genetics communities. Conservation Genetic Resources 3: 785-
4212 788.

4213 Jori F, Bastos DS. 2009. Role of wild suids in the epidemiology of African swine fever. Ecohealth 6:
4214 296-310.

4215 Jori F, Payne A, Stahl K, et al. 2016. Ecology of disease transmission at the interface between wild and
4216 domestic pig species in the old and the new world. Melletti M, Meijaard E, (Eds). Ecology, Evolution
4217 and Management of Wild Pigs and Peccaries. Implications for Conservation. Cambridge University
4218 Press.

4219 Jung JH, Shim KS, Na CS, et al. 2015. Studies on intramuscular fat percentage in live swine using real-
4220 time ultrasound to determine pork quality. Asian-Australasian Journal of Animal Sciences 28: 318-
4221 322.

4222 Júnior GF, Costa RB, De Camargo GM, et al. 2016. Genome scan for post-mortem carcass traits in
4223 Nellore cattle. Journal of Animal Science 94: 4087-4095.

4224 Kagira JM, Kanyari PWN, Maingi N, et al. 2010. Characteristics of the smallholder free-range pig
4225 production system in Western Kenya. Tropical Animal Health 42: 865-873.

4226 Kamuribo ED, Chenyambuga SW, Maken, VW, et al. 2011. Characteristics and production constraints
4227 of rural-based small-scale pig farming in Iringa region, Tanzania. Livestock Research for Rural
4228 Development 23: 172.

4229 Kanengoni AT, Chimonyo M, Erlwanger KH, et al. 2014. Growth performance, blood metabolic
4230 responses, and carcass characteristics of grower and finisher South African Windsnyer-type
4231 indigenous and Large White × Landrace crossbred pigs fed diets containing ensiled corncobs, Journal
4232 of Animal Science 92: 5739-5748.

4233 Kang K, Seo DW, Lee JB, et al. 2013. Identification of SNPs affecting porcine carcass weight with the
4234 60K SNP chip. Journal of Animal Science and Technology 55: 231-235.

Kardos M, Qvarnström A, Ellegren H. 2017. Inferring individual inbreeding and demographic history from segments of identity by descent in ficedula flycatcher genome sequences. Genetics 205: 1319-1334.

Karlsson EK, Baranowska I, Wade CM, et al. 2007. Efficient mapping of Mendelian traits in dogs through genome-wide association. Nature Genetics 39: 1321-1328.

Kem EH. 1993. Pig Production in South Africa. Irene: Agricultural Research Council.

Kemper KE, Reich CM, Bowman PJ, et al. 2015. Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. Genetics Selection Evolution 47: 29.

KeyghobadiN. 2007. The genetic implications of habitat fragmentation for animals. Canadian Journal of Zoology. 85: 1049-1064.

Khalilzadeh P, Rezaei HR, Fadakar D, et al. 2016. Contact zone of Asian and European Wild Boar at North West of Iran. PLoS ONE 11: e0159499.

Khanyile KS, Dzomba EF, Muchadeyi FC. 2015. Population genetic structure, linkage disequilibrium and effective population size of conserved and extensively raised chicken populations of Southern Africa. Frontiers in Genetics 6: 13.

Kim KI, Lee JH, Li K, et al. 2002. Phylogenetic relationships of Asian and European pig breeds determined by mitochondrial DNA d-loop sequence polymorphism. Animal Genetics 33: 19-25.

Kimura A, Namekata K, Guo X, et al. 2016. Neuroprotection, growth factors and BDNF-TrkB signalling in retinal degeneration. International Journal of Molecular Sciences 17: 1584.

Kirkpatrick FD. 2017. Crossbreeding in beef cattle. Programs in agriculture and natural resources. University of Tennessee Institute of Agriculture, Knoxville, TN.

Kirsten J, Blignaut C, Visser D. 2009. Pork chains in South Africa. Trienekens J, Petersen B, Wognum N. (Eds.). European pork chains: diversity and quality challengein consumerorientated production and distribution The Netherlands: Wageningen Academic Publishers.

Kirsten JF, van Zyl J. 1998. Defining small-scale farmers in the South African context. Agrekon 37: 551-562.

Kogelman LJ, Kadarmideen HN, Mark T, et al. 2013. An f2 pig resource population as a model for genetic studies of obesity and obesity-related diseases in humans: design and genetic parameters. Frontiers in Genetics 4: 29.

Krecek RC, Michael LM, Willingham III AL, et al. 2004. Questionnaire results from a community-based project on porcine cysticercosis in the Eastern Cape province of South Africa. Southeast Asian Journal of Tropical Medicine 35: 271-274.

Kriel, G. 2020. The future of pig genetics. Farmer's weekly, 01 September 2020. Available from: https://www.farmersweekly.co.za/animals/pigs/the-future-of-pig-genetics/. [Accessed: 14 December 2022].

Krige JE. 1988. (2nd Ed). The Social Systems of the Zulus. Pietermaritzburg: Shuter and Shooter.

4272 Krombeen SK. 2017. Neonatal growth measurements, umbilical concentrations, and differential gene
4273 expression in the placenta of swine in relation to placental efficiency. MSc. Clemson University,
4274 USA.

4275 Krüger DA, van Marle-Koster E, Theron HE. 2017. Comparison of on-farm progeny performances from
4276 local and imported boar semen used in the South African pig industry. South African Journal of
4277 Animal Science 475: 688-696.

4278 Krupa E, Wolfová M, Krupová Z, et al. 2020. Estimation of economic weights for number of teats and
4279 sperm quality traits in pigs. Journal of Animal Breeding and Genetics 137: 189-199.

4280 Kumar S, Deb R, Singh U, et al. 2015. Bovine circadian locomotor output cycles kaput (CLOCK) and
4281 clusterin (CLU) mRNA quantitation in ejaculated crossbred bull spermatozoa. Reproduction in
4282 Domestic Animals 50: 505-509.

4283 Kumaresan A, Bujarbaruah KM, Pathak KAet al. 2007. Performance of pigs reared under tradition tribal
4284 low input production system and chemical composition of non-conventional tropical plants used as
4285 pig feed. Livestock Science 107: 294-298.

4286 Kutwana HW, Gxasheka M, Tyasi TL. 2015. Body weight and morphological traits of Large White and
4287 Kolbroek pig breeds. International Journal of Advanced Research 3: 105-109.

4288 Lacy RC. 1995. Classification of genetic terms and their use in the management of captive populations.
4289 Zoo Biology 14: 565-577.

4290 Lai CQ. 2012. Adaptive genetic variation and population differences. Progress in Molecular Biology
4291 and Translational Science 108: 461-489.

4292 Lamsyah H, Rueda B, Baassi L, et al. 2009. Association of PTPN22 gene functional variants with
4293 development of pulmonary tuberculosis in Moroccan population. Tissue Antigens 74: 228-232.

4294 Lander ES, Schork NJ. 1994. Genetic dissection of complex traits. Science 265: 2037-2048.

4295 Larson G, Albarella U, Dobney K, et al. 2007. Ancient DNA, pig domestication, and the spread of the
4296 Neolithic into Europe. Proceedings of the National Academy of Sciences 104: 15276-15281.

4297 Larson G, Dobney K, Albarella U, et al. 2005. Worldwide phylogeography of Wild Boar reveals
4298 multiple centers of pig domestication. Science 307: 1618-1621.

4299 Larson G, Liu R, Zhao X, et al. 2010. Patterns of East Asian pig domestication, migration, and turnover
4300 revealed by modern and ancient DNA. Proceedings of the National Academy of Sciences of the
4301 United States of America 107: 7686-7691.

4302 Lashmar SF, Visser C. Marle-Köster E. 2016. SNP-based genetic diversity of South African commercial
4303 dairy and fiber goat breeds. Small Ruminant Research 136: 65-71.

4304 Laval G, Iannuccelli N, Legault C, et al. 2000. Genetic diversity of eleven European pig breeds. Genetic
4305 Selection Evolution 32: 187-203.

4306 Le TH, Christensen OF, Nielsen B, et al. 2017. Genome-wide association study for conformation traits
4307 in three Danish pig breeds. Genetic Selection Evolution 49: 12.

4308 Lee SH. 2020. Genetic diversity and ancestral study for Korean native pigs using 60K SNP chip.
4309 Animals 10: 760.

4310 Lee C, Day J, Goodman SM, et al. 2020. Genetic origins and diversity of bushpigs from Madagascar
4311 (Potamochoerus larvatus, family Suidae). Scientific Reports 10: 20629.

4312 Lee T, Shin DH, Cho S, et al. 2014. Genome-wide Association study of integrated meat quality-related
4313 traits of the Duroc pig breed. Asian-Australasian Journal of Animal Sciences. 27: 303-309.

4314 Lee W, Taye M, Kwon T, et al. 2017. Identifying candidate positive selection genes in Korean imported
4315 pig breeds. Genes Genomics 39: 557-565.

4316 Lekule FP, Kyvsgaard NC. 2003. Improving pig husbandry in tropical resource poor communities and
4317 its potential to reduce risk of porcine cysticercosis. Acta Tropica 87: 111- 117.

4318 Letunic I, Bork P. 2021. Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display
4319 and annotation. Nucleic Acids Research 49: W293-W296.

4320 Levathes L. 1994. When China ruled the seas: The treasure fleet of the Dragon Throne. New York:
4321 Sixteen Century Journal. pp. 1405-1433.

4322 Levy F, Antonovics J, Boynto JE. 1996. A population genetic analysis of chloroplast DNA in Phacella.
4323 Heredity 76: 143-155.

4324 Lewontin RC, Kojima K. 1960. The evolutionary dynamics of complex polymorphisms. Evolution 14:
4325 458-472.

4326 Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of the theory of the selective
4327 neutrality of polymorphisms. Genetics 74 175-195.

4328 Li K, Ren J, Xing Y, et al. 2009b. Quantitative trait loci for litter size and prenatal loss in a White Duroc
4329 x Chinese Erhualian resource population. Animal Genetics 40: 963-966.

4330 Li M, Tian S, Yeung C. et al. 2014. Whole-genome sequencing of Berkshire (European native pig)
4331 provides insights into its origin and domestication. Scientific Reports 4: 4678.

4332 Li X, Kim SW, Choi JS, et al. 2010. Investigation of porcine FABP3 and LEPR gene polymorphisms
4333 and mRNA expression for variation in intramuscular fat content. Molecular Biology Reports 37:
4334 3931- 3939.

4335 Li X, Kim SW, Do KT, et al. 2011. Analyses of porcine public SNPs in coding-gene regions by re-
4336 sequencing and phenotypic association studies. Molecular Biology Reports 38: 3805-3820.

4337 Li Y, Kim JJ. 2015. Multiple linkage disequilibrium mapping methods to validate additive quantitative
4338 trait loci in Korean native cattle (Hanwoo). Asian Australasian Journal of Animal Science 28: 926-
4339 935.

4340 Li Y, Mei S, Zhang X, et al. 2012. Identification of genome-wide copy number variations among diverse
4341 pig breeds by array CGH. BMC Genomics 13: 725.

4342 Liang Z, Hu L, Qin Y et al. 2019. Selection of optimal ancestry informative markers for classification
4343 and ancestry proportion estimation in pigs. Frontiers in Genetics 10: 183.

4344 Lim D, Lee S, Kim N, et al. 2013. Gene co- expression analysis to characterize genes related to marbling
4345 trait in Hanwoo (Korean) cattle. Asian Australasian Journal of Animal Science 26: 19-29.

4346 Linnaeus C. 1758. Systema naturae per regna tria naturae, secundum classes, ordines, genera, species,
4347 cum characteribus, differentiis, synonymis, locis. Editio decima, reformata: Holmiae 1-4: 1-824.

4348 Lindblad-Toh K, Wade CM, Mikkelsen TS, et al. 2005. Genome sequence, comparative analysis and
4349 haplotype structure of the domestic dog. Nature 438: 803.

4350 Lindgren D, Gea LD, Jefferson PA. 1996. Loss of genetic diversity monitored by status number. Silvae
4351 Genetics 45: 52-59.

4352 Lindholm-Perry AK, Butler AR, Kern RJ, et al. 2016. Differential gene expression in the duodenum,
4353 jejunum and ileum among crossbred beef steers with divergent gain and feed intake phenotypes.
4354 Animal Genetics 47: 408-427.

4355 Lindsey PA, Balme G, Becker M, et al. 2013. The bushmeat trade in African savannas: Impacts, drivers,
4356 and possible solutions. Biological Conservation 160: 80-96.

4357 Lionikas A, Cheng R, Lim JE, et al. 2010. Fine-mapping of muscle weight QTL in LG/J and SM/J
4358 intercrosses. Physiol. Genomics 42A: 33-38.

4359 Lionikas A, Meharg C, Derry JMJ, et al. 2012. Resolving candidate genes of mouse skeletal muscle
4360 QTL via RNA-Seq and expression network analyses. BMC Genomics 13: 592.

4361 Liu H, Nguyen YT, Nettleton D, et al. 2016. Post-weaning blood transcriptomic differences between
4362 Yorkshire pigs divergently selected for residual feed intake. BMC Genomics 17: 73.

4363 Liu Y. 2015. Fatty acids, inflammation and intestinal health in pigs. Journal of Animal Science and
4364 Biotechnology 6: 41.

4365 Lönnberg E. 1908. Remarks on some wart-hog skulls in the British Museum. Proceedings of the
4366 Zoological Society of London 78: 936-940.

4367 Lönnberg E. 1912. Mammals collected by the Swedish zoological expedition to British East Africa
4368 1911. Kungliga Svenska Vetenskapakademiens Handligar 48: 1-188.

4369 Lopes-Bote CJ. 1998. Sustained utilization of the Iberian pig breed. Meat Science 49: S17-S27.

4370 Lopes MS, Bastiaansen JW, Harlizius B, et al. 2014. A genome-wide association study reveals
4371 dominance effects on number of teats in pigs. PloS ONE 9: e105867.

4372 Lozada-Soto EA, Maltecca C, Wackel H, et al. 2021. Evidence for recombination variability in purebred
4373 swine populations. Journal of Animal Breeding and Genetics 138: 259-73.

4374 Ludwiczak A, Składanowska-Baryza J, Stanisz M. 2020. Effect of age and sex on the quality of offal
4375 and meat of the Wild Boar (Sus scrofa). Animals 10: 660.

4376 Luikart G, Allendorf FW. 1996 Mitochondrial DNA variation and genetic population structure in Rocky
4377 Mountain bighorn sheep (Ovis canadensis canadensis). Journal of Mammalogy 77: 109-123.

4378 Luikart G, England PR, Tallmon D, et al. 2003. The power and promise of population genomics: from
4379 genotyping to genome typing. Nature Reviews Genetics 4: 981-994.

Lunney JK. 2007. Advances in swine biomedical model genomics. International Journal of Biological Sciences. 3: 179-184.

Luo W, Cheng D, Chen S, et al. 2012. Genome-wide association analysis of meat quality traits in a porcine Large White x Minzhu intercross population. International Journal of Biological Sciences 8: 580-595.

Ma XJ, Zhang XL, Wang LX, et al. 2011. Studies on difference of immune and production indexes between Songliao black pig and Large White pig. China Animal Husbandry and Veterinary Medicine 38: 52-55.

Ma Y, Ding X, Qanbari S, et al. 2015. Properties of different selection signature statistics and a new strategy for combining them. Heredity 115: 426-36.

Madec F, Hurnik D, Porphyre V, et al. 2010. Good practices for biosecurity in the pig sector: Issues and options in developing and transition countries. FAO, Rome, Italy.

Madzimure J, Chimonyo M, Zander KK, et al, 2012. Diurnal heat-related physiological and behavioural responses in South African indigenous gilts. Journal of Arid Environments 87: 29-34.

Madzimure J, Chimonyo M, Zander KK, et al. 2012. Potential for using indigenous pigs in subsistence-oriented and market-oriented small-scale farming systems of Africa. Tropical Animal Health and Production 45: 135-142.

Magazines, C. 2012. Getting started with pigs - 4. Farmer's weekly, 25 June 2012. Available from: https://www.farmersweekly.co.za/farming-basics/how-to-livestock/getting-started-with-pigs-4/. [Accessed: 14 December 2022].

Maiorano AM, Cardoso DF, Carvalheiro R, et al. 2022. Signatures of selection in Nelore cattle revealed by whole-genome sequencing data. Genomics 114: 110304.

Makhanya LG. 2018. Phenotypic and reproductive characterisation of Kolbroek pigs. MTech dissertation, Central University of Technology, South Africa.

Makina SO, Muchadeyi FC, van Marle-Koöster E, et al. 2014. Genetic diversity and population structure among six cattle breeds in South Africa using a whole genome SNP panel. Frontiers in Genetics 5: 333.

Makina SO, Muchadeyi FC, van Marle-Köster E, et al. 2015. Genome-wide scan for selection signatures in six cattle breeds in South Africa. Genetics Selection Evolution 47: 92.

Malécot G. 1941. Etude mathématique des populations mendéliennes. Annales de l'Université de Lyon, Sciences A-4: 45-60.

Manchidi MJ. 2009. An evaluation of small-scale piggeries in Limpopo Province, South Africa. MPhil dissertation, Stellenbosch University.

Mapston M. 2007. Feral Hogs in Texas. Texas Wildlife Services. Available at https / /hdl .handle .net /1969 .1 /87218. [Accessed: 20 June 2022].

Marchelli P, Gallo LA. 2001. Genetic diversity and differentiation in a southern beech subjected to introgressive hybridization. Heredity 87: 284-93.

4417 Martínez AM, Periati VL, Domínguez MM, et al. 2012. Genetic characterization of autochthonous pig
4418 breeds from Spain with microsatellite markers. 7th International Symposium on the Mediterranean
4419 pig 101: 31-37.

4420 Martínez MAJ, Gasper DJ, Muciño MCC, et al. 2018. Suidae and Tayassuidae. Pathology of Wildlife
4421 and Zoo Animals: 207-228.

4422 Martínez-Montemayor MM, Hill GM, Raney NE, et al. 2008. Gene expression profiling in hepatic tissue
4423 of newly weaned pigs fed pharmacological zinc and phytase supplemented diets. BMC Genomics 9:
4424 412.

4425 Martínez-Montes AM, Muiños-Bühl A, Fernández A, et al. 2016. Deciphering the regulation of porcine
4426 genes influencing growth, fatness and yield-related traits through genetical genomics. Mammalian
4427 Genome 28: 130-142.

4428 Maselli V, Rippa D, Russo G, et al. 2014. Wild boars' social structure in the Mediterranean habitat.
4429 Italian Journal of Zoology 81: 610–617.

4430 Masenya MB, Mphaphathi ML, Mapeka MH, et al. 2011. Comparative study on semen characteristics
4431 of Kolbroek and Large White boars following computer aided sperm analysis (CASA). African
4432 Journal of Biotechnology 10: 14223-14229.

4433 Mashatise E, Hamudikuwanda H, Dzama K, et al. 2005. Socio-economic roles, traditional management
4434 systems and reproductive patterns of Mukota pigs in semi-arid north-eastern Zimbabwe. Bunda
4435 Journal of Agriculture, Environmental Science and Technology 3: 97-105.

4436 Mason IL, Maule JP. 1960. The indigenous livestock of eastern and Southern Africa. C.A.B. Technical
4437 Communication no 14. University Press, Aberdeen.

4438 Matabane MB, Nethenzheni P, Thomas R, et al. 2015. Status of the smallholder pig farming sector in
4439 Gauteng Province of South Africa. Applied Animal Husbandry and Rural Development 8: 9-25.

4440 Mayor C, Brudno M, Schwartz JR, et al. 2000. VISTA: Visualizing global DNA sequence alignments
4441 of arbitrary length. Bioinformatics 16: 1046-1047.

4442 McGinnity P, Prodöhl P, Ferguson A, et al. 2003 Fitness reduction and potential extinction of wild
4443 populations of Atlantic salmon, Salmo salar, as a result of interactions with escaped farm salmon.
4444 Proceedings of the Royal Society of London B 270: 2443-2450.

4445 McGraw K, List A. 2017. Chapter five-Erythropoietin receptor signalling and lipid rafts. Vitamin and
4446 Hormones 105: 79-100.

4447 McKay SD, Schnabel RD, Murdoch BM, et al. 2007 Whole genome linkage disequilibrium maps in
4448 cattle. BMC Genetics 8: 74.

4449 Mdladla K, Dzomba EF, Huson H, et al. 2016. Population genomic structure and linkage disequilibrium
4450 analysis of South African goat breeds using genome-wide SNP data. Animal Genetics 47: 471-482.

4451 Medzhitov R. 2008. Origin and physiological roles of inflammation. Nature 454: 428-435.

4452 Megens HJ, Crooijmans RP, Sancristobal M, et al. 2008. Biodiversity of pig breeds from China and
4453 Europe estimated from pooled DNA samples: differences in microsatellite variation between two
4454 areas of domestication. Genetics Selection Evolution 40: 103-128.

4455 Meissner HH, Scholtz MM, Palmer AR. 2013. Sustainability of the South African livestock sector
4456 towards 2050 Part 1: Worth and impact of the sector. South African Journal of Animal Science 43:
4457 282-297.

4458 Messad F, Louveau I, Koffi B, et al. 2019. Investigation of muscle transcriptomes using gradient
4459 boosting machine learning identifies molecular predictors of feed efficiency in growing pigs. BMC
4460 Genomics 20: 659.

4461 Meuwissen TH, Hayes BJ, Goddard ME. 2001 Prediction of total genetic value using genome-wide
4462 dense marker maps. Genetics 157: 1819-1829.

4463 Misztal I, Steyn Y, Lourenco DAL. 2022. Genomic evaluation with multibreed and crossbred data. JDS
4464 Communications 3: 156-159.

4465 Mohlatlole RP, Madoroba E, Muchadeyi FC et al. 2013. Virulence profiles of enterotoxigenic, shiga
4466 toxin and enteroaggregative Escherichia coli in South African pigs. Tropical Animal Health and
4467 Production 45: 399-405.

4468 Moioli B, D'Andrea S, De Grossi L, et al. 2016. Genomic scan for identifying candidate genes for
4469 paratuberculosis resistance in sheep. Animal Production Science 56: 1046-1055.

4470 Mokoele JM, Spencer BT, van Leengoed LA et al. 2014. Efficiency indices and indicators of poor
4471 performance among emerging small-scale pig farmers in the Limpopo Province, South Africa.
4472 Onderstepoort Journal of Veterinary Research 12:81.

4473 Mokry FB, Buzanskas ME, de Alvarenga Mudadu M, et al. 2014. Linkage disequilibrium and haplotype
4474 block structure in a composite beef cattle breed. BMC Genomics 15: S6.

4475 Molotsi AH, Cupido M, Hoffman LCH. 2021. Characterization of smallholder pig production systems
4476 in the Western Cape, South Africa. Tropical Animal Health and Production 53: 1-14.

4477 Morin PA, Luikart G, Wayne RK, et al. 2004. SNPs in ecology, evolution and conservation. Trends in
4478 Ecology and Evolution 19: 208-216.

4479 Moran B, Butler S, Moore S, et al. 2017. Differential gene expression in the endometrium reveals
4480 cytoskeletal and immunological genes in lactating dairy cows genetically divergent for fertility traits.
4481 Reproduction, Fertility and Development 29: 274-282.

4482 Moyers BT, Morrell PL, McKay JK. 2018. Genetic costs of domestication and improvement, Journal of
4483 Heredity 109: 103-116.

4484 Moyo D, Gomes M, Erlwanger KH. 2018. Comparison of the histology of the skin of the Windsnyer,
4485 Kolbroek and Large White pigs. Journal of the South African Veterinary Association 89: 1-10.

4486 Mtileni BJ, Muchadeyi FC, Maiwashe A, et al. 2009. Characterisation of production systems for
4487 indigenous chicken genetic resources of South Africa. Applied Animal Husbandry and Rural
4488 Development 2: 18-22.

4489 Mtileni BJ, Muchadeyi FC, Maiwashe A, et al. 2011. Genetic diversity and conservation of South
4490 African indigenous chicken populations. Journal of Animal Breeding and Genetics 128: 209-218.

4491 Mtileni BJ, Visser DP, Voordewind SF. 2006. The development of a computer management program
4492 for small- and medium-scale egg producers. The role of biotechnology in animal agriculture to

address poverty in Africa: Opportunities and challenges. In: Proceedings of the 4TH All Africa Conference on Animal Agriculture and the 31ST Annual Meeting of the Tanzania Society for Animal Production (TSAP).

Mujibi FD, Okoth E, Cheruiyot EK, et al. 2018. Genetic diversity, breed composition and admixture of Kenyan domestic pigs. PLoS ONE 13: e0190080.

Muñoz M, Bozzi R, García-Casco J, et al. 2019. Genomic diversity, linkage disequilibrium and selection signatures in European local pig breeds assessed with a high density SNP chip. Scientific Reports 9: 13546.

Munzhelele P, Oguttu J, Fasanmi OG, Fasina FO. 2017. Production constraints of smallholder pig farms in agro-ecological zones of Mpumalanga, South Africa. Tropical Animal Health and Production 49: 63-69.

Mutua F, Arimi S, Ogara W. 2010. Farmer perception on Indigenous pig farming in Kakamega District, western Kenya. Nordic Journal of African Studies 19: 43-57.

Muwanika VB, Nyakaana S, Siegismund HR, et al. 2003. Phylogeography and population structure of the common Warthog (Phacochoerus africanus) inferred from variation in mitochondrial DNA sequences and microsatellite loci. Journal of Heredity 91: 361-372.

Muwonge A, Johansen TB, Vigdis E, et al. 2012. Mycobacterium bovis infections in slaughter pigs in Mubende district, Uganda: a public health concern. BMC Veterinary Research 8: 168.

Nadeau NJ, Martin SH, Kozak KM et al. 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. Molecular Ecology 22: 814-826.

NAFU - National African Farmers Union. 2007. NAFU farmer headlines. NAFU, June Issue Newsletter, 18-20 November 2017, Pretoria, Republic of South Africa.

Naude RT, Visser DP. 1994. n Generiese kwalitatiewe benadering ten einde die doeltreffende produksie van verbruikersaanneemlike varkvleis te verseker. In: Proceedings of the 6de Nasionale SAVPO-Kongres te Elangeni Hotel, KwaZulu-Natal.

Ncobela CN. 2017. Response in growth performance, blood metabolites, nutrient digestibility, digesta characteristics and carcass characteristics of Windsnyer pigs fed increasing levels of potato hash silage. PhD thesis, University of Kwa-Zulu Natal, South Africa.

Nevrkla P, Václavková E, Rozkot M. 2021. The indigenous prestice black-pied pig breed differs from a commercial hybrid in growth intensity, carcass value and meat quality. Agriculture 11: 331.

Newton JR, De Santis C, Jerry DR. 2012. The gene expression response of the catadromous perciform barramundi Lates calcarifer to an acute heat stress. Journal of Fish Biology 81: 81-93.

Nicholas G. 1999. Kolbroek-the Unique Local Breed. Farmer's Weekly, August.

Nielsen R, Williamson S, Kim Y, et al. 2005. Genomic scans for selective sweeps using SNP data. Genome Research 15: 1566-1575.

Niewiadomska K, Kosicka-Gebska M, Gebski J, et al. 2020. Game Meat Consumption - Conscious Choice or Just a Game? Foods 9: 1357.

4530 Nissen S, Poulsen IH, Nejsum P, et al. 2011. Prevalence of gastrointestinal nematodes in growing pigs
4531 in Kabale District in Uganda. Tropical Animal Health 43: 567-572.

4532 Noce A, Amills M, Manunza A, et al. 2015. East African pigs have a complex Indian, Far Eastern and
4533 Western ancestry. Animal Genetics 46: 433-436.

4534 Notter DR. 1999. The importance of genetic diversity in livestock populations of the future. Journal of
4535 Animal Science 77: 61-9.

4536 Notter DR, Scherf B, Hoffmann I. 2013. Breeding of animals. Levin SA. (2nd Ed). Encyclopedia of
4537 Biodiversity. Elserivier, New York. pp. 636-649.

4538 Novianti I, Pitchford WS, Bottema CD. 2010. Beef cattle muscularity candidate genes. Indonesian
4539 journal of Animal Science Farm 20: 1-10.

4540 Nowacka-Woszuk J. 2021. The genetic background of hernia in pigs: A review. Livestock Science 244:
4541 104317.

4542 Nsengimana J, Baret P, Haley CS, et al. 2004. Linkage disequilibrium in the domesticated pig. Genetics
4543 166: 1395-1404.

4544 Nsoso SJ, Mannathoko GG, Modise K. 2006. Monitoring production, health and marketing of
4545 indigenous Tswana pigs in Ramotswa village of Botswana. Livestock Research for Rural
4546 Development 18: 125.

4547 Nyafu K. 2009. Warthogs as an introduced species in the Eastern Cape. MSc thesis, Nelson Mandela
4548 Metropolitan University, South Africa.

4549 O'Brien AMP, Mészáros G, Utsunomiya YT, et al. 2014. Linkage disequilibrium levels in Bos Indicus
4550 and Bos Taurus cattle using medium and high density SNP chip data and different minor allele
4551 frequency distributions. Livestock Science 166: 121-132.

4552 Ocampo LM, Leterme P, Buldgen A. 2005. A survey of pig production systems in the rain forest of the
4553 pacific coast of Colombia. Tropical Animal Health and Production 37: 315-326.

4554 Olalde I, Capote J, Del-Arco MC, et al. 2015. Ancient DNA sheds light on the ancestry of pre-Hispanic
4555 Canarian pigs. Genetics Selection Evolution 47: 40.

4556 Oleksyk TK, Smith MW, O'brien SJ. 2010. Genome-wide scans for footprints of natural selection.
4557 Philosophical Transactions of the Royal Society London B Biological Sciences 365: 185-205.

4558 Oliver WLR, Brisbin J, Takahami S. 1993. The Eurasian wild pig (Sus scrofa): A status survey and
4559 conservation action plan for pigs, peccaries and hippos. In Oliver LR. International Union for the
4560 Consedation of Nature, Gland: 112-121.

4561 Olivier, L. 2009. European pig genetic diversity: a minireview. Animal 3: 915-924.

4562 Olivieri BF, Mercadante ME, Cyrillo JN, et al. 2016. Genomic regions associated with feed efficiency
4563 indicator traits in an experimental Nellore cattle population. PLoS ONE 11: e0164390.

4564 Oliviero C, Junnikkala S, Peltoniemi O. 2019. The challenge of large litters on the immune system of
4565 the sow and the piglets. Reproduction in Domestic Animals 54: 12-21.

4566 Olstad K, Aasmundstad T, Kongsro J, et al. 2022. Osteochondrosis and other lesions in all intervertebral,
4567 articular process and rib joints from occiput to sacrum in pigs with poor back conformation, and
4568 relationship to juvenile kyphosis. BMC Veterinary Research 18: 44.

4569 Onzima RB, Upadhyay MR, Doekes HP, et al. 2018. Genome-wide characterization of selection
4570 signatures and runs of homozygosity in Ugandan goat breeds. Frontiers in Genetics 9: 318.

4571 Osei-Amponsah R, Skinner BM, Adjei DO, et al. 2017. Origin and phylogenetic status of the local
4572 Ashanti Dwarf pig (ADP) of Ghana based on genetic analysis. BMC Genomics 18: 193-205.

4573 Otyama PI, Wilkey A., Kulkarni R, et al. 2019. Evaluation of linkage disequilibrium, population
4574 structure, and genetic diversity in the US peanut mini core collection. BMC Genomics 20: 1-17.

4575 Ovcharenko I, Nobrega MA, Loots GG, et al. 2004. Browser: a tool for visualizing and accessing data
4576 from comparisons of multiple vertebrate genomes. Nucleic Acids Research 32: W280-W286.

4577 Ovilo C, Benítez R, Fernández A, et al. 2014. Longissimus dorsi transcriptome analysis of purebred and
4578 crossbred Iberian pigs differing in muscle characteristics. BMC Genomics 15: 413.

4579 Oyelami FO, Zhao Q, Xu Z, et al. 2020. Haplotype block analysis reveals candidate genes and QTLs
4580 for meat quality and disease resistance in Chinese Jiangquhai pig breed. Frontiers in Genetics 11:
4581 752.

4582 Palamara PF, Lencz T, Darvasi A, et al. 2012. Length distributions of identity by descent reveal fine-
4583 scale demographic history. American Journal of Human Genetics 91: 809-822.

4584 Pallas PS. 1766. (Ed) Miscellanea zoological quibus novae imprimis atque obscurae animalium species
4585 describuntur et observationibus iconibusque illustrantur. The Hague: Nabu Press. pp. 131-138.

4586 Parker Gaddis KL, Megonigal JH Jr, Clay JS, et al. 2018. Genome-wide association study for ketosis in
4587 US jerseys using producer recorded data. Journal of Dairy Science 101: 413-424.

4588 Patterson N, Richter DJ, Gnerre S, et al. 2006. Genetic evidence for complex speciation of humans and
4589 chimpanzees. Nature 441: 1103-1108.

4590 Patterson N, Moorjani P, Luo Y, et al. 2012. Ancient admixture in human history. Genetics 192: 1065-
4591 1093.

4592 Pearce F. 2005. The Protein Gap. Conservation in Practice 6, 117-123. https://doi.org/10.1111/J.1526-
4593 4629.2005.TB00178.X

4594 Penrith, ML, Bastos ADS, Etter EMC, et al. 2019. Epidemiology of African swine fever in Africa today:
4595 Sylvatic cycle versus socio-economic imperatives. Transboundary and Emerging Diseases 66: 672-
4596 686.

4597 Penrith ML, Vosloo W, Jori F, et al. 2013. African swine fever virus eradication in Africa. Virus
4598 Research 173: 228-246.

4599 Penrith ML, Vosloo W, Mather C. 2011. Classical swine fever (Hog cholera): Review of aspects relevant
4600 to control. Transboundary and Emerging Diseases 58: 187-196.

4601 Pham LD, Do DN, Nam LQ, et al. 2014. Molecular genetic diversity and genetic structure of Vietnamese
4602 indigenous pig populations. Journal of Animal Breeding and Genetics 131: 379-86.

4603 Phillips MS, Lawrence R, Sachidanandam R, et al. 2003. Chromosome-wide distribution of haplotype
4604 blocks and the role of recombination hot spots. Nature Genetics. 33: 382-387.

4605 Phogole SL. 2017. Characterization and conservation of local pig genetic resources in Sekhukhune
4606 district of Limpopo province. MSc dissertation, University of South Africa, South Africa.

4607 Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele
4608 frequency data. PLoS Genetics 8: e1002967.

4609 Pienaar L, Traub LN. 2015. Understanding the smallholder farmer in South Africa: towards a sustainable
4610 livelihoods classification. International Association of Agricultural Economists, 2015 Conference,
4611 Milan, Italy 212633.

4612 Pig Breeders Society of South Africa Members. 2022. Pig Stud Breeders Society. Available at:
4613 http://pigsa.co.za/p22/about-us/pig-breeders%E2%80%99-society-of-south-africa-members.html
4614 [Accessed: 15 July 2021].

4615 Pig Progress. 2021. Unrest in South Africa: Pig sector suffer losses. Natalie Berkhout. July 2021.
4616 Available from: https://www.pigprogress.net/pigs/unrest-in-south-africa-pig-sector-suffers-losses/
4617 [Accessed: 14 December 2022].

4618 Piórkowska K, ˙Zukowski K, Ropka-Molik K, et al. 2018. A comprehensive transcriptome analysis of
4619 skeletal muscles in two Polish pig breeds differing in fat and meat quality traits. Genetics and
4620 Molecular Biology 41: 125-136.

4621 Plug S, Badenhorst I. 2001. (Eds). The Distribution of Macromammals in Southern Africa over the Past
4622 30,000 Years as Reflected in Animal Remains from Archaeological Sites. Transvaal Museum
4623 Monographs 12. Isteg Scientific Publication. pp. 1-234.

4624 Ponsuksili S, Murani E, Brand B, et al. 2011. Integrating expression profiling and whole genome
4625 association for dissection of fat traits in a porcine model. Journal of Lipid Research 52: 668-678.

4626 Pool JE, Nielsen R. 2009. Inference of historical changes in migration rate from the lengths of migrant
4627 tracts. Genetics 181: 711-719.

4628 Porter V. (Ed). 1993. Pigs: A handbook to the breeds of the world. Mountfield, HK: Helm Information
4629 Ltd. pp. 1-256.

4630 Price EO. (Ed). 2002. Animal domestication and behavior. CABI Publishing. British Library, London,
4631 UK.

4632 Pritchard JK, Przeworski M. 2001. Linkage disequilibrium in humans: models and data. American
4633 Journal of Human Genetics 69: 1-14.

4634 Provan J, Maggs CA. Unique genetic variation at a species' rear edge is under threat from global climate
4635 change Proceedings of the Royal Society B: Biological Sciences 279: 39-47.

4636 Prufer K, Racimo F, Patterson N, et al. 2014. The complete genome sequence of a Neanderthal from the
4637 Altai Mountains. Nature 505: 43-49.

4638 Pugach IR, Matveyev A, Wollstein A, et al. 2011. Dating the age of admixture via wavelet transform
4639 analysis of genome-wide data. Genome Biology 12: R19.

4640 Puig-Oliveras A, Revilla M, Castelló A, et al. 2016. Expression-based GWAS identifies variants, gene
4641 interactions and key regulators affecting intramuscular fatty acid content and composition in porcine
4642 meat. Scientific Reports 6: 31803.

4643 Purcell S, Neale B, Todd-Brown K, et al. 2007. PLINK: a tool set for whole-genome association and
4644 population based linkage analyses. American Journal of Human Genetics 81: 559-575.

4645 Qanbari S, Simianer H. 2014. Mapping signatures of positive selection in the genome of livestock.
4646 Livestock Science 166: 133-143.

4647 Qin M, Li C, Li Z, et al. 2020. Genetic diversities and differentially selected regions between Shandong
4648 indigenous pig breeds and western pig breeds. Frontiers in Genetics 10: 1351.

4649 Qwabe SO, van Marle-Köster E, Visser C. 2013. Genetic diversity and population structure of the
4650 endangered Namaqua Afrikaner sheep. Tropical Animal Health and Production 45: 511-516.

4651 Ramírez O, Ojeda A, Tomas A, et al. 2009. Integrating Y-chromosome, mitochondrial, and autosomal
4652 data to analyse the origin of pig breeds. Molecular Biology and Evolution 26: 2061-2072.

4653 Ramos AM, Crooijmans RP, Affara NA, et al. 2009. Design of a high density SNP genotyping assay in
4654 the pig using SNPs identified and characterized by next generation sequencing technology. PLoS
4655 One 4: e6524.

4656 Ramsay K, Smuts M, Els HC. 2000. Adding value to South African landrace breeds conservation
4657 through utilisation. Animal Genetic Resources Information 27: 9-15.

4658 Ramsay KA, Reed DS, Bothma AJ, et al. 1994. Profitable and environmentally effective farming with
4659 early-domesticated livestock in Southern Africa. Pretoria: Department of Agriculture.

4660 Ramsay K, Harris L, Kotze A. 1998. (Eds). Landrace Breeds: South Africa's indigenous and locally
4661 developed farm animals. Farm Animal Conservation Trust, Pretoria.

4662 Ramsay K, Smuts M, Els HC. 2000. Adding value to South African landrace breeds conservation
4663 through utilisation. Animal Genetic Resources Information 27: 1-7.

4664 Ramstetter MD, Shenoy SA, Dyer TD, et al 2018. Inferring identical-by-descent sharing of sample
4665 ancestors promotes high-resolution relative detection. American Journal of Human Genetics 103: 30-
4666 44.

4667 Raschetti M, Castiglioni B, Caroli A, et al. 2013. SNP identification in swine candidate genes for meat
4668 quality. Livestock Science 155: 165-171.

4669 Raudsepp T, Finno CJ, Bellone RR, et al. 2019. Ten years of the horse reference genome: insights into
4670 equine biology, domestication and population dynamics in the post genome era. Animal Genetics 50:
4671 569-597.

4672 Reardon W, Mullen AM, Sweeney T, et al. 2010. Association of polymorphisms in candidate genes with
4673 colour, water-holding capacity, and composition traits in bovine M. longissimus and M.
4674 semimembranosus. Meat Science 86: 270-275.

4675 Rege JEO, Gibson JP. 2003. Animal genetic resources and economic development: issues in relation to
4676 economic valuation. Ecological Economics 45: 319-330.

Reich D, Thangaraj K, Patterson N, et al. 2009. Reconstructing Indian population history. Nature 461: 489-494.

Reimer C, Rubin CJ, Sharifi AR, et al.2018 Analysis of porcine body size variation using re-sequencing data of miniature and large pigs. BMC Genomics 19: 687.

Reimer C, Ha NT, Sharifi AR, et al. 2018. Analyses of the breed integrity of the Goettingen Minipig using pool-sequencing. In: Proceedings of the 11th World Congress of Genetics Applied to Livestock Production, Auckland, New Zealand.

Ren Z, Wang Y, Ren Y, et al. 2017. Enhancement of porcine intramuscular fat content by overexpression of the cytosolic form of phosphoenolpyruvate carboxykinase in skeletal muscle. Science Reports 7: 43746.

Revidatti MA, Gama LT, Martin Burriel I, et al. 2021. On the origins of American Criollo pigs: A common genetic background with a lasting Iberian signature. PloS ONE 16: e0251879.

Reyer H, Oster M, Magowan E, et al. 2017a. Strategies towards improved feed efficiency in pigs comprise molecular shifts in hepatic lipid and carbohydrate metabolism. International Journal of Molecular Sciences 18: 1674.

Reyer H, Shirali M, Ponsuksili S, et al. 2017b. Exploring the genetics of feed efficiency and feeding behaviour traits in a pig line highly selected for performance characteristics. Molecular Genetics and Genomics 292: 1001-1011.

Reynders, J. 2021. The impact of improved genetics on commercial pig production. Stockfarm 59, 11, 01 September 2021. Available from: https://journals.co.za/doi/epdf/10.10520/ejc-vp_stock_v11_n9_a19 / [Accessed: 14 December 2022].

Rhymer JM, Simberloff D. 1996. Extinction by hybridization and introgression. Annual Reviews 27: 83-109.

Rischkowsky B, Pilling D. 2007. The State of the World's Animal Genetic Resources for Food and Agriculture. Rome, Italy.

Robinson J, Penrith M. 2009. Pigs for profit: A manual for emerging pig farmers in South Africa. Menlo Park, South African Pork Producers, 200.

Rognes K. 2011. A review of the monophyly and composition of the Bengaliinae with the description of a new genus and species, and new evidence for the presence of Melanomyinae in the Afrotropical Region (Diptera, Calliphoridae). Zootaxa 2964: 1-60.

Romero-Suarez S, Shen J, Brotto L, et al. 2010. Muscle-specific inositide phosphatase (MIP/MTMR14) is reduced with age and its loss accelerates skeletal muscle aging process by altering calcium homeostasis. Journal of Aging 2: 504-513.

Rookmaaker, LC. (Ed). 1989. The Zoological Exploration of Southern Africa 1650-1790. Rotterdam: A.A Balkema.

Roosenvelt T, Heller E. (Eds). 1915. Life-histories of African game animals. In John Murray. London.

4713 Ropka-Molik K, Bereta A, ˙Zukowski K, et al. 2018. Screening for candidate genes related with
4714 histological microstructure, meat quality and carcass characteristic in pig based on RNA-seq data.
4715 Asian-Australasian Journal of Animal Sciences 31: 1565-1574.

4716 Ross JW, Hale BJ, Gabler NK, et al. 2015. Physiological consequences of heat stress in pigs. Animal
4717 Production Science 55: 1381-1390.

4718 Rothschild MF. 2010. Association of genetic markers with structural soundness and its relationship to
4719 gilt development and sow longevity. Research Report.

4720 Rothschild MF, Ruvinsky A. (2nd Eds). 1998. The Genetics of the Pig. CABI Publishing, Oxfordshire,
4721 UK.

4722 Rubin CJ, Megens HJ, Martinez Barrio A, et al. 2012. Strong signatures of selection in the domestic pig
4723 genome. Proceedings of the National Academy of Sciences of the United States of America 109:
4724 19529-19536.

4725 Rui L, Sun DX, Wang Y, et al. 2013. Fine mapping QTLs affecting milk production traits on BTA6 in
4726 Chinese Holstein with SNP markers. Journal of Integrative Agriculture 12: 110-117.

4727 Ryu J, Lee C. 2016. Genetic association of marbling score with intragenic nucleotide variants at
4728 selection signals of the bovine genome. Animal 10: 566-570.

4729 Saade G, Deblanc C, Bougon J, et al. 2020. Coinfections and their molecular consequences in the
4730 porcine respiratory tract. Veterinary Research 51: 80.

4731 Sabeti PC, Reich DE, Higgins JM, et al. 2002. Detecting recent positive selection in the human genome
4732 from haplotype structure. Nature 419: 832-837.

4733 Sabeti PC, Varilly P, Fry B, et al. 2007. Genome-wide detection and characterization of positive
4734 selection in human populations. Nature 449: 913-918.

4735 Sachidanandam R, Weissman D, Schmidt SC, et al. 2001. A map of human genome sequence variation
4736 containing 1.42 million single nucleotide polymorphisms. Nature 409: 928.

4737 Salehi A, Sobhani R, Aminafshar M, et al. 2015. Single nucleotide of FGF2 gene in Iranian Holstein
4738 proven bulls. Molecular Biology Research Communications 4: 57-62.

4739 Sales J, Kotrba R. 2013. Meat from Wild Boar (Sus scrofa L.): A review. Meat Science 94: 187-201.

4740 SanCristobal M, Chevalet C, Haley CS, et al. 2006. Genetic diversity within and between European pig
4741 breeds using microsatellite markers. Animal Genetics 37: 189-198.

4742 Sandenbergh L, Cloete SWP, Roodt-Wilding R, et al. 2016. Evaluation of the OvineSNP50 chip for use
4743 in four South African sheep breeds. South African Journal of Animal Science 46: 89-93.

4744 SAPPO. 2022. South African pork industry report: November 2021. Agri Orbit. Available at
4745 https://agriorbit.com/south-african-pork-industry-report-november-2021/. [Accessed: 20 June
4746 2022].

4747 Sargolzaei M, Schenkel FS, Jansen GB, et al. 2008. Extent of linkage disequilibrium in Holstein cattle
4748 in North America. Journal of Dairy Science 91: 2106-2117.

4749 SAS. 2013. Statistical Analysis Software. Users' Guide Statistics Version 9.4. SAS Institute Inc., Cary.

Sawyer B, Camp D. 2020. Real breeds of miniature pigs. Available at https://www.minipiginfo.com/mini-pig-breeds-and-registries.html. [Accessed: 15 July 2021].

Scandura M, Iacolina L, Apollonio M. 2011. Genetic diversity in the European Wild Boar Sus scrofa: phylogeography, population structure and wild domestic hybridization. Mammal Review 41: 125-137.

Schachler K, Minx JO, Sürie C, et al. 2020. Genetic characterisation of Mini-LEWE as resource population for experimental research. Berliner and Münchener Tierärztliche Wochenschrift 133.

Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. American Journal of Human Genetics 78: 629-644.

Scherf BD. 2000. World watch list for animal diversity. FAO. Rome, Italy.

Schlötterer C. 2003. Hitchhiking mapping – functional genomics from the population genetics perspective. Trends Genet 19: 32-38.

Schook LB, Beever JE, Rogers J, et al. Swine 2005. Genome Sequencing Consortium (SGSC): A strategic roadmap for sequencing the pig genome. Comparative and Functional Genomics 6: 251-255.

Schwartz K, Lawn RM, Wade DP. 2000. ABC1 Gene expression and ApoA-I-Mediated cholesterol efflux regulated by LXR. Biochemical and Biophysical Research Communications 274: 794-802.

Siddiq A, Gueorguiev M, Samson C, et al. 2007. Single nucleotide polymorphisms in the neuropeptide Y2 receptor (NPY2R) gene and association with severe obesity in French white subjects. Diabetologia 50: 574-584.

Simianer H, Kohn F. 2010. Genetic management of the Gottingen Minipig population. Journal of Pharmacological and Toxicological Methods 62: 221-226.

Singh PK, Gahtyari NC, Roy C, et al. 2021. Wheat Blast: A disease spreading by intercontinental jumps and its management strategies. Frontiers in Plant Science 12: 710707.

Skead CJ. 2007. Historical incidence of the larger land mammals in the broader Eastern Cape. In Boshoff AF, Kerley GIH, Lloyd PH. Centre for African Conservation Ecology, Nelson Mandela Metropolitan University, Port Elizabeth, South Africa.

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

Smith DB, Simmonds P, Members of the International Committee on the Taxonomy of Viruses Study Group, et al. 2014. Consensus proposals for classification of the family Hepeviridae. Journal of General Virology 95: 2223-2232.

Smith SP, Phillips JB, Johnson ML, et al. 2019. Genome-wide association analysis uncovers variants for reproductive variation across dog breeds and links to domestication. Evolution, Medicine, and Public Health: 93-103.

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

4788 Solé M. Gori AS, Faux P. et al. 2017. Age-based partitioning of individual genomic inbreeding levels
4789 in Belgian Blue cattle. Genetics Selection Evolution 49: 92.

4790 Somers MJ, Penzhorn BL, Rasa AE. 1994. Home range size use and dispersal of warthogs in the Eastern
4791 Cape, South Africa. Journal of African Zoology 108: 361-373.

4792 Somers MJ, Penzhorn BL. 1992. Reproduction in an introduced Warthog population in the Eastern Cape.
4793 South African Journal of Wildlife Research 22: 57-60.

4794 Somers MJ, Rasa OA, Penzhorn BL. 1995. Group structure and social behaviour of warthogs
4795 Phacochoerus aethiopicus. Acta Theriologica. 40: 257-281.

4796 Song H, Ye S, Jiang Y, et al. 2019. Using imputation-based whole-genome sequencing data to improve
4797 the accuracy of genomic prediction for combined populations in pigs. Genetics Selection Evolution
4798 51: 58.

4799 Song J, Yang D, Ruan J, et al. 2017. Production of immunodeficient rabbits by multiplex embryo
4800 transfer and multiplex gene targeting. Science Reports 7: 12202.

4801 Sosa-Madrid BS, Santacreu MA, Fontanesi L, et al. 2018. A genomic region on chromosome 17 has a
4802 major impact on litter size traits in rabbits. In: Proceedings of the World Congress on Genetics
4803 Applied to Livestock Production 11: 163.

4804 South African Studbook Association. 1971. South African Studbook Jubilee Edition. Dreyer Printers
4805 and Publishers, Bloemfontein, South Africa.

4806 Stachowiak M, Szczerbal I, Switonski M. 2016. Genetics of adiposity in large animal models for human
4807 obesity-studies on pigs and dogs. Progress in Molecular Biology and Translational Science 140: 233-
4808 70.

4809 Stam P. 1980. The distribution of the fraction of the genome identical by descent infinite random mating
4810 populations. Genetics Research 35: 131-155.

4811 Statista. 2021. Global number of pigs 2012-2021, by country. Available at
4812 https://www.statista.com/statistics/263963/number-of-pigs-worldwide-since-1990/. [Accessed: 18
4813 August 2021].

4814 Statista. 2022. Stock of pigs in South Africa from 2010 to 2020. Available at
4815 https://www.statista.com/statistics/263963/number-of-pigs-worldwide-since-1990/. [Accessed: 11
4816 December 2022].

4817 Stratz P, Baes C, Rückert C, et al. 2013. A two-step approach to map quantitative trait loci for meat
4818 quality in connected porcine F2 crosses considering main and epistatic effects. Animal Genetics 4:
4819 360.

4820 Stella A, Nicolazzi EL, Van Tassell CP. et al. 2018. AdaptMap: exploring goat diversity and adaptation.
4821 Genetics Selection Evolution 50: 61.

4822 Stratz P, Schmid M, Wellmann R, et al. 2018. Linkage disequilibrium pattern and genome-wide
4823 association mapping for meat traits in multiple porcine F2 crosses. Animal Genetics 49: 403-412.

4824 Streicher S. 2013. Personal Communication. CEO of South African Pork Producers Organisation.
4825 simon@sapork.com.

Subalini E, Silva GL, Demetawewa CM. 2010. Phenotypic characterization and production performance of village pigs in Sri Lanka. Tropical Agricultural Research 21:198-208.

Suchocki T, Wojdak-Maksymiec, K, Szyda J. 2016. Using gene networks to identify genes and pathways involved in milk production traits in Polish Holstein dairy cattle. Czech Journal of Animal Science 61: 526-538.

Supek F, Bosnjak M, Smuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS ONE 6: e21800.

Sved JA. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theoretical Population Biology 2: 125-141.

Swanepoel M, Leslie AJ, van der Rijst M, et al. 2016. Physical and chemical characteristics of warthog (Phacochoerus africanus) meat. African Journal of Wildlife Research 46: 2.

Swanepoel M, Schulze E, Cumming DHM. 2016. A conservation assessment of Phacochoerus africanus. Child MF, Roxburgh L, Do Linh San E. (Eds). The Red List of Mammals of South Africa, Swaziland and Lesotho. South African National Biodiversity Institute and Endangered Wildlife Trust, South Africa.

Swart H, Kotze A, Olivier PAS, et al. 2010. Microsatellite-based characterization of Southern African domestic pigs (Sus scrofa domestica). South African Journal of Animal Science 40: 121-132.

Sweett H, Miglior F, Livernois A, et al. 2018. Genome-wide association study to identify genomic regions and single nucleotide polymorphisms functionally associated with bull fertility. Journal of Animal Science 96: 138-139.

Switonski M, Stachowiak J, Cieslak M, et al. 2010. Genetics of fat tissue accumulation in pigs: a comparative approach. Journal of Applied Genetics 51: 153-168.

Taaibosch Piggery. 2021. Pig farming: Export of quality pig genetics from South Africa to Nigeria. October 2021. Available from: https://www.taaibosch.com/post/pig-farming-quality-export-of-pig-genetics-from-south-africa-to-nigeria. [Accessed: 14 December 2022].

Taberlet PA Valentini HR, Rezael S, et al. 2008. Are cattle, sheep and goats endangered species? Molecular Ecology 17: 275-284.

Tack J. 2018. Wild boar (Sus scrofa) populations in Europe: A scientific review of population trends and implications for management. European Landowners' Organization, Brussels, 56.

Tang H, Peng J, Wang P, et al. 2005. Estimation of individual admixture: analytical and study design considerations. Genetic Epidemiology 28: 289-301.

Tapio M, Marzanov N, Ozerov M, et al. 2006. Sheep mitochondrial DNA variation in European, Caucasian, and Central Asian areas. Molecular Biology and Evolution 23: 1776-1783.

Taverner MR., Dunkin AC. 1996. (Eds). Pig Production: World animal science. Elsevier Science Publishers. pp. 113-140.

Taylor AR, Echeverry DF, Anderson TJC, et al. 2020. Identity-by-descent with uncertainty characterises connectivity of Plasmodium falciparum populations on the Colombian-Pacific coast. PLoS Genetics 16: e1009101.

4864 Taylor G, Roese G. 2005. Breeds of pigs – Large White. Profitable and sustainable industries. Primefact
4865 62, November 2005.

4866 Team R. 2020. RStudio: Integrated development for R. RStudio Inc. (2015). Acknowledgements this
4867 work was supported by Fundação para a Ciência e Tecnologia (FCT), Azores.

4868 Templeton AR, Maxwell T, Posada D, et al. 2005. Tree scanning: a method for using haplotype trees in
4869 phenotype/genotype association studies. Genetics 169: 441-453.

4870 Terwilliger JD, Zöllner S, Laan M, et al. 1998. Mapping genes through the use of linkage disequilibrium
4871 generated by genetic drift: 'drift mapping' in small populations with no demographic expansion.
4872 Human Heredity 48: 138-154.

4873 Thamaga-Chitja JM, Morojele P. 2017. The context of smallholder farming in South Africa: towards a
4874 livelihood asset building framework. Journal of Human Ecology 45: 147-155.

4875 Thenius E, 1970. Zur Evolution und Verbreitungsgeschichte der Suidae (Artiodactyla, Mammalia).
4876 Zeitschrift für Saugetierkunde 35: 321-384.

4877 Thivhilaheli RN, Tsatsimpe M, Muller M, et al. 2020. Artificial insemination training program for
4878 smallholder pig farms in Gauteng Province, South Africa. doi:10.5539/jas.v12n12p89.

4879 Thomas AD, Kolbe KK. 1942. The wild pigs of South Africa: Their distribution and habits and their
4880 significance as agricultural pests and carriers of disease. Journal of the South African Veterinary
4881 Association 8: 1-11.

4882 Thompson EA. 2013. Identity by descent: Variation in meiosis, across genomes, and in
4883 populations.Genetics 194: 301-326.

4884 Tizioto PC, Taylor JF, Decker JE, et al. 2015. Detection of quantitative trait loci for mineral content of
4885 Nelore longissimus dorsi muscle. Genetic Selection Evolution 47: 15.

4886 Toosi A, Fernando RL, Dekkers JCM, 2010. Genomic selection in admixed and crossbred populations,
4887 Journal of Animal Science 88: 32-46.

4888 Tosser-Klopp G, Bardou P, Bouchez O, et al. 2014. Design and characterization of a 52K SNP Chip for
4889 Goats. PLoS ONE 9: e86227.

4890 Tóth EG, Tremblay F, Housset JM, et al. 2019. Geographic isolation and climatic variability contribute
4891 to genetic differentiation in fragmented populations of the long-lived subalpine conifer Pinus cembra
4892 L. in the western Alps. BMC Evolution Biology 19: 190.

4893 Treacy DA. 1976. A genetic analysis of the pedigree Landrace pig breed in Australia. Australian Journal
4894 of Experimental Agriculture and Animal Husbandary 16: 76-81.

4895 Trevisan L, Brum JS. 2020. Incidence of pale, soft and exudative (PSE) pork meat in reason of extrinsic
4896 stress factors. Anais da Academia Brasileira de Ciências 92: e20190086.

4897 Uimari P, Tapio M. 2011. Extent of linkage disequilibrium and effective population size in Finnish
4898 Landrace and Finnish Yorkshire pig breeds. Journal Animal Science 89: 609-614.

4899 Uimari P, Sironen AA. 2014. Combination of two variants in PRKAG3 is needed for a positive effect
4900 on meat quality in pigs. BMC Genetics 15: 29.

Uimari P, Sironen A, Sevon-Aimonen ML. 2011. Whole-genome SNP association analysis of reproduction traits in the Finnish Landrace pig breed. Genetics Selection Evolution 43: 42.

Vahedi SM, Ardestani SS, Do DN, et al. 2021. 149 Integration of selection signatures analyses and weighted single-step GWAS to prioritize candidate genes for body conformation traits in pigs, Journal of Animal Science 99: 76-77.

van Hossou SFU, Yin T, Scheper C, et al. 2021. Unraveling admixture, inbreeding, and recent selection signatures in West African indigenous cattle populations in Benin. Frontiers in Genetics 12: 657282.

van der Westhuizen L, MacNeil MD, Scholtz MM, et al. 2020. Genetic variability and relationships in nine South African cattle breeds using microsatellite markers. Tropical Animal Health and Production. 52: 177-184.

van Laere AS, Nguyen M, Braunschweig M, et al. 2003. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. Nature 425: 832-836.

van Tonder E. 2019. In search of the origins of the Kolbroek. Available at https://earthwormexpress.com/kolbroek-and-the-kune-kune/the-origins-of-the-kolbroek/. [Accessed: 15 July 2021].

Vatsiou AI, Bazin E, Gaggiotti OE. 2016. Detection of selective sweeps in structured populations: a comparison of recent methods. Molecular Ecology 25: 89-103.

Vaughan HW.1950. Breeds of livestock in America. Long's College Book Co.: Columbus, OH.

Venter J, Ehlers-Smith T, Seydeck A. 2016. A conservation assessment of Potamochoerus larvatus. Child MF, Roxburgh L, Do Linh San E. (Eds). The Red List of Mammals of South Africa, Swaziland and Lesotho. South African National Biodiversity Institute and Endangered Wildlife Trust, South Africa.

Verardo LL, Silva FF, Lopes MS, et al. 2016. Revealing new candidate genes for reproductive traits in pigs: Combining Bayesian GWAS and functional pathways. Genetic Selection Evolution 48: 9.

Verado LL, Silva FF, Varona L. et al. 2015. Bayesian GWAS and network analysis revealed new candidate genes for number of teats in pigs. Journal of Applied Genetics 56: 123-132.

Vercammen P, Seydack AHW, Oliver WLR. 1993. The Bush Pigs (Potamochoerus larvatus and P. porcus): Pigs, peccaries, and hippos: status survey and conservation action plan. Oliver WLR. IUCN SSC Pigs and Peccaries Specialist Group and IUCN SSC Hippo Specialist Group: 93-101.

Veroneze R, Lopes PS, Guimarães SEF, et al. 2013 Linkage disequilibrium and haplotype block structure in six commercial pig lines. Journal of Animal Science 91: 3493-3501.

Vigors S, O'Doherty JV, Bryan K, et al. 2019. A comparative analysis of the transcriptome profiles of liver and muscle tissue in pigs divergent for feed efficiency. BMC Genomics 20: 461.

Vigors S, Sweeney T, O'Shea CJ, et al. 2016. Pigs that are divergent in feed efficiency, differ in intestinal enzyme and nutrient transporter gene expression, nutrient digestibility and microbial activity. Animal 10: 1848-1855.

4937 Visser C, Lashmar SF, Van Marle-Köster E, et al. 2016. Genetic diversity and population structure in
4938 South African, French and Argentinian angora goats from genome-wide SNPData. PLoS ONE 11:
4939 e0154353.

4940 Visser C, van Marle-Köster E, Friedrich H. 2011. Parentage verification of South African Angora goats,
4941 using microsatellite markers. South African Journal of Animal Science 41: 250-255.

4942 Visser D. 2014. Modern Pig Production. Available at
4943 https://books.google.co.za/books?id=mpXorQEACAAJ. [Accessed 14 October 2016].

4944 Visser DP. 2012. (Ed). Die Kolbroek-'n Kaleidoskoop van Moontlikhede. Kejafa Knowledge Works
4945 Publisher.

4946 Visser DP, Browne A, Botes F, et al. 1993. Breeds and breeding programs. Pig Production in South
4947 Africa. Kemm EH. (Ed). Agricultural Research Council. V and R Printers, Pretoria. pp. 19-20.

4948 Visser DP, Kotze A. 1996. Status of genetic variation in purebred pig breeds in South Africa. In:
4949 Proceedings of the 7th All Africa Conference on Animal Agriculture, Pretoria.

4950 Voight BF, Kudaravalli S, Wen X, et al. 2006 A map of recent positive selection in the human genome.
4951 PLoS Biology 4: e72.

4952 Wang HT, Zhang Z, Fan M, et al. 2010. Expression of CD112 in colon carcinoma tissues and cell lines
4953 and their clinical significance. Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi 26: 477-479.

4954 Wang JY, Guo JF, Zhang Q, et al. 2011. Genetic diversity of Chinese indigenous pig breeds in Shandong
4955 province using microsatellite markers. Asian Australasian Journal of Animal Sciences 24: 28-36.

4956 Wang JY, Luo YR, Fu WX, et al. 2013. Genome-wide association studies for hematological traits in
4957 swine. Animal Genetics 44: 34-43.

4958 Wang K, Zhang L, Duan D, et al. 2021. Genomic Analysis Reveals Human-Mediated Introgression
4959 From European Commercial Pigs to Henan Indigenous Pigs. Frontiers in Genetics 12: 705803.

4960 Wang S, Li D, Zhu M, et al. 2019. A single nucleotide polymorphism of the porcine CXCL8 gene is
4961 associated with serum CXCL8 level. Italian Journal of Animal Science 18: 474-479.

4962 Wang X, Ma P, Liu J, et al. 2015. Genome wide association study in Chinese Holstein cows reveal two
4963 candidate gene for somatic cell score as an indicator for mastitis susceptibility. BMC Genetics 16:
4964 111.

4965 Wang Y, Zhao X, Wang C, et al. 2021. High-density single nucleotide polymorphism chip-based
4966 conservation genetic analysis of indigenous pig breeds from Shandong Province, China. Animal
4967 Bioscience 34: 1123-1133.

4968 Wang Z, Chen Q, Yang Y, et al. 2014. A genome-wide scan for selection signatures in Yorkshire and
4969 Landrace pigs based on sequencing data. Animal Genetics 45: 808-816.

4970 Watanabe T, Hayashi Y, Kimura J, et al. 1986. Pig mitochondrial DNA: polymorphism, restriction map
4971 orientation, and sequence data. Biochemical Genetics 24: 385-396.

4972 Webb EC, Phiri LM, Pieterse E. 2006. Gonadal development of commercial pig genotypes in South
4973 Africa. South African Journal of Animal Science 36: 26-29.

Weinberger MW, Qaim M. 2009. The food system transformation in developing countries: a disaggregate demand analysis for fruits and vegetables in Vietnam. Food Policy 34: 426-436.

White S. 2011. From globalized pig breeds to capitalist pigs: A study in animal cultures and evolutionary history. Environmental History 16: 94-120.

White TD, Harris JM. 1977. Suid evolution and correlation of African hominid localities. Science 198: 13-21.

Whitlock MC, McCauley DE. 1999. Indirect measures of gene flow and migration: FST not equal to 1/(4Nm + 1). Heredity 82: 117-125.

Whittemore CT. 2006a. Development and improvement of pigs by genetic selection. Kyriazakis I, Whittemore CT (3rd Eds.). Whittemore's Science and Practice of Pig Production. Oxford, UK: Blackwell Publishing Ltd. pp. 184-262.

Wiener P, Wilkinson S. 2011. Deciphering the genetic basis of animal domestication. Proceedings of the Royal Society B: Biological Sciences 278: 3161-3170.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370.

Weka R, Bwala D, Adedeji Y, et al. 2021 Tracing the Domestic Pigs in Africa. Kušec G, Kušec ID (Eds.). London, United Kingdom: IntechOpen, London. https://doi.org/10.5772/intechopen.95077

Wickham H. 2016. (Ed). ggplot2: Elegant graphics for data analysis. New York, NY: Springer.

Wilkinson S, Lu ZH, Megens HJ, et al. 2013. Signatures of diversifying selection in European pig breeds. PLoS Genetics 9: e1003453.

Willi Y, Van Buskirk J, Hoffman AA. 2006. Limits to the adaptive potential of small populations. Annual Review of Ecology, Evolution and Systematics 37: 433-458.

Wong GK, Liu B, Wang J. 2004. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature 432: 717-722.

Wright S. 1931. Evolution in Mendelian Populations. Genetics 16: 97-159.

Wright S. 1951. The genetical structure of populations. Annals of Eugenics 15: 323-54.

Wright S. 1978. Evolution and the genetics of populations. Vol. 4. Variability within and among Natural Populations. The University of Chicago Press. Chicago, IL.

Wu F, Sun H, Lu S, et al. 2020. Genetic diversity and selection signatures within Diannan small-ear pigs revealed by next-generation sequencing. Frontiers in Genetics 11: 733.

Wu GS, Yao YG, Qu KX, et al. 2007. Population phylogenomic analysis of mitochondrial DNA in Wild Boars and domestic pigs revealed multiple domestication events in East Asia. Genome Biology 8: R245.

Wu X, Zhang Q, Xu S, et al. 2016. Differential expression of six chicken genes associated with fatness traits in a divergently selected broiler population. Molecular and Cellular Probes 30: 1-5.

5009 Xiang R, Oddy VH, Archibald AL, et al. 2016. Epithelial, metabolic and innate immunity transcriptomic
5010 signatures differentiating the rumen from other sheep and mammalian gastrointestinal tract tissues.
5011 Peer Journal 4: e1762.

5012 Yang B, Cui L, Perez-Enciso M, et al. 2017. Genome-wide SNP data unveils the globalization of
5013 domestication pigs. Genetic Selection Evolution 49: 71.

5014 Yang SL, Wang ZG, Liu B, et al. 2003. Genetic variation and relationships of eighteen Chinese
5015 indigenous pig breeds. Genetics Selection Evolution 35: 657.

5016 Yazar S, Cuellar-Partida G, McKnight CM, et al. 2015. Genetic and environmental factors in
5017 conjunctival UV autofluorescence. JAMA Ophthalmology 133: 406-412.

5018 Yoshida GM, Lhorente JP, Carvalheiro R, et al. 2017. Bayesian genome-wide association analysis for
5019 body weight in farmed Atlantic salmon (Salmo salar L.). Animal Genetics 48: 698-703.

5020 Yu G, Xiang H, Wang J, et al. 2013. The phylogenetic status of typical Chinese native pigs: Analyzed
5021 by Asian and European pig mitochondrial genome sequences Journal of Animal Science and
5022 Biotechnology 4: 9.

5023 Zadik BJ. 2005. The Iberian pig in Spain and the Americas at the time of Columbus. MSc dissertation,
5024 University of California, USA.

5025 Zanga J, Chimonyo M, Kanengoni A. 2003. A comparison of the susceptibility of growing Mukota and
5026 Large White pig genotypes to infection with Ascaris suum. Veterinary Research Communications
5027 27 653-660.

5028 Zeder MA, Emshwiller E, Smith BD, et al. 2006. Documenting domestication: the intersection of
5029 genetics and archaeology. Trends in Genetics 22: 139-155.

5030 Zhang J, Rodríguez F, Navas MJ, et al. 2020. Fecal microbiota transplantation from Warthog to pig
5031 confirms the influence of the gut microbiota on African swine fever susceptibility. Science Reports
5032 10: 17605.

5033 Zhan J. 2016. Population genetics of plant pathogens. McDonald BA, Hoboken NJ. (Eds). John Wiley
5034 & Sons, Ltd. pp. 1-7.

5035 Zhang J, Jiao T, Zhao S. 2016. Genetic diversity in the mitochondrial DNA D-loop region of global
5036 swine (Sus scrofa) populations. Biochemical and Biophysical Research Communications 473:
5037 814e820.

5038 Zhang M, Li F, Ma XF, et al. 2019. Identification of differentially expressed genes and pathways
5039 between intramuscular and abdominal fat-derived preadipocyte differentiation of chickens in vitro.
5040 BMC Genomics 20: 743.

5041 Zhang W, Fan Z, Han E, et al. 2014. Hypoxia adaptations in the grey wolf (Canis lupus chanco) from
5042 Qinghai-Tibet Plateau. PLoS Genetics 10: e1004466.

5043 Zhang ZG. (Ed) 1986. Pig breeds in China. Shanghai: Shanghai Scientific and Technical Publishers. pp.
5044 73-277.

5045 Zhang Z, Xiao Q, Zhang QQ, et al. 2018. Genomic analysis reveals genes affecting distinct phenotypes
5046 among different Chinese and western pig breeds. Science Reports 8: 13352.

5047 Zhao F, McParland S, Kearney F, et al. 2015. Detection of selection signatures in dairy and beef cattle
5048 using high-density genomic information. Genetics Selection Evolution 47: 49.

5049 Zhao X, Chen S, Tan Z, et al. 2019. Transcriptome analysis of landrace pig subcutaneous preadipocytes
5050 during adipogenic differentiation. Genes (Basel) 10: 552.

5051 Zhou L, Ji J, Peng S, et al. 2016. A GWA study reveals genetic loci for body conformation traits in
5052 Chinese Laiwu pigs and its implications for human BMI. Mammalian Genome 27: 610-621.

5053 Zhou Y, Browning BL, Browning SR. 2020. Population specific recombination maps from segments of
5054 identity by descent. American Journal of Human Genetics 107: 137-148.

5055 Zhu JJ, Ramanathan P, Bishop EA, et al. 2019. Mechanisms of African swine fever virus pathogenesis
5056 and immune evasion inferred from gene expression changes in infected swine macrophages. PLoS
5057 ONE 14: e0223955.

5058 Zuo B. Xiong Y, Deng CY, et al. 2004. cDNA cloning, genomic structure and polymorphism of the
5059 porcineFHL3 gene. Animal Genetics 35: 230-233.

5060 Zuo B, Xiong Y, Yang H, et al. 2007. Full-length cDNA, expression pattern and association analysis of
5061 the porcine FHL3 gene. Asian-Australasian Journal of Animal Sciences 20: 1473-1477.
5062

APPENDIX A: SUPPLEMENTARY FILES

Supplementary files are available on the following link:

https://drive.google.com/file/d/1vYXcoYPm4pF8yQc1hczn7t67v5F_PbYP/view

Each chapter has its own folder as follows:

Chapter 3:

Supplementary Table 3.1: Supplementary Table 3.1 Average effective population values

Supplementary Table 3.2: Partitioning genetic variability

Supplementary Table 3.3: Pairwise genetic differentiation (F_{ST}) between South African and Burgoz-Paz et al. (2013) worldwide populations

Supplementary Figure 3.1: Cross validation plot

Supplementary Figure 3.2: Genome-wide Manhattan plot of F_{ST} among the pig populations of a) Alfred Nzo and Warthog, b) Duroc and Warthog, c) Kolbroek and Warthog, d) Large White and Warthog, e) South African landrace and Warthog, f) Windsnyer and Warthog, g) Indigenous and Duroc, h) Villages and Duroc, i) Villages and Kolbroek, j) South African Landrace with Large White and Indigenous, k) Indigenous and Vietnamese, l) Villages and Wild Boar, m) Villages and Vietnamese, n) Wild Boar and Duroc. The solid lines indicate the $F_{ST} \geq 0.8$ thresholds.

Chapter 4:

Supplementary Table 4.1: The highest haplotypes for the other eight populations

Supplementary Table 4.2: Unique haplotypes genes and QTLs for each population

Supplementary Table 4.3: Shared haplotype blocks between pig populations

Supplementary Table 4.4: Shared haplotypes genes and QTLs for between the nine populations

Supplementary Table 4.5: SNPs associated with high LD regions

Supplementary Table 4.6: QTLs associated with high region LD

Supplementary Figure 4.1: Ther number of haplotypes blocks for chromosomes 1-18

Supplementary Figure 4.2: Average block length for chromosomes 1-18

Supplementary Figure 4.3: Sum of block length for chromosomes 1 -18

Supplementary Figure 4.4: Number of haplotype blocks at varying length

Chapter 5:

Supplementary Table 5.1: Detailed *frappe* based estimates of proportions for South African pig populations

Supplementary Table 5.2: Migration events set the highest Mean (*f*) at 10

Supplementary Table 5.3: GO functional enrichment for the candidate genes associated with the ALN_CAP pig populations

Supplementary Table 5.4: GO functional enrichment for the candidate genes associated with the ALN_MOP pig populations

Supplementary Table 5.5: GO functional enrichment for the candidate genes associated with the ALN_ORT pig populations

Supplementary Table 5.6: GO functional enrichment for the candidate genes associated with the CAP_MOP pig populations

Supplementary Table 5.7: GO functional enrichment for the candidate genes associated with the CAP_ORT pig populations

Supplementary Table 5.8: GO functional enrichment for the candidate genes associated with the MOP_ORT pig populations

Supplementary Table 5.9: REVIGO enriched GO terms associated with all pig populations

Chapter 6:

Supplementary Table 6.1: Summaries of the number of potential regions detected in *iHS*

Supplementary Table 6.2: List of selected regions and candidate genes detected in *iHS* method associated with each pig population

Supplementary Table 6.3: Summaries of the number of potential regions for *XP-EHH*

Supplementary Table 6.4: Genomic regions under divergent selection identified by *XP-EHH* method and associated candidate genes

Supplementary Table 6.5: Genomic regions under divergent selection identified by *HapFLK* method and associated candidate genes

Supplementary Table 6.6: Enriched GO terms for significant genes identified using *iHS*, *XP-EHH* and *HapFLK*

Supplementary Table 6.6: Pathways associated genomic regions under signatures of selection