



Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

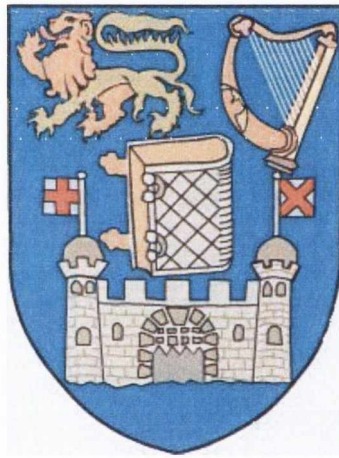
Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Genomewide admixture mapping of African cattle

Yonas Kassahun Hirutu



A thesis submitted to the University of Dublin for the degree of Doctor of Philosophy

Department of Genetics

University of Dublin

Trinity College

September 2011



Thesis 9645

Declaration

I hereby certify that this thesis has not been previously submitted for examination to this or any other university. The work described herein has been carried out by the author alone, except otherwise stated.

This thesis may be made available for consultation within the university library and may be photocopied or loaned to other libraries for the purpose of consultation.



Acknowledgements

I am heartily thankful to my supervisor, Dan Bradley, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject. This thesis would not have been possible without his extreme patient with me during the whole process.

I would also like to thank my parents (Agie and Kassahun) for me unconditional support and the love you have given me thought my life. Special thanks to Saron and Martha for waiting in patient, support and love you have given me. My sisters, brothers and members of the family (Tinbit, Mekdelawit, Migbar).

I owe my deepest gratitude to Gobena Ameni for site identification and coordinating the field work, field technical team members includes; Sorane Gemed, Nega Negash and veterinary graduating students. I also thank the regional bureau heads of Selalle regions for allowing us to work in the area and coordinating farmers for cattle examination. I also thank farmers in the region for their willingness and participate in the study. I would also like to extend my appreciation to Elena Hailu for kindly help during DNA extraction and sample shipment to the lab in Dublin. I would like also to thank Valeria Mattiangeli for helping me thought the study during sample and data processing. Valeria and Tony thank you so much for reading this thesis and giving me valuable comments. Ruth Freeman, thank you for giving me the chance to join the lab and helped to settle in Dublin.

I also acknowledge Aramuer Hansen research institute (AHRI) current director Abraham Aseffa, former director Howard Engers for their support and allowing me to take part in the bovine tuberculosis project and facilitating all the logistics during the sample collection process. I also thank all of AHRI staff members for their encouragement and support.

I would like to thank everyone in the all lab; Emma, Frauke, Lillian, Sarah, Tim, Matthew, Russell, Kevin. I would also extend my appreciation to Karsten for his help in Bioinformatics and the department of Genetics, TCD.

My friends in Dublin; Asa (for your continuous encouragement, helping me with perl programming and sharing phd life together), Jerry (the best flatmate and friend), Haile (thank you for looking after me all the time), Andrew, Linda, Kacia, Workeneh, Hibret and Arthur for wonderful and unforgettable time.

I am indebted to my friends and colleagues who supported me even when I disappear for a long time without contact especially to Ebba, Daniel, Hiwot, Zeyede, Amare, Endris, Dawit W/M, Liya, Kidist, Markos, Abebe, Wude, Tseihaynesh, Haimanot and Menbere. I am so happy that I am coming back to home and see you all.

Table of Contents

CHAPTER ONE: GENERAL INTRODUCTION.....	1
1.1 Introduction.....	2
1.1.1 Cattle origins.....	2
1.1.2 African cattle.....	3
1.1.3 The nature of cattle Admixture in Africa.....	4
1.1.4 Host genetic susceptibility to bovine tuberculosis.....	6
1.1.5 Case- referent strategies for identification of susceptibility loci to bovine tuberculosis.....	8
1.1.6 Admixture mapping analysis.....	11
1.1.7 Application of Admixture mapping analysis.....	14
CHAPTER TWO: GENOMEWIDE ADMIXTURE ANALYSIS OF AFRICAN CATTLE POPULATIONS	18
2.1 Introduction.....	19
2.1.2 Cattle genetic structure in Africa.....	19
2.1.3 Admixture mapping.....	21
2.2 Materials and Methods.....	24
2.2.1 Study Population and samples.....	24
2.2.2 Markers.....	24
2.2.3 Information content of the markers.....	25
2.2.4 Admixture analysis.....	25
2.2.5 Statistical Analysis.....	26
2.3 RESULTS.....	29
2.3.1 Informative markers.....	29
2.3.2 Genetic distance.....	32
2.3.3 Principal component analysis.....	33
2.3.4 Population Genetic Structure of Ethiopian zebu and Borgou.....	35
2.3.5 Age of admixture.....	37

2.3.6 Admixture mapping of African cattle hybrid zones	39
2.3.7 Genome mapping of specific regions of Ethiopian zebu cattle under selection.....	44
2.3.8 Genome mapping of specific regions of Borgou cattle under selection.....	56
2.4 Discussion	61
CHAPTER THREE: PHYLOGEOGRAPHY OF CATTLE	65
3.1 Introduction	66
3.2 Materials and Methods.....	68
3.2.1 Study populations	68
3.2.2 Ancestry Informative Markers	70
3.2.3 Population Structure and phylogeography	71
3.2.4 Individual admixture levels.....	71
3.2.5 Haplotype inference and measure of LD	72
3.2.6 Age of admixture	72
3.3 Results	73
3.3.1 Selection of Ancestry informative markers and genotype quality control	73
3.3.2 Genetic distance and phylogeography	76
3.3.3 Population structure and Admixture.....	79
3.3.4 Age of admixture	83
3.3.5 Haplotype structure and measure of LD	86
3.4 Discussion	90
CHAPTER FOUR: FIELD SAMPLE COLLECTION	93
4.1 Introduction	94
4.2 Materials and Methods.....	97
4.2.1 The study area	97
4.2.2 The Study sample	98
4.2.3 Comparative intradermal tuberculin test (CIDT)	100
4.2.4 DNA extraction	102
4.3 Results	103

4.4 Discussion	105
CHAPTER FIVE: GENOMEWIDE ADMIXTURE MAPPING SCAN FOR HOST SUSCEPTIBILITY TO BOVINE TUBERCULOSIS	107
5.1 Introduction	108
5.2 Material and methods	112
5.2.1 Study populations	112
5.2.2 Marker panel.....	112
5.2.3 Estimates of Ancestral allele frequency.....	112
5.2.4 SNPs and study sample quality control	112
5.2.5 Genomewide Admixture mapping scan.....	113
5.2.6 Statistical analysis for association of CIDT positivity and <i>Bos taurus</i> ancestry	114
5.3 Results	117
5.3.1 ANCESTRYMAP admixture mapping scan of tuberculin positivity.....	121
5.3.2 ADMIXMAP analysis of cases and controls for tuberculin skin test as a binary and continuous variable.....	126
5.3.3 Testing locus specific association with tuberculin skin test positivity.....	131
5.3.4 Possible list of associated genes for host immunity to infection	132
5.4 Discussion	135
CHAPTER SIX: GENOMEWIDE ADMIXTURE SCANS OF CATTLE PHENOTYPE TRAITS.....	139
6.1 Introduction	140
6.2 Materials and Methods	142
6.2.1 Study populations	142
6.2.1 Measurement of phenotype traits	142
6.2.1.1 Animal coat colour	142
6.2.1.2 Animal Body condition score	143
6.2.1.3 Grading dewlap and hump sizes.....	143
6.2.2 Genotype data for the ancestry informative markers.....	144
6.2.3 Genomewide ancestry association of phenotype traits.....	144

6.2.4 Statistical analysis.....	145
6.3 Results	146
6.3.1 Animal coat colour variants	146
6.3.2 Bovine QTL mapping for coat colour	154
6.3.3 Quantitative Admixture mapping for body condition score	158
6.3.4 Admixture mapping for dewlap and hump development.....	160
6.4 Discussion	163
General Discussion.....	166
CHAPTER SEVEN: LITERATURE CITED	170
7.1Reference	171
Appendix A: Genomewide locus specific -log P-value (Y-axis) for African <i>Bos taurus</i> and <i>Bos indicus</i> ancestry estimate of Ethiopian Zebu (ETZ) cattle using 9150 markers according to physical position (Mb) indicated on X-axis	i
Appendix B: Genomewide locus specific -log P-value (Y-axis) for African <i>Bos taurus</i> and <i>Bos indicus</i> ancestry estimate of Borgou (BOR) cattle using 9150 markers according to physical position (Mb) indicated on X-axis.....	xv

Figures

Figure 2.1 Contour map (Freeman et al. 2006a) for cattle genetic structure based on microsatellite allele frequencies; estimated <i>Bos taurus</i> ancestral proportion are plotted.....	20
Figure 2.2 Dot distribution showing allele frequency difference (δ) of 9150 ancestry informative markers	30
Figure 2.3 Dot distribution showing allele frequency difference (δ) of 7340 ancestry informative markers. δ values of selected top PCA correlated 7340 snps selected for comparative analysis.....	31
Figure 2.4 Principal component analysis representation of 7 populations of 156 cattle using all 39674 SNPs.....	33
Figure 2.5 Principal component analysis representation of 6 populations, after removing Holstein (HOL) from the analysis.....	35
Figure 2.6 Estimated Ln likelihood values for each number of population clusters (K) tested using 9150 ancestry informative markers.....	36
Figure 2.7 STRUCTURE clustering of seven breeds into two when K=2 used (top) showing the general subdivision into main taxa (<i>Bos taurus</i> and <i>Bos indicus</i>) and admixed population.....	37
Figure 2.8 Hapmix estimation of likelihood (y-axis) of age of admixture (number of generation) on the x-axis for Ethiopian zebu (ETZ) in blue and Borgou (BOR) in red. ..	38
Figure 2.9a Genomewide ancestry estimate (q) in Ethiopian zebu.	40
Figure 2.9b Genomewide ancestry estimate (q) in Borgou.	41
Figure 2.10 Manhattan plot of $-\log_{10}$ (P values) by chromosomal location of Ethiopian zebu (ETZ) cattle using 39674 SNPs (A) and using only ancestry informative (AIM) 9150 markers (B).	45
Figure 2.11 Ethiopian zebu (A) and Borgou (B) normal Q-Q plot for African <i>Bos taurus</i> ancestry estimate from $-\log$ P values using 9150 ancestry informative markers.....	46
Figure 2.12 Network analysis of genes associated with excess <i>Bos taurus</i> ancestry in Ethiopian zebu. Ingenuity Pathway Analysis (IPA) indicates known interactions that involve antigen presentation, cell-to-cell signaling and interaction, haematological system development and function.....	52

Figure 2.13. Network analysis of genes associated with excess <i>Bos taurus</i> ancestry in Ethiopian zebu. Ingenuity Pathway Analysis (IPA) indicates known interactions that involve in cell death, cell-to-cell signaling and interaction, cell signaling.	53
Figure 2.14. Network analysis of genes associated with excess <i>Bos indicus</i> ancestry in Ethiopian zebu. Ingenuity Pathway Analysis (IPA) indicates known interactions involved in Nucleic Acid Metabolism, Small Molecule Biochemistry, and Infectious Disease.	55
Figure 2.15. Manhattan plot of $-\log_{10}$ (P values) by chromosomal location of Borgou (BOR) cattle using 39674 SNPs (A) and using only ancestry informative (AIM) 9150 markers (B).	57
Figure 2.16. Network analysis of genes associated with excess <i>Bos taurus</i> ancestry in Borgou. Ingenuity Pathway Analysis (IPA) indicates known interactions that involve Nucleic Acid Metabolism, Small Molecule Biochemistry, and Infectious Disease.	59
Figure 2.17. Network analysis of genes associated with excess <i>Bos indicus</i> ancestry in Borgou. Ingenuity Pathway Analysis (IPA) indicates known interactions involved in Gene Expression, Infection Mechanism, Cardiovascular System Development and Function. .	60
Figure 3.1. The geographic origin of study populations.....	69
Figure 3.2. The physical positions of 768 SNPs across all the chromosomes chosen for the Goldengate assay.....	74
Figure 3.3. The allele frequency of taurus and zebu (primary xy axis), the corresposing delta values shown in the secondary y axis.	75
Figure 3.4 Neighbor joining phylogeny plot from the genetic distance matrix for 13 cattle populations.	78
Figure 3.5. Population structure inferred from clustering program STRUCTURE using K=2, 3 and 4..	80
Figure 3.6. Population structure in relation to the geographical distribution of the study populations.	81
Figure 3.7. Principal component analysis of 13 cattle breeds using 733 SNPs.....	82
Figure 3.8. Log likelihood curves for estimates of age of admixture among the hybrid populations..	84
Figure 3.9a. The 1 LOD interval estimate of age of admixture time for all hybrid cattle populations including a relative estimate for Sheko (SHK) derived from imported common genotypes from Hapmap data on BTA6.	85

Figure 3.9b. The 2 LOD interval estimate of age of admixture time for all hybrid cattle populations including a relative estimate for Sheko (SHK) derived from imported common genotypes from Hapmap data on BTA6.....	85
Figure 3.10 Haplotype block structure for each cattle breed based on BTA6.....	88
Figure 3.11. The Log r ² estimate value for pairwise LD measure for each cattle breeds. .	89
Figure 4.1. The distribution of major indigenous cattle breeds types in Ethiopia.	95
Figure 4.2. A map showing the location of study site.	98
Figure 4.3. An example of one herd in Selalle region showing Ethiopian zebu, Holstein-Friesian and their crossbreds under identical husbandry conditions (A).....	99
Figure 4.4 Photos of crossbred study population showing an intermediate taurus and zebu phenotypes.....	101
Figure 4.5. Animal testing for bovine tuberculosis based on intradermal tuberculin skin reaction using M bovis (PPD-B) and M avium (PPD-A) antigens (A) injected at two sites of the neck (B).....	101
Figure 4.6. Intradermal tuberculin testing result readouts after 72 hrs of injection. CIDT reactors showed induration at PPD-B site (A) and the non reactors showed no sign of skin thickness change for both injected sites (B).	101
Figure 4.7. Body condition score of the study samples	104
Figure 5.1. Allele frequencies within ancestral reference populations	117
Figure 5.3. The scatter plot of comparative intradermal tuberculin test (CIDT) readouts compared to the estimated <i>Bos taurus</i> ancestry for cases and controls	119
Figure 5.2. Box plot showing the genome wide average estimate of <i>Bos taurus</i> ancestry for cases (1) and controls (2).	120
Figure 5.4. The principal component analysis plot distributions of cases and controls...	120
Figure 5.5. Principal component analysis of cases and controls including the ancestral populations	121
Figure 5.6. ANCESTRYMAP risk model (x-axis) plotted Vs genomewide average LOD scores.....	122

Figure 5.7. The ANCESTRYMAP genomewide admixture scan for case-control analysis (z-score) and case only score (LOD) association of comparative intradermal tuberculin test (CIDT) positivity with ancestral origin.....	124
Figure 5.8. ADMIXMAP analysis using cross sectional and binary outcome of comparative intradermal tuberculin test (CIDT) positivity.	127
Figure 5.9. ADMIXMAP analysis using cross sectional and multiple outcome parameter accounting the binary parameters (case control) and comparative intradermal tuberculin skin test readouts as continuous variable	128
Figure 6.1. The number of animals counted for each coat colour category.....	146
Figure 6.2. Animal coat colour coding classified into eight groups.....	147
Figure 6.3. Admixture mapping of coat colour phenotype	150
Figure 6.4. Admixture mapping of combined quantitative dilution of coat colour phenotype highlighted for each chromosome (x-axis) and quantified based on Z-score (y-axis).	151
Figure 6.5. The significant admixture signal ($-\log P$ value =7.2) associated with red coat colour observed on chromosome 18 which has MC1R gene at the peak of the curve. ...	154
Figure 6.6. Coat colour QTL mapping using 1-LOD credible interval of ancestry associated regions on bovine chromosomes.....	156
Figure 6.7. Bovine QTL map for pigmentation traits for all autosomal chromosomes taken from the bovine QTL map viewer	157
Figure 6.8. Admixture mapping of body conformation trait (body condition)	158
Figure 6.9. Body condition QTL mapping using 1-LOD credible interval of ancestry associated regions on BTA2, BTA11 and BTA26 respectively.....	159
Figure 6.10. Admixture mapping of body conformation trait (dewlap development)	160
Figure 6.11. Admixture mapping of body conformation trait (hump development) for binary (1) and continuous outcome model (2).	161
Figure 6.12a. QTL regions of body conformation (dewlap development)	162
Figure 6.12b. QTL regions of body conformation (hump development) BTA2, BTA4, BTA10, BTA15, BTA19 and BTA20 using binary admixture model.	162

Figure 6.12c. QTL regions of body conformation (hump development) BTA1, BTA2, BTA6, BTA15, BTA18 and BTA19 using continuous variable admixture mode..... 162

Tables

Table 2.2a: List of Genes associated with Extreme outlier SNPs region based on excess African taurus ancestry estimate for Ethiopian zebu cattle..... 47

Table 2.2b: List of Genes associated with Extreme outlier SNPs region based on excess Zebu ancestry estimate for Ethiopian zebu cattle. 48

Table 2.3a. List of Genes associated with Extreme outlier SNPs region based on excess African taurus ancestry estimate for Borgou cattle..... 49

Table 2.3b. List of Genes associated with Extreme outlier SNPs region based on excess Zebu ancestry estimate for Borgou cattle..... 50

Table 3.1: Study populations and source of sampling locations..... 69

Table 3.2: The Fst matrix between cattle breeds..... 77

Table 4.1: Summary of base line information about 591 cases and controls study samples .. 104

Table 5.1: Summary report of ANCESTRYMAP scores for case only (LOD), case control z-score (CCS) and average LOD score for each chromosome (Chr_Local). 125

Table 5.2: The top markers based on either LOD or z-score values from ANCESTRYMAP analysis and the replication of the result using ADMIXMAP. 130

Table 5.3: Multiple logistic regression analysis of five SNPs showing the strongest genetic background dependant effects on CIDT positivity based on admixture mapping..... 132

Table 5.4: The list of genes extracted in the credible interval region of the five top chromosome regions..... 134

Table 6.1: The top admixture signals captured for genomewide correlations of phenotype traits 152

Abbreviations and Acronyms

SNP	Single nucleotide polymorphism
50k	Fifty thousand, Illumina 50k SNP chip
33k	Thirty three thousand, Bovine Hapmap consortium SNP chip
BTA	Chromosome, the bovine genome sequence assembly (Btau 4.0)
CIDT	Comparative intradermal tuberculin test
QTL	Quantitative trait loci
Mb	Mega base pairs
LD	Linkage disequilibrium
δ	Allele frequency difference, delta
AIM	Ancestry informative markers
K	Number of population clusters
IPA	Ingenuity pathway analysis
DAVID	The Database for Annotation, Visualization and Integrated Discovery
LOD	Logarithms (base 10) of odds
r^2	Correlation coefficient measure between pairs of loci
PPD	Purified protein derivative tuberculin
PPD-A	Purified protein derivative- avian
PPD-B	Purified protein derivative- bovine
BCG	Bacillus Calmette-Guérin (or Bacille Calmette-Guérin, BCG)
mtDNA	Mitochondrial DNA
GWAS	Genome wide association
MALD	Mapping admixture linkage disequilibrium
cM	Centimorgan
EM	Expectation-Maximization algorithm
PLINK	Open-source whole genome association analysis toolset
HWE	Hardy-Weinberg equilibrium

MCMC	Markov chain Monte Carlo
α	Admixture parameter, alpha
q	Admixture estimate
PCA	Principal component analysis
SD	Standard deviation
CI	Confidence interval
FDR	False discovery rate
UCSC	Genome Browser, Genome Bioinformatics Group of UC Santa Cruz
r	Recombination rate
LPS	Lipopolysaccharides
GO	Gene-Ontology
ng	Nanogram
NJ	Neighbor joining
Bp	Before present
BC	Before Christ
Km	Kilometer
μg	Microgram
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
mm	Millimeter
DTH	Delayed type hypersensitivity
BTB	Bovine tuberculosis
HMM	Hidden Markov model
OIE	World Organisation for Animal Health (Office Internationale des Epizooties)

Summary

Genome-wide SNP analyses of cattle populations have become practical. In parallel, advanced and sophisticated statistical packages make it practical to extract embedded genetic information; for example differential levels of admixture across the genome. In this study, a Bayesian based statistical approach implemented for admixture mapping analysis was used to examine Illumina 50k SNP genotype samples from the East (Ethiopian zebu) and West African (Borgou) hybrid zones. The analysis gave a quantitative measure of the degree of chromosomal mosaicism within the two hybrid zones, the dynamics of zebu genome introgression and identified regions with some evidence of the imprint of natural selection. 768 ancestry-informative SNPs from this high density bovine genotype data and other public resources were filtered for further work using a customised SNP assay.

Typing of these markers in 354 samples allowed a phylgeographic analysis with the added dimension of genome-wide analysis of admixture. The admixed populations from Africa and the Near East showed elevated linkage disequilibrium which correlated with visualised haplotype structure and also allowed plausible estimates of time depth of hybridisation.

The centre of the thesis is using an admixture mapping model also to highlight association of cattle genetic ancestry with host susceptibility to bovine tuberculosis. Ethiopian crossbreds of Ethiopian zebu (*Bos indicus*) and Holstein-Friesian (*Bos taurus*) are recently admixed and the parental populations have considerable breed disparities in bovine tuberculosis risk. These were genotyped with the customised genome-wide ancestry-informative SNP assay. The study implicated *Bos taurus* ancestral regions as associated with

bovine tuberculosis susceptibility. The most promising resides on BTA6 and includes a toll-like receptor gene cluster. Case only analysis supported this finding with additional significant scores on BTA11, BTA12 and BTA19.

Photographs taken on sampling the same hybrid samples allowed scoring for several additional phenotypes, animal coat colour, body condition and conformation, which were also investigated using the same genome-wide admixture mapping scan. This study identified regions on several bovine chromosome corresponding to previously mapped quantitative trait loci such as that for coat colour on BTA18 which contains MCR1, a major pigmentation gene responsible for red coat colour. The body condition trait also gave peaks that are supported by prior bovine quantitative trait mapping reports.

CHAPTER ONE: GENERAL INTRODUCTION

1.1 Introduction

1.1.1 Cattle origins

Domestic cattle, which have two extant species, *Bos taurus* and *Bos indicus*, descend from the extinct wild ox (*Bos primigenius*). The genetic evidence pertaining to the divergence of the two taxa, particularly from mtDNA sequencing, suggests separate sites were involved in their domestication (HIENDLEDER *et al.* 2008; LOFTUS *et al.* 1994a; LOFTUS *et al.* 1994b). *Bos taurus* has a probable origin in the Fertile Crescent and *Bos indicus* in the Indus valley. Their divergence time estimates vary from 330,000 years to 2 million years before present. In addition, molecular assessment of mtDNA sequence also gives clues to the possible routes of cattle expansion from domestication foci (ACHILLI *et al.* 2009; TROY *et al.* 2001).

Geographical peaks in nucleotide diversity within mtDNA studies confirm the probable centre of zebu domestication is resides within Indian subcontinent. *Bos indicus* expansion follows routes into southwest Asia, China, and Southeast Asia. The same approach has revealed the centre of *Bos taurus* domestication to be within Southwest Asia, from which major expansion follows routes into the European continent, Asia and perhaps Africa (AJMONE-MARSAN *et al.* 2010; CHEN *et al.* 2010b; MACHUGH *et al.* 1997).

Bos taurus has four major mtDNA haplotypes defined by specific nucleotide substitutions within the control region (T1,T2, and T3) of which T3 is the dominant type in European cattle (TROY *et al.* 2001). African cattle possess predominantly the T1 haplotype, a finding that supports somewhat a possible separate cattle domestication event in the continent which may date back to 9,000 years BP (BRADLEY *et al.* 1996; HANOTTE *et al.* 2002). An

alternate scenario is that the T1 haplotype and the equally geographically restricted East Asian predominant T4 haplotype are both products of the South west taurine domestication, followed by early migration and divergence of African and Eastern-Asia populations where a subset of the original diversity drifts to high frequency (ACHILLI *et al.* 2008; AJMONE-MARSAN *et al.* 2010).

1.1.2 African cattle

The existing data from genetics, pictorial representations and archaeology have indicated that the earliest domestic cattle in African continent were humpless, i.e. *Bos taurus* (BRADLEY *et al.* 1996). At present, Northern Africa and the tropical regions in the western and southern part of the continent have *Bos taurus* cattle types. A major zebu introduction through the Eastern Horn of Africa is a later event which and these cattle were favoured by their enhanced adaptation for a dry tropical climate and have achieved a wide distribution. There are additional advantages of *Bos indicus* cattle in that they have superior resistance to certain infectious diseases such as rinderpest – a major 19th century epidemic wiped out the majority of *Bos taurus* in affected areas while zebu population were relatively unaffected (HANOTTE *et al.* 2002; MURRAY and TRAIL 1984). Strikingly, the three African cattle types include African *Bos taurus*, sanga (a term usually used for East and southern African hybrids) and zebu have all taurine mtDNA. Zebu and sanga breeds have Y-chromosomes of mostly zebu origin (HANOTTE *et al.* 2000). This is consistent with male mediated zebu introduction in the African continent, in which introgression has taken place into herds of taurine origin.

A comprehensive phylogeographic analysis based on microsatellite allele diversity of African cattle breeds ascertained that the major recent introgression of zebu from Asia

was *via* the horn and the east coast of Africa. This event has followed by extensive zebu cattle expansion toward West Africa. It also illustrated the spread of pastoralism towards the South. The study also indicated a secondary *Bos taurus* influence from Europe and the near East (HANOTTE *et al.* 2002; MACHUGH *et al.* 1997).

1.1.3 The nature of cattle Admixture in Africa

Zebu genome introgression into African studied using mtDNA, autosomal and Y-chromosome data indicates that the population introduced into the continent has included crossing with and assimilation of earlier ingenious stock rather than its replacement. This was demonstrated by showing a continuum pattern of zebu genome decay from eastern to western and southern parts of Africa. There appear to be no zebu cattle in Africa that do not have some portion of taurine genetic influence (IBEAGHA-AWEMU *et al.* 2004).

Mitochondrial haplotypes assayed in west and east African zebu breeds have all been exclusively of *Bos taurus* types. Also East African animals of classically zebu morphology, have all exhibited reduced levels of zebu-specific microsatellite alleles relative to Indian cattle. This suggests taurine admixture throughout the nuclear genome. The Y-chromosome data showed convincing evidence that introgression was predominantly male-mediated (BRADLEY *et al.* 1998; HANOTTE *et al.* 2000; MACHUGH *et al.* 1997). The nature of cattle admixture in Africa is modified by a host adaptation to external factors. As mentioned earlier in this review, the impact of rinderpest in the lowland areas has contributed to a zebu phenotype expansion. Alternately, West African *Bos taurus* innate tolerance to the trypanosome parasite, which is the source of major cattle disease challenge along the central east-west belt of the continent, inhibits zebu expansion (BRADLEY *et al.* 1994; MATTIOLI *et al.* 2000). The present day cattle breed landscape in the

African continent exhibits a complex and dynamic admixture pattern influenced by adaptation to specific climatic and environmental factors.

Bos indicus genome introgression may have occurred in two waves. The first event is thought to have been of geographically limited impact with regard to most of the continent. The main focus for this introgression is most plausibly the Eastern and Northern part of Africa where the sanga breeds are thought to have been the outcome of a primary introgression. Currently these breeds are localized in the east and have expanded to the southern hemisphere of the continent. A second wave of zebu influx had a major, more widespread impact on zebu-taurus admixture. Hence, the distribution of *Bos indicus* introgression has a non uniform pattern. The region of strongest zebu influence is the East, the primary focus of introduction with Somalian, Ethiopian and Kenyan zebu breeds showing highest estimates of zebu genome content (HANOTTE *et al.* 2002; HANOTTE *et al.* 2000; MACHUGH *et al.* 1997). Interestingly, Malagasy zebu also shows an extreme value, probably related to an alternative origin given that island's unique links with Asia. Although here some *Bos taurus* breeds persisted in these regions, for example Sheko in Ethiopia, until recent decades, these are now clearly admixed. Zebu influence is also strong but decreasing in a transect to the south; Namibian, great lakes and South African breeds all retain substantial *Bos taurus* characteristics. The pattern toward West Africa is a dynamic one where the Sahelian belt has cattle of zebu morphology but with less indicine ancestry than to the East (ANDERUNG *et al.* 2007; FREEMAN *et al.* 2004). Breeds in southern, tropical regions of West and Central Africa are predominantly *Bos taurus* but with a gradient of indicine introgression; a process which continues to the present. North African cattle are predominantly *Bos taurus*. Clearly, *Bos indicus* genome introgression into Africa is not a replacement process but rather one that is layered

through time, continuing until the present. A relatively recent, but important process involves the input of European taurine cattle to the continent which are used in crossbreeding programmes to influence productivity.

1.1.4 Host genetic susceptibility to bovine tuberculosis

Bovine tuberculosis is a chronic disease caused by *Mycobacterium bovis*, which belongs to the *Mycobacterium tuberculosis* complex group. It is a major economic problem around the world and has a significant impact on cattle disease. It is also considered a potential zoonotic disease, posing a threat through cross infection to humans (BILAL *et al.* 2010; MICHEL *et al.* 2010; O'REILLY and DABORN 1995). The disease in cattle has many similarities to *Mycobacterium tuberculosis* infection in humans and there is significant phylogenetic proximity between the two mycobacterium species (BROSCH *et al.* 2000). As a result, a wealth of information on human tuberculosis may possibly be exploited for additional understanding of bovine tuberculosis. With human tuberculosis it has been showed in various studies that host genetics plays an important role. Evidence includes that from adoption studies, twin studies, genomewide linkage and population based case-control association analysis. Without ruling out the impact of socio-economic and environmental factors these have indicated that genetic factors in certain populations contribute to disease susceptibility (MOLLER *et al.* 2010).

A complex nature of genetic susceptibility to common infectious diseases including bovine tuberculosis is best explained by multiple genes targeted by weak selection, a dynamic process resulting in the stabilization of a particular trait. Field observation reports on the possible role of host variation in bovine tuberculosis have supported a genetic component (ALLEN *et al.* 2010; AMOS *et al.* 2010).

One type of relevant observation on bovine tuberculosis epidemiology is the suggestion that breed type contributes to disease distribution. According to a report from 1958 in Uganda, bovine tuberculosis incidence between 1931 and 1936 was 17% in Ankole cattle (native *Bos taurus*) while it was only 0.9% among zebu breeds (CARMICHAEL 1941; O'REILLY and DABORN 1995). More recently, a comparative Ethiopian study among European-origin *Bos taurus*, Ethiopian zebu (*Bos indicus*) and crossbreds has indicated that the disease prevalence was 22.2% in Holstein, 11.6% in zebu and 11.9% in crossbreds. Further analysis in the same study also showed that the degree of disease severity, ascertained by pathological examination, was significantly higher for *Bos taurus* than *Bos indicus* cattle (AMENI *et al.* 2007a; AMENI *et al.* 2006). One study on wildlife and livestock Deer in New-Zealand, which has a similar susceptibility pattern to domestic cattle (GRIFFIN *et al.* 2006) also gave evidence that on experimental low-dose inoculation of *M. bovis* bacterium the study subjects showed a variable pattern of host response which ranged from highly resistant to susceptible (GRIFFIN and MACKINTOSH 2000).

Heritability analyses, which estimate the extent to which phenotypic differences between animals are explained by genetic variation have been carried out. For example in south Siberian black and white cattle populations, a heritability of 0.06 -0.08 was reported for tuberculosis resistance. British dairy cattle have an estimated heritability for bovine tuberculosis resistance of 0.18 (BROTHERSTONE *et al.* 2010). An identical estimate of 0.18 was reported from observations in the Republic of Ireland. Interestingly, this latter study has also shown a strong positive genetic correlation between susceptibility to confirmed *M. bovis* infection and *M. bovis* purified protein derivate (PPD) responsiveness (BERMINGHAM *et al.* 2009). This implies that selection for animals resistant to PPD responsiveness will indirectly select resistant for tuberculosis infection. This also raises an important point

that accurate disease detection is essential to address the spectrum of host variation to disease susceptibility.

The most widely used animal testing tool for tuberculosis infection involves the use of comparative intradermal tuberculin in which an animal is injected with PPD and a measure is taken of dermal swelling because of a cell mediated immune response.

Although there are many limitations associated with single intradermal comparative tuberculin testing performance such as tuberculin doses variations, PPD preparations, site of application and interpretation schemes, intradermal comparative skin testing is still applicable for primarily screening (ALLEN *et al.* 2010; POLLOCK *et al.* 2001a; SCHILLER *et al.* 2010c). Variable findings on test sensitivity and specificity have been observed among independent studies from different regions. It has been suggested that, particularly in the cases with advanced disease stages, low sensitivity is possibly related to host anergy. Host immunosuppression which is related to stress is likely also a contributing factor to low tuberculin skin testing sensitivity (NEILL *et al.* 2001). Attempts to improve tuberculin skin testing by using additional tools have increased the sensitivity of disease identification; particularly the use of serological methods such as the interferon gamma (INF- γ) assay complement and enhance the basic test (SCHILLER *et al.* 2010a; SCHILLER *et al.* 2010c; WATERS *et al.* 2006).

1.1.5 Case- referent strategies for identification of susceptibility loci to bovine tuberculosis

Improved and accurate identification of the disease and phenotype definitions can greatly enhance the investigation of possible host genetic differences segregating between affected and control study populations. An additional, important factor is ensuring that

controls have had equal (or even greater) chance of exposure to pathogen as cases; this is difficult and yet crucial to correctly classifying subjects as having susceptible or resistant phenotypes (ALLEN *et al.* 2010).

Published investigations of bovine tuberculosis genetic susceptibility loci identification have been limited to a candidate gene testing strategy where a prior suspicion of biological involvement of a gene in the phenotype is a prerequisite. In this approach candidate genes that primarily interface with the host-pathogen immune system have been targeted. This includes the bovine MHC BoLA region which emerges from animal studies of variations in T cell responses to *M bovis* antigens between infected and healthy cattle (MEADE *et al.* 2007). Reports have also demonstrated that the dendritic cell-specific intracellular adhesion molecule (ICAM)-grabbing nonintergrin (DC-SIGN) receptor interacts and binds to *M bovis* Bacillus Calmette-Guérin (BCG) (GAGLIARDI *et al.* 2005). A similar finding has also been observed in human tuberculosis case-control association studies which suggest polymorphic variations in the DC-SIGN gene may be associated to tuberculosis resistance found sub-Saharan population (BURTON *et al.* 2007; SELVARAJ *et al.* 2009) but not in the Northern Africa (BEN-ALI *et al.* 2007). The NRAMP1 gene, which was first demonstrated in a mouse model to play a role in activation of macrophages during mycobacterial infections, has also been studied on cattle. The initial cattle study in Mexico and the USA found that infected animals did not have trait-associated genetic variation based on the NRAMP1 gene. This concluded a limited importance of NRAMP1 gene in susceptibility to infection with *M bovis* in cattle (BARTHEL *et al.* 2000).

In the light of the expected complexity in the genetics of infectious disease susceptibility, an alternate strategy which implements a genomewide association analysis (GWAS) is

merited. This approach uses genotyping of genetic polymorphisms throughout the genome without making assumptions about the genomic location of the causal variants. Unlike candidate gene approach, it eliminates the disadvantages of only a few polymorphisms are investigated (AMOS *et al.* 2010; ROSENBERG *et al.* 2010; WINKLER *et al.* 2010). The first human genome wide association study on tuberculosis susceptibility typed 333,754 SNPs in 2,237 cases and 3,122 controls from Gambia and Ghana which were diagnosed by medical history and clinical examination. This highlighted a new non-MHC susceptibility locus which was identified on chromosome 18q11.2 with a combined P-value of 6.8×10^{-9} and suggested that many further loci may be identifiable using GWAS in African populations (BOUMAN *et al.* 2006). However, there are major limitations to genome wide analysis. Firstly it has limited power to detect low-frequency variants with small effects on phenotype. As a result these kinds of studies in general typically require several thousands of case and control samples, and even with this sample size, it is clear that only a minority of the genetic component of complex traits is being accounted for by detectable loci (MANOLIO *et al.* 2009). Secondly, the analysis relies on using the LD between variants contributing to the phenotype and marker loci. Since LD decays quickly as distance between loci increases, a high marker density is also needed to ensure adequate power. In the case of cattle LD stretches over a greater distance than in humans which would likely reduce the marker density of SNPs needed. However, even with this the recommended number of markers for genomewide association analysis would be 800,000 SNPs for complete genome coverage and effective mapping (SMITH and O'BRIEN 2005; VILLA-ANGULO *et al.* 2009). Thus GWAS designs are extremely costly in that it requires high SNP density, large sample sizes and stringent statistical correction for massive multiple testing. A trade-off strategy requiring less resources but with high statistical

power to detect low to moderate effects would be an attractive alternative (AMOS *et al.* 2010; COOPER *et al.* 2008). Therefore, a lot of current attention is being given to mapping by admixture linkage disequilibrium (MALD) (WINKLER *et al.* 2010).

1.1.6 Admixture mapping analysis

Admixture refers to mixing of genomes of divergent parental taxa which can occur at the level of individuals and the level of populations. Admixture creates recombinants between divergent taxa where they come into geographic contact in hybrid zones (BUERKLE and LEXER 2008). A natural mapping strategy may be applied to study the pattern of gene flow across hybrid zones which can give insight into the genetic architecture of speciation. In effect hybrid zone mapping can highlight genetic regions that play a role in survival and reproduction - likely targets of selection (TEETER *et al.* 2008). Admixture mapping is an old concept but it is only in the late 80s that it regained new attention as a method using hybrid animals to exploit long range LD and limit the number of markers required for whole genome coverage.

As a basis, it is proposed that gene substitutions could cause complex phenotypes to evolve or differently distribute within specific geographic locations. Hybrids will have variable distributions of these variants and also phenotypes. Admixture mapping assesses linkage of a disease or trait with parental ancestry at each locus, defined as 0, 1, or 2 allele copies inherited from one ancestral population (MCKEIGUE 1998). Genomewide mapping using admixture has been shown to provide excellent statistical power, using less assumptions about transmission than linkage analysis and with higher sensitivity than genomewide association analysis. With only 1500-2500 ancestry-informative markers it is

possible to undertake a genomewide coverage analysis in humans and to resolve for small to moderate effects stretched within 10cM (WINKLER *et al.* 2010).

Admixture mapping typically takes advantage of the presence of a subset of genetic variants that significant frequency differences between parental strains, and these variants can be used to efficiently estimate local chromosomal ancestry. Recent admixture occurring between genetically distinct populations generates long-range linkage disequilibrium and hence distinct chromosomal segments originating from either ancestral populations that can be exploited to map disease genes and other important complex phenotypes that are more frequent in one ancestral population relative to the other (ZHU *et al.* 2008). A detailed outline of the design and analysis issues in admixture mapping studies are outlined by (HOGGART *et al.* 2004). These include, firstly, accounting for confounders as a result of population stratification which is overcome by either adjusting for parental admixture in a generalized linear model or by comparing observed locus ancestry with expected locus ancestry within each individual (MCKEIGUE 1998). Secondly, preset markers that show extreme allele frequency differentials between divergent populations qualify as informative markers. A lack of informative markers may be enhanced by the use of multipoint analysis that combine information from all markers on the chromosome to maximise information about ancestry at each locus (FALUSH *et al.* 2003). Thirdly, hidden values of ancestral specific allele frequencies may be inferred from both admixed and parental populations (HOGGART *et al.* 2004).

The models incorporated in admixture mapping methods are based on a hybrid of computationally intensive Bayesian and classical approaches. Three Bayesian programs, STRUCTURE (MONTANA and PRITCHARD 2004), ANCESTRYMAP (GREENSPAN *et al.*

2004) and ADMIXMAP (HOGGART *et al.* 2004) are the most commonly used packages that incorporate these computational methods for analysing admixture genotype data. The central element is the statistical model in which K subpopulations contribute to the gene pool of an admixed population, and variation of ancestry on each gamete is generated by K independent Poisson arrival processes, one for each subpopulation. The ratios between the intensities of these arrival processes are specified by the admixture proportion of the parent. The sum of intensities of the arrival processes can be interpreted as the average number of generations since admixture. This basic assumption also extends to include regression models for binary or quantitative traits. Two elegant computational innovations that have made feasible the analysis of large data sets with dense marker maps are the introduction of a Hidden Markov model forward-backward algorithm for sampling ancestry states on each chromosome (FALUSH *et al.* 2003) and a Rao-Blackwellized estimator that exploits this algorithm to calculate test statistics (GREENSPAN *et al.* 2004). The final test statistics from these programs estimate ancestral population frequencies across the genome which can be extended to use this information for association analysis (WINKLER *et al.* 2010).

1.1.7 Application of Admixture mapping analysis

Admixture mapping may be based on either a case-control or a case-only analysis to test for a relationship between disease risk and locus ancestry. In the former case the ancestry level for comparison is provided by an unaffected or random sample drawn from the same population. In the latter, the genomewide admixture level is estimated and used as a reference – this can be an advantage where negative phenotypes are difficult to diagnose, but the most secure approach is to combine both methods (MONTANA and HOGGART 2007; MUCHADEYI *et al.* 2007).

A number of genes have been discovered using admixture mapping analysis since 2005. Human genomewide admixture studies have been conducted successfully for a number of traits and diseases that occur at different rates in people of European, African or Latin American ancestry. These primarily focus on recently admixed North American populations such as African-American and Latinos, or less often, more ancient admixtures of Asian and African population. Reference populations may be assumed to be unadmixed European, African, Asian populations – each of which have very extensively genotyped samples included in the Hapmap database (<http://hapmap.ncbi.nlm.nih.gov/>) (SMITH and O'BRIEN 2005).

These studies illustrate the power of the approach. For example, a genomewide survey of African-American patients with chronic kidney/end stage kidney disease used 1331 markers and 891 cases and 1024 controls. This uncovered convincing evidence that the *MHY9* gene is involved in the higher predisposition to the disease conferred by African ancestry (KOPP *et al.* 2008). Similar successful studies, conducted on a comparable scale of other diseases related to increased risk due to African ancestry, include prostate cancer

(BOCK *et al.* 2009), Obesity (BASU *et al.* 2009; CHEN *et al.* 2010a), and hypertension (ZHU and COOPER 2007).

These reports exhibit some of the pre-requisites for successful admixture mapping. First, the trait must have a significantly different expression in one ancestral group relative to the other(s). Second, the genetic divergence between the contributing progenitor populations should be sufficient to enable strongly ancestry-diagnostic markers to be available throughout the genome. Third, admixture should be recent enough that sufficient linkage disequilibrium, or large enough chromosome portions of alternate ancestry, persist to make a genomewide screen using an affordable marker density feasible. The ~100,000 year divergence between Africans and Europeans, plus their ~ 400 years or 16 generation depth of admixture in the Americas, clearly satisfies these criteria. With cattle, the divergence between zebu and taurus exceeds any division within modern humans and with a time depth of greater than 200,000 years is more akin to that between humans and Neanderthals (GREEN *et al.* 2010; MURRAY *et al.* 2010). This ensures both a readily accessible supply of highly ancestry-informative SNPs and has resulted in many interesting traits with consistent and strong differences between the two cattle types. Many cattle populations in Africa, Asia, Australia and the Americas have a major component of zebu-taurus admixture in their origins, often providing different layers of time depths in the hybridisation processes. There are strong reasons to conclude that admixture mapping will be a valuable approach in the genetic dissection of important complex traits in cattle in the years ahead.

The importance of livestock resources in sub-Saharan Africa including Ethiopia is exceptionally high. Livestock productivity has been a prevailing issue and continues to pose problems in the continent (IBRAHIM and OLALOKU 2000). There are several factors

that contribute for the poor productivity of livestock in sub-Saharan Africa. The major ones include cattle diseases (for example bovine tuberculosis is endemic), poor nutrition, inefficient cattle husbandry systems, low genetic potential of indigenous cattle and a lack of comprehensive and multidimensional breed improvement programs for these populations (SHITAYE *et al.* 2007).

The dominant influence in African cattle genetics is that of crossbreeding between the two subspecies (HANOTTE *et al.* 2003). Examination of this phenomenon at molecular genetic precision can consequently make a contribution to production development.

African cattle genome analysis via a mapping admixture linkage disequilibrium (MALD) approach has the potential to accurately decipher the current population structure, to describe cattle admixture levels in sub-Saharan Africa and to refine estimates of the phenotypic effects of this hybridisation. This study may also outline non random ancestry correlates for bovine tuberculosis and other important phenotypes. The data could be extended to map out important genetic signatures of African cattle that can be used to conserve for selections methods in various breeding programs.

The multipurpose functions of livestock and the complex relationships between the biological, technical and social components of production require a systems approach, whereby animal breeding, nutrition, animal health, biotechnology knowhow, inputs and technologies are used to optimize resource use (KAASSCHIETER *et al.* 1992). Molecular population data will form a component of this, particularly in areas that are most affected by uncontrolled introduction of exotic breeds, continuous zebu introgression and weakening of indigenous African taurus adaptive values (KARUGIA *et al.* 2001; TAWAH *et*

al. 1997). Ideally, such information is shared among policy makers and farmers in the effort to alleviate the threat posed to native gene pools (HANOTTE *et al.* 2000).

A major aim of the work undertaken in this thesis is that through taking advantage of the extensive genetic divergence between *Bos taurus* and *Bos indicus*, one can make good inference about the possible ancestral correlates of various phenotypes and complex diseases. This will help to identify the genetic variants for the differential expression of complex traits, such as the tuberculosis susceptibility difference between *Bos taurus* and *Bos indicus*. The identification of DNA markers linked to such important traits could provide better tools for planned breeding and could be used to pre-select animals and increase genetic gain (SEIDEL 2010). Direct use of these data could directly benefit livestock productivity and progress to field applications in sub-Saharan Africa.

CHAPTER TWO: GENOMEWIDE ADMIXTURE ANALYSIS OF AFRICAN CATTLE POPULATIONS

2.1 Introduction

Modern day cattle consist of two major taxa, *Bos taurus* and *Bos indicus*, believed historically to have diverged more than 100 000 years ago as different forms of the wild ox, *Bos primigenius* (BISCARINI *et al.* 2010; LOFTUS *et al.* 1994b). Today more than 800 cattle breeds distributed all over the world have emerged from the first sites of domestication in the Near East for *Bos taurus* and south Asia for *Bos indicus* (BURT 2009). Some evidence has been presented suggesting Africa as another possible site of taurine domestication (BRADLEY *et al.* 1996), however, this has been weakened by subsequent data (ACHILLI *et al.* 2008). However, a markedly separate array of mitochondrial DNA (mtDNA) haplotypes on that continent suggests at the very least an early domestic divergence from the Near East and Europe.

There has been clear admixture of the two major taxa at different hybrid zones and this includes most of the African continent, particularly in the East (HANOTTE *et al.* 2000; MACHUGH *et al.* 1998). This, in concert with factors such as demographic bottlenecks, artificial selection and selective sweeps due to natural adaptation has shaped present day cattle populations in the region (GIBBS *et al.* 2009). One interesting aspect of investigation of hybrid zones is that it allows better understanding of cattle genetic structure and potentially may identify signals of selection for important traits.

2.1.2 Cattle genetic structure in Africa

The expansion of cattle across the African continent featured initially *Bos taurus* which were probably first introduced from the Near East possibly through Suez. Much later due

to trade and Arab movement, perhaps 2000-3000 years before present, the Horn of Africa served as a gateway for the introduction of zebu cattle into the region (HANOTTE *et al.* 2002). This introgression resulted in a profound diminishing of the *Bos taurus* genome at different levels throughout the African continent (figure 2.1).

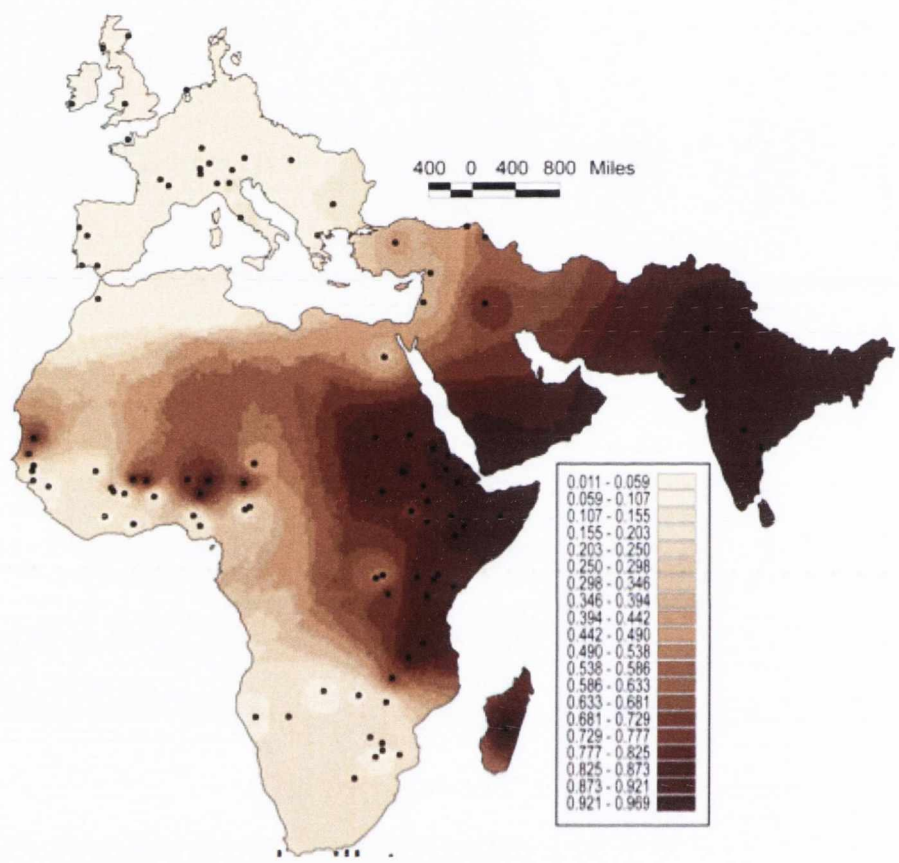


Figure 2.1 Contour map (FREEMAN *et al.* 2006a) for cattle genetic structure based on microsatellite allele frequencies; estimated *Bos taurus* ancestral proportion are plotted. The largest concentration of zebu ancestry was found in the Indian subcontinent, followed by East Africa and Madagascar. Little or no *Bos indicus* ancestry was encountered in Europe and West Africa.

Genome wide SNP data provide the current state of the art in population genetics analysis. These data have begun to emerge in cattle as a result of the bovine HapMap and parallel efforts (GIBBS *et al.* 2009; VAN TASSELL *et al.* 2008)

2.1.3 Admixture mapping

One opportunity in the study of hybrid zones is that of admixture mapping. This method uses Bayesian clustering and an EM algorithm to assign individuals into their respective cluster groups based on their frequency inferred using genotype data from either microsatellite or SNPs. Particularly, Bayesian clustering methods use prior information from modern parents to outline ancestral frequencies and predict the likelihood values for ancestral components within the test population in the genomic region under test. The prerequisite for this model is to choose informative markers, ideally showing high allele frequency difference between the contributing parental populations (PURCELL and SHAM 2004). This method takes advantage of single and neighbouring marker information to infer segments of chromosomal ancestry of each admixed individual. Consequently, one can visualize the mosaicism in ancestry along the genetic map. This in turn provides important information on the population structure and the nature of an admixture event (MONTANA and PRITCHARD 2004). In addition to illustrating the underlying genetic structure between and within the study populations, admixture mapping paves the way to look for possible regions under selections in a test population without having prior knowledge about any particular candidate region in the genome.

Admixture mapping is useful where the trait under study, (for example complex disease) involves two ancestral populations with divergent incidence of the traits. Among affected admixed individuals the trait frequency will be associated with inheritance of alleles at trait

genes from one of these populations. For example, near the loci predisposing to a disease affected individuals will have a higher probability of having inherited their chromosomes from the population in which the risk allele are more frequent. Admixture mapping will detect the signal generated from the affected individual taking into account controls and the general population structure (MONTANA and HOGGART 2007).

The application of admixture mapping has been well explored in human populations using a number of phenotypes and has delivered solid evidence, both reconfirming previously described genetic regions and also giving novel associations. This application of admixture mapping has involved finding complex traits associated with risk alleles in admixtures such as in African American populations (CHENG *et al.* 2009; ELBEIN *et al.* 2009; KOPP *et al.* 2008; ZHANG *et al.* 2008), and Latin American populations (ALDRICH *et al.* 2008; MAO *et al.* 2007; PRICE *et al.* 2007).

Analysis of Genetic admixture is more effective when it reflects migration and mating between individuals from genetically highly distinct parental populations (DARVASI and SHIFMAN 2005). The primary ancestral contributors to the current African cattle population are African *taurus*, European *taurus* and Asian *indicus*.

Previous analysis of African *Bos taurine* and *Bos indicus* cattle populations (DAYO *et al.* 2009; FREEMAN *et al.* 2004) suggest a strong genetic contribution through the level of admixture to disease resistance for many tropical infections like rinderpest and trypanosomiasis plus adaptation to harsh environmental conditions such as arid zones. Mapping these genes could give evidence on the molecular mechanisms behind various phenotypes differentiating between zebu and *taurus*.

The hypothesis behind this test is the assumption that African cattle were originally *Bos taurus* type (BRADLEY *et al.* 1996) and adapted to infectious diseases and other challenges on the continent over several thousand years before the influx of zebu animals. It is also known that zebu animals show adaptation such as survival in hot environments and withstanding extended drought conditions.

In this chapter an attempt was made to test within African cattle genomes whether the effects of zebu admixture throughout the genome are even or alternatively whether selection has acted to accelerate or retard zebu introgression at specific loci that have a bearing on local biological challenges.

2.2 Materials and Methods

2.2.1 Study Population and samples

Whole blood samples were collected from unrelated individual animals from Ethiopia. In this chapter Ethiopian zebu (ETZ) were screened. These breeds came from the highland central part of the country and are locally described as Arsi breeds. The second set of samples were from previously collected DNA material from West Africa, that include N'dama (NDA) and Somba (SOM) classified as pure African turines and Borgou (BOR), West African hybrid. Indian zebu were used as a parent population for *Bos indicus*, namely Tharparker (TRP) and Hariana (HAR). The third group were taken from the European Holstein-Friesian breed (HOL). DNA extraction was conducted using the standard protocol using an *ArchivePure* DNA purifying commercial kit. And the genotyping assay for 50K SNPs was performed using the Illumina platform using the commercial service provided by AROS Applied Biotechnology AS Ltd.

2.2.2 Markers

54000 markers were tested for consistency in quality control. All markers of unknown chromosomes were removed from the list and for computational convenience and in the absence of complete sex information, all autosomal chromosomes were analysed separately from X chromosome. Using the PLINK <http://pngu.mgh.harvard.edu/purcell/plink/> whole genome analysis tool, SNPs were removed and classified as low quality genotype data in the event of more than 10% missing data, low call rate, less than 1% minor allele frequency or failed Hardy-Weinberg equilibrium test (overall HWE, $P < 0.001$). The third criterion was to remove markers that

are tightly linked in parental populations which might interfere with the subsequent admixture linkage model analysis. After allele frequency, genotype and linkage disequilibrium (LD) based pruning 39671 SNPs were recovered for further analysis.

2.2.3 Information content of the markers

A test of informativeness was calculated from the allele frequency data of each marker. The parental populations were used to calculate the absolute value allele frequency difference (delta, δ) using the formula (ROSENBERG *et al.* 2003).

$$\Delta_i = \frac{|f_{i1} - f_{i2}|}{2}$$

, f_{i1} and f_{i2} are the i th allele frequencies of the two parental populations, *Bos indicus* and *Bos taurus*.

2.2.4 Admixture analysis

A linkage model using *STRUCTURE* Version2.3 (FALUSH *et al.* 2003) was applied to estimate allelic, individual and population level of admixture at different hybrid zones for unphased genotype for all autosomal informative markers with delta value at least 40%. The consistency of the program was evaluated at different numbers of burn-in and iterations going up to 100,000 for each case. The final run chosen for subsequent analysis consisted of a 10,000 burn-in period and additional 10,000 MCMC iterations after burn-in and was repeated three times. The program was run without assigning any prior information about the population. The admixture parameter (α) was inferred and a correlated allele frequency model was chosen. The remaining parameters in the program were assigned to default settings.

The optimal number of clusters (k) was examined from the Ln-likelihood outputs collected after each *STRUCTURE* run tested for $K=2,3,4$, and 5. A reasonable assumption was taken by combining Ln-likelihood estimates and the general population history of cattle.

Furthermore, primary stratification of the population using distance measures was done using *EIGNESTRAT* version 2.0 and the *Smartpca* package of the program was used to construct eigenvectors (CASTELNUOVO *et al.* 2006). The top PCA-correlated SNPs obtained based on a simple clustering algorithm implemented in the package were used to construct a second set of informative SNPs under population structure.

Time of admixture was estimated using the *HAPMIX* version 1.0 program which has reported suitability in dealing with older admixture events. Ethiopian zebu and Borgou samples were used as a test population. Phased genotype (using *fastPHASE* version 1.2 (BEARDEN *et al.* 2006) with African *taurus* and Indian zebu assigned as parental populations were used to predict ancestral chromosome segment blocks for the test population and time of admixture (PRICE *et al.* 2009).

2.2.5 Statistical Analysis

The *STRUCTURE* output for the ancestry estimate (q) using a linkage model was used to measure the amount of ancestral proportion for each marker for individual animals using genome wide data. An excess ancestry estimate for each locus was calculated for the test population by taking the difference from the average genome wide ancestry estimate.

The ratio of excess ancestry estimate to its standard deviation gave a z-score which was used to scale differences and identify markers with outlying ancestry estimates (ZHU *et al.* 2005).

$$Z_i = \frac{1}{N \Sigma} (q_i - M_i) \quad \begin{array}{l} q_i = \text{locus specific ancestry, } M_i = \text{genome wide ancestry,} \\ \text{SD} \quad \text{SD} = \text{standard deviation of locus specific ancestry} \end{array}$$

The nominal *P*-value for each specific SNP was calculated from the z-score. To avoid false positives and spurious assumptions, in this analysis the threshold score for significance were set to have at least z-score = ± 4.89 i.e. $P = 10^{-6}$. False discovery rate (FDR) testing was carried out from *P*-values and the corresponding q-value using the R package (qvalue, <http://cran.r-project.org/web/packages/qvalue/>). A significant score for an admixed population relative to ancestral group would indicate the presence alleles derived from one particular ancestral population more than expected (MONTANA and PRITCHARD 2004) and indicate regions affected by selection through admixture linkage disequilibrium with and advantageous variation between ancestral groups.

Top score values were used as references to identify significantly outlying positions and the pattern of scores in regions around the maximum were used to define a list of candidate genes in the region. The gene content of the outlying regions was based on their physical position accessed from the UCSC Genome Browser <http://genome.ucsc.edu/> for the bovine genome. The genes within statistically significant outlier SNP regions for either excess *Bos taurus* or *Bos indicus* ancestry for Ethiopian zebu and Borgou breeds were accessed for functional clustering according to Gene Ontology terms implemented in the DAVID bioinformatics web based tool (HUANG DA *et al.* 2009). In order to get

integrated information on the most significant functional list of genes and their possible interaction within the list and other important genes from public domain databases, functional and network analyses were performed with the Ingenuity Pathway Analysis (IPA) software v8.0 <http://www.ingenuity.com/>.

All other summary statistical calculations were performed using the R and SYSTAT software.

2.3 RESULTS

2.3.1 Informative markers

A total of 156 individual animals from seven breeds genotyped for diallelic SNPs were used for this study. Routine quality control testing was conducted for each marker. From a total of 52329 SNPs, based on call rate, very low parental minor allele frequency, high LD values ($r^2 > 0.8$) markers for parental population, more than 20% of genotypes were filtered out. The remaining 39674 markers for 156 individuals were used for further analysis.

Ideally informative markers are considered to have alleles present at high frequency in one of the parental populations and at very low frequency in the second population. An attempt was made here to select markers showing extreme allele frequency difference (δ) between parental populations (ROSENBERG *et al.* 2005).

In the study it was designed that the African *Bos taurus* parental population were represented by the west African N'Dama (NDA) and Somba (SOM) while Holstein (HOL) represented ancestry of European *Bos taurus*. The *Bos indicus* parental group was represented by the Indian zebu, Tharaparker (THR) and Haryana (HAR).

The delta value distribution between African *Bos taurus* Vs *Bos indicus*, and European *Bos taurus* Vs *Bos indicus* was plotted and shown in figure 2.2. It indicates that the markers range from non informative (0) to highly informative (close to 1). In this panel of markers 300 SNPS were highly informative with delta values of more than 0.9. A threshold delta value for assigning SNPs for ancestry informative markers (AIM) were set at least 0.40

between cattle breeds of African *Bos taurus* and *Bos indicus* or European *Bos taurus* and *Bos indicus*. A total 9150 markers were selected that were distributed uniformly across all autosomal chromosomes. The overall average delta value between *Bos taurus* and *Bos indicus* was 0.53 and the maximum $\delta = 1.00$.

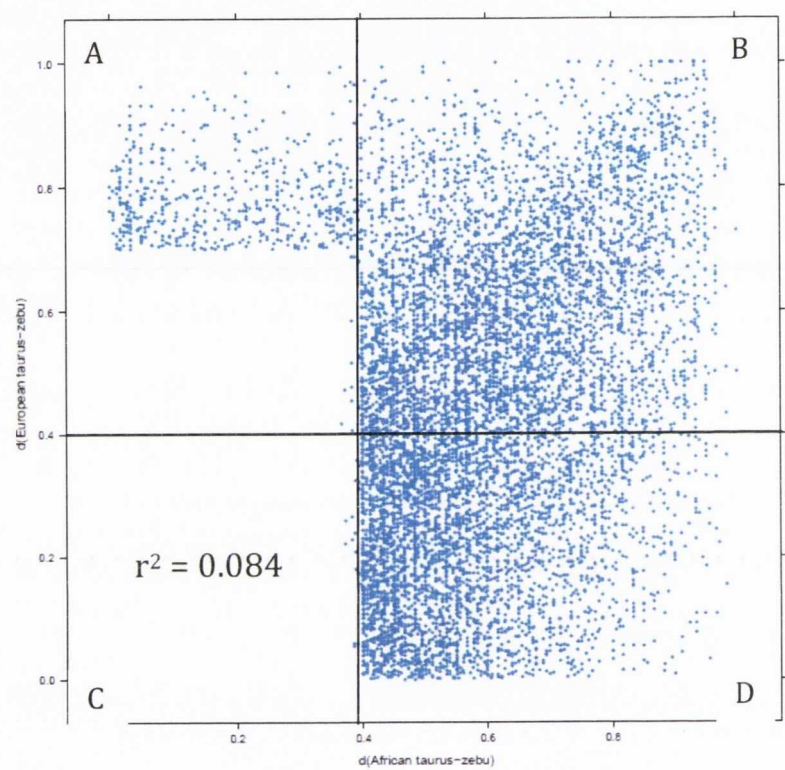


Figure 2.2 Dot distribution showing allele frequency difference (δ) of 9150 ancestry informative markers. δ values of selected 9150 SNPs are at least 0.40 for African *Bos taurus* Vs *Bos indicus* (x-axis) or 0.70 for European *Bos taurus* Vs *Bos indicus* (y-axis). The distribution of markers can be divided into four quadrants, markers highly informative for both African *Bos taurus* Vs *Bos indicus* and European *Bos taurus* Vs *Bos indicus* (B), markers highly informative for African *Bos taurus* Vs *Bos indicus* (D), markers highly informative for European *Bos taurus* Vs *Bos indicus* (A).

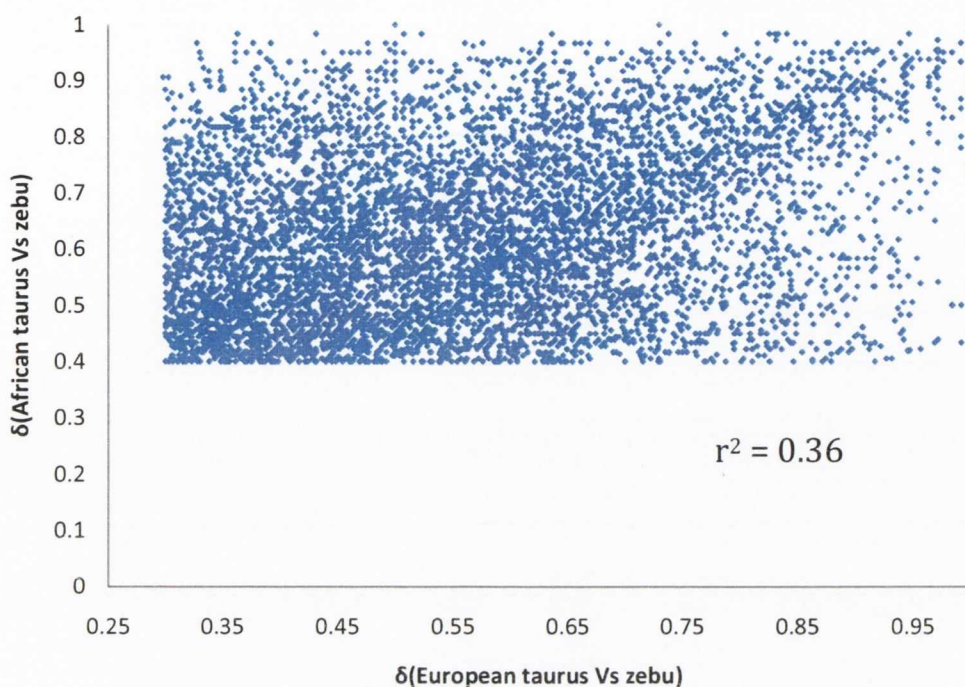


Figure 2.3 Dot distribution showing allele frequency difference (δ) of 7340 ancestry informative markers. δ values of selected top PCA correlated 7340 snps selected for comparative analysis. δ_{ez} (European *Bos taurus* Vs *Bos indicus*), δ_{az} (African *Bos taurus* Vs *Bos indicus*). SNPs δ value in this panel is above 0.40 for African *Bos taurus* Vs *Bos indicus* (y-axis) and above 0.30 in the case of European *Bos taurus* Vs *Bos indicus* (x-axis).

Marker choice was in proportion to the total length of each chromosome and was checked for maintaining a minimum of 40kb inter marker distance between neighbouring positions. Further evaluation of the markers was achieved through comparison of the corresponding delta values derived using the European and African taurus breeds. Taking all taurus together, it was possible to get informative markers to examine the possible structure between the African and European taurus. A positive correlation was shown for both delta values of African and European taurus vs zebu (figure 2.3). The mean minor allele frequency for *Bos taurus* was 0.275 and 0.1 for *Bos indicus*. It was believed that these markers are not tightly linked to have a background noise effect on admixture mapping programs used throughout the analysis.

2.3.2 Genetic distance

An initial analysis using all 39674 SNPs was used to estimate the pairwise genetic distance among all the populations used. Pairwise Fst measures indicate high values between the Indian zebu and both African and European *taurus*. The highest pairwise differentiation was noted between NDA and TRP with a value of 0.455 followed by 0.44 for NDA and HAR. The West African hybrid cattle population (BOR) showed intermediate Fst values with all the breeds but was closer to the West African taurine population, particularly to Somba (SOM) is *Bos taurus* cattle from the adjacent regions. On the other hand, the Ethiopian zebu is closer to the Indian zebu and its smallest pairwise Fst value was noted Vs. BOR with a value of 0.082. On average the Ethiopian zebu has 20% Fst value with all taurine populations while the value for Borgou is 12% (Table2.1).

Table 2.1: Pairwise Fst between populations, values based on 39674 SNPs.

	HOL	NDA	SOM	BOR	ETZ	TRP
NDA	0.243					
SOM	0.223	0.065				
BOR	0.185	0.105	0.070			
ETZ	0.232	0.259	0.230	0.082		
TRP	0.386	0.455	0.425	0.252	0.122	
HAR	0.372	0.440	0.410	0.233	0.099	0.058

2.3.3 Principal component analysis

Consistent observations emerged when individuals were plotted following principal component analysis. PC1 = 47.33% and PC2 = 25.03% were shown on the figure 2.4. The first and the second eigenvector clearly illustrated the genetic structure among all breeds and clustered them into five groups. Individuals belonging to the same breed were clustered in the same group with no significant outlier animals noted.

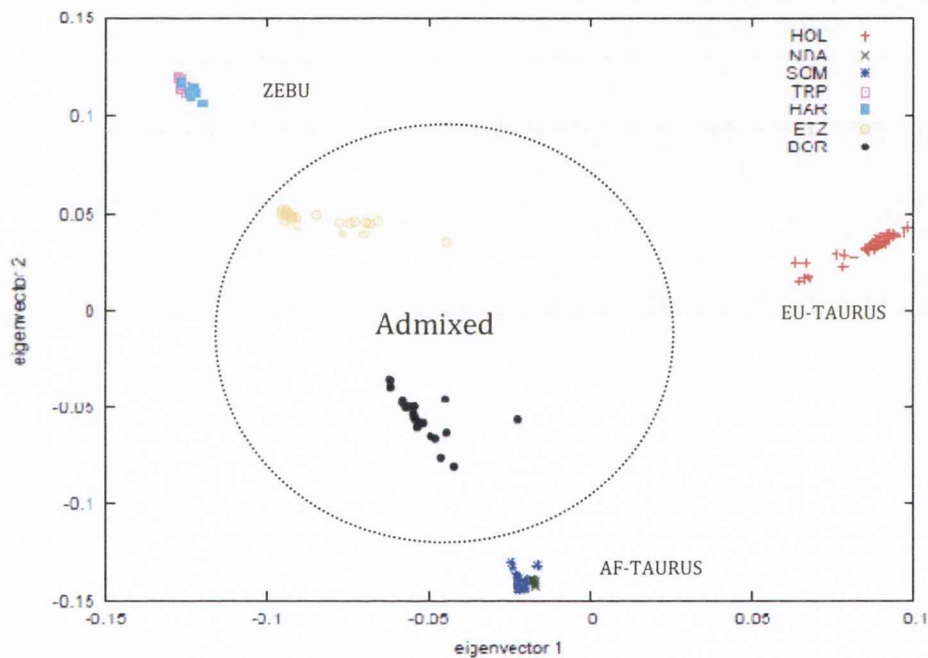


Figure 2.4 Principal component analysis representation of 7 populations of 156 cattle using all 39674 SNPs. The first and the second eigenvectors shows 47.3% and 25 % variation in total plotted on PCA showing individual and population variation among the study population. The dotted circle showed admixed populations of Ethiopian zebu (ETZ) and Borgou (BOR).

The principal component analysis showed that a more positive value on the first axis indicates a taurine cluster. The distribution of breeds in the African continent has concurs with a pattern from previous observations that, as we go from the east to the west in the continent the taurine genome becomes more dominant. The Ethiopian zebu showed a residual taurine genome and shared affinity with the West African breeds as the individuals stretched towards the Borgou. The overall observation suggests both Ethiopian zebu and Borgou represent admixed population as both breeds positioned intermediate between *Bos indicus* and *Bos taurus* animals. The second eigenvector further elaborates the divergence between African and European *Bos taurus* breeds (figure 2.4). The same data was used to construct a second PCA (PC1 = 56.75 % and PC2 = 7.48%) plot after removing Holstein-Friesian (HOL) with the assumption of these being an out group *Bos taurus* population. This further illustrates the intermediate, hence admixed nature of the Borgou and Ethiopian zebu breeds as shown in figure 2.5. It also gave support to the notion that the two ancestral populations contributing to both of these correspond to African *Bos taurus* and *Bos indicus*.

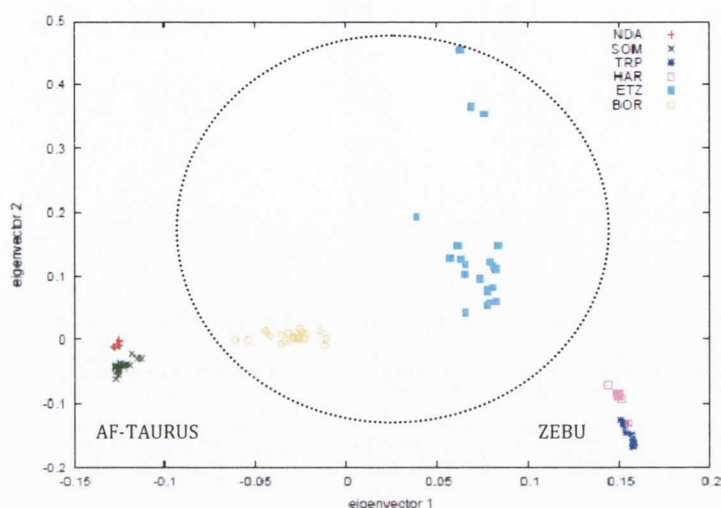


Figure 2.5 Principal component analysis representation of 6 populations, after removing Holstein (HOL) from the analysis. The first and the second eigenvectors shows 56.8% and 7.5 % variation in total plotted on PCA showing individual variations and population structure of African cattle with *Bos indicus*. The dotted circle showed admixed populations of Ethiopian zebu (ETZ) and Borgou (BOR).

2.3.4 Population Genetic Structure of Ethiopian zebu and Borgou

Population structure analysis using a Bayesian cluster analysis (*STRUCTURE*) shows the Asian, African and European ancestral content within breeds. The program was tested assuming two clusters ($k=2$) through to five clusters ($k=5$) and the Ln likelihood probability ratio for the optimal cluster was plotted for each assigned cluster. Based on Ln likelihood probability prediction as shown in figure 2.6, it was found three clusters ($K=3$) might be optimal, accounting for the substructure among African *Bos taurus*, European *Bos taurus* and *Bos indicus*. This effectively mirrors the result obtained from PCA. The plot in figure 5 shows that three clusters illustrate the continental populations. Hence $K=3$ can be assumed to usefully infer the population structure and $K=2$ was a logical reduction whenever the European *Bos taurus* (HOL) was removed the analysis.

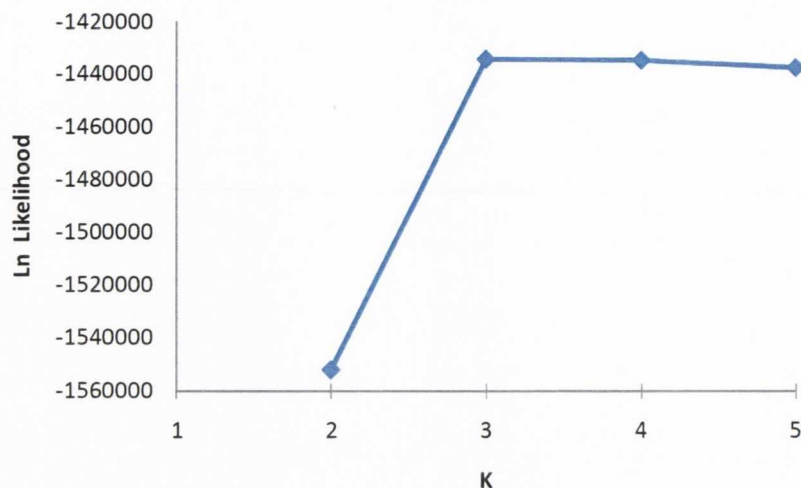
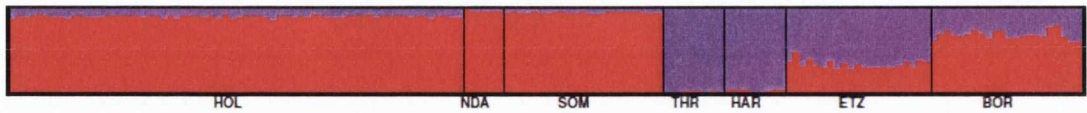
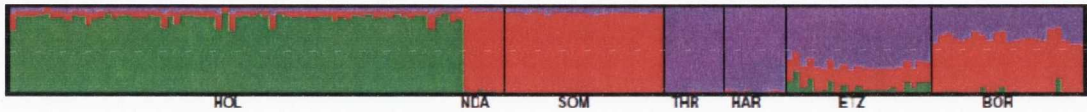


Figure 2.6 Estimated Ln likelihood values for each number of population clusters (K) tested using 9150 ancestry informative markers. The optimal population clusters inferred would be K=3 that correlates with European *Bos taurus*, African *Bos taurus* and *Bos indicus*.

In concordance with the PCA stratification result, the genetic admixture event for Ethiopian zebu and Borgou is well illustrated using *STRUCTURE* at both the population and individual levels of analysis as shown in figure 2.7. The first analysis at K=2, indicated the distinction between *Bos indicus* and *Bos taurus* where both African and European *taurus* clustered together. African and European *taurus* separate all the way from K=3 to K=5 and substructure within HOL becomes evident at K=4.



K=2



K=3

Figure 2.7 STRUCTURE clustering of seven breeds into two when K=2 used (top) showing the general subdivision into main taxa (*Bos taurus* and *Bos indicus*) and admixed population Ethiopian zebu (ETZ) and Borgou (BOR) have shared composition from the two groups. When k=3 (bottom) is applied, the substructure within *Bos taurus* cattle into European *Bos taurus* Holstein (HOL) and African *Bos taurus* Somba (SOM) and N'Dama (NDA).

2.3.5 Age of admixture

The estimate of admixture times of Ethiopian zebu (ETZ) and Borgou (BOR) populations were inferred from the *HAPMIX* program and STRUCTURE analysis. *HAPMIX* program is designed to utilize genetic information based on reconstructed haplotypes and estimate the age of admixture times of zebu and taurus in each test population. The program is particularly useful for in cases of ancient admixture events, it resolves by inferring small, ancient, ancestry segments of admixed population under study (PRICE *et al.* 2009). The log likelihood curve was plotted at different generation time as shown in Figure 2.8. For ETZ and BOR. Based on 39674 data, the 2LOD interval by interpolation from the highest log likelihood was used as optimal estimate of age of admixture time for ETZ and BOR.

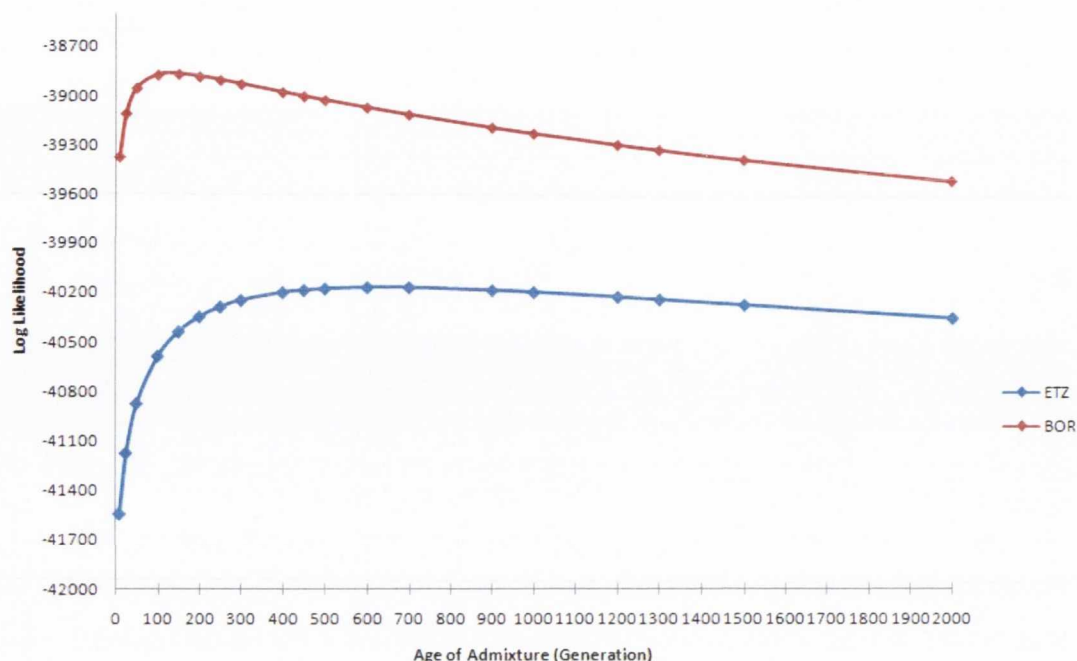


Figure 2.8 Hapmix estimation of likelihood (y-axis) of age of admixture (number of generation) on the x-axis for Ethiopian zebu (ETZ) in blue and Borgou (BOR) in red.

The 2 LOD interval estimate of generation time for ETZ was (626 – 710). Assuming 5 years per generation for cattle, admixture event is estimated to have happened in the range of 3130 to 3550 years ago. While the prediction for BOR showed a much younger admixture event with 2 LOD interval of (130 – 155) generation ago. Under this assumption age of admixture for BOR is in the range of 650 to 775 years ago.

The age of admixture estimate by STRUCTURE analysis using all 39K SNPs data was also used to compare the prediction made by *HAPMIX*. The STRUCTURE analysis performed under a linkage model plots the posterior distribution of recombination parameter (r), which could be converted to effective number of generation since the first admixture event (FALUSH *et al.* 2007). The mean r for ETZ cattle is 2.08 with 90% confidence interval (CI) of (2.06 – 2.11). In this model the age of admixture happened

about 208 (206 – 211) generation ago, or 1042 (1035 – 1055) years ago. Following similar model assumption for BOR, predict the mean r value of 1.34, and the 90% confidence interval (CI) is (1.32 – 1.35). STRUCTURE estimate of age of admixture time for BOR cattle is 134 (132 – 135) generation ago, or 668 (662 – 677) years ago. The difference between HAPMIX and STRUCTURE estimates could be probably due to the different algorithms implemented in both programs. In this aspect for example, STRUCTURE which the prediction heavily depends on the accuracy of recombination breakpoints estimates per genetic distance. Hence, the density of markers used influence the output in case of ancient admixture events. This could be the potential attribute for the observed discrepancy between the two models, particularly in the case of ETZ where the prediction of HAPMIX and STRUCTURE is very far apart. Another possibility for this variation could be the nature of zebu introgression in African continent which, in addition to its ancient occurrence is not a singular, linear process making the prediction far more complex.

2.3.6 Admixture mapping of African cattle hybrid zones

Genome wide marker specific evidence of excess ancestry was examined from the level of admixture estimate (q) for individual and marker specific measures under linkage model. Here the values generate posterior probability admixture estimates which in turn are used to measure whether portions of chromosome derive from a distinct background of either unusually high or low *Bos taurus/indicus* ancestry. The distribution of q for each chromosome using 9150 markers for Ethiopian zebu (ETZ) shown in figure 2.9a and similar distribution for Borgou is shown in figure 2.9b.

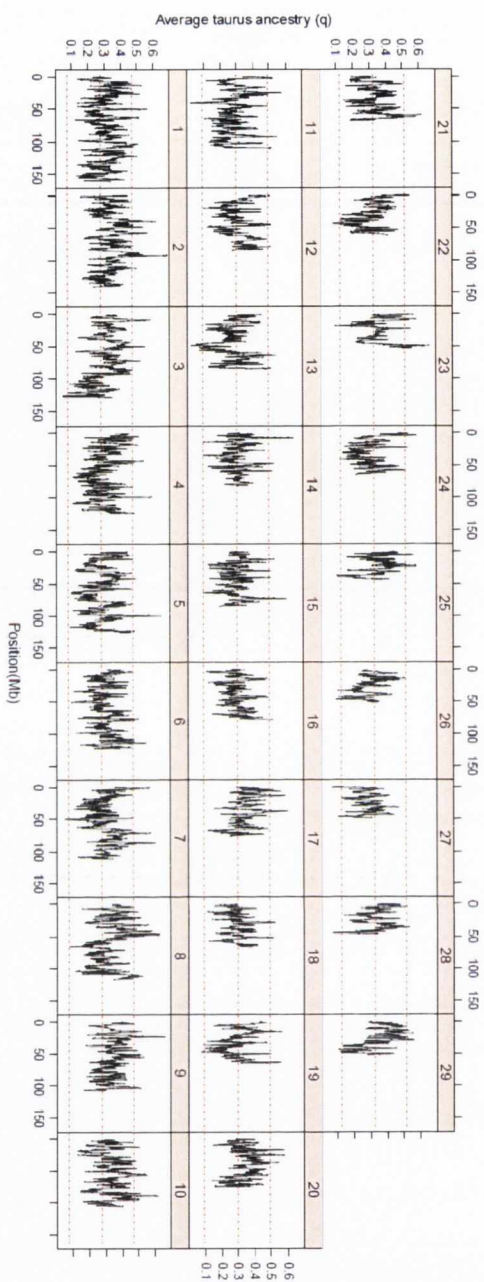


Figure 2.9a Genomewide ancestry estimate (q) in Ethiopian zebu. The estimate measured using ancestry informative markers, the middle red line for each chromosome box represent the genomewide average of *Bos taurus* ancestry. The degree of excess taurus ancestry variation is indicated by the upper and lower boundaries (top and bottom red lines).

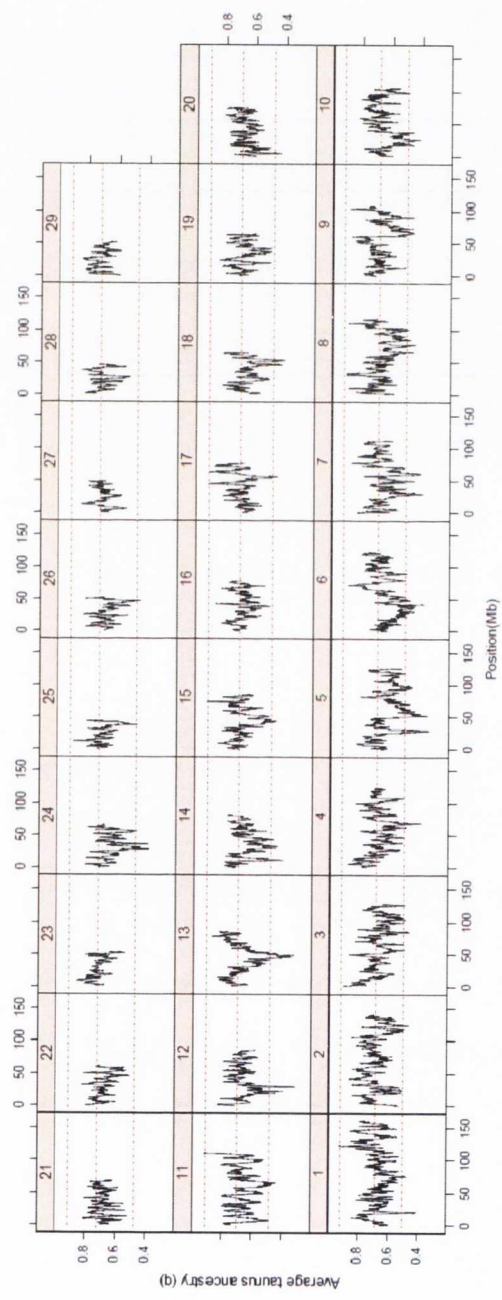


Figure 2.9b Genome-wide ancestry estimate (q) in Borgou. The estimate measured using ancestry informative markers, the middle red line for each chromosome box represent the genome-wide average of *Bos taurus* ancestry. The degree of excess taurus ancestry variation is indicated by the upper and lower boundaries (top and bottom red lines).

The overall mean European *Bos taurus* ancestry clustering level for Ethiopian zebu (ETZ) and Borgou (BOR) was 8.6% and 2.4% respectively. As expected for ETZ, *Bos indicus* ancestral estimates were highest, with a genome wide average of 68.4 % (65.55 -73.78, 95% CI). The average ancestral contribution of African *taurus* was 31.6% (28.75 – 36.9, 95% CI). The European *taurus* ancestry estimate was low with a genomewide mean value of 8.6%. The West African hybrid zone cattle population Borgou (BOR) ancestry estimates, unlike the East African hybrid zone gave African *Bos taurus* ancestry at 66.6% (62.94 -73.4, 95% CI) and a zebu ancestral clustering estimate of 33.3 % (29.74 - 40.2, 95%CI). Similar to the Ethiopian zebu, the proportion of European *taurus* influence is low at 10%. A pearson correlation matrix was constructed from genome wide ancestry estimates (q) for each marker to test the European *Bos taurus*, African *Bos taurus*, and *Bos indicus* relationships among the ancestral estimates. It was observed a high negative correlation for both ETZ and BOR ($r^2=-0.862$, -0.86) between the ancestral estimates values for African *taurus* and *indicus* population. The correlation was weak (ETZ $r^2=-0.47$, BOR -0.38) between European *Bos taurus* and *Bos indicus*. Similarly the coefficient were small (ETZ $r^2=-0.043$, BOR $r^2=-0.16$) for the ancestry estimates between African *taurus* and European *taurus*. Hence, in the tri-hybrid model, these findings further ascertain that the contribution of European *taurus* was not significant enough for independent treatment. Subsequently local ancestral estimates using STRUCTURE were carried out by removing the Holstein (HOL) population and assigning a two way admixture model from the contributing parental population (West African *Bos taurus* and *Bos indicus*). The admixture levels in two way admixture are q and $(1- q)$, therefore only the q distribution for African *Bos taurus* ancestry is used to illustrate the pattern in figure 9. Estimates of

genomewide excess ancestral scores for each individual marker relative to the two putative continental backgrounds of either African *Bos taurus* or *Bos indicus* are shown on broken lines and the middle broken lines approximates the genome wide average values of q .

An examination of Ethiopian zebu ancestral estimates for each locus across each chromosome reveals points of extreme departure from the genomewide mean value.

Clearly there are outlier regions showing very low *Bos indicus* accompanied by a corresponding very high *Bos taurus* ancestry and the distributions that crosses the upper and lower boundaries in the q distribution plot indicate these extreme markers.

Particularly visible outliers spots were observed in regions on chromosome 2, 8, 9, 10, 14, and 23 could be cited as high African *Bos taurus* ancestry chromosomes for ETZ. While excess *Bos indicus* ancestry spots can be observed along chromosome 7, 11 and 13.

Likewise, similar analysis for Borgou indicated a high excess ancestry was corresponds to a sudden depression in the *taurus* ancestry estimate as shown in the q boundary limits in figure 8. The highest excess African taurus ancestry can be found in chromosomes 1, 11, 15, and 17. Spots of regions showing excess *Bos indicus* ancestry are found in chromosomes 5, 6, 12 and 13.

These regions were further analysed in detail to assess if they might harbour some genes of biological interest that might render differential advantage for host survival and adaptation.

2.3.7 Genome mapping of specific regions of Ethiopian zebu cattle under selection

Further systematic analysis was carried out to identify the outliers more clearly and measure the significance of the signals measured at each marker. The observed P value distribution for each of the two possible ancestral populations as shown in figure 2.10 has indicated the points with extreme deviation from the standardized distribution.

The focus of the study was to reveal the *taurus* outliers in Ethiopian zebu breeds, this distribution skewed to the left and clearly showed the significant outliers in the distribution. An attempt was made to fit the African *taurus* ancestry estimate to a normal distribution through Q-Q plot from the z-score values (Figure 2.11). The QQ plot points out the outliers from the normal distribution and the fit was very poor. This observation provides further evidence that the significant outliers have influenced the fit to the normal distribution.

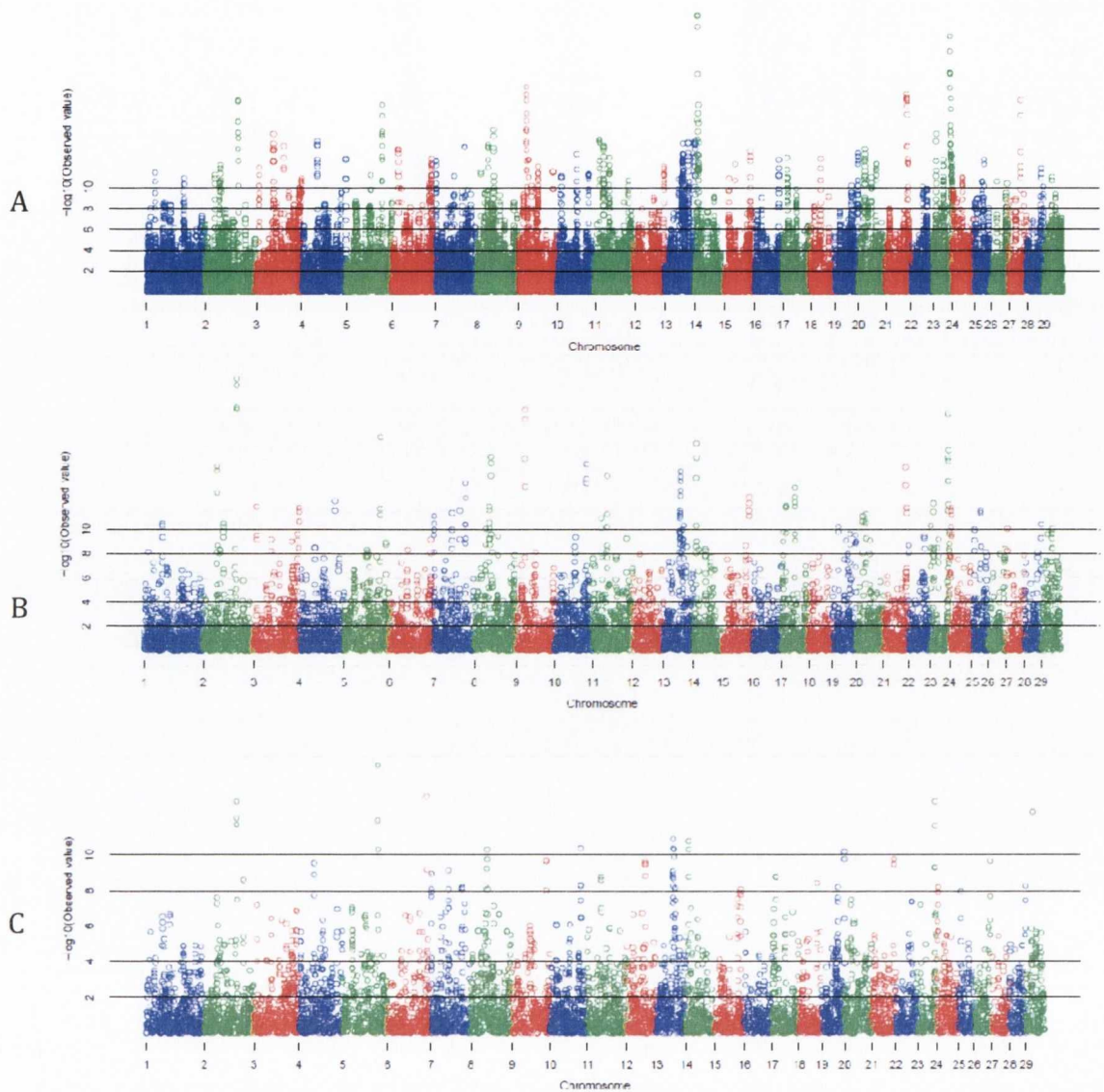


Figure 2.10 Manhattan plot of $-\log_{10}(P \text{ values})$ by chromosomal location of Ethiopian zebu (ETZ) cattle using 39674 SNPs (A) and using only ancestry informative (AIM) 9150 markers (B). A second independent SNP panel was formed from top PCA correlated markers that able to differentiate individual cattle into their clusters of origin. A total of 7340 markers were used as top PCA correlated panel and signal reproducibility compared with the results from AIM (C). The observed $-\log_{10} P$ value calculated each SNP indicate the excess ancestry at each chromosome position 1 to 29.

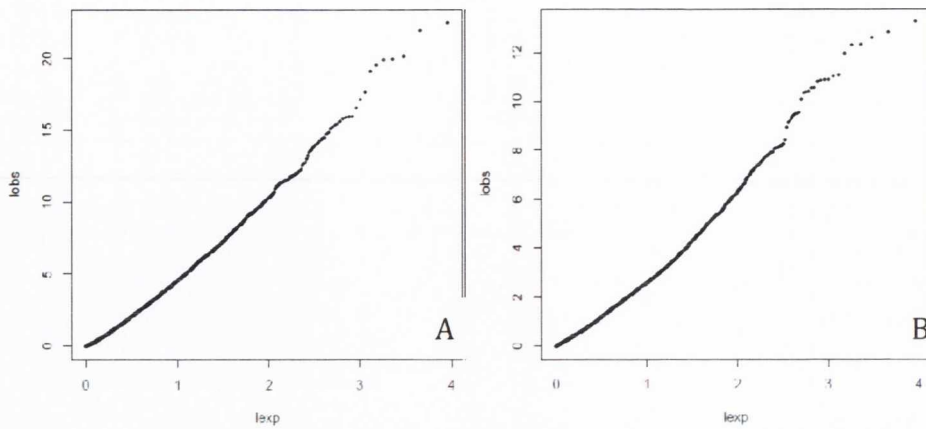


Figure 2.11 Ethiopian zebu (A) and Borgou (B) normal Q-Q plot for African *Bos taurus* ancestry estimate from $-\log P$ values using 9150 ancestry informative markers.

The whole genome admixture scan for ETZ adaptively differentiated loci with excess west African *Bos taurus* ancestry are summarized in table 2.2a with at least a z -score of 4.89 (P -value = 10^{-6}) and simultaneously above the threshold of FDR (<0.01) testing given by q -value. In total 311 (3.4%) of the 9150 ancestry informative markers displayed above this level of significance. The regions where these SNPs fall may display evidence of selection.

It is also interesting to note that the peak point contain more than one SNP and not just one outlier, ascertaining the confidence to make a final conclusion about a particular genetic region. Additional independent PCA-correlated 7340 SNPs panels also replicate the findings, suggesting those outlier genetic regions contain non-randomly high proportion of excess *Bos taurus* ancestry.

Table 2.2a: List of Genes associated with Extreme outlier SNPs region based on excess African taurus ancestry estimate for Ethiopian zebu cattle.

	Refere2nce top SNP	Position(bp)	z-score	P-value	q-value	Gene around the peak region	
1	BTA-113021-no-rs	49925040	6.64	3.23E-11	1.21E-09	ALCAM	
2	Hapmap55130-rs29021614	40579534	8.09	6.11E-16	1.10E-13	ACVR1,ACVR1C,CYTIP,ERMN,GALNT5	
2	Hapmap33966-BES2_Contig368_774	92419324	9.92	3.25E-23	1.05E-19	C2H2orf47,SPATS2L,KCTD18,SGOL2,AOX1,CLK1,NIF3L1,ORC2L,NDUFB3,CFLAR,CASP8,ACBD6,ALS2CR12,STRADB,ALS2CR11,ALS2CR4	
3	BTA-66943-no-rs	8678702	7.10	1.24E-12	8.72E-11	UAP1,SH2D18,C3H1orf226,C3H1orf111,NOS1AP,OLFM2B,ATF6,FCRLB,FCGR2,FCGR3A,FCGR3AFCGR2C,MGC128008,SHHC,MPZ,PVR4,ARHGAP30,USF1,PCP4L1,NRI13,TO MM40L,AP0A2,FCER1G,NDUF52,ADAMTS4,B4GALT3,PP0X,USP21,JFC1,DEDD,NTT1,PFDC2,KLHDC9,PVR14,F11R1TLN2,CD244,LYS,SLAMF7,CD48,SLAMF1,NGSTN,GOPA,PEX 19,DCAR,PEA15,CASO1,ATP1A4	
5	Hapmap34380-BES4_Contig293_1272	99378353	8.74	2.30E-18	9.30E-16	REGL,LM03,MGST1,DERA,COX6B1,STRAP,EP58	
7	ARS-BFGL-NGS-53233	1270450	6.85	7.50E-12	3.28E-10	CNOT6,GFPT2,MAPK9,RASGEF1C,RNF130,MIR340,TBC1D9B,C7H5orf45,SQSTM1,MGAT4B,LT4S,HNRNP11	
7	ARS-BFGL-NGS-101879	86846000	7.69	1.46E-14	1.63E-12	COX7C	
8	ARS-BFGL-NGS-5793	41457554	7.09	1.32E-12	9.08E-11	GLDC,UHRF2,IL33,MLANA,CD274,CBH9ORF46,INSL4,RCL1,AK3,CD37L1,PTAPDC2,SLC1A1	
8	BTB-01479628	45993522	8.30	1.08E-16	3.10E-14	DMRT2,DMRT1,DOCK8,TSPAN3,PGM5,FGM5	
9	BTA-26796-no-rs	23197139	9.31	1.26E-20	8.15E-18	UBE2CBP,PGM3,ME1	
10	UA-IFASA-7857	88848591	8.13	4.18E-16	7.95E-14	NPC2,ISCA2,LTBP2,LOC539472,RCF1,PROX2,DLST,RPS6KL1,PGF,EIF2B2,ACYP1,TMED10,FOS,DP2	
11	BTB-01903909	24235545	6.79	1.15E-11	4.71E-10	SLC8A1	
14	Hapmap30097-BTC-007678	7969429	8.61	7.16E-18	2.57E-15	ST3GAL1,NDRC1,TG,SLA,PHF20L1,LRRC6,KCNO3,OC90,EFR3A	
17	ARS-BFGL-BAC-31505	12905532	7.13	1.02E-12	7.33E-11	EDNRA,TTTC29,TTTC29,RNF146,P0U4F2,SLC10A7,LSM6	
17	Hapmap33069-BTA-161042	37376453	7.59	3.29E-14	3.22E-12	IL2IL21,CETN4,BBS12,FGF2,NUDT6,TRPC3	
19	ARS-BFGL-NGS-77575	15671997	6.54	6.16E-11	2.08E-09	CCT68,TMEM132E,CCL8,CCL11,CCL2,ACCN1	
19	BTB-01580865	64194808	6.31	2.79E-10	7.22E-09	WIPI1,ARSG,SLC16A6,AMTZ,RGS9,CCDC46,AP0H,PRKCA	
20	Hapmap57270-ss46526311	18062068	6.84	8.13E-12	3.46E-10	IP011,DJMT1L,KIF2A	
20	Hapmap53329-rs29023196	26144289	6.91	4.84E-12	2.24E-10	IL31RA,DDX4,SLC38A9,PPAP2A,SKIV2L2,MIR449C,MIR449B,MIR449A,GPX8,GZMA,gzma,ESM1,SNX18,HSPB3,ARL15	
21	ARS-BFGL-NGS-35461	59061456	8.05	8.10E-16	1.38E-13	CHGAT,TPK1,MGC137894,C21H14orf142,UBR7,PRIMA1,ASE2,MGC152353,OTUB2,DDX24,LOC514011,JSG12(B),IFI27,IFI27L2,SERPINA10,SERP1	
23	ARS-BFGL-BAC-29517	7841613	7.17	7.52E-13	5.66E-11	NA1,SERPINA11,SERPINA12,SERPINA5,SERPINA3,SERPINA3-2,SERPINA3-1,KRTCAP2,SERPINA3-6,SERPINA3- KLHL31,GCLC,DSB,BOLA-DYA,BOLA-DOB,TAP2,TAP2,PSMB8,TAP1,PSMB9,H2B,HLA-DMB,HLA-	
23	BTB-01175105	49345959	9.22	2.99E-20	1.61E-17	DMA,BRD2,COL11A2,RXB8,SLC39A7,HSD17B8,RING1,VP552,RPS48,T4,WDR46,PFN6,RGL2,TAPBP,DAXX,KIFC1,PHF1,CUTA,ZBTB9,BAK1 EEF1E1,MUTED,DSF,SSR11Y86,F13A1,NRN1,FARS2,LYRM4,RP40,CDYL	
24	ARS-BFGL-NGS-44005	4166030	7.05	1.73E-12	1.07E-10	CNDP1,CNDP2,CYB5A,C17H18orf55,FBXO15	
25	ARS-BFGL-NGS-105470	5941852	6.46	1.03E-10	3.14E-09	FOX1	
25	ARS-BFGL-NGS-29099	22935781	6.97	3.10E-12	1.62E-10	SCNN1B,COG7,GGA2,DCTN5,PLK1,CHP2,PRKCB,CACNG3	
29	Hapmap40781-BTA-65234	29796747	6.49	8.37E-11	2.68E-09	TMEM225,PANK3,TBRG1,SMAE,SPA17,NRGN,YSIG2,ESAM,ROBO4,HEPACAM,SLC37A2,TMEM218	

Table 2.2b: List of Genes associated with Extreme outlier SNPs region based on excess Zebu ancestry estimate for Ethiopian zebu cattle.

BT A	Reference top SNP	Position(bp)	z-score	P-value	q-value	Gene around the peak region
3	ARS-BFGL-NGS-6917	126589836	-6.95	3.59E-12	1.84E-10	ACXT, PPPIR7, SEPT2, STK25, BOK, THAP4, ATG4B, DTYMK, ING5, D2HGDH, PDCD1, CCDC37
7	ARS-BFGL-NGS-43299	50349478	-6.93	4.34E-12	2.10E-10	PRD12, WNT8A, NME5, KIF20A, CDC23, GPR43, CDC25C, FAM53C, REEP2, EGRI, ETP1, HSPA9, CTH, CTNNA1, LARTM2, SIL1, MATT3, PAIP2, PACAP, DNAJC18, ESCR, TMEM173, UBED2, CXKC5, IGF, C7H5orf32, PPONI1, HBEGR, SLC4A9, SRA1, APBB3, SLC35A4, CD14, NDUPA2, JK, WDR35, DNLI, HARS, HARS2, ZMAT2, SIL1
11	ARS-BFGL-NGS-11128	40729583	-7.87	3.48E-15	5.12E-13	EMI6, RTN4, RTN4, MGC139228, RPS27A, MTIF2, MGC148942, CCDC104, SMEK2, PNPT1, EFEMP1, MIR217, MIR216A, MIR216B, CCDC85A
13	BT A-115847-no-rs	48693198	-7.95	1.82E-15	2.80E-13	C9H12orf49, KLF6, PEKP, ADARB2, IDI1, GTPBP4, ZMYND11, PRNP, PRND, RASSF2, C13H20orf30, PCNA, CDS2, PROKR2, C13H20orf196, CHGB, TRMT6, MCM8, CRLS1
19	BT A-45668-no-rs	46696680	-6.34	2.27E-10	6.17E-09	NMT1, PLCD3, ACBD4, HEXIM1, HEXIM2, MAP3K14, ARHGAP27, RDM1, LYZL6, GOSR2, ARF4, CRHR1, MAPT
25	ARS-BFGL-NGS-41383	37326781	-5.79	7.00E-09	1.08E-07	YWHAQ, SRCRB4D, ZP3, DTX2, UPK3B, UPK3BL, POLR2J, ALKBH4, ORA12, PRRKIP1, EMID2, RABL5, FIS1, CLDN15
25	ARS-BFGL-NGS-32440	39309762	-5.64	1.69E-08	2.20E-07	BR13, CCZ1, RSPH10B, AIMP2, EIF2AK1, ANKRD61
28	ARS-BFGL-NGS-61191	44288288	-6.60	4.22E-11	1.55E-09	ERCC6, C28H10orf53, OGDHL, PARC, TIMM23, NCOA4, MSMB, MARCH8, ALOX5, ZNF22, RASSF4, C28H10ORF10

Table 2.3a. List of Genes associated with Extreme outlier SNPs region based on excess African taurus ancestry estimate for Borgou cattle.

BTA	Reference top SNP	Position(bp)	z-score	P- value	q-value	Gene around the peak
1	Hapmap33760-BTA-161469	122056533	5.39	6.91E-08	5.30E-06	HLTF, GYG1, CPA3, CPB1, AGTR1
11	Hapmap32029-BTA-127208	109959239	5.07	4.04E-07	2.21E-05	MRPS2, C11H9ORF116, OLFM1, MRPL41, ARRDCL1, EHMT1, CACNA1B
15	ARS-BFGL-NGS-40765	72978573	5.09	3.62E-07	2.03E-05	DGCR6, API5, MIR670, MIR129-2
17	BTB-01220555	45083414	5.05	4.36E-07	2.30E-05	GUCY1B3, TDO2, GUCY1B3, GUCY1A3

Table 2.3b. List of Genes associated with Extreme outlier SNPs region based on excess Zebu ancestry estimate for Borgou cattle.

BT A	Reference top SNP	Position(bp)	z-score	P-value	q-value	Genes
1	BT A-39438-no-rs	20513155	-5.43	5.74E-08	4.55E-06	MIR125B-2,
4	ARS-BFGL-NGS-20949	68200162	-5.83	5.62E-09	9.31E-07	NEUROD6,
5	Hapmap41950-BTA-72999	28774242	-6.81	9.52E-12	5.66E-09	TMCC3, NDUFA12,
5	Hapmap41813-BTA-27442	52747835	-6.50	8.21E-11	2.17E-08	IRAK3, TM6IM4,
6	Hapmap33430-BTC-037618	41300911	-6.16	7.51E-10	1.43E-07	SULT2, MIR218-1,
7	Hapmap57685-rs29016545	30070268	-5.80	6.71E-09	1.03E-06	PPIC, SNX24, SNX2,
7	Hapmap50476-BTA-79543	60350545	-5.51	3.63E-08	3.32E-06	FBXO38, HTRA,
10	BTB-00417947	27028539	-5.15	2.66E-07	1.60E-05	NDRG2, METT11D1,
11	Hapmap40862-BTA-100125	63983539	-4.96	7.15E-07	3.40E-05	TMEM17, MDH1,
12	Hapmap40891-BTA-115072	23105299	-5.77	7.91E-09	1.11E-06	NHLRC3, STOML3,
12	ARS-BFGL-NGS-76809	29014825	-7.52	5.33E-14	2.53E-10	NABP2L1, ZAR1L,
13	BTB-00527671	48392938	-7.23	4.67E-13	4.67E-10	LOC515358,
14	ARS-BFGL-NGS-104836	11377913	-5.64	1.68E-08	1.86E-06	MYC
14	Hapmap39744-BTA-33631	32573153	-4.96	7.08E-07	3.40E-05	PREX2, PRDM14,
18	ARS-BFGL-NGS-54266	45993979	-5.24	1.62E-07	1.03E-05	SCN1B, HPN, FXVD3,
18	UA-IFASA-7471	53666364	-5.38	7.33E-08	5.45E-06	RABAC1, CD79A,
20	Hapmap42635-BTA-68718	4745147	-4.92	8.74E-07	3.96E-05	STK10, UBTID2,
24	Hapmap57214-rs29016076	35667068	-5.68	1.32E-08	1.57E-06	MIR1-2, MIR133A-2,

A more focused plot around the peak region of $-\log_{10} P$ value integrated with UCSC reference genes around and near the significant point for each chromosome and selected plot is shown in table 2.2a&b. One notable result with excess *Bos taurus* ancestry for ETZ is a region on BTA23 which contains BoLA gene clusters, namely *BoLA-DYA*, *BoLA-DYB* and the human orthologue of *HLA-DMB*, *HLA-BMA* (appendix A). These clusters are categorized among non-classical classII bovine MHC genes located at BTA23q12-13. This plays an important role in host adaptive immune response in relation to antigen processing and presentation of peptides or polysaccharides via MHC classII. Further functional relationship was investigated using Ingenuity Pathway Analysis (IPA) in which HLA and all other genes with excess taurus ancestry were queried for possible interactions within the group. Among 25 molecular networks retrieved, the network that contains HLA genes is one of the top hit ($-\log_{10} P = 22$) and it contains 15 genes from the list as shown figure 2.12. This network elucidates general function interactions that are involved in antigen presentation, cell-to-cell signaling and interaction, haematological system development and function.

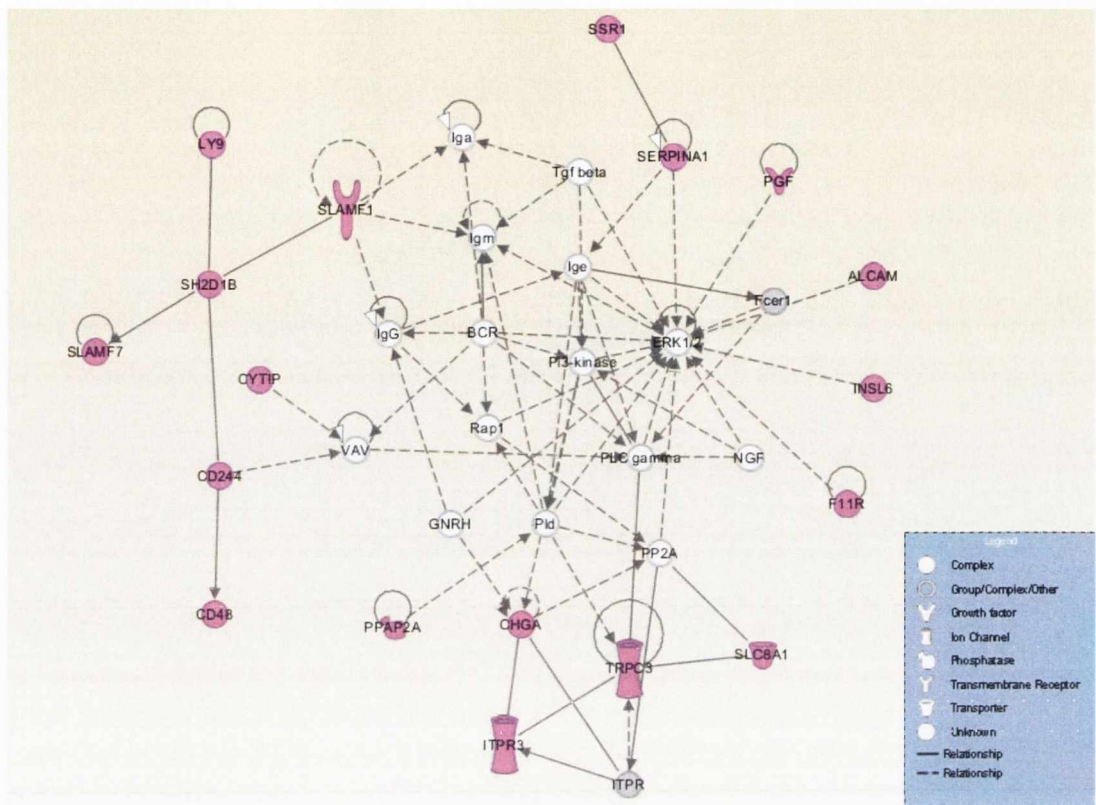


Figure 2.13. Network analysis of genes associated with excess *Bos taurus* ancestry in Ethiopian zebu. Ingenuity Pathway Analysis (IPA) indicates known interactions that involve in cell death, cell-to-cell signaling and interaction, cell signaling. Genes in colour are the subsets of the list that has all significant excess taurus ancestry (Table 2.2a).

Another list of genes includes CD274 found in BTA8 which is involved as a co-stimulatory signal essential for T-cell proliferation and production of *IL-10* and *INF-G*.

Under the same region important cytokine, *IL-33* which binds to a signal through *IL-1R* is also located. Clusters of another cytokine groups of *IL-2* and *IL-21* are also in close proximity of less than 1Mb from the peak marker in BTA17 with *P* value 3.29×10^{-14} (3.22×10^{-12}). These are important cytokines produced by T-cells in response to

antigenic/mitogenic stimulation and also have immunoregulatory function. In addition to immune related genes, cell signalling molecules that have role during infection (*SLAMF1*, *SLAMF7*, and *ALCAM*, figure 2.13), apoptotic genes such as *BAK1*, *CFLAR*, *ACVR1*, *CASP8*, *DEDD*, pigmentation related gene *MUTED* are examples found in excess taurus ancestry regions in Ethiopian zebu study samples.

A similar strategy of whole genome admixture scan analysis showed 203 (2.2%) SNPs with excess *Bos indicus* ancestry ($z\text{-score} \leq -4.89$, $P < 10^{-6}$) against the comparison to genomewide average value. The highest score was recorded for BTA13 and BTA11, for peak markers BTA-115847-no-rs ($P = 1.82 \times 10^{-15}$ ($q = 2.80 \times 10^{-13}$)), ARS-BFGL-NGS-11128 ($P = 3.48 \times 10^{-15}$ (5.12×10^{-13})) respectively. A complete account of all significant excess *Bos indicus* ancestry regions are listed in table 3b. However, one interesting gene, CD14 on BTA7 is harboured within the highest peak in the region ($P = 4.34 \times 10^{-12}$ ($q = 2.1 \times 10^{-10}$)). The protein encoded by this gene is heavily involved in mediation of innate immune response to bacterial lipopolysaccharides (LPS). IPA network analysis shows possible gene interactions and indicates the top network ($-\log_{10} P = 44$) shown in figure 12 which retrieves 23 genes from the list including the CD14 molecule. This gene could be a potential candidate region involved in selection within the East African hybrid zone for its functional relevance for host immune response.

2.3.8 Genome mapping of specific regions of Borgou cattle under selection

A similar strategy was applied for Borgou (BOR) samples and significant outlier regions that possibly are under selection for the West African cattle hybrid zone were screened (figure 2.15). The general depiction of the observed $-\log_{10} P$ values of ancestry estimates against the expected corresponding values presented in the Q-Q plot in figure 2.11 shows skewed outliers at the top of the tail of the distribution. Plots around those peaks regions showing significant excess *Bos taurus* or *Bos indicus* ancestry and the corresponding UCSC bovine reference genes are mapped in table 2.3a&b.

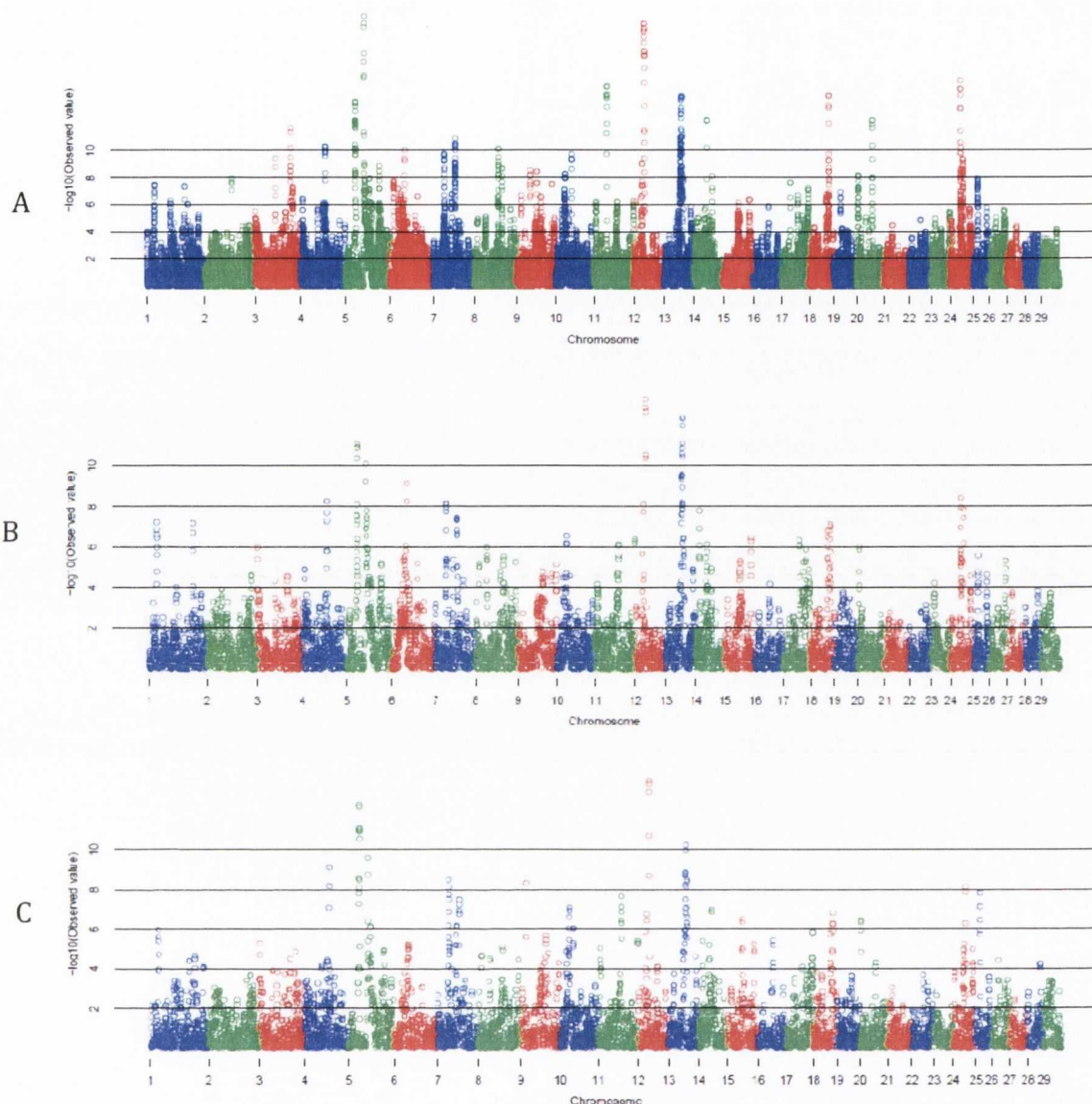


Figure 2.15. Manhattan plot of $-\log_{10} (P \text{ values})$ by chromosomal location of Borgou (BOR) cattle using 39674 SNPs (A) and using only ancestry informative (AIM) 9150 markers (B). A second independent SNP panel was formed from top PCA correlated markers that able to different individual cattle into their clusters of origin. A total of 7340 markers were used as top PCA correlated panel and signal reproducibility compared with the results from AIM (C). The observed $-\log_{10} P$ value calculated each SNP indicate the excess ancestry at each chromosome position 1 to 29.

The top excess African taurus ancestry for BOR in total 10 (0.1%) SNPs in total with minimum P value of $\leq 10^{-6}$. The highest score observed is on BTA1 with P value 6.9×10^{-8} ($q=5.3 \times 10^{-6}$) followed by BTA15 with P value of 3.62×10^{-7} (2×10^{-5}). A list of genes around the top reference SNPs is tabulated in table 2.3a. The genes in this list are functionally very broad. Clustering this group into gene ontology GO terms using the DAVID bioinformatics web tool showed an enrichment score for GO-terms such as intercellular signalling, blood circulation, vascular smooth muscle contraction, protein dimerization activity. The list of genes belonging to these GO-terms are *AGTR1*, *GUCY1B3* and *BUCY1A3*. Ingenuity Pathway Analysis (IPA) also showed a similar functional chart with one enriched major Network function, described as cellular compromise, nucleic acid metabolism and small molecule biochemistry, having 13 of 15 total genes (network on figure 2.16). A large number of markers showed excess zebu ancestry for BOR cattle as compared list of markers with excess taurus ancestry.

potential interaction within the list. A total of 25 interacting molecular networks were found and 16 Networks had significant scores ($-\log P$ value >10). The top Network with ($-\log P$ value =39) contains 24 of the genes from the excess zebu ancestry list. It is functionally involved in gene expression, infection mechanism, cardiovascular system development and function (Network figure 2.17). In this group it is interesting to note list of genes functionally related to host immunity, pigmentation, vitamin/nutrient metabolism and neurology.

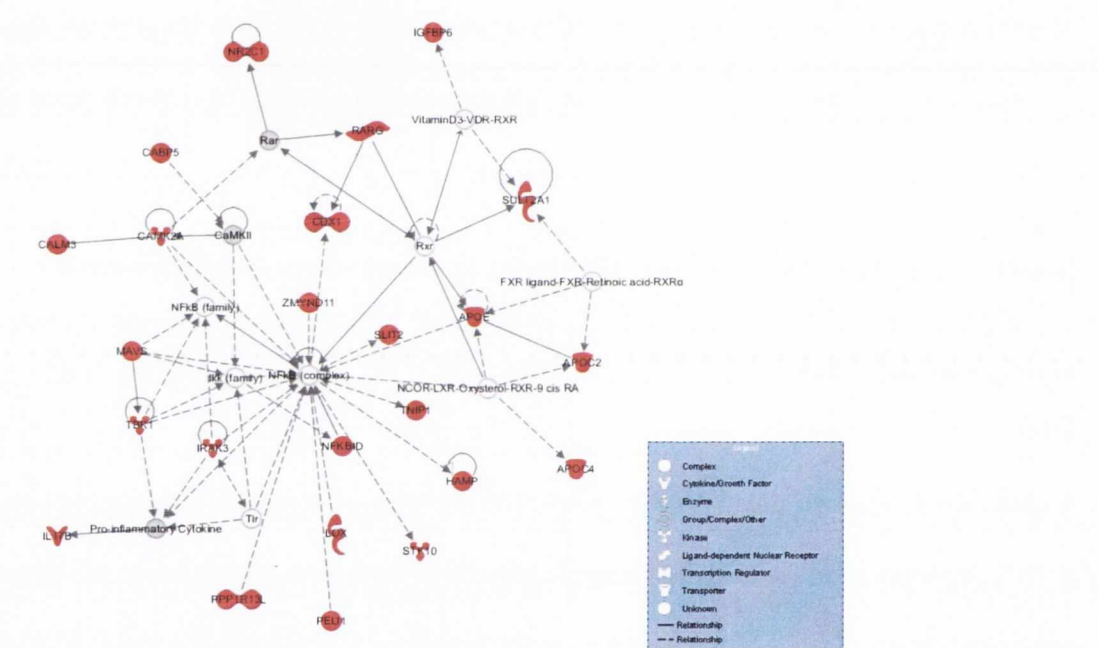


Figure 2.17. Network analysis of genes associated with excess *Bos indicus* ancestry in Borgou. Ingenuity Pathway Analysis (IPA) indicates known interactions involved in Gene Expression, Infection Mechanism, Cardiovascular System Development and Function. Genes in color are the subsets of the list that has all significant excess zebu ancestry (Table 2.3b).

2.4 Discussion

In this study, genomewide admixture levels of two African hybrid populations were investigated. These come from different hybrid zones. Principal component analysis and STRUCTURE consistently revealed higher levels of zebu introgression in Eastern Africa than in the West Africa hybrid zone. This estimation is consistent with previous studies based on mitochondrial DNA and Y chromosome data (FREEMAN *et al.* 2006a; HANOTTE *et al.* 2002; LOFTUS *et al.* 1999a; LOFTUS *et al.* 1994b). The top two eigenvectors resolve African, European *Bos taurus* from *Bos indicus*, and together with STRUCTURE assign individuals into three continental clusters that may be used to infer the possible nature of genetic introgression. Therefore, a three way admixture model is applicable to infer inter-continental cattle population structure and genetic introgression patterns. A two way admixture model in our study populations fits the previously described geography of admixture in African cattle.

The widely accepted history of cattle introduction/domestication in the African continent and further expansion to different regions within Africa using archeological and genetic data has suggested the primary domestic cattle inhabitants were *Bos taurus* probably originating mainly via Northern Africa from the Near East (LOFTUS *et al.* 1994a) with some possibility of independent African *Bos taurus* domestication within the continent. However, zebu introgression into the African continent was not a singular process (MACHUGH *et al.* 1997), possibly initiating 2,000-3000 years ago through East Africa (Horn of Africa) (HANOTTE *et al.* 2000). A second zebu wave into Africa is more recent,

partly related to a rinderpest epidemic, and is reported to be of high impact on the present day composition of modern cattle population in the continent (HANOTTE *et al.* 2002).

One of the most important issues raised along with the nature of introgression is the time estimate since the first admixture event. In this context, the admixture event is much older than most reports of admixture mapping done on human populations which often concern a 10 - 15 generation history such as that involved in the African American model (HOGGART *et al.* 2004; KOPP *et al.* 2008; REINER *et al.* 2005). Accurate estimation of recombination switches between segments of chromosome derived from the contributing distinct parental populations is a mainstay in the final resolution of age of admixture. Given more than 3000 years since the first event of zebu introgression into African continent, recombination events seem to fade LD with time (FREEMAN *et al.* 2006b). This might demand the need for highly dense markers and at the same time keeping the threshold inter-marker distance for optimal function of clustering programs used later in the analysis stage.

An admixture event in a population as result of multiple events causes serious computational challenges and accurate prediction has not been addressed very well in the currently available admixture mapping tools. The latest tool, *HAPMIX* has been reported to outperform any other existing programs in terms of extracting maximum information about the ancient ancestry in the test population (PRICE *et al.* 2009). We therefore assume our time of admixture prediction using *HAPMIX* is more plausible and it fits relatively well with previous reports and archeological evidence (LOFTUS *et al.* 1999a). This provides evidence that zebu introgression happened around 3 thousand years ago through Indian Ocean trade. The expansion of the zebu genome continued to the West and Southern

part of the African continent. A probable recent gene flow from pure African *Bos taurus* (Somba) in West Africa influences the observed genetic structure of modern day Borgou cattle. Certainly, the time depth estimate of the present study supports the later expansion of zebu into tropical West Africa where pure *Bos taurus* still exist. It is also important to note that this may be an amalgam of successive events as mentioned earlier in the text.

The evidence presented in this chapter, strongly supports that admixture is uneven throughout the genome in both East and West African cattle hybrid zones. Some of the genes listed as either excess *Bos taurus* or *Bos indicus* ancestry may be under selection for one or other ancestral parent. Particularly, genes related to have an impact on host immune response, nutrient /vitamin metabolism and pigmentation could be primary candidates (GAUTIER *et al.* 2009). The evidence given for specific selection in these regions gives some suggestion of adaptation to variable environmental conditions and improved immunity against pathogens (GIBBS *et al.* 2009). For example, that the bovine MHC in BTA23 is important in the general immune response in cattle and present a strong possibility of regional selection is well documented (HESS *et al.* 1999). The immune genes listed in this text including bovine MHC genes could be a starting point to analyse comparative host immunogenetic responses for various cattle diseases that occur at difference rates between *Bos taurus* and *Bos indicus* populations. A classical example is where taurine populations persist in subtropical regions of West Africa with an inherited resistance to trypanosomiasis that allows them to inhabit areas infected with the tsetse fly. Zebu cattle have no innate immunity /resistance to typanosomiasis and the population has only recently started to expand in these regions with assistance from better vector control measures and availability of treatments. Secondly, *Bos indicus* cattle breeds are more resistant to tick infestation than *Bos taurus* breeds and the profile of immune response

patterns between the two breeds during infestation has been implicated in the resistant phenotype (PIPER *et al.* 2009). These examples of disease models accounts the application admixture mapping and the possible role of immune related genes in shaping cattle population and provide comparative advantage under external environmental factors. An interesting disease-related candidate among these results is the *SLAM* (CD150) gene.

The *SLAM* gene (CD150) in BTA3 is a signaling lymphocyte activation molecule known to be a receptor molecule on the lymphoid cells for morbilliviruses including the cattle rinderpest virus (BARON 2005). The possible adaptive change of *SLAM* gene is response to the past rinderpest epidemics were investigated by JUNG *et al* (unpublished). In this report from resequencing of *SLAM* region able to highlight the possible selection signature exist in European *Bos taurus* cattle. Since rinderpest is a huge veterinary relevance in Africa (DOMENECH *et al.* 2010) and historical epidemics in the past has been decimated native population, this makes this gene one of the strongest potential candidates to study selection among African cattle populations.

CHAPTER THREE: PHYLOGEOGRAPHY OF CATTLE

3.1 Introduction

Understanding of cattle domestication, global dispersion and population structure have each been enriched by synthesis of genetic, archaeological, and historical data accumulated around the phylogeography of cattle populations. The current consensus suggests independent domestication events for *Bos taurus* and *Bos indicus* in the Near East and South Asia respectively (LOFTUS *et al.* 1999a; LOFTUS *et al.* 1999b; LOFTUS *et al.* 1994a; TROY *et al.* 2001). The distribution genetic diversity in other parts of the world has been influenced by many factors such as adaptation to specific eco-climatic regions and the degree of resistance to diseases and cultural preferences.

Genetic analysis and population structure of cattle has outlined the predomestication division of taurine and indicine cattle as well as later events of breed variations and admixture. Whereas much has been achieved using a limited number of markers the recent development of medium density marker coverage of bovine genome has given the opportunity to investigate population structure at an unprecedented level of resolution (GIBBS *et al.* 2009; VAN TASSELL *et al.* 2008). For example, haplotype block and LD analysis outlines the chromosomal segments that have undergone little historical recombination which has importance for designing association studies for various complex traits (CARDON and ABECASIS 2003).

Comparative study on the extent of LD and haplotype block structure of European and African cattle have showed a very similar LD pattern irrespective of their breed origin. In general very high LD at a short distance has been found in cattle samples (HAYES *et al.* 2007; HAYES *et al.* 2008). This is similarly observed at a much higher density SNP analysis

and breed coverage (GIBBS *et al.* 2009). Haplotype block studies therefore outlines degree cattle genomic mosaicism which may be attributed to variable recombination and mutation rates and genetic hitchhiking (KHATKAR *et al.* 2007). Genetic drift, inbreeding or gene flow are expected to influence the structure of the bovine genome and are directly related to the demographic history of the populations (HAYES *et al.* 2003); as a result integrated population structure, haplotype block and LD analysis offer the opportunity of a more comprehensive view of bovine phylogeography.

In this chapter the impact of cattle admixture on demographic history and the existing genetic structure of cattle are explored using a genomewide sample of SNP diversity in African, European, Near Eastern and Indian cattle.

3.2 Materials and Methods

3.2.1 Study populations

The cattle population in this study are samples collected from Africa, Asia, the Middle East and Europe. In order to facilitate this cross-sectional study, only unrelated individual cattle showing typical breed phenotypes in each geographical location were included. A total of 208 samples were from West and East African origin. 40 cattle samples from the Middle East, 70 Asian and 56 European animals were also included (figure 3.1). Sampling location and breed information is shown in table 3.1. Samples of 20 Holstein, 20 Gir, and 20 Brahman were kindly provided from the Bovine Hapmap consortium. The rest of the samples were previously collected cattle DNA archived in TCD.

The DNA samples of all animals were quantified using the fluorescence method, Qubit™. The DNA concentration of each sample was adjusted to between 400-700 ng for Goldengate genotyping on the Illumina platform. The assay was performed at the Wellcome Trust Centre for Human Genetics, Roosevelt Dr., Oxford, UK.

Table 3.1: Study populations and source of sampling locations

Breed name	Breed code	Geographic location of sampling region	Classification	Sample size
Thraparker	TRP	India	<i>Bos indicus</i>	7
Hariana	HAR	India	<i>Bos indicus</i>	10
Sahiwal	SWL	Pakistan	<i>Bos indicus</i>	8
Brahman	BRM	USA and Australia (Hapmap)	<i>Bos indicus</i>	20
Gir	GIR	Brazil (Hapmap)	<i>Bos indicus</i>	20
Boran	BRN	Kenya	<i>Bos indicus/Bos taurus</i>	8
Ethiopian zebu	ETZ	Ethiopia	<i>Bos indicus/Bos taurus</i>	40
Borgou	BOR	Benin	<i>Bos taurus/Bos indicus</i>	40
Kapsiki	KAP	Cameroon	<i>Bos taurus/Bos indicus</i>	39
Kuri	KUR	Chad	<i>Bos taurus/Bos indicus</i>	40
Taleshi	TAL	Iran	<i>Bos indicus/Bos taurus</i>	32
Djakore	DJK	Senegal	<i>Bos taurus/Bos indicus</i>	34
Holstein	HOL	Ireland,(USA and New zealand -Hapmap)	<i>Bos taurus</i>	56

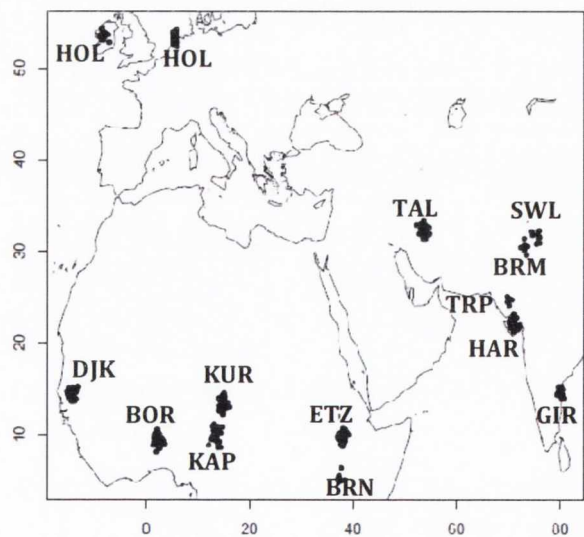


Figure 3.1. The geograhic origin of study populations. Tharaparer (TRP) ,Hariana (HAR), Sahiwal (SWL), Brahman (BRM), Gir(GIR), Boran(BRN), Ethiopian zebu (ETZ), Borgou(BOR), Kapsiki(KAP), Kuri(KUR), Taleshi(TAL), Djakore(DJK) and Holestein(HOL).

3.2.2 Ancestry Informative Markers

The absolute difference in allele frequency between pure *Bos taurus* and *Bos indicus* populations (δ) was used as a measure of informativeness of markers as described in chapter 2 (equation 1). The total number of markers selected for Goldengate assay design was 768 SNPs. Selection of informative markers involved first choosing the best available SNPs from those typed by the Bovine Hapmap consortium (GIBBS *et al.* 2009) which were reinforced by additional markers from the Illumina 54001 SNP (VAN TASSELL *et al.* 2008) chip data used to genotype 8 cattle populations from Africa, Asia and Europe, as described in chapter 2.

The Hapmap database constitutes 33K SNPs typed for over 500 animals derived from 20 breeds around the world. These markers include a substantial fraction discovered by resequencing of a zebu (Brahman) individual and comparison to the Hereford sequence. In contrast, the commercial 50K array had only Northern European breed centred SNP discovery and hence less likelihood of giving truly informative SNPs for *Bos indicus* versus *Bos taurus*. The following criteria were used for filtering informative markers from both data bases.

- A. Allele frequency difference (δ) between European *Bos taurus* and *Bos indicus* of 0.6 and above.
- B. Allele frequency difference (δ) between African *Bos taurus* and *Bos indicus* of 0.6 and above.
- C. Inter marker distance between two consecutive SNPs of approximately 4Mb achieving even coverage across all chromosomes.

- D. SNPs from other sources were included based on biological importance of the region
- E. A selection of packed SNPs from two densely typed chromosome regions to facilitate analysis of old admixture events was included.

3.2.3 Population Structure and phylogeography

The genetic distance between the breeds was calculated by the *smart_pca* program implemented in EIGENSTRAT (CASTELNUOVO *et al.* 2006). The F_{st} value matrix was used to construct an unrooted NJ-tree using the NEIGHBOR software of PHYLIP package (PERRIERE and GOUY 1996).

Population structure was inferred using STRUCTURE V2.3 (FALUSH *et al.* 2007; PRITCHARD *et al.* 2000) which was used without assigned prior information about location and breed. The model parameter set used was similar to the protocol used in chapter 2. Population structure was inferred based on pre-assigned possible numbers of clusters (K). The program was run for 10,000 MCMC and 10,000 burnin parameters. Converging of the Ln likelihood values were visually monitored to decide the optimal running time. The DISTRUCT program (ROSENBERG *et al.* 2001) was used to plot and visualize the outputs of STRUCTURE analysis.

3.2.4 Individual admixture levels

The level of *Bos taurus* and *Bos indicus* ancestry for individual cattle was inferred from STRUCTURE outputs from the pre-assigned two- way admixture model ($k=2$). The individual genotypes were used for principal component analysis; reducing a large number of measurements into few principal components by accounting for the largest amount of

variability in the data. EIGENSTRAT was used for the analysis and plots of each individual using coordinates of principal components were made.

3.2.5 Haplotype inference and measure of LD

The genotype data from chromosome 6 was filtered out for all breeds and PHASE 2.1 (STEPHENS *et al.* 2001) was used to construct the best inferred chromosome haplotypes. The PHASE program implements a Bayesian model for reconstructing haplotypes from population genetic data. The default setting, based on a general model for recombination rate variation was selected as described by (FALUSH *et al.* 2003). The level of mosaicism for all individual chromosomes was plotted from the reconstructed haplotypes using a Perl script for visual interpretation.

The measure of LD between two SNPs used was r^2 which is derived from Lewontin's D. In this study it is sensible to use r^2 over D' as the former is less liable to inflation when a small sample size is used. The value was measured for chromosome 6 of each breed using PLINK. The pattern of LD was estimated for 1Mb windows along the chromosome.

3.2.6 Age of admixture

Pure *Bos taurus* (Holstein-Friesian) and *Bos indicus* (Gir, Brahman, Hariana, and Tharaparker) breeds were assigned as reference populations and used to model admixed populations to estimate the age of admixture in the hybrid zones. The HAPMIX program (PRICE *et al.* 2009) was used with parameters similar to those explained in chapter 2. The genotype of East African *Bos taurus* (Sheko) data was imported from Hapmap database. Common markers of BTA6 were used to estimate age of admixture under the same strategy.

3.3 Results

3.3.1 Selection of Ancestry informative markers and genotype quality control

Selection of informative markers based on the formulated criteria and the locations of SNPs for each chromosome are shown in figure 3.2. The panel includes 652 SNPs from the Hapmap data set with extreme δ values between African/European *Bos taurus* and *Bos indicus*. The δ value for African *Bos taurus* Vs *Bos indicus* was calculated using N'dama (pure taurus) and pooled genotypes of zebu animals (Nelore, Gir and Brahman). The δ value for European *Bos taurus* Vs *Bos indicus* was obtained from comparison of Holstein (pure taurus) and pooled zebu animals. 93 SNPs were chosen from 50k Illumina array data from chromosome regions where there were no Hapmap SNP that fulfilled the selection criteria. An additional 18 SNPs were included from 50k data which have shown strong association with tuberculosis in a genomewide analysis of Holstein sires in Ireland. 5 SNPs were selected based on previous studies suggesting strong selection signals in regions that contain immunity related genes. Finally, a dense SNP sampling was included in two regions of chromosome 6 in order to observe patterns of close LD.

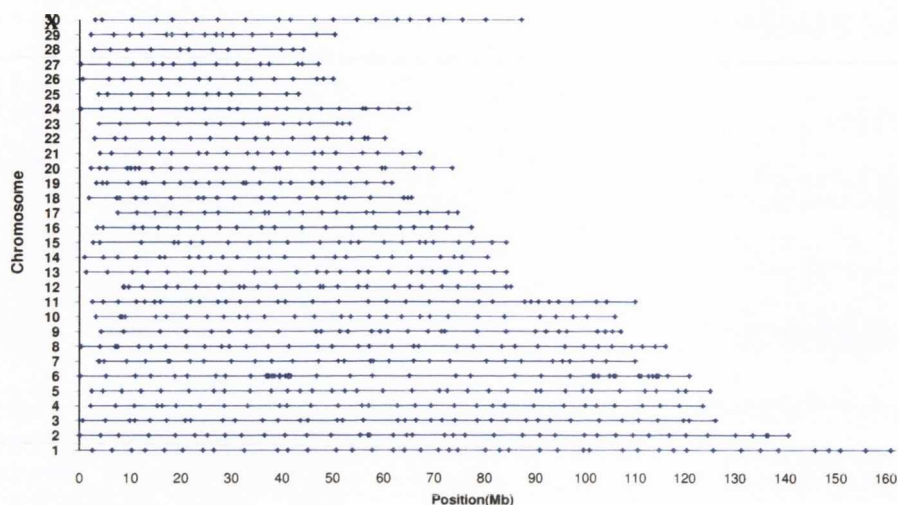


Figure 3.2. The physical positions of 768 SNPs across all the chromosomes chosen for the Goldengate assay. The horizontal lines represent chromosomes and the tick marks indicate the location of a SNP typed for the study. On BTA6 around 40Mb, and between 100 and 120 Mb high density SNP regions are observable.

Standard quality control tests were performed as described in chapter 2. A total of 337 out of 354 animals were used for further analysis. 17 samples were removed for low genotype data. 14 SNPs failed for all the samples and were removed from the data set. 21 SNPs were typed from the X-chromosome were also not used in the general analysis. The final genotype contains 733 autosomal SNPs in which BTA6 has 53 SNPs and the average total genotyping rate was 96.3%. The δ value distribution based on 733 SNPs is shown in figure 3.3.

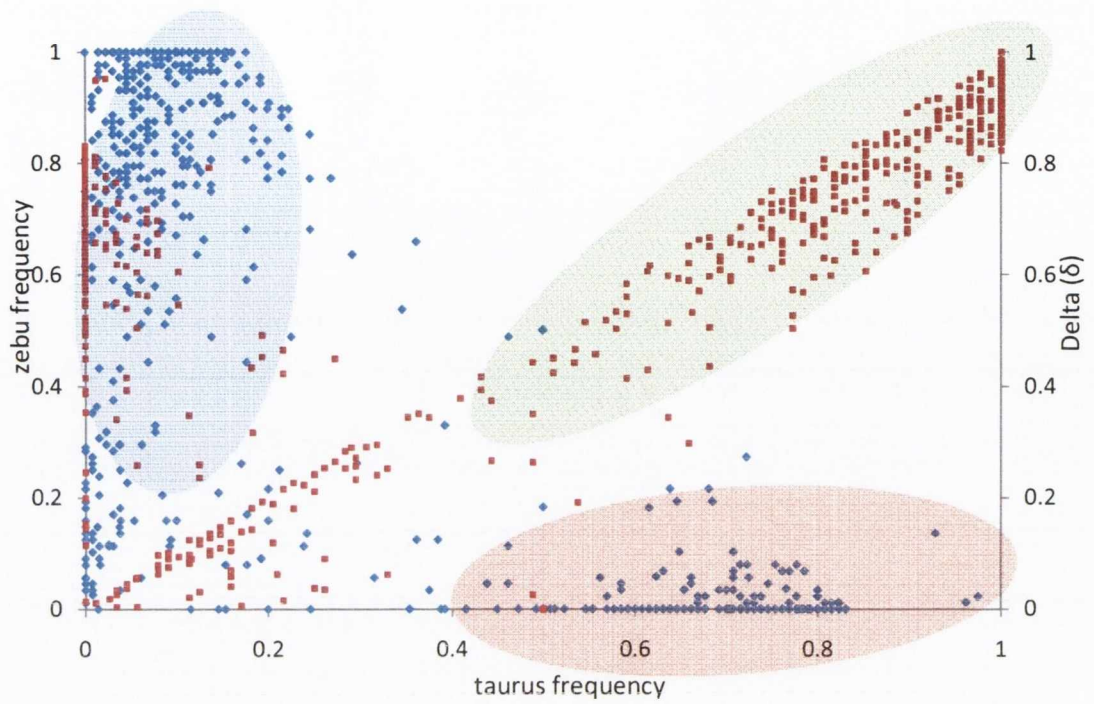


Figure 3.3. The allele frequency of taurus and zebu (primary xy axis), the corresponding delta values shown in the secondary y axis. Most of the markers are either high frequency zebu and low frequency taurus or vice versa (shaded blue and red). Top informative markers are highlighted in green.

3.3.2 Genetic distance and phylogeography

The genetic distances estimated between the sampled cattle populations are given in the Fst matrix shown in table 3.2. Genomewide Fst values within close geographic regions are low. The distance within Asian pure *Bos indicus* cattle reaches a maximum between Tharaparker and Sahiwal (0.082) and between Haryana and Gir (0.036). Similarly comparison between West African cattle populations also show relatively small values ranging from 0.024 for Borgou and Kapsiki to 0.158 for Kuri and Djakore. The Fst was 0.028 for Ethiopian zebu and Boran which are East African populations. However, there is substantial difference between East and West African populations. For example the Fst between Boran and Djakore was 0.409. It is also interesting to note that within the African continent the genetic distance linearly increases with geographic distance between the current locations of breed origin. The Near Eastern cattle sampled in this study, Taleshi are closer to the African hybrid cattle than the Asian *Bos indicus* and European cattle Holstein. As expected, the genetic distance between pure *Bos taurus* and *Bos indicus* has the highest value. The genetic distance among the continental cattle breeds is summarised in the NJ tree constructed based on the Fst matrix figure 3.4.

Table 3.2: The Fst matrix between cattle breeds

	HOL	DJK	TAL	KUR	KAP	BOR	ETZ	BRN	GIR	BRM	SHL	HAR
DJK	0.149											
TAL	0.425	0.277										
KUR	0.316	0.158	0.080									
KAP	0.262	0.104	0.120	0.042								
BOR	0.273	0.106	0.096	0.028	0.024							
ETZ	0.510	0.358	0.060	0.114	0.172	0.140						
BRN	0.557	0.409	0.096	0.154	0.220	0.187	0.028					
GIR	0.797	0.682	0.325	0.433	0.511	0.473	0.219	0.209				
BRM	0.714	0.588	0.227	0.328	0.407	0.370	0.131	0.122	0.059			
SHL	0.782	0.664	0.304	0.412	0.492	0.452	0.198	0.189	0.068	0.074		
HAR	0.775	0.656	0.290	0.400	0.479	0.442	0.184	0.172	0.036	0.041	0.049	
TRP	0.790	0.673	0.313	0.423	0.500	0.462	0.207	0.201	0.069	0.074	0.082	0.056

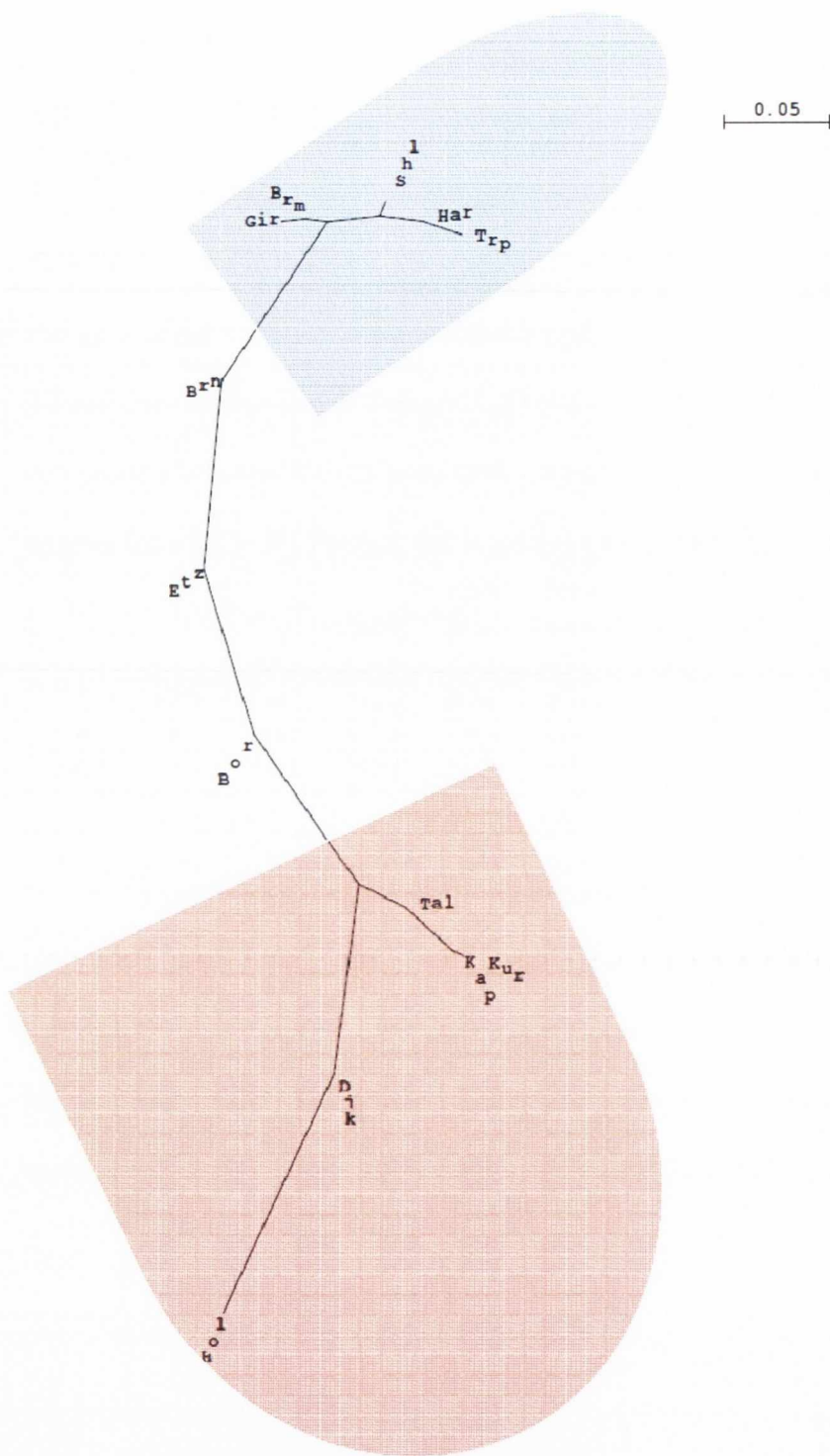


Figure 3.4 Neighbor joining phylogeny plot from the genetic distance matrix for 13 cattle populations. The coloured areas indicate the zebu (blue) and taurus (red) braches of the tree. Tharaparar (TRP) ,Hariana (HAR), Sahiwal (SWL), Brahman (BRM), Gir(GIR), Boran(BRN), Ethiopian zebu (ETZ), Borgou(BOR), Kapsiki(KAP), Kuri(KUR), Taleshi(TAL), Djakore(DJK) and Holestein(HOL).

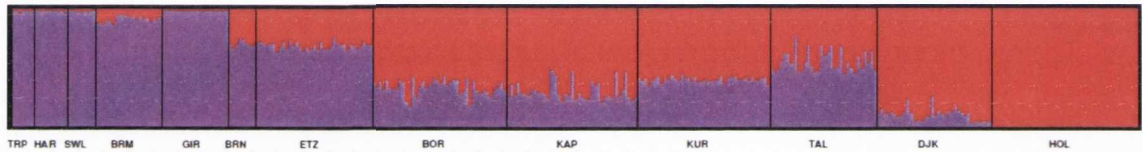
3.3.3 Population structure and Admixture

The clustering algorithm implemented in STRUCTURE can further elaborate on population structure and the patterns of *Bos indicus* and *Bos taurus* genetic admixture (figure 3.5). The posterior estimates for the two ancestor clustering model ($K=2$) shows that Tharaparker, Haryana, Sahiwal, Gir and Brahman represent a putative ancestral population of *Bos indicus*. Likewise, Holstein is assigned as a pure *Bos taurus* ancestral population. The remaining populations show variable composite levels of the two ancestral populations. The assignment $K=3$ marked out the substructure between African and European *Bos taurus*. Additional cluster assignment of $K=4$ seems to highlight the Near Eastern population component.

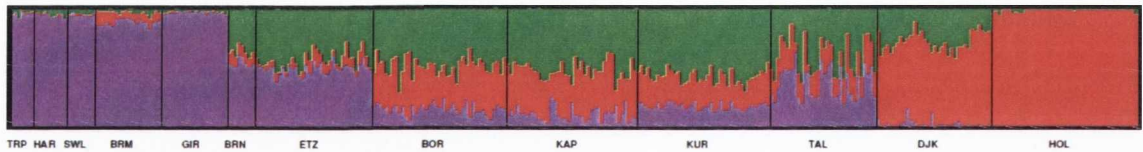
The two ancestor model infers the level of genetic ancestry of each population, interpretable as the level of *Bos taurus* (red colour) and *Bos indicus* (violet colour) of a particular breed (figure 3.5). African cattle have non uniform levels of admixture between zebu and taurus genetic ancestry. The East African cattle have the highest level of zebu ancestry (Ethiopian zebu, 68% and Boran, 72%). The level of taurus ancestry increases towards West Africa and the level reaches a maximum for Djakore (89%). Taleshi also displays an intermediate level of admixture between zebu (59%) and taurus (41%) ancestry. Unlike the other populations, the Taleshi show extensive individual variation of admixture level.

The principal component analysis based on the autosomal genotype data is better suited to observe how well individual animals cluster within breeds and how these compare across different geographic zones (figure 3.6). The first principal component explains 78% of the variation and differentiates between *Bos taurus* and *Bos indicus*. Along the first eigenvector one can visualize the level of ancestral taurus and zebu contribution to each individual which is consistent with the STRUCTURE analysis. The second eigenvector, accounting for 5% of the variation, indicates the substructure between African and European *Bos taurus* breeds (figure 3.7).

K=2



K=3



K=4

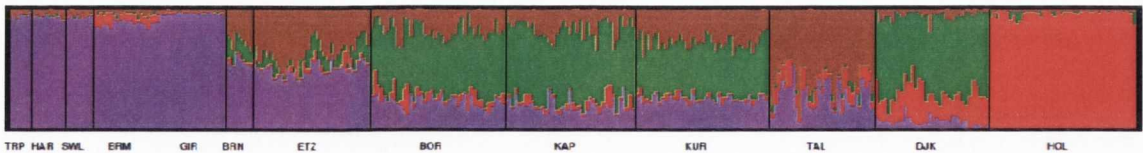


Figure 3.5. Population structure inferred from clustering program STRUCTURE using K=2, 3 and 4. Tharaparer (TRP),Haryana (HAR), Sahiwal (SWL), Brahman (BRM), Gir(GIR), Boran(BRN), Ethiopian zebu (ETZ), Borgou(BOR), Kapsiki(KAP), Kuri(KUR), Taleshi(TAL), Djakore(DJK) and Holestein(HOL).

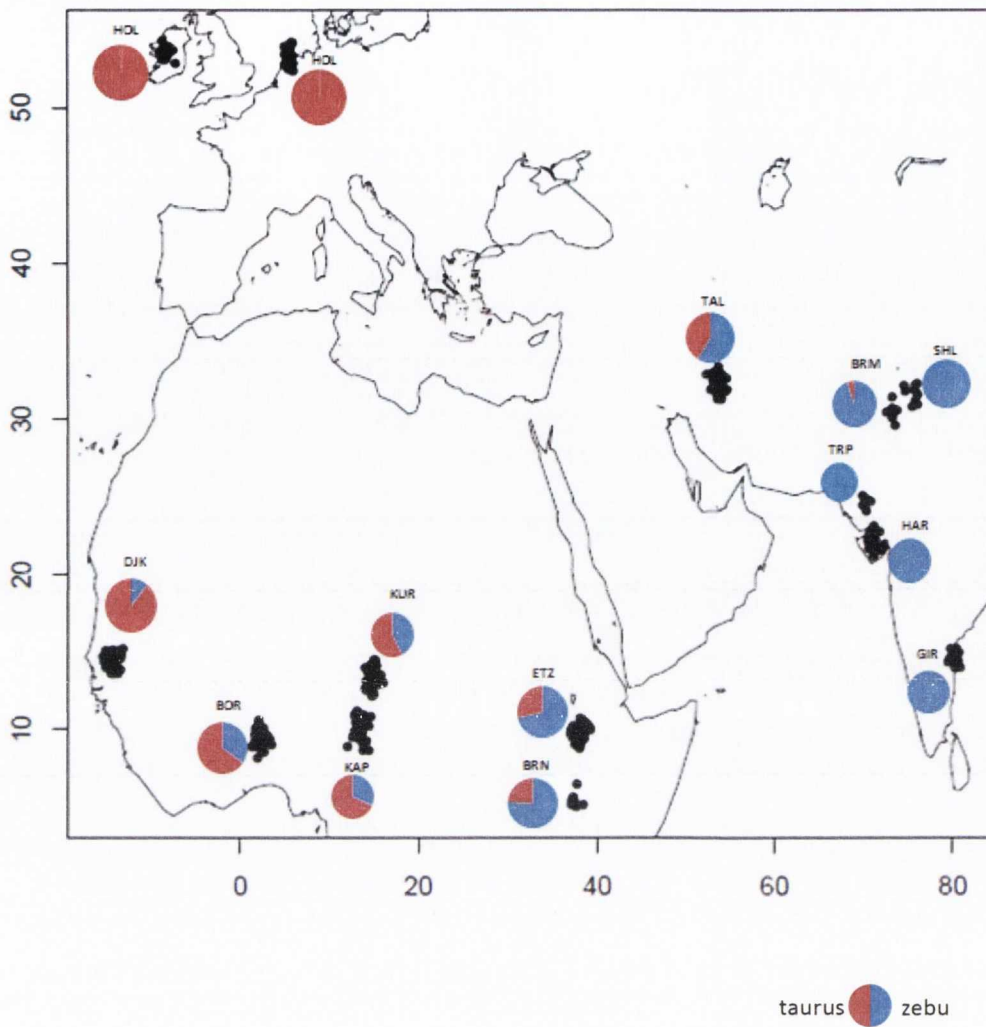


Figure 3.6. Population structure in relation to the geographical distribution of the study populations. The pie chart for the breeds indicate the taurus (red colored) and Zebu (blue colored) genetic ancestries in the population inferred from STRUCTURE program. Tharaparar (TRP), Hariana (HAR), Sahiwal (SWL), Brahman (BRM), Gir (GIR), Boran (BRN), Ethiopian zebu (ETZ), Borgou (BOR), Kapsiki (KAP), Kuri (KUR), Taleshi (TAL), Djakore (DJK) and Holestein (HOL).

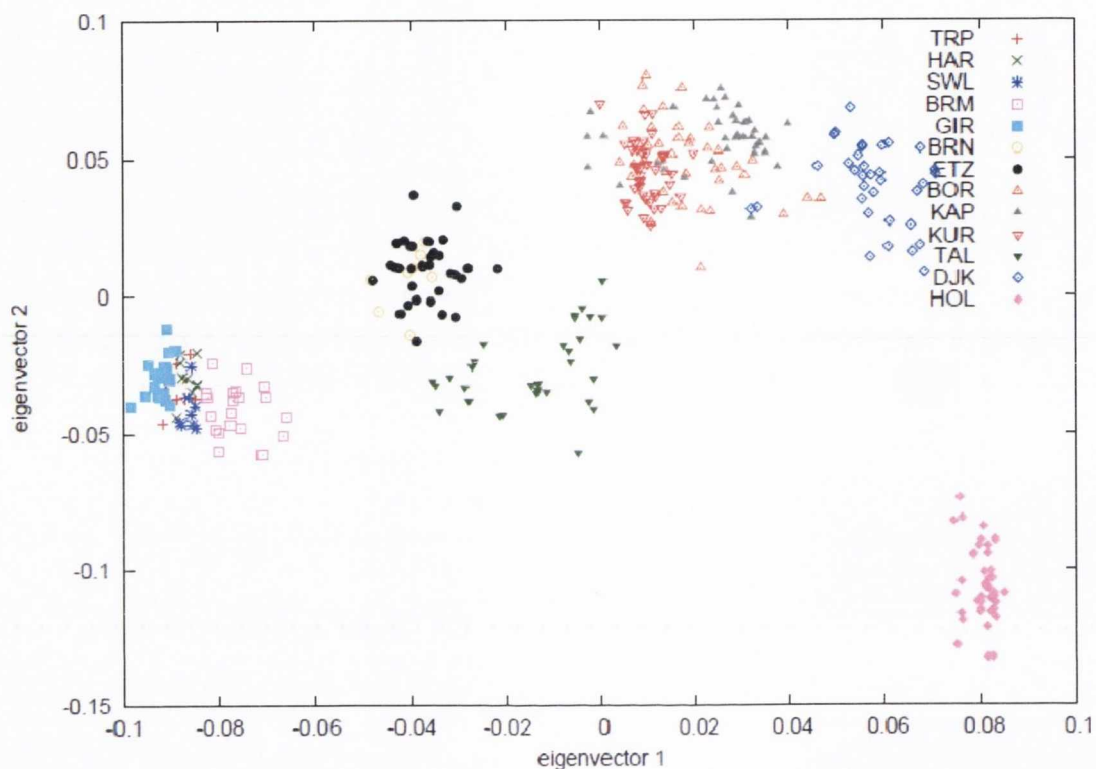


Figure 3.7. Principal component analysis of 13 cattle breeds using 733 SNPs. Animals are plotted according to their coordinates in the first two eigenvectors. The first eigenvector shows 78% of the variation and the second component explains 5% of variation.

Tharaparar (TRP) ,Hariana (HAR), Sahiwal (SWL), Brahman (BRM), Gir(GIR), Boran(BRN), Ethiopian zebu (ETZ), Borgou(BOR), Kapsiki(KAP), Kuri(KUR), Taleshi(TAL), Djakore(DJK) and Holestein(HOL).

3.3.4 Age of admixture

Age of admixture estimates for the hybrid cattle populations were predicted using analysis implemented in the HAPMIX programme. This software uses a Bayesian approach to assign ancestry to blocks of each chromosome. Age of ancestral admixture is related to the degree of mosaicism within haplotypes and posterior probabilities of this were generated for each breed sample. The plot indicated in figure 3.8 showed the distribution of likelihood values expressed as the number of generations since the first admixture event at a particular hybrid zone. In order to infer a sensible estimate the 1 LOD and 2 LOD intervals from the maximum likelihood values were estimated by extrapolation (figure 3.9a & 3.9b). The oldest admixture event was recorded for Taleshi where the curve did not reach a plateau in the 2000 generation interval calculated. The admixture event therefore could go as far back as 10,000 years before present assuming an average value of a generation to represent 5 years. Consistent with the result presented in chapter 2, the East African hybrid zone is much older than the West African cattle admixture event. The youngest age was predicted for Kapsiki (KAP) at 75 (22 – 311) years before present. When genotype data from Hapmap for one of the oldest hybrid cattle (Sheko) was imported and the common 43 SNPs were recovered including the dense regions from BTA6. The age of admixture estimate was 4500 (2134 – 7963) years before present which is consistent with other East African hybrids. The observed trend was interrupted for Djakore which gave an unresolved curve, which may be due to the fact that it shows a high level of *Bos taurus* ancestry at nearly 90%. Here, it is difficult to show a defined likelihood curve for the age of admixture.

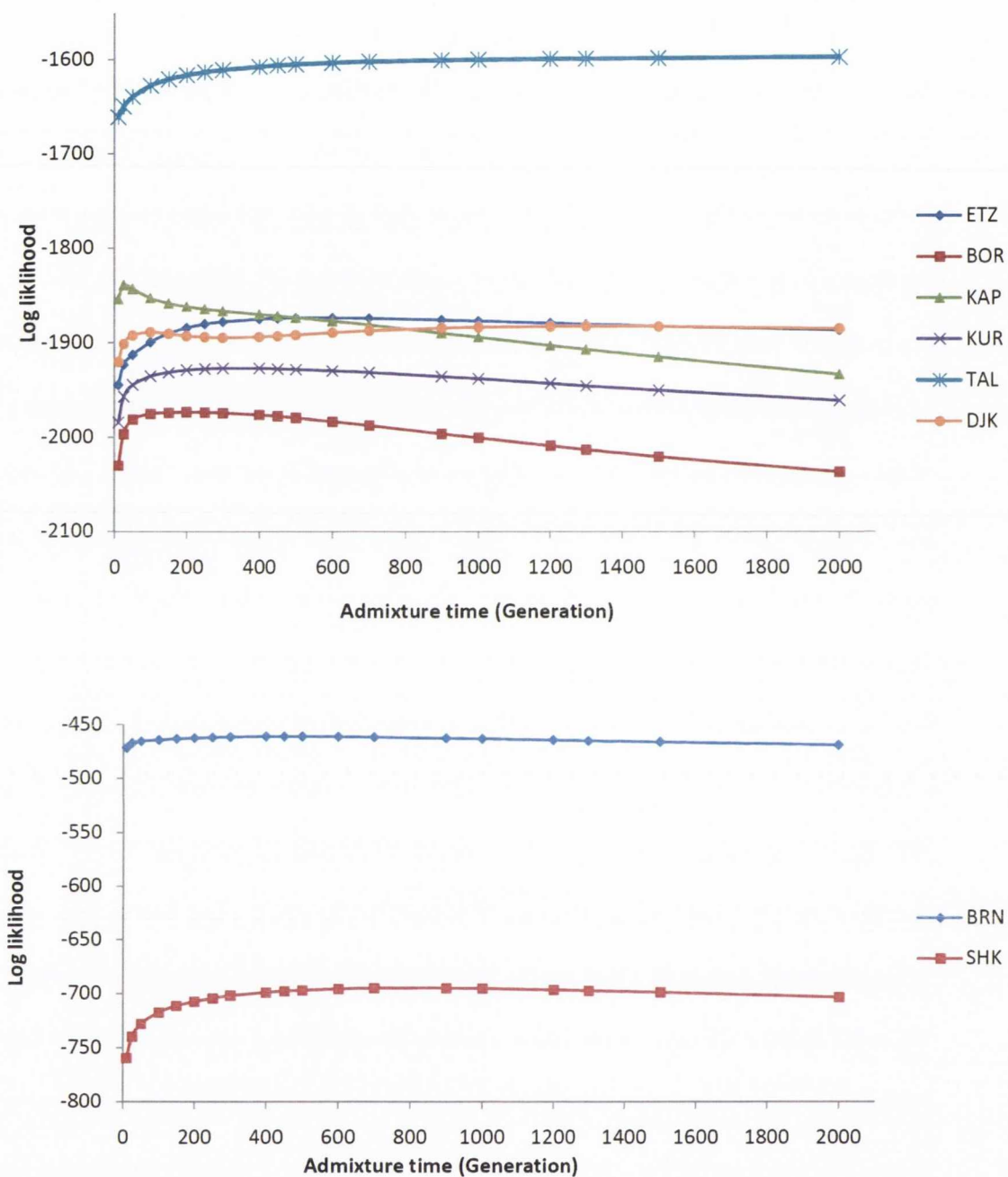


Figure 3.8. Log likelihood curves for estimates of age of admixture among the hybrid populations. Boran(BRN), Ethiopian zebu (ETZ), Borgou(BOR), Kapsiki(KAP), Kuri(KUR), Taleshi(TAL), Djakore(DJK) and Sheko(SHK).

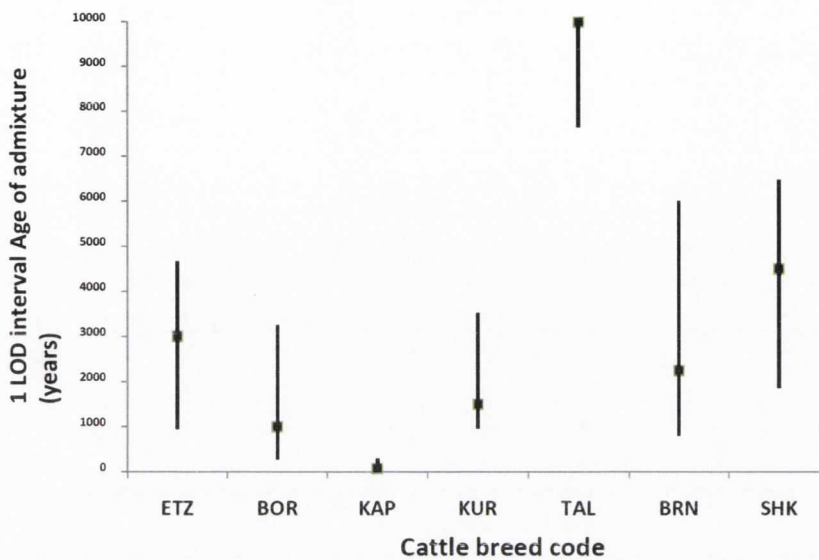


Figure 3.9a. The 1 LOD interval estimate of age of admixture time for all hybrid cattle populations including a relative estimate for Sheko (SHK) derived from imported common genotypes from Hapmap data on BTA6.

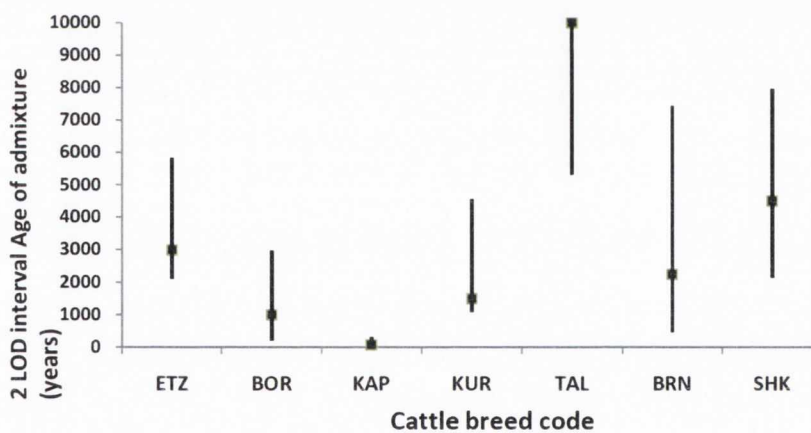


Figure 3.9b. The 2 LOD interval estimate of age of admixture time for all hybrid cattle populations including a relative estimate for Sheko (SHK) derived from imported common genotypes from Hapmap data on BTA6. Boran(BRN), Ethiopian zebu (ETZ), Borgou(BOR), Kapsiki(KAP), Kuri(KUR), Taleshi(TAL), Djakore(DJK) and Sheko(SHK).

3.3.5 Haplotype structure and measure of LD

Chromosome 6 has a higher density of SNP typing than other genome regions and here ancestral origins are estimated from contiguous haplotype regions along the chromosome within each individual. First PHASE was used to infer allelic phase and then alternate alleles were denoted as grey and red for each SNP (figure 3.10). Note that the colour coding does not represent alternates of *Bos indicus* or *Bos taurus* ancestry, but is rather assigned based on which is the major allele.

Here a visual comparison of the degree of ancestral mosaicism within haploblocks is possible and gives three obvious clusters. First haploblocks remain consistent within the unadmixed breeds of *Bos indicus* (GIR, BRM, HAR, TRP, and SWL) and *Bos taurus* (HOL, DJK). Second, within the more recent admixtures (BOR, KAP, and KUR) ancestral origins are mixed but tend to be consistent within each phased haplotype, especially within the densely typed regions. Third the Ethiopian zebu (ETZ), Boran (BRN) and particularly Taleshi (TAL) are ancient admixed populations and show marked disruption of ancestral haplotype blocks.

An additional means of considering haplotype integrity is examination of linkage disequilibrium (LD). The extent of pairwise LD between SNPs in the high density genotypes of BTA6 is measured by r^2 and plotted Vs physical distance (Kb) for each breed (figure 3.11). The $\log(r^2)$ pattern shows that LD drops sharply within 200kb but the nature of LD decay varies among the different samples. The extent of LD for Holstein and Taleshi was lowest while Kapsiki has the highest value at the smaller distances. The overall LD pattern was higher for West African cattle and intermediate for the pure *Bos indicus* populations with the exception of Brahman which has some recent admixture.

Hybrids as expected had elevated LD with some exception of Taleshi which was estimated as the oldest admixture.

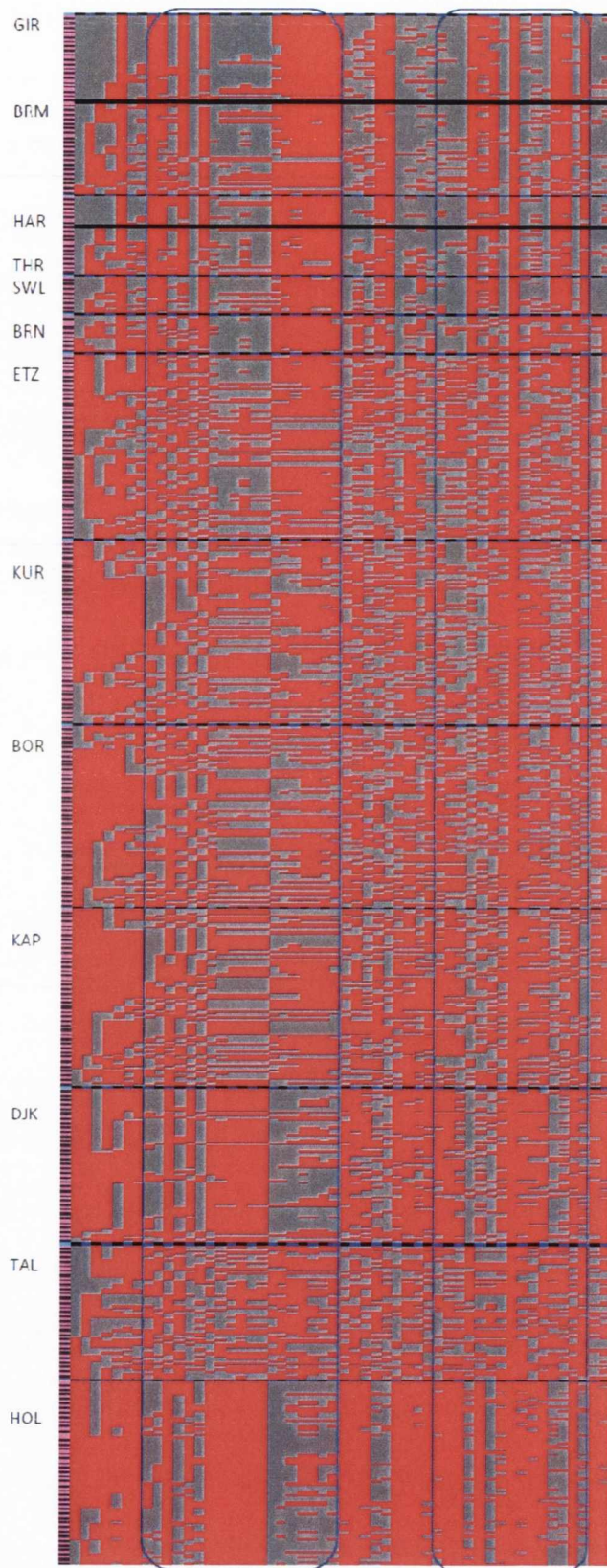


Figure 3.10 Haplotype block structure of 13 cattle breeds indicating one chromosome per row for each individual based on BTA6. The variation indicates alternates of major allele shown by red and grey color. Haplotype block variation among populations can be viewed along vertical line while within population variation could be observed along the horizontal axis. The gird region shows the two selected dense SNP locations. Tharaparer (TRP), Haryana (HAR), Sahiwal (SWL), Brahman (BRM), Gir (GIR), Boran (BRN), Ethiopian zebu (ETZ), Borgou (BOR), Kapsiki (KAP), Kuri (KUR), Taleshi (TAL), Djakore (DJK) and Holstein (HOL).

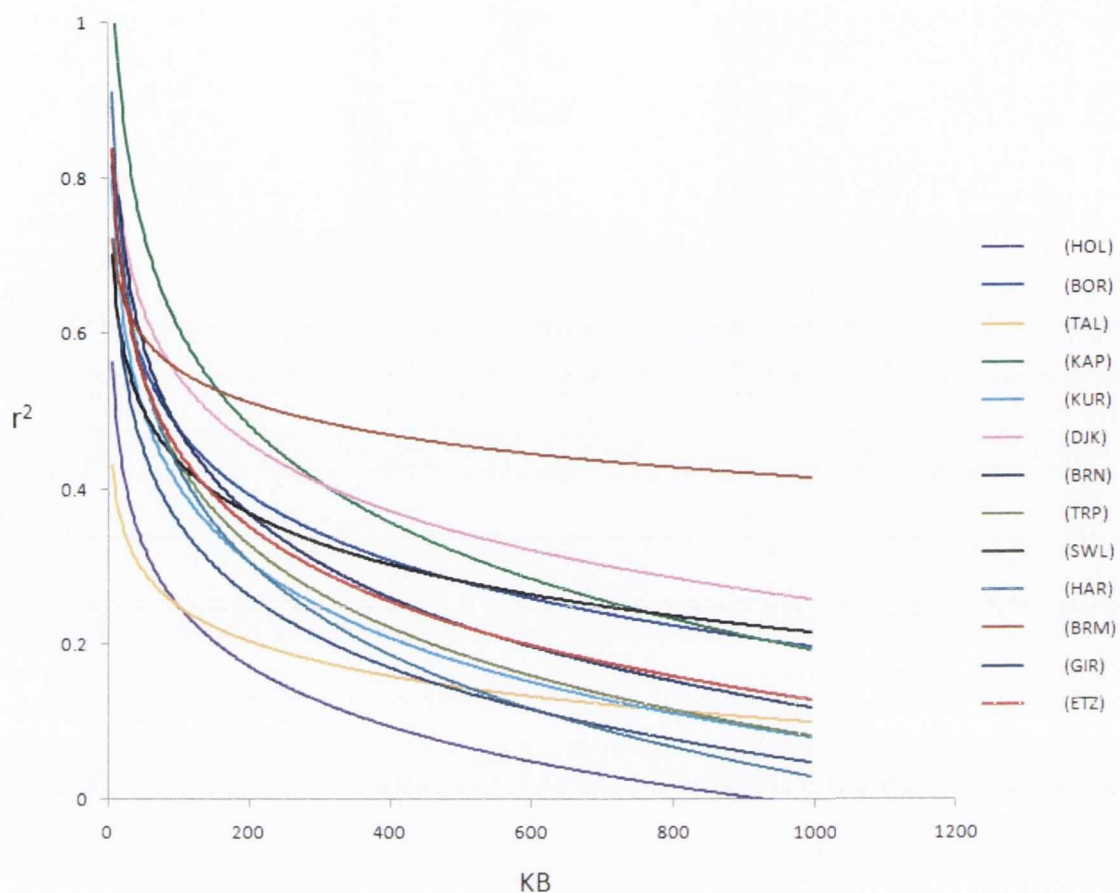


Figure 3.11. The Log r^2 estimate value for pairwise LD measure for each cattle breeds. Tharaparer (TRP), Hariana (HAR), Sahiwal (SWL), Brahman (BRM), Gir (GIR), Boran (BRN), Ethiopian zebu (ETZ), Borgou (BOR), Kapsiki (KAP), Kuri (KUR), Taleshi (TAL), Djakore (DJK) and Holstein (HOL).

3.4 Discussion

The genomewide analysis of indigenous cattle populations from various geographical regions demonstrates the population structure and the degree of admixture levels in hybrid zones. The informative markers used in the study clearly delineate *Bos taurus* and *Bos indicus* cattle and assist in placing their respective geographical origins. Secondly it also highlights the relationship of founder ancestral populations to admixed cattle at different hybrid zones.

The results obtained from genetic distance, population structure and principal component analysis marked out a substantial contribution of *Bos taurus* and *Bos indicus* in the Near East hybrid zone, interestingly with considerable variation in ancestry proportion between individuals. Metanalysis of microsatellite, mitochondrial and Y chromosomal data analysis (EDWARDS *et al.* 2007) suggested complex and possibly, sex biased zebu introgression to Near Eastern cattle.

The results here show marked reduction of *Bos indicus* ancestry levels over an East-West traverse of the African continent, a pattern that is well supported by previous studies (HANOTTE *et al.* 2002; HANOTTE *et al.* 2000). One of the possible factors given for low zebu admixture with distance across Africa from the East is relative susceptibility for sub-Saharan tropical diseases such as trypanosomiasis. The African *Bos taurus* in the west are innately tolerant of the disease. In symmetry, the differential advantage of *Bos indicus* cattle as relatively resistant to rinderpest has also probably contributed for their dominance in East Africa.

The analysis of LD here is comparable to the result presented based on the Hapmap database (VILLA-ANGULO *et al.* 2009). This suggested that cattle and human share similar patterns of LD with an average block scale of 100 kb. This study showed a slightly inflated figure which could be due to the difference in SNP density between the two studies. However another study on European *Bos taurus* (Angus) cattle has shown an average $r^2 = 0.2$ at 200kb which is similar to the observation presented in this chapter (HAYES *et al.* 2007).

The correlation of the degree of mosaicism with age of admixture has been documented previously (WIEHE *et al.* 2000). The data is in line with previous reports, that the Near East cattle represent the oldest hybrid zone while West African cattle have more intact ancestral haplotypes which is evidence for a recent admixture time depth (FREEMAN *et al.* 2006b; GAUTIER *et al.* 2007). The implication for association and mapping studies of such has been highlighted (BOUYER *et al.* 2007).

The annotation presented in this chapter is a quantitative measure for age of admixture in addition to comparative evidence for cattle admixture in the Near East, East Africa and West Africa. Zebu-taurus admixture in the Near-Eastern cattle populations from archaeological evidence is dated as far back as 9000- 10,000 BC possibly linked to ancient trade links. Historical adverse climatic events, for example the drought in Mesopotamia around 3000 – 4000 BP may have played significant role for introduction of arid-adapted zebu cattle into the Near East (GANOPOLSKI *et al.* 1998). The result presented for the admixture time using Iranian (TAL) cattle which is above 5000 years fits the archaeological narratives and the prior analysis based on X chromosome data showing the

Near East admixture as ancient as compared to other cattle hybrid zones (FREEMAN *et al.* 2006b).

Cattle admixture in African continent is later, the earliest domestic herds are considered to have been populated by *Bos taurus* type cattle (BRADLEY *et al.* 1996; EPSTEIN and MASON 1971) that first arrived through the Nile Delta around 6000 BC. An African independent centre of domestication, has also been postulated (BRADLEY *et al.* 1998; BRADLEY *et al.* 1996; HANOTTE *et al.* 2002; TROY *et al.* 2001). Archaeological evidence for zebu admixture dates to about 1500 BC, with a second introduction which was associated with Indian Ocean trade around 1000 BC. The horn of Africa is the epicentre of zebu ancestry in Africa (HANOTTE *et al.* 2002), indicating East Africa as the most ancient zone of cattle admixture. This is consistent with the finding in this study that Ethiopian zebu and Boran give the oldest admixture time estimates of 3000 (2099 – 5829) and 2250 (464 – 7426) respectively. Sheko which are believed to be the last remnants of East African *Bos taurus* cattle found in Ethiopia (albeit surrounded by indigenous zebu where they are under constant threat of losing taurine ancestry (HANOTTE *et al.* 2000)), has an estimated age of admixture 4500 (2134 – 7963). However, the zebu cattle expansion in the continent was never uniform and limited particularly along trypanosome free zones until very recently with the advent of prophylactic medicine. As a result the West African hybrid zone is clearly younger than the East African and Near East hybrid zones as shown in the case of Kapsiki 75 (22 – 311), Borgou 1000 (202 – 2966) and Kuri 1500 (1085-4544) populations.

CHAPTER FOUR: FIELD SAMPLE COLLECTION

4.1 Introduction

Ethiopia has the largest cattle population in the African continent. According to the Ethiopian central statistics authority (CSA) cattle census report in 2009/2010, the total cattle population for the country was estimated to be 50.8 million. The cattle herds in the country are dominated by the indigenous breed types which account for more than 99% of the total number (CSA 2010). Cross breeds account for less than 1% and are hybrids of indigenous with exotic (Holstein-Friesian and Jersey) cattle types. The most commonly known indigenous cattle breeds include Fogera, Horro, Arsi, Boran and Ogaden breeds (MOHAMMAD 2007). Breed type classification is based on the external phenotype of the animal, but population genetic structure among different breeds has been found to be minimal (DADI *et al.* 2008; LI *et al.* 2007). A similar finding was also observed using phylogeographical comparison between Arsi and Boran breeds (chapter 3). The Ethiopian central highland is predominantly stocked with the Arsi cattle breeds. They are believed to have been introduced in a recent zebu influx from West Asia forming a group of small shorthorn Abyssinian zebu (DAGRIS 2007). In this region cross breeding with imported *Bos taurus* cattle is practiced. The central highland of the country therefore has three main cattle types, the Arsi breeds (Ethiopian zebu/Abyssinian zebu), crossbreds and Holstein-Friesian.

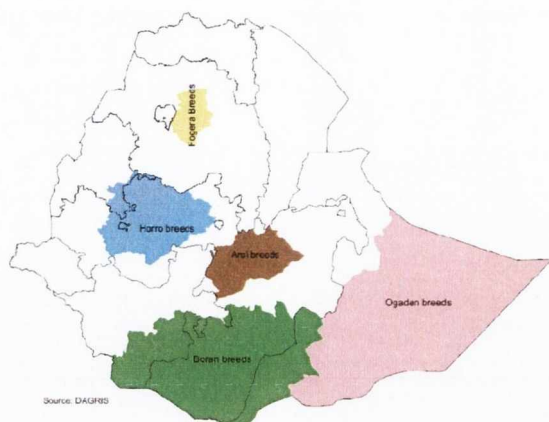


Figure 4.1. The distribution of major indigenous cattle breeds types in Ethiopia which includes Fogera, Horro, Arsi, Ogaden and Boran. The figure is taken from the Domestic Animal Genetic Resources Information System (DAGRIS) (MOHAMMAD 2007).

In recent history, exotic *Bos taurus* cattle, mainly Holstein-Friesian, were introduced to Ethiopia to increase livestock productivity in the country. The main focal distribution has been the central highland region where exotic breeds are used in intensive state and private farms as well as in small household cattle husbandry. Undoubtedly, productivity increases upon introduction of exotic *Bos taurus* cattle. However, indigenous cattle are well adapted to the local climatic conditions and relatively able to withstand endemic cattle diseases such as bovine tuberculosis and rinderpest.

Bovine tuberculosis is an economically important cattle disease. In Ethiopia the burden of the disease varies from place to place and husbandry settings (BERG *et al.* 2009). A report using intradermal comparative skin testing in various part of Ethiopia has found that overall bovine tuberculosis prevalence among indigenous zebu cattle was less than 5% (TSCHOPP *et al.* 2010). The disease showed higher reporting in the central highlands of the country where Holstein-Friesian and crossbreeds are more commonly used. Additional study from the central highland region of the country (Sellale) suggests that breed type

also plays a role in contracting tuberculosis infection (AMENI *et al.* 2007a). The disease affects Holstein-Friesian cattle more prevalently than the local Ethiopian zebu (Arsi) breeds. Crossbreds showed an intermediate level of incidence of bovine tuberculosis. The possibility of differential genetic susceptibility of *Bos indicus* versus *Bos taurus* cattle is documented for various tropical diseases including tuberculosis. To investigate the possible role of host genetics in bovine tuberculosis and other traits a sampling of hybrid animals was conducted for admixture mapping. In this chapter, the study population and field sample collection methods are described.

4.2 Materials and Methods

4.2.1 The study area

The study area selected for sample collection was located in the North West zone (Selalle region) in Ethiopia; its centred geographical coordinates are 9° 41' 0" North, 38° 8' 0" East (figure 4.2). The area is generally a highland plateau with an average altitude range between 2000 to 3500 meters above sea level. In this region small scale animal farming is widely practiced whereby dairy products are largely supplied to the capital city (Addis Ababa, located 120Km from Sellale town). The type of cattle in the region includes the local zebu (Arsi, Ethiopian zebu), Holstein-Friesian (European *Bos taurus*) and their crosses (zebu-taurus hybrid) (figure 4.3). These three types are found randomly distributed under the same type of cattle husbandry. In smallholder farming, which comprises more than 90%, farmers keep all breeds in the same herd. The average herd size among smallholder farmers is 10, of which half tend to be Holstein and/or crosses and the local zebu constitute the remainder. One of the very interesting features of the Sellale region is that it is possible to make a comparative study for various diseases of both veterinary and public health importance. It has also been possible to design genetic analysis models to investigate if diseases have skewed distribution between *Bos taurus* and *Bos indicus* cattle types.



Figure 4.2. A map showing the location of study site. Sellale region (B) is located in the North west of the capital city Addis ababa, Ethiopia (A).

4.2.2 The Study sample

The study design centred on a hybrid model (admixture mapping) that uses only crossbred animals. The selection process of each study animal involved an interview of the cattle owners to gather information about the genetic background of the candidate animal and a consensus vote among the field veterinarians to verify the information. A case record form was prepared for each sample and the following information was recorded.

- Full information and address of the owner
- Age, sex, physical condition score
- Breed history of the animal and if artificial insemination was used
- Picture of the animal on a side view with a study number showed on the photo which corresponds to the case record form

The inclusion criteria for each animal were to consider only hybrid animals with a breed history of at least two generations as indicated in figure 4.4. Hybrid animals from the first

generation that came from pure zebu (Ethiopian zebu) and taurus (Holstein) were excluded. Similarly if artificial insemination was used for an individual animal, it was also removed from the study group. After the animal breed history verification and selection, the body condition score was scored according to the standardized classification method (NICHOLSON 1986). If the animal is lean, it will be graded between 1 and 3, medium (graded between 4 and 6) and fat (7 to 9).



Figure 4.3. An example of one herd in Selalle region showing Ethiopian zebu, Holstein-Friesian and their crossbreds under identical husbandry conditions (A). The parental populations of the hybrid cattle are the European *Bos taurus* (Holstein-Friesian) pictured (B) and the Arsi bred (*Bos indicus*) an Ethiopian zebu (C).

4.2.3 Comparative intradermal tuberculin test (CIDT)

The hybrid animals were classified as reactor or non-reactor using a single comparative intradermal tuberculin test (CIDT) as a diagnostic tool for the presence of bovine tuberculosis infection. The test was performed via skin injection of purified protein derivatives (PPDs) which are crude extracts of *Mycobacterium bovis* (PPD-B) and *Mycobacterium avium* (PPD-A). Two sites on the mid-neck region of the animal were shaved approximately 12 cm apart from each other and the normal skin thickness was measured by a calliper. The antigens with each aliquot volume of 0.1 ml of PPD-B (2500 IU/ml, VLA) and PPD-A (2500 IU/ml, VLA) were injected into the sites as shown in figure 4.5. The skin thickness for each injected site were recorded after 72 hrs and compared to the original measurement taken prior to the skin test (figure 4.6). The animal were classified as reactor (CIDT positive) if the skin thickness increased by at least 4mm at PPD-B site than PPD-A taking the original skin thickness measure as a base line following the guideline of world organisation of animal health (OIE) standard (OIE 2004). In order to increase the sensitivity of the test without affecting the specificity, it is also recommended to consider an overall increase in PPD-B site greater than 2mm as a possible reactors animal in the Ethiopian context based on an extensive study in Selalle region (AMENI *et al.* 2010; AMENI *et al.* 2008). The animals were classified as non-reactors if the skin indurations measure at PPD-B site were below 2mm.



Figure 4.4 Photos of crossbred study population showing an intermediate taurus and zebu phenotypes.



PPD-A
PPD-B

Figure 4.5. Animal testing for bovine tuberculosis based on intradermal tuberculin skin reaction using *M bovis* (PPD-B) and *M avium* (PPD-A) antigens (A) injected at two sites of the neck (B).



Figure 4.6. Intradermal tuberculin testing result readouts after 72 hrs of injection. CIDT reactors showed induration at PPD-B site (A) and the non reactors showed no sign of skin thickness change for both injected sites (B).

4.2.4 DNA extraction

For the genetic analysis study 10 ml of whole blood was collected from the jugular vein of selected animals in two EDTA vacutainer tubes each containing 5ml volume. The blood samples were transported to from the field to Armauer Hansen Research Institute (AHRI) and kept at +4°C. DNA was extracted from 5ml blood using the ArchivePure DNA purification kit. The protocol followed assumed purification from 3ml whole blood after observing that most of the tubes contain less than 5ml. The purification steps briefly involved initial lysis of red blood cells and separation of pellet of white blood cells through centrifugation at 2000g. After suspending the white cells, a protein precipitation step followed and the supernatant containing the DNA was precipitated with 100% isopropanol and the pellet washed with 70% ethanol. A total concentration of 400µg/ml with a final yield of 100µg DNA was expected in this purification protocol.

4.3 Results

The field visits covered smallholder farming areas in the Selalle districts of Wuhcale and Degem. The total number of crossbred animals selected for screening by tuberculin test was 1705. The reactor animals in this group totalled 389 which gave 22% CIDT positivity. The number is higher than the prevalence report in the Selalle region (AMENI *et al.* 2008) for crossbred animals. In order to maximize the number of PPD reactor animals, the screening strategy used was to screen herds (identified by their owners) with previous reports of CIDT positivity. This study used information about the prevalence of bovine tuberculosis in the Selalle region between the years 2000 to 2007 based on more than 5400 animals (all breed types). This approach increased the number of cases obtained by 50%. Secondly, the value also increased due the lower cut-off value used for declaring CIDT positivity. When using >2mm skin indurations, a total of 91 extra animals grouped under cases than when using a >4mm cut-off value. Once the index cases were identified, a matched number of controls were assigned from non-reactor animals (CIDT negative). The inclusion criteria for the control group were to consider only animals that were found in a CIDT positive herd. The total number of controls included was 192. Table 4.1 summarises the general information about the selected 591 cases and controls for the subsequent genetic analysis. A total number of 508 samples were filtered from the initial 591 test groups that fulfilled the criterion for admixture mapping. The group contains 206 male and 385 female hybrid animals. The average body condition for the study group was very poor with the mean value score of 4 and the most frequent score was 3. Body condition classification categories of the animals are shown in figure 4.7.



Figure 4.7. Body condition score of the study samples classified as 1. Lean (< 3), 2. Medium (3 - 6) and 3. Fat (> 6)

Table 4.1: Summary of base line information about 591 cases and controls study samples

Factor	Category	Total No. (%)	Within Category CIDT+ (%)
Age	< 2	13 (2.2)	5 (38.5)
	2 - 5	273 (46.2)	153 (56.1)
	5 - 9	268 (45.3)	165 (61.6)
	> 9	37 (6.3)	23 (62.2)
Sex	Male	206 (34.9)	119 (57.8)
	Female	385 (65.1)	227 (65.6)
Body condition	Lean	266 (45.0)	159 (59.8)
	Medium	320 (54.1)	186 (58.1)
	Fat	5 (0.85)	1 (20)
Total (after quality control=508)	cases	346 (58.5)	346(58.5)
	controls	162 (41.5)	162 (41.5)

4.4 Discussion

The field sample collection successfully screened out 591 hybrid animals for admixture mapping analysis. These samples are appropriate for admixture mapping analysis assuming that accurate breed information was recorded from the interview and phenotype survey of the animal. This information was further checked for the study sample by calculating the estimates of generation time since admixture using a Bayesian likelihood posterior measure based on 768 SNP genotypes.

The figure showing 22% tuberculin positivity among the collected cross breed animals may not reflect the true prevalence data in the area. Previous reports in the Selalle regions showed that the CIDT positivity for cross-bred animals in the region is 11.9% while the figure was double for Holstein-Friesian (AMENI *et al.* 2007a; AMENI *et al.* 2007c; AMENI *et al.* 2008). The selected animals in this study were not based on a randomized sampling technique and also, an estimation of the burden of tuberculin positivity of the region was not the objective of the current study. The strategy followed here was to get a maximum sample size that fulfilled the selection criteria for admixture mapping of CIDT positivity. However, the cross sectional data does highlight the extent of bovine tuberculosis in the region and implications for animal and public health.

Single comparative intradermal tuberculin testing (CIDT) has long been used for an indirect diagnosis of bovine tuberculosis in cattle. Although the test sensitivity varies from place to places; for example in Europe tuberculin positive animals and confirmed cases of bovine tuberculosis are quite highly correlated (BERMINGHAM *et al.* 2009; SCHILLER *et al.* 2010c). The test efficacy in an African context could be different due to other factors such

as the presence of environmental *Mycobacterium spp.* whereby most animals are sensitized and probable host anergy might have decreased the sensitivity of CIDT. It was observed in the current study that it was very common to find PPD-A (*M. avium* reactors) in the population. Therefore, environmental *Mycobacterium spp.* has an impact on the overall decision of the CIDT status of the animal. Given the lack of accurate surveillance data on the transmission, it is difficult to correlate the number of reactor animals with *M. bovis* infection. The very nature of CIDT is that it captures the cellular immune repose in the context of delayed type hypersensitivity. During tuberculosis infection, it is well known that the cellular immune response is prominent at the early phase of infection. The later stage of the disease involves a typical upregulation of the humoral arm of immune response (BLANCO *et al.* 2009). As a result a good number of cases may be missed via CIDT. Newly developing diagnostic tools such as interferon gamma (INF- γ) (Bovigam) and lateral flow assay tests coupled with CIDT could enhance detection of infection at the wider range in the phase of infection (AMENI *et al.* 2010; SCHILLER *et al.* 2010b; SCHILLER *et al.* 2010c). In the context of this study, the specificity of the test measure in Sellale region was 96 to 99% (AMENI *et al.* 2008); the chance to assigning controls groups into cases was extremely rare.

This sample is the basis of the first study using a cattle admixture genetic mapping model to screen *Mycobacterium bovis* infected African cattle under natural conditions.

CHAPTER FIVE: GENOMEWIDE ADMIXTURE MAPPING SCAN FOR HOST SUSCEPTIBILITY TO BOVINE TUBERCULOSIS

5.1 Introduction

Bovine tuberculosis is a chronic debilitating respiratory infection caused by *Mycobacterium bovis*. The burden of bovine tuberculosis (BTB) remains significant on global agricultural industry and animal health. There are some reports that it can also be zoonotic which has a serious public health implication (Michel *et al.* 2010; Thoen *et al.* 2009). Infection in cattle leads to a spectrum of clinical outcomes which includes a protective cell-mediated immune response and immunopathogenic delayed type hypersensitivity (DTH) (POLLOCK *et al.* 2001b). The current testing strategies to identify infected cattle are the comparative intradermal tuberculin skin test (CIDT), the interferon gamma (INF- γ) assay and postmortem tuberculosis lesion detection in abattoir samples (SCHILLER *et al.* 2010a). Control of BTB infection requires understanding of the complex interaction between the host and the pathogen. There is a reasonable assumption that host genotype may play a role in disease outcome. Particularly, a comparative analysis of the genetic susceptibility patterns between the two divergent cattle subspecies, *Bos taurus* and *Bos indicus*, has some suggestion of differential disease risk (COLLINS-SCHRAMM *et al.* 2002). Although the observations vary among epidemiological zones, there is further evidence that host genetics play an important role (Allen *et al.* 2010). For example, the level of host susceptibility may vary within a single breed as studies conducted on Holstein-Friesian dairy cattle in Ireland suggest a heritability estimate up to 0.18 for confirmed BTB infected cases (Bermingham *et al.* 2009). Reports from the central highlands of Ethiopia, including Sellale where BTB is endemic, have shown the prevalence and pathology of the disease significantly skewed towards the *Bos taurus* (Holstein-Friesian) as compared to the local *Bos indicus* (Ethiopian zebu) cattle (AMENI *et al.* 2007a). Although cattle husbandry

practices have a pivotal role, here tuberculosis lesion severity and INF- γ responses were higher for European taurus than zebu cattle (Ameni *et al.* 2006). A consistent finding was reported (Demelash *et al.* 2009) from post-mortem examination of abattoir samples in Ethiopia that a higher prevalence bovine tuberculosis was found in European *Bos taurus* than *Bos indicus*.

Recently admixture mapping has gained attention in the context of finding genes associated with complex traits (Darvasi and Shifman 2005). It is a method that utilises recently admixed populations to localise disease susceptibility loci through the examination of admixture generated linkage disequilibrium. In this approach potential risk alleles diagnostic of one or other ancestral strain are detected if they are found more frequently among the affected admixed individuals than the healthy controls (Hoggart *et al.* 2004; Zhu *et al.* 2008). Admixture mapping analysis has been found to be more economical and powerful as compared to the exiting methods such as linkage analysis and association studies. Linkage analysis, which is based on pedigree mapping detects only moderate effects of major genes. Genomewide association analysis has good resolution power but needs an extremely dense coverage of makers and involves very large study sampling (Winkler *et al.* 2010; Zhu *et al.* 2004). Admixture mapping on the other hand has more power, does not need family based sampling, requires 200 to 500 fold fewer markers than normally needed, allows a flexible study design such that case only and/or case control can be used. The application of the model has been demonstrated in human populations by detecting ancestral associations to various complex traits.

Admixture scans for human diseases such as hypertension (Cheng *et al.* 2010b), diabetics (Elbein *et al.* 2009), prostate cancer (Bock *et al.* 2009), breast cancer (Fejerman *et al.* 2009),

end stage renal dysfunction (Kao *et al.* 2008; Kopp *et al.* 2008) and obesity (Basu *et al.* 2009; Cheng *et al.* 2009; Cheng *et al.* 2010a) have detected new susceptibility loci and also confirmed previously known candidate genetic regions. In these studies ancestral ethnic groups are identified as reference populations and used to identify the ancestral allele frequencies of informative markers for testing admixed populations for disease associations with ancestral origin. For example, for a recently admixed population such as African Americans, a two way admixture between the parental populations of West African and European Caucasian applied to 1500 to 2500 informative markers is sufficient for genomewide admixture scan. The sample size in such a study is dictated by the risk of a particular phenotype under investigation but is significantly fewer than the number needed in the case of genomewide association. The general guideline is that for cases where the risk due to one of the contributing ancestral population increases from 1.2 to 2, the admixture mapping requires a total sample size of 250 for the maximum risk model and 2500 for the smallest risk model (Winkler *et al.* 2010). A similar strategy can be implemented to detect ancestral associations of complex traits such as bovine tuberculosis among cattle populations.

The higher prevalence of bovine tuberculosis for within European *Bos taurus* cattle compared to the *Bos indicus* in an African context, makes admixture mapping a promising analysis. With genomewide SNP available for diverse cattle populations around the world, ancestral allele frequencies can be derived for a particular test populations and the selection of ancestry informative markers has become practical. The previous chapters have shown the practical use of Bovine Hapmap consortium and Illumina 50k SNP chip for selection of powerful informative SNPs for understanding of genetic selection and population structure within cattle populations. The current chapter utilises informative

genetic markers to highlight ancestral associations with bovine tuberculosis by applying an admixture mapping model.

5.2 Material and methods

5.2.1 Study populations

The study was carried out by screening 582 hybrid cattle in Ethiopia. The ancestral populations for these cattle were the local Ethiopian zebu and European Holstein-Friesian and genotypes were available for both (chapter 3). Visual observation and information obtained through interviews of the owners were used to include only crossbred animals. Furthermore, an attempt was made to screen out first generation crossbred animals and those that came from artificial insemination. The total analysis was carried using 389 comparative intradermal tuberculin (CIDT) test positive (reactors) and 193 CIDT test negative (non-reactors) animals.

5.2.2 Marker panel

The samples were typed using the Illumina BeadLab platform and 768 SNPs in the Golden Gate assay previously described in chapter 3.

5.2.3 Estimates of Ancestral allele frequency

The cattle breeds typed and analyzed in chapter 3 include European Holstein, Ethiopian zebu and Kenyan Boran which provided population samples within which to estimate ancestral allele frequencies. For this analysis, 44 Holstein and 44 African zebu (Ethiopian zebu and Boran) were included.

5.2.4 SNPs and study sample quality control

Iterative filtering of poor quality SNPs and samples were carried out using the feature implemented in the ANCESTRYMAP package (GREENSPAN *et al.* 2004). First, SNPs

missing with genotype data in the cases and controls were removed. Since an admixture scan is sensitive to data inconsistencies, it implements a series of inbuilt data checks. These include “mapcheck” that tests whether ancestry estimates obtained for each marker is correlated to the information obtained by adjacent markers. “freqcheck” which tests if the estimated ancestral frequency of an allele from the MCMC run fits with an observed frequency of a SNP in the crossbred animals. “leave1out” which removes the markers contributing most to the association and assesses whether the signal persists and “checkdup” which checks individual samples and removes these if they are duplicates or possibly related samples. Individuals with more than 80% genotype similarity were assigned to include only one sample. Finally after the exploratory run, case-control samples with an estimated European or African ancestry of more than 0.90 were excluded from further analysis.

5.2.5 Genomewide Admixture mapping scan

Local and genomewide ancestry estimates within cases and controls were achieved by ANCESTRYMAP which uses a Hidden Markov model (HMM) to combine the weak information about local ancestry that is provided by each marker, into a more confident estimate that takes into account the information from neighbouring markers. The HMM is nested within a Markov Chain Monte Carlo (MCMC) process, which accounts for uncertainty in the hidden ancestry state parameter. All MCMC runs used 100 burn-in and 200 iterations. Finally, the stability of the result was monitored through increasing the running parameters by a factor of 10. Therefore, a longer run of 1000 burn-in and 2000 follow-on iterations was carried out to check signals against the original parameters. A second Bayesian based admixture scan was also performed using ADMIXMAP (Hoggart

et al. 2004). It is a hybrid of Bayesian and classical approaches that utilizes MCMC methods to jointly estimate the admixture proportions for individuals and control for admixture induced confounding in association studies. Briefly, a Bayesian model is used to calculate the admixture values for individuals, based on the supplied prior distribution of allele frequencies which was specified by adding 0.5 to the observed counts in the ancestral populations. The posterior distribution of admixture and the ancestry specific allele frequencies at each locus, given the observed genotype and CIDT data, is generated by MCMC simulation. Linkage was assumed and analyses were conducted using prior allele frequencies of the founding populations together with pre-assigned number of populations ($K=2$). ADMIXMAP was performed using the default setting that allows for the specification of assortative mating populations and an uncorrelated allele frequency model. A triplicate burn-in of 400 and 1000 iterations were used and a final 10X run parameter was applied to check the stability of the signal.

5.2.6 Statistical analysis for association of CIDT positivity and *Bos taurus* ancestry

ANCESTRYMAP calculates individual ancestry estimates averaged across all individuals to identify genomic regions where there is enhanced ancestry from one of the parental populations, indicating the presence of an ancestry-associated disease gene nearby. For the binomial admixture scan a prior 27 risk model distribution assigned from 0.4 to 3.0. The overall association was presented by averaging across all the models. The model specified 0.4, 0.5, 0.6,.. 3.0 fold increased risk due to inheritance of one copy of an European ancestral allele for cases, with a control risk of 1 (GREENSPAN *et al.* 2004). In effect this model tests the association of CIDT positivity with European ancestry. The assumption

arises from previous studies both in the field and host immune response observations. The local Ethiopian zebu animals have been found to have at least 2 times reduced incidence compared to Holstein-Friesian animals. Cross-breeds were found to have intermediate levels of risk between the two pure populations for bovine tuberculosis under similar husbandry conditions (AMENI *et al.* 2007a).

The association was quantified based on the outputs of two scores from ANCESTRYMAP. First the locus-specific score is a case-only statistic where the likelihood of the genotype data at the locus under the risk model is compared to the likelihood of the genotype at the locus assuming that the locus is uncorrelated to the genotype. The log base 10 of the ratio of these two likelihoods gives a LOD score. A general recommendation for genome wide significance is $\text{LOD} > 5$ and $\text{LOD} > 4$ can be assumed to be genome wide suggestive. In line with this calculation, ANCESTRYMAP provides an overall account for association scores by taking the averaged likelihood ratio for associations across all loci in the genome and summarizing evidence of a risk locus anywhere in the genome. In this case, a genome wide score > 2 is taken as significant and a value > 1 as suggestive significant score (GREENSPAN *et al.* 2004).

The second score for an admixture scan is the case-control score calculated by comparing the locus specific deviations in European ancestry in cases versus controls at each locus across the genome. The score is taken as a z-score assuming that if there is no disease association, the score is expected to distribute under normality. The level of locus specific case-control statistical significance is taken as a z-score > 3 which correlates to uncorrected nominal $P < 2 \times 10^{-3}$.

A second independent Bayesian based and more computationally intensive admixture mapping method implemented in ADMIXMAP (Hoggart et al. 2004) was also used to investigate the consistency of the result. The output uses similar scoring methods described in ANCESTRYMAP for case only (affected only analysis) and case control analysis (ancestry association score). The final result was reported by Z- score and the corresponding P value for testing level of significance.

5.3 Results

The final set of markers after removing SNPs with low call rate, thinning the densely typed BTA6, and passing the internal quality control in ANCESTRYMAP includes 662 autosomal markers. The allele frequency of reference populations, zebu (Ethiopian zebu and Boran) and taurus (Holstein) used in the study are plotted in the figure 5.1. The plot indicates the predominance of high allele frequency differentials between the ancestral populations.

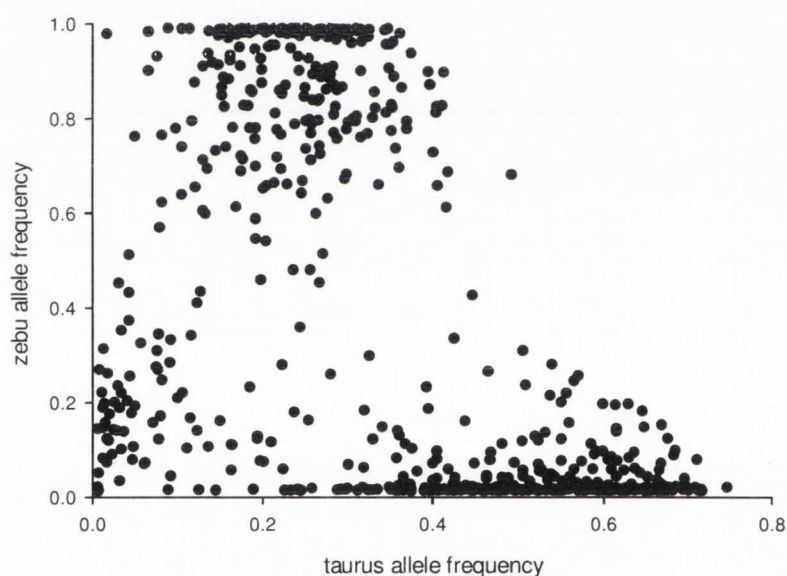


Figure 5.1. Allele frequencies within ancestral reference populations: zebu, Ethiopian zebu and Boran plotted on the y axis; Vs European taurus, Holstein plotted on the x axis. Note that most SNPs are either high frequency zebu and low frequency taurus or vice versa.

The admixture scan was based on a final set of 346 cases and 162 controls that fulfilled the entire selection criteria. Samples with an estimated percentage of ancestral zebu and taurus greater than 0.90 numbered 67 and were removed from further analysis. 16 samples were excluded due to very low genotype call rates, duplicates and closely related samples. In order to check the homogeneity of the cases and controls, EIGENSTRAT was used to verify that the final set of samples were all in one cluster. There was no visible population sub-structuring in the test population that needed to be accounted for the admixture scan as shown in figure 5.4. Furthermore, a second EIGENSTRAT analysis was constructed for the samples to ascertain population variation in relation to the parental populations. This principal component plot is shown in figure 5.5 and indicates clearly the relationship between the admixed and the parental European taurus (HOL) and the East African zebu (ETZ, BRN). The plot also shows that the crossbred animals have dominant zebu ancestry since most of the individuals are concentrated in close proximity to zebu parental populations. Again, no visible sorting of cases and controls was evident.

The estimates of average genomewide ancestry also support the visual observation of the principal component plot. The average taurus ancestry estimates for the study sample was 0.33 and the estimate was only slightly greater for cases than control groups as shown in figure 5.2. Prior to the admixture scan a simple regression analysis was performed to see if there is any obvious trend of association between the level of average taurus ancestry and the skin positive test for bovine tuberculosis. The relationship is as shown in figure 5.3 and there is a weak positive linear correlation between an individual taurus ancestry and the level of skin PPD readout. The relationship was not statistically significant with

$P=0.082$. The possible ancestry stratification between cases and controls can be viewed as a variation in individual admixtures which could be a source of statistical power for ancestry correlated phenotype or a confounder in causing false-positive results. Note, the central design of admixture mapping model is to detect genes that underlie ancestry differential in disease risk and control for stratification in cases-controls (Darvasi and Shifman 2005; Halder and Shriver 2003; Hoggart et al. 2004; Pritchard and Rosenberg 1999; Rosenberg and Nordborg 2006).

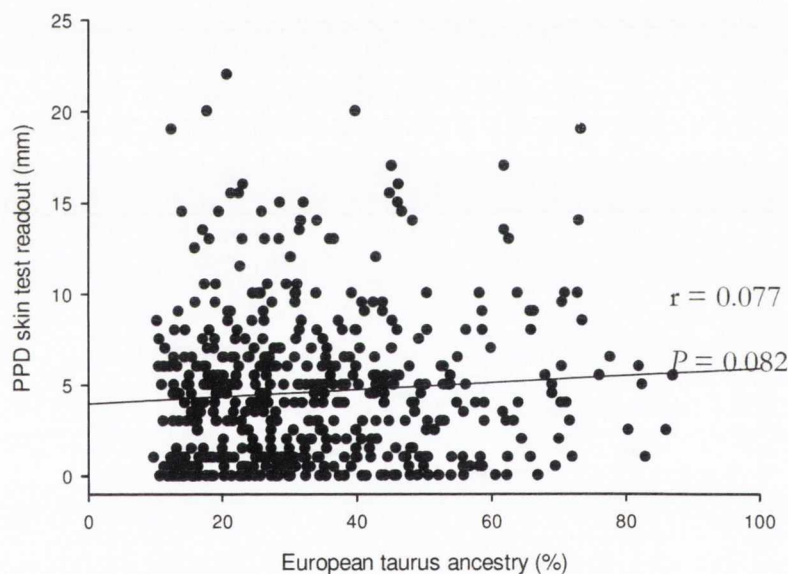


Figure 5.3. The scatter plot of comparative intradermal tuberculin test (CIDT) readouts compared to the estimated *Bos taurus* ancestry for cases and controls.

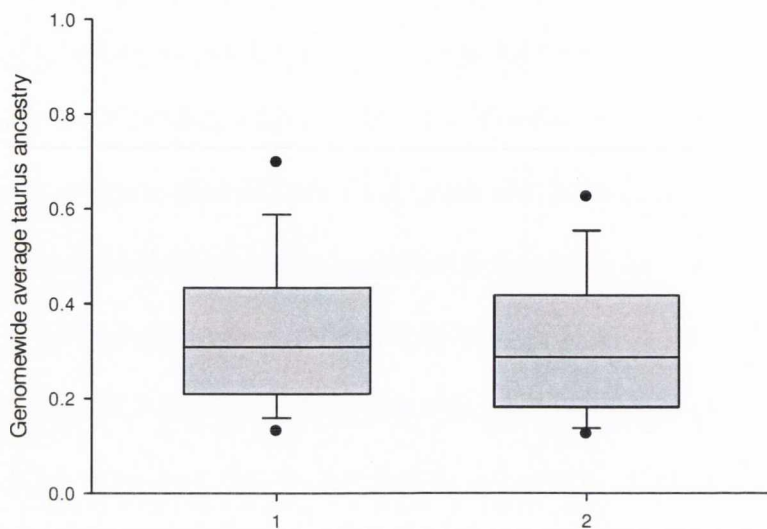


Figure 5.2. Box plot showing the genome wide average estimate of *Bos taurus* ancestry for cases (1) and controls (2). The bottom and top dots for the distribution indicates the 5 and 95 percentile values of taurus ancestry contributions.

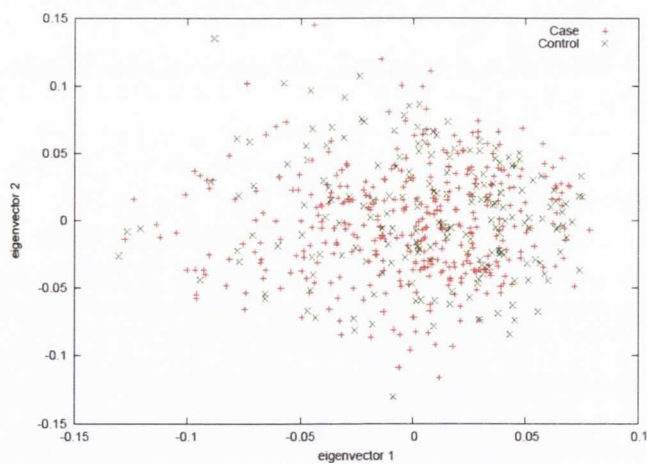


Figure 5.4. The principal component analysis plot distributions of cases and controls

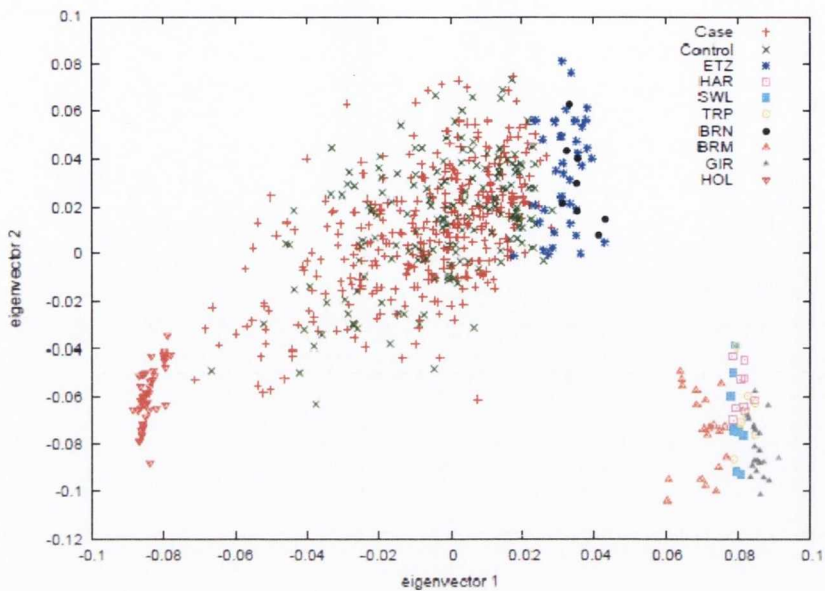


Figure 5.5. Principal component analysis of cases and controls including the ancestral populations of European taurus (Holstein, HOL), East African zebu (Ethiopian zebu(ETZ), Boran (BRN)), and non African zebu (Hariana (HAR), Sahiwal (SWL), Tharparker (TRP), Brahman (BRM), Gir (GIR)).

5.3.1 ANCESTRYMAP admixture mapping scan of tuberculin positivity

The association between ancestry and tuberculin reaction was tested using ANCESTRYMAP software. The program requires specification of risk models before carrying out the analysis. As shown in figure 5.6 equally likely risk models ranging from 0.4 to 3 fold increased risk for European taurus (Holstein) ancestry was tested. The maximum log-likelihood ratio (LOD) was 3.37 for a risk value of 1.6 which is attributed to having one copy of taurus allele. Furthermore, log-likelihood of the locus genome statistic averaged over all the markers in the genome was 2.57. The overall risk model analysis has given an evidence of association between ancestry and tuberculin positivity. The analysis was further extended to highlight locus specific scores for the binary

outcome which was done for case only analysis (LOD score) and case control analysis (z-score).

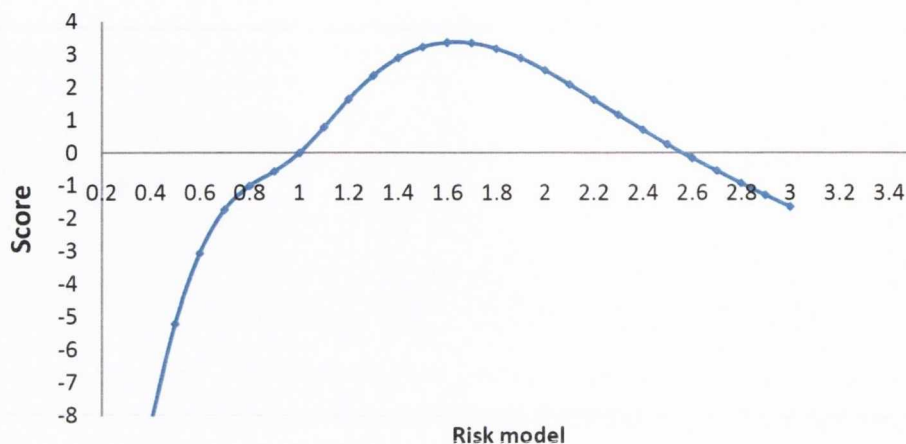


Figure 5.6. ANCESTRYMAP risk model (x-axis) plotted Vs genomewide average LOD scores. The maximum likelihood risk model is 1.6 with a genomewide score LOD of value 3.37 (y-axis). The average LOD score for all risk models is 2.57.

The admixture mapping scan based on the cases and controls for the CIDT skin testing is plotted in figure 5.7. Values for both the case control comparison (z-score) and case only analysis (LOD score) are given. This detected top signals for genome wide association at BTA-5, BTA-6, BTA-11, BTA-12 and BTA-19 shown in figure 5.7. The strongest association found for case-control comparison (z-score) was on BTA-6 for the marker BTA76573 with a z-score value of 3.03 and corresponds to a nominal $P=2.4 \times 10^{-3}$. This locus also has a strong suggestive admixture generated signal for case only analysis with a LOD score of 4.52. The second high score was on BTA-5 for the marker BTA149284 and z-score 2.92 ($P=3.5 \times 10^{-3}$); here the LOD score for case only analysis was not significant. With case only analysis the highest signal observed was for the maker

rs29019760 on BTA-11 which corresponds to a significant locus specific LOD=6.10. The peaks on BTA-12 for the marker BTA152656 and on BTA-19 for BTA133013 gave suggestive evidence of association with LOD scores of 3.86 and 3.79 respectively. The summary scores for each chromosome are shown in Table 5.1. The case control and case only analyses generally follow similar patterns as shown in the plot. The strongest peak with corroboration from both analyses was found on BTA6. The other selected regions were supported by either case control (BTA5) or case only analysis (BTA11, BTA12, and BTA19).

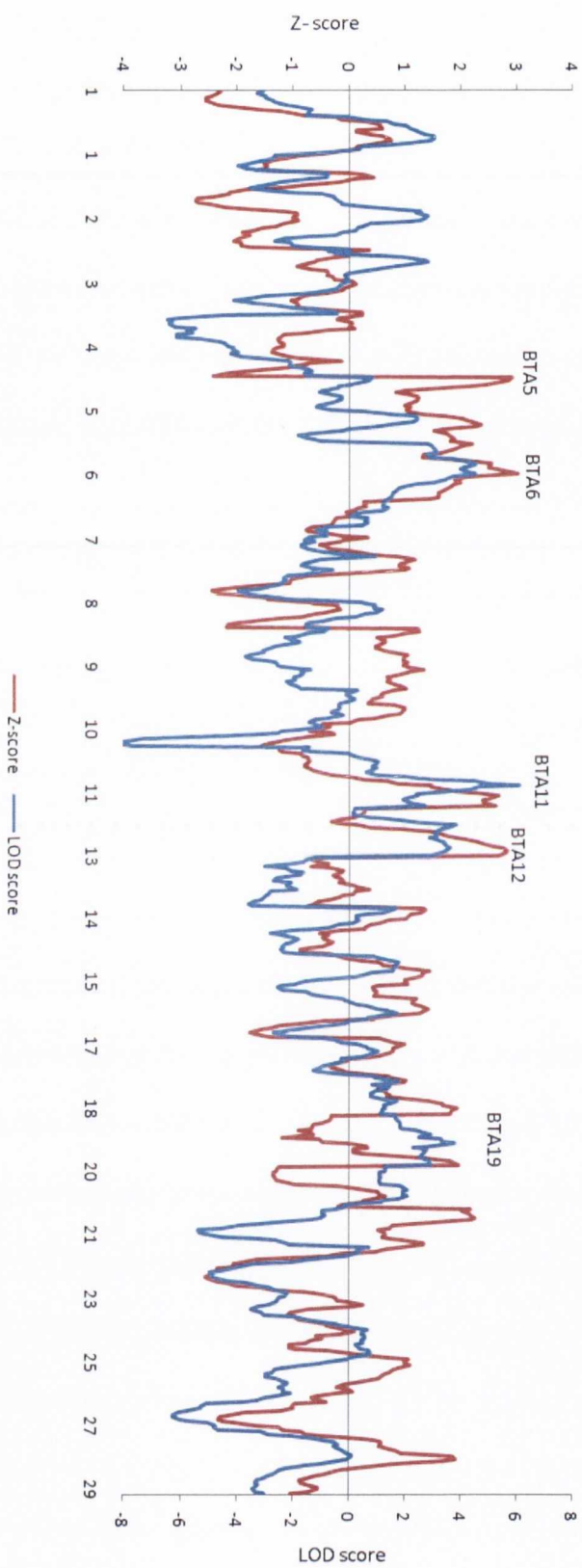


Figure 5.7. The ANCESTRYMAP genome-wide admixture scan for case-control analysis (z-score) and case only score (LOD) association of comparative intradermal tuberculin test (CIDT) positivity with ancestral origin. The case control analysis has no significant negative z-score value.

Table 5.1: Summary report of ANCESTRYMAP scores for case only (LOD), case control z-score (CCS) and average LOD score for each chromosome (Chr_Local).

BTA	LOD_MAX ^a	CCS_MAX ^b	CCS_MIN ^c	Chr_LOCAL ^d
1	3.16	0.77	-2.56	2.12
2	3.07	0.43	-2.75	2.08
3	2.87	0.31	-1.02	1.65
4	-1.25	0.17	-2.39	-2.10
5	2.51	2.91	0.85	1.65
6	4.54	3.04	1.34	3.89
7	1.43	0.82	-0.75	0.57
8	1.21	1.17	-2.44	0.35
9	-0.72	1.35	0.48	-1.62
10	0.31	1.02	-1.53	-0.42
11	6.02	2.67	-1.23	4.63
12	3.83	2.85	-0.28	3.25
13	-1.16	0.37	-0.72	-1.88
14	1.69	1.37	-0.86	0.57
15	1.79	1.45	-0.12	1.15
16	1.70	1.43	-1.75	0.85
17	1.04	1.00	-0.34	0.43
18	1.81	1.96	0.42	1.30
19	3.75	1.97	-1.10	2.92
20	2.09	0.65	-1.34	1.71
21	-0.48	2.28	0.54	-1.22
22	0.76	0.91	0.33	-2.46
23	-2.17	0.27	0.33	-2.83
24	0.80	0.22	0.34	-0.05
25	0.09	1.10	0.34	-1.68
26	-2.06	0.57	0.33	-3.08
27	-1.99	-0.46	0.32	-4.83
28	0.14	1.94	0.35	-0.35
29	-1.00	0.85	0.34	-2.73

^a Maximum locus genome statistic score

^{b,c} The maximum and minimum case control statistic scores

^d Log likelihood of the locus genome statistic score obtained by averaging over all the markers on that chromosome

5.3.2 ADMIXMAP analysis of cases and controls for tuberculin skin test as a binary and continuous variable

An additional alternative Bayesian likelihood method was applied that is implemented in ADMIXMAP for checking the consistency of the result under a binary model and also to extend the analysis using the tuberculin skin test results as a continuous variable. The application also implements the use of covariates to adjust the admixture scan through multiple regression analysis. The admixture scan was undertaken assuming multiple outcomes through regression on tuberculin skin test as a binary outcome and also the diameter of tuberculin readout as a continuous variable. The parameters were adjusted for age, sex and body condition of individual test animals. This analysis supports the results obtained from ANCESTRYMAP as shown in figure 5.8 (binary outcome) and figure 5.9 when applying a continuous multiple outcome and adjusted for covariates. The results from binary and continuous models were very similar. For the purpose of comparison between ANCESTRYMAP and ADMIXMAP, the binary model scores were referred. ADMIXMAP analysis also provides case control scores and case only scores and all the significant or suggestive signals for possible association for skin test positive animals remain consistent with the ANCESTRYMAP outputs. It is important to note here that unlike ANCESTRYMAP where the case only analysis output was given in the context of LOD score, ADMIXMAP gave a final inference about the level of association for case only analysis under a standard normal distribution score (z-score). Therefore, the numerical values of case only analysis obtained from ANCESTRYMAP and ADMIXMAP may not match directly. But after separate interpretation, the two tests could be compared using the level of statistical significance in a region.

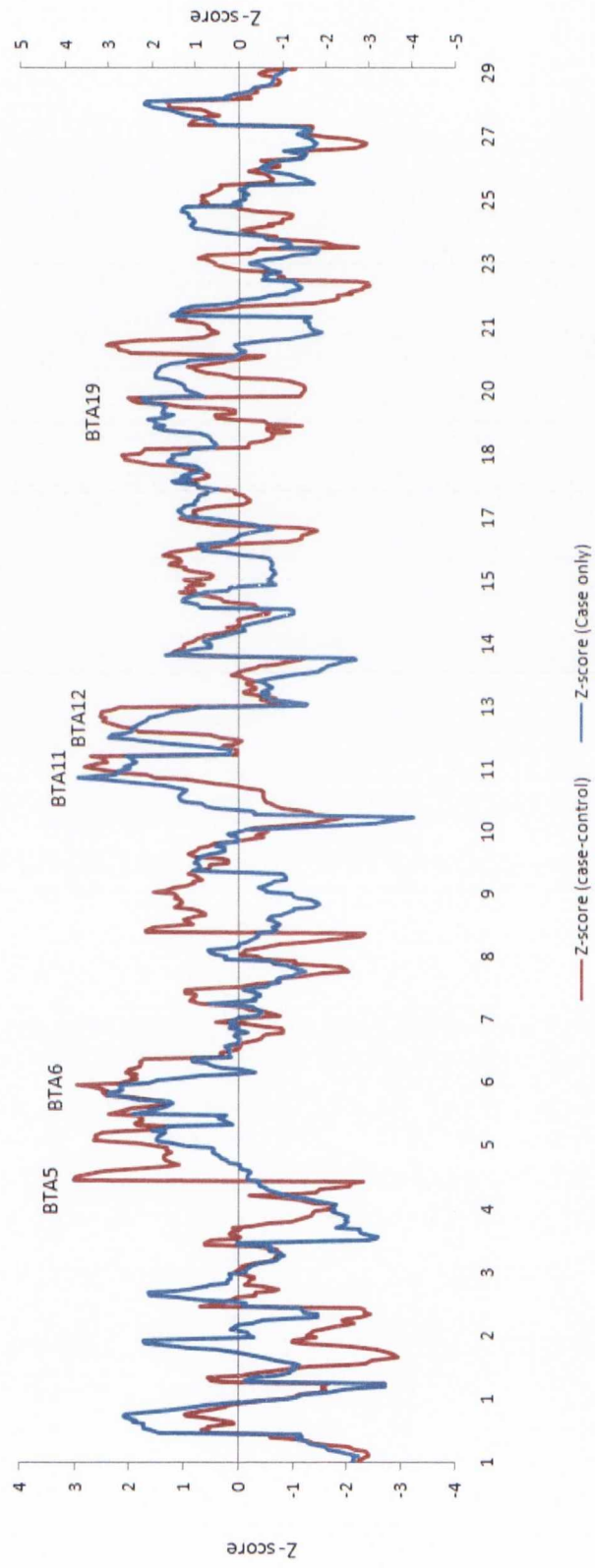


Figure 5.8. ADMIXMAP analysis using cross sectional and binary outcome of comparative intradermal tuberculin test (CIDT) positivity.

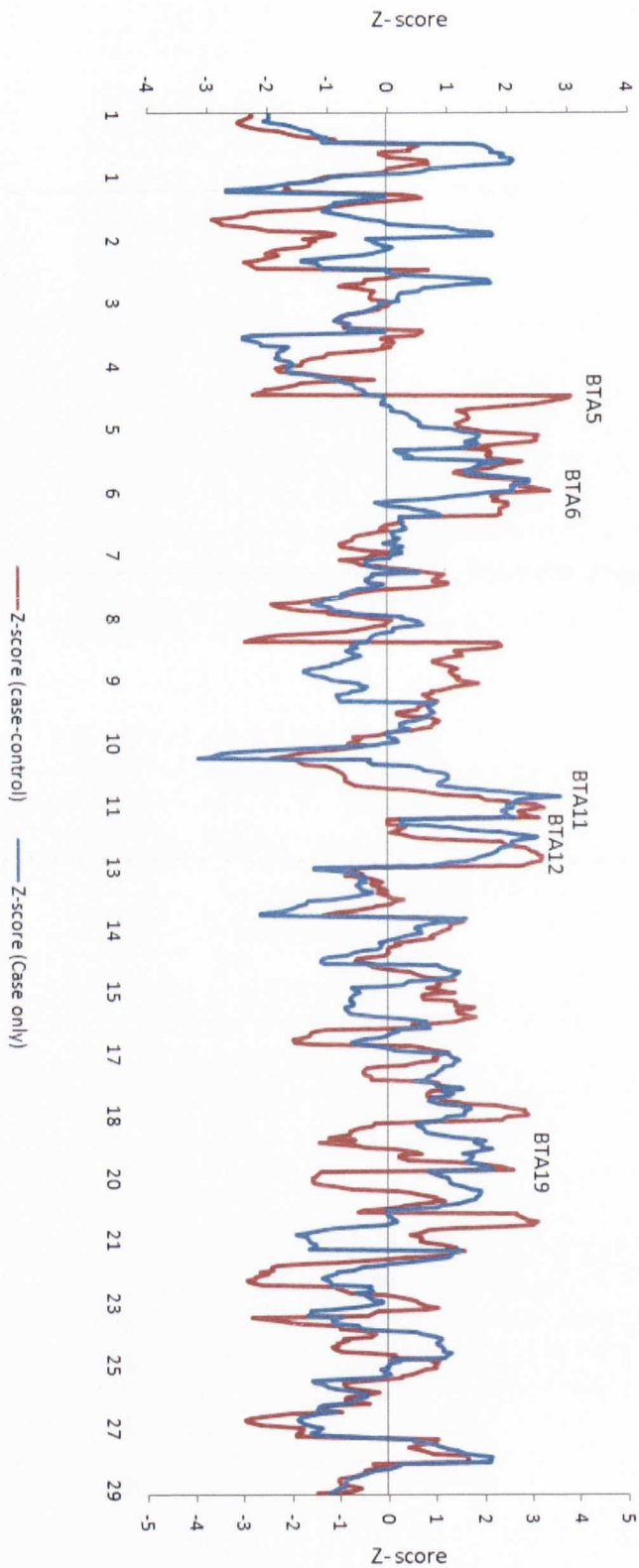


Figure 5.9. ADMIXMAP analysis using cross sectional and multiple outcome parameter accounting the binary parameters (case control) and comparative intradermal tuberculin skin test readouts as continuous variable.

The case control analysis (described as ancestry association score in ADMIXMAP) gave a top z-score in BTA-5 (BTA149284) of 3.05 ($P = 2.3 \times 10^{-3}$) and BTA-6 (BTA76573) with z-score 2.69 ($P = 7.2 \times 10^{-3}$). The case only analysis (affected only score) gave a peak admixture value of on BTA-11 (rs29019760, z-score = 3.69) and BTA-12 (BTA152656, z-score = 2.98) which were both similar to the ANCESTRYMAP values. An exception was BTA-19 (BTA133013) where the value obtained had a weak statistical significance (z-score = 2.14, P value = 3.2×10^{-2}) and hence did not replicate (Table 5.2).

Table 5.2: The top markers based on either LOD or z-score values from ANCESTRYMAP analysis and the replication of the result using ADMIXMAP.

BTA	Peak marker	Position (bp)	ANCESTRYMAP		ADMIXMAP (binary, z-score)		ADMIXMAP(cont, z-score)	
			local LOD	z-score	Affected	Case control	Affected	Case control
5	BTAA149284	2436485	0.84	2.92	-0.18	3.02	-0.07	3.05
6	BTAA76573	65162299	4.52	3.03	2.68	2.96	2.58	2.69
11	rs29019760	64371895	6.10	1.48	3.69	1.55	3.58	1.51
12	BTAA152656	35296048	3.86	0.15	2.98	0.34	3.09	0.43
19	BTAA133013	35619640	3.79	0.03	2.14	0.11	2.14	0.13

5.3.3 Testing locus specific association with tuberculin skin test positivity

The local estimate of taurus ancestry was extracted for all cases and controls at the top admixture markers on BTA-5, BTA-6, BTA-11, BTA-12 and BTA-19. In order to investigate if local ancestry has an effect on tuberculin positivity, transformed values of standardized PPD skin readouts with individual animal genome wide ancestry interactions were investigated as covariants. A multiple logistic regression model was applied to determine if there was evidence for residual association with local ancestry after adjustment for global ancestry. The binary variable of either case or control was assigned as the dependant variable. The local taurus ancestry for each animal, expressed as the presence of 0, 1, or 2 allele copy for the specific makers, was used as the independent variable. The transformed quantitative skin test, adjusted for global genome wide ancestry, was used as a covariate in the logistic regression model. Consistent residual association of local ancestry with tuberculin reactor animals was observed for markers on each chromosome BTA-6 (BTA76573), BTA-11 (rs29019760), BTA-12 (BTA152656) and BTA-19 (BTA133013) as shown in table 3. For all of these markers the regression coefficient was a positive value, indicating that having additional copies of the taurus allele increases the likelihood of being tuberculin positive. This may roughly be interpreted a higher chance that the animal is susceptible to bovine tuberculosis. The highest test of significance for the multiple logistic regression analysis was for rs29019760 (1.7×10^{-5}) on BTA-11 which coincides with the trend observed for LOD scores from the ANCESTRYMAP analysis. The test did not show significant association for BTA-5 (BTA149284) with a P value of 0.194 which did not give a significant LOD score value from the admixture scan that may be due to the local ancestry (0.35) was low when

compared to the average genome wide ancestry estimate for all the cases (0.36). Although the case control analysis indicated a significant value, it was not consistent with a case only analysis.

Table 5.3: Multiple logistic regression analysis of five SNPs showing the strongest genetic background dependant effects on CIDT positivity based on admixture mapping.

BTA	Peak marker	Position (bp)	Reg.coef. ^a	OR (95%CI)	P value
5	BTA149284	2436485	-0.28	0.75 (0.49 - 1.16)	0.19
6	BTA76573	65162299	0.63	1.89 (1.23 - 2.88)	3.69x10-3
11	rs29019760	64371895	0.95	2.59 (1.68 - 3.99)	1.70x10-5
12	BTA152656	35296048	0.59	1.81 (1.18 - 2.78)	6.79x10-3
19	BTA133013	35619640	0.67	1.96 (1.27 - 3.00)	2.19x10-3

^a Reg.coef. Regression coefficient

5.3.4 Possible list of associated genes for host immunity to infection

The chromosome regions that contain the five top admixture scan findings were examined to list the possible genes associated with the disease. The Baylor 4.0/bosTau4 build of the bovine genome data base from UCSC browser (<http://genome.ucsc.edu/>) was utilized to extract genes from the associated region. It was identified by using the top admixture scan as a reference point and spanned around to include approximately 95% CI of the extrapolated likelihood curve found by taking 1-LOD support interval (Manichaikul et al. 2006). The list of genes harboured in the top candidate regions are presented in table 5.4.

The strongest case control ($P=2.4 \times 10^{-3}$) and case only score ($LOD = 4.52$) has a candidate region of 11.7 Mb (53,432,843 - 65,162,299) and contains the toll-like receptor genes TLR1, TLR6, TLR10 and the RHOH gene. TLRs are important in bacterial pattern recognition and are known to be cellular receptors that play pivotal roles in host innate immune responses (Khor 2009; Ryffel et al. 2005). RHOH has a role in the development of T cell lymphocytes via modulation of T cell receptor signaling (Wang et al. 2011).

The chromosome 11 admixture peak which has the highest LOD score of 6.10 but a non significant z-score for case control analysis stretched in a large segment from 56,910,615 to 81,528,883 and covered 24 Mb in total. The region contains genes such as PELI1 (Pellino homolog 1) which is a ubiquitin ligase needed for the transmission of TRIF-dependent TLR signals making it vital for innate immunity and inflammation in general (Chang et al. 2009).

Although not a very strong admixture peak, BTA19 which covers 32,317,789-35,619,640 Mb, has a signaling molecule gene ADPORA2B. This takes part in the regulation of T-lymphocytes which results induced deactivation cell death (Himer et al. 2010).

Table 5.4: The list of genes extracted in the credible interval region of the five top chromosome regions.

Reference Genes					
BT A	Peak marker	Reference Position (bp)	Upper and lower boundary of physical position	LOD (ANC)	z-score (ANC)
5	BT A149284	2436485	2436485 - 8427297	0.84	2.92
ATXN17B, GLIPR11, GLIPR1, KRR1, PHIDA1, NAPI11, CSR2, NAV3					
6	BT A76573	65162299	53432843 - 65162299	4.52	3.03
LOC514424, RELL1, TRC1D1, TLRI0, TLR6, TLRI, TMEM156, RPL9, LIAS, UGDH, G6H4orf34, UBE2K, RHOF, CHRNA9, APBB2, LCHL1, LMCH1, BEND4, ATP8A1					
11	rs29019760	64371895	56910615 - 81528883	6.10	1.48
CTNNA2, REG3G, REG3A, AHS42, XPO1, CCT4, COMMD1, B3GNT2, TMEM17, MDH1, UGP2, VPS54, PELL1, SLCA14, RAB1A, ACTR2, HDAC8, MEIS1, ETAA1, CID, WDR92, PNQ1, PPP3R1, CNRIP1, PLEK, APLF, PROKR1, ARHGAP25, BAP10, GKN1, GPTT1, NEU1, ANXA4, GMCL1, SNRNP27, MXD1, PCBP1, C11H2orf42, TIA1, PCYOX1, SNRPC, EHD3, CAPN14, GALNT14, CAPN13, LBA, YPEL5, GLIP4, SPDYA, PPP1CB, FOSL2, BRG, RBKS, MRPL33, SUPTL2, GPN1, ZNF512, FNDC4, KRTCAP3, NRBPI, PPATG, ZNF513, SNX17, EHE2B4, NPV17, UCN, TRIM54, DNAGC5G, SLCA3A3, APR3, SLCA5A6, TCF23, PREB, ABHD1, CGREF1, KHK, EMLIN1, AGBL5, TMEM214, MAPRE3, DPYSL5, MGCI39367, CIB4, MGCI28001, OTOF, C11H2orf49, EPTT, HADHB, HADHA, RAB10, KIF3C, DTNR, MIR1301, POMC, DNAGC27, CENPO, LOC617088, MGCI33709, LOC100125925, PFN4, TP53I3, SF3B14, FKBP1B, UBXN2A, KLHL29, C2orf43, RHOB, SDC1, LAPTM4A					
12	BT A152656	35296048	32537680 - 43406630	3.86	0.15
USP12, RPL21, WASF3, RN16, SHISA2, ATP8A2, NUPL1, MTNR6, SPATA13, C1QTNP9, TNFRSF19, SCGC, EPHA1, ZDHHC20, MIRP3, SKA3, SAPI8, N6AMT2, IFT88, GJB6, GJB6, CRY11, GJB2, GJA3, ZMYM2, ZMYM5, PSPCI					
19	BT A133013	35619640	32317780 - 35619640	3.79	0.03
HS3ST1A1, COX10, HS3STB1, PMP22, TEK13, CDR14, FAM188, ZNF286A, TRPV2, LOC281370, PIG1, NCOR1, ZSWIN7, ADORA2B, AKAP10, ULK2, ALDH3A1, SLCA7A2, ALDH3A2, RNFI12, MFAP4, MAPK7, BBD1, GRAP, SLCA5A10, PRPSAP2, SHMT1, TOP3A, SMCRT, ILG11, DRG2, ATPAF2, LRRC48, TOMH12					

5.4 Discussion

In this chapter whole genome admixture mapping was carried out to search for genomic regions associated with the comparative intradermal tuberculin (CIDT) positive trait in hybrid animals. This skin test is the standard immunodiagnostic procedure used for the identification of bovine tuberculosis infected cattle. The results suggest that chromosomes 6 and 11 have a greater local taurus ancestry region in CIDT positive cattle. These two regions could play important roles in the susceptibility pattern of bovine tuberculosis infection in the admixed animals typical of East Africa and beyond. The admixture results also indicated that affected populations have greater taurus ancestry associated risk with overall genome wide significant score of 3.37 and it was well above the threshold level (2.0). The genomewide score is instrumental to detect evidence for whether there an association between the trait and ancestry anywhere in the genome (GREENSPAN *et al.* 2004).

An admixture mapping approach is facilitated when both mapped traits and test alleles show significant difference in frequency between the populations ancestral to the test samples and is a powerful approach (HOGGART *et al.* 2004; MUCHADEYI *et al.* 2007; ZHU *et al.* 2008). This allows of a relatively small number of informative markers spread evenly across the chromosomes for genome wide inference. Particularly, for young admixture populations such as used in this study, the numbers of makers needed are limited for an initial detection of signals. Recently admixed populations have higher levels of admixture generated linkage disequilibrium compared to ancestral populations (Falush *et al.* 2003; Halder and Shriver 2003; Hoggart *et al.* 2004). This makes the limited number of makers

used in this study plausible as the test population has a recent admixture history with an average of 3 to 4 generations.

The sample size here is also within the range recommended for admixture mapping given the clear risk difference for bovine tuberculosis between ancestral taurus and zebu animals. Previous studies reported that crossbred animals in the current study area have intermediate level of tuberculosis prevalence. In the report the European cattle (Holstein) were at least 2 times more susceptible than the local breeds (Ethiopian zebu) (AMENI *et al.* 2007a). In admixture mapping studies performed on African American human populations with an average time of admixture 10 -13 generations ago, a sample size of less than 500 was sufficient to effectively scan in a population where the differential risk due to one of the ancestry was higher than 1.6 (Winkler *et al.* 2010). In the current analysis the bulk of markers were highly informative compared to human panels (although it was a challenge in marker selection given the Ethiopian zebu are not pure *Bos indicus* and some markers were less informative than previously anticipated distribution plot of parental population in Figure 1). Also, the population has very recent admixture history, risk of bovine tuberculosis infection is more than twice as great in *Bos taurus* breeds than *Bos indicus* and the contribution of European *Bos taurus* ancestry in the general population was smaller (0.33) than *Bos indicus* ancestry. Considering all these parameters, the results presented using 662 markers for 508 sample size allow a credible genome scan.

The most important analysis is the z-score summary of case-control differences as this employs the total information in the sample. This yielded significant score on BTA6 and BTA5. Suggestive trait associations with genome of these regions and some additional

suggestion were given from the case only analysis for the associations at BTA11, BTA12 and BTA19.

The chromosome 6 peak harbors important genes related to mycobacterium infections. The role of TLR1, TLR6, TLR10 and other TLRs for mycobacterial infection is well documented in terms of their critical importance for innate immunity (Doherty and Arditi 2004; Heldwein and Fenton 2002; Heldwein et al. 2003; Korbel et al. 2008; Moller et al. 2010; Reiling et al. 2002). They are part of the Toll gene family that is involved in bacterial pattern recognition (pathogen associated molecular patterns, PAMPs) including bovine tuberculosis which initiates the modulation of an innate immune response (Moller et al. 2010; Quesniaux et al. 2004). Mutations have been implied to confer either resistance or increased susceptibility to infectious diseases in human. It has also been demonstrated that natural variation in some of these molecules affect differential susceptibility to infectious diseases (e.g. invasive bacterial infections (Coats et al. 2003), human tuberculosis (BRYC *et al.* 2010; SELVARAJ *et al.* 2010), and malaria (Corr and O'Neill 2009) specific to the binding spectrum of the TLRs involved (Khor 2009). Several other regions among the strongest peaks in the case control admixture scan and showing consistent signal among independent analyses also give potential candidate genes for showing differential levels of susceptibility to bovine tuberculosis infection. Peli1 is found in chromosome 11 and is a ubiquitin ligase needed for the transmission of TRIF-dependent TLR signals (Chang et al. 2009). Although no direct link is reported on ADORA2B gene and tuberculosis infection, some evidence has emerged regarding its role in malaria infection which is also an intracellular parasite (Auburn et al. 2010).

Follow up fine mapping as focus analysis of more markers in the region could help identify the possible variants associated with bovine tuberculosis. However, this might pose a challenge when an applying admixture model in a population that has a very recent admixture history. Recently admixed individuals display large ancestral blocks that come from one of the contributing parents. Multiple generations are required to reach the optimal state of mosaicism for fine mapping analysis. This may overcome in part by extending the study population, or by complementing this work in other African populations with more ancestral zebu-taurus admixtures.

CHAPTER SIX: GENOMEWIDE ADMIXTURE SCANS OF CATTLE PHENOTYPE TRAITS

6.1 Introduction

Cattle phenotype characterization has been utilized greatly to increase the breeding value of cattle and to upgrade economically important traits. Various traits such as dairy quality, meat quality, body (carcass) weight and disease resistance traits are among the most favoured characters (IBEAGHA-AWEMU *et al.* 2008). Although most characters are polygenic in nature and environmental factors play a vital role, it has been possible to map genes influencing many of those desired traits within domestic cattle to specific regions of the genome (<http://www.animalgenome.org/cgi-bin/QTLdb/BT/>) (ASHWELL *et al.* 2001). The most common strategies used to detect genetic markers related to a particular trait, or quantitative trait loci (QTL), is by using linkage analysis of pedigrees and genome wide association analysis of unrelated individuals.

Linkage analysis in cattle have dealt with the mapping of loci for various important traits. There are a plethora of regions of cattle genome that are linked to QTL. For example traits associated with milk (MARTINEZ-ROYO *et al.* 2010), body weight, carcass traits (TAKASUGA *et al.* 2007; YOKOUCHI *et al.* 2009), and coat colour (MATUKUMALLI *et al.* 2009) all have QTL containing regions described.

The availability of crossbreds of divergent parental populations offers a powerful admixture mapping approach for initial detection of QTL by because of the substantial long-range LD created in such crosses. Divergent ancestral strains also increase the availability of informative markers and variance in phenotype (BRAUNSCHWEIG 2010).

Here admixture based QTL analysis is applied to map several phenotype traits correlated with zebu versus taurus ancestry. In this strategy it is possible to identify alleles with minor or modest phenotypes effects without relying on pedigree information. The most important feature of admixture mapping based QTL analysis is that it has a capacity to account for population stratification (McKEIGUE 2005).

Admixture mapping can function as a bridge between pedigree-based linkage analysis and genome wide association studies since it has less disequilibrium than linkage but more than association-based analysis (McKEIGUE 2005). Recently created admixture between genetically differentiated populations such as those used in this study can demonstrate the wider utility of this model within cattle for a variety of phenotype traits that may possibly be correlated differentially to the parental subspecies. Admixed populations have a high level of LD for the loci located as far apart as 10 -20 cM and genetic substructure effects (inter individual variations), are the major factors for a greater statistical power to detect QTL (HALDER and SHRIVER 2003).

6.2 Materials and Methods

6.2.1 Study populations

The subjects in the present study were from the 591 crossbred Ethiopian cattle used for the genetic association of bovine tuberculosis described in chapter 5. The sampling and examination procedure for the study population is described in chapter 4. Subjects included in this study were those with complete baseline information about the individual animal. The animals were selected without regard to the information about their disease status for tuberculosis.

6.2.1 Measurement of phenotype traits

6.2.1.1 Animal coat colour

During the field visit to Selalle, photographs of each individual animal were taken using a digital camera showing clearly the coat colour of the animal. The study number (code) of the animal was also attached to the animal during each photo shot such that the number matches to the case record form. The lateral side view of the animal was photographed, showing the conformation of the participants. Phenotypic coat colour scoring was performed for the animals based on categories and scales definitions. Eight possible categories were formulated. Binary assignment for each colour was used to indicate presence (coded 1) and absence (coded 0). The colour categories include White, Off-white, Light-grey, Grey, Light-red, Red, Black and Spot following the assignment made by (HIROOKA *et al.* 2002). Furthermore, the groups in each colour categories were merged together to formulate a quantitative colour assignment using a modified method explained

by (GUTIERREZ-GIL *et al.* 2007). Quantitative coat colour was classed using gradient dilution categories and assigned as not diluted (black and red), intermediate diluted (grey and light red) and completely diluted if the animal had a white coat. Whenever there were individuals that were difficult to assign based on the formatted coat colour category, the animal was coded as missing during analysis.

6.2.1.2 Animal Body condition score

The general information about individual animals that were used as a covariant at different stages of admixture mapping in chapter 5 were utilized here and included as a covariant factors in the analysis. One of the parameters used to investigate a possible quantitative trait related to either *Bos taurus* or *Bos indicus* ancestry was animal body condition. The classification as indicated in chapter 4 involved scoring each animal from the scale of 1 to 9 as described in (NICHOLSON 1986). The standardized value of the body score was used as a quantitative value for admixture mapping analysis. A second generalized format was also applied to use a three way classification of animal body condition. The animal body condition was coded 1 (lean) if the body score is less than 3 and coded 2 (medium) if the body condition is between 3 and 6. The animal was classified as fat and coded 3 if the body condition score is above 6. The values were used as a continuous variable for admixture mapping to scan for quantitative traits associated with ancestry.

6.2.1.3 Grading dewlap and hump sizes

The external features of cattle dewlap and hump for pure *Bos taurus* and *Bos indicus* shows a distinct pattern. *Bos taurus* animals are humpless and the extent of neck dewlap is absent. While a pure *Bos indicus* animal's main external phenotype is the presence of hump and

well defined dewlap growth. The study subjects were scored for both features from the scale of 1 to 4. The score 1 being completely *Bos taurus* type, i.e. no hump and dewlap seen on the subject. The score value of 4 represents completely *Bos indicus* features of well developed sizes of dewlap and hump. Each score was made independently for the two features and to minimize subjective variations and biased score assignment a double entry strategy was followed. The score values were used as a continuous variable for measuring quantitative traits associated with ancestry. A binary trait was also derived from the scores and coded 1 if an animal has above average developed dewlap and hump. The animals with below average growth levels of dewlap and hump were coded as 0 for the binary outcome variable of admixture mapping analysis.

6.2.2 Genotype data for the ancestry informative markers

The genotype data from the 768 SNPs used in the Goldengate assay were imported to analyse the phenotype variation among the hybrid cattle. These markers were selected based on the degree of information they provide about the genetic difference between *Bos taurus* and *Bos indicus* cattle as detailed in chapter 3. This made it possible to leverage the same data base to investigate the genetic correlates of the additional traits with ancestry. These markers were informative enough to capture if there is a skewed distribution in ancestry that might contribute to the observed phenotype variation under consideration. In this analysis 662 SNPs were included which qualified for the admixture mapping analysis in chapter 5 after quality control filtering.

6.2.3 Genomewide ancestry association of phenotype traits

The genotype data were analysed for each trait under the binary and continuous models implemented in the ADMIXMAP program. This Bayesian-based program highlights

genomic regions associated with an increased *Bos taurus* or *Bos indicus* ancestry in relation to phenotype traits. The phenotype traits scanned for admixture mapping included animal coat colour, body condition, the size of dewlap and hump. The analysis was carried out for each of these traits by adjusting the observation for covariates such as age and sex.

6.2.4 Statistical analysis

The ADMIXMAP analysis provides a z-score for the association of both binary and continuous phenotypes and the measured level of ancestry. In this study all the outputs are referenced in relation to excess *Bos indicus* ancestry. As a result a significant zebu ancestry association gives a z-score > 2.5 and z-score > 2 is interpreted as a suggestive value. The sign changes but with the same interpretation for significant and suggestive excess *Bos taurus* ancestry associations. The corresponding *P* values were used to assess the level of association.

6.3 Results

6.3.1 Animal coat colour variants

The genotype data was imported for 500 admixed test samples for the analysis of coat colour association and ancestry. 18 animals did not have pictures taken and were assigned missing for the trait. The ADMIXMAP analysis was performed by putting the local Ethiopian zebu (Arsi breed) as one of the reference population. Ethiopian zebu have a variable coat colour: mostly dark brown but may also be found with a wide spectrum of black, red and spotted derivatives of colours. The second ancestral population are the European *Bos taurus*, Holstein-Friesian, that are exclusively black and white. The crossbred animals showed a very diverse segregation for the coat colour phenotype (figure 6.2).

Figure 6.1 shows the number of animals expressing the phenotype using eight coat colour coding indexes.

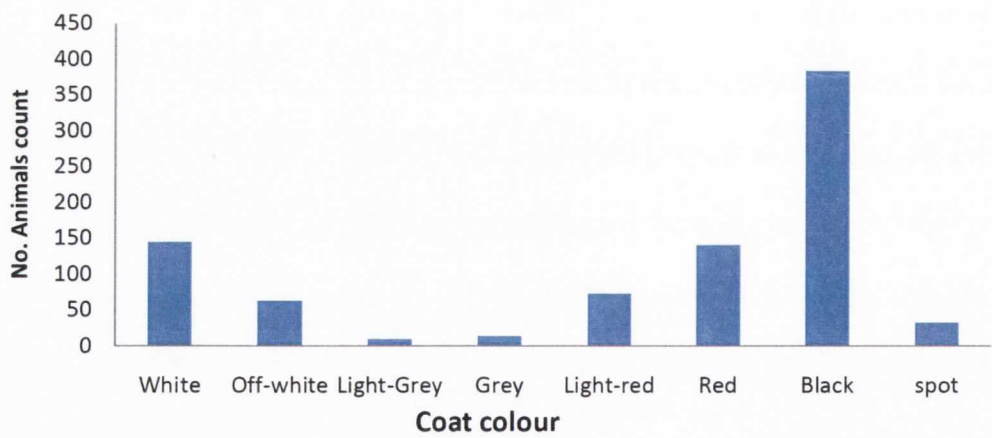


Figure 6.1. The number of animals counted for each coat colour category. The classification is based on the study populations containing 482 hybrid animals of *Bos taurus* and *Bos indicus* for at least two generations.

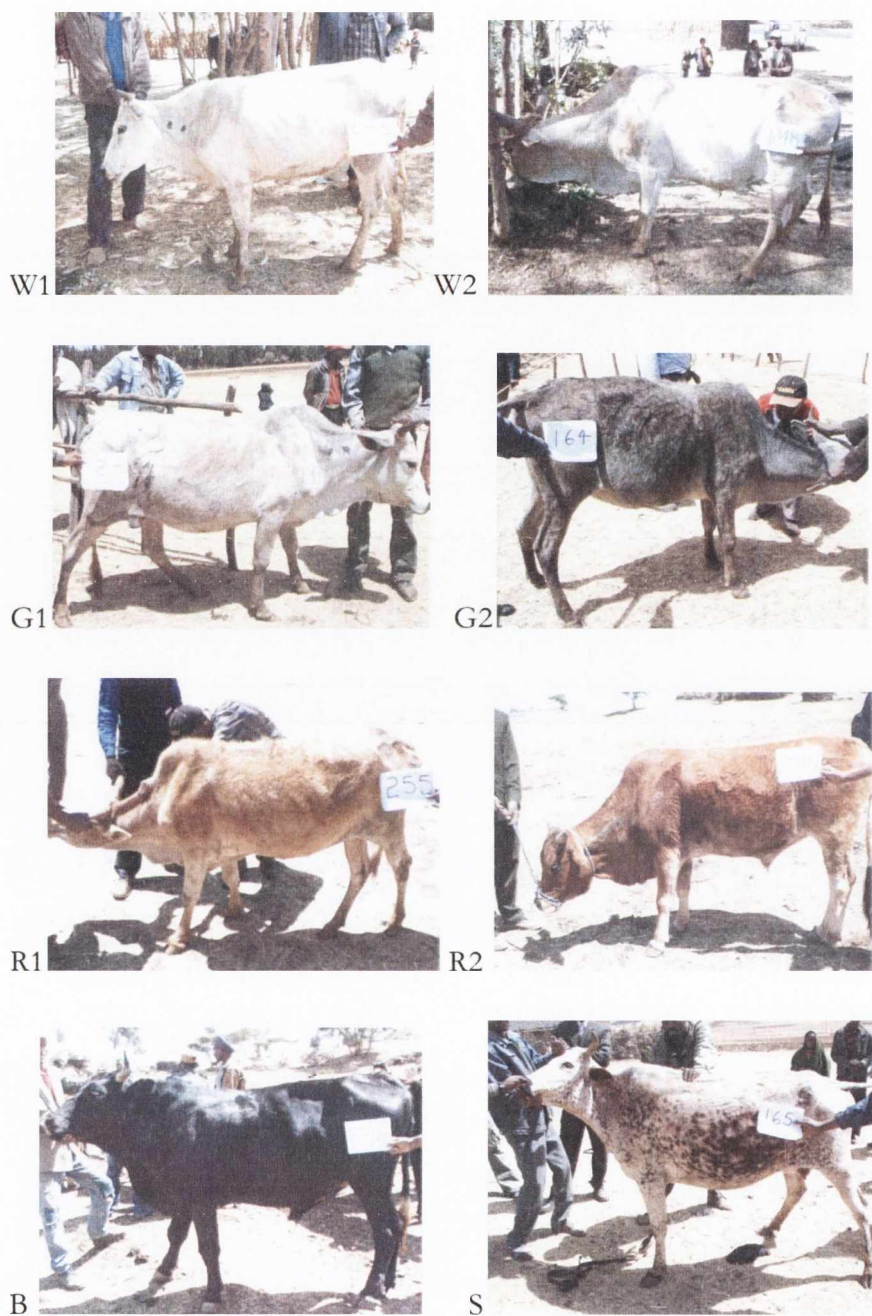
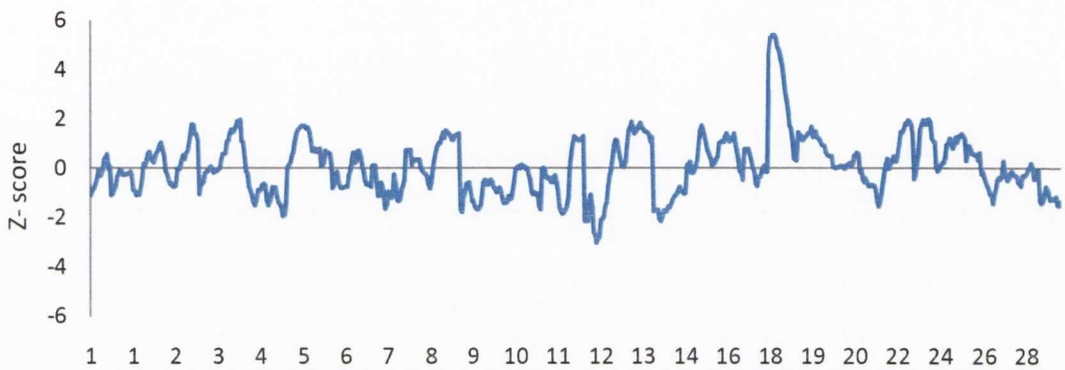


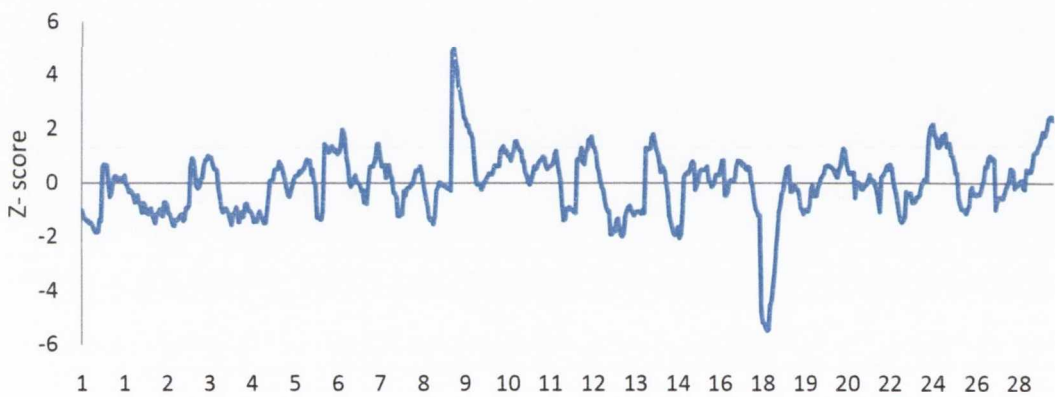
Figure 6.2. Animal coat colour coding classified into eight groups. W1- white, W2-Off white, G1-light grey, G2-Grey, R1-light red, R2-Red, B-Black and S- Spot.

The binary admixture scan was performed for each coat colour group using the colour guideline illustrated in figure 5.2. For a single trait variable five admixture mapping analyses were made after putting closely related colour variants in the same group. The first group contains subjects with light red and red coat coloured which were 76 in total compared to the rest of the animals (red vs Not red). The second group compared the presence of black (341) to not black coat colours, the third group compared light-gray and grey (which were 20) to the not gray animals. Similarly White (50) vs not white, and spotted coat (33 animals) vs not spotted coat skin were scanned for the initial admixture mapping analysis (figure 6.3). The most significant observation was made at chromosome 18 for the comparison between red and not-red animals. The z-score for the regions was 5.4 with a P value of (6.3×10^{-8}) and the peak centred at the melanocortin 1 receptor (*MC1R*) gene which is known to regulate the level of Tyrosinase enzyme production which is responsible for the switch between pheomelanin (red pigment) and eumelanin (black pigment) production (figure 6.5). The dominant *MC1R* gene is responsible for black coat colour and the recessive allele results in red coat colour (SEO *et al.* 2007). The z-score was a positive value indicating that the red coated phenotype was as a result of excess zebu ancestry. Similarly the same region was significant for the analysis of black vs not-black groups. However, the z-score gave a negative value (z-score = -5.5; P value = 4.2×10^{-8}) showing that the region has an excess taurus ancestry. There was also another peak on chromosome 9 where a positive z-score of 4.9 (P value = 6.4×10^{-7}) was registered indicating significant zebu ancestry region may also be associated with the black coat colour phenotype. The remaining colour variants also showed some signals of association for ancestry. However, the intensity of the signals was not as strong as the observation

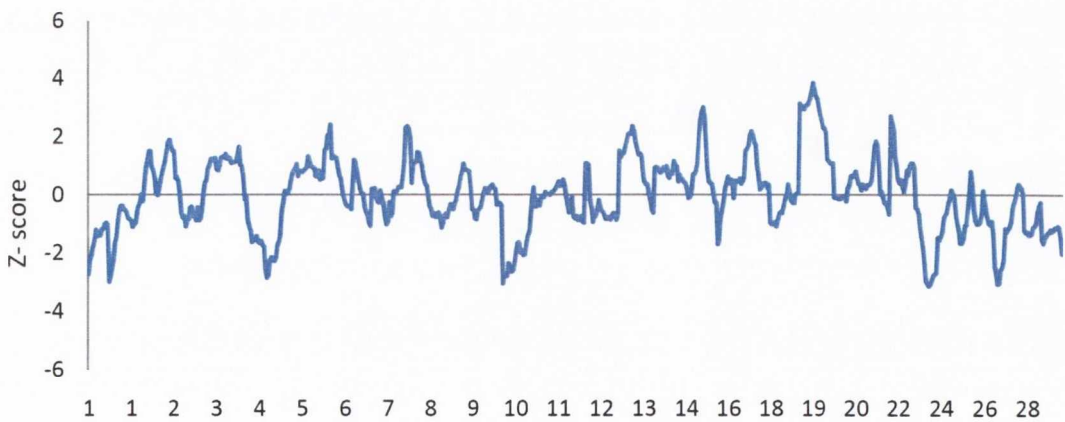
made for the red and black colours. A summary of significant regions and the scores for the binary association of skin colour with ancestry is shown in table 6.1.



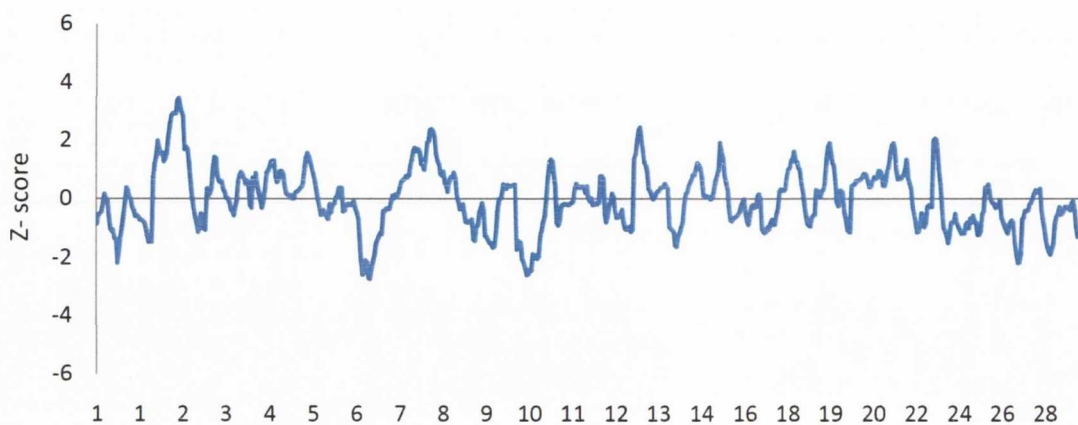
3a. Red vs Not-red



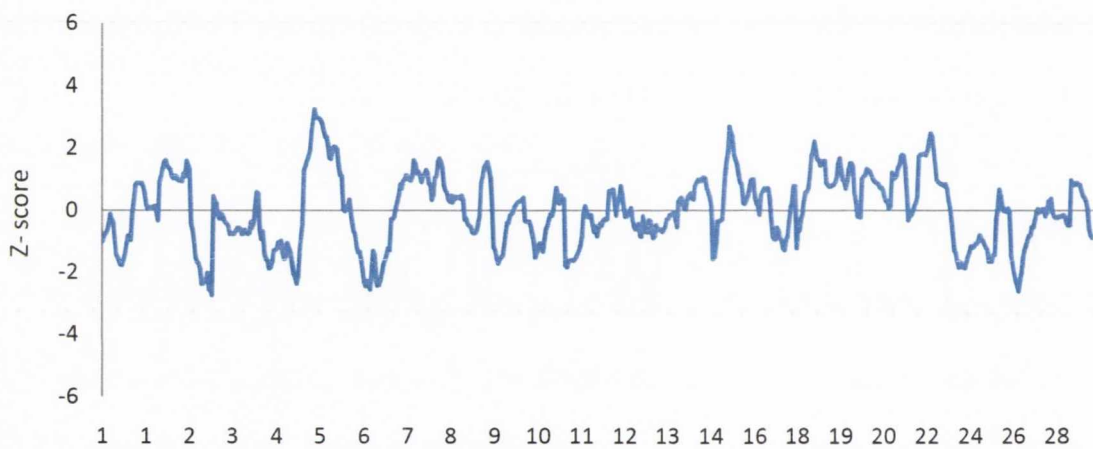
3b. Black vs Not-black



3c. Grey vs Not-gray



3d. White vs Not-white



3e. Spotted vs Not-spotted

Figure 6.3. Admixture mapping of coat colour phenotype for each chromosome (x-axis) showing the Z-score (y-axis). 3a. Red vs Not-red 3b. Black Vs Not-black 3c. Grey Vs Not-grey 3d. White vs Not-white 3e. Spot vs Not-spot

In order to investigate the dilution effect on coat colour and to identify possible genes responsible for the effect, a systematic quantitative method was devised based on black, red and white colours. The combined dilution quantitative scores included 5 animals that showed completely diluted (White, coded 1), 81 partially diluted (light-grey, grey, light-red and coded 2), and 396 completely undiluted animals (black, red and coded 3). The quantitative admixture scan highlighted three strong signals on chromosome 9, 18 and 19 as shown in the figure 6.4.

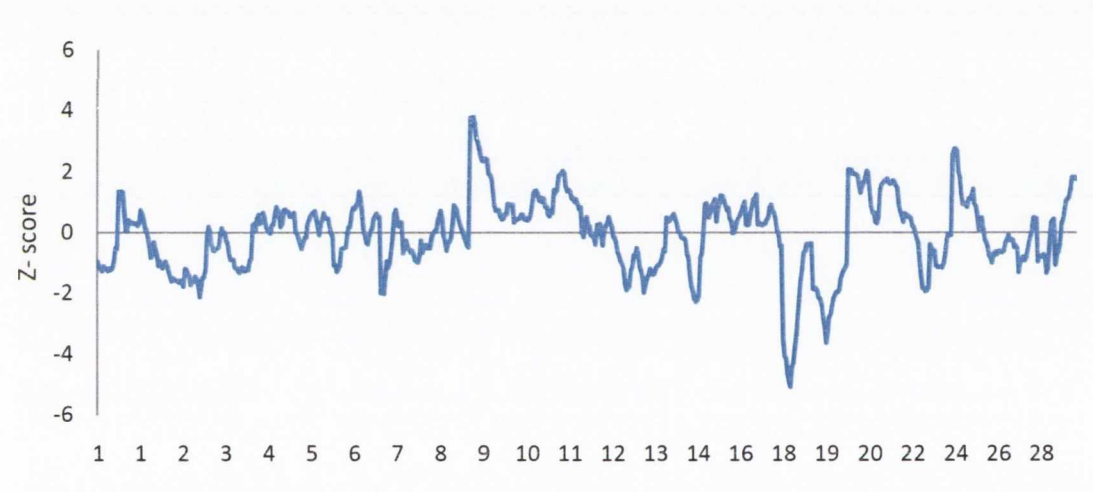


Figure 6.4. Admixture mapping of combined quantitative dilution of coat colour phenotype highlighted for each chromosome (x-axis) and quantified based on Z-score (y-axis).

Table 6.1: The top admixture signals captured for genomewide correlations of phenotype traits

Phenotype	Variation	Outcome variable	BTA	Peak position(bp)	Nearest marker	Ancestry	z-score	P value
Coat colour	Red	Binary	18	11002597	BTA44427	<i>B.indicus</i>	5.40	6.3x10 ⁻⁸
			12	35296048	BTA152656	<i>B.taurus</i>	-3.02	2.6x10 ⁻³
	Black	Binary	9	8771708	BTA159047	<i>B.indicus</i>	4.98	6.4x10 ⁻⁷
			18	13030708	BTA132064	<i>B.taurus</i>	-5.48	4.2x10 ⁻⁸
	Grey	Binary	19	25598691	BTA155051	<i>B.indicus</i>	3.84	1.2x10 ⁻⁴
			1	64174818	BTA153744	<i>B.taurus</i>	-2.98	2.9x10 ⁻³
			4	66290526	BTA70886	<i>B.taurus</i>	-2.84	4.6x10 ⁻³
			10	3208775	BTA149082	<i>B.taurus</i>	-3.04	2.4x10 ⁻³
			15	41084554	BTA36820	<i>B.indicus</i>	2.99	2.7x10 ⁻³
			22	2942820	BTA136233	<i>B.indicus</i>	2.69	7.0x10 ⁻³
			23	40251495	rs29027406	<i>B.taurus</i>	-3.13	1.7x10 ⁻³
			26	38419991	BES5509942	<i>B.taurus</i>	-3.05	2.3x10 ⁻³
	White	Binary	2	65744631	BTA134039	<i>B.indicus</i>	3.43	5.9x10 ⁻⁴
			6	91138928	BTA77209	<i>B.taurus</i>	-2.73	6.3x10 ⁻³
			10	25244764	BTA125260	<i>B.taurus</i>	-2.60	9.3x10 ⁻³
	Spot	Binary	5	28774242	BTA72999	<i>B.indicus</i>	3.25	1.2x10 ⁻³
			2	140621929	BTA134746	<i>B.taurus</i>	-2.72	6.5x10 ⁻³
			6	57806018	BTA18387	<i>B.taurus</i>	-2.54	1.1x10 ⁻²
			15	55011002	BTA150498	<i>B.indicus</i>	2.69	7.2x10 ⁻³
			26	19831953	BTA138776	<i>B.taurus</i>	-2.60	9.3x10 ⁻³
	Colour	Continuous	9	15930978	BTA146275	<i>B.indicus</i>	3.80	1.5x10 ⁻⁴
			18	18072666	BTA132100	<i>B.taurus</i>	-5.07	4.0x10 ⁻⁷
			19	25598691	BTA155051	<i>B.taurus</i>	-3.61	3.1x10 ⁻⁴
			24	10808268	BTA28187	<i>B.indicus</i>	2.75	5.9x10 ⁻³
Body	lean,medium,fat	Continuous	26	23510130	BTA60882	<i>B.indicus</i>	2.74	6.1x10 ⁻³
			2	116969661	BTA134534	<i>B.taurus</i>	-2.66	7.9x10 ⁻³
		Binary	26	23510130	BTA60882	<i>B.indicus</i>	2.80	5.2x10 ⁻³
			2	93963498	BTA101823	<i>B.taurus</i>	-2.92	3.5x10 ⁻³
			11	60549611	BTA154095	<i>B.taurus</i>	-2.79	5.3x10 ⁻³
Dewlap		Binary	18	1822486	BTA131863	<i>B.indicus</i>	3.17	1.5x10 ⁻³
			4	95706599	BTA142388	<i>B.taurus</i>	-3.71	2.1x10 ⁻⁴
		Continuous	9	71435003	BTA146623	<i>B.indicus</i>	2.72	6.5x10 ⁻³

Hump		Binary	10	9058148	BTA27577	<i>B.taurus</i>	-3.36	7.8x10 ⁻⁴
			19	25598691	BTA155051	<i>B.indicus</i>	2.94	3.2x10 ⁻³
			2	72164867	BTA151467	<i>B.indicus</i>	2.87	4.1x10 ⁻³
			4	95706599	BTA142388	<i>B.taurus</i>	-2.94	3.3x10 ⁻³
			15	41084554	BTA36820	<i>B.indicus</i>	2.89	3.9x10 ⁻³
			20	46453651	BTA55581	<i>B.indicus</i>	2.65	7.9x10 ⁻³
		Continuous	6	120914184	BTA04482	<i>B.indicus</i>	3.40	6.7x10 ⁻⁴
			1	2596843	BTA123586	<i>B.indicus</i>	2.55	1.0x10 ⁻²
			2	109524324	BTA101152	<i>B.indicus</i>	2.78	5.4x10 ⁻³
			15	55011002	BTA150498	<i>B.taurus</i>	-2.64	8.4x10 ⁻³
			19	25598691	BTA155051	<i>B.taurus</i>	-2.54	1.1x10 ⁻²
			18	1822486	BTA131863	<i>B.indicus</i>	2.56	1.1x10 ⁻²
			16	53743308	BTA39603	<i>B.taurus</i>	-2.99	2.5x10 ⁻³

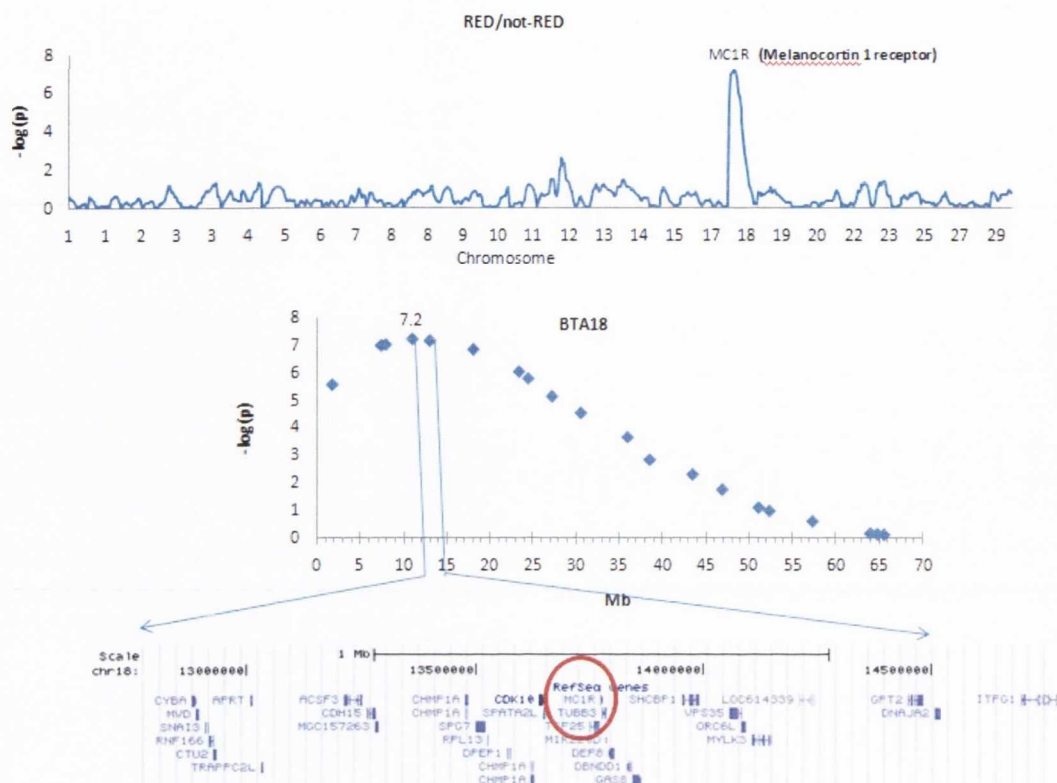


Figure 6.5. The significant admixture signal ($-\log P$ value = 7.2) associated with red coat colour observed on chromosome 18 which has *MC1R* gene at the peak of the curve.

6.3.2 Bovine QTL mapping for coat colour

The ADMIXMAP significant scores for coat colour traits shown in table 6.1 were used as a reference points for constructing 1 – LOD credible intervals for a possible QTL regions based on the bovine B.tau4 map build. The coat colour QTL regions are plotted in figure 6.6 and compared to bovine pigmentation QTL (<http://www.animalgenome.org/cgi-bin/QTLdb/BT/>) for all autosomal chromosomes in figure 6.7. The main cattle pigmentation QTL regions found in BTA6 and BTA18 with significance from admixture mapping were consistent with the bovine QTL map. A chromosome 6 QTL was significant for spotting at 57.08 Mb with 1 –LOD interval of 26.80 Mb – 71.07 Mb. White

versus non white coat colour admixture mapping analysis has two peaks at 57.08 Mb (53.14 Mb – 99.07 Mb) and at 91.13 Mb (58.40 Mb – 138.34 Mb). In this chromosome there are six pigmentation QTL regions highlighted on the bovine map on figure 6.7 that are related to spotting and coat colour. The region contains the pigmentation related gene KIT within the interval. The chromosome 18 QTL map has three peaks for red: 11 Mb (7.27 Mb – 14.27 Mb), black 13.03 Mb (10.98 Mb – 23.38 Mb) and coat colour dilution 18.07 Mb (4.7 Mb – 25.22 Mb) which also coincide with the bovine pigmentation map that has four QTL in the region. As shown in the figure 6.5 this regions contains the important pigmentation gene MCR1 which is found in the credible interval and is reported to cause red and black coat colour variation. Coincidental QTL regions were identified here on BTA15 41.08 Mb (24.13 Mb – 69.02 Mb) for grey coat colour variation and the bovine QTL map for pigmentation of hoof colour variations. The pigmentation QTL found on BTA2 for spotting is also in close proximity to the corresponding bovine QTL map position which contains the RPE gene. The QTL region highlighted for the spotting trait on BTA5 28.77 Mb (23.84 Mb – 46.79 Mb) is not described on bovine QTL map on (figure 6.7) but is also in close proximity to KITLG gene. KITLG is the ligand for the KIT gene and has been shown to have phenotype variation for the roan coat colour (DESEPULVEDA *et al.* 1995).

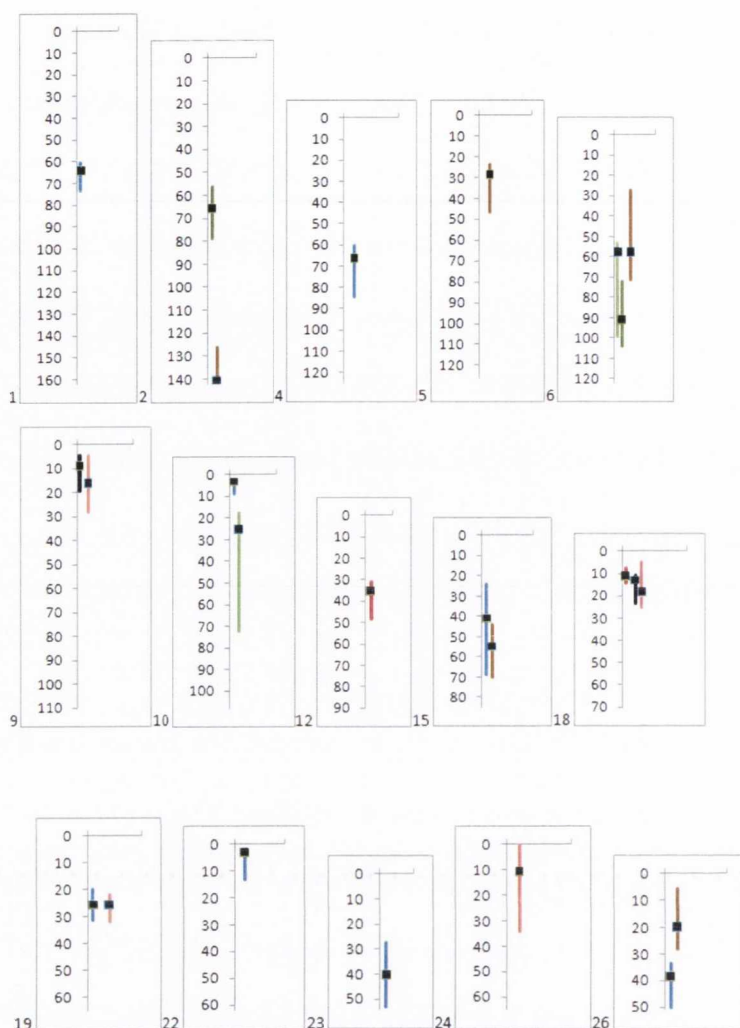


Figure 6.6. Coat colour QTL mapping using 1-LOD credible interval of ancestry associated regions on bovine chromosomes.



6.3.3 Quantitative Admixture mapping for body condition score

The admixture mapping for body condition was performed using two quantitative scores. The first one was using the numerical scaling of body condition from the lowest (1) to highest (7) scores given to the individual subject. The second approach was to categorize the scores in to three levels that include lean (coded 1), medium (coded 2) and fat (coded 3). There were 230 lean, 266 medium and 4 fat body scored in the group. The admixture mapping showed very similar patterns for both analyses but the signals were slightly stronger when 1 to 7 scoring method was used (figure 6.8). The strongest signals were observed for BTA2 (z-score = -2.92, P value = 3.5×10^{-3}), BTA11 (z-score = -2.79, P value = 5.3×10^{-3}) and BTA26 (z-score = 2.80, P value = 5.2×10^{-3}) as shown in table1. The association z-score is a negative value for BTA2 and BTA11 which indicates an excess *Bos taurus* ancestry correlated to this phenotype. While the value was positive for BTA26 showing an excess *Bos indicus* ancestry correlates for animal body condition. Body condition phenotype QTL map with 1 –LOD credible interval for these chromosomes are shown in figure 6.9.

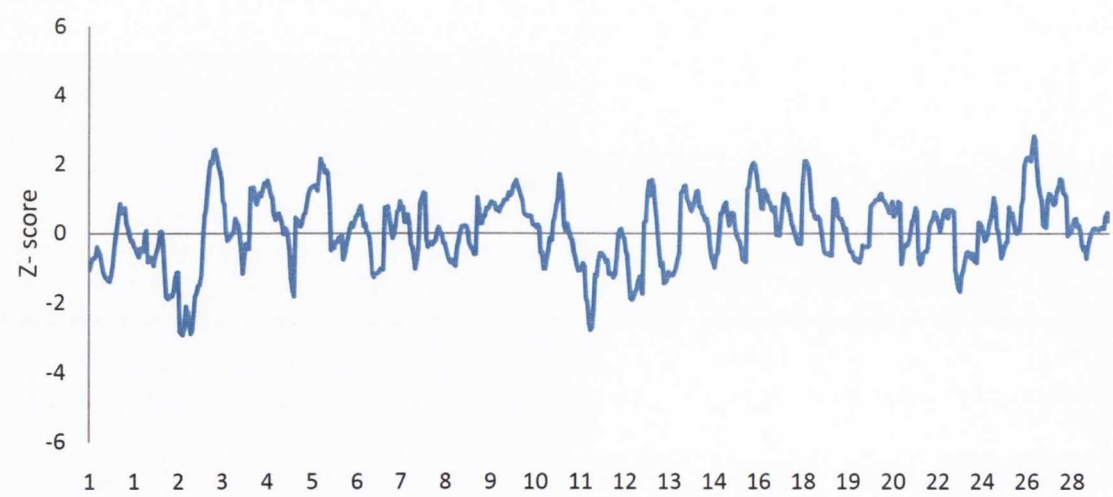


Figure 6.8. Admixture mapping of body conformation trait (body condition)

Admixture mapping significant association scores for animal body condition were compared to the bovine QTL maps for body and carcass weight traits of the same chromosomes (figure 9). The top score on BTA2 at 116.96 Mb with 1 –LOD credible interval range between 110.42 Mb to 120.23 Mb has six QTL described for body and carcass weight. Intervals on BTA11 at 60.54 Mb (52.24 Mb – 87.72 Mb) and BTA26 at 23.51 Mb (17.17 Mb – 25.70 Mb) contain four and three described QTL for body and carcass weight respectively.

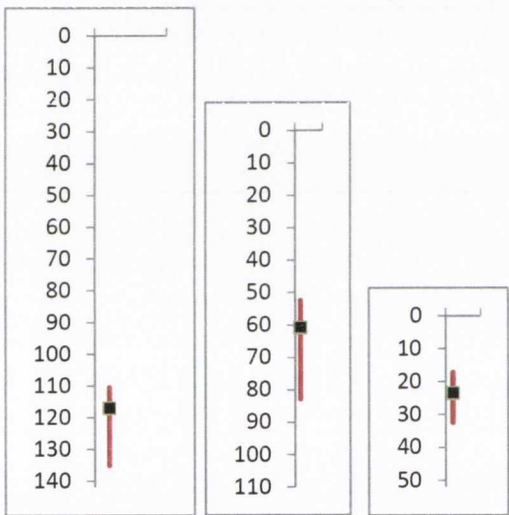


Figure 6.9. Body condition QTL mapping using 1-LOD credible interval of ancestry associated regions on BTA2, BTA11 and BTA26 respectively.

6.3.4 Admixture mapping for dewlap and hump development

A significant admixture mapping signal for binary scored dewlap size was observed on chromosome 4 with top z-score of -3.7 (P value = 2.1×10^{-4}) with a second region on chromosome 18 with z-score 3.2 (P value = 1.5×10^{-3}), shown in figure 6.10. The dewlap size scored as a continuous trait for quantitative admixture scan gave a signal on chromosome 9 with a z-score of 2.7 (P value = 6.5×10^{-3}).



Figure 6.10. Admixture mapping of body conformation trait (dewlap development)

ADMIXMAP analysis also showed a significant signal linking ancestry and the extent of hump growth. For the binary recoded analysis the top score related to a significant taurus ancestry region on chromosome 10 with z-score -3.4 ($P=7.8 \times 10^{-4}$) and the top zebu ancestry associated score was z-score = 2.9 ($P=3.2 \times 10^{-3}$) on chromosome 19. The use of a categorical variable format for this phenotype, gave significant top regions on chromosome 6 (z-score= 3.4, $P = 6.7 \times 10^{-4}$) and on chromosome 16 (z-score = -3 $P = 2.8 \times 10^{-3}$) (figure 6.11). The possible QTL regions around the significant markers using 1 –

LOD credible intervals for ancestry association for dewlap and hump conformation are shown in figure 6.12a & 6.12b respectively.

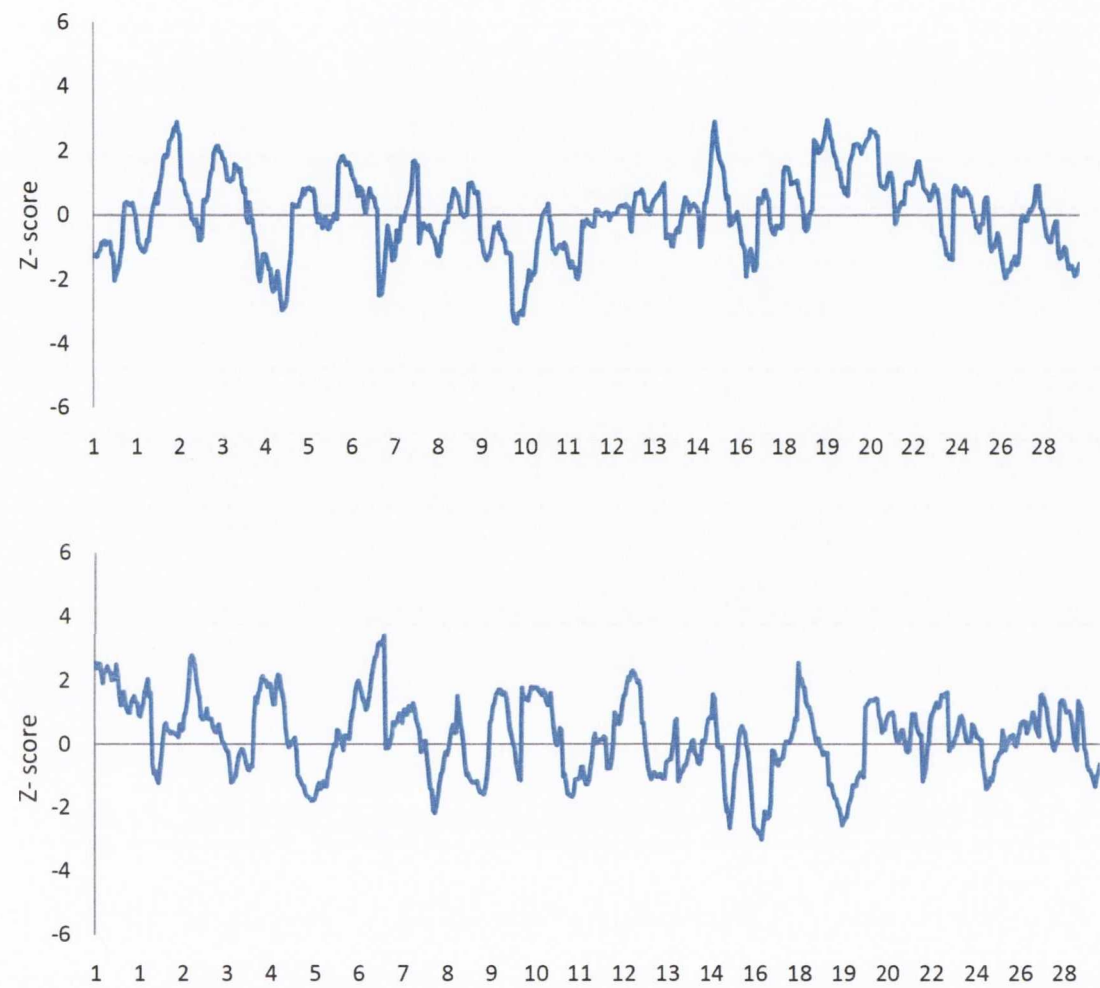


Figure 6.11. Admixture mapping of body conformation trait (hump development) for binary (1) and continuous outcome model (2).

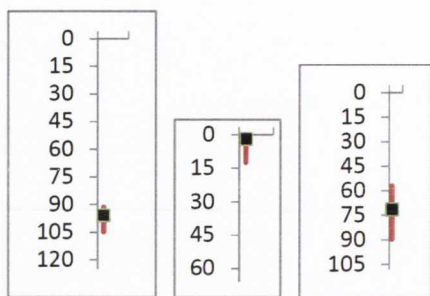


Figure 6.12a. QTL regions of body conformation (dewlap development) BTA4, BTA19 and BTA9

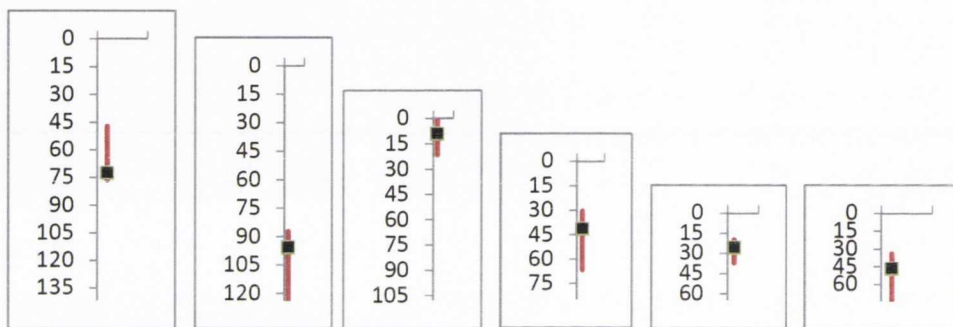


Figure 6.12b. QTL regions of body conformation (hump development) BTA2, BTA4, BTA10, BTA15, BTA19 and BTA20 using binary admixture model.

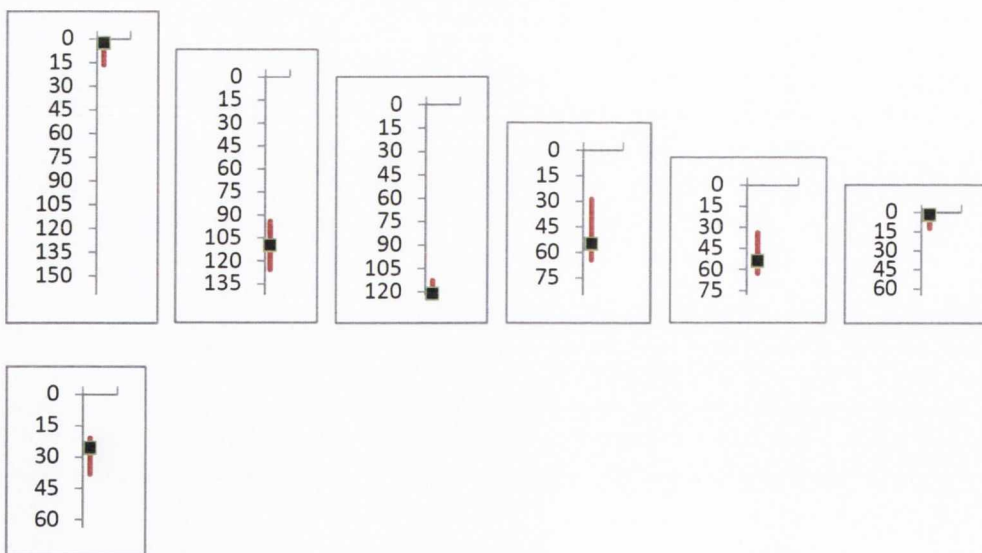


Figure 6.12c. QTL regions of body conformation (hump development) BTA1, BTA2, BTA6, BTA15, BTA18 and BTA19 using continuous variable admixture mode

6.4 Discussion

Ancestry based analysis of admixed cattle populations sampled from Ethiopia that were crossbreds of European *Bos taurus* (Holstein-Friesian) and *Bos indicus* (Ethiopian zebu) have illustrated an extra dimension in detection of QTL influencing various phenotype traits. It has clearly shown that ancestry estimates of either taurus or indicus were highly correlated to animal coat colour and other body conformation parameters suggesting strong genetic factors involved in the observed differences.

Numerous genetic regions were highlighted that could contain genetic variants affecting animal coat colour, body condition and the extent of dewlap and hump development. The most significant genetic regions for the pigmentation traits were detected on BTA5, BTA6, BTA9, BTA15, BTA18 and BTA19. Variants in these regions could be responsible for the presence or absence of a specific coat colour. The other possibility is that genes in the highlighted positions could be involved in the regulation of the level of expression of a particular colour phenotype. The top region that came out in this analysis was found in BTA18 that contains the *MCR1* gene which is directly linked to red and black pigmentation (SEO *et al.* 2007). A similar finding was reported from a backcross linkage analysis of Holstein X Jersey and Holstein X Charolais population where *MCR1* was highly correlated to a switch between red and black coat colour (GUTIERREZ-GIL *et al.* 2007; LIU *et al.* 2009). A family based analysis of red variants of rare phenotype in Holstein also mapped to the *MCR1* gene which was found to be co-segregated with another pigmentation gene (beta-defensin 103) on BTA27 (DREGER and SCHMUTZ 2010). The significant regions also coincide with some additional cattle QTL regions for

pigmentations. For example in BTA 6, which has numerous colour related genes, has two pigmentation QTL regions that are also described in other study. Consistent finding was reported on BTA6 QTL for a variation on spotting of animal coat (CASAS *et al.* 1999; LIU *et al.* 2009; MORRIS *et al.* 2009). This region contains a receptor tyrosinase kinase (KIT) locus which is associated with the de-pigmentation phenotype in mammals (CASAS *et al.* 1999). Interesting another mapped variant came out in BTA5 where the candidate region is found in close proximity to the *KITLG* gene. This is functionally related to the roan type coat colour in cattle and a similar observation was made in spotted/roan groups in this analysis (SEITZ *et al.* 1999). A QTL region on BTA15 has been described for hoof pigmentation in cattle (LIU *et al.* 2009), and here a similar peak near the region was detected for excess *Bos indicus* ancestry relating to grey coat colour.

The coat colour admixture mapping signals for BTA9 and BTA19 were very strong. There were no previously described pigmentation genes that could be associated with these significant scores. The possible explanation for this observation could be that these may contain some new genes functionally related to pigmentation. It has been documented that pigmentation genes are also involved in functions such as precursors for nerve cells, since melanocytes originate in the neural crest (HORNYAK 2006), and other pleiotropic effects that can be deleterious (JACKSON 1994).

Admixture mapping based on the animal body condition score and body conformations gave some significant regions on BTA2, BTA11 and BTA 26 which also have previously described bovine QTL related to animal body weight and carcass weight (KNEELAND *et al.* 2004; MACNEIL and GROSZ 2002; MCCLURE *et al.* 2010). There are no described QTL

for cattle hump and dewlap sizes that could be used to check for the highest peaks regions detected BTA4, BTA 6, and BTA10.

Admixture mapping analysis is a powerful mapping approach and potentially complements other methods such as linkage and genomewide association analysis. It is particularly suited for divergent populations with large allele frequency differences and for traits and characters that are found at different rates in ancestral populations.

General Discussion

In this thesis the main objectives were to use genomewide SNPs to assess the impact of admixture at different regions of cattle genome, to detect genes of importance and to assess the nature of admixture in African cattle. In order to achieve the planned objectives, informative markers utilising the genetic divergence existing between *Bos taurus* and *Bos indicus* were selected to give a uniform coverage across the whole genome. The main tasks completed may be summarized in four steps. The first task detailed the nature of admixture in East African zebu using the Illumina 50k SNPs genomewide assay. This gave a first detailed description of the genomic distribution of ancestral influences from this millennia old hybridisation and suggested the effects of natural selection at specific locations. Second a custom design of Golden Gate 768 SNPs assay gave a cost effective means for a genomewide admixture scan and was initially tested in several admixed populations collected from various geographic regions. This allowed the most accurate assessment of this phenomenon to date and also yielded estimates of the time depth of admixtures which were compatible with archaeological knowledge. Third, the Golden Gate genome wide scan for ancestry was used to search for associations with bovine tuberculosis susceptibility in hybrid animals, giving a suggestive positive signal in the region of a toll-like receptor (*TLR*) gene cluster. These data also gave refined estimates of the extent that European cattle (*Bos taurus*) are more susceptible than African zebu cattle (*Bos indicus*). Fourth, the same genomic data were used to map additional traits such as pigmentation and conformation. A clear mapping of a red color phenotype locus to the MC1R (a known pigmentation gene) region gave a positive quality control for the

genotype data and tested the statistical algorithm implemented in admixture mapping analysis for trait identification. Overall, the results shown in this work further illustrate the utility of admixture mapping models which have been established in various human population studies.

Admixture mapping using human genetic data has been increasingly used to decipher novel genetic associations to various complex traits as well as to review previously known genes. Many successful studies have been reported, mainly focused on chronic illnesses such as kidney failure, obesity, blood pressure and various forms of cancer (WINKLER *et al.* 2010). Limited genetic variants have also emerged in the field of infectious diseases and promise more interesting findings in the years ahead. This will involve further refinement of the algorithms and statistical analyses, plus increasingly accurate case definition and detection are important components in refining methods. Although the relevant genetic typing technology has been improving at an exponential rate, the existing statistical models require considerable computational time and often lack the capability needed for efficient and flexible user defined bioinformatics analysis. Estimation of posterior distributions for various parameters remains a significant computational expense related to the complexity of the models.

There were some challenges in the admixture mapping work presented in this thesis. For example it was a challenge to accurately outline the nature of ancient admixture events. The models suffer in cases, as with the samples here, involving a three way admixture which has taken place in layered and complex events of introgression. Multiple MCMC models were needed to test explore what were consistent with historical records. Secondly, an initial plan to use pure West African *Bos taurus* genotype data imported from

a previously typed data base from the bovine Hapmap consortium was not feasible due to the challenge faced by platform incompatibility.

Marker density was limited in this mapping project by budget concerns. However, in the identification of genetic regions correlated with the various traits here, including susceptibility to bovine tuberculosis, the number of markers used here were sufficient because of the small number of generation times since admixture in the Ethiopian-European hybrid cattle. However, an average scale of 4MB inter-marker distance inevitably gives peaks that contain numerous genes and hence are limited in the information required to pin point functional variant genes. Fine mapping admixture analysis is required but may not be an ideal next step in the same population with its typical 3 - 4 generation admixture. However, this challenge could be overcome by applying the same approach coupled with fine mapping admixture analysis in a population where there is older admixture event. A final challenge arises from the disease classification for bovine tuberculosis which was entirely dependent on CIDT read outs; the issue of sensitivity may have had some impact. Given the amount of animals screened in the area, it could be possible that valuable samples were dropped due to the lack of sensitivity of skin testing. It is important to supplement animal screening based on more sensitive serological tests and if possible to include a subset of population with a post-mortem showing positive pathology for tuberculosis lesions and to incorporate animals tested with the gold standard method (culture isolation of *Mycobacterium bovis*).

Despite the limitations of the current work, it points toward important information that with development could eventually be translated to improve livestock sector in Ethiopia and sub-Saharan Africa. For example, such work could work toward the molecular

prediction of breeding potential within African breeds for various complex traits and could contribute to enhancing and conserve the genetic resources represented by the locally adapted breeds in the continent.

Genetic mapping technology has innovated quickly and in this post-genomic era we have an explosion of livestock genotype data. The advent of next generation sequencing has pushed the capacity for assaying genetic variation even further. This technology affords the opportunity for re-sequencing of genomes as a standard technique and will enhance our understanding of genetic differences that impinge on health and disease conditions. In parallel, the companion computational methods and means of extracting vital information from this massive data base are also progressing. This coming era will undoubtedly afford opportunities to undertake genomic researches in a more comprehensive and focused approach to address various complex issues in the areas addressed in this thesis. I hope to be involved in this process particularly in the area of genomic analysis of host-infectious disease interactions relevant to sub-Saharan Africa.

CHAPTER SEVEN: LITERATURE CITED

7.1Reference

- ACHILLI, A., S. BONFIGLIO, A. OLIVIERI, A. MALUSA, M. PALA *et al.*, 2009 The Multifaceted Origin of Taurine Cattle Reflected by the Mitochondrial Genome. *PLoS ONE* 4(6):e5753.
- ACHILLI, A., A. OLIVIERI, M. PELLECCIA, C. UBOLDI, L. COLLI *et al.*, 2008 Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Curr Biol* 18: R157-158.
- AJMONE-MARSAN, P., J. F. GARCIA, J. A. LENSTRA and G. CONSORTIUM, 2010 On the Origin of Cattle: How Aurochs Became Cattle and Colonized the World. *Evolutionary Anthropology* 19: 148-157.
- ALDRICH, M. C., S. SELVIN, H. M. HANSEN, L. F. BARCELLOS, M. R. WRENSCH *et al.*, 2008 Comparison of statistical methods for estimating genetic admixture in a lung cancer study of African Americans and Latinos. *Am J Epidemiol* 168: 1035-1046.
- ALLEN, A. R., G. MINOZZI, E. J. GLASS, R. A. SKUCE, S. W. J. MCDOWELL *et al.*, 2010 Bovine tuberculosis: the genetic basis of host susceptibility. *Proceedings of the Royal Society B-Biological Sciences* 277: 2737-2745.
- AMENI, G., A. ASEFFA, H. ENGERS, D. YOUNG, S. GORDON *et al.*, 2007a High prevalence and increased severity of pathology of bovine tuberculosis in Holsteins compared to zebu breeds under field cattle husbandry in central Ethiopia. *Clinical and Vaccine Immunology* 14: 1356-1361.
- AMENI, G., A. ASEFFA, H. ENGERS, D. YOUNG, G. HEWINSON *et al.*, 2006 Cattle husbandry in Ethiopia is a predominant factor affecting the pathology of bovine tuberculosis and gamma interferon responses to mycobacterial antigens. *Clinical and Vaccine Immunology* 13: 1030-1036.
- AMENI, G., A. ASEFFA, G. HEWINSON and M. VORDERMEIER, 2010 Comparison of different testing schemes to increase the detection *Mycobacterium bovis* infection in Ethiopian cattle. *Trop Anim Health Prod* 42: 375-383.
- AMENI, G., A. ASEFFA, A. SIRAK, H. ENGERS, D. B. YOUNG *et al.*, 2007c Effect of skin testing and segregation on the prevalence of bovine tuberculosis, and molecular typing of *Mycobacterium bovis*, in Ethiopia. *Veterinary Record* 161: 782-786.
- AMENI, G., G. HEWINSON, A. ASEFFA, D. YOUNG and M. VORDERMEIER, 2008 Appraisal of interpretation criteria for the comparative intradermal tuberculin test for diagnosis of tuberculosis in cattle in central Ethiopia. *Clin Vaccine Immunol* 15: 1272-1276.

- AMOS, W., E. DRISCOLL and J. I. HOFFMAN, 2010 Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proc Biol Sci* 278: 1183-1188.
- ANDERUNG, C., L. HELLBORG, J. SEDDON, O. HANOTTE and A. GOTHERSTROM, 2007 Investigation of X- and Y-specific single nucleotide polymorphisms in taurine (*Bos taurus*) and indicine (*Bos indicus*) cattle. *Anim Genet* 38: 595-600.
- ASHWELL, M. S., C. P. VAN TASSELL and T. S. SONSTEGARD, 2001 A genome scan to identify quantitative trait loci affecting economically important traits in a US Holstein population. *J Dairy Sci* 84: 2535-2542.
- AUBURN, S., A. E. FRY, T. G. CLARK, S. CAMPINO, M. DIAKITE *et al.*, 2010 Further Evidence Supporting a Role for Gs Signal Transduction in Severe Malaria Pathogenesis. *PLoS ONE* 5: -.
- BARON, M. D., 2005 Wild-type Rinderpest virus uses SLAM (CD150) as its receptor. *J Gen Virol* 86: 1753-1757.
- BARTHEL, R., J. A. PIEDRAHITA, D. N. MCMURRAY, J. PAYEUR, D. BACA *et al.*, 2000 Pathologic findings and association of *Mycobacterium bovis* infection with the bovine NRAMP1 gene in cattle from herds with naturally occurring tuberculosis. *Am J Vet Res* 61: 1140-1144.
- BASU, A., H. TANG, D. ARNETT, C. C. GU, T. MOSLEY *et al.*, 2009 Admixture mapping of quantitative trait loci for BMI in African Americans: evidence for loci on chromosomes 3q, 5q, and 15q. *Obesity (Silver Spring)* 17: 1226-1231.
- BEARDEN, I. G., D. BEAVIS, C. BESLIU, B. BUDICK, H. BOGGILD *et al.*, 2006 Forward rapidity results from RHIC. *Revista Mexicana De Fisica* 52: 112-118.
- BEN-ALI, M., L. B. BARREIRO, A. CHABBOU, R. HALTITI, E. BRAHAM *et al.*, 2007 Promoter and neck region length variation of DC-SIGN is not associated with susceptibility to tuberculosis in Tunisian patients. *Hum Immunol* 68: 908-912.
- BERG, S., R. FIRDESSA, M. HABTAMU, E. GADISA, A. MENGISTU *et al.*, 2009 The burden of mycobacterial disease in ethiopian cattle: implications for public health. *PLoS ONE* 4: e5068.
- BERMINGHAM, M. L., S. J. MORE, M. GOOD, A. R. CROMIE, I. M. HIGGINS *et al.*, 2009 Genetics of tuberculosis in Irish Holstein-Friesian dairy herds. *J Dairy Sci* 92: 3447-3456.
- BILAL, S., M. IQBAL, P. MURPHY and J. POWER, 2010 Human bovine tuberculosis - remains in the differential. *J Med Microbiol* 59: 1379-1382.

- BISCARINI, F., H. BOVENHUIS, J. A. VAN ARENDONK, H. K. PARMENTIER, A. P. JUNGERIUS *et al.*, 2010 Across-line SNP association study of innate and adaptive immune response in laying hens. *Anim Genet* 41: 26-38.
- BLANCO, F. C., P. SCHIERLOH, M. V. BIANCO, K. CAIMI, V. MEIKLE *et al.*, 2009 Study of the immunological profile towards *Mycobacterium bovis* antigens in naturally infected cattle. *Microbiol Immunol* 53: 460-467.
- BOCK, C. H., A. G. SCHWARTZ, J. J. RUTERBUSCH, A. M. LEVIN, C. NESLUND-DUDAS *et al.*, 2009 Results from a prostate cancer admixture mapping study in African-American men. *Hum Genet* 126: 637-642.
- BOUMAN, H. A., O. ULLOA, D. J. SCANLAN, K. ZWIRGLMAIER, W. K. LI *et al.*, 2006 Oceanographic basis of the global surface distribution of *Prochlorococcus* ecotypes. *Science* 312: 918-921.
- BOUYER, J., S. RAVEL, J. P. DUJARDIN, T. DE MEEUS, L. VIAL *et al.*, 2007 Population structuring of *Glossina palpalis gambiensis* (Diptera: Glossinidae) according to landscape fragmentation in the Mouhoun river, Burkina Faso. *J Med Entomol* 44: 788-795.
- BRADLEY, D. G., R. T. LOFTUS, P. CUNNINGHAM and D. E. MACHUGH, 1998 Genetics and domestic cattle origins. *Evolutionary Anthropology* 6: 79-86.
- BRADLEY, D. G., D. E. MACHUGH, P. CUNNINGHAM and R. T. LOFTUS, 1996 Mitochondrial diversity and the origins of African and European cattle. *Proc Natl Acad Sci U S A* 93: 5131-5135.
- BRADLEY, D. G., D. E. MACHUGH, R. T. LOFTUS, R. S. SOW, C. H. HOSTE *et al.*, 1994 Zebu-taurine variation in Y chromosomal DNA: a sensitive assay for genetic introgression in west African trypanotolerant cattle populations. *Anim Genet* 25: 7-12.
- BRAUNSCHWEIG, M. H., 2010 Mutations in the bovine ABCG2 and the ovine MSTN gene added to the few quantitative trait nucleotides identified in farm animals: a mini-review. *J Appl Genet* 51: 289-297.
- BROSCH, R., S. V. GORDON, A. PYM, K. EIGLMEIER, T. GARNIER *et al.*, 2000 Comparative genomics of the mycobacteria. *Int J Med Microbiol* 290: 143-152.
- BROTHERSTONE, S., I. M. WHITE, M. COFFEY, S. H. DOWNS, A. P. MITCHELL *et al.*, 2010 Evidence of genetic resistance of cattle to infection with *Mycobacterium bovis*. *J Dairy Sci* 93: 1234-1242.
- BRYC, K., C. VELEZ, T. KARAFET, A. MORENO-ESTRADA, A. REYNOLDS *et al.*, 2010 Colloquium paper: genome-wide patterns of population structure and admixture

- among Hispanic/Latino populations. *Proc Natl Acad Sci U S A* 107 Suppl 2: 8954-8961.
- BUERKLE, C. A., and C. LEXER, 2008 Admixture as the basis for genetic mapping. *Trends in Ecology & Evolution* 23: 686-694.
- BURT, D. W., 2009 The cattle genome reveals its secrets. *J Biol* 8: 36.
- BURTON, P. R., D. G. CLAYTON, L. R. CARDON, N. CRADDOCK, P. DELOUKAS *et al.*, 2007 Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* 39: 1329-1337.
- CARDON, L. R., and G. R. ABECASIS, 2003 Using haplotype blocks to map human complex trait loci. *Trends Genet* 19: 135-140.
- CARMICHAEL, J., 1941 Bovine tuberculosis in the tropics with special reference to Uganda. Part 11. Experimental tuberculosis in zebu cattle. *Vet J* 97: 329-339.
- CASAS, E., W. BARENDSE, J. E. BEEVER, B. M. BURNS, S. K. DAVIS *et al.*, 1999 Bovine chromosome 4 workshop: consensus and comprehensive linkage maps. *Anim Genet* 30: 375-377.
- CASTELNUOVO, E., J. THOMPSON-COON, M. PITT, M. CRAMP, U. SIEBERT *et al.*, 2006 The cost-effectiveness of testing for hepatitis C in former injecting drug users. *Health Technol Assess* 10: iii-iv, ix-xii, 1-93.
- CHANG, M., W. JIN and S. C. SUN, 2009 Peli1 facilitates TRIF-dependent Toll-like receptor signaling and proinflammatory cytokine production. *Nat Immunol* 10: 1089-1095.
- CHEN, G., D. SHRINER, J. ZHOU, A. DOUMATEY, H. HUANG *et al.*, 2010a Development of admixture mapping panels for African Americans from commercial high-density SNP arrays. *BMC Genomics* 11: 417.
- CHEN, S., B. Z. LIN, M. BAIG, B. MITRA, R. J. LOPES *et al.*, 2010b Zebu cattle are an exclusive legacy of the South Asia neolithic. *Mol Biol Evol* 27: 1-6.
- CHENG, C. Y., W. H. KAO, N. PATTERSON, A. TANDON, C. A. HAIMAN *et al.*, 2009 Admixture mapping of 15,280 African Americans identifies obesity susceptibility loci on chromosomes 5 and X. *PLoS Genet* 5: e1000490.
- CHENG, C. Y., D. REICH, J. CORESH, E. BOERWINKLE, N. PATTERSON *et al.*, 2010a Admixture mapping of obesity-related traits in African Americans: the Atherosclerosis Risk in Communities (ARIC) Study. *Obesity (Silver Spring)* 18: 563-572.

- CHENG, C. Y., D. REICH, T. Y. WONG, R. KLEIN, B. E. KLEIN *et al.*, 2010b Admixture mapping scans identify a locus affecting retinal vascular caliber in hypertensive African Americans: the Atherosclerosis Risk in Communities (ARIC) study. *PLoS Genet* 6: e1000908.
- COATS, S. R., R. A. REIFE, B. W. BAINBRIDGE, T. T. PHAM and R. P. DARVEAU, 2003 *Porphyromonas gingivalis* lipopolysaccharide antagonizes *Escherichia coli* lipopolysaccharide at toll-like receptor 4 in human endothelial cells. *Infect Immun* 71: 6799-6807.
- COLLINS-SCHRAMM, H. E., C. M. PHILLIPS, D. J. OPERARIO, J. S. LEE, J. L. WEBER *et al.*, 2002 Ethnic-difference markers for use in mapping by admixture linkage disequilibrium. *Am J Hum Genet* 70: 737-750.
- COOPER, R. S., B. TAYO and X. ZHU, 2008 Genome-wide association studies: implications for multiethnic samples. *Hum Mol Genet* 17: R151-155.
- CORR, S. C., and L. A. O'NEILL, 2009 Genetic variation in Toll-like receptor signalling and the risk of inflammatory and immune diseases. *J Innate Immun* 1: 350-357.
- CSA, 2010 Report on livestock and livestock characteristics (private peasant holdings), pp. 12 in *Agricultural sample survey 2009/2010*.
- DADI, H., M. TIBBO, Y. TAKAHASHI, K. NOMURA, H. HANADA *et al.*, 2008 Microsatellite analysis reveals high genetic diversity but low genetic structure in Ethiopian indigenous cattle populations. *Anim Genet* 39: 425-431.
- DAGRIS, 2007 Domestic Animal Genetic Resources Information System, <http://dagris.ilri.cgiar.org/display.asp?ID=30>, pp. in *General Breed Information*.
- DARVASI, A., and S. SHIFMAN, 2005 The beauty of admixture. *Nat Genet* 37: 118-119.
- DAYO, G. K., S. THEVENON, D. BERTHIER, K. MOAZAMI-GOUDARZI, C. DENIS *et al.*, 2009 Detection of selection signatures within candidate regions underlying trypanotolerance in outbred cattle populations. *Mol Ecol* 18: 1801-1813.
- DEMELASH, B., F. INANGOLET, J. OLOYA, B. ASSEGED, M. BADASO *et al.*, 2009 Prevalence of Bovine tuberculosis in Ethiopian slaughter cattle based on post-mortem examination. *Tropical Animal Health and Production* 41: 755-765.
- DESEPULVEDA, P., J. L. GUENET and J. J. PANTHIER, 1995 Phenotypic Reversions at the W/Kit Locus Mediated by Mitotic Recombination in Mice. *Molecular and Cellular Biology* 15: 5898-5905.
- DOHERTY, T. M., and M. ARDITI, 2004 TB, or not TB: that is the question -- does TLR signaling hold the answer? *J Clin Invest* 114: 1699-1703.

- DOMENECH, J., J. LUBROTH and K. SUMPTION, 2010 Immune protection in animals: the examples of rinderpest and foot-and-mouth disease. *J Comp Pathol* 142 Suppl 1: S120-124.
- DREGER, D. L., and S. M. SCHMUTZ, 2010 The variant red coat colour phenotype of Holstein cattle maps to BTA27. *Anim Genet* 41: 109-112.
- EDWARDS, C. J., J. F. BAIRD and D. E. MACHUGH, 2007 Taurine and zebu admixture in Near Eastern cattle: a comparison of mitochondrial, autosomal and Y-chromosomal data. *Anim Genet* 38: 520-524.
- ELBEIN, S. C., S. K. DAS, D. M. HALLMAN, C. L. HANIS and S. J. HASSTEDT, 2009 Genome-wide linkage and admixture mapping of type 2 diabetes in African American families from the American Diabetes Association GENNID (Genetics of NIDDM) Study Cohort. *Diabetes* 58: 268-274.
- EPSTEIN, H., and I. L. MASON, 1971 *The origin of the domestic animals of Africa*. Edition Leipzig, Leipzig.
- FALUSH, D., M. STEPHENS and J. K. PRITCHARD, 2003 Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.
- FALUSH, D., M. STEPHENS and J. K. PRITCHARD, 2007 Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7: 574-578.
- FEJERMAN, L., C. A. HAIMAN, D. REICH, A. TANDON, R. C. DEO *et al.*, 2009 An admixture scan in 1,484 African American women with breast cancer. *Cancer Epidemiol Biomarkers Prev* 18: 3110-3117.
- FREEMAN, A. R., D. G. BRADLEY, S. NAGDA, J. P. GIBSON and O. HANOTTE, 2006a Combination of multiple microsatellite data sets to investigate genetic diversity and admixture of domestic cattle. *Anim Genet* 37: 1-9.
- FREEMAN, A. R., C. J. HOGGART, O. HANOTTE and D. G. BRADLEY, 2006b Assessing the relative ages of admixture in the bovine hybrid zones of Africa and the Near East using X chromosome haplotype mosaicism. *Genetics* 173: 1503-1510.
- FREEMAN, A. R., C. M. MEGHEN, D. E. MACHUGH, R. T. LOFTUS, M. D. ACHUKWI *et al.*, 2004 Admixture and diversity in West African cattle populations. *Mol Ecol* 13: 3477-3487.
- GAGLIARDI, M. C., R. TELONI, F. GIANNONI, M. PARDINI, V. SARGENTINI *et al.*, 2005 *Mycobacterium bovis* Bacillus Calmette-Guerin infects DC-SIGN- dendritic cell and causes the inhibition of IL-12 and the enhancement of IL-10 production. *J Leukoc Biol* 78: 106-113.

- GANOPOLSKI, A., C. KUBATZKI, M. CLAUSSEN, V. V. BROVKIN and V. V. PETOUKHOV, 1998 The influence of vegetation-atmosphere-ocean interaction on climate during the mid-holocene. *Science* 280: 1916-1919.
- GAUTIER, M., T. FARAUT, K. MOAZAMI-GOUDARZI, V. NAVRATIL, M. FOGGIO *et al.*, 2007 Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics* 177: 1059-1070.
- GAUTIER, M., L. FLORI, A. RIEBLER, F. JAFFREZIC, D. LALOE *et al.*, 2009 A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics* 10: 550.
- GIBBS, R. A., J. F. TAYLOR, C. P. VAN TASSELL, W. BARENDSE, K. A. EVERSOLE *et al.*, 2009 Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324: 528-532.
- GREEN, R. E., J. KRAUSE, A. W. BRIGGS, T. MARICIC, U. STENZEL *et al.*, 2010 A draft sequence of the Neandertal genome. *Science* 328: 710-722.
- GREENSPAN, G., D. GEIGER, F. GOTCH, M. BOWER, S. PATTERSON *et al.*, 2004 Model-based inference of recombination hotspots in a highly variable oncogene [corrected]. *J Mol Evol* 58: 239-251.
- GRIFFIN, J. F. T., and C. G. MACKINTOSH, 2000 Tuberculosis in deer: Perceptions, problems and progress. *Veterinary Journal* 160: 202-219.
- GRIFFIN, J. F. T., C. R. RODGERS, S. LIGGETT and C. G. MACKINTOSH, 2006 Tuberculosis in ruminants: Characteristics of intra-tonsillar *Mycobacterium bovis* infection models in cattle and deer. *Tuberculosis* 86: 404-418.
- GUTIERREZ-GIL, B., P. WIENER and J. L. WILLIAMS, 2007 Genetic effects on coat colour in cattle: dilution of eumelanin and pheomelanin pigments in an F2-Backcross Charolais x Holstein population. *BMC Genet* 8: 56.
- HALDER, I., and M. D. SHRIVER, 2003 Measuring and using admixture to study the genetics of complex diseases. *Hum Genomics* 1: 52-62.
- HANOTTE, O., D. G. BRADLEY, J. W. OCHIENG, Y. VERJEE, E. W. HILL *et al.*, 2002 African pastoralism: genetic imprints of origins and migrations. *Science* 296: 336-339.
- HANOTTE, O., Y. RONIN, M. AGABA, P. NILSSON, A. GELHAUS *et al.*, 2003 Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proc Natl Acad Sci U S A* 100: 7443-7448.

- HANOTTE, O., C. L. TAWAH, D. G. BRADLEY, M. OKOMO, Y. VERJEE *et al.*, 2000 Geographic distribution and frequency of a taurine *Bos taurus* and an indicine *Bos indicus* Y specific allele amongst sub-saharan African cattle breeds. *Mol Ecol* 9: 387-396.
- HAYES, B. J., A. J. CHAMBERLAIN, H. MCPARTLAN, I. MACLEOD, L. SETHURAMAN *et al.*, 2007 Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. *Genet Res* 89: 215-220.
- HAYES, B. J., S. LIEN, H. NILSEN, H. G. OLSEN, P. BERG *et al.*, 2008 The origin of selection signatures on bovine chromosome 6. *Anim Genet* 39: 105-111.
- HAYES, B. J., P. M. VISSCHER, H. C. MCPARTLAN and M. E. GODDARD, 2003 Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res* 13: 635-643.
- HELDWEIN, K. A., and M. J. FENTON, 2002 The role of Toll-like receptors in immunity against mycobacterial infection. *Microbes Infect* 4: 937-944.
- HELDWEIN, K. A., M. D. LIANG, T. K. ANDRESEN, K. E. THOMAS, A. M. MARTY *et al.*, 2003 TLR2 and TLR4 serve distinct roles in the host immune response against *Mycobacterium bovis* BCG. *J Leukoc Biol* 74: 277-286.
- HESS, M., T. GOLDAMMER, A. GELHAUS, K. RIED, G. RAPPOLD *et al.*, 1999 Physical assignment of the bovine MHC class IIa and class IIb genes. *Cytogenet Cell Genet* 85: 244-247.
- HIENDLEDER, S., H. LEWALSKI and A. JANKE, 2008 Complete mitochondrial genomes of *Bos taurus* and *Bos indicus* provide new insights into intra-species variation, taxonomy and domestication. *Cytogenet Genome Res* 120: 150-156.
- HIMER, L., B. CSOKA, Z. SELMECZY, B. KOSCSO, T. POCZA *et al.*, 2010 Adenosine A2A receptor activation protects CD4+ T lymphocytes against activation-induced cell death. *Faseb J* 24: 2631-2640.
- HIROOKA, H., D. J. DE KONING, J. A. VAN ARENDONK, B. HARLIZIUS, P. N. DE GROOT *et al.*, 2002 Genome scan reveals new coat color loci in exotic pig cross. *J Hered* 93: 1-8.
- HOGGART, C. J., M. D. SHRIVER, R. A. KITTLES, D. G. CLAYTON and P. M. MCKEIGUE, 2004 Design and analysis of admixture mapping studies. *Am J Hum Genet* 74: 965-978.
- HORNYAK, T. J., 2006 The developmental biology of melanocytes and its application to understanding human congenital disorders of pigmentation. *Adv Dermatol* 22: 201-218.

- HUANG DA, W., B. T. SHERMAN and R. A. LEMPICKI, 2009 Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44-57.
- IBEAGHA-AWEMU, E. M., O. C. JANN, C. WEIMANN and G. ERHARDT, 2004 Genetic diversity, introgression and relationships among West/Central African cattle breeds. *Genet Sel Evol* 36: 673-690.
- IBEAGHA-AWEMU, E. M., P. KGWATALALA and X. ZHAO, 2008 A critical analysis of production-associated DNA polymorphisms in the genes of cattle, goat, sheep, and pig. *Mamm Genome* 19: 591-617.
- IBRAHIM, H., and E. OLALOKU, 2000 Improving cattle for milk, meat and traction. ILRI Manual 4. ILRI (International Livestock Research Institute), pp. 135, Nairobi, Kenya.
- JACKSON, I. J., 1994 Molecular and developmental genetics of mouse coat color. *Annu Rev Genet* 28: 189-217.
- KAASSCHIETER, G. A., R. DE JONG, J. B. SCHIERE and D. ZWART, 1992 Towards a sustainable livestock production in developing countries and the importance of animal health strategy therein. *Vet Q.* 14: 66 - 75.
- KAO, W. H., M. J. KLAG, L. A. MEONI, D. REICH, Y. BERTHIER-SCHAAD *et al.*, 2008 MYH9 is associated with nondiabetic end-stage renal disease in African Americans. *Nat Genet* 40: 1185-1192.
- KARUGIA, J. T., O. A. MWAI, R. KAITO, C. B. WOLLNY, A. G. DRUCKER *et al.*, 2001 Economic analysis of crossbreeding programs in sub-Saharan Africa: A conceptual framework and Kenyan case. *Animal Genetic Resources Research Report No. 2.*, pp., Nairobi, Kenya: ILRI.
- KHATKAR, M. S., K. R. ZENGER, M. HOBBS, R. J. HAWKEN, J. A. CAVANAGH *et al.*, 2007 A primary assembly of a bovine haplotype block map based on a 15,036-single-nucleotide polymorphism panel genotyped in holstein-friesian cattle. *Genetics* 176: 763-772.
- KHOR, C. C., 2009 Genotyping methods to analyse polymorphisms in Toll-like receptors and disease. *Methods Mol Biol* 517: 297-309.
- KNEELAND, J., C. LI, J. BASARAB, W. M. SNELLING, B. BENKEL *et al.*, 2004 Identification and fine mapping of quantitative trait loci for growth traits on bovine chromosomes 2, 6, 14, 19, 21, and 23 within one commercial line of *Bos taurus*. *J Anim Sci* 82: 3405-3414.

- KOPP, J. B., M. W. SMITH, G. W. NELSON, R. C. JOHNSON, B. I. FREEDMAN *et al.*, 2008 MYH9 is a major-effect risk gene for focal segmental glomerulosclerosis. *Nat Genet* 40: 1175-1184.
- KORBEL, D. S., B. E. SCHNEIDER and U. E. SCHAIBLE, 2008 Innate immunity in tuberculosis: myths and truth. *Microbes Infect* 10: 995-1004.
- LI, M. H., M. ZERABRUK, O. VANGEN, I. OLSAKER and J. KANTANEN, 2007 Reduced genetic structure of north Ethiopian cattle revealed by Y-chromosome analysis. *Heredity* 98: 214-221.
- LIPTAK, A. R., B. T. SULLIVAN, L. E. HENKES, M. P. WIJAYAGUNAWARDANE, A. MIYAMOTO *et al.*, 2005 Cooperative expression of monocyte chemoattractant protein 1 within the bovine corpus luteum: evidence of immune cell-endothelial cell interactions in a coculture system. *Biol Reprod* 72: 1169-1176.
- LIU, L., B. HARRIS, M. KEEHAN and Y. ZHANG, 2009 Genome scan of pigmentation traits in Friesian-Jersey crossbred cattle. *J Genet Genomics* 36: 661-666.
- LOFTUS, R. T., O. ERTUGRUL, A. H. HARBA, M. A. EL-BARODY, D. E. MACHUGH *et al.*, 1999a A microsatellite survey of cattle from a centre of origin: the Near East. *Mol Ecol* 8: 2015-2022.
- LOFTUS, R. T., D. E. MACHUGH, D. G. BRADLEY, P. M. SHARP and P. CUNNINGHAM, 1994a Evidence for two independent domestications of cattle. *Proc Natl Acad Sci U S A* 91: 2757-2761.
- LOFTUS, R. T., D. E. MACHUGH, L. O. NGERE, D. S. BALAIN, A. M. BADI *et al.*, 1994b Mitochondrial genetic variation in European, African and Indian cattle populations. *Anim Genet* 25: 265-271.
- MACHUGH, D. E., R. T. LOFTUS, P. CUNNINGHAM and D. G. BRADLEY, 1998 Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Anim Genet* 29: 333-340.
- MACHUGH, D. E., M. D. SHRIVER, R. T. LOFTUS, P. CUNNINGHAM and D. G. BRADLEY, 1997 Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* 146: 1071-1086.
- MACNEIL, M. D., and M. D. GROSZ, 2002 Genome-wide scans for QTL affecting carcass traits in Hereford x composite double backcross populations. *J Anim Sci* 80: 2316-2324.
- MANICHAIKUL, A., J. DUPUIS, S. SEN and K. W. BROMAN, 2006 Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. *Genetics* 174: 481-489.

- MANOLIO, T. A., F. S. COLLINS, N. J. COX, D. B. GOLDSTEIN, L. A. HINDORFF *et al.*, 2009 Finding the missing heritability of complex diseases. *Nature* 461: 747-753.
- MAO, X., A. W. BIGHAM, R. MEI, G. GUTIERREZ, K. M. WEISS *et al.*, 2007 A genomewide admixture mapping panel for Hispanic/Latino populations. *Am J Hum Genet* 80: 1171-1178.
- MARTINEZ-ROYO, A., L. ORDOVAS, P. ZARAGOZA, J. ALTARRIBA, M. SERRANO *et al.*, 2010 The bovine annexin 9 gene (ANXA9) is significantly associated with milk-fat yield in a Spanish Holstein-Friesian population. *Res Vet Sci* 88: 452-455.
- MATTIOLI, R. C., V. S. PANDEY, M. MURRAY and J. L. FITZPATRICK, 2000 Immunogenetic influences on tick resistance in African cattle with particular reference to trypanotolerant N'Dama (*Bos taurus*) and trypanosusceptible Gobra zebu (*Bos indicus*) cattle. *Acta Tropica* 75: 263-277.
- MATUKUMALLI, L. K., C. T. LAWLEY, R. D. SCHNABEL, J. F. TAYLOR, M. F. ALLAN *et al.*, 2009 Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* 4: e5350.
- MCCLURE, M. C., N. S. MORSCI, R. D. SCHNABEL, J. W. KIM, P. YAO *et al.*, 2010 A genome scan for quantitative trait loci influencing carcass, post-natal growth and reproductive traits in commercial Angus cattle. *Anim Genet* 41: 597-607.
- MCKEIGUE, P. M., 1998 Mapping genes that underlie ethnic differences in disease risk: methods for detecting linkage in admixed populations, by conditioning on parental admixture. *Am J Hum Genet* 63: 241-251.
- MCKEIGUE, P. M., 2005 Prospects for admixture mapping of complex traits. *Am J Hum Genet* 76: 1-7.
- MEADE, K. G., E. GORMLEY, M. B. DOYLE, T. FITZSIMONS, C. O'FARRELLY *et al.*, 2007 Innate gene repression associated with *Mycobacterium bovis* infection in cattle: toward a gene signature of disease. *BMC Genomics* 8: 400.
- MICHEL, A. L., B. MULLER and P. D. VAN HELDEN, 2010 *Mycobacterium bovis* at the animal-human interface: a problem, or not? *Vet Microbiol* 140: 371-381.
- MOHAMMAD, J. A., N. TAYE, G., 2007 Geographic distribution of cattle and shoats populations and their market supply sheds in Ethiopia, pp. 26 in *Discussion Paper No.2. Improving market opportunities. ILRI(International Livestock Research Institute)*, Nairobi, Kenya.
- MOLLER, M., E. DE WIT and E. G. HOAL, 2010 Past, present and future directions in human genetic susceptibility to tuberculosis. *FEMS Immunol Med Microbiol* 58: 3-26.

- MONTANA, G., and C. HOGGART, 2007 Statistical software for gene mapping by admixture linkage disequilibrium. *Brief Bioinform* 8: 393-395.
- MONTANA, G., and J. K. PRITCHARD, 2004 Statistical tests for admixture mapping with case-control and cases-only data. *Am J Hum Genet* 75: 771-789.
- MORRIS, C. A., W. S. PITCHFORD, N. G. CULLEN, A. K. ESMAILIZADEH, S. M. HICKEY *et al.*, 2009 Quantitative trait loci for live animal and carcass composition traits in Jersey and Limousin back-cross cattle finished on pasture or feedlot. *Anim Genet* 40: 648-654.
- MUCHADEYI, F. C., H. EDING, C. B. WOLLNY, E. GROENEVELD, S. M. MAKUZA *et al.*, 2007 Absence of population substructuring in Zimbabwe chicken ecotypes inferred using microsatellite analysis. *Anim Genet* 38: 332-339.
- MURRAY, C., E. HUERTA-SANCHEZ, F. CASEY and D. G. BRADLEY, 2010 Cattle demographic history modelled from autosomal sequence variation. *Philos Trans R Soc Lond B Biol Sci* 365: 2531-2539.
- MURRAY, M., and J. C. M. TRAIL, 1984 Genetic-Resistance to Animal Trypanosomiasis in Africa. *Preventive Veterinary Medicine* 2: 541-551.
- NEILL, S. D., D. G. BRYSON and J. M. POLLOCK, 2001 Pathogenesis of tuberculosis in cattle. *Tuberculosis (Edinb)* 81: 79-86.
- NICHOLSON, M. J. B., M.H., 1986 A Guide to Condition Scoring of Zebu Cattle, pp. 29pp. INTERNATIONAL LIVESTOCK CENTRE FOR AFRICA, ADDIS ABABA , ETHIOPIA.
- O'REILLY, L. M., and C. J. DABORN, 1995 The epidemiology of *Mycobacterium bovis* infections in animals and man: a review. *Tuber Lung Dis* 76 Suppl 1: 1-46.
- OIE, 2004 World organisation for Animal Health: Manual of Diagnostic Tests and Vaccines for Terrestrial Animals, pp. 2.3.3 in *BOVINE TUBERCULOSIS*.
- PERRIERE, G., and M. GOUY, 1996 WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie* 78: 364-369.
- PIPER, E. K., N. N. JONSSON, C. GONDRO, A. E. LEW-TABOR, P. MOOLHUIZEN *et al.*, 2009 Immunological profiles of *Bos taurus* and *Bos indicus* cattle infested with the cattle tick, *Rhipicephalus (Boophilus) microplus*. *Clin Vaccine Immunol* 16: 1074-1086.
- POLLOCK, J. M., B. M. BUDDLE and P. ANDERSEN, 2001a Towards more accurate diagnosis of bovine tuberculosis using defined antigens. *Tuberculosis* 81: 65-69.

- POLLOCK, J. M., J. MCNAIR, M. D. WELSH, R. M. GIRVIN, H. E. KENNEDY *et al.*, 2001b
Immune responses in bovine tuberculosis. *Tuberculosis (Edinb)* 81: 103-107.
- PRICE, A. L., N. PATTERSON, F. YU, D. R. COX, A. WALISZEWSKA *et al.*, 2007 A
genomewide admixture map for Latino populations. *Am J Hum Genet* 80: 1024-1036.
- PRICE, A. L., A. TANDON, N. PATTERSON, K. C. BARNES, N. RAFAELS *et al.*, 2009
Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS Genet* 5: e1000519.
- PRITCHARD, J. K., and N. A. ROSENBERG, 1999 Use of unlinked genetic markers to detect
population stratification in association studies. *Am J Hum Genet* 65: 220-228.
- PRITCHARD, J. K., M. STEPHENS and P. DONNELLY, 2000 Inference of population
structure using multilocus genotype data. *Genetics* 155: 945-959.
- PURCELL, S., and P. SHAM, 2004 Properties of structured association approaches to
detecting population stratification. *Hum Hered* 58: 93-107.
- QUESNIAUX, V., C. FREMOND, M. JACOBS, S. PARIDA, D. NICOLLE *et al.*, 2004 Toll-like
receptor pathways in the immune responses to mycobacteria. *Microbes Infect* 6:
946-959.
- REILING, N., C. HOLSCHER, A. FEHRENBACH, S. KROGER, C. J. KIRSCHNING *et al.*, 2002
Cutting edge: Toll-like receptor (TLR)2- and TLR4-mediated pathogen
recognition in resistance to airborne infection with *Mycobacterium tuberculosis*. *J Immunol* 169: 3480-3484.
- REINER, A. P., E. ZIV, D. L. LIND, C. M. NIEVERGELT, N. J. SCHORK *et al.*, 2005
Population structure, admixture, and aging-related phenotypes in African
American adults: the Cardiovascular Health Study. *Am J Hum Genet* 76: 463-477.
- ROSENBERG, N. A., T. BURKE, K. ELO, M. W. FELDMAN, P. J. FREIDLIN *et al.*, 2001
Empirical evaluation of genetic clustering methods using multilocus genotypes
from 20 chicken breeds. *Genetics* 159: 699-713.
- ROSENBERG, N. A., L. HUANG, E. M. JEWETT, Z. A. SZPIECH, I. JANKOVIC *et al.*, 2010
Genome-wide association studies in diverse populations. *Nat Rev Genet* 11: 356-366.
- ROSENBERG, N. A., L. M. LI, R. WARD and J. K. PRITCHARD, 2003 Informativeness of
genetic markers for inference of ancestry. *Am J Hum Genet* 73: 1402-1422.
- ROSENBERG, N. A., S. MAHAJAN, S. RAMACHANDRAN, C. ZHAO, J. K. PRITCHARD *et al.*,
2005 Clines, clusters, and the effect of study design on the inference of human
population structure. *PLoS Genet* 1: e70.

- ROSENBERG, N. A., and M. NORDBORG, 2006 A general population-genetic model for the production by population structure of spurious genotype-phenotype associations in discrete, admixed or spatially distributed populations. *Genetics* 173: 1665-1678.
- RYFFEL, B., C. FREMOND, M. JACOBS, S. PARIDA, T. BOTHA *et al.*, 2005 Innate immunity to mycobacterial infection in mice: critical role for toll-like receptors. *Tuberculosis (Edinb)* 85: 395-405.
- SCHILLER, I., B. OESCH, H. M. VORDERMEIER, M. V. PALMER, B. N. HARRIS *et al.*, 2010a Bovine Tuberculosis: A Review of Current and Emerging Diagnostic Techniques in View of their Relevance for Disease Control and Eradication. *Transboundary and Emerging Diseases* 57: 205-220.
- SCHILLER, I., H. M. VORDERMEIER, W. R. WATERS, A. KYBURZ, M. CAGIOLA *et al.*, 2010b Comparison of tuberculin activity using the interferon-gamma assay for the diagnosis of bovine tuberculosis. *Veterinary Record* 167: 322-326.
- SCHILLER, I., H. M. VORDERMEIER, W. R. WATERS, A. O. WHELAN, M. COAD *et al.*, 2010c Bovine tuberculosis: Effect of the tuberculin skin test on in vitro interferon gamma responses. *Veterinary Immunology and Immunopathology* 136: 1-11.
- SEIDEL, G. E., 2010 Brief introduction to whole-genome selection in cattle using single nucleotide polymorphisms. *Reprod Fertil Dev* 22: 138-144.
- SEITZ, J. J., S. M. SCHMUTZ, T. D. THUE and F. C. BUCHANAN, 1999 A missense mutation in the bovine MGF gene is associated with the roan phenotype in Belgian Blue and Shorthorn cattle. *Mamm Genome* 10: 710-712.
- SELVARAJ, P., K. ALAGARASU, S. SWAMINATHAN, M. HARISHANKAR and G. NARENDRAN, 2009 CD209 gene polymorphisms in South Indian HIV and HIV-TB patients. *Infect Genet Evol* 9: 256-262.
- SELVARAJ, P., M. HARISHANKAR, B. SINGH, M. S. JAWAHAR and V. V. BANUREKHA, 2010 Toll-like receptor and TIRAP gene polymorphisms in pulmonary tuberculosis patients of South India. *Tuberculosis (Edinb)* 90: 306-310.
- SEO, K., T. R. MOHANTY, T. CHOI and I. HWANG, 2007 Biology of epidermal and hair pigmentation in cattle: a mini-review. *Vet Dermatol* 18: 392-400.
- SHITAYE, J. E., W. TSEGAYE and I. PAVLIK, 2007 Bovine tuberculosis infection in animal and human populations in Ethiopia: a review. *Veterinari Medicina* 52: 317-332.
- SMITH, M. W., and S. J. O'BRIEN, 2005 Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. *Nat Rev Genet* 6: 623-632.

- STEPHENS, M., N. J. SMITH and P. DONNELLY, 2001 A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 68: 978-989.
- TAKASUGA, A., T. WATANABE, Y. MIZOGUCHI, T. HIRANO, N. IHARA *et al.*, 2007 Identification of bovine QTL for growth and carcass traits in Japanese Black cattle by replication and identical-by-descent mapping. *Mamm Genome* 18: 125-136.
- TAWAH, C. L., J. E. O. REGE and G. S. ABOAGYE, 1997 A close look at a rare African breed - the Kuri cattle of Lake Chad Basin: origin, distribution, production and adaptive characteristics. *South African Journal of Animal Science* 27: 31-40.
- TEETER, K. C., B. A. PAYSEUR, L. W. HARRIS, M. A. BAKEWELL, L. M. THIBODEAU *et al.*, 2008 Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Res* 18: 67-76.
- THOEN, C. O., P. A. LOBUE, D. A. ENARSON, J. B. KANEENE and I. N. DE KANTOR, 2009 Tuberculosis: a re-emerging disease in animals and humans. *Vet Ital* 45: 135-181.
- TROY, C. S., D. E. MACHUGH, J. F. BAILEY, D. A. MAGEE, R. T. LOFTUS *et al.*, 2001 Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410: 1088-1091.
- TSCHOPP, R., E. SCHELLING, J. HATTENDORF, D. YOUNG, A. ASEFFA *et al.*, 2010 Repeated cross-sectional skin testing for bovine tuberculosis in cattle kept in a traditional husbandry system in Ethiopia. *Veterinary Record* 167: 250-256.
- VAN TASSELL, C. P., T. P. SMITH, L. K. MATUKUMALLI, J. F. TAYLOR, R. D. SCHNABEL *et al.*, 2008 SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat Methods* 5: 247-252.
- VILLA-ANGULO, R., L. K. MATUKUMALLI, C. A. GILL, J. CHOI, C. P. VAN TASSELL *et al.*, 2009 High-resolution haplotype block structure in the cattle genome. *BMC Genet* 10: 19.
- WANG, H., X. ZENG, Z. FAN and B. LIM, 2011 RhoH modulates pre-TCR and TCR signalling by regulating LCK. *Cell Signal* 23: 249-258.
- WATERS, W. R., M. V. PALMER, T. C. THACKER, J. P. BANNANTINE, H. M. VORDERMEIER *et al.*, 2006 Early antibody responses to experimental *Mycobacterium bovis* infection of cattle. *Clin Vaccine Immunol* 13: 648-654.
- WIEHE, T., J. MOUNTAIN, P. PARHAM and M. SLATKIN, 2000 Distinguishing recombination and intragenic gene conversion by linkage disequilibrium patterns. *Genet Res* 75: 61-73.

- WINKLER, C. A., G. W. NELSON and M. W. SMITH, 2010 Admixture mapping comes of age. *Annu Rev Genomics Hum Genet* 11: 65-89.
- YOKOUCHI, K., Y. MIZOGUCHI, T. WATANABE, E. IWAMOTO, Y. SUGIMOTO *et al.*, 2009 Identification of a 3.7-Mb region for a marbling QTL on bovine chromosome 4 by identical-by-descent and association analysis. *Anim Genet* 40: 945-951.
- ZHANG, Q., C. E. LEWIS, L. E. WAGENKNECHT, R. H. MYERS, J. S. PANKOW *et al.*, 2008 Genome-wide admixture mapping for coronary artery calcification in African Americans: the NHLBI Family Heart Study. *Genet Epidemiol* 32: 264-272.
- ZHU, X., and R. S. COOPER, 2007 Admixture mapping provides evidence of association of the VNN1 gene with hypertension. *PLoS One* 2: e1244.
- ZHU, X., R. S. COOPER and R. C. ELSTON, 2004 Linkage analysis of a complex disease through use of admixed populations. *Am J Hum Genet* 74: 1136-1153.
- ZHU, X., A. LUKE, R. S. COOPER, T. QUERTERMOUS, C. HANIS *et al.*, 2005 Admixture mapping for hypertension loci with genome-scan markers. *Nat Genet* 37: 177-181.
- ZHU, X., H. TANG and N. RISCH, 2008 Admixture mapping and the role of population structure for localizing disease genes. *Adv Genet* 60: 547-569.

Appendix A: Genomewide locus specific -log P-value (Y-axis) for African *Bos taurus* and *Bos indicus* ancestry estimate of Ethiopian Zebu (ETZ) cattle using 9150 markers according to physical position (Mb) indicated on X-axis

