

**School of Biomedical Sciences**

**Characterisation of the Djallonke Sheep Breed in Ghana using Molecular Markers**

**Mohammed Yaro**

**This thesis is presented for the Degree of  
Doctor of Philosophy  
of  
Curtin University**

**August 2017**

# Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Animal Ethics (For projects involving animal use) The research presented and reported in this thesis was conducted in compliance with the National Health and Medical Research Council Australian code for the care and use of animals for scientific purposes 8th edition (2013). The proposed research study received animal ethics approval from the Curtin University Animal Ethics Committee, Approval Number # AEC\_2014\_35

Signature:  .....

Date: 30/08/2017

# Abstract

This research sought to use molecular markers to characterise the Djallonke sheep breed in Ghana. An additional aim was to analyse the Djallonke sheep genome for signatures of adaptive selection in relation to the many adaptive attributes, particularly host resistance to African animal trypanosomiasis and gastrointestinal nematode infection. Molecular characterisation of the class IIa Major Histocompatibility complex (MHC) genes in the Djallonke sheep was also undertaken because of the importance of the MHC to immunity and adaptive traits.

The Djallonke sheep is the most ancient sheep breed found in Africa, and is present in at least 14 countries. The Djallonke sheep are a popular breed due to their natural resistance to many livestock diseases. In spite of the importance of Djallonke sheep, genetic analyses of them has been very limited and the Djallonke sheep populations in Ghana have not yet been subjected to any genetic characterisation. Furthermore, recent studies have shown a concerning loss of important survival traits amongst the Djallonke populations (e.g. resistance to parasites). This loss has been attributed to the introgression of trypanosusceptible Sahelian genetics, due to a practice of indiscriminate cross breeding with the considerably larger framed Sahelian sheep.

Therefore, blood was collected from Djallonke and Sahelian sheep from two sheep breeding stations in Ghana. Genomic DNA was extracted from these samples for molecular genetic studies.

The first part of this work involved a comparative whole genome analysis for the Djallonke and Sahelian sheep breeds using a “sequence and pool” methodology. Five unrelated individuals of Djallonke and Sahelian sheep were sequenced at a combined greater than 22-fold average coverage for each breed. High levels (~ 96%) of variant similarities were found between the two breeds, a reflection of their shared demographic histories. In addition, significant numbers of breed specific single nucleotide polymorphisms (SNPs), insertions and deletions (INDELs) were identified for the Djallonke (~ 0.33 million) and Sahelian (~ 0.16 million) and will provide targets for developing efficient breed specific molecular identification markers. A total of 11.1 million and 10.9 million SNPs were identified in the Djallonke and Sahelian breeds, of which approximately 15% (~ 2million) and 16% (~1.8 million)

respectively, have not been previously reported in sheep. The Djallonke and Sahelian sheep both constitute significant and unique genetic resources, and contribute to world sheep diversity. This study represents the first ever whole genome variant analysis in the two sheep breeds, and will be of fundamental importance for the sustainable management of these two sheep breeds.

The second part of this study undertook a selection signature analysis of the whole genome variant dataset for the Djallonke and Sahelian sheep breeds. The high variant rates in Djallonke and Sahelian, coupled with the high accuracy of the sequencing process as determined by the ratios of transition and transversion mutations, means that this is an optimal dataset for selection signature analysis. Multiple regions of reduced heterozygosity (i.e. increased homozygosity) were found, 70 of these co-localised with genomic regions known to harbour genes that mediate disease resistance, immune response and adaptation in sheep or cattle. Of particular interest are the regions of reduced heterozygosity that co-localised with previously reported genes for resistance to Haemonchosis and Trypanosomiasis. These reduced heterozygosity regions are likely to be a consequence of genetic hitchhiking and are probably signatures of selection for these diseases. This study provides a valuable insight for elucidating the underlying mutations and mechanisms of the many adaptive traits in the Djallonke breed. This study presents the first support to putative regions harbouring QTL for trypanotolerance, resistance to *H. contortus* infection, and adaptation to a harsh tropical climate in these two sheep breeds. This knowledge is fundamental to the sustainable breeding and utilisation of the Djallonke sheep, and has long term implications for food security and poverty alleviation in all 14 affected countries in the region.

The third part of this study is a comparative analysis of MHC class IIa region in a population of 100 Djallonke and 100 Sahelian sheep. Variations within the class IIa MHC region in sheep have been associated with resistance to parasites. The Djallonke sheep breed is known to be more resistant to parasitic infection than the Sahelian breed. These contrasting resistance traits have been proven via many parasitological and haematological studies, but no genetic analyses of this region have previously been conducted. Therefore, the rationale for this study was to analyse, for the first time, the allelic variation within the MHC class II region in these two breeds. Sequencing-based genotyping of three polymorphic class II genes in the two sheep populations revealed high numbers (24) of novel alleles at two of the loci. High numbers of breed specific alleles were also identified for both the Djallonke (11) and Sahelian

(7) breeds. Evolutionary distance analysis showed that the Djallonke breed has the higher diversity at all three loci analysed. Higher diversity at the class II region is associated with increased resistance to parasitic infection. This analysis represents the first characterisation of the class II MHC region in these two breeds, and supports the hypothesis of the existence of molecular evidence for the contrasting phenotypes in these two sheep breeds. This information offers new opportunities such as the application of molecular tools for the sustainable management of these two important indigenous sheep breeds.

The fourth part of this study presents a comparative population genetic analysis of 100 Djallonke and 100 Sahelian sheep from Ghana. Statistical analyses for allelic frequency, heterozygosity, linkage disequilibrium, genic differentiation, and departure from Hardy Weinberg equilibrium were also undertaken. To understand the demographic factors shaping the MHC class II region in these two sheep populations, the Nei-Gojobori codon-based selection test was conducted for three class II loci. Thirteen alleles were under significant selection in the Djallonke population in comparison to eleven alleles in the Sahelian population, with only four common alleles under the same direction of selection. This study provides the first information on population genetic analysis of the MHC class II region in the Djallonke and Sahelian sheep. These results offer valuable insights to some of the demographic factors have shaped the genomic MHC architecture in these two sheep populations in this region.

The overall aim of this study: to characterise the Djallonke sheep breed of Ghana using molecular markers, was accomplished, and additionally the Sahelian sheep was similarly characterised. The genetic information obtained from these two indigenous Ghanaian and African sheep breeds with contrasting phenotypic traits facilitated a wide range of comparative analyses. In summary, substantial genetic similarities were identified between the two sheep, but also significant differences. These findings provided the first evidence of genetic basis of the contrasting phenotypes between the Djallonke and Sahelian breeds. This knowledge is fundamental to the sustainable management of these two sheep breeds and will be valuable for studies into disease resistance in sheep.

# Table of Contents

Declaration.....	ii	
Abstract.....	iii	
Table of Contents.....	vi	
List of Figures.....	xi	
List of Tables.....	xv	
List of Appendices.....	xvi	
List of Abbreviations.....	xix	
Acknowledgements.....	xxiii	
Publications and awards associated with this research work.....	xxiv	
Chapter 1	Molecular identification of livestock breeds: a tool for modern conservation biology.....	1
1.1.1	Introduction.....	1
1.1.2	Key Scientific, Cultural and Economic Importance of global livestock genetic resources.....	4
1.1.3	Threats to Global Livestock Genetic Resources.....	5
1.1.4	Assessment of Livestock Genetic Diversity and Conservation Status.....	9
1.1.4.1	Variable Number of Tandem Repeats (VNTRS).....	12
1.1.4.2	Single Nucleotide Polymorphism (SNP) Markers.....	13
1.1.5	Genomic Methods for Breed Prediction.....	14
1.1.6	Modelling: The Way Forward.....	23
1.1.7	Conclusions.....	25
1.2	Combatting African Animal Trypanosomiasis (AAT) in livestock: The potential role of trypanotolerance.....	26
1.2.1	Introduction.....	26
1.2.2	A Century of Intervention against AAT.....	28
1.2.3	Genetic Basis for trypanotolerance.....	31
1.2.4	Mechanism of trypanosome infection and trypanotolerance.....	37
1.2.5	Trypanotolerance in the context of Africa livestock production.....	41
1.2.6	Conclusions.....	42
1.3	The key role Class II Major histocompatibility complex in parasitic infection of domestic sheep.....	43
1.3.1	Introduction.....	43
1.3.2	Functions of the MHC class II genes.....	44
1.3.3	Mechanisms of balancing selection at the MHC.....	45

1.3.4	Contrasting immunological phenotype in Djallonke and Sahelian sheep .....	47
1.3.5	Conclusions .....	48
1.4	Research aims .....	48
Chapter 2	Comparative analyses of whole genomic variant characteristics .....	50
2.1	Introduction .....	50
2.2	Methods .....	52
2.2.1	Sheep .....	52
2.2.2	Blood Collection & DNA extraction .....	52
2.2.3	Library Construction and Sequencing .....	53
2.2.4	Mapping and Pre-processing of Reads .....	53
2.2.5	Variant calling pipeline .....	53
2.2.6	Annotation and functional analysis of genomic variants .....	54
2.3	Results .....	54
2.3.1	Whole genome sequencing and mapping .....	54
2.3.2	Availability of data and materials .....	55
2.3.3	Comparison of the chromosomal distribution of all identified Autosomal Variants of the Djallonke and Sahelian Genomes .....	55
2.3.4	Annotation of variants in Djallonke and Sahelian Sheep .....	57
2.3.5	Comparison of the distribution of autosomal variants by genomic regions in the Djallonke and Sahelian sheep .....	59
2.4	Discussion .....	61
2.4.1	Variant characteristics and Novelty in Djallonke and Sahelian sheep .....	61
2.4.2	The potential role of “other variant “genomic variants categories to sustainable management domestic livestock .....	64
2.5	Conclusions .....	66
Chapter 3	Analysis of whole genomes of Djallonke and Sahelian sheep reveals co-localisation of regions of reduced heterozygosity with candidate genes for disease resistance and adaptation to tropical climatic conditions .....	67
3.1	Introduction .....	67
3.2	Methods .....	69
3.2.1	Detection of Signatures of selection .....	70
3.3	Results .....	71
3.3.1	Signatures of Selection in Djallonke and Sahelian Sheep .....	71
3.3.1.1	HomSI Output legend for genomic sweeps (Adapted from Görmez <i>et al.</i> , 2014) .....	71
3.3.2	Genomic evidence for tolerance to trypanosomiasis in Djallonke and Sahelian sheep breeds .....	74
3.3.3	Genomic Evidence for resistance to Haemonchosis in Djallonke and Sahelian sheep .....	79

3.3.4	Genomic evidence for adaptation to tropical climatic conditions in the Djallonke and Sahelian sheep breeds .....	81
3.4	Discussion.....	83
3.4.1	Signatures of Selection in Djallonke and Sahelian Sheep .....	83
3.4.2	Genomic evidence for trypanotolerance in Djallonke and Sahelian sheep .....	85
3.4.3	Genomic Evidence for resistance to Haemonchosis in Djallonke and Sahelian sheep.....	87
3.4.4	Genomic evidence for adaptation to tropical climatic conditions in Djallonke and Sahelian sheep breeds.....	88
3.5	Conclusions .....	89
Chapter 4	Molecular characterisation of Ovar-MHC class II region reveals novel alleles in the Djallonke and Sahelian sheep breeds of Ghana .....	91
4.1	Introduction .....	91
4.2	Methods .....	93
4.2.1	Blood collection and white blood cell isolation .....	93
4.2.2	Purification of sheep genomic DNA from buffy coats.....	93
4.2.3	Preparation of Media, Buffers and solutions.....	94
4.2.3.1	Preparation of LB agar plates with Ampicillin, IPTG and X-GAL .....	94
4.2.4	Agarose gel electrophoresis.....	94
4.2.5	Polymerase chain reaction (PCR).....	95
4.2.6	Optimised PCR conditions for <i>DRB1</i> , <i>DQA1</i> , and <i>DQA2</i> .....	95
4.2.7	Clean up of PCR amplicons .....	96
4.2.8	Ligation and Transformation of PCR amplicons.....	96
4.2.8.1	Ligation of PCR amplicons .....	96
4.2.8.2	Transformation of Ligation products .....	96
4.2.8.3	Screening, selection and purification of recombinant plasmid containing <i>E. coli</i> colonies .....	97
4.2.9	Sequencing of DNA amplicons and Plasmids.....	97
4.2.9.1	Analyses of sequenced PCR amplicon data .....	97
4.2.10	Phylogenetic analysis .....	98
4.2.11	Genotype discovery curves.....	99
4.3	Results .....	99
4.3.1	Diversity of MHC Class IIa <i>DRB1</i> , <i>DQA1</i> and <i>DQA2</i> genes in Djallonke and Sahelian sheep of Ghana .....	99
4.3.1.1	Availability of data.....	99
4.3.2	Genotype Discovery Curve for Sampling of Djallonke and Sahelian Sheep Populations.....	100
4.3.2.1	Molecular phylogenetic re-construction for unique coding sequences at the <i>DRB1</i> , <i>DQA2</i> and <i>DQA1</i> loci.....	102

4.3.3	Classification of MHC class IIa alleles in Djallonke and Sahelian sheep using the MHC-IPD Nomenclature guidelines .....	119
4.3.3.1	Phylogenetics of the <i>DRB1</i> , <i>DQA2</i> and <i>DQAI</i> alleles in Djallonke and Sahelian sheep populations based on MHC-IPD nomenclature classification	126
4.3.3.1.1	<i>OLA-DRB1</i> coding sequence alignment .....	126
4.3.3.1.2	<i>OLA-DQA2</i> coding sequence alignment .....	127
4.3.3.1.3	<i>OLA-DQAI</i> coding sequence alignment .....	127
4.4	Discussion.....	137
4.5	Conclusions .....	144
Chapter 5	Genetic diversity of MHC class II region in Djallonke and Sahelian sheep populations .....	145
5.1	Introduction .....	145
5.2	Methods .....	146
5.2.1	Test for Signatures of Selection .....	147
5.3	Results .....	147
5.3.1	Heterozygosity, Inbreeding, Hardy Weinberg Equilibrium and Linkage disequilibrium in the Djallonke and Sahelian populations.....	155
5.3.2	Selection Signature at the MHC class II in Djallonke and Sahelian sheep populations .....	157
5.4	Discussion.....	158
5.5	Conclusions .....	163
Chapter 6	General Discussion.....	164
6.1	Molecular adaptive signatures of Selection in Djallonke and Sahelian Sheep....	165
6.2	Molecular characterisation of Ovar-MHC class II region reveals novel alleles in the Djallonke and Sahelian sheep breeds of Ghana.....	166
6.3	Genetic diversity of MHC class II region in Djallonke and Sahelian populations	168
6.4	Future directions.....	169
References.....		172
<hr/>		
APPENDICES		202
<hr/>		
Appendix A	Images of sampling regions and animals .....	202
Appendix B	Protocols for preparing media, buffers and solutions.....	204
B.1	NanoDrop DNA analysis.....	206
B.2	ExoSAP PCR clean up protocol .....	206
B.3	Favorgen PCR clean up Protocol.....	206
Appendix C	Additional results for chapter 3 .....	208
C.1	HomSI images of genomic regions associated with resistance to <i>H. contortus</i> and other GIT parasites in Djallonke and Sahelian sheep breeds .....	208

C.1.1	Contrasting signatures of selection in genetic regions associated with resistance to Haemonchosis and other GIT parasite the Djallonke (high ROH) and Sahelian (low ROH) sheep breeds .....	208
C.1.2	Similar signatures of selection in genetic regions associated with resistance to Haemonchosis and other GIT parasite the Djallonke (high ROH) and Sahelian (high ROH) sheep breeds .....	210
C.2	HomSI analyses of genomic regions associated with adaptive selection in Djallonke and Sahelian sheep breeds .....	211
C.2.1	Contrasting signatures of adaptive selection in genomic regions associated with the Djallonke (high ROH) and Sahelian (low ROH) sheep breeds .....	211
C.2.2	Similar signatures of selection in genomic regions associated with adaptation in both the Djallonke (high ROH) and Sahelian (high ROH) sheep breeds .....	214
Appendix D	Additional results for Chapter 4 .....	216
D.1	Maximum Likelihood Neighbour Joining Tree (ML-NJT) Model for Djallonke the sheep .....	216
Appendix E	GenBank accessions for MHC class II genes generated in this studies ....	222
Appendix F	Co-Authors consent for thesis associated publications .....	271

# List of Figures

Figure 1.1 Distribution of three economically important trypanotolerant livestock breeds within the African animal trypanosomiasis-tsetse fly endemic region of Africa.....	30
Figure 2.1 Comparison of the numbers and proportion of SNPs , insertions and deletions for Djallonke and Sahelian sheep with respect to Oar v3.1.....	58
Figure 2.2 Comparison of variants distribution in the genomes of Djallonke and Sahelian sheep breeds.....	60
Figure 2.3 Similar proportions of variant distribution in exonic & other genic categories in the genomes of Djallonke and Sahelian sheep.....	61
Figure 3.1 Distribution of Djallonke and Sahelian sheep in relation to the Trypanosomiasis-Tsetse fly endemic zone of Africa .....	68
Figure 3.2 Prominent but similar run of homozygosity (see legend 3.3.1.1) at non QTL Genic regions on chromosome 13 (co-localised with <i>BTBD3</i> gene) and 20 (co-localised with <i>CDKALI</i> gene) in both sheep breeds.....	72
Figure 3.3 An example of non-QTL intergenic regions showing high homozygosity (>80 kb) at chromosomes 1 and 4 only in the Djallonke sheep in striking contrast to the Sahelian sheep breeds .....	73
Figure 3.4 Non QTL genic regions of contrasting homozygosity between Djallonke(blue) and Sahelian (white region represents stretch homozygous for the opposite allele) sheep on Chromosome 20 denoting breed specific alleles .....	73
Figure 3.5 Comparison of the sweep characteristics within four previously identified trypanotolerance-associated genes in Djallonke and Sahelian sheep breeds ..	75
Figure 3.6 Differential Sweep Characteristics at the trypanotolerant associated <i>ARHGAP15</i> gene and <i>STX7</i> gene of Djallonke (showing more extended homozygosity) and Sahelian (less homozygosity) sheep breeds.....	75
Figure 3.7 Two trypanotolerance associated QTL regions showing higher homozygosity for the <i>CTSS</i> and <i>CD19</i> genes for Djallonke in comparison to the Sahelian sheep ..	76
Figure 3.8 Comparison of the sweep characteristics at three genes previously shown to be associated with resistance to <i>Haemonchus contortus</i> higher homozygosity (blue) in the Djallonke sheep.....	76
Figure 3.9 Comparison of the sweep characteristics at two genes previously shown to be associated with resistance to <i>Haemonchus contortus</i> showing higher (blue) in the Djallonke sheep.....	77
Figure 3.10 Comparison of the sweep characteristics in Djallonke and Sahelian sheep at three known QTL associated with adaptive selection.....	78
Figure 3.11 HomSI analysis adjacent to <i>ALDH1A3</i> , <i>PLCB1</i> , <i>BATF2</i> & <i>ELF2</i> genes showing contrasting sweeps in Djallonke (higher homozygosity) and Sahelian .....	78
Figure 4.1 Genotype discovery curve for <i>DRB1</i> , <i>DQA2</i> and <i>DQA1</i> alleles for 100 randomly sampled individual Sahelian sheep .....	101
Figure 4.2 Genotype discovery curve for <i>DRB1</i> , <i>DQA2</i> and <i>DQA1</i> alleles for 100 randomly sampled Djallonke sheep .....	102

- Figure 4.3 Best model Maximum likelihood phylogenetic tree for combined functional and unique *DRBI* exon 2 coding sequences in the two sheep populations. Tree was constructed to scale with 85 amino acid sequences. The tree shows the topography of evolutionary relationship of all unique sequences. The numbers at the nodes represent bootstrap confidence based on 500 replications..... 107
- Figure 4.4 Best model Maximum likelihood phylogenetic tree for functional and unique *DRBI* exon 2 coding sequences in the Djallonke sheep population. Tree was constructed to scale with 58 amino acid long sequences. The IPD name was used for unique sequences that matched previous publications. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications..... 108
- Figure 4.5 Best model Maximum likelihood phylogenetic tree for functional and unique *DRBI* exon 2 coding sequences in the Sahelian sheep population. Tree was constructed to scale with 40 amino acid long sequences. The IPD name was used for unique sequences that matched previous publications. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications ..... 109
- Figure 4.6 Best model Maximum likelihood phylogenetic tree for combined functional and unique *DQA2* exon 2 coding sequences in the two sheep populations. Tree was constructed to scale with 41 amino acid long sequences. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications..... 111
- Figure 4.7 Best model Maximum likelihood phylogenetic tree for functional and unique *DQA2* exon 2 coding sequences in the Djallonke sheep population. Tree was constructed to scale with 38 amino acid long sequences. The IPD name was used for unique sequences that matched previous publications. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications..... 112
- Figure 4.8 Best model Maximum likelihood phylogenetic tree for functional and unique *DQA2* exon 2 coding sequences in the Sahelian sheep population. The IPD name was used for unique sequences that matched previous publications. Tree was constructed to scale with 34 amino acid long sequences. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications..... 113
- Figure 4.9 Best model Maximum likelihood phylogenetic tree for combined functional and unique *DQAI* exon 2 coding sequences in the two sheep populations. Tree was constructed to scale with 55 amino acid long sequences. The IPD name was used for unique sequences that matched previous publications. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications..... 116
- Figure 4.10 Best model Maximum likelihood phylogenetic tree for functional and unique *DQAI* exon 2 coding sequences in the Djallonke sheep population. Tree was constructed to scale with 42 amino acid long sequences. The GenBank name was used for unique sequences that matched previous publications. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications..... 117

- Figure 4.11 Best model Maximum likelihood phylogenetic tree for functional and unique *DQA1* exon 2 coding sequences in the Sahelian sheep population. Tree was constructed to scale with 22 amino acid long sequences. The GenBank name was used for unique sequences that matched previous publications. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications..... 118
- Figure 4.12 Phylogenetic Tree of 26 *DRB1* Alleles (A-Z) in Djallonke and Sahelian sheep using a codon\_start=1 translation frame for the *DRB1\*0401\_1* allele. The numbers at the nodes of each cluster denote the percentage support for that cluster in the constructed evolutionary relationship per 100 bootstrap replicates. The codon\_start=1 translated allele group are highlighted in yellow showing very significant deviation (100% bootstrap support) from all other *DRB1* alleles and the *BoLA-DRB3* (outgroup) is highlighted in orange..... 129
- Figure 4.13 Poisson Neighbour joining tree of *DRB1* exon 2 CDS (A to Z) in Djallonke and Sahelian sheep using a codon\_start=1 translation frame for the *DRB1\*0401* allele (highlighted in yellow). The cattle *BoLA-DRB3* (NCBI accession BTU78548.1) as an outgroup (orange), clusters with the *Ovar-DRB1* alleles. The scaled evolutionary distance tree revealed a very prominent deviation of the \*0401 group from remaining *Ovar-DRB1* alleles..... 130
- Figure 4.14 Reconstructed phylogenetic Tree of the 26 *DRB1* Alleles (A to Z) in Djallonke and Sahelian sheep showing the codon\_start=3 translation of *DRB1\*0401* alleles (prefixed with letters N, O, P, & Q). The numbers at the nodes of each cluster denote the percentage support for that cluster per 100 bootstrap replicates. The *Ovar-DRB1\*0901* is shown as the most distinctive allele. The outgroup *BoLA-DRB3* is prefixed with the letter U..... 131
- Figure 4.15 Poisson Neighbour joining tree of *DRB1* exon 2 CDS (A to Z) in Djallonke and Sahelian sheep using a codon\_start=3 translation frame for the *DRB1\*0401* allele (highlighted in yellow). The cattle *BoLA-DRB3* (NCBI accession BTU78548.1) as an outgroup (orange), clusters with the *Ovar-DRB1* alleles. The scaled evolutionary distance tree revealed that the \*0401 group now clusters with all other *Ovar-DRB1* alleles. The *ovar-DRB1\*0901* group (highlighted in yellow and orange) is now the most distinct allele in the tree..... 132
- Figure 4.16 Poisson Phylogenetic Tree of the 13 *DQA2* Alleles (A to M) in Djallonke and Sahelian sheep with respective GenBank references. The numbers at the nodes of each cluster denote the percentage support for that cluster in the re-constructed evolutionary relationship per 100 bootstrapped replicates. The *OLA-DQA2\*08011* (C) and *OLA-DQA2\*08012* (M) alleles cluster on the same node and two other alleles (H & I) also cluster on the same node. Bootstrap support is variable (23 to 100%), but is predominantly very high. .... 134
- Figure 4.17 Neighbour Joining Phylogenetic Tree of *DQA1* Alleles (A to M) in Djallonke and Sahelian sheep with respective GenBank references. The numbers at the nodes of each cluster denote the percentage support for that cluster in the re-constructed evolutionary relationship per 100 bootstrapped replicates. Two alleles (K&L) cluster on the same note with very high (100%) bootstrap support, and two other alleles (M & C) cluster with moderately high (66%) bootstrap support..... 136
- Figure 5.1 Comparison of the *Ovar-DRB1* Allele frequency in Djallonke and Sahelian sheep populations..... 151
- Figure 5.2 Comparison of the *OLA-DQA2* Allele frequency in Djallonke and Sahelian sheep populations..... 153

Figure 5.3 Comparison of the *DQA1* Allele frequency in Djallonke and Sahelian sheep populations..... 155

# List of Tables

Table 1.1 Present status of efforts to identify livestock breeds using genomic methods.....	17
Table 1.2 Present status of efforts to identify quantitative trait loci (QTL) for trypanotolerance in African cattle breeds and experimental mice.....	35
Table 2.1 Summary of sequencing metrics for Djallonke and Sahelian sheep data generated on an Ion Proton Sequencer (ThermoFisher).....	55
Table 2.2 Comparison of the identified Autosomal variants' per chromosome of the Djallonke and the Sahelian Sheep Genomes .....	56
Table 2.3 Autosomal SNP Transitions and transversion ratio in Djallonke and Sahelian Dataset .....	56
Table 2.4 Pairwise comparison of Genomic SNPs and INDELS computed for Sahelian and Djallonke sheep.....	57
Table 2.5 Total Autosomal variants identified showing the novel variants components in Djallonke and Sahelian sheep .....	58
Table 2.6 Comparisons of Autosomal variants by type and functional categories in Djallonke and Sahelian Genomes.....	59
Table 2.7 Comparison of the ratio of total autosomal SNPs to three consequential genomic variants categories identified in this study and other livestock species.....	64
Table 3.1 Trypanotolerance candidate genes co-localised with signatures of selection in Djallonke and Sahelian sheep .....	79
Table 3.2 QTL for Resistance to Haemonchosis and other GIT parasites co-localised with signatures of selection in Djallonke and Sahelian sheep .....	80
Table 3.3 Candidate genes for adaptive selection co-localised with regions of high homozygosity in Djallonke and Sahelian sheep .....	82
Table 4.1 The Oligonucleotide(Oligo) primers used for genotyping the exon2 DRB1, DQA1 and DQA2 loci .....	95
Table 4.2 Unique DRB1 exon 2 amino acid sequences identified in Djallonke and Sahelian sheep populations (Presumed functional coding sequences (n= 83) with non functional coding sequences containing premature stop codons highlighted in yellow (n =2); Prefix: DJ and SA denote Djallonke and Sahelian samples respectively).....	104
Table 4.3 Unique DQA2 exon 2 amino acid sequences identified in Djallonke and Sahelian sheep populations (n= 41).....	110
Table 4.4 Unique DQA1 exon 2 amino acid sequences identified in Djallonke and Sahelian sheep populations (Presumed functional coding sequences (n= 52) with non functional coding sequences containing premature stop codons highlighted in yellow (n =3)) .....	114
Table 4.5 Twenty six DRB1 allele groups in Djallonke and Sahelian sheep populations based on IPD-MHC nomenclature.....	120
Table 4.6 DQA2 allele groups in Djallonke and Sahelian sheep populations based on IPD-MHC nomenclature.....	122

Table 4.7 DQA1 allele groups in Djallonke and Sahelian sheep populations based on IPD-MHC nomenclature.....	123
Table 4.8 Muscle alignment of the 26 DRB1 Alleles (A to Z) in Djallonke and Sahelian sheep using codon_start 1 translation frame for DRB1*0401 showing misalignment with the remaining DRB1 alleles.....	124
Table 4.9 Muscle alignment of the 26 DRB1 Alleles using one member per group with (A to Z) in the two sheep populations using codon_start 3 translated DRB1*0401 showing alignment with all the remaining DRB1 alleles.....	125
Table 4.10 Muscle alignment of the 13 DQA2 Alleles (A to M) in Djallonke and Sahelian sheep populations with respective GenBank references.....	133
Table 4.11 Muscle alignment of the 13 DQA1 Alleles (A to M) in Djallonke and Sahelian sheep populations with respective GenBank references.....	135
Table 4.12 Comparison of the known MHC Class IIa alleles in the Djallonke and Sahelian with reports from other sheep breeds.....	141
Table 5.1 Maximum likelihood estimates of allele frequencies of DRB1 in 100 Djallonke and 100 Sahelian sheep from Ghana.....	149
Table 5.2 Maximum likelihood Estimates of Allele frequencies of DQA2 in 100 Djallonke and 100 Sahelian sheep from Ghana.....	152
Table 5.3 Maximum likelihood Estimates of Allele frequencies of DQA1 in 100 Djallonke and 100 Sahelian sheep from Ghana.....	154
Table 5.4 The Levene's correction estimates for number of observed and expected heterozygotes in the Djallonke and Sahelian populations.....	156
Table 5.5 Weir & Cockerham estimates of population average coefficient of Inbreeding (Fis) at all loci for Djallonke and Sahelian sheep.....	156
Table 5.6 Class II MHC alleles under selection in Djallonke and Sahelian populations....	158

## List of Appendices

Appendix Figure 1 Sampling location for the Djallonke (Ejura) and Sahelian(Pong Tamale) sheep cohorts.....	202
Appendix Figure 2 Adult female (left) and male (right) Djallonke Sheep breed ( <a href="http://www.fao.org/wairdocs/ilri/x5468e/x5468e04.htm">http://www.fao.org/wairdocs/ilri/x5468e/x5468e04.htm</a> accessed on 21.10.2014).....	203
Appendix Figure 3 Adult Male (left) and female (right) Sahelian sheep breed ( <a href="http://www.ansi.okstate.edu/breeds/sheep/saheltype">http://www.ansi.okstate.edu/breeds/sheep/saheltype</a> ).....	203
Appendix Figure 4 HomSI analysis adjacent to MHC II-DRB1, CHI3L2, LRP8 and DENND2D genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH).....	208
Appendix Figure 5 HomSI analysis adjacent to LAMC1, RELN and FCER2 genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH).....	209
Appendix Figure 6. HomSI analysis adjacent to NSUN2, CXCL12, ABCB9, SOX9 and AREG genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH).....	209

Appendix Figure 7 HomSI region adjacent to PAK4, CXCR6, HRH1, UBE2N and KIT genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH) .....	210
Appendix Figure 8 HomSI analysis adjacent to PIK3CD, IL17RB & MUC15 genes showing similar sweeps in Djallonke (high ROH) and Sahelian (high ROH). .....	210
Appendix Figure 9 HomSI analysis adjacent to ALCAM, GNAI3 & IL7 genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH).....	211
Appendix Figure 10 HomSI analysis adjacent to FGF5, TSAPAN12, HSPA1A & KRIT1 genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH). .....	211
Appendix Figure 11 HomSI analysis adjacent to IL12RB2, LMLN, ABHD2 & PGRMC2 genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH). .....	212
Appendix Figure 12 HomSI analysis adjacent to GLB1, OR2AG1 PRLR & PCDH9 genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH). .....	212
Appendix Figure 13 HomSI analysis adjacent to ALDH1A3, PLCB1, BATF2 & ELF2 genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH). .....	213
Appendix Figure 14 HomSI analysis adjacent to SUCLG2, OLR1, SYNJ1 & OR2AG1 genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH). .....	213
Appendix Figure 15 HomSI analysis adjacent to MITF, NFATC2 & IL1R1 genes showing similar sweeps in Djallonke (high ROH) and Sahelian (low ROH). .....	214
Appendix Figure 16 HomSI analysis adjacent to ALK, NPR2, IFITM3, PDK2, ATP12A & BMP4 genes showing similar sweeps in Djallonke (high ROH) and Sahelian (low ROH). .....	215
Appendix Figure 17 HomSI analysis adjacent to NF1, ABCG2, FOXG1 & APC2 genes showing similar sweeps in Djallonke (high ROH) and Sahelian (low ROH). .....	215
Appendix Figure 18 Re-constructed Neighbour Joining Tree (NJT) Model for the functional and unique DRB1 exon 2 coding sequences in the Djallonke sheep population showing conserved clustering characteristic as the ML JTT tree Model (chapter 4 Figure 4.2).....	216
Appendix Figure 19 Re-constructed Neighbour Joining Tree (NJT) Model for the functional and unique DRB1 exon 2 coding sequences in the Sahelian sheep population showing conserved clustering characteristic as the ML JTT tree Model (chapter 4 Figure 4.3).....	217
Appendix Figure 20 Re-constructed Neighbour Joining Tree (NJT) Model for the functional and unique DQA2 exon 2 coding sequences in the Djallonke sheep population showing conserved clustering characteristic as the ML JTT tree Model (chapter 4 Figure 4.5).....	218
Appendix Figure 21 Re-constructed Neighbour Joining Tree (NJT) Model for the functional and unique DQA2 exon 2 coding sequences in the Sahelian sheep population showing conserved clustering characteristic as the ML JTT tree Model (chapter 4 Figure 4.6).....	219

Appendix Figure 22 Re-constructed Neighbour Joining Tree (NJT) Model for the functional and unique DQA1 exon 2 coding sequences in the Djallonke sheep population showing conserved clustering characteristic as the ML JJT tree Model (chapter 4 Figure 4.8).....	220
Appendix Figure 23 Re-constructed Neighbour Joining Tree (NJT) Model for the functional and unique DQA2 exon 2 coding sequences in the Sahelian sheep population showing conserved clustering characteristic as the ML JJT tree Model (chapter 4 Figure 4.9).....	221
Appendix Table B.1 Recipes and protocols for reagents used for ligation and transformation	204
Appendix Table B.2 Optimised PCR master mix for DRB1, DQA1 and DQA2 for HS MyTaq polymerase 50 µl PCR reaction .....	205
Appendix Table B.3 Optimised DRB1 PCR cycling conditions for MyTaq HS .....	205
Appendix Table B.4 Template requirement for MacroGen Incorporation standard sequencing	207

## List of Abbreviations

<b><i>ABCB9</i></b>	ATP binding cassette subfamily B member 9
<b><i>ABCG2</i></b>	ATP binding cassette subfamily G member 2
<b><i>ABHD2</i></b>	Abhydrolase domain containing 2
<b><i>ALDH1A3</i></b>	Aldehyde dehydrogenase 1 family member A3
<b><i>ALCAM</i></b>	Activated leukocyte cell adhesion molecule
<b><i>ALK</i></b>	Anaplastic lymphoma receptor tyrosine kinase
<b><i>APC</i></b>	APC, WNT signaling pathway regulator
<b><i>ATP12A</i></b>	ATPase H <sup>+</sup> /K <sup>+</sup> transporting non-gastric alpha2 subunit
<b><i>ATP2B1</i></b>	ATPase plasma membrane Ca <sup>2+</sup> transporting 1
<b><i>AREG</i></b>	Amphiregulin
<b><i>ARHGAP15</i></b>	Rho GTPase activating protein 15
<b><i>BATF2</i></b>	Basic leucine zipper ATF-like transcription factor 2
<b><i>BMP2</i></b>	Bone morphogenetic protein 2;
<b><i>CD19</i></b>	CD19 molecule
<b><i>CHI3L2</i></b>	Chitinase 3 like 2
<b><i>CHIA</i></b>	Chitinase, acidic
<b><i>CTSS</i></b>	Cathepsin S
<b><i>CXCL2</i></b>	C-X-C motif chemokine ligand 2
<b><i>CXCR6</i></b>	C-X-C motif chemokine receptor 6
<b><i>DENND2D</i></b>	DENN domain containing 2D
<b><i>ELF2</i></b>	E74 like ETS transcription factor 2
<b><i>FecB</i></b>	Fecundity, Booroola
<b><i>FCER2</i></b>	Fc fragment of IgE receptor II
<b><i>FGF5</i></b>	Fibroblast growth factor 5
<b><i>FOXP1</i></b>	Forkhead box G1
<b><i>GLB1</i></b>	Galactosidase beta 1
<b><i>GNAI3</i></b>	G protein subunit alpha i3
<b><i>HRH1</i></b>	Histamine receptor H1
<b><i>HSPA1A</i></b>	Heat shock protein family A
<b><i>IFNG</i></b>	Interferon gamma

<b><i>IL7</i></b>	Interleukin 7
<b><i>IL1RI</i></b>	Interleukin 1 receptor type 1
<b><i>IL12RB2</i></b>	Interleukin 12 receptor subunit beta 2
<b><i>IL17RB</i></b>	Interleukin 17 receptor B
<b><i>IL20RA</i></b>	Interleukin 20 receptor subunit alpha
<b><i>IFTM10</i></b>	Interferon induced transmembrane protein 10
<b><i>KIT</i></b>	KIT proto-oncogene receptor tyrosine kinase
<b><i>KRIT1</i></b>	KRIT1, ankyrin repeat containing
<b><i>LAMC1</i></b>	Laminin subunit gamma 1
<b><i>LMLN</i></b>	Leishmanolysin like peptidase
<b><i>LRP8</i></b>	LDL receptor related protein 8
<b><i>INHBA</i></b>	Inhibin beta A chain
<b><i>MC1R</i></b>	melanocortin 1 receptor
<b><i>MHC-DRB1</i></b>	Major histocompatibility complex, class II, DR beta 1
<b><i>MITF</i></b>	Melanogenesis associated transcription factor
<b><i>MSRB3</i></b>	Methionine sulfoxide reductase B3
<b><i>MUC15</i></b>	Mucin 15 cell surface associated
<b><i>NF1</i></b>	Neurofibromin 1
<b><i>NFATC2</i></b>	Nuclear factor of activated T-cells 2
<b><i>NPR2</i></b>	Natriuretic peptide receptor 2
<b><i>NSUN2</i></b>	NOP2/Sun RNA methyltransferase family member 2
<b><i>OLR1</i></b>	Oxidized low density lipoprotein receptor 1
<b><i>OR2AG1</i></b>	Olfactory receptor family 2 subfamily AG member 1
<b><i>PAK4</i></b>	P21 (RAC1) activated kinase 4
<b><i>PCDH9</i></b>	Protocadherin 9
<b><i>PDK2</i></b>	Pyruvate dehydrogenase kinase 2
<b><i>PGRMC2</i></b>	Progesterone receptor membrane component 2
<b><i>PIK3CD</i></b>	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta
<b><i>PLCB1</i></b>	Phospholipase C beta 1
<b><i>PPP1R12A</i></b>	Protein phosphatase 1 regulatory subunit 12A
<b><i>PRLR</i></b>	Prolactin receptor
<b><i>SCAMP1</i></b>	Secretory carrier membrane protein 1
<b><i>SOX9</i></b>	SRY-box 9

<b><i>STX7</i></b>	Syntaxin 7
<b><i>SUCLG2</i></b>	Succinate-CoA ligase GDP-forming beta subunit
<b><i>SUGT1</i></b>	SGT1 homolog
<b><i>SYNJI</i></b>	Synaptojanin 1
<b><i>TICAM1</i></b>	Toll like receptor adaptor molecule 1
<b><i>TRHDE</i></b>	Thyrotropin releasing hormone degrading enzyme
<b><i>TSPAN12:</i></b>	Tetraspanin 12
<b><i>RAB35</i></b>	Member RAS oncogene family
<b><i>RELN</i></b>	Reelin; AREG: amphiregulin
<b><i>UBE2N</i></b>	Ubiquitin conjugating enzyme E2 N
<b>AAT</b>	African animal trypanosomiasis
<b>ADP</b>	Asian domestic pigs
<b>AFLP</b>	Amplified fragment length polymorphism
<b>AnGR</b>	Animal or livestock genetic resources
<b>AWB</b>	Asian Wild Boer
<b>BWA</b>	Burrows-Wheeler Aligner
<b>CDS</b>	Coding sequence
<b>DAD-IS</b>	Domestic Animal Diversity Information System
<b>DAGRIS</b>	Domestic Animal Genetic Resource Information System
<b>DAVID</b>	Database for annotation visualization and integrated discovery
<b>dbSNP</b>	Database single Nucleotide polymorphism
<b>DJ</b>	Djallonke
<b><i>E. coli</i></b>	<i>Escherichia coli</i>
<b>EDTA</b>	Ethylenediaminetetraacetic acid
<b>EFABIS</b>	European farm animal biodiversity information system
<b>FAO</b>	Food and Agriculture Organisation
<b>GATK</b>	Genome analysis tool kit
<b>GBS</b>	Genotyping by sequencing
<b>GLOBALDIV</b>	Global View of Livestock Biodiversity and Conservation
<b>HAN</b>	Hanwoo
<b>HomSI</b>	Homozygosity Stretch Identifier
<b>IAEA</b>	International Atomic Energy Agency
<b>IGV</b>	Integrative genomics viewer

<b>ILRI</b>	International Livestock Research Institute
<b>Indels</b>	Insertions and deletions
<b>ISAG</b>	International society for Animal genetics
<b>Kbp</b>	Kilo base pair
<b>LB</b>	Luria Bertani
<b>LD</b>	Linkage disequilibrium
<b>Mbp</b>	Mega base pair
<b>MHC</b>	Major Histocompatibility complex
<b>MtDNA</b>	Mitochondria deoxyribonucleic acid
<b>NCBI</b>	National center for biotechnology information
<b>NGS</b>	Next generation sequencing
<b>QTL</b>	Quantitative trait loci
<b>PCA</b>	Principal component analysis
<b>Rpm</b>	Revolution per minute
<b>RAPD</b>	Random Polymorphic DNA
<b>RFLP</b>	Restriction fragment length polymorphism
<b>SA</b>	Sahelian
<b>SAVE foundation</b>	Safeguard for Agricultural Varieties in Europe Foundation
<b>SNPs</b>	Single nucleotide polymorphisms
<b>SOC</b>	Super optimal broth with catabolite repression
<b>SSA</b>	Sub Saharan Africa
<b>Ts</b>	Transition
<b>Tv</b>	Transversion
<b>UCSC</b>	University of California Santa Cruz
<b>UN</b>	United Nations
<b>VTNR</b>	Variable number tandem repeat
<b>WGS</b>	Whole genome sequencing
<b>X-GAL</b>	5-bromo-4-chloro-indolyl-B-D-galactopyranoside
<b>YAN</b>	Yanbian

# Acknowledgements

As I reflect on my educational journey to this point, I realised the invaluable support that I have been lucky to have received from countless number of people. It is practically impossible to list all their names here, but I will take this opportunity to sincerely thank David M. Groth, Kylie A. Munyard and Eleanor Morgan for their patience, support and guidance throughout my studentship. I am not completely sure how you did it, but your collective efforts proved perfectly synergistic for me. I will carry this experience for the rest of my life. My appreciation also goes to Mike J. Stear for his “vote of confidence” since 2010. Special thanks also to Curtin University (CIPRS scholarship) for funding this research. Also, to my laboratory colleagues Miheer, Himel, Imran, Gae, Cory, Andrew Hosie, Audrey, Joanne Gardner, and Stephanie Coulson, I say thank you. Thank you to Abdul Wahed, Jamiu and Sam Siem for your special friendship. My appreciation also goes to Dr Ibn Rahman, the farm managers and staff members of the Ejura and Pong Tamale sheep breeding stations of Ghana for their overwhelming support during the blood sampling from sheep in 2015. Finally, to my dear wife Bilqiss for your support and understanding throughout this program.

“The sun will rise tomorrow. It always does, and all the wishing in the world for the way things were, or for what they could have been, won't change that. It won't change how things are.” (Elizabeth Scott)

# Publications and awards associated with this research work

## AWARDS

1. International Society for Animal Genetics Travel Grant Award **2017, Dublin, Ireland**

Topic presented: Molecular characterisation of Ovar-MHC class II region reveals novel alleles in the Djallonke and Sahelian sheep breeds of Ghana

2. Best Poster Presenter Award 2015, Mark Liveris' Health Sciences Research Seminar

Topic presented: Combating African Animal trypanosomiasis: The potential role of trypanotolerance

3. Best Poster Award 2015, Mark Liveris' Health Sciences Research Seminar, **2015**

Topic presented: Combating African Animal trypanosomiasis: The potential role of trypanotolerance

## Refereed Journal Publications

1. **Yaro, M**, K.A. Munyard, E. Morgan, R.J.N. Allcock, M.J. Stear, D.M. Groth

(**under review BMC Genomics 2017**)

Whole genome analysis of pooled sequences from Djallonke and Sahelian Sheep of Ghana reveals co-localisation of regions of reduced heterozygosity with candidate genes for disease resistance and adaptation to a tropical environment. **Impact factor 4.38**

2. **Yaro, M.**, Munyard, K.A., Stear, M.J., Groth, D.M., (2017)

Molecular identification of livestock breeds: a tool for modern conservation biology. *Biological Reviews, Camb. Philos. Soc.* 2017, 92(2):993-1010. **Impact factor 11.615**

3. **Yaro, M.**, Munyard, K.A., Stear, M.J., & Groth, D.M. (2016). Combating African Animal trypanosomiasis: The potential role of trypanotolerance. *Veterinary Parasitology* 225 (2016) 43–52. **Impact factor 2.69**
  
4. **Yaro, M.**, Munyard, K. A., Morgan, E., Allcock, R. J., Stear, M. J., & Groth, D. M. (2016). P4041 Pooled whole-genome sequencing reveals molecular signatures of natural adaptive selection in Djallonke sheep of Ghana. *Journal of Animal Science*, 94(7supplement4), 98-99. <http://dx.doi.org/10.2134/jas2016.947supplement498a>  
**Impact factor 2.014**

# **Chapter 1      Molecular identification of livestock breeds: a tool for modern conservation biology**

*Some of the work and ideas in this review have already been published, i.e. Yaro et al. (2017) Molecular identification of livestock breeds: a tool for modern conservation biology. Biological Reviews, Cambridge Philosophical Society, 92(2), 993-1010. <http://dx.doi.org/10.1111/brv.12265> and Yaro et al. (2016) Combatting African Animal trypanosomiasis: the potential role of trypanotolerance, Veterinary Parasitology, 225, 43-52. <http://dx.doi.org/http://dx.doi.org/10.1016/j.vetpar.2016.05.003>*

*The first part of chapter 1 reviews the genesis and current status of global domestic livestock genetic diversity, and highlights its crucial importance to human development. In particular, the exceeding relevance of between breed diversity over inter species diversity, with respect to conservation biology, is discussed. A strong argument was made for the need for molecular characterisation of livestock breeds to facilitate their reliable identification. The second section of chapter 1 chronicles the perpetual challenges faced by the livestock production industry due to African Animal Trypanosomiasis (AAT). The potential role of exploiting the natural resistance of some indigenous livestock, typically, the Djallonke sheep, as an integral part of the multifaceted solution for AAT in the context of African animal production systems was strongly supported. The third and final section of chapter 1 reviews the key role of the MHC in the host's resistance to parasitic diseases with a specific focus on the domestic sheep. An analysis of published literature suggests that MHC class II genes in sheep play a major in parasitic infections.*

## **1.1.1      Introduction**

A large amount of the genetic variation present in wild animal lineages prior to domestication has been conserved during the domestication process, and persists within the respective domesticates (Dobney & Larson, 2006). Currently, most of these wild lineages are either extinct or critically endangered (Taberlet et al., 2011). Over the 12,000 years since farm animals were first domesticated, their genetic make-up has undergone subtle adaptation due to both natural (speciation) and artificial (domestication/breeding) selection pressures exerted by their specific environments and human activity, respectively (Banik, Pankaj, & Naskar, 2015; Hoffmann & Scherf, 2005; Jensen, 2006; Mignon-Grasteau et al., 2005; Morris, 2006; Naskar, Gowane, & Chopra, 2015; Price, 1999; Vigne, 2011; Zeder et al., 2006). These selection pressures have culminated in the development of a rich global domestic animal diversity with

thousands of breeds (Ajmone-Marsan & The Globaldiv Consortium, 2010; Groeneveld et al., 2010). Each of these breeds is characterised by their unique morphology and productivity related to specific environmental and applied farming conditions (Lopes et al., 2015; Shand, 1997). A livestock breed can be generally defined as either a homogenous group with unique and identifiable phenotypic features that distinguish it from other subgroups within the same species, or a homogenous group for which geographical isolation from other groups of the same species has resulted in their acceptance as unique entities (FAO, 2000; Rege, 2003). Recently, a more refined definition of a breed concept which encompasses the history of the livestock was proposed by Felius, Theunissen, and Lenstra (2014) and Tixier-Boichard (2014). The scope of this new definition conforms to current practical reality, as not all breeds by definition actually represent unique genetic resources. Breeds can therefore be regarded as the unit of management for livestock instead of the unit of conservation so as to make it a more useful instrument for conservation purposes (Felius et al., 2014; Groeneveld et al., 2010).

A recent report on livestock breed diversity stated that there were 7,202 local breeds (breeds found in only one country), 509 regional trans-boundary breeds (breeds found in different countries within one region) and 551 international trans-boundary breeds (breeds found in different countries in different continents) (FAO, 2013). These breed classifications cover the seven main mammalian livestock species (sheep, goats, cattle, pigs, buffalo, horses, and asses/donkeys), four main avian livestock species (chicken, turkeys, ducks, and geese) and eight minor livestock species (alpacas, yaks, llamas, camels, elephants, musk oxen, and guinea pigs). However, since the concept of selective breeding only emerged in the last 200 years, and subsequently through more intensive selection in the last few decades, domestic animal diversity has been under sustained threat of significant erosion (Ajmone-Marsan & The Globaldiv Consortium, 2010; Köhler-Rollefson, 1997). In 2012, an analysis of data from 182 countries by the Global Databank for farm animal genetic resources revealed that approximately 8% of all farm animal (local, regional trans-boundary and international trans-boundary) breeds could already be considered extinct, 22% were at varying degrees of extinction risk, and the risk status of 34% was unknown (FAO, 2013). The report, which was an update of the previous 2010 edition, brings the total farm animal extinctions to a staggering 12% since 1999. This is quite significant given the fact that a total of only 16% extinctions was recorded in the preceding century (1900–1999) (FAO, 2013). The report also indicated that only approximately 36% of global farm animal genetic resources were not at any risk of immediate extinction.

This growing threat to the world's animal genetic resources was recognised by the Food and Agriculture Organisation (FAO) of the United Nations (UN) as an emerging global challenge, and this recognition has led to the ratification by 109 countries, in 2007, of the Interlaken Declaration on world animal genetic resources (Rischkowsky, Pilling, & Commission on Genetic Resources for Food Agriculture, 2007). The Interlaken Declaration was the first global action plan specifically aimed at conserving our current animal genetic resources. The declaration called for urgent and prompt measures to be undertaken to mitigate the risk of large-scale loss of defined breeds in the face of challenges such as increasing human population, climate change and emerging diseases. It was also envisaged that such intervention, when successful, would also make a significant contribution to Millennium Development Goals 1 and 7: eradication of extreme poverty and hunger, and ensuring environmental sustainability, respectively. The Millennium Development Goals (or agenda) are a blueprint of eight goals referred to as the UN Millennium Declaration, which was commissioned by the UN general assembly in September, 2000 (United Nations, 2000). The objective of the declaration is to galvanize unprecedented efforts from all member countries to reverse the poverty, hunger and disease affecting billions of people around the world within a 15-year time frame. Despite the historic breakthrough at the Interlaken Summit, little progress has been made so far, especially in developing countries, due to several factors, the most prominent being a general lack of technical capacity and financial resources (FAO, 2007).

The Domestic Animal Diversity Information System (DAD-IS) is an information and communication tool that was set up to coordinate management strategies developed for domestic animal diversity at global, regional and national levels. This system has challenges, especially regarding the quality of entries from developing countries (Tixier-Boichard, 2014). Most of the data submitted, especially from Africa, requires regular updating to make them relevant to the current situation. For example 48% and 53% of mammalian and avian breeds recorded in DAD-IS were found to lack sufficient demographic information necessary for the assessment of their precise risk status (Groeneveld et al., 2010). Furthermore, 87% of entries regarding breed demographics were found to be based on a survey or census, thus presenting a significant limitation, and might be unreliable (Groeneveld et al., 2010).

In recognition of these and other shortcomings in attempts at addressing global animal genetic resource erosion issues, the European Union has recently commissioned a three-year global programme named 'The GLOBALDIV Project' (Ajmone-Marsan & The Globaldiv Consortium, 2010). The GLOBALDIV project also known as 'Global View of Livestock Biodiversity and Conservation' had representations from the FAO of the UN, the International

Livestock Research Institute (ILRI), the International Atomic Energy Agency (IAEA), and 34 individual international researchers from key institutions that are working in areas related to the characterisation of farm animal genetic resources (Globaldiv consortium, 2010). The main aim of this project is to integrate and disseminate the experience of past, large-scale, biodiversity projects and to review the main drivers of biodiversity loss, and then to implement strategies for the conservation of farm animal genetic diversity. Notable among the recommendations of the GLOBADIV project is the need for amalgamation of the disciplines of genetics, socioeconomics and geographic information science for efficient valuation of domestic animal genetic resources.

Currently, improved geo-referencing methods, for example global positioning systems (GPS), are being used as part of a range of measures to provide better production-environment descriptors (Groeneveld et al., 2010). However, because of the dynamic nature of domestic livestock diversity, it is now obvious that more innovative interventions are required to provide precise information on breed structure and status and effectively halt the rapid loss of global livestock genetic diversity. For any livestock breed considered to be at risk, it is recommended that the monitoring of population trends in terms of population size and structure must be carried out at least once per generation (Groeneveld et al., 2010). More recently, it was indicated that currently available data are inadequate for the ascertainment of the real extent of domestic animal genetic erosion (Bruford et al., 2015). The development of breed-specific identification tools for each characterised livestock breed will not only facilitate the process of regular monitoring of population trends and demographics, but also promote conservation.

This review summarises our knowledge of (i) the key importance of domestic animal genetic resources, (ii) the threats to this resource diversity, (iii) the current status of domestic animal genetic resources, and (iv) conservation methods, with specific emphasis on a molecular genetics approach. We conclude with an assessment of the potential development and use of reliable breed identification tools for livestock breeds for enhancing modern conservation biology studies and preservation of livestock breed diversity.

### **1.1.2 Key Scientific, Cultural and Economic Importance of global livestock genetic resources**

Domestic livestock are known directly to provide food and livelihoods to more than 90% of the 1.97 billion people who live on less than one US\$ a day (Anderson, 2003). With a total global asset value of US\$ 1.4 trillion, domestic livestock is reported to contribute 33% and

17% to global protein and kilocalorie consumption, respectively (Herrero et al., 2009). In many developing countries, apart from the provision of food and income, livestock transactions also develop and foster meaningful and emotional social relationships between and among communities (McCorkle & James, 1996). The so-called minor livestock species, although fewer in population number and distribution, are typically of critical importance in terms of cultural heritage and for the livelihood of their owners (McCorkle & James, 1996; York & Mancus, 2013). For instance, draught-animal power plays an essential role in the livelihoods of marginal communities in many developing countries in Asia, sub-Saharan Africa, and Latin America (Barrett, 1992; Lawrence & Pearson, 2002; Teweldmehidin & Conroy, 2010). In addition to these traditional important uses, several species of animals are now used as models in toxicology studies to ascertain the hazard level to humans of prospective drugs (Olson et al., 2000). For example, the miniature pig was identified as an ideal non-primate model for chromosomal abnormalities, skin cell therapy and neural stem cell studies (Vodička et al., 2005). Also a strain of rabbit referred to as ‘Watanabe heritable hyperlipidemic’ was found to be a good model for the study of human myocardial infarction (Shiomi et al., 2003). It has been recommended that comparative medicine, which entails disease studies across animals and human species, holds the key to efficient prevention and control strategies for many zoonotic diseases (Kahn, 2006). Livestock diversity should not only be considered on the basis of global food security, but also as having critical cultural, economic and scientific importance, both currently and into the future.

### **1.1.3 Threats to Global Livestock Genetic Resources**

The global domestic animal or livestock genetic resources (AnGR) are defined as the sum total of animal species, breeds and strains that currently are, or may be, of future economic, scientific and cultural heritage importance to humans. For the purpose of conservation it is usually breed diversity rather than species diversity that is of greater importance (Philipsson, Rege, & Okeyo, 2006). According to the latest report by the commission on animal genetic resources the percentage of local livestock breeds considered to be at risk of extinction increased by two percentage points between 2010 and 2012 (FAO, 2013). This outlook on the prevailing extinction rate of livestock, although alarming, is likely to be an under-estimation of the actual situation, especially in relation to estimates for developing regions of the world such as sub-Saharan Africa (Rege & Gibson, 2003).

The loss of livestock genetic diversity reduces the range of opportunities available to confront the challenges of unpredictable future events, such as climate change, social change, disease epidemics, selection errors, and many others (Anderson & Centonze, 2007). Some less-common or rare breeds of livestock may also contain rare gene variants that provide greater resistance/resilience to disease and parasites. For many years, the Djallonke sheep and N'Dama cattle breeds of West Africa were regarded as less-desirable livestock because of their generally lower productivity, until these breeds were found to possess an inherent ability to resist the debilitating African animal trypanosomiasis disease (Dolan, 1987; Geerts et al., 2009; Goossens et al., 1999; Mwai et al., 2015; Tano et al., 2003). These breeds have since gained popularity in the region, particularly in the trypanosomiasis endemic areas, prompting their inclusion in strategies for mitigating the effects of the disease (Murray et al., 1984; Naessens, Teale, & Sileghem, 2002). Several quantitative trait loci studies for trypanotolerance in these breeds have been undertaken to facilitate this process (Dayo et al., 2012; Gautier et al., 2009; Hanotte et al., 2003; Kemp & Teale, 1998). In another example, an approximate 50% increase in weaning rate was attained when the Booroola *FecB* gene of the lower productivity small Garole sheep breed was introgressed into the highly productive but low fecundity Deccani sheep breed in India (Marshall et al., 2011). Furthermore, behavioural traits such as ability to use a greater range of food sources (which may not normally be efficiently digested in the more common commercial breeds), tolerance of heat and/or cold, and even differences in mothering abilities are all important heritable traits that should be preserved. As many of these unique breeds are in developing countries and as they are not currently adequately characterised, it is therefore important to conserve as much as possible of this existing genetic diversity, because we do not know its genetic potential (Mwai et al., 2015). While extinction is a natural process due to the presence of inferior traits (for example, the Djallonke sheep breed not being commercially desirable), until these breeds are fully genetically characterised, it is not known what genetic potential we are losing for future generations that face different challenges. Many of the presently uncharacterised breeds might serve as important genetic reservoirs for future selection options (Ciani et al., 2013). A more developed strategy of conservation such as has been employed in the preservation of plant germplasm is probably critical for future sustained food security (Sachs, 2009). Furthermore, the plasticity of genotype can be enormous in certain cases, and may not be dependent on the long-term conservation of rare breeds or breed groups. As likelihood that rare alleles and genetic variation within individuals and within a breed could certainly be as large, if not larger than between breeds. The plasticity of genotypes in such cases could be explored for the long-term adaptive fitness of livestock systems.

There is a wide spectrum of interrelated man-made and natural factors that pose varying levels of threats to global AnGR (Philipsson et al., 2011; Rege & Gibson, 2003). The factors that are responsible for the erosion of genetic diversity are often a function of the size of the population under consideration (Barbato et al., 2015). Generally, the smaller a livestock population, the greater is its vulnerability to extinction (Biscarini et al., 2015; Henson, 1992; Ramstad et al., 2004). Human factors offer the greatest threat to global livestock diversity (Biscarini et al., 2015; Frankham, 1995). The human factors include, but are not limited to; intensive selective breeding, overexploitation, political instability and wars (Goe & Stranzinger, 2002), indiscriminate crossbreeding (Alvarez et al., 2009; Wollny, 2003) and general neglect or lack of breeding programmes (Rege et al., 2011; Wollny, 2003). Interestingly, these human factors vary across both developed and developing regions of the world. In the developed regions, the threat to livestock diversity is mostly associated with overexploitation such as specialised breeding in response to dynamic socioeconomic pressures (Groeneveld et al., 2010). This trend is also expedited partially by often misguided or inappropriate application of advanced molecular biology technologies (Tisdell, 2003). Conversely in developing countries, the main factors are a general neglect of livestock and or poorly structured breeding programmes driven in part by lack of technical knowledge and financial resources (Alvarez et al., 2009; Biscarini et al., 2015; Philipsson et al., 2011). In the face of this clear dimorphism, it is of utmost importance to take measures necessary to minimise the ‘Swanson dominance effect’ (Tisdell, 2003). The Swanson dominance effect refers to a phenomenon in which the choices made by the earliest developing societies influence the later pattern of development in later societies. There have been reports of livestock keepers in parts of sub-Saharan Africa abandoning their locally adapted breeds in favour of specialised potentially highly productive, but non-adapted exotic breeds, thereby leading to a decline in diversity (Groeneveld et al., 2010; Wollny, 2003). Nonetheless, regardless of the region of the world, general increases in human population tend to impact negatively on livestock diversity.

Natural events that have commonly been cited as major causes of erosion of livestock genetic resources include tsunamis, earthquakes, hurricanes, droughts, disease epidemics, famine and floods (Prentice & Anzar, 2011). In the past two or more decades, climate change has emerged as a higher-level driving force for reduction in animal genetic resources (AnGR) (Nardone et al., 2010; Thornton et al., 2009). Many reports have described the expected impact of climate change on livestock production systems and diversity (Banik et al., 2015; Herrero et al., 2009; Hoffmann, 2010; Kantanen et al., 2015; McMichael et al., 2007; Naskar et al., 2015). This is mainly because of the direct and indirect implications of climate change on both the frequencies

and intensities of most of the causative factors for genetic erosion mentioned previously (Naskar et al., 2015). The irony, however, is that a few livestock species contribute significantly to climate change, as they contribute about a fifth of global greenhouse gas emissions (Garnett, 2009; Gavrilova et al., 2010; McMichael et al.; Shields & Orme-Evans, 2015).

Natural and human-made evolutionary forces either directly or indirectly can cause a reduction in the effective population size ( $N_e$ ) of a livestock breeding population. Therefore, the genetic variability of subsequent populations is drastically reduced because it is derived from the genetic constitution of the few survivors remaining from the original population (Allendorf, 1986). In population genetic studies, these reductions in population size are referred to as bottlenecks. A population that passes through a bottleneck loses alleles and usually shows reduced average heterozygosity (Allendorf, 1986; Nei, Maruyama, & Chakraborty, 1975), but could also lead, temporarily, to an increase in heterozygosity if more rare alleles are lost in the process (Hundertmark & Van Daele, 2010; Luikart & Cornuet, 1998). This temporary increase in heterozygosity occurs only if the loss of the rare alleles due to the bottleneck event (mutation-drift equilibrium) has more effect on the expected heterozygosity of a given set of alleles than what is to be expected of that set of alleles under a Hardy-Weinberg equilibrium. However, it is the overall decrease in genetic variation of the population post-bottleneck events that is of major relevance. Regardless of the cause of a bottleneck, it may take many generations to restore the original level of heterozygosity through new mutations (Chakraborty & Nei, 1977). Generally, the impact of a bottleneck is logically more profound on small breeding populations because of the larger correlative effect of the resultant diminished genetic variability on population fitness compared to large breeding populations. In population genetic studies, a bottleneck effect is referred to as a founder effect if it is associated with the founding of a new population (Dlugosch & Parker, 2008; Ramstad et al., 2004; Templeton, 1980). Random events such as founder and bottleneck effects that imperfectly eliminate genes and reduce variability within a population are also described as genetic drift (Newman & Pilson, 1997; Ramstad et al., 2004). Reduction in heterozygosity in a livestock population can be associated with decline in fitness of individual members, as is often the case in wild populations (Worley et al., 2010). This is because within small populations, the rate of inbreeding is much higher and consequently there is higher likelihood of the expression of deleterious recessives in a homozygous state. The expression of deleterious alleles has adverse effects on the livestock population, often presenting as reduced production, reproduction and survival (Dlugosch & Parker, 2008; Lacy, 1997). Frankham (1995) and Lacy (1997) have described the positive correlation between inbreeding and risk of extinction. The effective population size model

takes into account important population variables such as age and structure, inbreeding rates, genetic drift, genetic diversity and sex ratio. For example, a population of four males and four females constitutes the same effective population size as that of 100 females and only two males (Henson, 1992). Therefore, the effective population size is the preferred indicator of livestock conservation risk status (Dlugosch & Parker, 2008; Nei et al., 1975). In a breed regeneration programme, the effective population size can be enhanced by equalising the male to female ratio, standardising litter size and longevity within the breeding population, so as to ensure that each animal contributes equally to the next generation. However, it is apparent that the estimation of the effective population size and subsequent determination of its conservation status for a given breed is limited by the lack of availability of a reliable breed identification tool for any specific breed.

#### **1.1.4 Assessment of Livestock Genetic Diversity and Conservation Status**

In order to manage livestock genetic resources sustainably a comprehensive knowledge of diversity within and between breed populations is required (Groeneveld et al., 2010). A major step towards standardising the assessment criteria for livestock breed conservation status was the establishment of a universal classification framework by the FAO for categorising risk status. The current classification of livestock conservation risk status contains seven categories: extinct, critical, critical-maintained, endangered, endangered-maintained, not at risk, and unknown (FAO, 2013). Regular assessment of genetic conservation status of livestock is of fundamental importance to prevent genetic erosion and to preserve diversity.

Key to achieving an effective assessment of livestock conservation status is a reliable mode of identification of members of a target breed. There are two broad methods for identifying individual members of a livestock breed, and their merits and demerits have been discussed thoroughly elsewhere (Agaviezor et al., 2012; Ashley & Dow, 1994; Birteeb et al., 2012). These methods comprise phenotypic and molecular identification techniques. Traditionally phenotypic identification has been used to identify the breed of an individual in livestock genetic diversity studies. The phenotypic variables usually used comprise physical features (e.g. shape of horn, ears, body measurements, colour, etc.), production traits (e.g. growth parameters), reproductive traits (e.g. fecundity) and survival traits (e.g. disease resistance, drought resistance) (Brinks et al., 1964; Gwakisa, Kemp, & Teale, 1994; Reverter et al., 2003). These methods are used extensively not only because they are inexpensive and often do not require the use of sophisticated equipment, but also may be useful criteria to some breed

societies. However, the major disadvantage is that the genetic diversity is observed only at the phenotypic level and this does not always correspond to actual diversity at the DNA level (Feliuss et al., 2014).

It is possible to find different phenotypes with similar genotypes, typically due to genotype–environment interactions, for example as observed in Brazilian sheep breeds (Paiva et al., 2005) and Egyptian sheep breeds (Ali, 2003). Similar phenotypes with different genotypes also occur, as observed between the West African Djallonke sheep and F1 Djallonke–Sahelian crossbreeds (Alvarez, Capote, et al., 2012; 2009; Wafula et al., 2005). As a result, the use of molecular tools in many assessment studies of genetic diversity in different regions of the world revealed varying degrees of unexpected introgression and admixture in livestock populations. These studies include the Djallonke sheep breed of sub-Saharan Africa (Alvarez et al., 2009; Wafula et al., 2005), Herdwick sheep of the United Kingdom (Bowles, Carson, & Isaac, 2014) and alpaca and llama of Latin America (Kadwell et al., 2001). This obvious shortcoming has rendered the use of phenotypic methods in isolation as unreliable for determination of livestock breeds for the purpose of genetic diversity studies.

In livestock genetic diversity studies, the molecular method for determining breed identity entails two main approaches based upon either protein markers or DNA markers (Ferguson et al., 1995; McMahon, Teeling, & Höglund, 2014). Protein markers, also referred to as allozymes, are based on the characteristic polymorphism of the blood group systems, leucocyte antigens and enzymes (Dodgson, Cheng, & Okimoto, 1997). This molecular method employs these protein markers to estimate genetic variability in livestock populations as well as phylogenetic relationships between breeds (Pépin & Nguyen, 1994; Witko-Sarsat et al., 1996). Although better than the phenotypic method, the use of protein markers is too expensive for a large number of loci, and lacks the power to resolve differences between closely related breeds, because of limits of detection of genetic variation (Engel et al., 1996; Ferguson et al., 1995; Toro, Fernández, & Caballero, 2009). The use of DNA markers is the most reliable molecular method for assessment of genetic diversity (Liu & Cordes, 2004). Nuclear and mitochondrial DNA marker analyses have revealed detailed information on many domestication events, such as their timing and location (Bruford, Bradley, & Luikart, 2003; Zhao et al., 2013). DNA marker analyses provide an added opportunity for investigating the genetic composition of both extinct and endangered breeds without destructive sampling.

There are seven principal DNA marker techniques commonly used for livestock diversity studies (Sunnucks, 2000). These seven DNA marker techniques have been discussed thoroughly and their advantages and disadvantages are well documented. These techniques are:

restriction fragment length polymorphism (RFLP) (Beckmann & Soller, 1983, 1986; Thurston et al., 2002), mitochondrial DNA barcoding (mtDNA) (Avisé et al., 1987; Avisé & Ellis, 1986; Harrison, 1989; Kocher et al., 1989; Zhang & Hewitt, 1996), random amplified polymorphic DNA (RAPD) (Ali et al., 2004; Dodgson et al., 1997; Koh et al., 1998; Levin, Crittenden, & Dodgson, 1993), amplified fragment length polymorphism technique (AFLP) (Blears et al., 1998; Parsons & Shaw, 2001; Savelkoul et al., 1999), Y-chromosome technique (Bruford et al., 2003; Zeder et al., 2006), variable number of tandem repeats (VNTR) (minisatellite and microsatellite markers) (Chistiakov, Hellemans, & Volckaert, 2006; Fan & Chu, 2007; Lopes et al., 2015; Zane, Bargelloni, & Patarnello, 2002) and single nucleotide polymorphism (SNP) based techniques (Andersson & Georges, 2004; Liu & Cordes, 2004; McMahon et al., 2014; Morin, Luikart, & Wayne, 2004; Tixier-Boichard, 2014; Vignal et al., 2002). The latter two DNA techniques are the most popular.

The advancement of DNA technologies during the past three decades, and particularly since 2007 when high-throughput next-generation sequencing became readily available, is revolutionising livestock population genetics studies (Helyar et al., 2011; Schlotterer et al., 2014). This revolution is expedited by the concomitant advancement in bioinformatics tools and pipelines (Kofler, Nolte, & Schlotterer, 2016). DNA markers have been used not only for diversity studies but also for molecular characterisation of numerous livestock breeds worldwide (Agaviezor et al., 2012; Al-Atiyat, Salameh, & Tabbaa, 2014; Alvarez, Capote, et al., 2012; Bowles et al., 2014; Chenyambuga et al., 2004; Mukesh et al., 2004). The dramatic reduction in the cost of DNA markers has facilitated their greater use by researchers. AFLP and RAPD markers are both bi-allelic and dominant in nature, and hence are less informative and also have low reproducibility compared to the other markers (Vignal et al., 2002). These characteristics have rendered them less popular for most animal-based molecular genetic studies. RFLP markers are bi-allelic and co-dominant, and were famously used in the first large-scale mapping of the human genome. However, RFLPs have now been superseded by more informative microsatellite markers, a type of VNTR for both animal and human genome studies. In turn, microsatellite markers have been largely supplanted by single nucleotide polymorphism (SNP) arrays. mtDNA along with microsatellite markers were once popular molecular genetic techniques of choice for evolutionary and ecological studies, however the molecular information provided by mtDNA markers is limited to only maternally inherited loci (Morin et al., 2004). The use of mtDNA techniques, in combination with archaeological data, has provided precise information on most of the important centres of domestication for the main livestock species around the world (Bruford et al., 2003; Globaldiv consortium, 2010;

Guo et al., 2006; Zeder et al., 2006). Similarly limited, the use of Y-chromosomal haplotype markers elucidates specific molecular information only on paternally inherited traits (Luikart et al., 2006). The VNTR and the SNP techniques will be discussed in greater detail below because of their current wider application compared with the other molecular markers.

#### **1.1.4.1 Variable Number of Tandem Repeats (VNTRS)**

The application of VNTRs for assessment of genetic variation, sub-structuring and hybridisation in natural populations has been reviewed in great detail previously (Bruford & Wayne, 1993; Chistiakov et al., 2006; Fan & Chu, 2007; Sunnucks, 2000). The VNTR technique is based on the abundance of tandem repeats of simple sequences of nucleotides throughout the eukaryotic genome (Takezaki & Nei, 2008). These VNTRs have been categorised into minisatellites and microsatellites according to the number of nucleotides per motif of repeats. VNTRs of between 1 and 6 nucleotide base pair units are referred to as microsatellites (Ashley & Dow, 1994; Chistiakov et al., 2006; Fan & Chu, 2007), whereas a range of between 10 and 60 nucleotide base pair units is regarded as a minisatellite (Ashley & Dow, 1994; Wasko & Galetti, 2003). Whereas minisatellites are concentrated towards the telomere of chromosomes, microsatellites are randomly distributed in chromosomes. Microsatellite markers are highly polymorphic, co-dominant markers of relatively small size, and hence are more amenable to polymerase chain reaction (PCR) typing than are minisatellites (Zane et al., 2002). Also, in comparison to the RFLP and RAPD techniques, the genetic basis of variability is readily apparent for microsatellites. Most microsatellites are located in non-coding regions of the genome (Chistiakov et al., 2006). Generally, microsatellite primers developed for one species of livestock are broadly applicable to other closely related species. For example, microsatellite markers developed for studies in bovine species are applicable to caprine and ovine species (Engel et al., 1996). This versatility has led to the popularity of microsatellite maps for economically important livestock species (Sunnucks, 2000).

Microsatellites have been used in linkage mapping in diverse organisms, for example in the bovine genome (Barendse et al., 1997), porcine genome (Rohrer et al., 1994), human genome (Dib et al., 1996), and ovine genome (Maddox et al., 2001). Microsatellites have also been employed for the identification of quantitative trait loci (QTL) in major livestock species, for example, carcass composition and growth rate in cattle (Casas et al., 2000), back fat thickness and intramuscular fat in pigs (Rohrer & Keele, 1998) and intestinal parasitic infection in sheep (Davies et al., 2006). Other population genetics studies accomplished with microsatellite

markers include the determination of evolutionary relationships (Alvarez, Capote, et al., 2012; Buchanan et al., 1994; Vanhala et al., 1998), estimation of pedigree errors (Visscher et al., 2002) and determination of genetic diversity among livestock populations (Alvarez, Capote, et al., 2012; Alvarez et al., 2009; Curković et al., 2015; Marletta et al., 2006; Medugorac et al., 2011; Wafula et al., 2005). The genetic distance between individuals within a livestock population indicates the suitability of an individual for conservation purposes. Individuals within the same breed with the widest differences in genetic distances are deemed most suitable candidates for conservation programmes. The estimates of genetic distances are also relevant for the determination of divergence time and construction of phylogenies (Takezaki & Nei, 1996). Prior to the use of SNP markers, microsatellites were the most popular and efficient technique for genetic-diversity investigation, not only in livestock but also in humans. Microsatellites continue to be seen as a method of choice for many researchers in breeding programs, particularly in third-world and developing countries, due to their low cost, relative ease of analysis and requirement for relatively unsophisticated scientific equipment (Rege et al., 2011). Whereas newer technologies offer better prospects, the enabling supporting infrastructure is often not available in developing regions of world. For example, the analysis of large-scale genomic data requires reliable and fast internet access for web-based reference sequence information, which currently cannot be guaranteed in many sub-Saharan African countries (Gulati, 2008; Oyelaran-Oyeyinka & Lal, 2005). The same can also be said of the availability and reliability of electric power supplies necessary to support cryobanking of important genetic materials (Deichmann et al., 2011; Wolde-Rufael, 2006). Given the levels of existing infrastructure and human technical capacity in many developing countries, significant investment is required to implement some of the recent genomic technologies for sustainable livestock production and conservation (Rege et al., 2011).

#### **1.1.4.2 Single Nucleotide Polymorphism (SNP) Markers**

The growing importance of SNP marker applications in molecular genetics has been reviewed in detail by Barbato et al. (2015), Broxham (2015), Goddard and Hayes (2009), Vignal et al. (2002), Hamblin, Warburton, and Buckler (2007) and Morin et al. (2004). SNPs represent a location within a DNA sequence for which more than one nucleotide type is present within a given population (Morin et al., 2004). In a strict molecular sense, SNPs are base substitutions within nucleotide sequences, and the very high density of their occurrence in the genomes of eukaryotes, including livestock, has been of great significance in population genomics studies

(Goddard & Hayes, 2009; Vignal et al., 2002). Although SNPs are usually bi-allelic (sometimes tri-allelic or quadri-allelic) co-dominant molecular markers, their high density permits, more than any other technique, very detailed information to be elucidated on genome dynamics within a study population (Hamblin et al., 2007; Morin et al., 2004). Furthermore, they provide deeper insight than microsatellites with respect to linkage disequilibrium and haplotype diversity, pedigree information and past demographic events, such as bottlenecks within a target population (Clarke et al., 2014; Gautier et al., 2007; Helyar et al., 2011; Morin et al., 2004; The Bovine HapMap Consortium, 2009; Vignal et al., 2002). SNP markers also allow for standardised data recording, and are stable over generations if selected from neutral genomic loci (Tixier-Boichard, 2014). These features of SNP markers are opening opportunities for wider applications of SNP markers in understanding of livestock genetic architecture, such as precise identification of genomic regions that control traits of economic and survival importance (Kohn et al., 2006; Matukumalli et al., 2009) and ultimately genomic selection (Choi et al., 2015; Clarke et al., 2014; Goddard & Hayes, 2009). These advances in genetic marker application for use in population genetic studies will not only enhance the development of improved livestock production systems, but most importantly will facilitate the development of efficient conservation strategies.

#### **1.1.5 Genomic Methods for Breed Prediction**

The unique genetic structure of livestock breeds, shaped by their demographic history of natural and artificial selection, provides a basis for the assignment of an individual to a particular breed (Bertolini et al., 2015). The large numbers of SNPs identified in various domestic animal species have been used to develop species-specific standard technology products referred to as BeadChips or SNP chips (Wilkinson et al., 2011). These SNP chips are commercially available, and have been designed to amplify genome-wide SNP loci rapidly in an automated platform to generate large-scale genomic SNP data for analysis. Examples are the Illumina ovine SNP 50 BeadChip and the Illumina bovine SNP50 BeadChip developed for sheep and cattle, respectively (Bertolini et al., 2015; Dodds et al., 2014). An analysis of the data generated using SNP chip technologies has shown that it is possible to assign an individual animal correctly to a specific breed (Table 1.1). Moderate- to high-density SNP genotyping assays are frequently used to capture common genomic variations within breed populations (Bertolini et al., 2015; Broxham, 2015; Dodds et al., 2014; Frkonja et al., 2012; Kijas et al., 2009; Lewis et al., 2011; Rolf et al., 2014; Sasazaki et al., 2007; Suekawa et al., 2010;

Wilkinson et al., 2011). Bioinformatics and statistical tools such as STRUCTURE, principal component analysis (PCA) and discriminant analysis have been widely applied to these SNP data sets with varying levels of success (Gilbert et al., 2012; Herrero-Medrano et al., 2013; Hubisz et al., 2009; Schwartz & McKelvey, 2009). Unlike STRUCTURE and PCA, discriminant analysis is not considered a multivariate statistical method for assignment of individuals to a population. Furthermore, the discriminant analysis does not permit the fractional or mixed prediction of individuals in a subject population (Dodds et al., 2014). This is a major limitation for studies that require multiple predictors, making it less suitable for breed predictions (Dodds et al., 2014). However, with continuous advances in bioinformatics tools, many more tools are becoming available for this kind of analysis. Recently, two different analyses of Illumina OvineSNP50 genotyped data were used to assign four New Zealand sheep breeds correctly with high prediction accuracy (0.85–0.97) (Dodds et al., 2014). The two methods used were a regression analysis with a genomic selection algorithm that employed allele frequencies, and genomic Best Linear Unbiased Prediction (gBLUP) estimates respectively, derived from a pure-bred subset of each sheep breed population. These estimates were then used as the training data set for respective breed predictions in each of the four populations. The two methods produced different prediction accuracies that depended upon the breed structure of the subject populations. It was concluded that the accuracy of breed prediction was enhanced if the composition of the training set is representative of the breed diversity within a subject population (Dodds et al., 2014). A recent study of the genomics of cattle in the USA beef industry has supported this conclusion, particularly for predictions in multi-breed beef cattle populations (Rolf et al., 2014). The accuracy of predictions obtained from both methods was similar to that recorded with STRUCTURE (Dodds et al., 2014). However, unlike these two methods, STRUCTURE does not provide a prediction equation for subsequent breed prediction in a subject population (Dodds et al., 2014). STRUCTURE analysis of data also has low reproducibility (Gilbert et al., 2012). In spite of these drawbacks, the STRUCTURE algorithm has been used extensively in clustering of genetic data (Falush, Stephens, & Pritchard, 2007; Hubisz et al., 2009; Schwartz & McKelvey, 2009).

Principal component analysis (PCA) is also a powerful multivariate tool that facilitates the elicitation of unknown population clusters (Lewis et al., 2011). When applied to genomic data, it has been found to group individuals of the same breed together (Dodds et al., 2014). The use of a combination of ancestry-informative marker metrics and PCA using 30,501 SNPs on the Bovine HapMap accurately predicted 19 cattle breeds (Lewis et al., 2011). This result led to the conclusion that a carefully selected panel of 250–500 SNPs from the Bovine HapMap data

set was sufficient for correct breed assignment. Justifiably, Lewis et al. (2011) also conceded that the sensitivity and the resolving power of their approach would be higher if applied to denser genomic data than the Bovine HapMap data set used. This view is also supported by Wilkinson et al. (2011). However, the PCA result does not readily translate to the actual breed proportion estimates in mixed breeds. Hence, it is more suitable for the verification of a pure-breed member (Dodds et al., 2014). Nonetheless, this combined tool approach is said to be suitable for the reliable tracing of breed-specific branded products in the meat industry (Lewis et al., 2011). Prior to this study, SNP-based markers derived from the AFLP technique were used for distinguishing between Australian and Japanese beef (Sasazaki et al., 2007). Although the AFLP-derived markers were of low resolution, the power was sufficient to discriminate cattle breeds from the two countries. A more stringent breed-specific SNP marker panel was later developed from Bovine 50K SNP BeadChip data and this was able to discriminate between Japanese and America cattle products (Suekawa et al., 2010). The higher efficiency marker panel was developed in response to an outbreak of bovine spongiform encephalopathy (BSE) in the USA, and comprised only six highly informative breed-specific SNPs (Suekawa et al., 2010). More recently, the use of 48 and 96 highly informative SNP markers derived from a combined PCA in combination with a ranking algorithm (random forests) of Illumina Bovine SNP50 BeadChip genotyped data, correctly assigned four Italian cattle breeds (Bertolini et al., 2015). A few of the highly informative SNPs used in that study were found to be located in loci associated within important quantitative traits for some cattle breeds. A systematic assessment of the efficiency of four different methods for identifying population-informative SNPs from the SNP50 BeadChip data set showed no gain of further assignment power beyond the use of more than 200 SNPs in a panel, for all the approaches (Wilkinson et al., 2011). Wilkinson et al. (2011) also showed that a panel of 60 SNP markers was the minimum required for successful prediction of the cattle breeds investigated. However, more genetic markers (in excess of 200) will be required successfully to assign closely related breeds and far fewer for distantly related breeds. Hence, Wilkinson et al. (2011) provided evidence that the number of SNPs required for correct assignment of an individual to a breed is directly proportional to the genetic heterogeneity or homogeneity of the sampled population. In a more recent study, two separate panels of SNPs derived from 21 different sheep breeds from Italy and Slovenia were used to assign all the sheep correctly to the breeds (Dimauro et al., 2015). This study combined three different types of discriminant analyses on an Illumina Ovine SNP50 genotyped data set from all 21 breeds to produce a reduced panel of 108 and 110 SNP markers.

**Table 1.1 Present status of efforts to identify livestock breeds using genomic methods**

Species	Genotyping platform	Method	Reference
<b>Italian sheep breeds</b>	Illumina OvineSNP50	Stepwise discriminant analysis Canonical discriminant analysis Discriminant analysis GENECLASS 2 software POWERMARKER software	(Dimauro et al., 2015)
<b>Italian cattle breeds</b>	Illumina Bovine50 BeadChip	Principal component analysis, random forest regression	(Bertolini et al., 2015)
<b>New Zealand sheep breeds</b>	Illumina OvineSNP50	Regression and genomic BLUP STRUCTURE algorithm	(Dodds et al., 2014)
<b>Korean native and Holstein cattle breeds</b>	Illumina HiSeq 2000	Mapping of Reads to Bovine Genome Assembly UMD 3.1 Samtools-0.1.18 MPILEUP GATK ver. 2.4	(Choi et al., 2014)
<b>Yunnan (South China) chicken breeds</b>	Microsatellite markers	Nei's genetic distance Hardy Weinberg analysis GENALEX 6 Software Weir & Cockerham's $F_{ST}$	(Huo et al., 2014)
<b>Swiss cattle breeds</b>	Illumina Bovine50 BeadChip	STRUCTURE algorithm BayesB Partial least squares regression	(Frkonja et al., 2012)
<b>19 worldwide cattle breeds</b>	Bovine HapMap data set SNP marker	Principal component analysis Nearest neighbour classification Algorithm	(Lewis et al., 2011)
<b>European cattle breeds</b>	Illumina Bovine50 BeadChip	Weir & Cockerham's $F_{ST}$ , Delta, Wright's $F_{ST}$ Principal component analysis methods	(Wilkinson et al., 2011)
<b>Japanese and USA cattle breeds</b>	Illumina Bovine50 BeadChip	Allelic frequency method BLAST programs	(Suekawa et al., 2010)

Species	Genotyping platform	Method	Reference
		PCR restriction fragment length polymorphism	
<b>Italian cattle breeds</b>	Microsatellite markers	STRUCTURE algorithm Wright's F-statistics MolKin V3.0 software	(Bozzi et al., 2009)
<b>Local European and Asian chicken breeds</b>	Microsatellite markers	Nei's genetic distance Hardy Weinberg analysis GENECLASS 2 software Multiple CO-inertia analysis Weir & Cockerham's $F_{ST}$	(Berthouly et al., 2008)
<b>Japanese and Australian cattle breeds</b>	Microsatellite markers	Polymerase chain reaction–amplified fragment length polymorphism	(Sasazaki et al., 2007)
<b>Iberian pig breeds</b>	Microsatellite markers	PHYLIP software package Markov chain Monte Carlo methods	(Fabuel et al., 2004)
<b>Asian and European pig breeds</b>	Mitochondria DNA D-loop assay	Polymerase chain reaction Clustal W software	(Kim et al., 2002)

Further advances in genomics have also shown that the use of SNP chip technology is prone to ascertainment biases because the discovery SNP panels are derived from small numbers of individuals from selected populations that are not representative of all the species populations (Albrechtsen, Nielsen, & Nielsen, 2010; Foll, Beaumont, & Gaggiotti, 2008). The bias in sheep was particularly evident for African and Asian breeds as the SNPs were originally identified from European breeds (Kijas et al., 2012). Such ascertainment biases are likely to skew inferences determined from genotyped data such as allele frequency spectra and genetic differentiation of subject populations.

Next-generation sequencing (NGS) of genomic DNA will provide denser SNP data than the Illumina OvineSNP50 BeadChip or any medium-density marker set that has previously been used in this type of study, and does not suffer from ascertainment bias. It is important to note that NGS data have inherent challenges arising from alignment and sequencing errors, but these are smaller in comparison to the biases of the SNP-chip genotyped data (Albrechtsen et al., 2010). For example, a whole genome of an animal from a Korean cattle breed that was sequenced on the Illumina HiSeq 2000 platform resulted in more than 10 million SNPs being

identified, 54% of which were novel (Choi et al., 2014). Furthermore, another study showed a reduction in prediction accuracy when a SNP data set derived from Illumina Bovine SNP50 was replaced with that derived from an Illumina SNP3K genotyped data set (Kuehn et al., 2011). The higher resolving power of NGS has been shown to capture more rare breed-specific polymorphisms or more informative polymorphisms (with higher confidence) than bovine SNP50 BeadChip genotyping (Choi et al., 2015; Lee et al., 2013).

Collectively, these studies suggest that the use of high-density data will enable the real possibility of developing a smaller panel containing the most informative breed-specific SNPs having the highest sensitivity for resolving breed differences. Therefore, analyses and use of such NGS data will lead to more accurate breed predictions, and the NGS of individual genomes at high coverage has been referred to as the 'gold standard' for generating quality data (Schlotterer et al., 2014). In spite of the dramatic reduction in the cost of NGS, the cost of sequencing the large number of individuals required for population studies of this nature, at high coverage, is still economically prohibitive. However, it has also been shown that NGS of pools of individuals at a moderate coverage could provide a cost-effective and efficient alternative technique for generating very high density SNP data sets across a genome, compared with NGS of non-pooled individuals (Gautier et al., 2013; Kofler et al., 2016; Schlotterer et al., 2014). Another promising cost-effective approach that was applied successfully to some plant species is referred to as genotyping by sequencing (GBS), and is based on the sequencing and analysing of more informative regions of the genome rather than the whole genome (Elshire et al., 2011). GBS is fast, highly specific and exceedingly reproducible, and could be used to complement the pool-sequencing approach through confirmatory testing where the need arises. For association-mapping studies, the analysis of pooled-sequenced data has more statistical power than SNP arrays (Futschik & Schlotterer, 2010; Gautier et al., 2013; Kofler, Pandey, & Schlotterer, 2011). The high-density SNP data generated by this pool strategy was shown to facilitate the discovery of more accurate allelic frequency estimates across a genome (Futschik & Schlotterer, 2010). The advantages of a pooled-sequencing technique over individual sequencing have been reviewed previously (Schlotterer et al., 2014). NGS sequencing of pools of unrelated individual purebreds from a subject population therefore will enable the identification the most informative breed-specific SNPs. The rationale for sampling unrelated animals instead of related ones is to enable the capturing of a wide spectrum of within-breed genetic diversity of the subject population. This strategy is necessary to minimise the introduction of ascertainment bias into the subsequent

breed identification tool (McTavish & Hillis, 2015). A carefully selected panel of SNPs derived from the identified breed-specific SNPs can serve as an efficient breed identification tool.

Conservation of AnGR comprises all the management practices carried out to preserve the pool of genetic diversity of livestock for the purposes of meeting current and future needs of humans (Rege & Gibson, 2003). The relevance of conservation of AnGR has been discussed from several different perspectives, including economic evaluation as a basis for AnGR conservation decisions (Anderson, 2003; Drucker, Gomez, & Anderson, 2001), the role of cryopreservation, reproductive technologies and genetic resource banks for AnGR conservation strategies (Hiemstra, van der Lende, & Woelders, 2006b; Holt & Pickard, 1999; Mara et al., 2013), information on population kinships as a basis for AnGR conservation decisions (Eding & Meuwissen, 2001) and the challenge of conserving indigenous AnGR diversity (Mendelsohn, 2003). Each breed of livestock consists of unique sets of genes resulting from evolutionary events and diverse selection pressures imposed by the environment combined with the activities of humans over time. It is therefore difficult, if not impossible, to replace lost breeds of livestock, because those unique evolutionary processes cannot be re-created. There has been a general consensus on three critical approaches regarding the conservation of domestic livestock breeds: sustainable utilisation of available livestock breeds; appropriate diversity-based improvement strategies for livestock breeds; and development of appropriate assessment and preservation strategies (FAO, 2000; Hammond, 1999; Koehler-Rollefson & Meyer; Notter, Mariante, & Sheng, 1994; Thornton et al., 2007). In addition to these approaches, the FAO has also recommended the regular monitoring of livestock breed conservation status (FAO, 2013).

Currently the two main methods of AnGR conservation applied are the *in situ* and the *ex situ* methods. The applicability of both conservation methods, and their respective merits and demerits has been reported extensively (Boettcher et al., 2010; Hammond, 1994; Henson, 1992; Mara et al., 2013; Rege & Gibson, 2003). *In situ* conservation can best be described as the sustainable breeding of an endangered livestock breed in their normal adaptive production environment, or as close to it as practically possible, to conserve genetic diversity over a long period (Andrabi & Maxwell, 2007; Henson, 1992). Notable features of *in situ* conservation therefore include selection and mating programmes that retain genetic variation within the target group, as well as management of the ecosystem to sustain their production. The basic requirements for *in situ* conservation programmes are generally readily available and affordable globally. There is a distinct difference between developed and developing countries regarding the minimum number of individuals required to commence an *in situ* programme. This is typically due to general differences in the efficiency of management of their respective

livestock production systems. For example, whereas the minimum number for major livestock breeds (i.e. cattle, sheep, goats, pigs) required for *in situ* conservation is 100–1,000 breeding females in developed countries, no fewer than 5,000 breeding females is recommended for developing countries (Signorello & Pappalardo, 2003). Simon (1999) reported 500 breeding females for pigs and goats, 750 for cattle, and 1,500 for sheep for European breeds. It has been recommended that, ideally, for unrelated animals a minimum of 25 males and 50 females is sufficient to commence an *in situ* conservation programme, because the possible loss of genetic variability is estimated to be less than 1% per generation (Henson, 1992; Mara et al., 2013). However, recent advances in the field of genomics have enabled the elucidation of abundant genomic information *via* high-throughput sequencing technologies and analysis, so a re-evaluation of these recommended numbers required for conservation programmes is overdue. This is because more accurate population genetic parameters such as allele frequencies can be computed for target populations to allow for more precise determination of these numbers. There are a number of flagship *in situ* conservation programmes in place to conserve and improve some disease-resistant breeds of livestock in some African countries, for example, N'Dama cattle in the republic of Guinea (Yapi-Gnaoré, Dagnogo, & Oya, 2003), Djallonke sheep in Ghana and Cote D'Ivoire (Kosgey & Okeyo, 2007), and Tswana sheep in Botswana (Henson, 1992). The unique advantage of the *in situ* conservation method is that the target livestock breed continues to be utilised in the process. However, the danger is that the target livestock breed remains susceptible to uncertain demographic threats such as natural disasters and disease epidemics.

The *ex situ* livestock conservation method is the preservation of endangered livestock outside their normal production systems (Henson, 1992; Hiemstra et al., 2006a). This method is normally applied to target groups that are faced with imminent extinction, and hence requires the use of high-level expertise and technology. The three main *ex situ* methods are cryopreservation, farm park conservation, and breed pools or composite preservation. Cryopreservation, also referred to as *in vitro ex situ* is undoubtedly the most popular of the *ex situ* approaches to conservation of AnGR (Hiemstra et al., 2006a). This approach involves the cryopreservation of eggs, semen and or embryos of endangered or threatened animals in genome banks for use in managing diversity or regenerating the population decades, or even centuries, later (Chen, Zhang, & Yu, 2008; Hanks, 2001; Russo et al., 2007; Xiao-Yong et al., 2008). Cryogenic storage of carefully evaluated genetic material from a target breed population is also seen as an insurance policy against future loss. The merits and demerits of using these approaches have been discussed previously (Boettcher et al., 2010; Munro & Adams, 1991;

Philipsson et al., 2006; Pintado & Hourcade, 2011; Ruane & Sonnino, 2011). The application of cryopreservation formerly depended only on assisted reproductive techniques such as artificial insemination and embryo transfer technologies. However, recent advances in reproductive biotechnologies including semen sexing, embryo micromanipulation and *in vitro* fertilisation have the potential to revolutionise the livestock cryopreservation approach (O'Brien, Steinman, & Robeck, 2009; Prentice & Anzar, 2011). Cryopreserved genetic materials are shielded from the influence of unfavourable environmental conditions in existence in the normal production ecosystems. Regeneration of a breed through only preserved semen requires a number of back crosses (Andrabi & Maxwell, 2007). However, the exact genetic composition of an original breed after going through adaptive selection is not recoverable with only cryopreserved semen that was collected before the adaptation process. In practice, the *in situ* and *ex situ* conservation methods are not mutually exclusive because the cryopreservation approach can be used to complement the *in situ* method to achieve better regeneration of endangered populations. A range of combinations of *in situ* and *ex situ* conservation methods are being applied in a now-popular integrated conservation approach (de Souza et al.; Hiemstra et al., 2006a). It has been recommended that a stock of cryopreserved semen from 25 unrelated sires is sufficient to provide a reasonable diversity for an endangered population (Bruns & Glodek, 1999; Mara et al., 2013).

The farm-park *ex situ* conservation approach is similar to the *in situ* conservation approach, except that the targeted breeds are preserved outside their normal production environment in a specialised institutional setting, also referred to as an Ark-farm (Simon, 1999). Farm-park animals are usually also subjected to more stringent management regimes to conserve natural levels of genetic variability within each species (Chesser, Smith, & Brisbin, 1980). A notable feature of the farm-park approach is its popularity in attracting tourists, and hence creating awareness of the need to conserve endangered animals. The Cotswold farm park in the UK is an example where rare breeds of sheep, goats, cattle, pigs and horses are being conserved, and it attracts more than 100,000 visitors yearly (Henson, 1992). The breed-pool preservation programme is unique in the sense that it involves the breeding together of a pool of two to four rare breeds with similar characteristics, and subsequently managing their offspring to conserve genetic variation (Henson, 1992). It is, however, recommended that the breed characteristics of each of the rare breeds is well ascertained prior to commencing a breed-pool programme (Santos et al., 2013). This method is particularly suitable for genes that control obvious morphological traits and extreme quantitative traits such as coat colour and prolificacy, respectively. Although this approach conserves useful genes from the pool, individual breeds

are lost in the process. An example of the breed-pool approach is the conservation programme of four rare desert goat breeds in the north eastern part of Brazil (Henson, 1992).

Given that no single conservation method is capable of solving the myriad of challenges of domestic animal genetic resource erosion, an integrated conservation approach has been advocated to provide greater efficiency (Rege & Gibson, 2003).

### **1.1.6 Modelling: The Way Forward**

The successful domestication of animals represents a pivotal historic event in the cultural and demographic development of humans. The importance of global domestic livestock diversity to human wellbeing is now well appreciated. This is evident from the globally coordinated efforts directed at halting the decline in AnGR as well as the sustainable utilisation of available livestock resources as discussed herein. These global initiatives have yielded several interventions which are being implemented at the international, regional and local level (Ajmone-Marsan & The Globaldiv Consortium, 2010; FAO, 2013). The main global body is the DAD-IS that coordinates regional bodies, for example the European Farm Animal Biodiversity Information System (EFABIS) and the Domestic Animal Genetic Resource Information System (DAGRIS) for European and African regions, respectively. The regional bodies in turn coordinate the local or national bodies, which essentially are the individual member states of the FAO of the UN. These efforts are being supplemented by the activities of other important organisations, prominent members including the GLOBALDIV, the International Society for Animal Genetics (ISAG), the SAVE foundation and several livestock breed societies worldwide (Broxham, 2015). Some notable progress has been made towards reducing the rate of erosion of global AnGR. For example, the status information for global mammalian and avian livestock breeds in the DAD-IS has increased from 43% and 39%, respectively in 2009, to 57% and 48%, respectively, in 2012 (FAO, 2013). The effective monitoring of breed conservation status of livestock requires at least one census per generation of that target breed (FAO, 2007; Groeneveld et al., 2010). A specific breed identification tool for each livestock breed will expedite this exercise. However, pivotal to the success of these conservation efforts is the reliability of genetic identification of individual members within a target breed. The advancement in molecular technology in the last two decades has significantly increased our understanding of the population genetics of domestic animals. It is apparent that the molecular characterisation of all domestic livestock breeds, particularly in developing countries, is a pre-requisite for their sustainable utilisation and conservation. This

is because characterisation at the molecular level provides precise information for determination of the actual population characteristics such as genetic variation and effective population size (Luikart et al., 2003). Currently, many domestic livestock breeds, particularly those in the developing countries, have yet to be characterised due to myriad issues including the lack of financial and technological capacity. A recent report indicated that the risk status of 36% of all populations of the local livestock breeds still remains unknown (FAO, 2013).

The main molecular technique used for most livestock genetic characterisation was microsatellite markers [for example, in Spanish native cattle breeds (Martín-Burriel, García-Muro, & Zaragoza, 1999), Aberdeen Angus cattle breeds (Vasconcellos et al., 2003), Austrian sheep breeds (Baumung et al., 2006) and indigenous goats in sub-Saharan Africa (Chenyambuga et al., 2004)]. Although highly informative, the current panels of microsatellites used for analyses are not capable of elucidating all the information required regarding breed variation in livestock (Toro et al., 2009). Recently, it is becoming more evident that SNP analysis is more suited for the high-throughput genotyping that is required to elucidate greater molecular insights such as historic signatures of selection (Qanbari et al., 2014), phenotypic variations within livestock breeds (Groenen et al., 2011) as well as linkage disequilibrium over short physical distances (Kijas et al., 2014). The availability and accessibility of comprehensive databases of genomic data for various uses has also facilitated population genetic studies globally, for example the National Centre for Biotechnology Information (NCBI) (Sayers et al., 2011), the Livestock Animal Quantitative Trait Loci database (Hu et al., 2013), and the University of California Santa Cruz (UCSC) genome browser (Dreszer et al., 2012).

The challenge now is to use these enhanced insights and understanding of molecular methods to develop breed-specific identification tools that are easily applicable to populations of livestock in different ecosystems. Such a breed-specific tool can be developed through identification and characterisation of unique phylogenomic SNPs in next-generation sequenced pooled genomic DNA from a selected representative small group of pure-bred individuals. The lessons derived from the ascertainment bias of genetic markers indicate that, for the purpose of conservation, it will be more suitable to develop a robust specific identification assay for each target breed rather than one assay for the identification of different breeds. This assertion does not discount the continued relevance of the use of common marker sets across multiple breeds or even species in other molecular studies such as investigation of QTL for economically important production and disease-resistant traits. However, for a target breed population the selection of breed-specific SNPs from neutral regions of the genome would guard against loss of efficiency of the SNP assay developed over time through direct selection or hitchhiking

effects. It is important to add that it is not always obvious which regions of the genome are under the influence of selection. In a recent study, an annotation of SNPs derived from the WGS Korean cattle breed using the bovine reference genome led to the suggestion that fixed, breed-specific SNPs might be useful for breed identification (Choi et al., 2015). That study described breed-specific fixation of many SNPs.

A growing number of software tools are being developed for the analysis of pooled-sequenced data (Kofler et al., 2016; Li & Durbin, 2010; Li et al., 2009). Read alignment of next-generation sequenced pooled genomic data to reference genomes has been achieved using the Burrows–Wheeler Tool (Kofler et al., 2011; Li & Durbin, 2010). The aligned reads are converted to a compatible pileup file format with SAMtools for subsequent analyses with the PoPoolation algorithm (Kofler et al., 2016; Li et al., 2009). Pooled-genomic sequenced data have been successfully analysed with PoPoolation accurately and efficiently to identify allele frequencies and population differentiation parameters of subject populations (Kofler et al., 2016; Kofler et al., 2011). Other analytical methods successfully applied include modified versions of popular genetic estimators such as the Watterson’s  $\theta$  and Tajima’s  $\pi$  analyses (Futschik & Schlötterer, 2010; Gautier & Naves, 2011).

These tools enable the efficient analyses of whole-genome sequenced SNP data sets such as mapping to appropriate reference genomes, SNP calling and annotation. Additional benefits of such data sets include use for the investigation of traits of economic and adaptive importance for the breed (Choi et al., 2015; Qanbari et al., 2014). Information from discovered phylogenomic SNPs can then be used to develop breed-specific SNP assays for the easy and precise identification of pure-bred members from mixed populations of breeds in various ecosystems. These tools will not only facilitate the timely diagnosis of the conservation status of livestock breeds, but will also permit the regular monitoring of endangered breed populations, particularly in developing countries where the lack of technical and financial capacity is reported to be a major impediment.

### **1.1.7 Conclusions**

- (1) Maintaining global domestic animal genetic diversity is important to human wellbeing.
- (2) Breed-specific molecular identification tools are urgently needed to allow the reliable and expeditious identification of individual members of any given breed; this is a pre-requisite for sustainable utilisation and conservation of any breed.

(3) A growing number of studies have established that whole-genome sequencing of pools of individuals within a group or breed provides a great deal of information on genetic variation across the whole genome even when performed at relatively low coverage, but also at considerably lower cost (Clarke et al., 2014; Gautier et al., 2013; Kim et al., 2010; Kofler et al., 2016). This will be a cost-effective technique for the identification of breed-specific phylogenomic SNPs within a target breed for the purposes of developing breed-specific molecular identification tools.

## **1.2 Combatting African Animal Trypanosomiasis (AAT) in livestock: The potential role of trypanotolerance**

### **1.2.1 Introduction**

African trypanosomiasis is a chronic debilitating disease caused by extracellular flagellate trypanosome protozoans (*Trypanosoma* species) and is spread mainly by the infected Tsetse fly vector (Diptera: Glossinidae) (Bruce, 1915; Brun et al., 2010; Hill et al., 2005; Mony & Matthews, 2015). The disease affects a wide range of mammalian species; including humans (Büscher et al.; Jha et al.; Matovu et al., 2001; Mwai et al., 2015; Seck et al., 2010). The three main trypanosome species endemic to Africa are *Trypanosoma vivax* (Dutonella) and *Trypanosoma congolense* (Nanomonas) that mainly infect livestock, and *Trypanosoma brucei* (Trypanozoon) which infects both humans and animals (Bezie et al., 2014; Kato et al., 2015; Nakayima et al., 2012). *T. brucei* has three sub-species of which two, *T. b. gambiense*, and *T. b. rhodesiense* infect humans, whereas the third, *T. b. brucei* infects domestic and wild animals (Munday, Settimo, & de Koning, 2015; Sima et al., 2011; Welburn et al., 2001; Welburn, Maudlin, & Simarro, 2009). The distribution of the tsetse fly (*Glossina* spp.) vector correlates closely with the prevalence of trypanosome parasites in the 10 million km<sup>2</sup> prevalence zone, thereby facilitating the entrenchment of the disease (Black & Seed, 2002; Stijlemans, De Baetselier, et al., 2017). Trypanosome parasites invade the host's lymphatic system, the blood circulation, and eventually migrate to the brain to cause a broad range of pathologies, most commonly severe anaemia, weight loss, abortion, and cachexia. It can kill the host if left untreated (Matovu et al., 2001; Sima et al., 2011; Stijlemans et al., 2015), and indeed millions of trypanosomiasis related deaths are recorded in livestock each year (Smetko et al., 2015). Other symptoms that have been reported for African animal trypanosomiasis (AAT) include

infertility, sleeping disorders, emaciation, pica, splenomegaly, paralysis, neuroendocrine dysfunctions and coma (Courtin et al., 2008; Steverding, 2008).

For several decades, AAT has been considered a neglected tropical disease and as such, remains endemic in 37 of 54 African countries. The affected areas cover approximately 10 million km<sup>2</sup> of arable land mass, and AAT has been shown to significantly reduce productivity in over 150 million cattle and 260 million sheep and goats (Baker, 1995; Jahnke et al., 1988; Leigh, Emikpe, & Ogunsola, 2014; Nyimba et al., 2015). AAT has a very significant combined economic and health burden in this sub-Saharan African (SSA) region (Habila et al., 2012; Namangala, 2012). The disease also has an additional impact on crop agriculture, human settlement and welfare, as approximately 7 million km<sup>2</sup> of the region is rendered unsuitable for mixed crop-livestock farming systems (Dagnachew et al., 2015; Peregrine, 1994). AAT is estimated to cause annual losses of more than US\$ 4.5 billion dollars through direct and indirect agricultural production losses (Dagnachew & Bezie, 2015; Leigh et al., 2014; Sanni et al., 2013). It is not surprising that as a direct consequence of the confounding effects of AAT on the general development of SSA, that 21 of the countries for which Trypanosomiasis is endemic are deemed to be amongst the world's 25 poorest countries (Shaw, 2009), with 32 considered highly indebted (IAEA, 2002).

Over several decades, the use of just a few therapeutic drugs for AAT that have limited efficacy against the parasites, but which are highly toxic to the host, has fueled the widespread emergence of drug resistance across the region (Delespaux & de Koning, 2007; Delespaux et al., 2008; Matovu et al., 2001). The continued lack of a suitable vaccine for the disease has also facilitated an over-reliance on these drugs (Tsegaye, Dagnachew, & Terefe, 2015). Furthermore, ongoing efforts directed at controlling the Tsetse fly vector in SSA have been largely ineffectual (Goossens et al., 1999; Hendrickx, de La Rocque, & Mattioli, 2004; Holmes, 1997; Torr & Vale, 2015). These factors, coupled with persistent political instability and armed conflicts have ensured that AAT persists across the region (Brun et al., 2010; Geerts et al., 2001).

Certain African livestock breeds such as Djallonke sheep and Taurine cattle, which entered Africa from the near east around 5000 BC and 7000 BC respectively, have developed innate tolerance to AAT, probably as a result of natural selection pressures (Dolan, 1987; Muigai & Hanotte, 2013; Naessens, 2006). The innate ability of these livestock breeds to survive and remain productive under AAT challenge, with very low mortality and without the use of trypanocidal drugs, is referred to as trypanotolerance. To illustrate the potential relevance of

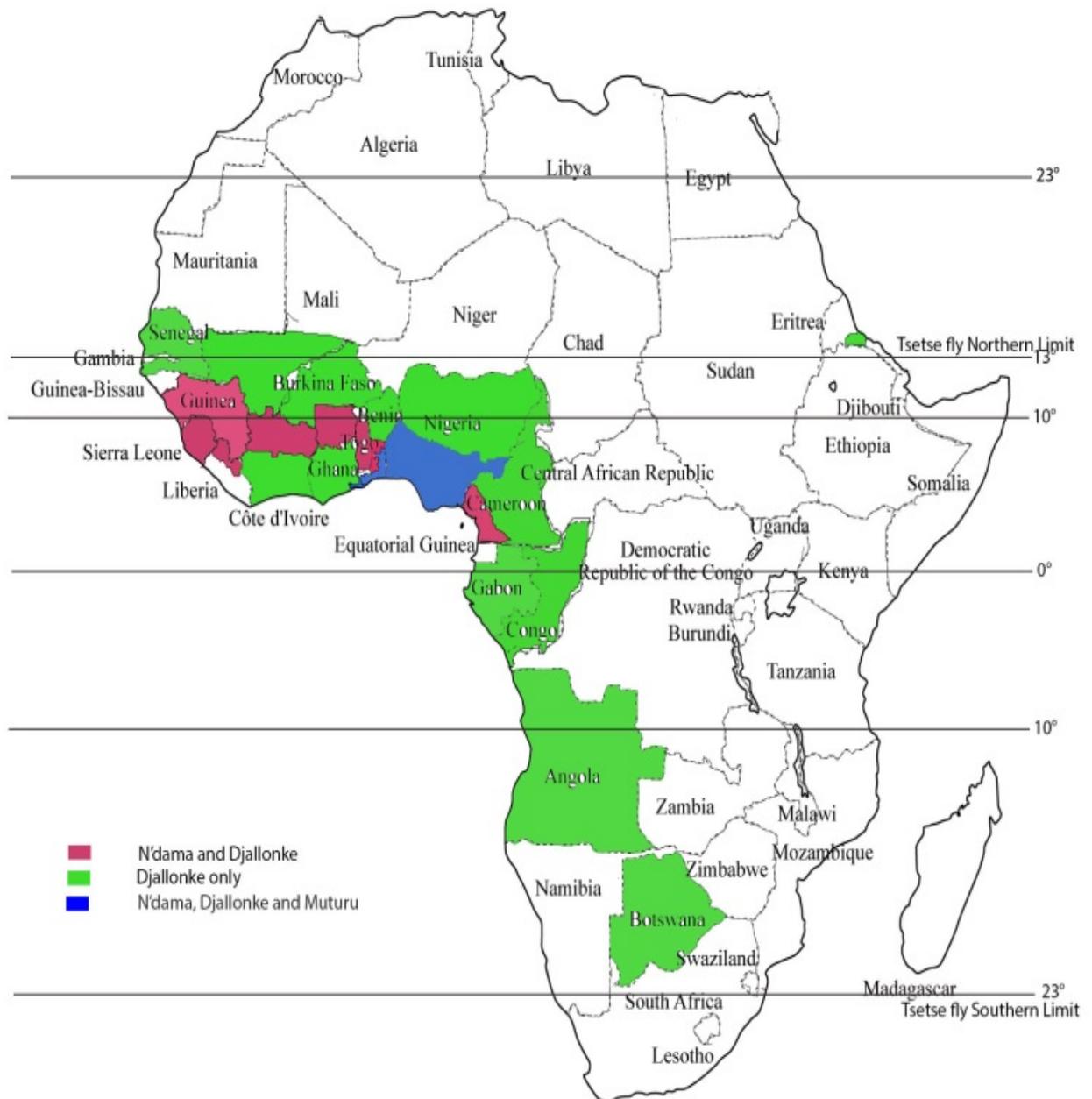
trypanotolerant breeds in SSA, figure 1 indicates the geographical distribution of three important trypanotolerant breeds of livestock within the Tsetse fly – AAT region of Africa. Trypanotolerance has been described as an economical and sustainable option for combating AAT (Geerts et al., 2009; Goossens et al., 1997; Murray, Morrison, & Whitelaw, 1982). If systematically implemented as a control strategy, utilizing trypanotolerance could have a major positive effect on long-term food security for the region (Osaer et al., 1994). In this review, we detail the key challenges remaining after a century of intervention against AAT, and the new insights on the genetics and mechanisms of trypanotolerance. Finally, we discuss the potential benefits of identifying the genetic mechanisms of trypanotolerance and harnessing livestock trypanotolerance through introgression within the context of SSA livestock systems.

### **1.2.2 A Century of Intervention against AAT**

The history of intervention programs for African trypanosomiasis involves the contributions of parasitologists, zoologists, entomologists, veterinarians and clinicians. However, with regards to AAT the landmark events are the findings of Bruce and Evans in the last decade of the 19<sup>th</sup> century (Bruce, 1915; Cox, 2004). Between 1891 and 1898, Evans identified *T. evansi* in equine spp. and Bruce identified *T. brucei* in cattle (Cox, 2004). In 1909, Bruce also identified the Tsetse fly as the vector transmitting trypanosome parasites (Bruce, 1915). These significant findings marked the beginning of attempts to combat AAT using a variety of measures, and the next 100 years was spent trying to eradicate this disease, with little success (Steverding, 2008). Throughout the 20<sup>th</sup> century, there were several attempts to control AAT through controlling the transmitting Tsetse fly vector. These control methods included; the sterile insect release technique, the destruction of fly habitat, the use of Tsetse traps, the use of insecticide-treated livestock, and coordinated mass spraying of insecticide (Doyle, Moloo, & Borowy, 1984; Hendrickx et al., 2004; Hill et al., 2005; Holmes, 1997; Torr & Vale, 2015). These interventions yielded limited positive outcomes against AAT, but have often been associated with negative environmental consequences including insecticide pollution of water bodies and deforestation (Goossens et al., 1999; Hendrickx et al., 2004; Holmes, 1997; Torr & Vale, 2015). Other attempts at curbing AAT through targeting the parasite using anti-AAT drugs have not produced the expected results due to the rapid development of trypanocidal resistance (Alsford et al., 2013; Kaufmann et al., 1992). In 2008, 17 SSA countries reported veterinary trypanocidal drug resistance problems, and by early 2015 this number had risen to 21 countries (Delespoux et al., 2008; Tsegaye et al., 2015). This resistance was expedited in part by the

reliance on predominantly three drugs for treating AAT over 50 years (Delespaux et al., 2008; Geerts et al., 2001; Munday et al., 2015; Peregrine & Mamman, 1993). Development of resistance and cross-resistance of trypanosomes to these drugs was further expedited as these drugs have similar chemical compositions (Peregrine, 1994). Furthermore, these three AAT drugs have high host toxicity, and have shown limited efficacy (Matovu et al., 2001; Peregrine & Mamman, 1993; Steverding, 2015). Other factors contributing towards drug resistance include the high degree of re-infection rates among treated livestock, and significant levels of misuse of trypanocides by farmers as a consequence of the deregulation and privatization of veterinary services (Geerts et al., 2001). In 2008, a report indicated that, out of an estimated 35 million doses of veterinary trypanocidal drugs administered, diminazene aceturate, isomethamidium chloride and ethidium bromide accounted for 33%, 40% and 26% respectively (Delespaux et al., 2008). Peregrine and Mamman (1993) reviewed the causes and mechanisms for parasite resistance to each of the drugs used against AAT. Drug resistance continues to interfere with effective therapeutic management of AAT, and is reported to be responsible for many widespread outbreaks of Trypanosomiasis, in different parts of the region, that did not respond to standard chemotherapeutic regimens (Holmes, 1997; Mamoudou et al., 2008).

The hope of developing an effective vaccine based on the surface glycoprotein antigens of trypanosomes remains particularly elusive due to the complexity of the parasite's antigenic repertoire (Hill et al., 2005; McCulloch, Rudenko, & Borst, 1997). Horn (2014), Manna et al. (2014) and Taylor and Rudenko (2006) have provided comprehensive reviews on the mechanisms of trypanosome antigenic variation. Although an increased understanding of the structure and mechanism of this antigenic variation of trypanosome parasites has occurred over the past 40 years, to date no vaccine is available (Manna et al., 2014; McCulloch & Field, 2015; Mony & Matthews, 2015; Stijlemans, De Baetselier, et al., 2017). These factors are the main reasons why most research efforts aimed at developing a vaccine for trypanosomiasis have since shifted from variable surface (VSG) antigen towards the identification of other, invariant structural components of the parasite (Alsford et al., 2013; Taylor, 1998; Tsegaye et al., 2015).



**Figure 1.1 Distribution of three economically important trypanotolerant livestock breeds within the African animal trypanosomiasis-tsetse fly endemic region of Africa**

Towards the late 1990s, the recognition of the systematic failure of existing Trypanosomiasis control methods led African scientists to set up a regionally coordinated initiative to tackle the disease (Maudlin, 2006). In 2000, the Pan-Africa Tsetse and Trypanosomiasis Eradication Campaign (PATTEC), was endorsed at the 37<sup>th</sup> African Union summit in Togo (Kabayo, 2002). Central to that new initiative was a “Money Map” of the region that shows the anticipated

benefit of eradication of trypanosomiasis in US\$ per a Km<sup>2</sup> of land area over 20 years to illustrate the magnitude of economic loss caused by the disease (Shaw, 2009). PATTEC received support from multinational agencies including the World Health Organisation, the Food and Agriculture Organisation and the International Atomic Energy Agency. Unprecedented support in the fight against Trypanosomiasis also came from a long list of non-governmental organisations and pharmaceutical companies including the Bill and Melinda Gates Foundation, Bayer, Bristol-Mayers Squibb, and Aventis Pharma (WHO, 2002). PATTEC, although stipulating a holistic and integrated approach to combatting trypanosomiasis, was criticised for overly focusing on Tsetse fly eradication at the expense of other methods such as the use of trypanotolerant breeds (IAEA, 2002). Furthermore, the Tsetse fly control module employed in the program was criticized as unsuitable and unrealistic for application to the 10 million km<sup>2</sup> affected area of SSA, although it was used successfully in the isolated and small region of Zanzibar (about 2650 km<sup>2</sup>) (IAEA, 2002; Kabayo, 2002).

The reality is, that despite all these different control measures targeted at AAT, the disease still persists, and is not likely to be completely eradicated soon (Dolan, 1987; Magez & Radwanska, 2014; Nakayima et al., 2012). Therefore, exploiting other measures such as natural trypanotolerance of livestock is imperative, particularly for the vast majority of resource-poor smallholder livestock keepers in the region for whom the existing chemotherapy and vector control programs are neither accessible nor affordable (Cooper et al., 2008).

### **1.2.3 Genetic Basis for trypanotolerance**

Trypanotolerance, an inherent ability in certain breeds of livestock to withstand trypanosomiasis, was first reported as early as 1904 and 1913 in West and East Africa respectively (Bruce, 1915; Dolan, 1987; Murray et al., 1982). A growing number of studies have suggested the use of trypanotolerance as an economically sustainable option for tackling AAT (Mattioli et al., 2000; Murray et al., 1982; Naessens, 2006). Despite this, the extent, the mechanisms and the effects of trypanotolerance as a control measure for the disease have remained largely unexplored and poorly understood (Dolan, 1987; Kosgey & Okeyo, 2007). Research findings from experimental and naturally infected livestock populations have confirmed that trypanotolerance is an heritable trait in breeds such as N'Dama cattle (Hanotte et al., 2003; Murray & Black, 1985; Naessens et al., 2002; Namangala, 2012; Trail et al., 1991), Muturu and Baoule cattle (Naessens et al., 2002), Djallonke sheep and the west African dwarf goat (Geerts et al., 2009; Osaer et al., 1994). For example, in a natural Trypanosomiasis

challenge study using N'Dama cattle in Gabon, an estimated heritability of 0.64 was observed for the trait controlling anaemia, as measured by the average packed red blood cell volume, a characteristic feature of trypanotolerance (Trail et al., 1991). Different crosses of trypanotolerant and trypanosusceptible breeds in both cattle and sheep have produced phenotypes with varying intermediate degrees of tolerance to the disease, indicating a complex and possibly varied genetic control mechanism (Goossens et al., 1997; Goossens et al., 1998; Goossens et al., 1999; Murray & Trail, 1984). For example, crosses between trypanotolerant Djallonke sheep and trypanosusceptible Sahelian sheep (Goossens et al., 1999), and trypanotolerant N'Dama and trypanosusceptible Boran cattle (Hill et al., 2005) have been studied. Although many previous authors have used the words resistance to Trypanosomiasis and Trypanotolerance interchangeably (Iraqi et al., 2000; Murray & Black, 1985; Ogunremi & Tabel, 1995), in the strict sense of the word, the concept of tolerance differs from the concept of complete resistance to diseases. Similar to nematode infection in small ruminants (Bishop, 2012), tolerance is more equivalent to resilience or relative resistance, and depicts an ability of the livestock to survive and maintain productivity in the face of the disease challenge. Geerts et al. (2009), however, argued for the use of resistance for trypanotolerant N'Dama cattle because the high frequency of natural "self-cure" from the infection in the breed. The Cape Buffalo (*Syncerus caffer*) is the only ruminant species in the region that is known to have an absolute resistance to infection from all species of trypanosomes (Namangala, 2012). However, the Cape Buffalo with an estimated population of only one million, has been shown to be a reservoir for the spread of many other important livestock diseases in the region such as bovine tuberculosis, brucellosis, anthrax and foot and mouth disease (Michel & Bengis, 2012). Hence, these features of the Cape Buffalo make it unattractive to be considered in the mix of control interventions for AAT.

Although there has been no absolute determination of the genetic mode of inheritance of trypanotolerance in cattle and sheep, several studies in cattle and even mice have suggested that only a small number of genes of major effect are involved combined with many other genes having only small effects (Kemp & Teale, 1998; Ogunremi & Tabel, 1995). A recent study using a trypanotolerant breed (N'Dama) crossed with a susceptible breed (Boran) showed that females were more trypanotolerant than males and that the F1 crosses had an intermediate response compared to the resistant N'Dama and the susceptible Boran breeds (Orange et al., 2012). This is suggestive of the possibility of several fixed alleles being present in the trypanotolerant breeds.

Genetic investigations using microsatellites (Dayo et al., 2009; Hanotte et al., 2003), and single nucleotide polymorphisms (SNPs) (Gautier et al., 2009) have identified numerous quantitative trait loci (QTL) target regions for trypanotolerance in N'Dama cattle. A follow-up investigation within these identified regions using a denser panel of microsatellites identified a polymorphic allele in one of the previously identified candidate regions that was associated with anaemia control (Dayo et al., 2012). Another QTL study using advanced intercross mouse lines showed the presence of 14 QTL markers, however, these effectively only spanned three chromosomes (1, 5 and 17) with peaks observed in 5 regions (one on chromosome 17, one on chromosome 5 and three on chromosome 1) (Iraqi et al., 2000).

Table 1.2 shows the current status of efforts to identify QTL for trypanotolerance in African cattle breeds and experimental mice. Conspicuously missing in the list is trypanotolerant QTL investigation in small ruminants (i.e. sheep and goats). Given the huge economic importance of small ruminants in the SSA region (Geerts et al., 2009; Kosgey et al., 2006; Kosgey & Okeyo, 2007), it is surprising that there are no trypanotolerance QTL studies recorded in trypanotolerant sheep and goat breeds. Although many of these early findings provided novel molecular genetic insights into trypanotolerance, the size of the resolved QTL regions were still quite large (within the range of 10 to 40 centiMorgan (cM) i.e. approximately 10 to 40 million base pairs) for their general application in a structured introgression programme using selection markers (Georges, 2007; Iraqi et al., 2000; Noyes et al., 2011). Furthermore, the identified QTLs in some of these studies were only able to explain a small fraction of the observed genetic heritability (Courtin et al., 2008; Georges, 2007). Therefore, most of the early trypanotolerance QTL studies require further refinement mapping to locate the precise causal mutations or quantitative trait nucleotides (QTN) so as to increase their adaptation into current breeding practices. Meuwissen, Hayes, and Goddard (2016) provided a comprehensive review on the limitation of the previous traditional QTL studies in livestock for application in genomic selection. Prior to these studies, a number of studies using transcriptome analyses also suggested the involvement of large number of genes in trypanotolerance in N'Dama cattle (Berthier et al., 2003; Maillard et al., 2005), in mice (Fisher et al., 2007), and in N'Dama and Boran cattle (O'Gorman et al., 2009). However, a major multiple genomic study that combined the analyses of transcriptomes, single nucleotide polymorphisms and expressed sequence tag libraries within a previously identified trypanotolerant QTL region reported only two candidate genes (Noyes et al., 2011). These trypanotolerance QTL studies have typically used linkage-map based units of cM, however, these will need to be converted to actual genomic coordinates for them to be applicable from a current genomic typing perspective.

So far there are no trypanotolerance studies which have used a whole genome, next generation sequencing approach for the detection of putative loci for trypanotolerance. The identification of putative genomic variation responsible for trypanotolerance in livestock will pave the way for the widespread application of genomic selection (Hayes & Goddard, 2010). Recently, it has been shown that the use of next generation sequenced pools of individual genomes could be a good method for providing high resolution power for identification of candidate genes or variation signatures of adaptive evolution than the use of microsatellite markers or SNP chips (Bergland et al., 2014; Kofler et al., 2016).

**Table 1.2 Present status of efforts to identify quantitative trait loci (QTL) for trypanotolerance in African cattle breeds and experimental mice**

<b>Species</b>	<b>Molecular tool</b>	<b>Main finding</b>	<b>Reference</b>
<b>N'Dama and Boran</b>	477 Microsatellite markers	19 QTL	(Hanotte et al., 2003)
<b>N'Dama cattle</b>	Transcriptomics –Serial analysis of gene expression (SAGE)	187 genes (up- and down-regulated)	(Berthier et al., 2003)
<b>N'Dama cattle</b>	Transcriptomics -SAGE	82 genes (up- and down-regulated)	(Maillard et al., 2005)
<b>Mice</b>	Microarray	32 candidate genes	(Fisher et al., 2007)
<b>9 West African cattle breeds</b>	36,320 SNPs (single nucleotide polymorphisms)	53 genomic regions (42 genes)	(Gautier et al., 2009)
<b>N'Dama and Boran cattle</b>	Bovine Long Oligonucleotide Microarray	32 genes	(O'Gorman et al., 2009)
<b>N'Dama and Boran cattle</b>	92 Microsatellite Markers	2 trypanotolerance-related Outlier Loci	(Dayo et al., 2009)
<b>Taurine and Zebu cattle</b>	64 Microsatellite Markers	One QTL of trypanotolerance-associated trait	(Dayo et al., 2012)
<b>BALB/c and CcS-11 mice</b>	Microsatellite markers	4 QTL of trypanosusceptibility	(Sima et al., 2011)
<b>N'Dama cattle and Boran cattle</b>	Multiple genomic tools; Transcriptomics, 21,034 SNPs- and expressed sequence tags- library sequencing of previously identified trypanotolerant QTL	2 candidate genes	(Noyes et al., 2011)

For example, a re-sequenced pool of the whole genomes of 50 to 100 *Drosophila melanogaster* precisely identified seasonally associated polymorphisms in the flies (Bergland et al., 2014). Similarly, the trypanotolerant breeds have co-evolved with the trypanosome parasites over several millennia (Murray & Black, 1985; Mwai et al., 2015), it is therefore expected that signatures of selection for resistance should exist in their genomes. Consequently, a whole genome sequencing approach, because of its significantly higher resolution, may be a suitable way to identify the actual causative genomic variations for trypanotolerance. A recent study, for instance, has shown that a whole-genome re-sequencing and analysis of a small group of cattle revealed genomic selective sweeps harboring domestication related candidate genes such as the KIT, MITF and MC1R genes (Qanbari et al., 2014). This result was obtained by using two complementary statistical tools; the integrated haplotype homozygosity score (for detecting ongoing selection) and composite of likelihood ratio (for detecting completed sweeps and fixation of advantageous alleles) to scan the approximately 15 million SNPs generated from the whole genome sequence data. More recently, a whole genome sequenced dataset (17 million SNPs) analysis for selective sweeps in a native Korean beef cattle breed was also accomplished using a Z-score of the pooled heterozygosity statistic (Choi et al., 2015). Subsequently, the annotation of the identified non-synonymous SNPs using the database for annotation visualization and integrated discovery (DAVID), (Huang da, Sherman, & Lempicki, 2009)) and the animal QTL database has led to the precise identification of the PPP1R12A gene. This gene is known to be associated with intramuscular fat and was localized in one of the several sweep regions (Choi et al., 2015).

It is apparent that the past QTL studies for trypanotolerance in livestock have not provided adequate genomic information necessary for important applications such as marker assisted selective breeding. However, recent advances in genomic sequencing and computational bioinformatics have produced tools with higher power, capable of resolving more refined trypanotolerance QTL (Kofler et al., 2016; Li & Durbin, 2010; Li et al., 2009; Rubin et al., 2010). An improved strategy to identify putative trypanotolerance loci could be achieved through the explorative analysis of next generation sequenced pools of genomic DNA from small representative group each of trypanotolerant (e.g. Djallonke sheep) and trypanosusceptible (e.g. Sahelian sheep) livestock breeds sampled from the same geographic region. The pooled-next generation sequencing method has been shown to be not only more powerful, but also a more cost-effective approach than the traditional Genome-wide association study (GWAS) approach (Gautier et al., 2013; Schlotterer et al., 2014). It is important to ensure that each sampled group consists of unrelated individuals in order to minimize false prediction

due to relatedness. Given that the trypanotolerance trait most likely resulted from natural selection over millennia, it is expected that detectable selection signatures (marked by regions of homozygosity) will be present in the genome of trypanotolerant breeds. Currently, it has been shown that genome wide selection signatures can be accurately identified from whole genome sequenced data using robust systematic analytical tools (Choi et al., 2015; Kardos et al., 2015; Qanbari et al., 2014). Furthermore, there are many algorithms available for controlling false discovery rates in the analysis of whole genome datasets for selection, and have recently been comprehensively reviewed (Francois et al., 2016).

Collectively, trypanotolerance appears to be a complex multi-genic trait, and is possibly controlled by many QTL across the genome (Berthier et al., 2003; Fisher et al., 2007; Gautier et al., 2009; Hanotte et al., 2003; Maillard et al., 2005; Noyes et al., 2011). Currently, a number of precision bioinformatics tools are capable of identifying unique genome-wide mutations such as those expected to be present in trypanotolerant breeds, but absent in trypanosusceptible breeds and the species reference genome. These tools include the genome analysis tool kit (GATK), SAMtools, and the Burrows-Wheeler Aligner (BWA), for example (Choi et al., 2015; Li et al., 2009). A subsequent functional annotation and analysis of prospective trypanotolerant candidate mutations can be facilitated by utilizing the free online resources such as the animal QTL database (Hu et al., 2013), a repository of curated livestock genomic information, the snpEff program and the DAVID tool (Choi et al., 2015; Huang da et al., 2009). These tools constitute only a fraction of the number of new genomic and bioinformatics procedures currently available by an ongoing relentless march of this rapidly evolving field. As has been accomplished for some other important mammalian traits, we are of the opinion that these methodologies may provide answers to some of the key research questions relating to this important neglected tropical disease.

#### **1.2.4 Mechanism of trypanosome infection and trypanotolerance**

The hallmark of trypanosomiasis pathology is the remarkable ability of trypanosome parasites to elicit responses from both the host innate and acquired immune systems (Magez & Radwanska, 2014; Mansfield, Paulnock, & Hedberg, 2014; Naessens et al., 2002). Trypanosome species are extracellular parasites with more than 1,000 different VSG genes and pseudogenes, of which only one is transcribed at any one time. These numerous VSG genes and pseudogenes allow a trypanosome to produce successive waves of up to  $10^7$  different VSG antigens, and this mechanism allows sub-populations of the parasite to evade any humoral

adaptive immunity of the host (Cnops, De Trez, et al., 2015; Donelson, 2003; Manna et al., 2014; Mansfield et al., 2014; Taylor & Rudenko, 2006). This classical immune escape mechanism makes trypanosome species infection difficult to control once established (Cnops, Magez, & De Trez, 2015; Donelson, 2003; Taylor & Rudenko, 2006). Matthews, McCulloch, and Morrison (2015) provided a comprehensive review of the within-host dynamics that facilitate trypanosome species infection including a novel quorum sensing mechanism. Also, the variety of trypanosome immune escape mechanisms that aid the chronicity of trypanosomiasis has also been recently reviewed (Cnops, Magez, et al., 2015). More recently, evidence of the role of hybridization of genetically divergent trypanosome populations in compounding the virulence, pathogenicity as well as drug resistance of trypanosomes has been reported in African livestock (Tihon et al., 2017).

Earlier reports have suggested that the host's adaptive immunity in the form of VSG-specific B- and T- lymphocytes was mainly responsible for trypanotolerance in trypanotolerant breeds (Taylor, 1998). However, more recent evidence points to the host's innate immunity in the form of activated macrophages as key to trypanotolerance (Liu et al., 2015; Mansfield et al., 2014). Recently, it has been shown that the adaptive immune response to trypanosome species infection is short-lived and is effective only against a sub-population of the possible VSG types (Cnops, De Trez, et al., 2015; Magez & Radwanska, 2014). The report also indicates that effective immunological memory is not developed by the host during trypanosome infection, and hence the successive waves of infection destroy the host's B cell compartment, leading to the failure of adaptive immunity (Magez & Radwanska, 2014; Stijlemans, Radwanska, et al., 2017). The failure of the host adaptive immunity is marked haematologically by dramatically increased levels of parasitaemia and concomitant anaemia, which is progressively followed by several of the pathological symptoms previously mentioned in this review. These pathological symptoms, if left untreated, may culminate in the death of the host (Goossens et al., 1998). Conversely, trypanotolerant breeds exhibit a capacity to control this characteristic anaemia and parasitaemia that accompanies the trypanosome infection (Goossens et al., 1998; Murray et al., 1982; Murray et al., 1984). This capacity to control anaemia is crucial for trypanotolerance, and permits the host to remain productive under disease challenge (Naessens, 2006; Naessens et al., 2002; Trail et al., 1990; Trail et al., 1991). Generally, trypanotolerant livestock thrive better with low to medium intensity challenge than with high intensity parasite challenge (Holmes, 1997). In parts of west and central Africa, where the AAT challenge is very high, Diminazene chemotherapy has been used to help trypanotolerant breeds maintain desirable levels of production (Peregrine & Mamman, 1993). The high intensity of the disease challenge

in these parts of Africa has completely excluded the farming of trypanosusceptible breeds, given the enormous cost of the trypanocides that would be required (Peregrine & Mamman, 1993).

Similar to resistance to other parasitic infections of ruminants (Coop & Holmes, 1996; Wallace et al., 1995), trypanotolerance has also been found to be enhanced in hosts with a high plane of nutrition and *vice versa* (Coop & Kyriazakis, 1999, 2001; Cunningham-Rundles, McNeeley, & Moon, 2005; Van Houtert & Sykes, 1996). Conversely, the presence of inter-current parasitic infection in the host reduces trypanotolerance, as do physiological stress factors such as gestation (Coop & Kyriazakis, 1999; Murray et al., 1982; Murray et al., 1984). The deleterious effect of mixed infections of AAT and other parasites, particularly gastrointestinal helminths, in trypanotolerant breeds have been very well documented (Goossens et al., 1997; Goossens et al., 1999; Kaufmann et al., 1992; Okaiyeto et al., 2010). In Djallonke sheep, an untreated mixed infection is generally characterised haematologically by a rapid fall in packed red blood cell volume, high levels of eosinophils, immunosuppression, weight loss and mortality (Goossens et al., 1999; Okaiyeto et al., 2010). A similar situation occurs in N'Dama cattle (Kaufmann et al., 1992). In an experimental infection of Djallonke sheep, more severe acute symptoms were observed when infection with trypanosomes was followed by concurrent *Haemonchus* infection (Goossens et al., 1997). Conversely, a more chronic form of the disease was observed when the sequence of infection was reversed (Goossens et al., 1997; Kaufmann et al., 1992). This observation suggests that AAT infection has an immunosuppressive effect on the Djallonke sheep, and renders it more susceptible to subsequent infection by *Haemonchus*. In an another study, abrogation of immunity to *Heligosomoides polygyrus* in previously immunized mice was attributed to the immunosuppressive effect of a concurrent trypanosome infection (Fakae et al., 1997). This phenomenon of the host's immune response to one type of parasite, compromising its ability to mount an appropriate immune response against a secondary infection has been referred to as negative feedback (Jolles, Beechler, & Dolan, 2015).

Recent studies have suggested that the associated immunosuppression is due to the extensive apoptosis of the splenic B cell compartment in the host during trypanosome species infection (Cnops, De Trez, et al., 2015; Magez & Radwanska, 2014; Radwanska et al., 2008). The mixed infection phenomenon presents an important obstacle to exploring trypanotolerance in most livestock production systems in SSA (Alvarez, Traore, et al., 2012). However, trypanotolerant livestock also exhibit relative resistance to other important livestock diseases such as helminthiasis, anaplasmosis, babesiosis and heartwater (Murray et al., 1982; Murray et al.,

1984; Tano et al., 2003). A recent report linked the resistance to nematode infection trait in the Caribbean hair sheep to historical introgression of the trypanotolerant Djallonke genetics (Spangler et al., 2017). Attempts to achieve genetic introgression of trypanotolerance through indiscriminate crossbreeding with trypanosusceptible breeds is common in livestock production systems in many parts of SSA, but leads to the dilution of the trait (Alvarez, Traore, et al., 2012; Bradley et al., 1994; Geerts et al., 2009; Kosgey & Okeyo, 2007). Morris (2007) reports that a considerable number of trypanotolerant N'Dama and western African shorthorn cattle have already been introduced into 19 countries within the central African region in response to the disease challenge. The use of trypanotolerant breeds in the mainly trypanosomiasis endemic regions of SSA (where typically the production of trypanosusceptible animals is already limited by the disease prevalence) has minimal epidemiological implications in terms of the danger of becoming a reservoir for the disease to non-tolerant livestock. In the hypothetical scenario, where both tolerant and susceptible breeds co-exist in a trypanosomiasis endemic area, it is natural that both will be exposed to the same risk of infection through the indiscriminate bites of infected tsetse fly vectors. While it is still alive, an infected trypanosusceptible animal presents the same risk to a trypanotolerant breed as does the reverse situation. This is because it will only require an infected tsetse fly vector to transmit the disease from an infected trypanosusceptible animal to a non-infected trypanotolerant animal and vice versa. The most likely outcome is that the trypanotolerant animal will outlive the trypanosusceptible animal due to its inherent ability to withstand the disease. This explains the reason why resource-poor livestock keepers within this trypanosomiasis endemic region, in their quest to prevent the loss of their trypanosusceptible breeds, have rationally resorted to crossbreeding with the trypanotolerant breeds. It is important that to note that these farmers conduct the crossbreeding without the specific knowledge of the trypanotolerant QTL or genes (Berthier et al., 2016). The desire to keep trypanosusceptible breeds in a trypanosomiasis endemic region is also partly due to their inherently bigger body sizes, for example the trypanosusceptible Sahelian sheep is much bigger than the trypanotolerant Djallonke sheep (Yaro, Ayizanga, & Abdul-Rahman, 2012). But, the larger body size of trypanotolerant breeds does not directly translate to higher productivity under trypanosomiasis challenge. It has been reported that under the condition of AAT challenge, the trypanotolerant Djallonke sheep recorded higher productivity of approximately 42% over trypanosusceptible sheep breeds (Geerts et al., 2009). More recently, a microsatellite and SNP marker analysis within trypanosomiasis candidate regions of trypanotolerant and trypanosusceptible cattle confirmed significant levels of admixture of the two breeds (Smetko et al., 2015). Most of these

indigenous African trypanotolerant livestock breeds may now be endangered due to indiscriminate crossbreeding and breed replacement, and there is a high risk of these adaptive traits being diluted and lost forever (Mwai et al., 2015). This excessive genetic introgression poses a major challenge to the sustainable use of the trypanotolerant trait for the control of AAT if deliberate programmes for the preservation and development of trypanotolerant breeds are not first developed. However, when relocating trypanotolerant breeds to trypanosomiasis-free regions of the continent, or any part of the world, standard screening and quarantine precautions need to be applied to ensure that the animals are not harbouring any trypanosomes. If this precaution is not observed, there will be some risk of introducing the disease to these regions. However, under this scenario the risk will still be minor because AAT is mainly transmitted by the tsetse fly vector, the distribution for which is known to be confined to only the trypanosomiasis endemic region.

### **1.2.5 Trypanotolerance in the context of Africa livestock production**

The existing AAT control and eradication programs have not, and might never, reach the level of effective implementation required to rid the region of this complex disease. This is because the main livestock systems in SSA are smallholder pastoral, agro-pastoral, mixed crop–livestock and peri-urban livestock systems (Kosgey & Okeyo, 2007; Tano et al., 2003; Zougmore, Traoré, & Mbodj, 2015). These systems are characterised by low input, low technology, and with a considerable level of subsistence production (Jahnke & Jahnke, 1982). There is a general lack of capacity for local, mostly smallholder, livestock farmers to diagnose or treat such a complex disease (Geerts et al., 2001; Murray et al., 1984).

Over the last century, the sheer complexity and diversity of socio-politico-cultural elements of the different countries within the SSA region have constrained the effective coordination of many anti-trypanosomiasis programs between member states, and also with regional and international partners (Black et al., 1985; Dzingirai et al., 2017; Ford, 2007; Smith, Taylor, & Kingsley, 2015). The persistence of civil conflicts and wars within the SSA region has been directly linked to the re-emergence of the disease in many affected countries (Ford, 2007). For example; the trypanosomiasis epidemic outbreak of Uganda in the late 1970's resulted from political turbulence, and the outbreak in Angola that started in 1975 during the post-independence civil war period. The presence of large populations of wildlife reservoirs of trypanosome parasite in some parts of the region has also worked against the total eradication of the disease (Murray et al., 1982).

The appropriate use of trypanotolerant breeds would mitigate the high annual losses incurred by the vast majority of these smallholder livestock farmers (Tano et al., 2003). Mitigating losses is of particular importance because the improvement of smallholder agricultural productivity in Africa is fundamental to overcoming the problem of poverty in the region (Babikir et al., 2015). Exploiting host genetics to help control disease is a widely accepted concept in developed countries (Piedrafita et al., 2010; Raadsma & Fullard, 2006), and host resistance has been extensively studied in breeds such as the Scottish Blackface sheep (Stear et al., 1997). There are ongoing extensive and equally successful breeding programs for disease resistance in Australian Brahman cattle (Frisch, O'Neill, & Kelly, 2000), and in sheep in Australia and New Zealand (Van der Werf, 2007). For example, AUD\$ 8 billion in extra earnings was realized within 30 years through the exploitation of parasite resistance genes of Brahman cattle via a landmark cross-breeding program within the Northern Australian beef industry (Frisch et al., 2000; Morris, 2007). Most research and application has involved parasitic diseases, but Bishop and MacKenzie (2003) provide a comprehensive model framework for the utilization of disease resistance for the control of bacterial and viral infections in livestock.

In the context of SSA livestock production systems, measures would need to be put in place to preserve the genetic purity of known trypanotolerant breeds, and to protect those breeds from excessive genetic introgression by trypanosusceptible breeds (Geerts et al., 2009). We propose that a molecular characterisation of all trypanotolerant livestock breeds for the purpose of reliable breed identification will also expedite future genetic improvement programs for trypanotolerant breeds. This effort can take the form of establishment and coordination of regionally or nationally supported satellite open nucleus breeding stations for pure trypanotolerant breeds across the SSA region to sustain this intervention.

### **1.2.6 Conclusions**

After a century of intervention against AAT, the disease persists, as does its staggering impact on the livelihoods of the population in SSA. A diverse range of interconnected factors has contributed to the entrenchment of the disease. Given the current perspective, imminent eradication of this disease does not seem a possibility. Although advances in scientific technologies have been accompanied by a greater understanding of the mechanisms of the disease over this period, the existing control measures remain largely inadequate. We are of the view that an improved outcome in the battle against AAT will require a more holistic

approach that is dynamic and context-specific to the different livestock production systems across SSA. Therefore, management of the disease including the development of structured programs for the use of trypanotolerant breeds will be a more realistic and achievable objective. The exploitation of naturally trypanotolerant breeds of livestock will not only add an economically sustainable option to the mix of interventions, but is also compatible with livestock production systems in the region.

### **1.3 The key role Class II Major histocompatibility complex in parasitic infection of domestic sheep**

#### **1.3.1 Introduction**

The major histocompatibility complex (MHC) is a gene dense, and the most polymorphic, region in the vertebrate genome (Hedrick, 1998; Klein et al., 1993). The MHC was first discovered in the 1930's, and is regarded as being central to the control of both innate and adaptive immune response in vertebrates, as well as mate choice (Horton et al., 2004). Although the MHC also contains non-functional genes, the majority of studies focus on the functionally transcribed MHC genes for obvious reasons (Robinson et al., 2003). The human MHC, also referred to as the human leucocyte antigen (HLA) is the most extensively studied of all the mammalian species (Parham, Adams, & Arnett, 1995). The HLA complex is located on chromosome 6, and is comprised of telomeric, centromeric and central regions referred to as the class I, II and III genes, respectively (Hughes & Yeager, 1998). A standard nomenclature for the HLA genes has been established (Robinson et al., 2015) and all identified sequences are submitted to a dedicated database available at <http://www.ebi.ac.uk/ipd/imgt/hla/>. The size of HLA is approximately 4 Megabases long, and contains at least 220 genes (Robinson et al., 2003). The fundamental MHC structure is largely conserved in mammalian genomes (Dukkipati et al., 2006). Uniquely, the centromeric class II region in ruminant species is subdivided into the class IIa and IIb regions, which are separated by 18.5MB of non-MHC genes (Amills et al., 1998; Gao et al., 2010; Siva Subramaniam et al., 2015). It was suggested that an ancestral chromosomal inversion that predates the speciation of Bovidae might be the cause of the unique MHC structure in ruminants (Gao et al., 2010). The sheep MHC is located on chromosome 20 (MÄKinen et al., 1989), and that of cattle is located on chromosome 23 (conserved synteny to the sheep chromosome 20) (Burt, 2009). Nomenclature for sheep MHC alleles and selected domestic animals has also been established, and is hosted at the same online database (<https://www.ebi.ac.uk/ipd/mhc/>) as the HLA alleles. In sheep, the MHC is

also referred to as the OLA or Ovar-MHC (Dukkipati et al., 2006), and was first reported about four decades ago (Millot, 1978). However, until the beginning of this millennium, the structure of Ovar-MHC was poorly characterised (Amills et al., 1998). In recent years, the structural organisation has received growing attention, with the publication of the genomic structure of the class I (Siva Subramaniam et al., 2015), class II (Lee et al., 2011), class III (Qin et al., 2008) as well the complete Ovar-MHC region (Gao et al., 2010). The information provided by these publications heralded greater understanding of the genomic organisation of the Ovar-MHC. For example, a total of 177 genes have now been identified in the Ovar-MHC, spanning approximately 2.4 mega bases (Gao et al., 2010).

### **1.3.2 Functions of the MHC class II genes**

In terms of functionality, the vertebrate MHC genes control the mechanisms that determine the resistance or susceptibility to infection by the host (Gutierrez-Espeleta et al., 2001). In Sheep, many studies have associated variations at the MHC to gastrointestinal nematode infection (Davies et al., 2006; Hickford et al., 2011; Sayers et al., 2005), footrot (Ennen et al., 2009; Escayg, Hickford, & Bullock, 1997; Gelasakis et al., 2013), and other less economically important livestock viral infections (Konnai, Takeshima, et al., 2003; Nagaoka et al., 1999). The telomeric class I and centromeric class II of the MHC are reported as the most important regions with respect to response to infection (Dukkipati et al., 2006). However, most of the association studies were linked to variations within the class II region (Bryja et al., 2006; Hassan et al., 2011; Meyer-Lucht & Sommer, 2005; Richman, Herrera, & Nash, 2001). The class II genes in sheep are known to code for specialised heterodimeric glycoprotein molecules on the surface of four different immune cells. The four immune cells include; dendrites, macrophages, B lymphocytes and thymus glands (Herrmann-Hoesing et al., 2008). The MHC class II molecules are capable of binding and presenting a wide range of endogenous and exogenous antigens to CD4<sup>+</sup> T lymphocytes (Ballingall et al., 2011; Herrmann-Hoesing et al., 2008). This ability is primarily due to the extraordinary variation within the class II genes in general, but particularly so, in the exon II segment (Charbonnel et al., 2010; Meyer-Lucht & Sommer, 2005). Therefore, most studies have focused on specific variations within the exon II of the Class II genes. The structure of the class II region of the MHC in humans is composed of DP, DQ and DR sections, whereas the class II Ovar-MHC shares the *DQ* and *DR*, but lacks the DP section (Dukkipati et al., 2006). Furthermore, the sheep class IIa loci have been shown to be hypervariable, and five transcribed class II genes have been widely reported in literature

(Ballingall, Fardoe, & McKeever, 2008; Dukkipati et al., 2006; Herrmann-Hoesing et al., 2008; Hickford, Zhou, & Fang, 2007; Hickford et al., 2004). These five class IIa sheep MHC genes are the *DQB2*, *DQA2*, *DQB1*, *DQA1* and *DRB1* genes, and cover approximately, 86.8 kb of the region (Herrmann-Hoesing et al., 2008). A sixth class IIa gene, the *DRA*, is reported to be less polymorphic (Schwaiger et al., 1996). The non-coding portion of the class IIa region in sheep is reported to be highly conserved between many mammalian species, and is considered important to the functioning of immune cells (Herrmann-Hoesing et al., 2008). However, this region has not received as much research attention as have the class II genes. In sheep, most of the disease association studies reported have been linked variations within three of the 5 MHC loci; *DRB1*, *DQA1*, & *DQA2* (Hickford et al., 2011; Keane et al., 2007; Sayers et al., 2005).

### **1.3.3 Mechanisms of balancing selection at the MHC**

New variations at the MHC are thought to be generated by two main mechanisms, namely; point mutations and micro-recombination events (Edwards & Hedrick, 1998; Schierup, Mikkelsen, & Hein, 2001). Some of these variations have been shown to predate speciation as they are present in different lineages (Ballingall et al., 2015; Bollmer, Vargas, & Parker, 2007; Bryja et al., 2006). Evidence of conserved orthologues of specific MHC class II alleles have been identified among sheep, goats and cattle species (Ballingall et al., 2015). There is a general agreement in the literature that the extensive variation at the MHC region is maintained by balancing selection (Charbonnel & Pemberton, 2005; Doherty & Zinkernagel, 1975; Edwards et al., 1997; Hedrick, 1998; Hughes & Yeager, 1998; Marmesat et al., 2017; Parham & Ohta, 1996). However, a plethora of selection theories have been proposed as the mechanism for the maintenance of this balance at the MHC (Charbonnel & Pemberton, 2005). The three most well supported among these mechanisms include pathogen mediated selection (Charbonnel & Pemberton, 2005; Hedrick, 1998; Hughes & Yeager, 1998; Parham et al., 1995), maternal-foetal interaction (Hedrick, 1998), and negative-assortative mating or sexual selection (Edwards & Hedrick, 1998; Hedrick, 1998). All these three mechanisms are plausible, and are well supported in the literature. However, it is important to note that all of these mechanisms could be working in concert at any point, depending on the prevailing evolutionary factors that maintain the variation at the MHC (Hedrick, 1998). Indeed, different selection mechanisms have been detected simultaneously in different class II MHC genes (Tollenaere et al., 2008). Of these three mechanisms, the maternal-foetal interaction is limited to mammalian species because live bearing is absent in reptiles and birds. The evidence of the importance of this

mechanism was established in the study of spontaneous abortions in humans, in which specific maternal-foetal HLA loci compatibility were required for successful implantation (Thomas et al., 1986). The mode for pathogen mediated selection has been reported in many species including human malaria infection, sheep *Teladorsia species* infection, and Marek's disease in chickens (Adrian et al., 1994; Charbonnel & Pemberton, 2005; Hedrick, 1998; Plachy, Pink, & Hála, 1992). Negative assortative mating in a population is simply defined as the occurrence of more non-random mating among non-like members within the population than expected. In humans, the role of negative assortative mating has been of lesser prominence as the other two mechanisms (pathogen mediated & maternal-foetal interactions) primarily due to the practical difficulty in obtaining sufficient samples to generate high statistical power (Hedrick, 1998). Nonetheless, evidence of negative assortative mating has been found in limited HLA studies (Jin, Speed, & Thomson, 1995; Rosenberg, Cooperman, & Payne, 1983). Conversely, evidence of negative-assortative mating has been well demonstrated in mouse MHC studies (Brown & Eklund, 1994; Hedrick, 1998; Yamazaki et al., 1988). Of these, three popular theories of maintenance of balancing selection at the MHC, the pathogen mediated mechanism is the most well investigated in sheep. Many studies have demonstrated specific variations at MHC to a directly to the interactions between pathogen and host (Bryja et al., 2006; Dukkupati et al., 2006). These differences have been in the form of specific allelic, genotypic and haplotypic differences (Hassan et al., 2011; Keane et al., 2007). The three different types of pathogen mediated balancing selection reported are heterozygote advantage (Hedrick, 1998)(ref), rare allele advantage (also referred to as the frequency dependent mode) (Charbonnel & Pemberton, 2005; Hedrick, 1998) and Spatio-temporal or "the time and space" dependent selection mode (Charbonnel & Pemberton, 2005; Hedrick, 1998). All three modes are well supported in the literature (Charbonnel & Pemberton, 2005; Doherty & Zinkernagel, 1975). The heterozygote advantage mode, as the name implies, is said to prevail when each of the two alleles within the MHC locus in a heterozygote is capable of conferring immunity to a different group of pathogens (Charbonnel & Pemberton, 2005). On this basis, the heterozygotes present an advantage over the homozygote. The pre-condition of the rare allele advantage mode of maintenance of balancing selection is when a new variant confers immune protection to a prevailing disease demographic of the ecosystem. Such variants are favoured by demographic selection pressure, and persist within the target population at higher frequencies over time as observed in a study of voles (Bryja et al., 2006). The "time and space" pathogen mode of balancing selection occurs in situations when the disease events are spatio-temporary such as epidemics (Charbonnel & Pemberton, 2005; Hedrick, 1998). The survivors of disease

epidemics will harbour the associated MHC variations. However, this mode pathogen mediated selection is considered as part of the heterozygote advantage mode (Hedrick, 1998). The extraordinary mechanisms of balancing selection permit the capacity to maintain high levels of variation at the MHC, even in the face of fluctuations in the selection demographic factors. Surprisingly high levels of variation were recorded at the MHC loci in comparison to non-MHC loci in an otherwise monomorphic fox population (Aguilar et al., 2004) and in a population of bighorn sheep (Gutierrez-Espeleta et al., 2001). However, lower than expected levels of MHC variation reported in an endangered population of penguins was attributed to a past population bottleneck event (Bollmer et al., 2007). Therefore, a low MHC variation can also result from an extreme bottleneck event or a low intensity of selection (Aguilar et al., 2004; Bollmer et al., 2007).

#### **1.3.4 Contrasting immunological phenotype in Djallonke and Sahelian sheep**

Djallonke and Sahelian sheep are two of the most important sheep breeds in sub Saharan African (Brahi et al., 2015), a region, characterised by low input extensive livestock production systems (Cooper et al., 2008; Kosgey & Okeyo, 2007). A major constraint to these livestock production system is parasitic diseases (Kosgey & Okeyo, 2007). The Djallonke sheep breed is the most ancient African sheep breed (Muigai & Hanotte, 2013), and has been shown to be resistant to wide range of tropical livestock diseases, particularly, to the animal trypanosomiasis and intestinal helminths in comparison to the Sahelian sheep (Goossens et al., 1997; Goossens et al., 1999; Traoré et al., 2017). Most of these studies involved comparative haematological and parasitological investigation in the two sheep breeds with respect to parasitic infections. Resistance to parasitic diseases in indigenous SSA livestock has been ascribed to adaptive selection over a millennium of co-evolution with these parasites (Mwai et al., 2015). Therefore, one genetic region of the host that is most likely contributing to the acquired immunity is the MHC region. In sheep, resistance to infections, including nematodes, has been associated with variation within genes in the MHC class II region (Hickford et al., 2011; Sayers et al., 2005; Schwaiger et al., 1995). Although the class II MHC is the main immune control centre for parasitic infections (Charbonnel et al., 2010), the extent to which the MHC genotype in these two sheep breeds has influenced their respective phenotypes in relation to parasitic infections remains unexplored. Unravelling of the genetic basis for diversity between breeds is of crucial importance for its sustainable utilisation and conservation. Therefore, an analysis of the MHC, particularly the class II region, in these two

sheep breeds will offer key molecular insight into the observed differential disease phenotypes. Finally, this information is important in understanding the true value and distinctiveness of these genetic resources, and will facilitate the application of molecular tools in the management of both sheep.

### **1.3.5 Conclusions**

Molecular characterisation of the immune genes is important for understanding the genetic basis of observed immunological phenotypes in livestock. The MHC-disease nexus is probably, one of the most complex and multifaceted dynamics of any biological system. Based on the available literature, the class II region of MHC in sheep appears to be an important region of interest region for control of the majority of parasitic diseases in sheep. Therefore, genetic analysis of this region in the Djallonke and Sahelian sheep will facilitate the identification and subsequent association of specific MHC class II alleles or polymorphisms to resistance or susceptibility to parasitic diseases. This information is fundamental to the development of efficient and sustainable management of the two sheep breeds.

## **1.4 Research aims**

Overall, given the current lack of genetic information on the relatively disease resistant Djallonke sheep and disease susceptible Sahelian sheep of Ghana, and in the light of the demonstrated key importance of the MHC class II genes in resistance to parasitic infection in sheep, the specific aims of this research include:

To first conduct a comparative whole genome variant analysis between the Djallonke and Sahelian sheep breeds using the “sequence and pool” strategy to identify both common variants and breed specific variants which will provide targets for the development of molecular identification tools.

To perform an analysis of signatures of selection of the whole genome variant datasets for the Djallonke and Sahelian sheep breeds to find genetic evidence of their contrasting phenotypes.

To characterise the genomic architecture of the class II MHC region in these two breeds, and identify both common and breed specific alleles that could be associated with the differences in resistance to parasitic diseases in these two sheep breeds.

To perform a comparative population genetics analysis of the MHC class II genes in a population of Djallonke and Sahelian sheep in order to understand how the demographic factors have shaped this genomic region in the two sheep populations.

## Chapter 2 Comparative analyses of whole genomic variant characteristics

*This chapter describes the comparative analysis of whole genomic sequences from Djallonke and Sahelian sheep breeds using a strategy of “sequence and pool” as proposed in chapter 1. Five unrelated Djallonke and Sahelian sheep sourced from Ghana were individually sequenced, pooled and giving a greater than 22-fold average coverage for each breed. High levels (~ 96%) of variant similarities were found between the two breeds, probably reflecting their shared demographic histories. In addition, significant numbers of potential breed specific single nucleotide polymorphisms (SNPs), insertions and deletions (INDELs) were identified for the Djallonke (~ 0.33 million) and Sahelian (~ 0.16 million) breeds, and these will provide targets for the future development of breed specific molecular identification markers. Of the 11.1 million and 10.9 million SNPs identified in the Djallonke and Sahelian breeds, approximately 15% (~ 2 million) and 16% (~1.8 million) respectively, have not been previously observed in sheep (in reference to the sheep dbSNP of the Ensembl variation release 85 Ovis aries). The Djallonke and Sahelian sheep constitute a significant and unique genetic resource, and results from this study contribute to an understanding of world sheep diversity. This study represents the first whole genomic variant analysis in these two sheep breeds, and will be of fundamental importance for the sustainable management of these two sheep breeds.*

### 2.1 Introduction

The Djallonke sheep breed is of high socioeconomic and cultural significance in at least 14 countries within the sub-Saharan African (SSA) region (International Livestock Research Institute, 2005). The Djallonke breed is regarded as the most ancient of the approximately 170 sheep breeds in Africa (Muigai & Hanotte, 2013). The Djallonke was thought to have originated in western Asia in about 7000BCE (5000BC), and entered Africa through the Isthmus of Suez (International Livestock Research Institute, 2005). Recent studies, based on archaeological information, have supported a date range between 7000 and 7500BCE for their entry into Africa (Muigai & Hanotte, 2013). The Djallonke, a “thin tailed sheep breed”, inhabits the regions south of latitude 14° N, encompassing central and western Africa.

A long history of cross breeding between Djallonke and Sahelian sheep has led to a trend of introgression of Sahelian genetics into the Djallonke breed as described in chapter 1. This manifests as a loss of the adaptive characteristics within the introgressed Djallonke sheep (Geerts et al., 2009; Goossens et al., 1999; Traoré et al., 2012). This high rate of cross breeding

has been reported to be driven by many factors as described previously in chapter 1. This long history of crossbreeding has led to various crossbred populations of Djallonke and Sahelian sheep. Phenotypically, crossbred animals can more closely resemble either the pure Djallonke or pure Sahelian breed, depending on their composition. Most of the farmers in the region are smallholder peasants and are largely illiterate, and therefore record keeping has historically been through oral tradition. As a consequence, through the passage of time and several generations of stock keepers, most of the information regarding the actual breed composition or status for many animal groups has been lost. Currently, the mode of identification of both breeds is through the use of visual physical characteristics, and in most cases, by consensus amongst the regional farmers (Birteeb et al., 2012). Most breed diversity studies, as well as breed status reports, are use these identification criteria (Asamoah-Boaheng & Sam, 2016). Given the long history of introgression, these criteria are therefore unreliable for planning and decision making in relation to the sustainable management of these important animal genetic resources. The characterisation of both sheep breeds at the molecular level will facilitate the application of modern genomic tools for sustainable breeding and utilisation of these two important genetic resources.

In spite of the importance of the Djallonke sheep to the region, genetic studies have been scarce (Alvarez et al., 2009; Traoré et al., 2012; Wafula et al., 2005). There are no records of whole genome variant characterisation in either of the Djallonke or Sahelian breeds. To the best of our knowledge this is the first report on the application of whole genome sequence analysis in these two important sheep breeds.

The objectives presented in this chapter are;

- i) To identify and perform a comparative analysis of whole genome variants (SNPs and indels) in the Djallonke and Sahelian sheep breeds.
- ii) To annotate the identified variants for both Djallonke and Sahelian sheep in relation to the Ensembl variation dbSNP for sheep (release 85) and then perform a comparative analysis of the functional characteristics of the variants within and between the two sheep breeds.

## **2.2 Methods**

### **2.2.1 Sheep**

Five unrelated sheep from each of the Djallonke (DJ) and Sahelian (SA) breeds were identified and selected from two National sheep breeding stations in Ghana (see appendix A, for the geographic location of the breeding stations, pictures of Djallonke & Sahelian. The Djallonke sheep were selected from the “National Open Nucleus Breeding Station for Djallonke Sheep” in Ejura in the Ashanti region (longitude 01° 28’W and latitude 06° 41’ N), and included 4 ewes and 1 ram. The Sahelian sheep were selected from the “National sheep breeding station in Pong Tamale” (longitude 00° 54’W and latitude 09° 38’ N) and comprised 5 ewes. All sheep sampled had attained reproductive maturity (13-24 months old) and were selected in consultation with the managers of the regional breeding stations to represent animals that were true to breed type (phenotypically similar to the breed ideal). The republic of Ghana is located within longitudes 03° 11’W and 01°11’ E and latitudes 04° 44’N and 11° 11’ N. It is bordered on the north, east and west by Burkina Faso, Togo and Côte d’Ivoire, respectively, and about 550km of the Atlantic Ocean coastline comprises the southern boundary. Ghana enjoys approximately 200 days of sunshine in a year, with a mean annual temperature of 28°C (annual range is between 15°C to 28°C). The mean annual precipitation is about 787 mm. The average relative humidity is 85%.

### **2.2.2 Blood Collection & DNA extraction**

From each of the 10 sheep approximately 9ml of blood was collected via jugular venipuncture into labelled vacutainers containing the disodium salt of ethylene diamine tetra-acetate (EDTA). The samples were transported at 4° C to the laboratory. The whole blood samples were subjected to centrifugation at 800 x g for 3 minutes at room temperature (15-25°C) using a Beckman centrifuge with the rotor bucket brake set to “off”. The buffy coat was aspirated from the distinct middle layer using a disposable 5ml pipette and transferred into 5ml screw cap tubes for genomic DNA extraction, and was either extracted immediately or stored at -20°C.

Genomic DNA was extracted from each sample using the Zymo Quick-gDNA<sup>TM</sup> MiniPrep DNA purification Kit (per the manufacturer’s protocol). DNA quality was assessed using agarose gel electrophoresis (1% w/v in TAE) and by Nanodrop spectrophotometry.

### **2.2.3 Library Construction and Sequencing**

For each sample, 100ng of DNA was sheared, to generate a broad range of DNA fragments with sizes ranging from 100-1000bp, using the Covaris S2 System. The NEB Next Ultra kit (New England Biosciences) was then used to process and ligate the DNA fragments to T-overhang adaptors with different barcodes assigned for each animal (Integrated DNA Technologies, Singapore). Fragments of approximately 300bp were then size-selected using the E-gel system (Invitrogen), and fragments were purified using AMPure XP SPRI beads (Beckman).

Equimolar amounts of each library were then combined and amplified using emulsion PCR on the Ion Chef system (ThermoFisher Scientific). The amplified libraries generated were sequenced on a P1 chip using the Ion Proton™ system (ThermoFisher Scientific).

### **2.2.4 Mapping and Pre-processing of Reads**

Base calling, de-multiplexing, quality control (QC) and alignment pre-processing were performed using the TorrentSuite 4.6 on a Torrent Server (ThermoFisher Scientific) as per the optimized protocol used by Yuan, Xu, and Leung (2016). Briefly, polyclonal and uniformly low-quality reads were removed and the reads trimmed from the 3' end only. Mapping to the sheep reference genome Oar v3.1 (University of California, Santa Cruz (UCSC)) was performed within the TorrentSuite 4.6 software, using the Torrent mapping alignment program (TMAP). For each sheep breed, individual BAM files were generated, merged and sorted using SAMtools v0.1.19-44428cd (Li et al., 2009). Coverage analysis was performed for both the individual and combined sequence datasets using automated plugins within the TorrentSuite 4.6 software package.

### **2.2.5 Variant calling pipeline**

Pooled BAM files were sorted and indexed using SAMtools v0.1.19-44428cd. Duplicate reads were removed using Picard Tools v1.122. Genome Analysis Tool Kit version 3.2.2 (GATK) RealignerTargetCreator and IndelRealigner were used to produce realignments of the pooled bam files for each breed. GATK HaplotypeCaller using GVCF mode was used to call intermediate genome-wide variants separately for the pooled DJ genomes and pooled SA genomes, resulting in two pooled genomic variant call format (gvcf) files. GATK GenotypeGVCFs was then used to perform a joint genotyping of these two pooled gvcf files for the Djallonke and Sahelian datasets using minimum standard confidence thresholds for both

calling and emitting. Called and emitting variables were set at 30 (-stand\_call\_conf 30 -stand\_emit\_conf 30) to produce a composite pooled variant call format (vcf) file for each breed (Pooled-Sheep VCF). This analysis procedure was selected to ensure good quality variant calling for the dataset and reduce false discovery rates. Finally, VCFtools (v0.1.15) (Danecek et al., 2011) was used to extract individual Djallonke and Sahelian samples from the composite joint genotyped vcf into separate vcf files, which were then used for further downstream analyses.

## **2.2.6 Annotation and functional analysis of genomic variants**

As there were unequal numbers of males and females used in this study for the two breeds, for the purpose of a balanced comparison, autosomal chromosomes were extracted from each of the pooled Djallonke and Sahelian vcf files using VCF tools v0.1.15. The SnpSift v4.2 “annotate” command was used to identify known SNPs in the pooled Djallonke and Sahelian vcf files, using the sheep dbSNP of the Ensembl release variation 85 *Ovis aries* VCF file as the database ([Ftp://ftp.ensembl.org/pub/release-85/variation/vcf/ovis\\_aries/](Ftp://ftp.ensembl.org/pub/release-85/variation/vcf/ovis_aries/)) (Cunningham et al., 2015). SnpEff v4.2 (Cingolani et al., 2012) was also used for functional annotation of identified autosomal SNPs in both the Djallonke and Sahelian genomes based against the Ensembl sheep genome assembly Oar\_v3.1.82. Pairwise comparison of Genomic SNPs and Indels for the Sahelian and Djallonke sheep genomes was computed in a UNIX environment using the BEDTools suite (v. 2.26.0) by (Quinlan & Hall, 2010).

## **2.3 Results**

### **2.3.1 Whole genome sequencing and mapping**

Genomic DNA from five unrelated individuals representing the Djallonke and Sahelian breeds were separately sequenced, and the sequencing reads pooled according to breed. After filtering and trimming, an average of 10% and 13% of the reads were excluded due to low quality scores, and 26% and 28% excluded due to polyclonality for the Djallonke and Sahelian samples, respectively. Coverage analysis was then performed on a total (post QC) dataset containing 73 Gbp of sequenced data, comprising 404,755,012 pooled reads (average read length 185.4 nt) and 57.6 Gbp comprising 303,136,043 reads (average read length 176.6 nt) for Djallonke and Sahelian breeds respectively (Table 2.1). The depth of coverage for each genome pool was estimated to be 27.90x and 22.01x for the Djallonke and Sahelian breeds respectively, and each

covered ~97% of sheep reference assembly v3.1. Key genomic parameters were computed to ascertain the accuracy of the sequenced data in this study. The autosomal SNP transition to transversion ratios observed for the Djallonke (2.47) and Sahelian (2.48) datasets, respectively, were both similar to each other (Table 2.3). The computed missense to silent SNP proportions were also similar within the Djallonke and Sahelian sheep datasets (0.69 and 0.68, respectively) (Table 2.6). The two datasets also had equal insertion to deletion ratios (0.38) (Figure 2.1).

### 2.3.2 Availability of data and materials

All the variant call format files for this project are available through the European variant archive of the European Bioinformatics Institute with the accession number PRJEB15642.

**Table 2.1 Summary of sequencing metrics for Djallonke and Sahelian sheep data generated on an Ion Proton Sequencer (ThermoFisher)**

Sheep	Raw reads	Polyclonal	Av. Low quality	Final post QC reads	Av. read length	Total data (Gb)	Mapping rates	Depth of coverage
<b>Djallonke</b>	625572175	26%	10%	414862202	185.4	76		
<b>Mapped reads</b>				404755012		73	<b>98%</b>	27.9X
<b>Sahelian</b>	568776210	28%	13%	347594060	176.6	61.5		
<b>Mapped reads</b>				334136043		57.6	<b>96%</b>	22.01X

### 2.3.3 Comparison of the chromosomal distribution of all identified Autosomal Variants of the Djallonke and Sahelian Genomes

The distribution of the variants (i.e. the sum of SNPs, Insertions and Deletions) was similar across all the autosomes, with chromosomes 11 and 26 showing the highest and lowest frequencies of variants for both the Djallonke and Sahelian sheep genomes, respectively (Table 2.2). A total of 12,821,836 and 12,654,761 variants were identified in Djallonke and Sahelian sheep, respectively, with 12,556,638 (96.30%) shared between the two genomes. A further pairwise variant comparison of the two datasets showed that 324,760 (2.49%) variants were specific to the Djallonke breed, whereas 158,085 (1.21%) variants were specific to the Sahelian breed. Therefore, the total number of variants identified for the two breeds is 13,039,483. In

addition, pairwise analysis indicated that 242,572 SNPs and 82,609 indels were specific to the Djallonke, whereas 120,652 SNPs and 37,762 indels were unique for the Sahelian (Table 2.4).

**Table 2.2 Comparison of the identified Autosomal variants' per chromosome of the Djallonke and the Sahelian Sheep Genomes**

Chromosome	Length (bases)	Djallonke		Sahelian	
		Variants	Ratio of nucleotides to variants	Variants	Ratio of nucleotides to variants
1	275,612,895	1,415,883	194	1,397,799	197
2	248,993,846	1,252,169	198	1,236,471	201
3	224,283,230	1,126,647	199	1,112,239	201
4	119,255,633	614,660	194	606,652	196
5	107,901,688	547,117	197	540,304	199
6	117,031,472	638,104	183	629,707	185
7	100,079,507	521,899	191	514,989	194
8	90,695,168	463,424	195	457,602	198
9	94,726,778	508,149	186	501,452	188
10	86,447,213	458,020	188	451,926	191
11	62,248,096	309,854	200	305,905	203
12	79,100,223	418,404	189	412,899	191
13	83,079,144	420,072	197	414,577	200
14	62,722,625	313,626	199	309,445	202
15	80,923,592	431,999	187	426,246	189
16	71,719,816	397,765	180	392,498	182
17	72,286,588	390,027	185	384,875	187
18	68,604,602	361,551	189	356,643	192
19	60,464,314	314,693	192	310,777	194
20	51,176,841	272,206	188	268,369	190
21	50,073,674	279,334	179	275,226	181
22	50,832,532	279,610	181	275,533	184
23	62,330,649	342,228	182	337,640	184
24	42,034,648	233,669	179	230,626	182
25	45,367,442	259,240	175	256,130	177
26	44,077,779	251,486	175	248,131	177
<b>Total</b>	2,452,069,995	12,821,836		12,654,761	
<b>Average</b>			191		193

**Table 2.3 Autosomal SNP Transitions and transversion ratio in Djallonke and Sahelian Dataset**

	Djallonke	Sahelian
Transitions (Ts)	9,141,413	8,893,550
Transversions (Tv)	3,695,613	3,583,318
Ts/Tv	2.4736	2.4819

**Table 2.4 Pairwise comparison of Genomic SNPs and INDELS computed for Sahelian and Djallonke sheep**

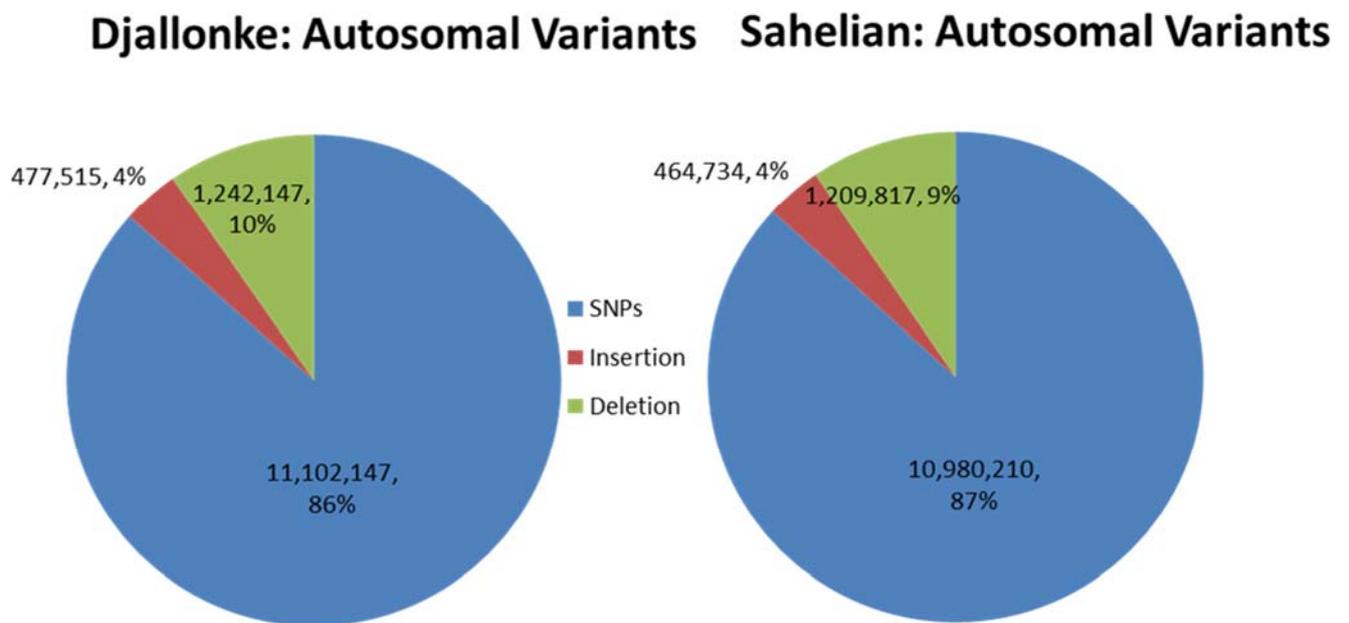
	<b>Djallonke</b>	<b>Sahelian</b>	<b>Commonality</b>
<b>Total SNPs</b>	11,102,147	10,980,210	10,852,501
<b>Breed-specific SNPs</b>	<b>249,646</b>	<b>127,709</b>	
<b>Total Indels</b>	1,719,689	1,673,551	1,668,901
<b>Breed Specific Indels</b>	<b>82,609</b>	<b>37,762</b>	

### 2.3.4 Annotation of variants in Djallonke and Sahelian Sheep

The snpEff-snpSift annotation of the total variants for both Djallonke and Sahelian sheep in relation to the Ensembl variation dbSNP for sheep (release 85) showed that 10,764,740 (83.97%) and 10,671,465 (84.33%) of the total variants, respectively (Table 2.5), were previously published in the database, suggesting that 2,0570,096 (Djallonke 16.03%) and 1,983,296 (Sahelian 15.67%) of the observed variants are new. The total variants for each of the sheep breeds can be further subdivided into SNPs and Indels. There were similar proportions of SNPs, insertions and deletions for Djallonke (86%: 4%: 10%) compared to Sahelian (87%: 4%: 9%) (Figure 2.1). An average rate of 1 SNP per 221 bp and 1 SNP per 223 bp, was observed for the Djallonke and Sahelian sheep respectively.

**Table 2.5 Total Autosomal variants identified showing the novel variants components in Djallonke and Sahelian sheep**

	Total variants	Variants already present in Ensembl (%)	Novel variants (%)
<b>Djallonke</b>	12,821,836	10,764,740 (83.97%)	2,057,096 (16.03%)
<b>Sahelian</b>	12,654,761	10,671,465 (84.33%)	1,983,296 (15.67%)



**Figure 2.1 Comparison of the numbers and proportion of SNPs , insertions and deletions for Djallonke and Sahelian sheep with respect to Oar v3.1**

A similar analysis of total SNPs, revealed that 10,460,303 (94.22%) and 10,372,765 (94.47%) SNPs for the Djallonke and Sahelian sheep, respectively, to be present in Ensembl variation dbSNP build 147. This suggests that 641,844 (5.78%) and 607,445 (5.53%) of the observed SNPs for the Djallonke and the Sahelian sheep, respectively, are previously unidentified sheep SNPs.

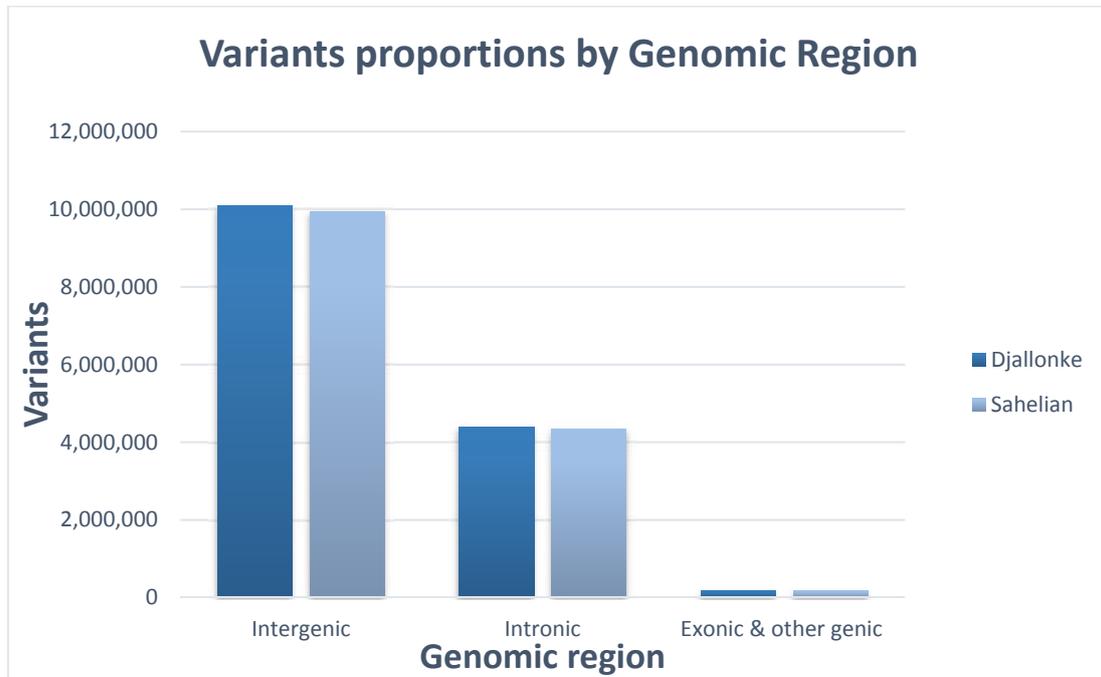
**Table 2.6 Comparisons of Autosomal variants by type and functional categories in Djallonke and Sahelian Genomes**

<b>Variants</b>	<b>Djallonke</b>	<b>Sahelian</b>
<b>Functional class</b>		
<b>Missense</b>	33,984 (40.7%)	33,395 (40.5%)
<b>Nonsense</b>	344 (0.41%)	338 (0.41%)
<b>Silent</b>	49,171 (58.9%)	48,732 (59.1%)
<b>Mutations by type</b>		
<b>Intergenic region</b>	8,861,516	8,747,769
<b>Upstream gene Variants</b>	611,224	601,689
<b>Downstream gene Variants</b>	605,975	596,609
<b>Intronic Variants</b>	4,413,052	4,350,066
<b>Intragenic Variants</b>	199	181
<b>3 prime UTR Variants</b>	29,548	28,988
<b>5 prime UTR Premature start codon gain Variants</b>	681	672
<b>5 prime UTR Truncation</b>	2	2
<b>5 prime UTR Variants</b>	5,412	5,307
<b>Splice acceptor Variants</b>	1,507	1,454
<b>Splice donor Variants</b>	1,640	1,553
<b>Splice region Variants</b>	14,544	14,262
<b>Exon Synonymous Variants</b>	49,141	48,702
<b>Non synonymous Variants</b>	33,902	33,313
<b>Non coding exon Variants</b>	22,865	22,315
<b>Non coding transcript Variants</b>	52	52
<b>Exon loss Variants</b>	12	12
<b>Initiator codon Variants</b>	7	8
<b>Start lost</b>	97	92
<b>Stop gained</b>	541	528
<b>Stop lost</b>	44	43
<b>Stop retained Variants</b>	28	28
<b>Transcript ablation</b>	1	1
<b>In frame deletion</b>	349	330
<b>in frame insertion</b>	674	646
<b>Disruptive in frame deletion</b>	476	447
<b>Disruptive in frame insertion</b>	607	592
<b>Frameshift Variants</b>	16,997	16,661

### **2.3.5 Comparison of the distribution of autosomal variants by genomic regions in the Djallonke and Sahelian sheep**

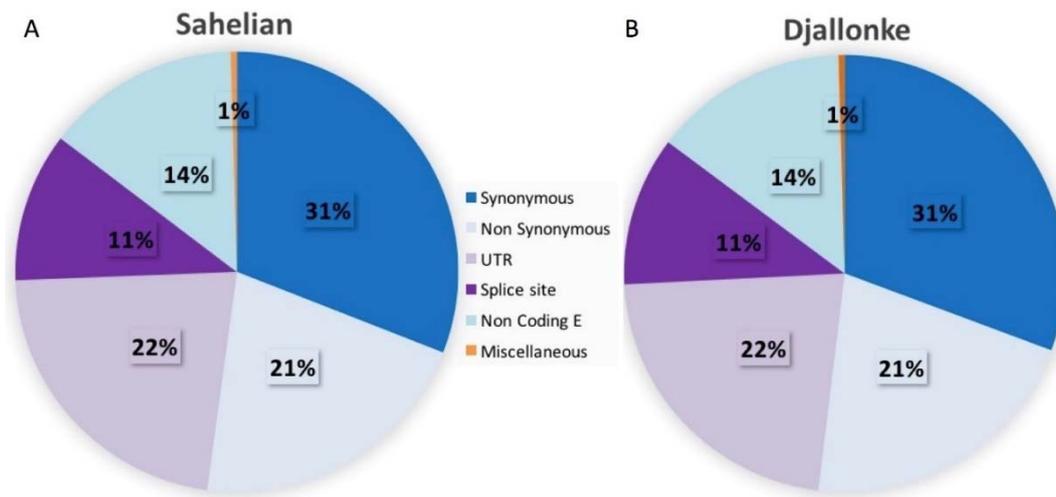
Annotation of all SNPs using SnpEff v4.2 shows that a majority of the variants in both sheep breeds were intergenic and intronic (Table 2.6), with approximately 1% located in the remaining genic regions (i.e. untranslated regions, exons, and splice sites). Although the Djallonke sheep recorded higher numbers of variants for all regions of the genome in

comparison to the Sahelian sheep, the ratios of the variants in these three regions (intergenic, intronic and other genic regions including exons) were similar in percentage terms for Djallonke (68.78%: 30.04%: 1.18%) and Sahelian (68.81%: 30.02%: 1.17%) (Figure 2.2).



**Figure 2.2 Comparison of variants distribution in the genomes of Djallonke and Sahelian sheep breeds**

Although the Djallonke had higher absolute numbers of SNPs, a comparative analysis of the exonic & “other” genic category revealed relatively equal proportions (in percentage terms) of the constituents synonymous, non-synonymous, splice site, untranslated regions and miscellaneous variants for the two breeds (Figure 2.3).



**Figure 2.3 Similar proportions of variant distribution in exonic & other genic categories in the genomes of Djallonke and Sahelian sheep**

## 2.4 Discussion

### 2.4.1 Variant characteristics and Novelty in Djallonke and Sahelian sheep

A high number of previously undescribed variants were identified; 2,0570,096 (16.03%) and 1,983,296 (15.67%) in the Djallonke and the Sahelian sheep (Table 2.5), respectively, which confirms the uniqueness of these breeds as a genetic resource with respect to world sheep diversity. More than 0.5 million SNPs in each of the two sheep breeds were previously unreported (Djallonke: 641,844 SNPs and Sahelian: 607,445 SNPs) in the sheep dbSNP of Ensembl variation release 85. The high number of previously unreported variants identified is mainly because this sheep reference assembly was generated from the Texel sheep breed which has completely different demographic history to the Djallonke and Sahelian sheep breeds (The International Sheep Genomics et al., 2010). The transition to transversion ratio determined for both Djallonke (2.47) and Sahelian (2.48) are very similar to the observed values in other mammalian genomes: e.g. 2.26 for the cattle whole genome (Choi et al., 2015), ~2.30 in pig autosomal SNPs (Bianco et al., 2015), 2.13 for human intergenic SNPs (Gudbjartsson et al., 2015), and 2.81 for human exonic SNPs (Guo et al., 2012). These comparable ratios support the reliability of the sequenced datasets obtained in this study, and they are therefore expected to contain low numbers of false positives (Type 1 errors) caused by random sequencing errors. The numbers of breed specific SNPs and Indels obtained by pairwise comparison of the

Djallonke (242,572 SNPs, 82,609 Indels) and the Sahelian (120,652 SNPs, 37,762 Indels) indicates that considerable sequence differences exist between these two breeds (Table 2.4). These breed specific variants are strong candidate targets for the development of breed specific markers to facilitate the sustainable management of these breeds. These results corroborate recent reports that indigenous African small ruminants may constitute “treasure troves” that could aid in confronting emerging livestock diseases as well other global challenges, such as the uncertain consequences of climate change (Mwacharo et al., 2016). As recent reports indicate that most of the indigenous African livestock breeds are endangered, there is a need to develop more pragmatic steps for their sustainable utilisation and conservation, (Mwai et al., 2015). The high frequency of SNPs within each breed (>1 SNP per 148 base pairs), coupled with the comparable distribution of the variants across all the autosomal chromosomes (Table 2.2) should provide sufficient density of potential markers for the reliable elucidation of rare breed specific alleles (Choi et al., 2015). The high numbers breed specific SNPs obtained in this study could help explain the uniqueness of each of these two sheep breeds. In humans, it has been shown that NGS of a group of individuals, each at a lower coverage (e.g. 5x), is a cost-effective and efficient strategy for the elucidation of quality genomic variants, including rare disease variants, in comparison to just deep sequencing (e.g. 30x) of one individual of that group (Lee et al., 2014). The improved detection of high quality variants is because the NGS data for the group of individuals permits the leveraging of the sequence variation across the individuals within that group (Le & Durbin, 2011; Li et al., 2011). Furthermore, it has been demonstrated that a scalable reference panel can be developed based upon the information derived from the group NGS dataset and this can be used for imputation of the additional individuals (Li et al., 2011). The potential drawback of this strategy of lower coverage group NGS is that it is more likely to suffer more from sequencing errors, thus resulting in decreased accuracy (Lee et al., 2014). However, evidence from the genomic parameters shows that the Djallonke and Sahelian datasets in this study appear not to have been affected by this drawback. The combined high sequencing coverage statistics obtained for the Djallonke and for the Sahelian genomes (i.e. > 97% and > 22x), has been shown to be more than adequate for “high-confidence” variant calling (Choi et al., 2014). The advantage that NGS has over medium or high density SNP genotype datasets is that it provides higher resolution and power for the detection of selection signatures (Boitard et al., 2016). For example, the Illumina Ovine 50KSNP BeadChip and Illumina Bovine HD 800KSNP BeadChip only provide a SNP density of approximately one SNP for every 5 million and 3 million bases, respectively. In humans, it has been shown that, in comparison to the use of a dense SNP array, low coverage (0.1-1.0x)

NGS is capable of elucidating more common and rare sequence variations (Pasaniuc et al., 2012), as well as uncaptured rare disease variants (Gilly et al., 2016). Furthermore, the use of these types of commercial markers on breeds that have not been included in the original training set used for the marker development introduces the possibility of ascertainment bias into the analysis. Indeed, there is evidence of ascertainment bias with respect to African and Asian sheep breeds in an analysis of genetic diversity using the ovine SNP50 beadchip (Kijas et al., 2012).

The categorised proportions of the types of SNP in the intergenic, intronic and the remaining genic regions (including exons) for the two genomes (Figure 2.2) are similar to the proportions recorded in Korean cattle breeds (Choi et al., 2015). The exonic regions, although containing the least number of SNPs, represent the most important subset of SNPs, because they are more useful for the elucidation of subtle phenotypic differences among different populations than intronic and intergenic SNPs (Zhan et al., 2015). Population-specific, rare exonic SNPs have been shown to be the most consequential determinants of fitness traits in humans (Gu et al., 2015; Gudmundsson et al., 2012; Kelly et al., 2017; Rivas et al., 2011). Fixed non-synonymous SNPs, which are described as SNPs for which only one allele (of a given locus) is present in a population, are of major interest in identifying breed or population specific traits (Choi et al., 2015).

**Table 2.7 Comparison of the ratio of total autosomal SNPs to three consequential genomic variants categories identified in this study and other livestock species**

<b>Livestock species</b>	<b>Total SNPs</b>	<b>Start lost x10<sup>-5</sup></b>	<b>Stop gained x10<sup>-5</sup></b>	<b>Stop lost x10<sup>-5</sup></b>	<b>Reference</b>
<b>Djallonke</b>	11,102,147	0.87	4.9	0.40	This study
<b>Sahelian</b>	10,980,210	0.83	4.8	0.39	This study
<b>HAN</b>	13,544,560	0.25	2.9	0.12	(Choi et al., 2015)
<b>YAN</b>	15,857,687	0.27	3.0	0.15	(Choi et al., 2015)
<b>ADP</b>	25,427,907	0.29	3.4	-	(Bianco et al., 2015)
<b>AWB</b>	34,471,527	0.28	3.9	-	(Bianco et al., 2015)
<b>EDP</b>	29,073,261	0.43	12.5	-	(Bianco et al., 2015)
<b>EWB</b>	11,602,230	0.40	6.4	-	(Bianco et al., 2015)

#### **2.4.2 The potential role of “other variant “genomic variants categories to sustainable management domestic livestock**

The proportions of start lost, stop gained and stop lost variants in the genome are important because of their ability to directly alter both the function and stability of affected proteins (Savas, Tuzmen, & Ozcelik, 2006). The frequencies of such variants and other functional variant categories in a target population (Qanbari & Simianer, 2014) are expected to be shaped by the demographic history, in particular, the nature (i.e. purifying, positive, or neutral) and intensity of the selection factors. In humans, the frequency of premature stop codons gained have shown population specificity (Savas et al., 2006). Furthermore, human premature stop codons have been associated with a range of biological functions including accelerated

degradation of transcripts, and deletion of amino acids (Savas et al., 2006). These category of variants are, therefore, of high evolutionary significance, and the full extent of their functional role will be better understood when the data from the ongoing domestic animal functional annotation programs, including the functional annotation of the animal genome (www.FAANG.org) project is fully analysed (Tuggle et al., 2016). Until more information on the functional annotation of these groups of variants becomes available, discussions are limited, and this is probably why many NGS studies in domestic animals have not discussed these categories of variants (Kardos et al., 2015; Kawahara-Miki et al., 2011; Lee et al., 2013).

As the total variants determined using NGS of a genome can be affected by the average depth of sequencing coverage, the ratio of total SNPs to different variant types (e.g. start lost, stop gained & stop lost variants) was determined in three species to test if this was the case in this study. Table 2.7 presents the comparisons of the ratio of total autosomal SNPs to three different (start lost, stop gained & stop lost variants) consequential variants in this study to those from six groups comprising two cattle breeds and four groupings of domestic pigs and wild boars from Asia and Europe. The Djallonke and Sahelian have approximately 2-3 times ( $\sim 0.85 \times 10^{-5}$ ) the number of start gain variants than the two indigenous Korean cattle breeds (Hanwoo (HAN) & Yanbian (YAN)), the groups of domestic pigs and boars from Asia; Asian Domestic Pigs (ADP) & Asian wild Boars (AWB), and of European origin; European Domestic Pigs (EDP) & European Wild Boars (EWB) (Table 2.7). With the exception of the EDP and EWB, the Djallonke and Sahelian genomes show higher relative proportions of stop gained variants than observed in the HAN, YAN, ADP and AWB genomes. The EDP and EWB are reported to have lower effective population size in comparison to the ADP and AWB (Bianco et al., 2015), and that is probably why they showed the highest proportion of stop gained variants. Furthermore, the EWB has been intensively selected for a variety of production traits, resulting in dramatic changes in phenotype (Wilkinson et al., 2013). The combination of lower effective population size and the effect of extreme selection might have resulted in the accumulation of the relatively high proportion of stop gained variants ( $12.5 \times 10^{-5}$ ) than all the other groups analysed (Table 2.7). Similarly, the Djallonke and Sahelian breeds recorded a higher proportion of stop loss variants ( $\sim 0.4 \times 10^{-5}$ ) than the HAN and YAN cattle breeds (Table 2.7). The HAN cattle, which recorded the lowest proportion of stop loss variants ( $\sim 0.12 \times 10^{-5}$ ) has been subjected to more than three decades of intensive artificial selection for beef traits to become the representative beef breed of Korea (Choi et al., 2015). However, the indigenous YAN cattle breed, similarly to the indigenous Djallonke and Sahelian sheep breeds, are not been under such intense artificial pressure. A plausible explanation for the observation of the lowest

proportion of stop lost variants in YAN could be ascribed to the consequence of intense selection leading to reduced diversity in the selected population due probably to excessive culling of poor performing animals from the breeding program. Besides the functionality of these category of variants, it will be equally important to research how or if their occurrence and frequency in domesticates are affected by species, breeds or population of origin. For instance in bacterial populations, the frequency of stop codons has been shown to be positively correlated to the GC content (Povolotskaya et al., 2012), and it has been suggested that the types (TAA, TAG & TGA) and numbers of stop codons could be important to bacterial evolution (Wong et al., 2008).

The understanding of all these genomic dynamics and functionality of all these variants, not only SNPs and indels, could facilitate the sustainable management of global livestock genetic resources.

## **2.5 Conclusions**

We have conducted a whole genome variant analysis of the Djallonke and Sahelian sheep breeds and identified over 2 million novel genomic variants. This large number of novel variants suggests that the two sheep breeds represent unique genetic resources, and hence are important for world sheep diversity. The considerable number of breed specific SNPs identified in Djallonke and Sahelian sheep provide targets for development of breed specific markers that can be used for the sustainable management of each breed.

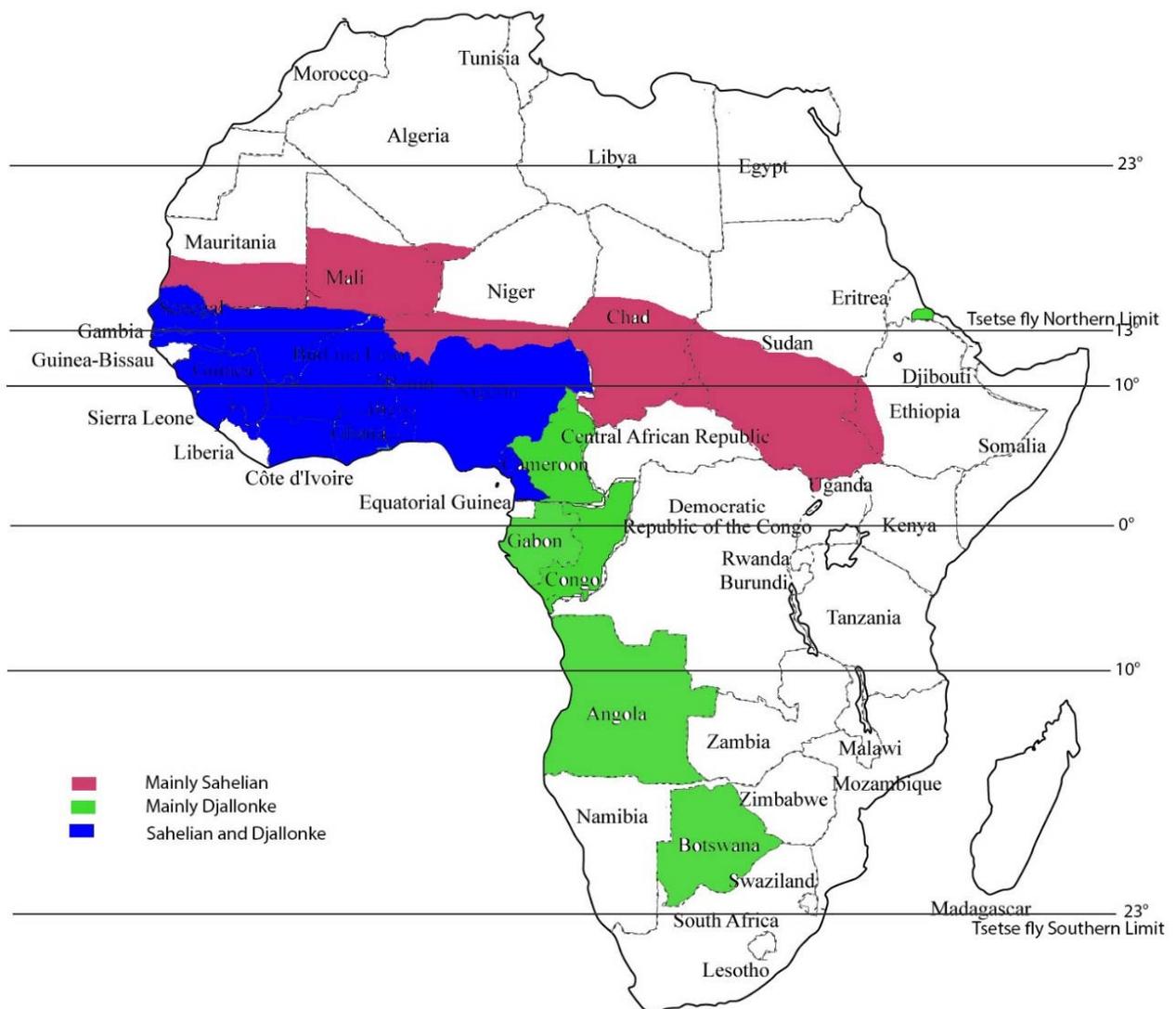
## **Chapter 3 Analysis of whole genomes of Djallonke and Sahelian sheep reveals co-localisation of regions of reduced heterozygosity with candidate genes for disease resistance and adaptation to tropical climatic conditions**

*This chapter presents a selection signature analysis of the whole genome variant datasets for the Djallonke and Sahelian sheep breeds that were described in chapter two. The high variant rates in Djallonke and Sahelian, coupled with the good accuracy of the sequencing process as determined by the ratios of transition and transversions, means that this is an optimal dataset for selection signature analysis. Multiple regions of extended run of homozygosity were found, 70 of these co-localised with genomic regions known to harbour genes that mediate disease resistance, immune response and adaptation in sheep or cattle. Of particular interest are the regions of homozygosity that were co-localised with previously reported genes for resistance to Haemonchosis and Trypanosomiasis. These homozygous regions are probably signatures of selection associated with these diseases and/ or a consequence of genetic hitchhiking. This study provides valuable insight for elucidating the genetic variations and mechanisms of the numerous adaptive traits in the Djallonke breed. This knowledge is fundamental to the sustainable breeding and utilisation of the Djallonke sheep, and has long term implications for food security and poverty alleviation in all 14 affected countries in the region.*

### **3.1 Introduction**

Djallonke sheep are recognised for their natural ability to withstand a harsh, hot and humid tropical climate, where they are faced with the challenges of persistent drought, diseases and feed scarcity (Wafula et al., 2005; Yaro et al., 2012). The most important of these attributes is an adaptation to the diseases trypanosomiasis and haemonchosis. Adaptation is probably a consequence of natural selection over several millennia (Dolan, 1987; Muigai & Hanotte, 2013; Naessens, 2006). The innate ability of the Djallonke sheep to survive and remain productive even under trypanosome challenge, resulting in very low mortality and without the use of trypanocidal drugs is referred to as trypanotolerance (Murray & Trail, 1984; Orange et al., 2012). The co-distribution of Djallonke sheep populations with the trypanosomiasis vector tsetse fly populations within the 10 million km<sup>2</sup> of the SSA region (Figure 3.1) underscores the

critical importance of this sheep breed as a resource for poor farmers in the region (Wafula et al., 2005). These adaptive characteristics of the Djallonke sheep enable the predominantly small-scale farmers within the region to successfully rear, and derive their sustenance from this breed without resorting to the use of veterinary drugs, which are neither accessible nor affordable for many of these farmers. Due to the high annual economic losses caused by trypanosomiasis to the SSA region, as described in chapter 1, and the accompanied inadequacies of a century of unsuccessful interventions against trypanosomiasis, the integration of natural trypanotolerance of livestock has been advocated by a growing number of studies (Geerts et al., 2009; Goossens et al., 1997; Namangala, 2012).



**Figure 3.1 Distribution of Djallonke and Sahelian sheep in relation to the Trypanosomiasis-Tsetse fly endemic zone of Africa**

*Haemonchus contortus* is regarded to be the most pathogenic and the most economically important gastrointestinal parasite in sheep production worldwide (Benavides, Sonstegard, & Van Tassell, 2016; Marshall et al., 2013). *H. contortus* is the predominant gastrointestinal helminth that affects sheep in the SSA region (Goossens et al., 1997; Preston & Allonby, 1979; Traoré et al., 2017). However, economic analyses of the impacts of the disease in this region are scarce. Globally, parasitic diseases of sheep and cattle annually cause several billions of dollars (USD\$) in economic losses to the animal production industry (Roeber, Jex, & Gasser, 2013).

The larger framed (45 kg to 65 kg), but more disease susceptible Sahelian sheep breed cohabits a large part of the SSA with the Djallonke sheep breed (20 kg to 30 kg), particularly the Savanna and Sahel regions spanning at least nine countries of West Africa, (Figure 3.1). Hence, these contrasting characteristics between the two sheep breeds provide conflicting economic drivers for the farmers in the region, and have been the basis of indiscriminate crossbreeding between the two breeds in order to produce larger sheep with an “intermediate resistance” (Goossens et al., 1999). The details of the negative effects of these widespread indiscriminate crossbreeding the two sheep breeds, notably the dilution of the disease resistant trait in the Djallonke, have been described in chapter 1.

Given that these disease resistance traits, particularly, trypanotolerance and helminth resistance, resulted from natural selection over millennia, it is hypothesised that there will be detectable regional genomic selective sweeps (marked by extended regions of homozygosity) corresponding to regions of the genomes influencing for resistance.

The objectives presented in this chapter include;

- i) to use the whole genomic variants for Djallonke and Sahelian sheep (presented in chapter 2) to identify molecular signatures of selection in both genomes.
- ii) to analyse the key differences and similarities between the genomic architectures of these two sheep breeds that could be used as potential targets for developing breed specific markers.

## 3.2 Methods

Whole genomic variant data for Djallonke and Sahelian sheep as variant call format (VCF) files (from chapter 2) were used for the analysis of selection signatures implemented in Homozygous Stretch Identifier software (HomSI, <http://www.igbam.bilgem.tubitak.gov.tr/en/software/HomSI/index.html>).

### 3.2.1 Detection of Signatures of selection

HomSI was used to identify signatures of selection in both genomes using the protocol established by Gormez, Bakir-Gungor, and Sagiroglu (2014). The Djallonke genome was designated as the index case and compared against the Sahelian as the unaffected case for input settings for the HomSI analysis. Analysis of genomic signatures of selection (also known as selective sweeps or high homozygosity) in Djallonke and Sahelian sheep using HomSI, with the stringent settings of 5Mb window size and 10 kb sliding size, permits the capturing of a wide spectrum of different sizes of signatures of selections throughout the genome (Bayrakli et al., 2015; Gormez et al., 2014; Kancheva et al., 2016; Tuncer et al., 2015). Integrative Genomics Viewer (IGV 2.3.46, [www.BroadInstitute.org](http://www.BroadInstitute.org)) was used to view vcf and BAM file tracks aligned to the sheep reference genome Oar v3.1, selecting regions, based on genomic coordinates, of potential selective sweeps (regions of contrasting homozygosity) identified by HomSI in order to identify candidate genes.

The prominent selection signatures identified were investigated using IGV for the specific genes of interest within a swept region. A modification of the methods used by Brown et al. (2013) and Kim et al. (2016) was used to identify candidate genes within an identified swept region. For every selective sweep in the HomSI output, the co-localised candidate gene/genes were inferred from the available Ensembl (Release 85 and 86) annotated sheep reference assembly (version 3.1) or from the conserved syntony for other mammalian genomes from the Ensembl genome database (Cunningham et al., 2015; Herrero et al., 2016; Yates et al., 2016). The approach presented here for investigating genetic evidence of trypanotolerance and resistance to haemonchosis was directly linking identified selection sweeps in this dataset to the reported quantitative trait loci in the database for animal genome QTL (Hu, Park, & Reecy, 2016) and other previously published QTL studies for the two traits. There have been several previous genomic investigations of resistance to nematode infection including *H. contortus* in multiple sheep breeds and these results were compared with our Djallonke and Sahelian sheep results (Benavides et al., 2015; Guo et al., 2016; McRae et al., 2014; Periasamy et al., 2014; Yang et al., 2015). In contrast, there has been no previous genomic investigation of trypanosomiasis in any sheep breed; therefore, comparison was made with the reported trypanotolerance QTL studies in Ndama cattle (Dayo et al., 2012; Hanotte et al., 2003; O'Gorman et al., 2009). Functional enrichment analysis of the genes identified within regions of high homozygosity was performed using the gene ontology information obtained from the Ensembl database releases 85 & 89 (Aken et al., 2016; Cunningham et al., 2015).

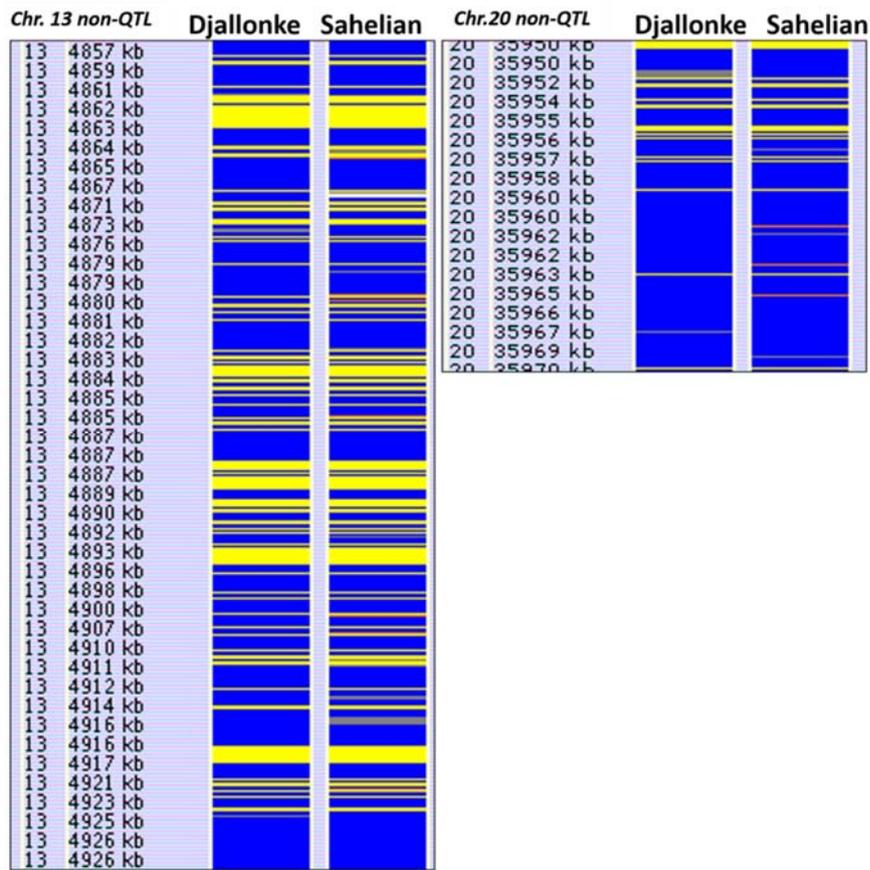
### 3.3 Results

#### 3.3.1 Signatures of Selection in Djallonke and Sahelian Sheep

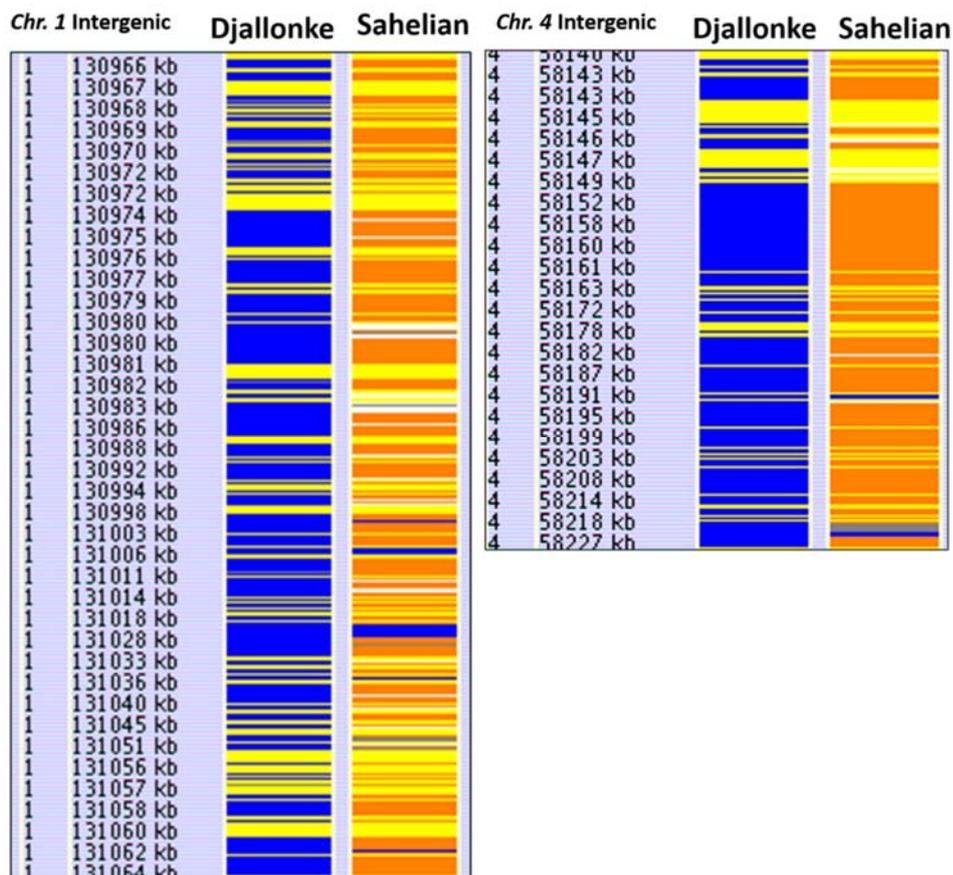
Using HomSI analysis, approximately 11.1 and 10.9 million of identified autosomal SNPs from the Djallonke and Sahelian breeds (Chapter 2 Table 2.4), respectively were resolved into about 50,000 detection windows for each breed. Interestingly, signatures of selection (regions of high homozygosity) of varying sizes were observed within both breeds across many regions of the genome, including both genic and intergenic regions. As presented in the HomSI software legend (3.3.1.1), at any genomic region of interest for the two sheep breeds, a more extensive stretch of blue colour denotes higher homozygosity in that sheep. Conversely, less extensive blue represents lower homozygosity. The detection efficiency of this strategy was further enhanced by the relatively high density of SNPs within this dataset (1:221 bases and 1:223 bases for Djallonke and Sahelian, respectively). Seventy of the regions containing selection signatures were co-localised with known QTL in sheep and other ruminant species. There were also several non-QTL associated genic regions that show high homozygosity in both breeds: e.g. Chromosome 20: 35,950,000-35,970,000 (20kb) and Chromosome 13: 4,847,000-4,926,000 (79kb), and are co-localised with the *CDKALI* (20: 35,755,504-36,365,789) and *BTBD3* (13: 4,900,219-4,905,937) genes, respectively (Figure 3.2). There were also non-QTL associated intergenic regions that showed high homozygosity in only one breed: e.g. Chromosome 1: 130,966,000-131,072,000 (106kb) and Chromosome 4: 58,143,000-58,227,000 (84), both show high homozygosity only in the Djallonke (Figure 3.3). Regions of completely contrasting selective sweeps were also identified. These are regions where the Djallonke was observed to be homozygous for a given allele (blue) and the Sahelian homozygous for the alternative allele (white): e.g. the *TRHDE* gene (Figure 3.10) and three non-QTL regions on chromosome 20 (Figure 3.4).

##### 3.3.1.1 HomSI Output legend for genomic sweeps (Adapted from Görmez *et al.*, 2014)

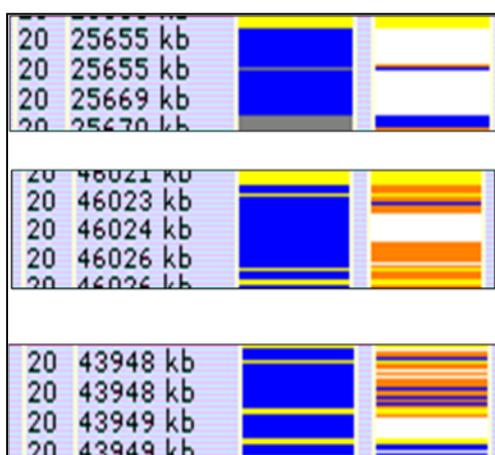
Index Case	Other Samples	Index Color	Other Sample Color
Homozygous	Homozygous	Blue	Blue
	Contrasting Homozygous		White
	Heterozygous		Orange
	No Call		Gray
Heterozygous	-	Yellow	Yellow
No Call	Homozygous	Gray	Blue
	Heterozygous		Orange
	No Call		Gray



**Figure 3.2** Prominent but similar run of homozygosity (see legend 3.3.1.1) at non QTL Genic regions on chromosome 13 (co-localised with *BTBD3* gene) and 20 (co-localised with *CDKALI* gene) in both sheep breeds



**Figure 3.3** An example of non-QTL intergenic regions showing high homozygosity (>80 kb) at chromosomes 1 and 4 only in the Djallonke sheep in striking contrast to the Sahelian sheep breeds



**Figure 3.4** Non QTL genic regions of contrasting homozygosity between Djallonke(blue) and Sahelian (white region represents stretch homozygous for the opposite allele) sheep on Chromosome 20 denoting breed specific alleles

### 3.3.2 Genomic evidence for tolerance to trypanosomiasis in Djallonke and Sahelian sheep breeds

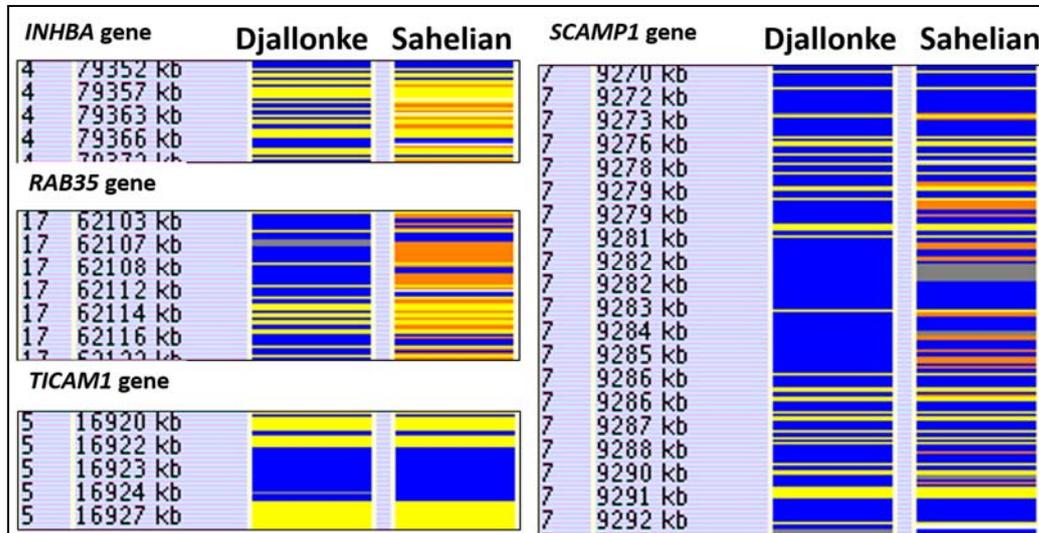
Eight of genomic regions of high homozygosity observed in this study were in regions previously identified to contain trypanotolerance QTL candidate genes (Table 3.1) (Dayo et al., 2012; Dayo et al., 2009; O'Gorman et al., 2009). Six of the eight previously reported QTL associated regions (*CTSS*, *ARHGAP15*, *INHBA*, *STX7*, *RAB35*, *CD19*) co-localised with the selective sweep regions identified in the Djallonke genome only (Figure 3.5, Figure 3.6 & Figure 3.7), and the other two; *SCAMP1* and *TICAM1* were co-localised within possible selective sweeps observed in both the Djallonke and Sahelian genomes (Figure 3.5).

HomSI output for the region adjacent to the four candidate genes associated with trypanotolerance is shown in Figure 3.5. These are the *INHBA*, *RAB35*, *TICAM1*, and *SCAMP1* genes. The Djallonke sheep appear to show higher homozygosity (blue) than the Sahelian sheep in both the *INHBA* and *RAB35* gene regions, but both breeds showing relatively high homozygosity (blue) across an approximately 2-kb stretch (16922kb-16924kb) at the *TICAM1* gene (Figure 3.5). The Sahelian breed shows a region of high heterozygosity (orange) between 9279 kb to 9286 kb within the *SCAMP1* gene region. In contrast, the Djallonke shows high homozygosity at the same region (blue). The HomSI analysis for the remaining two trypanotolerant associated QTL regions (*CD19* & *CTSS*) of low heterozygosity in the Djallonke sheep are highlighted in Figure 3.6.

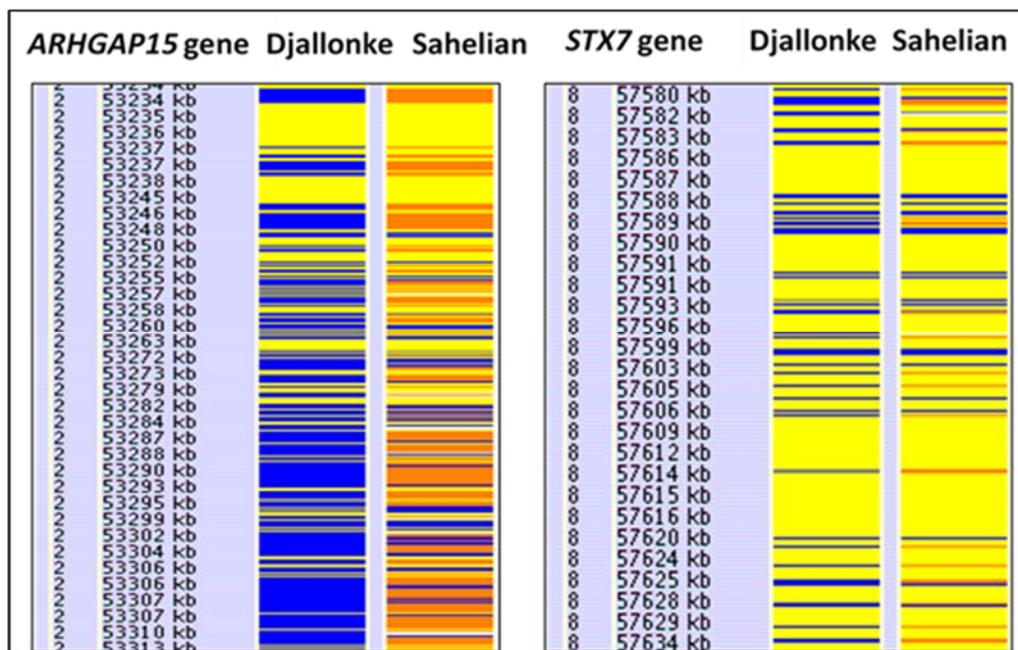
Similarly, Figure 3.8 and Figure 3.9 show five of the 25 candidate gene regions associated with resistance to *H. contortus* infection and other GIT parasites: *IFNG*, *CHIA*, *SUGT1*, *IL20RA* and *ATP2B1*. In each case, the Djallonke sheep has higher, in comparison to the Sahelian sheep. The HomSI output of the Djallonke and Sahelian sheep for the remaining 20 QTL regions that were associated with resistance to GIT parasites are shown in appendix C.1.

Output from HomSI analysis at seven of the 37 candidate gene regions associated with adaptive selection (Table 3.3; *MSRB3*, *TRHDE*, *FGF5*, *ALDH1A3*, *PLCB1*, *ELF2* and *BATF2*) is shown in Figure 3.10 and Figure 3.11. There are clearly observed differences between these genes with respect to the polymorphism patterns. The Djallonke sheep have higher homozygosity (blue) at the region containing the *MSRB3* gene while the Sahelian is more polymorphic (Figure 3.10). Interestingly, within the *TRDHE* gene region, the two sheep breeds show contrasting (blue versus white) run of homozygosity (Figure 3.10). This suggests that the two breeds are fixed for alternative alleles and haplotypes over much of this region. The Djallonke genome shows relatively high homozygosity (blue) at regions surrounding the *FGF5*, *ALDH1A3*,

*PLCB1*, *ELF2* and *BATF2* genes in comparison to the Sahelian genome (Figure 3.10 & Figure 3.11).



**Figure 3.5 Comparison of the sweep characteristics within four previously identified trypanotolerance-associated genes in Djallonke and Sahelian sheep breeds**



**Figure 3.6 Differential Sweep Characteristics at the trypanotolerant associated *ARHGAP15* gene and *STX7* gene of Djallonke (showing more extended homozygosity) and Sahelian (less homozygosity) sheep breeds**

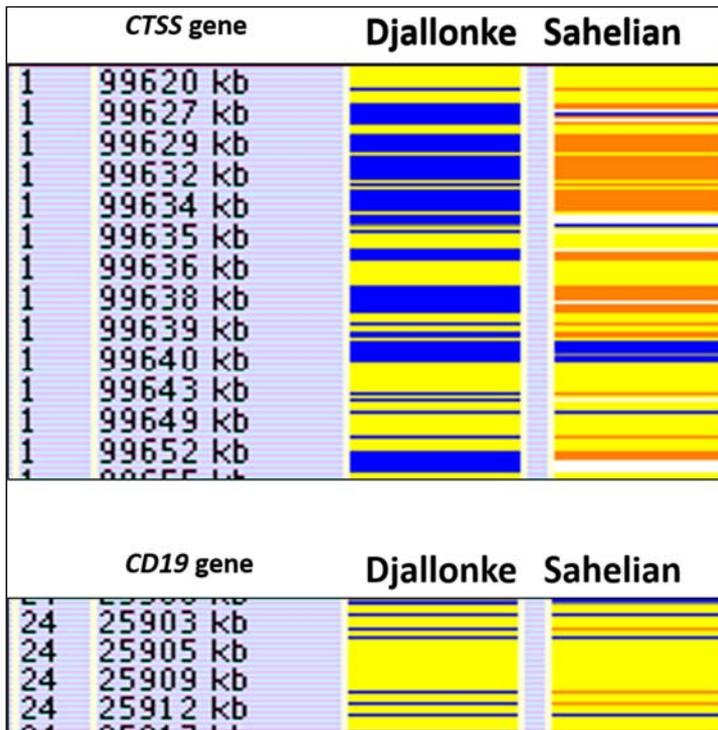


Figure 3.7 Two trypanotolerance associated QTL regions showing higher homozygosity for the *CTSS* and *CD19* genes for Djallonke in comparison to the Sahelian sheep

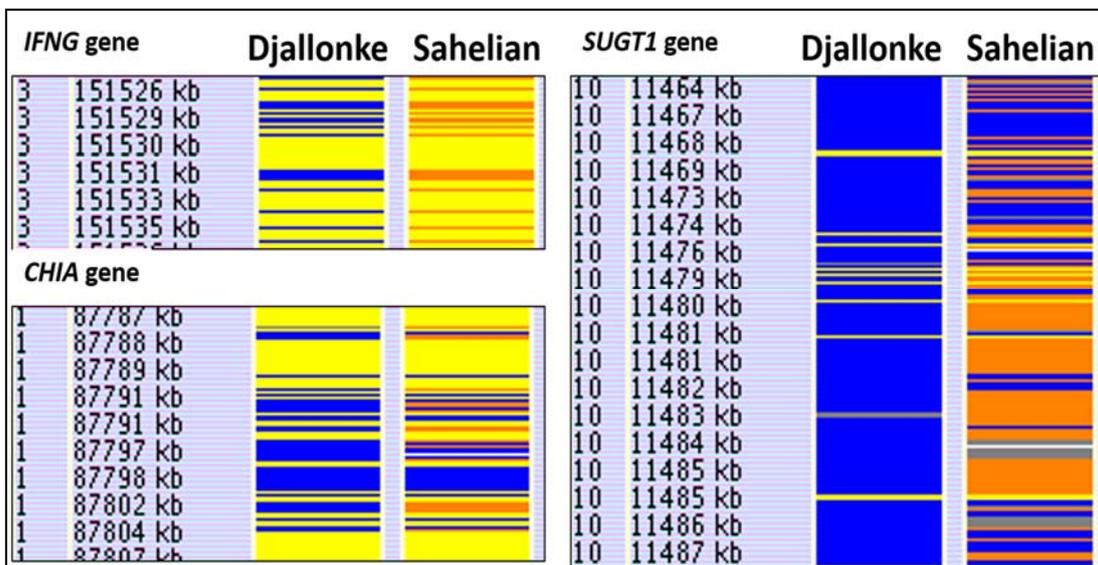


Figure 3.8 Comparison of the sweep characteristics at three genes previously shown to be associated with resistance to *Haemonchus contortus* higher homozygosity (blue) in the Djallonke sheep

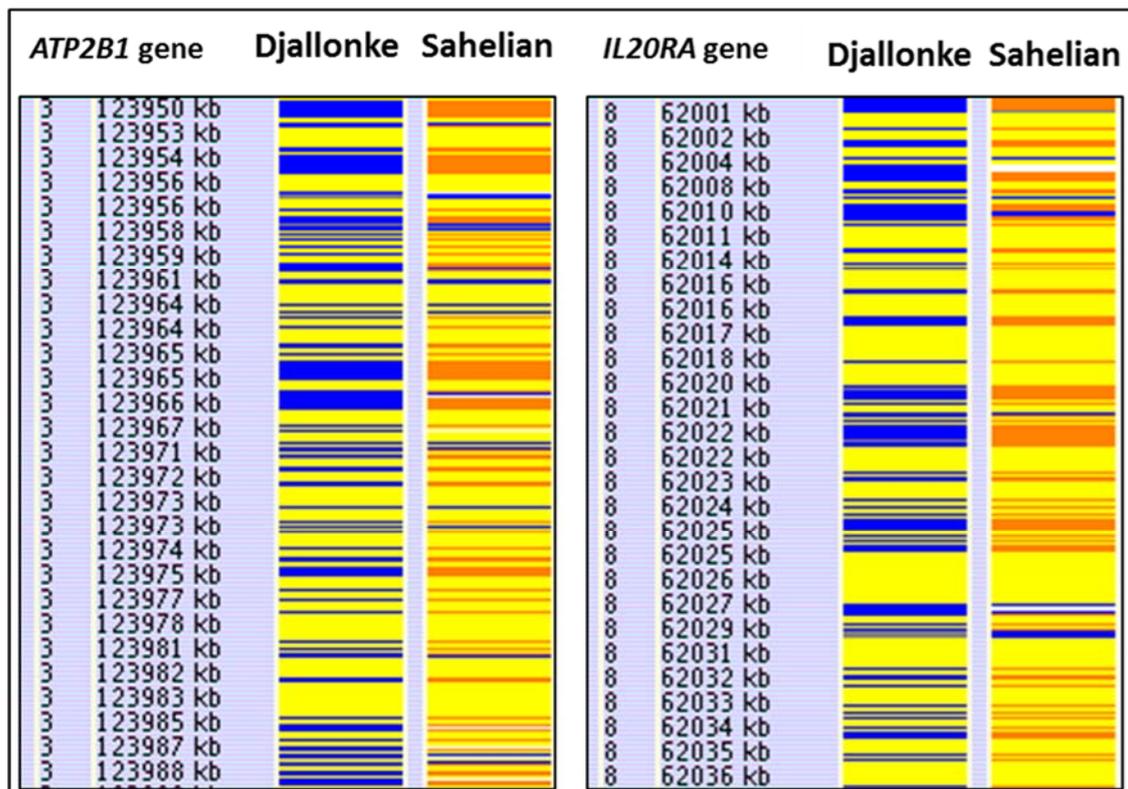


Figure 3.9 Comparison of the sweep characteristics at two genes previously shown to be associated with resistance to *Haemonchus contortus* showing higher (blue) in the Djallonke sheep

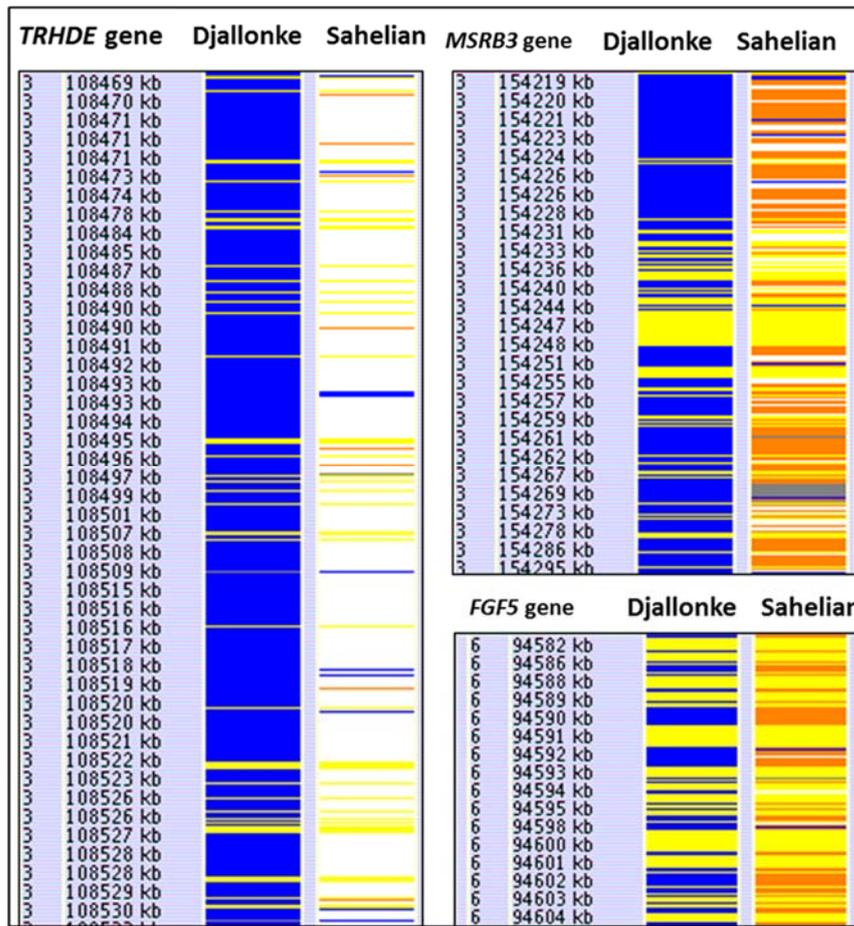


Figure 3.10 Comparison of the sweep characteristics in Djallonke and Sahelian sheep at three known QTL associated with adaptive selection

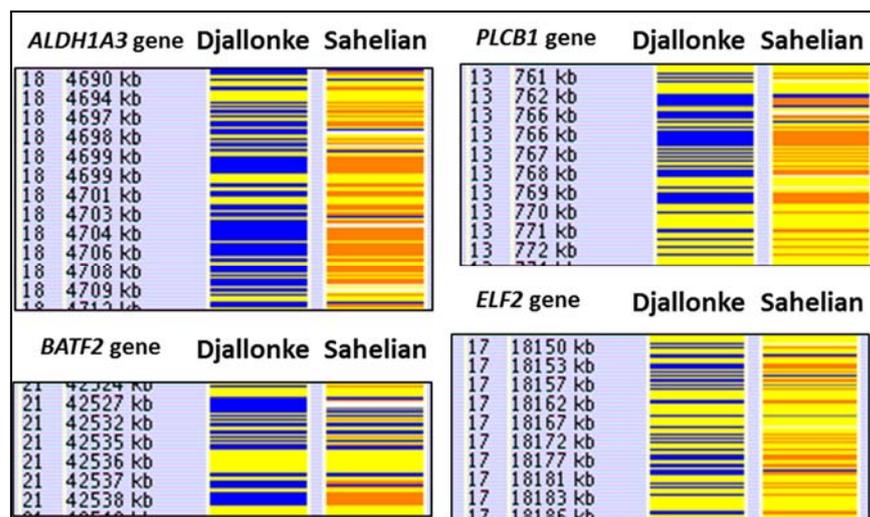


Figure 3.11 HomSI analysis adjacent to *ALDH1A3*, *PLCB1*, *BATF2* & *ELF2* genes showing contrasting sweeps in Djallonke (higher homozygosity) and Sahelian

**Table 3.1 Trypanotolerance candidate genes co-localised with signatures of selection in Djallonke and Sahelian sheep**

Sheep Breed	Chr	Genomic	Candidate genes	Orthologous loci		Reference
		coordinates(bp) v3.1		in Cattle	UMD v 3.1	
DJ	1	99,623,159- 99,650,475	<i>CTSS</i>	3:20,024,302 20,047,228	-	(O'Gorman et al., 2009)
DJ	2	165,154,368- 165,495,812	<i>ARHGAP15</i>	2:53,065,587 53,732,838	-	(Noyes et al., 2011)
DJ	4	79,355,697- 79,366,930	<i>INHBA</i>	4:79,986,254 79,997,754	-	(Dayo et al., 2012)
Both	7	9,272,393- 9,359,760	<i>SCAMP1</i>	10:9,369,310 9,520,700	-	(O'Gorman et al., 2009)
Both	5	16,922,626- 16,924,761	<i>TICAM1</i>	7:20,547,964 20,550,264	-	(Noyes et al., 2011)
DJ	8	57,581,762- 57,631,158	<i>STX7</i>	9:71,381,757 71,455,585	-	(O'Gorman et al., 2009)
DJ	17	62,105,682- 62,115,444	<i>RAB35</i>	17:64,724,244 64,742,928	-	(O'Gorman et al., 2009)
DJ	24	25,905,178- 25,915,071	<i>CD19</i>	25:164,039 26,169,956	-	(O'Gorman et al., 2009)

### 3.3.3 Genomic Evidence for resistance to Haemonchosis in Djallonke and Sahelian sheep

There were 25 genomic regions of extended runs of homozygosity observed in this study that co-localised with previously reported 25 candidate genes for Haemonchosis resistance and for other gastrointestinal parasites (Table 3.2). Three of these 25 genes (*IL20RA*, *PIK3CD*, *RELN*) have also been associated with resistance to other gastrointestinal nematodes such as

*Trichostrongyle* species, *Teladorsagia circumcincta* and other *Nematodirus* species (Benavides et al., 2015; Periasamy et al., 2014). Twenty-two of these regions of extended runs of homozygosity were observed only in Djallonke sheep (Table 3.2). The remaining three regions co-localised with candidate three genes (*PIK3CD*, *MUC15* & *IL17RB*) were observed in both the Djallonke and Sahelian breeds.

**Table 3.2 QTL for Resistance to Haemonchosis and other GIT parasites co-localised with signatures of selection in Djallonke and Sahelian sheep**

Sheep breed	Chr.	Genomic coordinate	Candidate Gene	Trait Inference	Reference
DJ	20	25,400,738-25,402,966	<i>MHCII-DRB1</i>	GIT parasites	(Schwaiger et al., 1995)
DJ	1	27,524,283-27,601,025	<i>LRP8</i>	<i>H. contortus</i>	(Benavides et al., 2016)
DJ	1	87,657,990-87,674,113	<i>DENND2D</i>	<i>H. contortus</i>	(McRae et al., 2014)
DJ	1	87,710,082-87,728,410	<i>CHI3L2</i>	<i>H. contortus</i>	(McRae et al., 2014)
DJ	1	87,788,905-87,811,255	<i>CHIA</i>	<i>H. contortus</i>	(McRae et al., 2014)
DJ	3	151,527,165-151,535,188	<i>IFNG</i>	GIT Nematodes	(Coltman et al., 2001)
DJ	3	123,851,175-125,982,479	<i>ATP2B1</i>	<i>H. contortus</i>	(Benavides et al., 2016)
DJ	8	62,006,022-62,039,859	<i>IL20RA</i>	<i>H. contortus</i> + others	(Periasamy et al., 2014)
Both	12	41,923,922-41,973,979	<i>PIK3CD</i>	<i>H. contortus</i> + others	(Periasamy et al., 2014)
DJ	12	62,163,057-62,283,591	<i>LAMC1</i>	<i>H. contortus</i>	(Benavides et al., 2016)
Both	15	55,404,807-55,417,235	<i>MUC15</i>	<i>H. contortus</i>	(Benavides et al., 2016)
DJ	17	52,168,248-52,191,479	<i>ABCB9</i>	<i>H. contortus</i> , homeostasis	(Yang et al., 2015)
DJ	10	11,465,154-11,505,274	<i>SUGT1</i>	<i>H. contortus</i> , homeostasis	(Yang et al., 2015)
DJ	14	48,062,306-48,071,138	<i>PAK4</i>	<i>H. contortus</i> , homeostasis	(Yang et al., 2015)
DJ	14	14,206,784-14,215,316	<i>FCER2</i>	<i>H. contortus</i> , homeostasis	(Yang et al., 2015)
DJ	4	44,668,137-45,205,142	<i>RELN</i>	<i>H. contortus</i> , homeostasis	(McRae et al., 2014)
DJ	6	89,053,717-89,061,196	<i>AREG</i>	<i>H. contortus</i>	(Guo et al., 2016)
DJ	11	57,796,766-57,800,093	<i>SOX9</i>	<i>H. contortus</i>	(Benavides et al., 2016)
DJ	6	70,189,729-70,234,612	<i>KIT</i>	<i>H. contortus</i>	(Guo et al., 2016)

Sheep breed	Chr.	Genomic coordinate	Candidate Gene	Trait Inference	Reference
DJ	16	66,450,575-66,482,861	<i>NSUN2</i>	<i>H. contortus</i>	(McRae et al., 2014)
Both	19	47,044,394-47,059,322	<i>IL17RB</i>	<i>H. contortus</i>	(Guo et al., 2016)
DJ	19	55,173,564-55,175,033	<i>HRH1</i>	<i>H. contortus</i>	(McRae et al., 2014)
DJ	25	44,535,996-44,543,183	<i>CXCL12</i>	<i>H. contortus</i>	(Guo et al., 2016)
DJ	19	53,290,059-53,291,096	<i>CXCR6</i>	<i>H. contortus</i>	(Guo et al., 2016)
DJ	20	34,055,201-34,055,659	<i>UBE2N</i>	<i>H. contortus</i>	(Benavides et al., 2016)

### 3.3.4 Genomic evidence for adaptation to tropical climatic conditions in the Djallonke and Sahelian sheep breeds

In addition to the regions of high homozygosity that have been previously observed to be associated with trypanotolerance and resistance to haemonchosis associated QTL, there were genomic regions showing high homozygosity that include genes shown to be involved in immune responses and natural adaptation (Table 3.3). A total of 37 previously reported adaptive candidate genes were observed to fall within these genomic regions of low heterozygosity in the Djallonke sheep dataset, including 13 that were also observed in the Sahelian sheep dataset (Table 3.3). HomSI analysis for the Djallonke and Sahelian sheep surrounding seven of these of these 37 QTL associated regions (*TRHDE*, *MSRB3*, *FGF5*, *ALH1A3*, *PLCB1*, *BATF2*, & *ELF2* genes) is shown on Figure 3.10 and Table 3.3. The remaining 30 regions for the Djallonke and Sahelian are shown in appendix C.2.

**Table 3.3 Candidate genes for adaptive selection co-localised with regions of high homozygosity in Djallonke and Sahelian sheep**

Sheep Breed	Chr.	Genomic coordinates	Candidate gene	Trait Inference	Reference
DJ	1	85,955,810-86,011,841	<i>GNAI3</i>	Melanogenesis (Thermo-tolerance)	(Kim et al., 2016)
DJ	1	188,388,916-188,441,236	<i>LMLN</i>	Melanogenesis	(Kim et al., 2016)
DJ	1	42,584,598-42,656,778	<i>IL12RB2</i>	Immune functions	(Roffler et al., 2016)
DJ	1	168,393,395-168,624,986	<i>ALCAM</i>	Immune functions	(Roffler et al., 2016)
DJ	1	121,075,675-121,168,117	<i>SYNJ1</i>	Phosphatidylinositol dephosphorylation	(Roffler et al., 2016)
Both	2	52,423,842-52,445,175	<i>NPR2</i>	Skeletal Morphology and body size	(Kijas et al., 2012)
DJ	4	9,433,282-9,465,962	<i>KRIT1</i>	Regulation of endothelial cell proliferation and migration	(Roffler et al., 2016)
DJ	4	85,316,865-85,381,180	<i>TSPAN12</i>	Regulation of signal transduction of cell surface receptors	(Roffler et al., 2016)
Both	5	41,256,802-41,272,546	<i>APC2</i>	Immune functions (Tumour suppressor)	(Roffler et al., 2016)
Both	6	36,514,210-36,556,824	<i>ABCG2</i>	Urea Metabolism (Homeostasis)	(Kijas et al., 2012)
DJ	6	94,584,400-94,605,575	<i>FGF5</i>	Regulation of fibroblast growth factor receptor	(Kijas et al., 2012)
Both	7	63,450,344-63,456,226	<i>BMP4</i>	Body size and development	(Kim et al., 2016)
DJ	3	204,447,104-204,461,390	<i>OLR1</i>	Internalization, degradation of oxidised low density lipoprotein by endothelial cells	(Roffler et al., 2016)
DJ	3	108,235,641-108,685,027	<i>TRHDE</i>	Regulation of appetite and digestion	(Kim et al., 2016)
DJ	3	154,219,234-154,397,986	<i>MSRB3</i>	Regulation of response to oxidative stress	(Kijas et al., 2012)
Both	3	35,907,955-36,031,445	<i>ALK</i>	Immune function (Protein phosphorylation)	(Kim et al., 2016)
Both	3	99,472,045-99,509,159	<i>IL1R1</i>	Immune function	(Kim et al., 2016)
DJ	9	54,817,997-54,825,977	<i>IL7</i>	Immune function	(Kim et al., 2016)
Both	10	36,838,524-36,858,872	<i>ATP12A</i>	Homeostasis (Potassium and Sodium)	(Kim et al., 2016)
DJ	10	40,800,056-40,821,770	<i>PCDH9</i>	Homophilic cell adhesion	(Kim et al., 2016)
Both	11	36,083,204-36,098,540	<i>PDK2</i>	Homeostasis	(Kijas et al., 2012)
Both	11	18,245,395-18,411,418	<i>NF1</i>	Homeostasis	(Kijas et al., 2012)

Sheep Breed	Chr.	Genomic coordinates	Candidate gene	Trait Inference	Reference
<b>Both</b>	13	78,815,423-78,893,076	<i>NFATC2</i>	Immune function	(Kijas et al., 2012)
<b>DJ</b>	13	666,266-1,154,524	<i>PLCB1</i>	Thermotolerance	(Kim et al., 2016)
<b>DJ</b>	15	45,551,281-45,552,222	<i>OR2AG1</i>	Response to stimulus	(Kijas et al., 2012)
<b>DJ</b>	16	38,969,273-39,028,126	<i>PRLR</i>	Reproduction	(Kijas et al., 2012)
<b>DJ</b>	17	18,131,831-18,226,233	<i>ELF2</i>	Regulation of transcription	(Kim et al., 2016)
<b>DJ</b>	17	29,240,707-29,257,289	<i>PGRMC2</i>	Reproduction	(Kim et al., 2016)
<b>DJ</b>	18	19,723,286-19,802,578	<i>ABHD2</i>	Wound healing	(Kijas et al., 2012)
<b>DJ</b>	18	4,690,980-4,728,935	<i>ALDH1A3</i>	Energy, digestive Metabolism	(Kim et al., 2016)
<b>Both</b>	18	38,107,388-38,110,333	<i>FOXG1</i>	Regulation of transcription	(Kijas et al., 2012)
<b>Both</b>	19	31,583,789-31,811,540	<i>MITF</i>	Melanogenesis	(Kijas et al., 2012)
<b>DJ</b>	19	7,255,507-7,331,066	<i>GLB1</i>	Cellular metabolism	(Kijas et al., 2012)
<b>DJ</b>	19	33,852,131-34,140,194	<i>SUCLG2</i>	Cellular metabolism	(Kim et al., 2016)
<b>DJ</b>	20	26,649,266-26,651,191	<i>HSPA1A</i>	Homeostasis	(Kadarmideen, Watson-Haigh, & Andronicos, 2011)
<b>Both</b>	21	49,011,232-49,012,130	<i>IFITM10</i>	Immune functions	(Roffler et al., 2016)
<b>DJ</b>	21	42,526,284-42,531,851	<i>BATF2</i>	Immune functions	(Roffler et al., 2016)

### 3.4 Discussion

#### 3.4.1 Signatures of Selection in Djallonke and Sahelian Sheep

The scan of the genome (using HomSI with a sliding window of 10,000 base pairs) permitted the identification of selection signatures in greater detail than other sliding window algorithms such as the “Integrated haplotype homozygosity score (iHS)” (Gautier et al., 2013; Lee et al., 2013) and “the composite of likelihood ratio (CLR)” statistics (Boitard et al., 2016; Qanbari et al., 2014). Furthermore, it has been shown that the iHS only detects “ongoing sweeps” and CLR only detects “completed sweeps” within the target genome (Qanbari et al., 2014). Additionally, potential selection sweeps identified using HomSI have a higher resolution (with sliding windows of 10 kb) in comparison with other methods (with sliding windows of 50kb

to 100kb) (Ai et al., 2015; Gautier & Naves, 2011; Kardos et al., 2015). This higher resolution facilitates a more detailed analysis of the selective sweep characteristics, and identification of the specific genomic coordinates for each swept region across the genome (Gormez et al., 2014). Conversely, large detection sliding windows are prone to an overestimation of the sweep outliers, resulting in increased false positives (Fuentes-Pardo & Ruzzante, 2017). However, the HomSI algorithm, cannot distinguish between completed and ongoing sweeps. Nevertheless, potential signatures of selection were identified in known QTL and non-QTL regions (genic and intergenic) within both breeds. Some of the non-QTL genic regions show similar sweep characteristics and these may explain some of the interesting and common traits of these two under-studied sheep breeds: e.g. the 20 kb and 79 kb swept regions co-localised with the *CDKALI* and *BTBD3* genes on chromosomes 20 and 13, respectively (Figure 3.2). Interestingly, the gene ontology terms for the *CDKALI* and *BTBD3* genes revealed a range of key biological and molecular functions including; dendrite morphogenesis, cerebral cortex development, as well as the maintenance of translational fidelity (Aken et al., 2016). Furthermore, there are three interesting regions of contrasting sweeps between the two sheep genomes at chromosome 20 that have close proximity (<1MB) to many known associated QTL for immune response (including the MHC genes) (Figure 3.4). In particular, one of these three swept regions (chromosome 20: 25,655,000-25,670,000) partially overlaps with the *DQA2* gene (chromosome 20: 25,666,699-25,676,792), and could be of potential relevance in understanding the genetic basis of important contrasting resistance to parasitic infection in these sheep breeds. This is because, in sheep, variations within the *DQA2* gene have been associated with resistance or susceptibility to gastrointestinal nematode infection (Hickford et al., 2011) and footrot infections (Ennen et al., 2009; Gelasakis et al., 2013). The remaining two swept regions on chromosome 20 (20: 46,021,000-46,027,000 & 20: 43,948,000-43,949,000) are located within approximately 1Mb of the *SLC35B3* gene (20: 46,139,978-46,160,425) and the *MAK-201* gene (20: 44,348,622-44,349,634), respectively, and their relevance is not immediately apparent. Although in general, LD decay is reported to be high over shorter physical distances in sheep (< 300 kb) than in other domesticates (Al-Mamun et al., 2015; Kijas et al., 2014), LD has been found to persist over longer physical distances (1 mb to 50 mb) in some sheep breeds (Chitneedi et al., 2017; García-Gómez et al., 2012). Furthermore, it is expected that mutations within these swept regions, which are a consequence of natural selection over time, will most likely be beneficial rather than deleterious due to the purifying effect of natural selection (Andersson, 2001; Gibson, 2012).

### 3.4.2 Genomic evidence for trypanotolerance in Djallonke and Sahelian sheep

Of the eight candidate trypanotolerance genes observed to be within selective sweeps in this study (Table 3.1), five (*STX7*, *SCAMPI*, *RAB35*, *CD19*, *CTSS*) were also previously identified in a gene expression study using peripheral blood mononuclear cells from trypanotolerant Ndama cattle experimentally infected with the parasite (O'Gorman et al., 2009). Whereas four of these five candidate genes were shown to be co-localised within selection signatures for the Djallonke genome only (Figure 3.5 Figure 3.6 & Figure 3.7), the fifth (*SCAMPI*), was co-localised with sweeps in both sheep breeds (Figure 3.5). The similar sweep observation at the *SCAMPI* region could be explained by two different hypotheses. Firstly, the Sahelian genomes might also be undergoing trypanosome mediated selection, albeit to a lesser degree in comparison to the Djallonke, and this is consistent with the theory of pathogen mediated selection (Hedrick, 1998). Secondly, this observation could also be attributed to a degree of historical admixture between these two breeds as indicated by a recent mitogenomic study (Brahi et al., 2015). Indeed, a recent study of trypanotolerant and non trypanotolerant cattle breeds from West Africa has shown significant admixture within trypanotolerance loci (Smetko et al., 2015). Furthermore, the currently reported trypanotolerant associated QTL in the animal QTL database cannot be considered to be exhaustive in the light of the limitations of the detection methods used, and given the peculiar genetics of complex quantitative traits (as reviewed in chapter 1.2). In particular, trypanotolerance is thought to be a complex polygenic trait (Gautier et al., 2009; Hanotte et al., 2003; Noyes et al., 2011). QTL's identified in the past could suffer from the "Beavis effect" due to an over-estimation of the effect of a few significant markers, resulting in a lowered repeatability of results (Meuwissen et al., 2016). This effect is similar to the "missing heritability" phenomenon encountered in the early mapping of disease associated variants in humans (Eichler et al., 2010; Zuk et al., 2012). This limitation was generalised for most of the early quantitative trait mapping studies performed on livestock, and is the main reason why these studies have not led to widespread applications such as marker assisted selection (Crawford et al., 2006; Meuwissen et al., 2016). However, the continuous advancement in molecular tools over recent years offers increased opportunities to refine these reported QTL, and then use them for agricultural applications. For example, the high resolution of genomic sweeps characteristic of HomSI outputs observed in this study will facilitate higher precision QTL mapping in structured association studies as demonstrated by Gormez et al. (2014).

Interestingly, a sixth candidate trypanotolerance gene, the *INHBA* gene (Figure 3.5), was found within swept regions but only within the Djallonke dataset. It is the most significant potential trypanotolerance QTL in the Animal QTL database so far identified (Hu et al., 2016). The *INHBA* gene was previously identified through a fine mapping analysis of four *a priori* trypanotolerance associated QTLs in 360 Ndama cattle infected under natural infection conditions (Dayo et al., 2012). The *INHBA* gene regulates the differentiation of hematopoietic cells in mammals (Bult et al., 2015; Eppig et al., 2015; Johansson & Wiles, 1995; Smith et al., 2014). This is consistent with a hypothesised mechanism of trypanotolerance, and demonstrated that the trait is strongly associated with the host's capacity to efficiently control anaemia (Dayo et al., 2012; Naessens, 2006; Trail et al., 1991).

The last two of the eight trypanotolerance candidate genes that fall within the selection regions identified in this study are *ARHGAP15* and *TICAM1* (Figure 3.5 & Table 3.1). These two genes were also previously identified in a combined transcriptomic and selective sweep study of infected trypanotolerant Ndama and Boran cattle (Noyes et al., 2011). Both genes co-localise with signatures of selection within the Djallonke dataset (Figure 3.5 & Table 3.1), whereas only the *TICAM1* gene (Figure 3.5) was observed within the selection signatures identified within the Sahelian dataset.

Most of the previously reported trypanotolerance associated QTL studies have used a lower density of molecular markers than was used in this study (Dayo et al., 2012; Dayo et al., 2009; Hanotte et al., 2003; O'Gorman et al., 2009), and therefore are of lower resolution, so the sizes of the reported QTL are quite large and not always suitable for downstream molecular applications. However, the very high density of genomic SNPs in this study permitted further insight into the genomic architecture within these previously reported trypanotolerance associated QTL. Comparative analysis of sweep characteristics of the Djallonke and Sahelian datasets for some of the previously reported trypanotolerance associated QTL revealed several co-localised selective sweeps of varying sizes (as small as 2 Kilo base resolution). This resolution permits detailed analysis of the swept regions that include shared, unshared or completely contrasting sweeps between the two genomes. This revelation supports an assertion that previously reported QTL require refinement to enhance the identification of actual causative mutations for this trait. Furthermore, although trypanotolerance is a complex quantitative trait, and is controlled by many genes, it is highly unlikely that all of the variants captured in these regions are causative variants. It is more likely that some variants are in linkage disequilibrium (LD) with the actual causal variants, and are selected by the hitch-hiking effect of the positive selection sweep over time (Qanbari et al., 2014). The high resolution of

sweep regions suggests that fine-mapping of causative variants could be possible. The identification of causative mutations will simplify a wide variety of downstream applications such as marker assisted breeding or genomic selection. We hypothesise that identification of causative trypanotolerance variants will require a more comprehensive approach with a larger number of animals than used in this study, and one that integrates these results with phenotypic data (such as actual infection state) from the two breeds. Other current genomic techniques that could complement this approach include Genotyping by Sequencing (GBS) (of *a priori* identified candidate genes) (Dodds et al., 2015; Zhan et al., 2011), as well as highly sensitive epigenomic analysis such as ATAC-seq (Buenrostro et al., 2015).

### **3.4.3 Genomic Evidence for resistance to Haemonchosis in Djallonke and Sahelian sheep**

Gene ontology terms analysis revealed that the 25 haemonchosis and other GIT parasite associated QTL regions (Table 3.2) are involved in multiple molecular functions and biological processes, such as immune response and chemotaxis (*IL20RA*, *IL17RB*, *FCER2*, *HRH1*), response to pain and tissue homeostasis (*RELN*, *SOX9*), and protein coding, binding, methylation and phosphorylation (*ATP2B1*, *SOX9*, *MUC15*, *UBE2N*, *LRP8*, *RELN*, *NSUN2*, *LAMC1*, *ABCB9*, *PIK3CD*, *SUGTI*, *PAK4*) (Cunningham et al., 2015; Yates et al., 2016). Other functions of the identified candidate genes include calcium binding and transport (*LRP8*, *LAMC1*) and carbohydrate metabolism (*CHI3L2*, *CHIA*) (Cunningham et al., 2015; Yates et al., 2016).

Six of the genes that fell within the selection regions identified in this study (*LRP8*, *ATP2B1*, *LAMC1*, *SOX9*, *MUC15*, *UBE2N*) were among the genes associated with resistance to *H. contortus* infection in a recent GWAS study of a backcross population of Red Maasai and Dorper sheep under natural infection conditions (Benavides et al., 2015). A further six genes (*CHI3L2*, *CHIA*, *DENND2D*, *RELN*, *NSUN2*, and *HRH1*) were among the top 1% of candidate genes for resistance and susceptibility to gastrointestinal nematodes in divergent populations of Romney and Perendale sheep breeds (McRae et al., 2014). Two of the genes (*IL20RA*, *PIK3CD*) are associated with resistance to experimental challenge with *H. contortus* (Periasamy et al., 2014). In a gene expression study using deliberately infected Chinese Hu sheep, four of the genes (*ABCB9*, *SUGTI*, *PAK4*, *FCER2*) were found to contribute to the key immunological response (Yang et al., 2015). More recently, five of the genes (*AREG*, *KIT*, *IL17RB*, *CXCL12*, *CLCR6*)(Table 3.2) were also found to be expressed in *H. contortus* resistant Canarian hair sheep (Guo et al., 2016). A recent study has also attributed the inherent resistance

to gastrointestinal parasites in the Caribbean hair sheep to an historical admixture with the Djallonke sheep from West Africa (Spangler et al., 2017).

#### **3.4.4 Genomic evidence for adaptation to tropical climatic conditions in Djallonke and Sahelian sheep breeds**

The 37 genome regions identified to have low heterozygosity, and which lie within regions previously shown to be associated with adaptive responses (Table 3.3) contained the following genes involved in immune functions (e.g. *IL12RB2*, *ALCAM*, *APC2*, *IL1R*, *IL7*), homeostasis (e.g. *HSPA1A*, *ATP12A*, *PDK2*, *NF1*, *ABCG2*), melanogenesis/ thermotolerance (*GNAI3*, *LMLN*, *PLB1*, *MITF*) and cellular and digestive metabolism (*GLB1*, *SUCLG2*, *TRHDE*, *OLR1*) (Kijas et al., 2012; Kim et al., 2016; Roffler et al., 2016). These could be due to chance or as a consequence of natural selection for resistance to disease, heat tolerance and the ability to exist on poor quality diets within the harsh, hot and humid climatic conditions faced by these sheep breeds.

Twelve of the 37 regions (*NPR2*, *ABCG2*, *FGF5*, *MSRB3*, *PDK2*, *NF1*, *NFATC2*, *OR2AG1*, *PRLR*, *ABHD2*, *MITF*, *GLB1*) were also reported in the top 0.1% of candidate genes identified in a genome-wide study for signatures of recent adaptive selection in 74 different sheep breeds that had been selected from various regions of the world, (Kijas et al., 2012).

Fourteen of the 37 genes (*GNAI3*, *LMLN*, *BMP4*, *TRHDE*, *ALK*, *IL1R1*, *IL7*, *ATP12A*, *PCDH9*, *PLCB1*, *ELF2*, *PGRMC2*, *ALDH1A3*, and *SUCLG2*) were also among those reported as possible candidate adaptive genes in the indigenous Egyptian sheep and goat breeds (Kim et al., 2016). The regions included genes for thermotolerance (*GNAI3*, *PLCB1*), energy and digestive metabolism (*TRHDE*, *ALDH1A3*), body size and development (*BMP4*) and immune responses (*IL1R1*, *IL7*) (Kim et al., 2016). Ghana and Egypt have a similar mean annual temperature of 28°C. In addition, the main sheep production systems in Ghana and Egypt have been the “low input extensive” type, with virtually no artificial selection for any specific production trait. Typically, sheep are grazed on low quality fodder or crop residues and are at the mercy of the harsh climate throughout the year. Therefore, the Djallonke and Sahelian sheep breeds in Ghana are likely to have remained quite primitive and are largely unimproved. Hence the shared adaptive candidate genes possibly under selection in these livestock breeds are likely due to natural adaption to similar environmental selective forces that prevail within these two countries.

More recently, in a study of natural local environmental adaptation, 10 of the genes within the sweep regions in this study (*IL12RBB2*, *ALCAM*, *SYNJI*, *KRIT1*, *TSPAN12*, *APC*, *OLR1*,

*IFITM10*, and *BATF2*) were reported as being important for adaptation in Dall sheep (*Ovis dalli dalli*) (Roffler et al., 2016). Roffler and colleagues' approach combined targeted resequencing of *a priori* identified candidate adaptive genes of immunity and metabolism in domestic sheep (*Ovis aries*) and bighorn sheep (*Ovis canadensis*) to develop a panel of SNP markers (Roffler et al., 2016). The SNP markers were then used to test a population of 476 Dall sheep using three different methods, including SNP chip genotyping to identify the candidate genes. Similar to the case of the Djallonke sheep, the Dall sheep had been subjected to many centuries of natural selection in the wild with limited human intervention. In contrast to the tropical climatic conditions for Djallonke and Sahelian sheep, the Dall sheep breed evolved under Arctic and sub Arctic climatic challenges, which explains why the common swept regions contained genes mainly associated with immune function and not heat tolerance (Table 3.3).

Some of the suggested common signatures of selection observed between the Djallonke and Sahelian sheep in this study (Table 3.3) could be attributed to exposure to similar environmental selection pressures over several centuries. Historical admixture has been reported in Djallonke sheep populations in different regions of SSA (Alvarez, Traore, et al., 2012; Alvarez et al., 2009; Brahi et al., 2015; Wafula et al., 2005). Brahi and colleagues (2015) identified historical admixture between the Djallonke and Sahelian sheep breeds using mitogenomic analysis. The high number of shared autosomal SNPs (~10.8 million) between the Djallonke and Sahelian sheep breeds observed in this study supports this earlier report. In contrast, both breeds differ by more than 10 million autosomal SNPs when compared to the reference assembly (Oar v3.1) derived from the Texel breed, which has a different demographic history to the two sheep breeds in this study (The International Sheep Genomics et al., 2010).

### **3.5 Conclusions**

This study provides the first support to putative regions harbouring possible QTL for trypanotolerance, resistance to *H. contortus* infection and adaptation to a harsh tropical climate in these two sheep breeds. The genomic evidence of trypanotolerance, inferred from conserved orthologues of trypanotolerant N'Dama cattle, showcases evidence of similar adaptive selection response for a common disease in two different ruminant species. These similar effects observed in common gene regions of both cattle and sheep strongly suggest that the effects are real, and an expanded study for both species will help to refine these results. Given that both the Djallonke and Sahelian breeds are of high socioeconomic importance in more

than 14 countries within the SSA, a regionally coordinated action plan to preserve their genetic diversity is needed for their sustainable utilisation.

## **Chapter 4    Molecular characterisation of Ovar-MHC class II region reveals novel alleles in the Djallonke and Sahelian sheep breeds of Ghana**

*This chapter presents a comparative analysis of the MHC class IIa genes in a population of 100 Djallonke and 100 Sahelian sheep. Variation within the genes from the class IIa region in sheep have previously been associated with host resistance to parasites. Historically, the Djallonke sheep is known to be more resistant to parasitic infection compared to the Sahelian. These contrasting resistance traits have been demonstrated via many parasitological and haematological studies, but no genetic analysis of the MHC region for either breed has been performed. Therefore, the rationale for this study was to analyse and determine, for the first time, the allelic variations within some MHC class II genes in these two breeds. Sequencing-based genotyping of three polymorphic class II genes in the two sheep populations revealed high numbers of new alleles at two of these loci. Evolutionary distance analysis shows that the Djallonke has higher allele diversity at all three loci studied. It has been suggested that higher diversity at the class II region is associated with increased resistance to parasitic infection. This analysis presents an initial characterisation of the class II MHC genes in these two breeds, and provides valuable baseline information for future association studies in these breeds with respect to parasitic diseases. Furthermore, the finding of alleles in one breed that are not present the other gives a possibility of a breed identification system and, when applied alongside information described from whole genome analysis, will enable molecular breed delineation. The information presented also offers new research opportunities in these breeds such as the application of molecular tools for the sustainable management of these two important indigenous sheep breeds.*

### **4.1    Introduction**

The Djallonke and the Sahelian sheep breeds are of high socio-economic and cultural significance in at least 14 countries within the sub Saharan Africa (SSA) (Geerts et al., 2009; Muigai & Hanotte, 2013). The Djallonke sheep breed is the most ancient African sheep breed, and is known for resistance against most livestock diseases in the region, particularly African Animal trypanosomiasis and haemonchosis (Geerts et al., 2001; Goossens et al., 1997; Goossens et al., 1999). In contrast, the Sahelian breed lacks this disease resistance

characteristic, but has a body size about twice that of the Djallonke. This dichotomy has been the basis of conflicting interests for small holder livestock farmers within the SSA region. The conflicting interests resulted in a long history of indiscriminate crossbreeding, and its consequent excessive genetic introgression is threatening the unique diversity of most African indigenous livestock breeds (Mwai et al., 2015). For example, significant levels of admixture have recently been reported in the candidate regions for trypanosomiasis in trypanotolerant and non trypanotolerant indigenous African cattle breeds (Smetko et al., 2015). In the light of these challenges, it has become imperative to characterise not only the Djallonke sheep, but also the Sahelian sheep at the molecular level as a fundamental step for sustainable breeding and conservation of the two breeds. This is because molecular characterisation is fundamental to the elucidation of the underlying genetic basis of the unique characteristics of the breeds, including the differential resistance to disease. In SSA, two major diseases that are known to significantly affect ruminant livestock are African animal trypanosomiasis and nematode infections (Reviewed in Chapter 1.2). Nematode infections and trypanosomiasis have remained the main disease challenge to livestock due to multiple factors discussed in the literature review. The predominantly free range (pastoral, agro-pastoral & integrated crop-livestock) systems of production (Kosgey & Okeyo, 2007; Tano et al., 2003; Teweldmehidin & Conroy, 2010), also means that livestock, including sheep, are continuously exposed to nematode infection. Furthermore, due to the agrarian nature of the farming systems, farmers lack the capacity to manage complex diseases including haemonchosis (Geerts et al., 2001; Murray et al., 1984). Exploring the natural resistance to diseases, including nematode infection, is compatible with the low input type of farming systems, and will provide a more economically sustainable control option for the disease control. Although the challenges of alternative nematode control measures in livestock, such as resistance to anti-helminthic drugs, have become more apparent, such alternatives are neither available nor affordable to the majority of resource poor smallholder farmers in SSA. As reviewed in chapter 1.2, breeding for enhanced disease resistance in livestock is an established concept, and has been well documented in cattle and sheep globally.

Variation within MHC class II genes has been associated with resistance to parasitic diseases in sheep (Davies et al., 2006; Hassan et al., 2011; Hickford et al., 2007). The sheep class II region has been shown to be subdivided into class IIa and class IIb (Lee et al., 2012). Several genes in the MHC class IIa region in sheep encode glycoprotein molecules that present exogenous and endogenous antigens to circulating CD4<sup>+</sup> T cells to elicit immune responses

(Dukkipati et al., 2006; Fremont et al., 1996). The *DRB1*, *DQA2* and *DQA1* MHC class IIa genes are known to be hypervariable in sheep (Dukkipati et al., 2006; Herrmann-Hoesing et al., 2008). Furthermore, within each of these class II MHC loci, the second exon has been reported to be the most polymorphic, making these loci the primary region of interest in MHC class II studies (Ballingall et al., 2010; Konnai, Nagaoka, et al., 2003).

In this context, the complete exon II for each of the *DRB1*, *DQA2* and *DQA1* loci were genotyped in populations of Djallonke and Sahelian sheep with the main aim of characterising the allele diversity at the three loci. This is the first characterisation study of MHC class II genes in these two sheep breeds, and the information gained will be valuable for future association studies with respect to MHC determined susceptibility or resistance to disease.

## **4.2 Methods**

### **4.2.1 Blood collection and white blood cell isolation**

Approximately 9 mL blood was sampled from 100 Djallonke sheep and 100 Sahelian from the same populations described in section 2.2.1 of chapter 2, using the same blood cell isolation protocol described in section 2.2.2.

### **4.2.2 Purification of sheep genomic DNA from buffy coats**

Genomic DNA was extracted from sheep buffy coats (prepared as described in chapter 2.2) using the Zymo Quick-gDNA<sup>TM</sup> MiniPrep KIT (Catalogue number D3025; Zymo Research Corporation, Australia) according to the manufacturer's protocol, with one modification aimed at increasing the yield of DNA. For each sample, 200  $\mu$ L of buffy coat (instead of 100  $\mu$ L as prescribed by the manufacturer) was mixed with 400  $\mu$ L genomic lysis buffer and left overnight (instead of 10 minutes as prescribed by the manufacturer) at room temperature to allow for efficient lysis of blood cells. The rest of the extraction procedure was done according to the manufacturer's protocol. The quantity and quality of extracted DNA was analysed using the NanoDrop 1000 spectrophotometer version 3.1.8 (Thermo SCIENTIFIC), per the manufacturer's guidelines. The purified DNA was stored at -20°C.

### **4.2.3 Preparation of Media, Buffers and solutions**

The methods used to prepare isopropyl- $\beta$ -D thiogalactopyranoside (IPTG), 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside (X-GAL), ampicillin sodium salt, super optimal broth with catabolite repression (SOC) medium, TAE buffer, and Luria Bertani (LB) agar and broth medium are shown in Appendix Table B.1. X-GAL, ampicillin and IPTG were used to identify transformed *E coli* cells via blue-white colony screening. LB broth was used for sub-culturing individual transformed competent cells from single colonies. TAE at 1X concentration was used as the buffer for agarose gel electrophoresis.

#### **4.2.3.1 Preparation of LB agar plates with Ampicillin, IPTG and X-GAL**

Agar plates were prepared under aseptic conditions within a class II biosafety cabinet (Catalogue number 212524, GELAIRE, AUSTRALA). Approximately 15 ml of warm molten LB agar was poured into a standard plastic petri-dish, and allowed to solidify with the UV light on. When solidified, 120  $\mu$ l of master mix, composed of 40  $\mu$ l ampicillin (50 mg/mL), 20  $\mu$ l of IPTG (50 mg/mL) and 60  $\mu$ l of X-GAL (20 mg/mL) was spread on the surface of each agar plate using a sterile Lazy-L spreader (catalogue number Z376779PAK, SIGMA ALDRICH, Australia). The plates were dried briefly at room temperature, then stored at 4°C.

#### **4.2.4 Agarose gel electrophoresis**

Purified genomic DNA, plasmid DNA and PCR amplicons were visualised by agarose gel electrophoresis to assess for degradation, transformed plasmid and amplification, respectively. All electrophoresis was performed using a Bio Rad PowerPac basic electrophoresis system as described by Brody et al. (2004). Briefly, 1% agarose (Bioline, Australia; for genomic and plasmid DNA) or 2% agarose (for PCR amplicons) with Gel Green nucleic acid stain (Fisher Biotec, Australia) and a final concentration of 1X Bromophenol blue DNA loading buffer (Catalogue number: R0611, ThermoFisher Scientific, Australia) was used. A 100-base pair ladder (Promega, USA) was used for quantification of all PCR amplicons because the product sizes ranged between 250 bp to 1 kb. Electrophoresis results were visualized using a UV Transilluminator (Fisher Biotec-Australia).

#### 4.2.5 Polymerase chain reaction (PCR)

PCR amplification was performed for the three MHC class IIa loci (*DRB1*, *DQA1* and *DQA2*) using genomic DNA isolated from Djallonke and Sahelian sheep white blood cells. The MHC primers used, as described by Hickford et al. (2004), Ballingall et al. (2010) and Ali et al. (2017) are listed in Table 4.1 (GeneWorks, Australia). MyTaq™ HS DNA Polymerase (catalogue number BIO-21111, Bionline, Australia) and the Applied Biosystems Thermocycler Veriti (Thermo Fisher Scientific) were used for all PCR reactions.

**Table 4.1 The Oligonucleotide(Oligo) primers used for genotyping the exon2 *DRB1*, *DQA1* and *DQA2* loci**

Primer name	Sequence 5' to 3'	Target region	Amplicon size (bp)
<i>DRB1</i> Forward	attagcctctccccaggagtc	Intron 1	368
<i>DRB1</i> Reverse	cacacacacaactgctccaca	Intron 2	
<i>DQA1</i> Forward	actggccacaaatgaagcccacaa	Intron 1	525
<i>DQA1</i> Reverse	agaaggcagaagatgagggttcag	Intron 2	
<i>DQA2</i> Forward	cacatgttacagtgcaaaarcage	Intron 1	831
<i>DQA2</i> Reverse	ccctcycaccaacgtttccag	Intron 2	

#### 4.2.6 Optimised PCR conditions for *DRB1*, *DQA1*, and *DQA2*

The optimised PCR cycling conditions for *DRB1*, *DQA1* (NiKDQA1) and *DQA2* are described in appendix Table B3. PCR cycling conditions were same for the three loci, except that 38 cycles at an annealing temperature of 62°C was used for the *DQA1* and *DQA2* compared to 36 cycles at 58°C for the *DRB1*. The optimised volumes of the MyTaq reaction buffer, primers and of the MyTaq Polymerase enzyme for *DQA1* and *DQA2* were the same as for the *DRB1* master-mix described in appendix Table B2. The strategy applied to determine the quantities of DNA and Ultrapure water required were the same as for the *DRB1* primers.

#### **4.2.7 Clean up of PCR amplicons**

PCR amplicons intended for direct sequencing were purified to remove excess dNTPs and unused primers prior to sequencing. The FavorPrep PCR clean-up kit (Catalogue number FAPCK 001-1; Favorgen Biotec Corporation, Fisher Biotec, Australia) and the EXOSAP PCR clean-up kit (BioLabs, New England, Australia) were used as per the manufacturers' guidelines and as described in sections (appendix B2 & B3). PCR amplicons to be cloned were not subject to this clean-up process.

#### **4.2.8 Ligation and Transformation of PCR amplicons**

##### **4.2.8.1 Ligation of PCR amplicons**

The pGEM-T Easy Vector System I (Promega Corporation, Australia) was used to clone PCR amplicons according to the manufacturer's protocol. Briefly, a 10  $\mu$ L ligation reaction mix containing 5  $\mu$ L of 2x rapid ligation buffer, 1  $\mu$ L of pGEM-T Easy Vector, 1  $\mu$ L of T4 DNA ligase, the required amount of each DNA amplicon based on a computed metrics using the Biomath calculator ([www.promega.com/biomath](http://www.promega.com/biomath)) was added, and where required, the remaining volume was made up to 10  $\mu$ L with ultrapure water. All reaction mixtures were prepared on ice. A positive control was prepared using 2  $\mu$ L of the pGEM-T control DNA insert provided, 5  $\mu$ L of 2x rapid ligation buffer, 1  $\mu$ L of the pGEM-T Easy vector, 1  $\mu$ L of T4 ligase and 2  $\mu$ L of ultrapure water. All reaction mixes were incubated at 4°C overnight.

##### **4.2.8.2 Transformation of Ligation products**

High efficiency competent *E. coli* cells (transformation efficiency approximately  $1 \times 10^8$  colony forming units per microgram of DNA) were used for the transformation reaction. A standard heat-shock transformation procedure was used as per the Promega transformation protocol. Briefly, plates of LB media agar with ampicillin, X-Gal and IPTG were prepared according to the method used by Inoue, Nojima, and Okayama (1990) and were pre-incubated at 37°C for 20 minutes prior to inoculation. Ligation reactions were centrifuged for approximately 5 seconds, then 2  $\mu$ L of each sample or 0.1 ng of uncut pGEM plasmid (positive control), was mixed with 50  $\mu$ L of *E. coli* competent cells (Promega, Australia) on ice for 20 minutes. Then the mixtures were heat-shocked for 45-50 seconds in a 42°C water bath, followed by incubation on ice for 2 minutes. Room temperature SOC broth (600-900  $\mu$ L) was added to all reactions. The mixtures were incubated at 37°C for 90 minutes with shaking at 150 rpm. Then 100  $\mu$ L of

each transformation mixture was plated on duplicate LB agar plates. The inoculated plates were incubated at 37°C for 18 hours.

#### **4.2.8.3 Screening, selection and purification of recombinant plasmid containing *E. coli* colonies**

For each plate of transformed colonies, 5 white colonies were separately sub-cultured in 5 mL of Luria Bertani (LB) media in 50 mL Falcon tubes, and incubated at 37°C with shaking at 180 revolutions per minute (rpm) for 18 hours on a multi-functional orbital shaker (Catalogue number PSU-20i; Fisher Biotec, Australia). For each sample, 1 mL of culture was aliquoted into a 1.5 mL graduated microfuge tube and centrifuged at 10,000 rcf for 60 seconds to separate the cells from the residual media. Supernatants were discarded, and any residual supernatant aspirated using a clean sterile pipette. Plasmids were then extracted from the cell pellets using a FavorPrep™ Plasmid extraction mini kit (catalogue number FAPDE 100; Favorgen Biotech Corp., Australia) using the manufacturer's protocol, except that ultrapure water was used for the DNA elution step. The quality and quantity of plasmids were checked with agarose gel electrophoresis and NanoDrop analyses as described in appendix B1.

#### **4.2.9 Sequencing of DNA amplicons and Plasmids**

All the plasmid DNA products and un-cloned PCR products were prepared according to the sequencing requirements of Macrogen Incorporated (South Korea) available at [http://dna.macrogen.com/eng/support/ces/guide/ces\\_sample\\_submission.jsp](http://dna.macrogen.com/eng/support/ces/guide/ces_sample_submission.jsp). Sequencing of PCR amplicons and of cloned amplicons was performed using 3730XL sequencers (Applied Biosystems) at Macrogen Incorporated. The pUC/M13 forward and reverse universal sequencing primers were used for sequencing all pGEM-T Easy Plasmid vector cloned amplicons, whereas the specific primers for template amplification were used for the sequencing of respective PCR amplicons.

##### **4.2.9.1 Analyses of sequenced PCR amplicon data**

Sequenced data from both cloned and un-cloned amplicons were classified using NCBI nucleotide blast search (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and the Immunopolymorphism database for major histocompatibility complex (IPD-MHC) ([www.ebi.ac.uk/ipd/mhc/](http://www.ebi.ac.uk/ipd/mhc/)). *De novo* assembly implemented in Geneious v8.1.2 was used to separately group the cloned nucleotide sequences for each of the three MHC class IIa genes

(Kearse et al., 2012). Grouped sequences for each of the MHC class II genes were manually examined using the Geneious software display and each sequence within a group subjected to nucleotide blast against the NCBI database to obtain the best GenBank matched accession number (Morgulis et al., 2008; Zhang et al., 2000). Annotation information of the best matched GenBank accession was then used to assign the sequences to specific alleles. The nucleotide sequences were translated to amino acid sequences and examined for variations. Translated amino acid sequences are referred to with the respective GenBank nucleotide accession numbers for clarity. New alleles were determined using the criteria set out in the IPD-MHC (Maccari et al., 2017; Robinson et al., 2015)

#### **4.2.10 Phylogenetic analysis**

Multiple sequence alignments of coding region for each locus were obtained separately using SEAVIEW software version 4.0. (Gouy, Guindon, & Gascuel, 2010) and MUSCLE (Edgar, 2004). Multiple sequence alignments for all *DRB1*, *DQA2* and *DQA2* exon 2 alleles were performed using the translation frame of the respective GenBank and IPD reference accessions. For each locus, a representative sequence per allele group and the respective standard NCBI accession as reference were included in the alignment. A second translation frame (codon\_start=3), presumed to be correct, was then used to re-translate the *DRB1*\*0401 allele group including an additional multiple alignment with remaining DRB1 alleles for comparison. Phylogenetic trees were constructed for each locus using three models of phylogeny namely; Maximum Likelihood (ML), Maximum Parsimony (MP) and Poisson's Neighbour-Joining Tree (PNJT) in MEGA7 (Jones, Taylor, & Thornton, 1992; Kumar, Stecher, & Tamura, 2016). Best protein model analysis for ML tree in MEGA was used to determine the best tree model for each locus. The tree model with lowest Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) scores was selected for each of the loci. Trees were also constructed in SEAVIEW software for comparison. To determine the reliability of branches in the trees, bootstrap analysis with 500 replicates was used as the default parameter in the constructions. All the trees were constructed to scale and the units of evolutionary distance were computed with the Poisson correction method (Zuckerkandl, 1965). Evolutionary distances for each sheep population were then computed from the NJT phylogeny.

#### **4.2.11 Genotype discovery curves**

To assess if the number of sheep sampled was adequate for finding all the genotypic diversity in the populations, genotype discovery curves were constructed for all three loci for the two populations. Plots of the cumulative number of new genotypes sampled for each locus (y-axis) and the total number of sampled sheep (x-axis) were constructed using Microsoft excel version 2010. The genotype discovery curves are based on a random sampling procedure that was used to sample from the sheep populations.

### **4.3 Results**

#### **4.3.1 Diversity of MHC Class IIa *DRB1*, *DQA1* and *DQA2* genes in Djallonke and Sahelian sheep of Ghana**

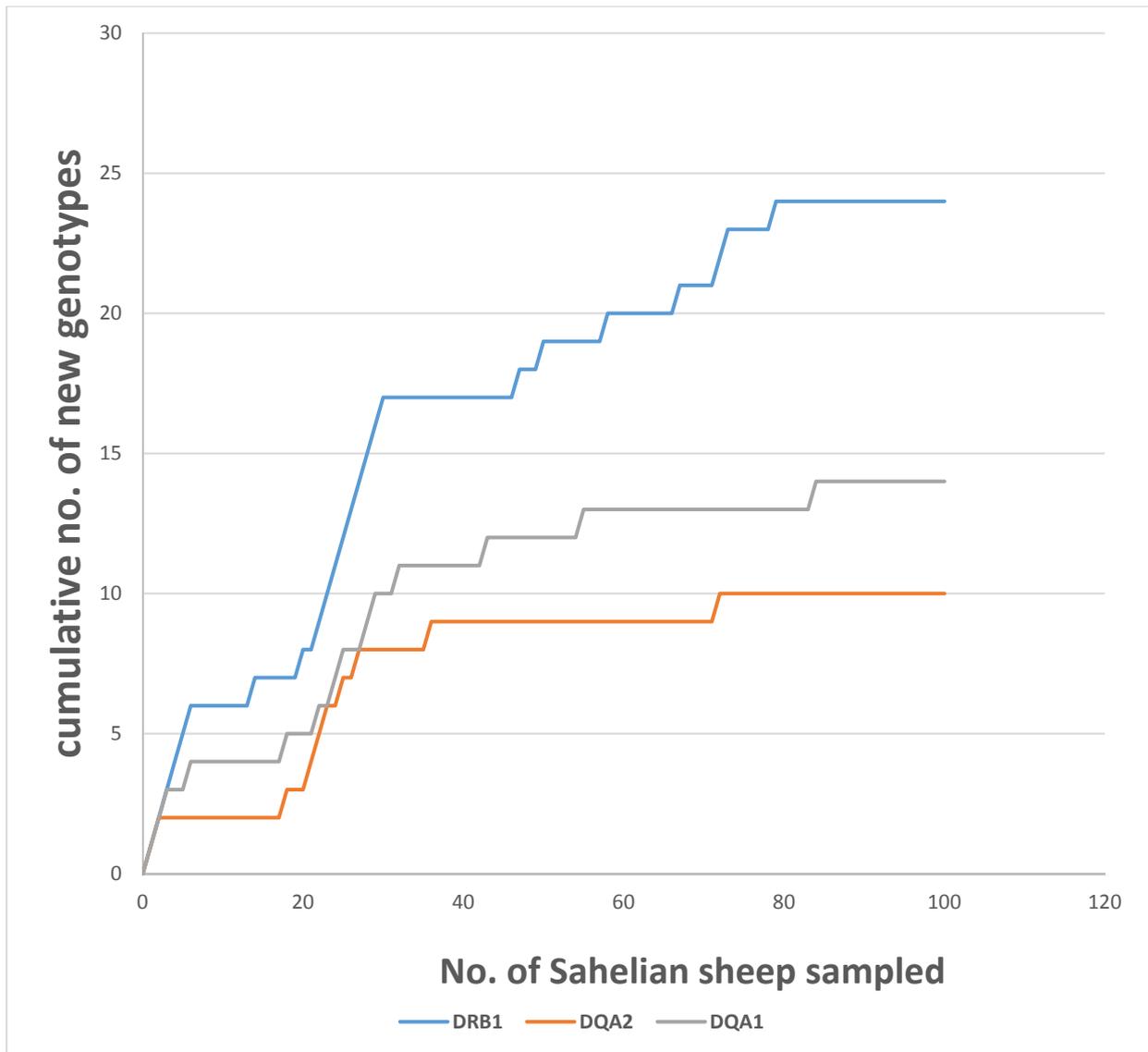
A total of 1,381 (913 clones) MHC class IIa curated nucleotide sequences were obtained, comprising 553 (380 clones) from *DRB1* (369bp), 336 (257 clones) from *DQA2* (821bp) and 492 (276 clones) from *DQA1* (525bp), from 100 Djallonke and 100 Sahelian sheep. Comparison of the putative sequences to the respective standard accessions from GenBank (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) and the MHC-IPD ([www.ebi.ac.uk/ipd/mhc/](http://www.ebi.ac.uk/ipd/mhc/)) revealed high numbers of sequence variations (nucleotides & amino acids) between the two breeds (Maccari et al., 2017; Robinson et al., 2015). On the basis of the numbers and positions of CDS mutations in the amino acid sequences, 85 putative unique *DRB1* alleles were identified, of which two alleles contained premature stop codons, and are therefore presumed to be non functional (Table 4.2). Similarly, 41 putative unique *DQA2* sequences were identified, although none contained a premature stop codon (Table 4.3). The *DQA1* sequences included 55 putative unique alleles, 52 of which were presumed functional, and the remaining 3 alleles had premature stop codons (Table 4.4).

##### **4.3.1.1 Availability of data**

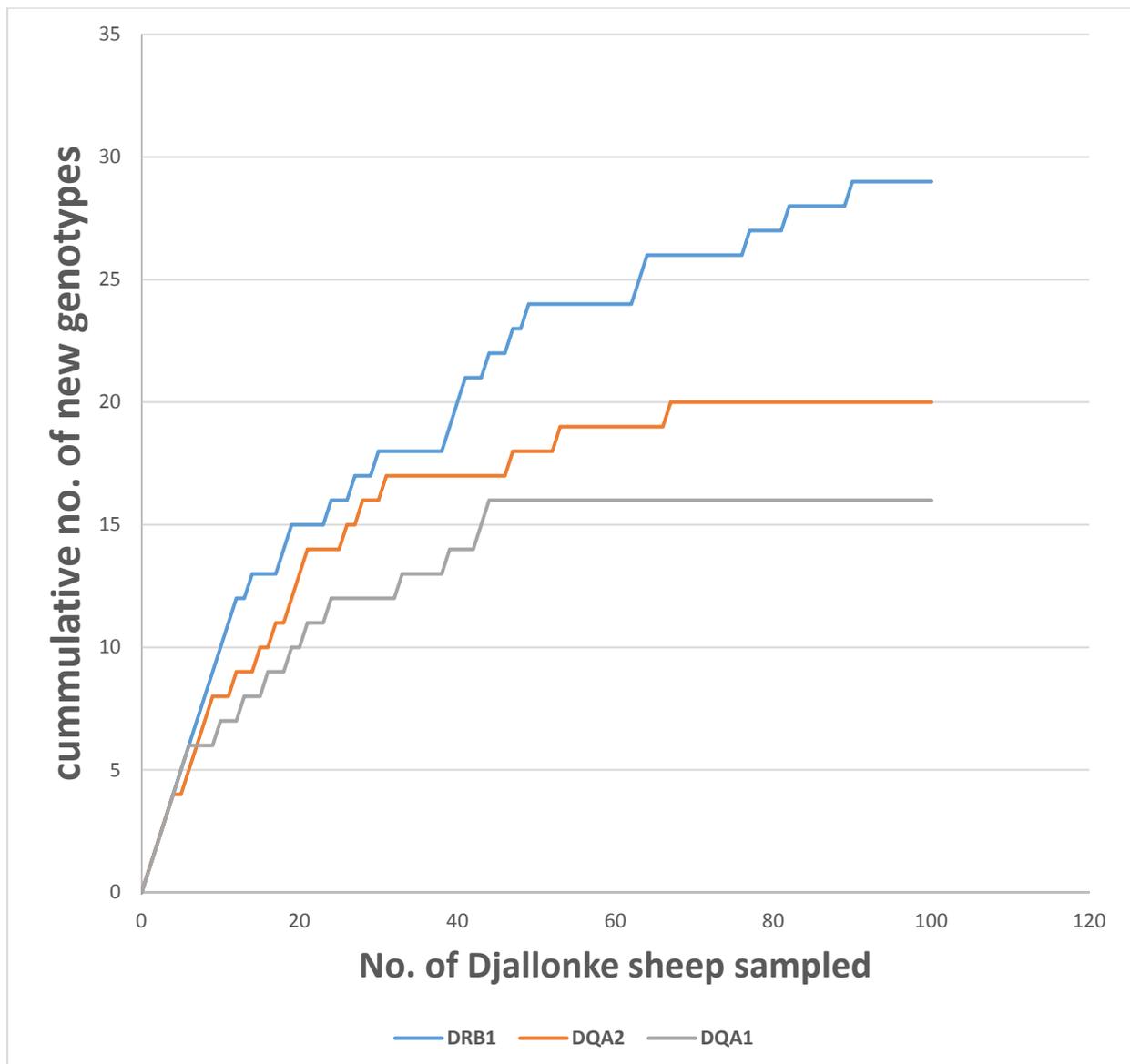
A total of 151 GenBank accession numbers comprising of a first batch of 72 GenBank accession numbers for *DRB1* alleles and second batch of 81 GenBank accession numbers for the *DRB1* alleles produced from this work are listed in appendix E

### 4.3.2 Genotype Discovery Curve for Sampling of Djallonke and Sahelian Sheep Populations

Having determined the genotypes for all the individual samples, the genotype data for the three loci *DRB1*, *DQA2* and *DQAI* were used to plot genotype discovery curves for the 100 Djallonke sheep (Figure 4.2) and the 100 Sahelian sheep (Figure 4.1). The curves constructed for all three loci from the two sheep populations show similar trends with an initial steep gradient that begins plateauing towards the end of the plot. The discovery curves for *DRB1* genotypes in both sheep breeds showed the steepest gradient. The curve for *DQAI* showed the lowest gradient for the Djallonke breed, but the one for *DQA2* had the lowest gradient for the Sahelian breed. The Djallonke *DRB1* curve begins to plateau at approximately the 95<sup>th</sup> sample compared to approximately the 80<sup>th</sup> sample for the Sahelian. Interestingly, the curves for *DQA2* plateaued at the same point (approximately the 70<sup>th</sup> sample) for both sheep breeds. The curves for *DQAI* showed greater between breed differences, with the Djallonke plateauing at the 45<sup>th</sup> sample compared to the 85<sup>th</sup> sample within the Sahelian population. With the exception of the Djallonke *DRB1* discovery curve, which plateaued at 95% of sample size, all of the discovery curves plateaued within 85% of the sample size studied.



**Figure 4.1** Genotype discovery curve for *DRB1*, *DQA2* and *DQA1* alleles for 100 randomly sampled individual Sahelian sheep



**Figure 4.2 Genotype discovery curve for *DRB1*, *DQA2* and *DQA1* alleles for 100 randomly sampled Djallonke sheep**

#### **4.3.2.1 Molecular phylogenetic re-construction for unique coding sequences at the *DRB1*, *DQA2* and *DQA1* loci**

Maximum Likelihood phylogenetic tree model test for unique functional *DRB1*, *DQA2* and *DQA1* amino acid sequences predicted the Jones Taylor & Thornton (ML JTT) amino acid substitution model with Gamma distribution as the best model for the three loci (Jones et al., 1992; Kumar et al., 2016). The ML JTT phylogeny for the combined unique functional sequences at each locus for the two sheep populations revealed the complexity of inter- and intra- population evolutionary relationship between the identified alleles (Figure 4.3, Figure 4.6 & Figure 4.9). Analysis of the combined loci phylogenetic trees show three patterns of

clustering; mixed clusters of the Djallonke and Sahelian, Djallonke-only clusters and Sahelian-only clusters. The population-specific ML JTT phylogeny at each locus more clearly demonstrates an evolutionary history of alleles for each population (Figure 4.4, Figure 4.5, Figure 4.7, Figure 4.8, Figure 4.10 & Figure 4.11). This is because, in the constructed phylogeny of the combined loci, unique alleles that were common for the two populations were only represented by one sequence in the tree. That step was taken to avoid redundancy of sequences so as to make the tree clearer.

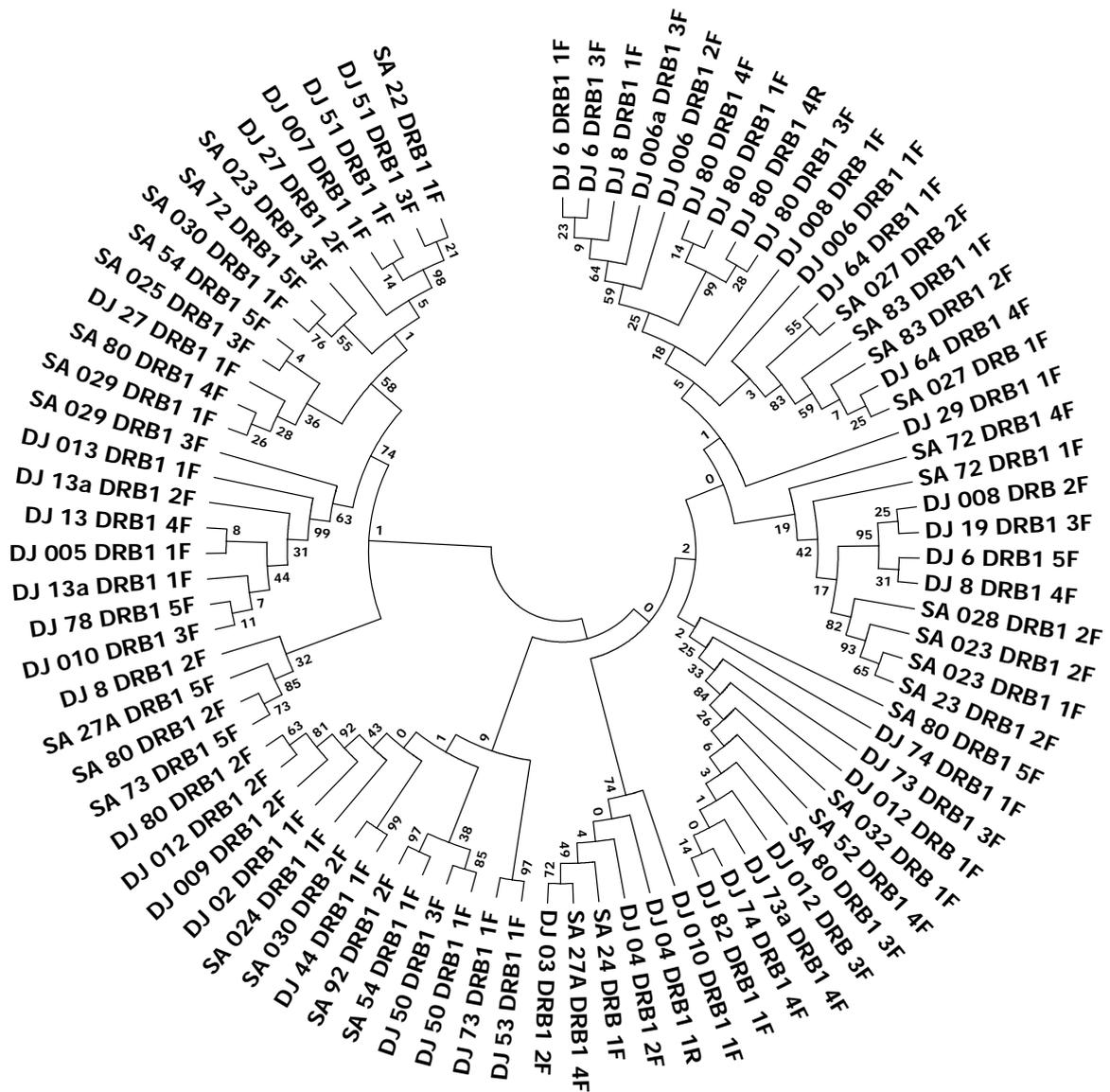
A re-constructed phylogeny based on the Neighbour Joining (NJ) method (Felsenstein, 1985; Jones et al., 1992; Kumar et al., 2016; Saitou & Nei, 1987) for the the same population-specific sequences showed that the clustering characteristic of alleles was generally conserved between the two models (ML JTT and ML NJT), albeit bootstrap values varied (Appendix D Figure 1 to Appendix D Figure 6). A total of 89, 82 and 82 amino acid substitution sites were observed in the *DRB1*, *DQA2* and *DQAI* datasets respectively, and used for the evolutionary distance analysis. Evolutionary distance analysis using the combined population functional loci revealed that the combined population unique *DRB1* exon 2 sequences were the most variable with a total branch length of approximately 1.90 units (Figure 4.3), followed by the *DQA2* locus with a total of 1.72 units (Figure 4.6), and the *DQAI* locus was the least variable with a total of 1.26 units (Figure 4.9). Population specific evolutionary distance analyses showed that the Djallonke population had higher variation at the *DRB1* locus, with approximate total branch length of approximately 1.62 units (Figure 4.4) in comparison to 1.36 units (Figure 4.5) for the Sahelian population. At the *DQA2* locus, the evolutionary distance for the Djallonke population was 1.44 units (Figure 4.7), again higher than the 0.81 units (Figure 4.8) obtained for the Sahelian population. The *DQAI* locus showed the least variation in both sheep populations with the Djallonke and Sahelian recording a total evolutionary distance of 0.97 units (Figure 4.10) and 0.77 units (Figure 4.11) respectively.

**Table 4.2 Unique *DRB1* exon 2 amino acid sequences identified in Djallonke and Sahelian sheep populations (Presumed functional coding sequences (n= 83) with non functional coding sequences containing premature stop codons highlighted in yellow (n =2); Prefix: DJ and SA denote Djallonke and Sahelian samples respectively)**

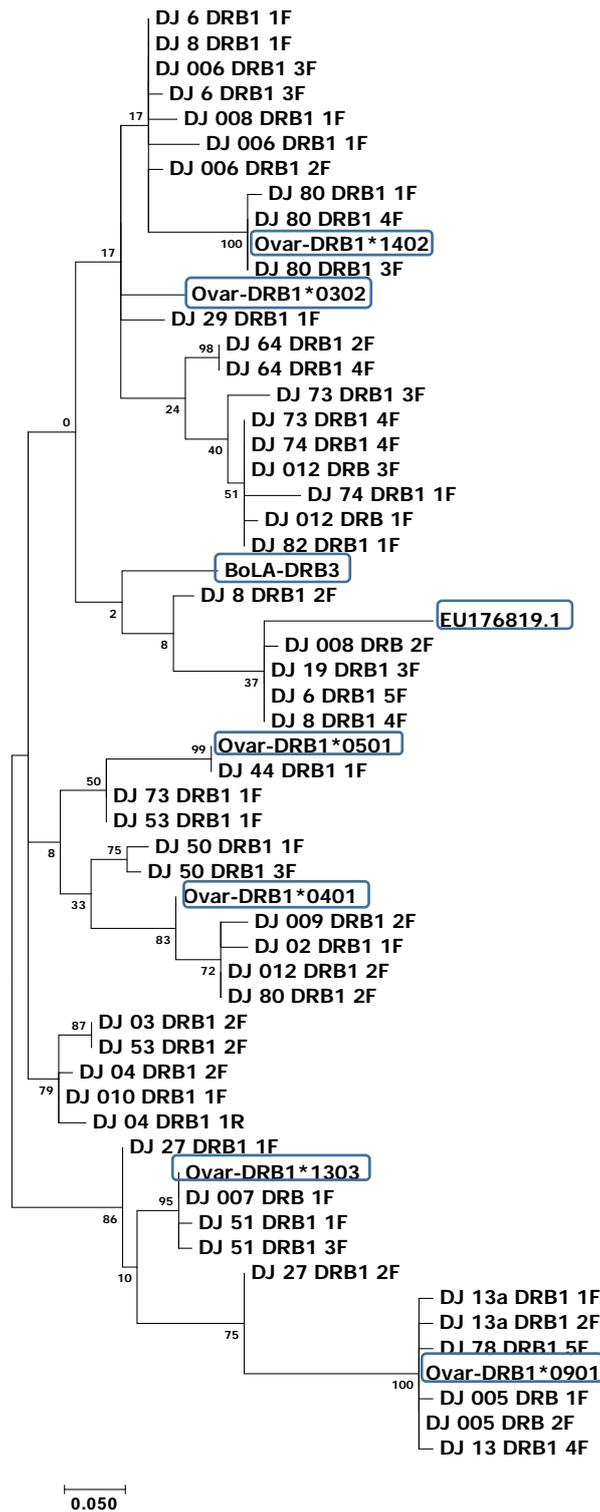
DJ_005_DRB1_1F	HFLEYRSEC	HFFNGTERVR	LLERYFHNGE	EFARFSDSWG	EFRAVTELGR	PAAEQWNSQK	NILEQKRAEV	NTVCRHNYGV	FESFAAQR
DJ_010_DRB1_3F									V
DJ_013_DRB1_1F			C			Q			V
DJ_13_DRB1_4F								R	V
DJ_13a_DRB1_1F					A				V
DJ_13a_DRB1_2F						R			V
DJ_78_DRB1_5F			K						V
SA_029_DRB1_3F							DF.R.TA	D.Y	I.SV
DJ_007_DRB1_1F	TKK		F.D.YT	NV		D	E.R.TA	D.Y	I.SV
DJ_27_DRB1_1F	SK		F.D.YT	NV		D	DF.R.TA	D.Y	I.SV
DJ_27_DRB1_2F	SK		F.D.YT	NV		D			V
DJ_51_DRB1_1F	TKK		F.D.YT	NV	P	D	E.R.TA	D.Y	I.SV
DJ_51_DRB1_3F	TKK		F.D.YT	NV		D	E.R.TA	D.Y.S	I.SV
SA_023_DRB1_3F	SK		F.D.YT	NV		D	DF.R.TA	D.Y	I.TV
SA_025_DRB1_3F	ST		F.D.YT	NV		D	DF.R.TA	D.Y	I.SV
SA_029_DRB1_1F	SK	G	F.D.YT	NV		D	DF.R.TA	D.Y	I.SV
SA_030_DRB1_1F	SK	W	F.D.YT	NV		D	DF.R.TA	D.Y	G.TV
SA_22_DRB1_1F	TKK		F.D.YT.R	NV		D	E.R.TA	D.Y	I.SV
SA_54_DRB1_5F	SK	S	F.D.YT	NV		D	DF.R.TA	D.Y	I.SV
SA_72_DRB1_5F	SK		F.D.YT	NV		D	DF.R.TA	D.Y	G.TV
SA_80_DRB1_4F	SK		F.D.YT	NV		D	DF.R.TA	D.Y.W	I.SV
SA_52_DRB1_4F	ST		F.D.Y	YV	Y.A	RS.Y	E.RR.T	D.Y	I.TV
SA_72_DRB1_1F	TKK	R.S	F.D.Y	Y	Y.A	D.KY	E.RR.T	D.Y	G.TV
SA_72_DRB1_4F	SK		F.D	Y	Y.A	D.KY	E.RR.T	D.Y	G.TV
SA_80_DRB1_3F	ST		F.D.Y	YV	Y.A	RS.Y	E.RR.T	D.Y	I.TV
SA_80_DRB1_5F	ST		F.D.Y	YV	Y.A	D.KY	E.RR.T	D.Y	I.SV
SA_83_DRB1_1F	HK		F.D.Y	YV	A	QS.Y	EL.RR.T	D.Y	I.SV
SA_83_DRB1_2F	L.HK		F.D.Y	YV	A	QS.Y	EL.RR.T	D.Y	I.SV
SA_27A_DRB1_4F	ST		F.D.Y	YV	Y.A	D.KY	DF.RA.A	D.Y	G.TV
DJ_006_DRB1_1F	HK		F.D.Y	YV.N	A	E.KY	E.RR.T	D.Y	G.TV
DJ_006_DRB1_2F	HK		F.D.Y	YV	A	E.KY	G.RR.T	D.Y	I.SV
DJ_006a_DRB1_3F	HK		F.D.Y	YV	A	E.KY	E.RR.T	D.Y	I.SV
DJ_008_DRB1_1F	HK		F.D.Y	YV	A	G.KYR	E.RR.T	D.Y	I.SV
DJ_008_DRB1_2F	TKK	R.S	F.D.Y	Y	Y.A	RS.Y	E.P.R.A	D.Y	G.TV
DJ_012_DRB1_1F	R	ST	F.D.Y	YV	Y.A	RS.Y	E.RR.T	D.Y	I.TV
DJ_012_DRB1_3F	ST		F.D.Y	YV	Y.A	RS.Y	E.RR.T	D.Y	I.TV
DJ_19_DRB1_3F	TKK	R.S	F.D.Y	Y	Y.A	RS.Y	E.R.A	D.Y	G.TV
DJ_29_DRB1_1F			F.D.Y	YV	Y.A	D.KY	EV.RR.T	D.Y	I.SV
DJ_53_DRB1_1F	AK	R	F.D.Y	YV	Y.A	D.KY	DF.R.N	D.Y	I.SV

DJ_005DRB1_1	FHFLEYRSEC	HFFNGTERVR	LLERYFHNGE	EFARFDSDWG	EFRAVTELGR	PAAEQWNSQK	NILEQKRAEV	NTVCRHNYGV	FESFAAQRR
DJ_64_DRB1_1F	HK		F.D.Y	.CV	AG	QS.Y	EL.RR.T	D.Y	SV
DJ_64_DRB1_4F	HK		F.D.Y	.YV	A	QS.Y	EL.RR.T	D.Y	SV
DJ_6_DRB1_1F	HK		F.D.Y	.YV	A	.E.KY	E.RR.T	D.Y	I.SV
DJ_6_DRB1_3F	HK		F.D.Y	.YV	A	.E.KY	E.RR.T	D.Y.C	I.SV
DJ_6_DRB1_5F	TKK	R.S	F.D.Y	.Y	Y.A	RS.Y	E.R.A	D.Y	G.TV
DJ_73_DRB1_1F	AK	R	F.Y	.YV	Y.A	.D.KY	DF.R.N	D.Y	SV
DJ_73_DRB1_3F	AK	R	F.Y	.YV	Y.A	RS.Y	E.RR.T	D.Y	TV
DJ_73a_DRB1_4F	ST		F.D.Y	.YV	Y.A	RS.Y	E.RR.T	D.Y	TV
DJ_74_DRB1_1F	ST	S	F.D.Y	.YV	YG.A	RS.Y	E.RR.T	D.Y	I.SV
DJ_74_DRB1_4F	ST		F.D.Y	.YV	Y.A	RS.Y	E.RR.T	D.Y	TV
DJ_80_DRB1_4F	TKK		F.Y	.YV	A	.E.KY	E.SR.TA	D.Y	I.SV
DJ_82_DRB1_1F	ST		F.D.Y	.YV	Y.A	RS.Y	E.RR.T	D.Y	TV
DJ_8_DRB1_1F	HK		F.D.Y	.YV	A	.E.KY	E.RR.T	D.Y	I.SV
DJ_8_DRB1_4F	TKK	R.S	F.D.Y	.Y	Y.A	RS.Y	E.R.A	D.Y	G.TV
SA_027_DRB_1F	HK		F.D.Y	.YV	A	QS.Y	EL.RR.T	D.Y	SV
SA_027_DRB_2F	HK		F.Y	.TL	A	QS.Y	EL.RR.T	D.Y	SV
SA_032_DRB_1F	ST		F.D.Y	.YV	Y.A	RS.Y	E.RR.T	D.Y	TV
DJ_009_DRB1_2F	HK	R.S	Y.D.Y	.YV.N	A	RS.Y	DF.T	D.Y	IV.SV
DJ_010_DRB1_1F	ST		F.D.Y	.YV	Y.A	.D.KY	DF.RA.A	D.Y	I.SV
DJ_012_DRB1_2F	HK	R.S	Y.D.Y	.YV.N	Y.A	RS.Y	DF.T	D.Y	I.SV
DJ_02_DRB1_1F	HK	R.S	Y.D.Y	.YV.N	Y.A	RS.Y	EL.T	D.Y	I.SV
DJ_04_DRB1_1R	ST		S.D.Y	.YV	Y.A	.D.KY	DF.RA.A	D.Y.G	I.SV
DJ_04_DRB1_2F	ST		F.D.Y	.YV	Y.A	.D.KY	DF.RA.A	D.Y	I.G.SV
DJ_50_DRB1_1F		R.S	Y.D.Y	.YV	Y.A	.D.KY	EL.R.N	D.Y	I.SV
DJ_50_DRB1_3F	ST	R.S	Y.D.Y	.YV	Y.A	.D.KY	EL.R.N	D.Y	I.SV
DJ_80_DRB1_2F	HK	R.S	Y.D.Y	.YV.N	Y.A	RS.Y	DF.T	D.Y	I.SV
SA_54_DRB1_1F	AK	S	Y.D.Y	.TL	Y.A	.D.KY	E.R.N	D.Y	I.SV
SA_92_DRB1_2F	AK	S	Y.D.Y	.TL	Y.A	.D.KY	G.R.N	D.Y	I.SV
DJ_44_DRB1_1F	AK	R	F.Y	.TL	Y.A	.D.KY	EL.R.N	D.Y	G.TV
SA_030_DRB_2F	AK	R	F.Y	.TL	Y.A	.D.KY	EL.R.N	D.Y.R	G.TV
SA_024_DRB1_1F	AK	R	F.Y	.YV.N	Y.A	RS.Y	EL.R.N	D.Y	I.SV
DJ_80_DRB1_1F	TKK		F.Y	.YV	A	.E.KY	E.SR.TA	D.H	I.SV
DJ_80_DRB1_3F	TKK		F.Y	.YV	A	.E.KY	E.SR.TA	D.Y	I.SV
DJ_8_DRB1_2F	HK		F.D.Y	.YV	A	RS.Y	E.R.A	D.Y	G.G.TV
DJ_03_DRB1_2F	ST		F.D.Y	.YV	Y.A	.D.KY	DF.RA.A	D.Y	G.TV
SA_023_DRB1_1F	TKK	RSS	F.D	.TL	Y.A	.D.KY	DF.RA.A	D.Y	I.TV
SA_023_DRB1_2F	T.K	R.S	F.D	.TL	Y.A	.D.KY	DF.RA.A	D.Y	I.TV
SA_028_DRB1_2F	TKK	R.S	F.D	.Y	Y.A	.D.KY	DF.RA.A	D.Y	G.TV
SA_23_DRB1_2F	TKK	R.S	F.D	.TL	Y.A	.D.KY	DF.RA.A	D.Y	I.TV

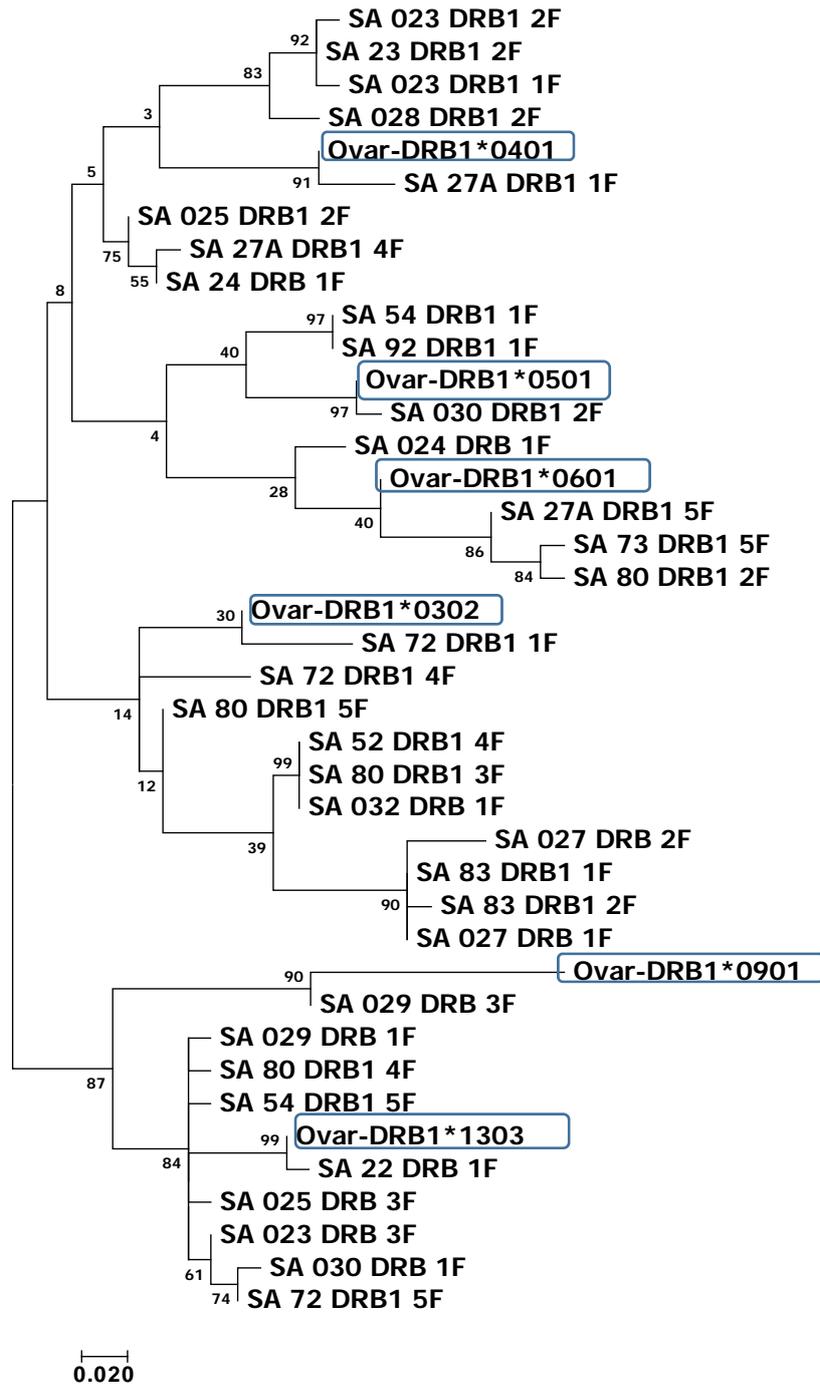
DJ005\_DRB1\_1F HFLEYRSEC HFFNGTERVR LLERYFHNGE EFARFDSDWG EFRAVTELGR PAAEQWNSQK NILEQKRAEV NTVCRHNYGV FESFAAQRR  
 SA\_24\_DRB\_1F .....ST... ..... F.D...Y... .YV..... .Y...A.... .D.KY..... DF..RA..A. D.Y..... G...SV...  
 SA\_27A\_DRB1\_5F.....AK... R..... F.D...Y... .YV..... .....A.... RS..Y..... DF..RA..A. D.Y..... G...TV...  
 SA\_73\_DRB1\_5F .....AKR.. ..... F.D...Y... .YV.....S .....A.... RS..Y..... DF..RA..A. D.Y..... G...TV...  
 SA\_80\_DRB1\_2F .....AKR.. R..... F.D..SY... .YV.....S .....A.... RS..Y..... DF..RA..A. D.Y..... G...TV...  
 DJ\_005\_DRB12F .....\* ..... ..... ..... ..... ..... ..... .....V...  
 SA\_27A\_DRB12F .....AK..C R..... F.D...Y... .YV..... .....A\*... RS..Y..... DF..RA..A. D.Y.....R. G...T....



**Figure 4.3** Best model Maximum likelihood phylogenetic tree for combined functional and unique *DRB1* exon 2 coding sequences in the two sheep populations. Tree was constructed to scale with 85 amino acid sequences. The tree shows the topography of evolutionary relationship of all unique sequences. The numbers at the nodes represent bootstrap confidence based on 500 replications



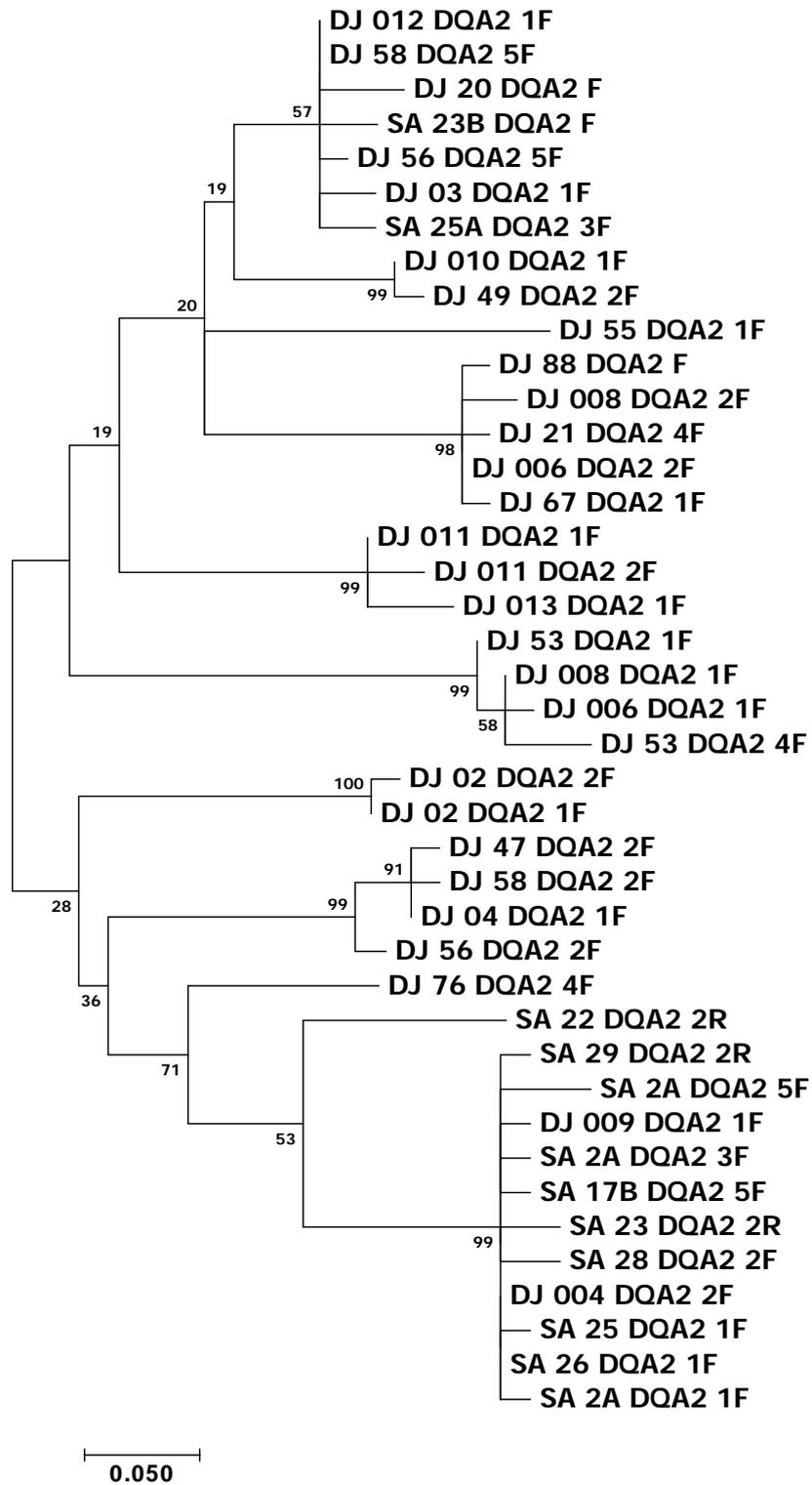
**Figure 4.4** Best model Maximum likelihood phylogenetic tree for functional and unique *DRB1* exon 2 coding sequences in the Djallonke sheep population. Tree was constructed to scale with 58 amino acid long sequences. The IPD name was used for unique sequences that matched previous publications. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications.



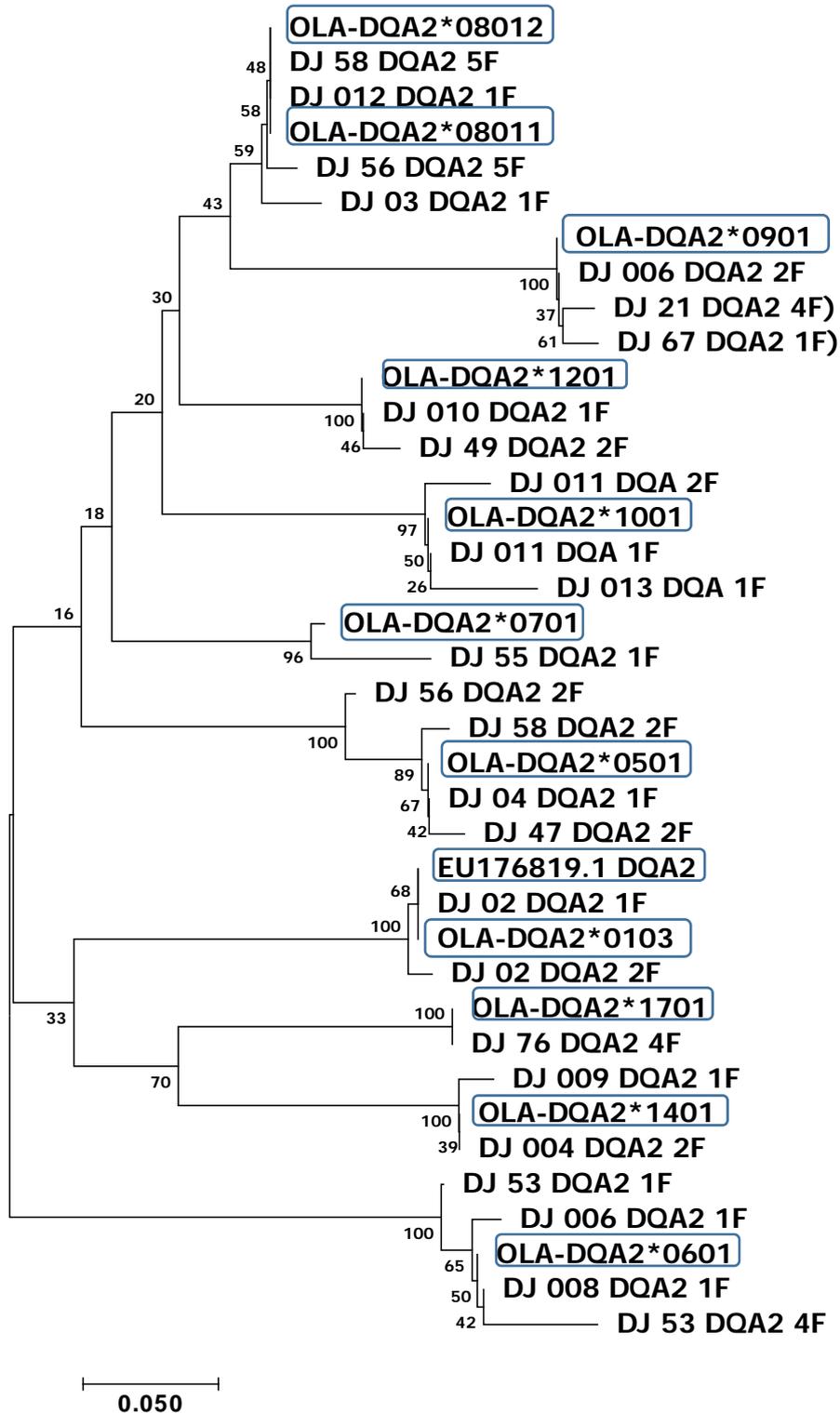
**Figure 4.5** Best model Maximum likelihood phylogenetic tree for functional and unique *DRB1* exon 2 coding sequences in the Sahelian sheep population. Tree was constructed to scale with 40 amino acid long sequences. The IPD name was used for unique sequences that matched previous publications. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications

Table 4.3 Unique DQA2 exon 2 amino acid sequences identified in Djallonke and Sahelian sheep populations (n= 41)

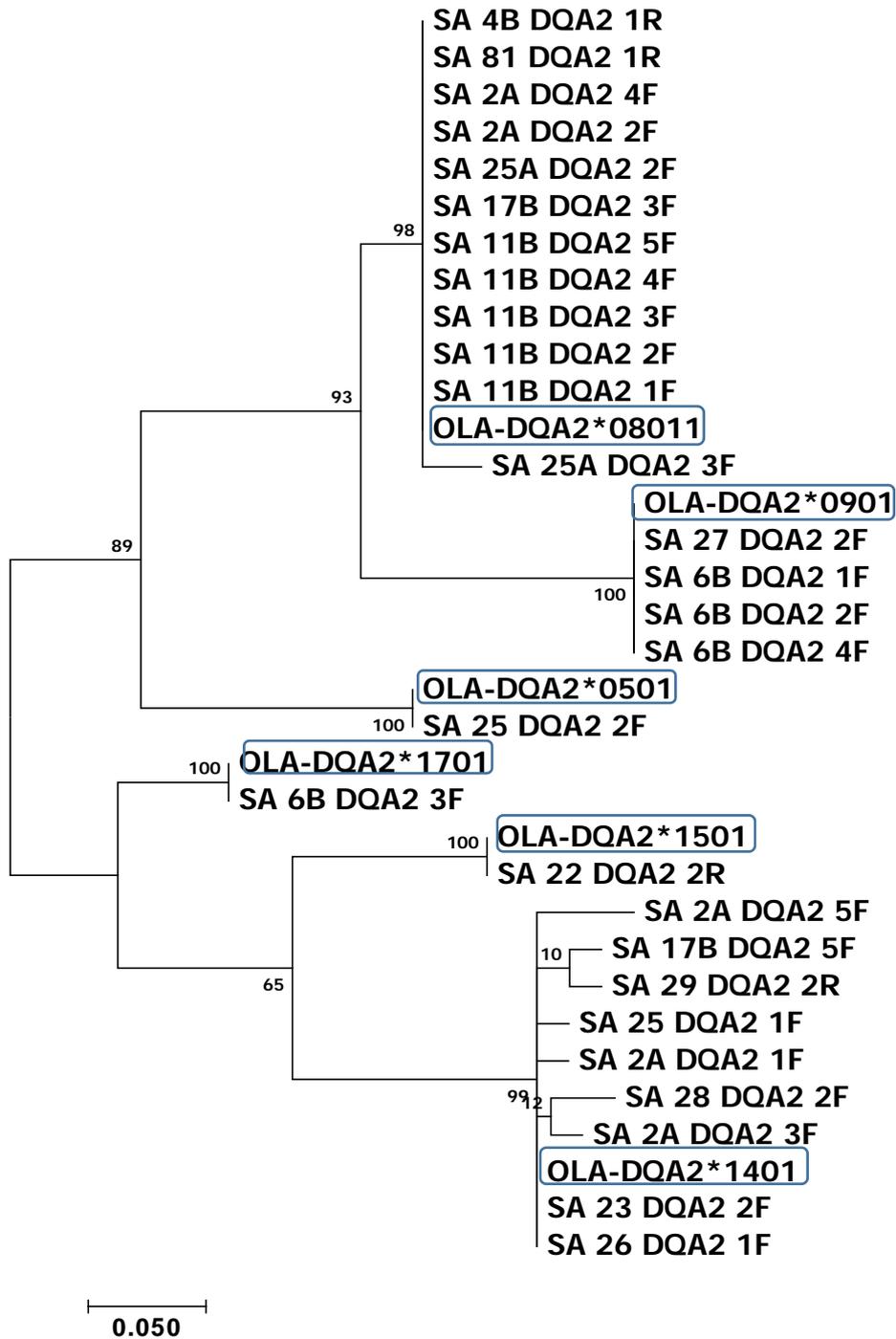
DJ_004_DQA2_2F	DHIGTYGTFD	YQSHGPSGQY	IHEFDGDEQL	YVDLEKKETV	WRLPMFDGL-	SFDPQRALSN	IATAKHNLDR	LTKWYNFTPV	IN
DJ_009_DQA2_1F	.....	.....	.....	.....	.....	.....	D.....	.....	..
SA_17B_DQA2_5F	.....	.....	R.....	.....	.....	.....	.....	.....	..
SA_23_DQA2_2R	.....	.....	.....	.....	.....	.....	S.....	S.....	..
SA_25_DQA2_1F	.....	P.....	.....	.....	.....	.....	.....	.....	..
SA_26_DQA2_1F	.....	.....	.....	.....	.....	.....	.....	.....	..
SA_28_DQA2_2F	.....	C.....	.....	P.....	.....	.....	.....	.....	..
SA_29_DQA2_2R	.....	.....	.....	.....	.....	.....	M.....	.....	..
SA_2A_DQA2_1F	.....	.....	.....	.....	.....	.....	.....	.....	L..
SA_2A_DQA2_3F	.....	.....	.....	.....	.....	.....	.....	P.....	..
SA_2A_DQA2_5F	.....	.....	G.....	.....	S.....	.....	.....	C.....	..
DJ_006_DQA2_1F	V...AE.	.....SE.	TQ...E.L.	.....	.....GQFA	G.HI.V.	T...S.V	M.....	..
DJ_008_DQA2_1F	V...AE.	.....SE.	TQ...E.L.	.....	.....GQFA	G.HI.V.	T...V	M.....	..
DJ_53_DQA2_1F	V...AE.	.....SE.	TQ...E.L.	.....	.....GQFA	G.HI.V.	T...V	M.R.....	..
DJ_53_DQA2_4F	RV.C.AE.	.....SE.	TQ...E.L.	.....	.....GQFA	G.HI.V.	V.T...V	M.....	..
DJ_012_DQA2_1F	F.S...TI	.....F	TQ...LF	.....	.....SQFA	G...G...	A.....I	RS.S...	..
DJ_03_DQA2_1F	F.S...TI	.....F	TQ...L.	.....	.....SQFA	G...G...S	A.....I	RS.S...	..
DJ_20_DQA2_F	F.S...TI	.....F	TQ...NKLF	K.....	.....SQFA	G...G...	A.....I	RS.S...	..
DJ_56_DQA2_5F	F.S...TI	.....F	TQ...LF	.....	.....SQFA	G...G...	A.....I	RS.S.A..	..
SA_23B_DQA2_F	F.S...TI	.....F	TQ...LF	.....	.....SQFA	G...G...I	AS.....I	RS.S...	..
SA_25A_DQA2_3F	F.S...TI	.....F	TQ...LF	.....	.....SQFA	G...G...	A...S.V	RS.S...	..
DJ_006_DQA2_2F	S...TI	.....S.	TQ...LF	.....	.....SQFA	G.NI.D.N.	PA....GI	RS.....	..
DJ_008_DQA2_2F	S...TI	.....S.	TQ...LF	.....	.....SQFA	G.NV.D.N.	PA....GI	RS.....	T.
DJ_21_DQA2_4F	S...TI	.....S.	TQ...LF	.....	.....SQFA	G.NI.D.N.	PA.R...GI	RS.....	..
DJ_67_DQA2_1F	S...TI	N.....S.	TQ...LF	.....	.....SQFA	G.NI.D.N.	PA....GI	RS.....	..
DJ_88_DQA2_F	S...TI	.....S.	TQ...KLF	.....	.....SQFA	G.NI.D.N.	PA....GI	RS.....	..
DJ_04_DQA2_1F	V.C.S.II	.....F	T...L.	.....	.....I.GE.T	....G...	T...I	CS.C...	..
DJ_47_DQA2_2F	V.C.S.II	.....F	T...L.	E.....	.....I.GE.T	....G...	T...I	CS.C...	..
DJ_56_DQA2_2F	F.S...II	.....F	T...L.	.....	.....I.GE.T	....G...	T...I	CS.C...	..
DJ_58_DQA2_2F	A.C.S.II	.....F	T...L.	.....	.....I.GE.T	....G...	T...I	CS.C...	..
DJ_011_DQA2_1F	V.S.....	.....F	TQ...L.	G.....	.....SQFA	D...G.R.	T.DT..I	RS.S...	..
DJ_011_DQA2_2F	V.S.....	.....F	TQ...L.	G.....	.....V.SQFA	D...G.RD	T.DT..I	RS.S...	..
DJ_013_DQA2_1F	V.S.A..	.....F	TQ...G.S.	G.....	.....SQFA	D...G.R.	T.DT..I	RS.S...	..
DJ_010_DQA2_1F	F.S...EI	.....TQ	...LF	G.....	.....SQFA	G...G...E	T.Q...I	RS...A..	..
DJ_49_DQA2_2F	F.S...EI	.....TQ	...LF	GE.....	.....SQFA	G...G...E	T.Q...I	RS...A..	..
DJ_02_DQA2_2F	V.I.A.S	.....T	...LF	G.....	.....GEFT	....G...E	K.QT..I	MI.RS....	..
DJ_02_DQA2_1F	V.I.A..	.....T	...LF	G.....	.....GEFT	....G...E	K.QT..I	MI.RS....	..
DJ_76_DQA2_4F	V...N.	.....L	...RF	.....	.....GE.I	....G...	V.AS....I	RS.....	..
DJ_55_DQA2_1F	.....VI	.....T	R.LF	.....	.....SQFA	GS...G...	T...D..I	M.Q.H.S...	..
SA_22_DQA2_2R	V.I.....	.....E.	L...EF	.....	.....E.R	R...G.N.	.....I	R.....	..
DJ_58_DQA2_5F	F.S...TI	.....F	TQ...LF	.....	.....SQFA	G...G...	A.....I	RS.S...	..



**Figure 4.6** Best model Maximum likelihood phylogenetic tree for combined functional and unique *DQA2* exon 2 coding sequences in the two sheep populations. Tree was constructed to scale with 41 amino acid long sequences. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications.



**Figure 4.7** Best model Maximum likelihood phylogenetic tree for functional and unique *DQA2* exon 2 coding sequences in the Djallonke sheep population. Tree was constructed to scale with 38 amino acid long sequences. The IPD name was used for unique sequences that matched previous publications. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications

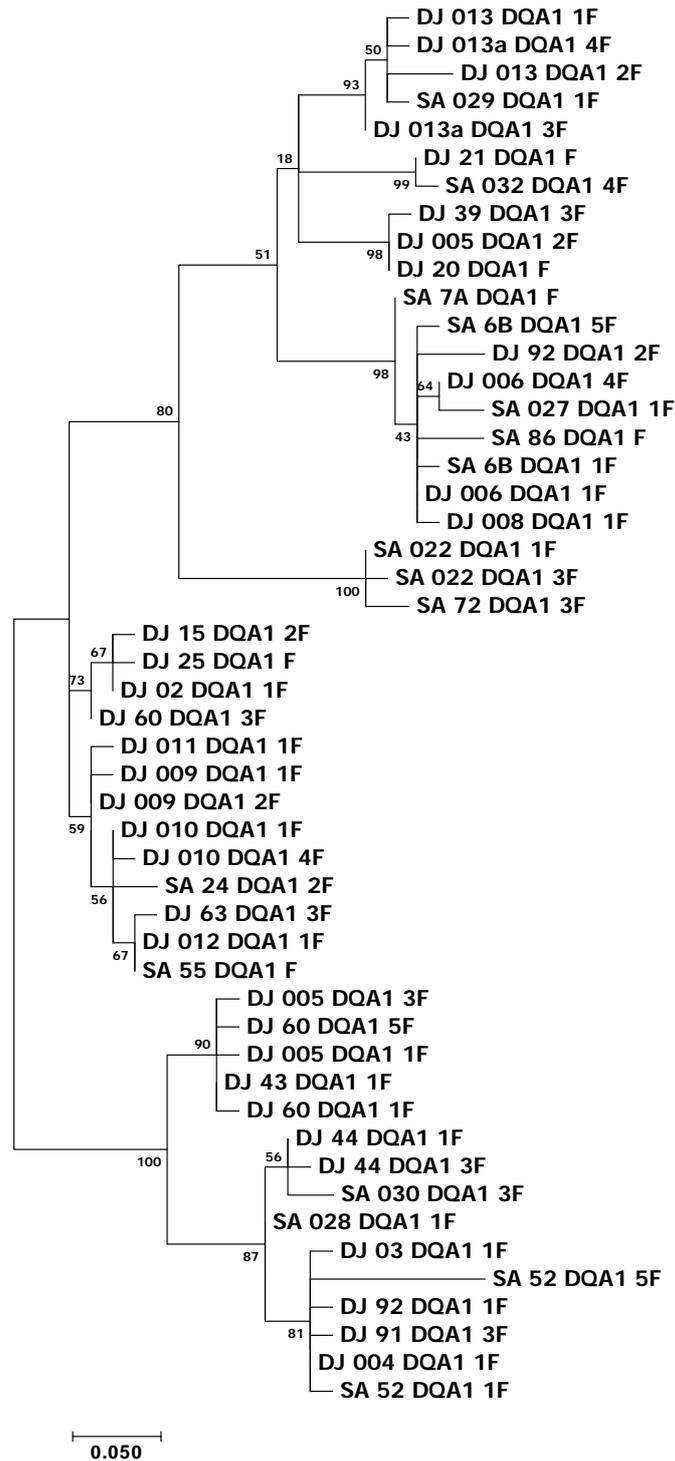


**Figure 4.8** Best model Maximum likelihood phylogenetic tree for functional and unique *DQA2* exon 2 coding sequences in the Sahelian sheep population. The IPD name was used for unique sequences that matched previous publications. Tree was constructed to scale with 34 amino acid long sequences. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications

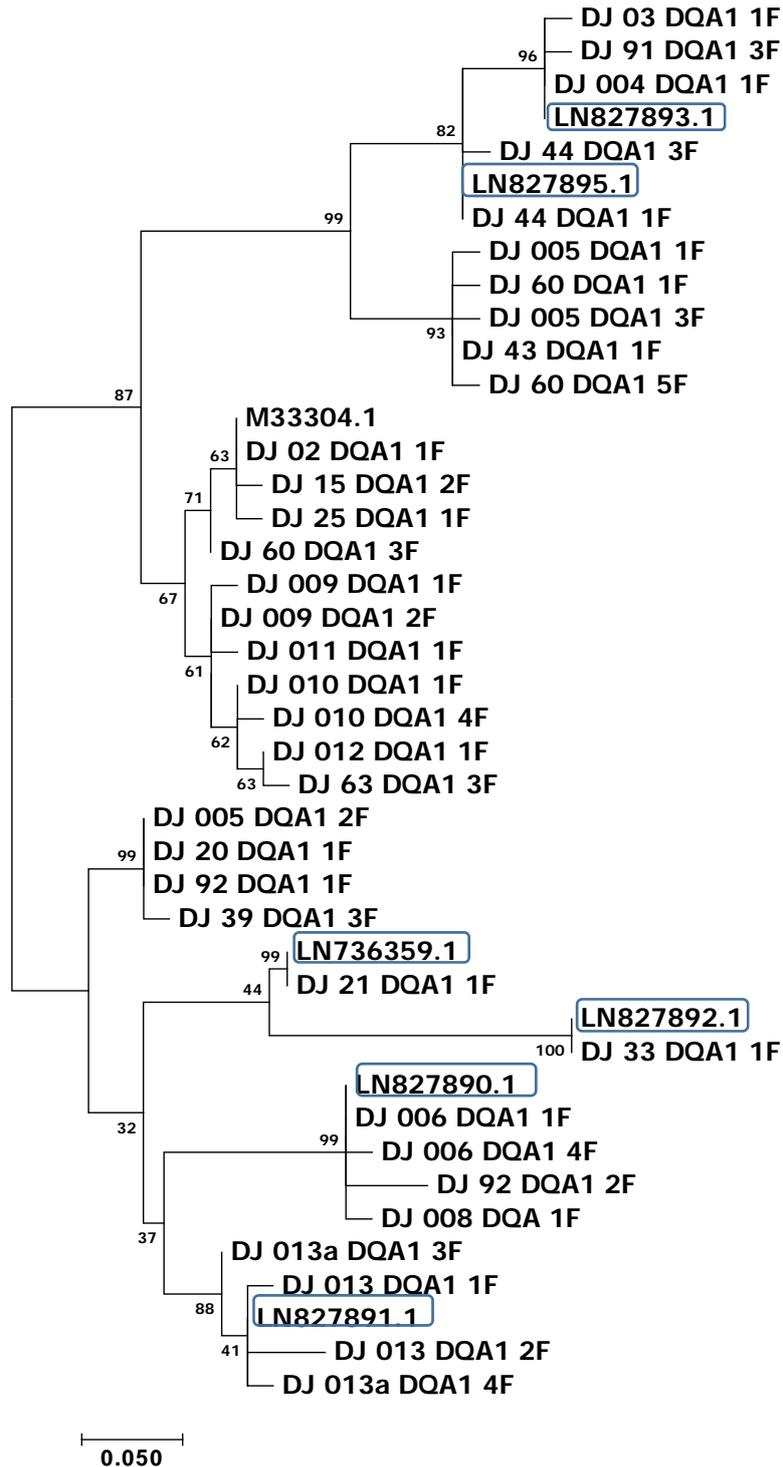
Table 4.4 Unique *DQA1* exon 2 amino acid sequences identified in Djallonke and Sahelian sheep populations (Presumed functional coding sequences (n= 52) with non functional coding sequences containing premature stop codons highlighted in yellow (n =3))

DJ_006_DQA1_1F	DHIAAYGINV	YHRYGPSGYY	THEFDGDEEF	YVDLEKRETV	WHLPMFSKFR	RFDPQGALRN	IAAAKHNLEV	LIQDSNSTAA	SN
DJ_006_DQA1_4F				P					
DJ_008_DQA1_1F			L						
DJ_92_DQA1_2F			L		I	D			
SA_7A_DQA1_F					V				
SA_86_DQA1_F			N.K			K			
SA_6B_DQA1_1F				G					
SA_6B_DQA1_5F							V		
SA_027_DQA1_1F	T			R					
DJ_005_DQA1_2F		QT		K	V.Q		T	I	T.R
DJ_009_DQA1_1F	GI.V	QT.P			R.E.T	S	T	I	R
DJ_009_DQA1_2F	GI.V	QT			R.E.T	S	T	I	R
DJ_010_DQA1_1F	GI.V	QT			R.E.T	S	T	I	M
DJ_010_DQA1_4F	GI.V	QT		M	R.E.T	S	T	I	M
DJ_011_DQA1_1F	GI.V	QT			R.E.T	S.C	T	I	R
DJ_012_DQA1_1F	GI.V	QT			R.E.T	S	TT	I	M
DJ_02_DQA1_1F	GT.V.I	QT			R.E.T	S	TV	I	R
DJ_15_DQA1_2F	GTH.V.I	QT			R.E.T	S	TV	I	R
DJ_20_DQA1_F		QT		K	V.Q		T	I	T.R
DJ_25_DQA1_F	GT.V.I	QT		H	R.E.T	S	TV	I	R
DJ_39_DQA1_3F		QT		G.K	V.Q		T	I	T.R
DJ_60_DQA1_3F	GT.V.I	QT			R.E.T	S	T	I	R
DJ_63_DQA1_3F	GI.V	QT	L		R.E.T	S	TT	I	M
SA_24_DQA1_2F	GI.V	QT			C.E.T	S	T	I	M
DJ_013_DQA1_1F		S			V.Q	S	TV		T.R
DJ_013_DQA1_2F	R.V	S		G	V.Q	S	TV		R
DJ_013a_DQA1_3F		S			V.Q	G	TV		R
DJ_013a_DQA1_4F		S			V.Q	S	TV	G	R
SA_029_DQA1_1F		S		K	V.Q	S	TV		R
DJ_004_DQA1_1F	GT.V	QT.S		Q	R.E.EVG	Y.A.F	TF	L	M.R
DJ_03_DQA1_1F	GT.V	QT.S		Q	R.E.EVG	Y.A.F	TF	L	M.R
DJ_91_DQA1_3F	GT.V	QT.S		Q	R.E.EVG	Y.A.F	TF.R	L	M.R
DJ_92_DQA1_1F	GT.V	QT.S		Q	R.E.EVG	Y.A.F	TF	VL	M.R
SA_52_DQA1_1F	GT.V	QT.S		VQ	R.E.EVG	Y.A.F	TF	L	M.R
SA_52_DQA1_5F	GT.V	QT.S.H		Q	R.E.EVG	Y.A.F	TF	L	M.R.TLLLL
005_DQA1_1F	GT.V	QT.F		Q	R.E.G	D.F	TL	L	I.R
DJ_005_DQA1_3F	GT.V	QT.F		Q	R.G.G	D.F	TL	L	I.R
DJ_43_DQA1_1F	GT.V	QT.F		Q	R.G	D.F	TL	L	I.R
DJ_44_DQA1_1F	GT.V	QT			R.EVG	Y.F	TF	L	M.R

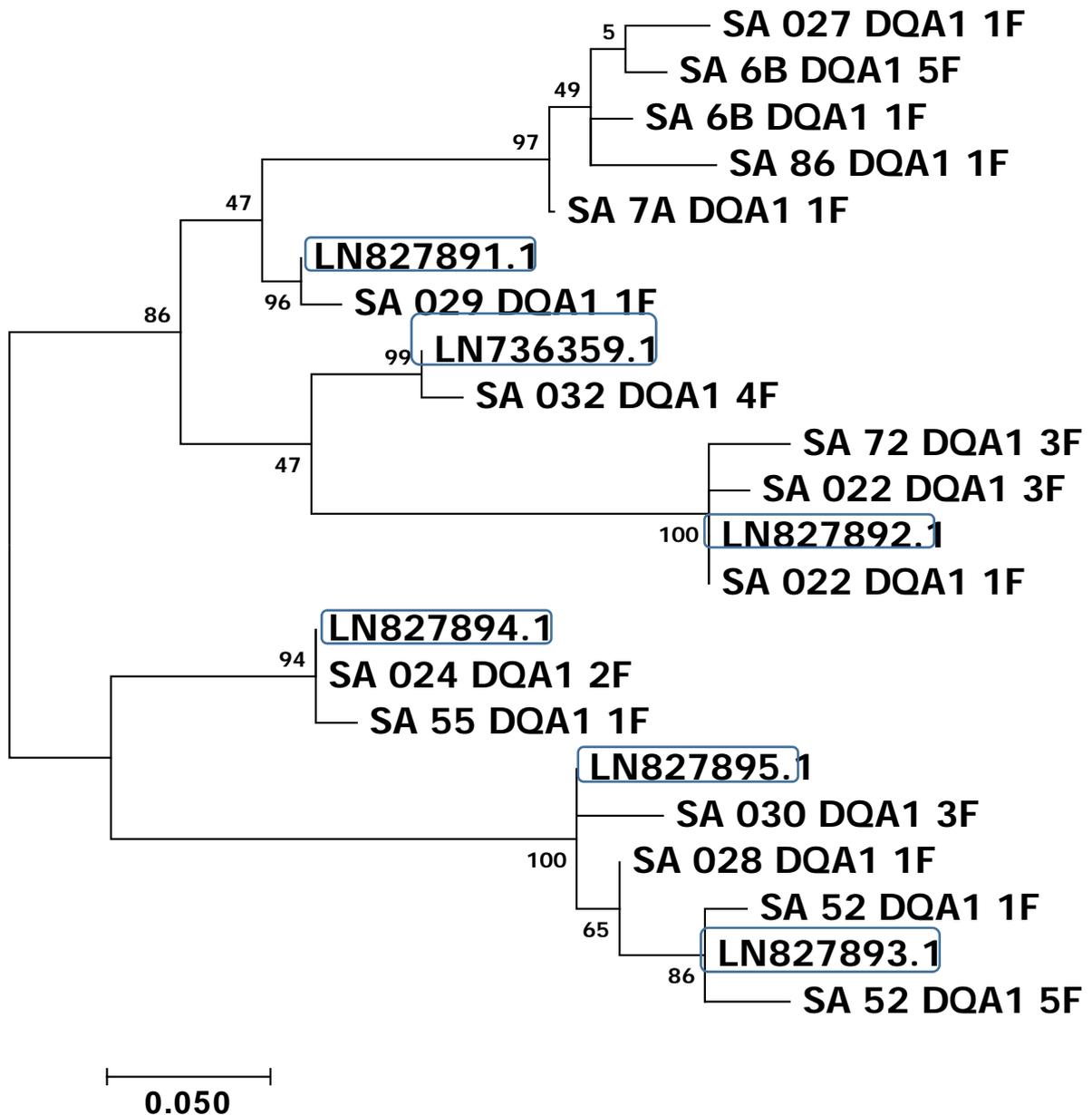
DJ006\_DQA1\_1F DHIAAYGINV YHRYGPGSGYY THEFDGDEEF YVDLEKRETV WHLPMPFKFR RFDPQALRN IAAAKHNLEV LIQDSNSTAA SN  
 DJ\_44\_DQA1\_3F ...GT..V.. .QT..... .G.. .R.....EVG Y....F.... ..TF.....L M..R..... T.  
 DJ\_60\_DQA1\_1F ...GT..V.. .QT.....F .....Q. .... .R.....G D....L.... ..TL.....L I..R..... T.  
 DJ\_60\_DQA1\_5F ...GT..VS. .QT.....F .....Q. .... .R.....G D....F.... ..TL.....L I..R..... T.  
 SA\_028\_DQA1\_1F...GT..V.. .QT..... .Q. .... .R.....EVG Y....F.... ..TF.....L M..R..... T.  
 SA\_030\_DQA1\_3F...GT..V.. .QT..... .L..... .R.....EVG Y....F.... ..TF....GL M..R..... T.  
 DJ\_21\_DQA1\_F .....S..... .R..V..Q.. ..... .VG.QS..I ...R..... T.  
 SA\_032\_DQA1\_4F.....S..... .R..VS.Q.. ..... .VG.QS..I ...R..... T.  
 SA\_022\_DQA1\_1F.....S..... .Q. ....K... .R..E....T S..... M.SG.QT..I M..R..... T.  
 SA\_022\_DQA1\_3F.....S..... .Q. ....K... .R..E....T S..... M.SG.QT..I M.RR..... T.  
 SA\_72\_DQA1\_3F .....S..... .Q. ....K... .RP.E....T SS..... M.SG.QT..I M..R..... T.  
 SA\_55\_DQA1\_F ...GI..V.. .QT..... .R..E....T S..... ..TT.....I ...M..... T.  
 DJ\_39\_DQA1\_2F ..... .MNLMEMKSS TWTWKR.RLS GICLCLVNLE GLTLRVH\*ET \*LQRNIWRS RFKGPTLLLL PT  
 DJ\_91\_DQA1\_2F ...GT\*.V.. .....S..... .Q. .E.....-ET VW..PMPSEV .Y---F-DAQ F.L...A--- T..HNLE.MI QR  
 DJ\_92\_DQA1\_4F .....M.. .S..... .....-ET VWR.PVFSQF RR---F-DPQ G.L...A--- VG.QS\*EI.I QR



**Figure 4.9** Best model Maximum likelihood phylogenetic tree for combined functional and unique *DQA1* exon 2 coding sequences in the two sheep populations. Tree was constructed to scale with 55 amino acid long sequences. The IPD name was used for unique sequences that matched previous publications. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications.



**Figure 4.10** Best model Maximum likelihood phylogenetic tree for functional and unique *DQA1* exon 2 coding sequences in the Djallonke sheep population. Tree was constructed to scale with 42 amino acid long sequences. The GenBank name was used for unique sequences that matched previous publications. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications



**Figure 4.11** Best model Maximum likelihood phylogenetic tree for functional and unique *DQA1* exon 2 coding sequences in the Sahelian sheep population. Tree was constructed to scale with 22 amino acid long sequences. The GenBank name was used for unique sequences that matched previous publications. The branch lengths are equivalent to the amino acid substitutions per site and the numbers at the nodes represent bootstrap confidence based on 500 replications

### 4.3.3 Classification of MHC class IIa alleles in Djallonke and Sahelian sheep using the MHC-IPD Nomenclature guidelines

Using the MHC-IPD nomenclature guidelines in association with GenBank reference sequences (Maccari et al., 2017; Robinson et al., 2015), an assignment of the amino acid sequences generated in this study to generate a putative allele designation, yielded 26 *DRB1* alleles, 13 *DQA2* alleles and 13 *DQA1* alleles from the Djallonke and Sahelian sheep breeds.

Of the 26 *DRB1* allele groups observed (Table 4.5), seven matched previously published ovine *DRB1* sequences and therefore can be referred to by their MHC-IPD accession numbers, or their respective GenBank accession number where applicable. Each of the remaining 19 alleles differed within the CDS, having at least five amino acid differences compared to the best NCBI blast match, and therefore were described as “variants” of their respective matched allele at the MHC-IPD accession and/or the GenBank ID. These 19 alleles have not been previously described, and may possibly represent breed specific alleles. Interestingly, one of the allele “variant” groups appeared to best match the *BoLA-DRB3* (BTU78548.1) of cattle, and differed only by 8 CDS amino acid substitutions. Equal numbers of *DRB1* alleles (21) were found within each of the two sheep breeds. Five of the *DRB1* alleles (*Ovar-DRB1\*0901*, *Ovar-DRB1\*1303* variant-2, *Ovar-DRB1\*0302* variant-4, *Ovar-DRB1\*0302* variant-5 and *Ovar-DRB1\*0401* variant-2) were found only in the Djallonke population; whereas five other alleles (*Ovar-DRB1\*0302* variant-7, *Ovar-DRB1\*0302* variant-8, *Ovar-DRB1\*0401* variant-3, *DRB1* EU176819.1 variant-2 and *DRB1* EU176819.1 variant-3) were present in the Sahelian population. The allele group *DRB1\*0401* was interesting because it had a different translation frame (codon\_start=1) for the 270-nucleotide sequence. Furthermore, unlike all the other *DRB1* allele groups that are 89 amino acids long, this codon\_start=1 translated *DRB1\*0401* allele has a 90-amino acid long predicted protein. A presumed correct translation frame with codon starting at position 3 was used to re-translate all the *DRB1\*0401* sequences for comparison. The two groups of the differently translated *DRB1\*0401* amino acids sequences were analysed further to ascertain the correct *DRB1\*0401* amino acid sequence using multiple sequence alignments (Table 4.8 & Table 4.9) and phylogeny (Figure 4.12, Figure 4.13, Figure 4.14 & Figure 4.15).

The 13 *DQA2* alleles matched previously reported sheep *DQA2* alleles, and were denoted by their IPD or GenBank ID numbers (Table 4.6). The number of non-synonymous CDS mutations for each of the *DQA2* alleles in both sheep populations when compared to the previously published versions was less than five amino acids, hence no new *DQA2* alleles were

found, as per the IPD nomenclature criteria. Of the 13 *DQA2* alleles identified, 12 were found to occur within the Djallonke breed, whereas only 8 were found within the Sahelian breed. The allele *OLA-DQA2\*1501* was identified only in the Sahelian population while four other alleles were found only in the Djallonke population (*OLA-DQA2\*1001*, *OLA-DQA2\*0103*, *OLA-DQA2\*0812*, and *OLA-DQA2-EU176819.1*). The allele *OLA-DQA2\*1401*, unlike all the other *DQA2* alleles is only 81 amino acids in length.

For the 13 *DQA1* alleles (Table 4.7) amplified using the *DQA1* primer set, eight have been designated according to their GenBank ID. The remaining five alleles differ by five or more non-synonymous mutations with respect to the best BLAST matches, and were labelled as variants of their respective closest matched accession within the database (as listed in Table 4.7). An equal number of *DQA1* alleles (11) were found in both sheep populations. Two of the *DQA1* alleles (M33304.1 *variant-2* and LN827895.1 *variant-1*) were identified only in the Djallonke, and two other alleles (LN827893.1 and LN827894.1 *variant-3*) were identified only in the Sahelian.

For some samples, *DRB1*, *DQA2*, and/or *DQA1* PCR did not produce any detectable product despite optimisation attempts, and were therefore defined as possible null alleles.

**Table 4.5 Twenty six *DRB1* allele groups in Djallonke and Sahelian sheep populations based on IPD-MHC nomenclature**

<i>DRB1</i> Allele number	Official name/ designated name	Accession number	Designated code	Number of Amino acids
1	<i>Ovar-DRB1*0901</i>	FN543119.1	A	89
2	*0901 variant 1		Z	89
3	<i>Ovar-DRB1*1303</i>	FR751085.1	B	89
4	*1303 variant 1		C	89
5	*1303 variant 2		D	89
6	<i>Ovar-DRB1*0302</i>	AM884913.1	E	89
7	*0302 variant 1		F	89
8	*0302 variant 2		G	89
9	*0302 variant 3		H	89
10	*0302 variant 4		I	89

<b>DRB1 Allele number</b>	<b>Official name/ designated name</b>	<b>Accession number</b>	<b>Designated code</b>	<b>Number of Amino acids</b>
11	*0302 variant 5		J	89
12	*0302 variant 6		K	89
13	*0302 variant 7		L	89
14	*0302 variant 8		M	89
15	<b><i>Ovar-DRB1*0401</i></b>	<b>AM884914.1</b>	N	90/89*
16	*0401 variant 1		O	90/89*
17	*0401 variant 2		P	90/89*
18	*0401 variant 3		Q	90/89*
19	<b><i>Ovar-DRB1*0501</i></b>	<b>FN543116.1</b>	<b>R</b>	<b>89</b>
-	<i>Ovar-DRB1*0601</i>	FN543117.1	<b>No sample</b>	-
20	*0601 variant 1		S	89
21	<b><i>Ovar-DRB1*1402</i></b>	<b>FN393733.1</b>	<b>T</b>	<b>89</b>
-	<b><i>BOLA-DRB3</i></b>	<b>BTU78548.1</b>	<b>No Sample</b>	-
22	<i>BOLA-DRB3</i> Variant 1		U	89
23	<b>EU176819.1 DRB1</b>	<b>EU176819.1</b>	V	89
24	EU176819.1 variant 1		W	89
25	EU176819.1 variant 2		X	89
26	EU176819.1 variant 3		<b>Y</b>	<b>89</b>

**Table 4.6 *DQA2* allele groups in Djallonke and Sahelian sheep populations based on IPD-MHC nomenclature**

<b>DQA2 Allele number</b>	<b>Official allele name</b>	<b>Accession number</b>	<b>Designated code</b>	<b>Number of Amino acids</b>
<b>1</b>	<i>OLA-DQA2*1401</i>	AY312392.1	A	81
<b>2</b>	<i>OLA-DQA2*0601</i>	AY312381.1	B	82
<b>3</b>	<i>OLA-DQA2*08011</i>	AY312385.1	C	82
<b>4</b>	<i>OLA-DQA2*0901</i>	AY312387.1	D	82
<b>5</b>	<i>OLA-DQA2*0501</i>	AY312380.1	E	82
<b>6</b>	<i>OLA-DQA2*1001</i>	AY312388.1	F	82
<b>7</b>	<i>OLA-DQA2*1201</i>	AY312390.1	G	82
<b>8</b>	<i>OLA-DQA2*0103</i>	AY312377.1	H	82
<b>9</b>	EU176819.1 DQA2	EU176819.1	I	82
<b>10</b>	<i>OLA-DQA2*1701</i>	AY312395.1	J	82
<b>11</b>	<i>OLA-DQA2*0701</i>	AY312383.1	K	82
<b>12</b>	<i>OLA-DQA2*1501</i>	AY312393.1	L	82
<b>13</b>	<i>OLA-DQA2*08012</i>	AY312386.1	M	82

**Table 4.7 *DQA1* allele groups in Djallonke and Sahelian sheep populations based on IPD-MHC nomenclature**

<b>DQA1 Allele number</b>	<b>Official Allele name /designated name</b>	<b>Designated code</b>	<b>Number of Amino acids</b>
<b>1</b>	<b>LN827890.1</b>	<b>A</b>	<b>82</b>
<b>2</b>	<b>M33304.1</b>	<b>B</b>	<b>82</b>
<b>3</b>	M33304.1 variant 1	C	82
<b>4</b>	M33304.1 variant 2	D	82
<b>5</b>	<b>LN827891.1</b>	<b>E</b>	82
<b>6</b>	<b>LN827893.1</b>	<b>F</b>	<b>82</b>
<b>7</b>	LN827893.1 variant 1	G	<b>82</b>
<b>8</b>	LN827893.1 variant 2	H	82
<b>9</b>	<b>LN827895.1</b>	<b>I</b>	82
<b>10</b>	LN827895.1 variant 1	J	82
<b>11</b>	<b>LN736359.1</b>	<b>K</b>	<b>82</b>
<b>12</b>	<b>LN827892.1</b>	<b>L</b>	<b>82</b>
<b>13</b>	<b>LN827894.1</b>	<b>M</b>	<b>82</b>

Table 4.8 Muscle alignment of the 26 *DRB1* Alleles (A to Z) in Djallonke and Sahelian sheep using codon\_start 1 translation frame for *DRB1\*0401* showing misalignment with the remaining DRB1 alleles

A_ovar-DRB1*0901	HFLEYRSEC	HFFNGTERVR	LLERYFHNGE	EFARFDSDWG	EFRA--VTEL	GRPAAEQWNS	QKNILEQKRA	EVNT---VC	RHNYGVFE-S	
A_DJ_78_DRB1_5F	.....	.....	...K.....	.....	.....	.....	.....	.....	.....	
Z_SA_029_DRB1_3F	.....	.....	.....	.....	.....	.....	..DF..R..T	A.D....Y.	.....I...	
B_Ovar-DRB1*1303	.....TKK..	.....	F.D..YT...	.NV.....	.....	..D.....	..E...R..T	A.D....Y.	.....I...	
SA_52_DRB1_3F	.....SK...	.....	F.D..YT...	.NV.....	.....	..D.....	..DF..R..T	A.D....Y.	.....I...	
C_SA_54_DRB1_5F	.....SK...	..S.....	F.D..YT...	.NV.....	.....	..D.....	..DF..R..T	A.D....Y.	.....I...	
D_DJ_27_DRB1_2F	.....SK...	.....	F.D..YT...	.NV.....	.....	..D.....	.....	.....	.....	
E_Ovar-DRB1*0302	.....TKK..	R.S.....	F.D..Y...	.YV.....	.Y....A...	..D.KY...	..E...RR.T	..D....Y.	.....I...	
SA_72_DRB1_1F	.....TKK..	R.S.....	F.D..Y...	.Y.....	.Y....A...	..D.KY...	..E...RR.T	..D....Y.	.....G...	
F_SA_80_DRB1_5F	.....ST...	.....	F.D..Y...	.YV.....	.Y....A...	..D.KY...	..E...RR.T	..D....Y.	.....I...	
G_DJ_08_DRB1_2F	.....TKK..	R.S.....	F.D..Y...	.Y.....	.Y....A...	..RS..Y...	..E.P.R...	A.D....Y.	.....G...	
H_SA_96_DRB1_1F	.....ST...	.....	F.D..Y...	.YV.....	.Y....A...	..RS..Y...	..E...RR.T	..D....Y.	.....	
I_DJ_29_DRB1_1F	.....	.....	F.D..Y...	.YV.....	.Y....A...	..D.KY...	..EV..RR.T	..D....Y.	.....I...	
J_DJ_53_DRB1_1F	.....AK...	R.....	F...Y...	.YV.....	.Y....A...	..D.KY...	..DF..R...	N.D....Y.	.....	
K_DJ_64_DRB1_1F	.....HK...	.....	F.D..Y...	.CV.....	.....AG.	..QS..Y...	..EL..RR.T	..D....Y.	.....	
L_SA_27A_DRB1_4F	.....ST...	.....	F.D..Y...	.YV.....	.Y....A...	..D.KY...	..DF..RA..	A.D....Y.	.....G...	
M_SA_52_DRB1_1F	.....SK...	.....	F.D..YT...	.NV.....	.....	..D.....	..DF..R..T	A.D....Y.	.....I...	
N_Ovar-DRB1*0401	-----	-----	-----	-----	HIS.S	II..SV.SPT	..SGCGT.TD	TSIMEKSTC.	STT.GASTER	WPSW.GRTP.
N_DJ_02_DRB1_1F	-----	-----	-----	-----	HIS.S	II..SV.SPT	..SGCGT.TD	TSIMEKSTC.	STT.GASTER	WPSW.GGAP.
O_DJ_04_DRB1_2F	-----	-----	-----	-----	HIS.S	IL..SVISST	..SGCGS.TD	TSIMEKSTC.	STA.GASTER	WPSW.GRTP.
P_DJ_50_DRB1_1F	-----	-----	-----	-----	HIS.S	IIG.SV.SPT	..SGCGT.TD	TSIMEKSTC.	STA.GASTER	WPSW.GRTP.
Q_SA_7A_DRB1_1F	-----	-----	-----	-----	HIS.S	ML..SVISPT	..SGCGT.TD	TSIMEKRPC.	STA.GASTER	WPSW.GRTP.
R_ovar-DRB1*0501	.....AK...	R.....	F...Y...	.TL.....	.Y....A...	..D.KY...	..EL..R...	N.D....Y.	.....G...	
R_SA_25A_DRB1_2F	.....AK...	R.....	F...Y...	.TL.....	.Y....A...	..D.KY...	..EL..R...	N.D....Y.	.....G...	
ovar-DRB1*0601	.....AK...	R.....	F...Y...	.YV.....	.Y....A...	..RS..Y...	..DF..R...	N.D....Y.	.....G...	
S_SA_024_DRB1_1F	.....AK...	R.....	F...Y...	.YV..N...	.Y....A...	..RS..Y...	..EL..R...	N.D....Y.	.....I...	
T_Ovar-DRB1*1402	.....TKK..	.....	.....Y...	.YV.....	.....A...	..E.KY...	..E...SR.T	A.D....Y.	.....I...	
DJ_80_DRB1_1F	.....TKK..	.....	.....Y...	.YV.....	.....A...	..E.KY...	..E...SR.T	A.D....H.	.....I...	
BTU_BoLA-DRB3	.....CKR..	.....	F.D..Y...	.YV.....	.....	..D.KY...	..E...RE..	Y.D....Y.	.....G...	
U_DJ_8_DRB1_2F	.....HK...	.....	F.D..Y...	.YV.....	.....A...	..RS..Y...	..E...R...	A.D....Y.	.....G..G	
v_EU176819.1_DRB1	.....TKK..	R.S.....	F.D.....	.TL.....	.Y....A...	..D..Y...	..DF..RA..	A.D....Y.	.....G...	
V_SA_23_DRB1_2F	.....TKK..	R.S.....	F.D.....	.TL.....	.Y....A...	..D.KY...	..DF..RA..	A.D....Y.	.....I...	
W_DJ_53_DRB1_2F	.....ST...	.....	F.D..Y...	.YV.....	.Y....A...	..D.KY...	..DF..RA..	A.D....Y.	.....G...	
X_SA_27A_DRB1_5F	.....AK...	R.....	F.D..Y...	.YV.....	.....A...	..RS..Y...	..DF..RA..	A.D....Y.	.....G...	
Y_SA_73_DRB1_5F	.....AKR..	.....	F.D..Y...	.YV.....	S.....A...	..RS..Y...	..DF..RA..	A.D....Y.	.....G...	

Table 4.9 Muscle alignment of the 26 *DRB1* Alleles using one member per group with (A to Z) in the two sheep populations using codon\_start 3 translated *DRB1\*0401* showing alignment with all the remaining *DRB1* alleles

A_ovar-DRB1*0901	HFLEYRSEC	HFFNGTERVR	LLERYFHNGE	EFARFDSDWG	EFRAVTELGR	PAAEQWNSQK	NILEQKRAEV	NTVCRHNYGV	FESFAVQRR
A_DJ_78_DRB1_5F	.....	.....	...K.....	.....	.....	.....	.....	.....	.....
Z_SA_029_DRB1_3F	.....	.....	.....	.....	.....	.....	DF..R..TA	D.Y.....	I...S....
B_Over-DRB1*1303	....TKK..	.....	F.D..YT..	.NV.....	.....	.D.....	E...R..TA	D.Y.....	I...S....
SA_52_DRB1_3F	....SK...	.....	F.D..YT..	.NV.....	.....	.D.....	DF..R..TA	D.Y.....	I...S....
C_SA_54_DRB1_5F	....SK...	..S.....	F.D..YT..	.NV.....	.....	.D.....	DF..R..TA	D.Y.....	I...S....
D_DJ_27_DRB1_2F	....SK...	.....	F.D..YT..	.NV.....	.....	.D.....	.....	D.....	.....
E_Over-DRB1*0302	....TKK..	R.S.....	F.D..Y...	.YV.....	.Y..A...	.D.KY....	E...RR.T.	D.Y.....	I...S....
SA_72_DRB1_1F	....TKK..	R.S.....	F.D.....	.Y.....	.Y..A...	.D.KY....	E...RR.T.	D.Y.....	G...T....
F_SA_80_DRB1_5F	....ST...	.....	F.D..Y...	.YV.....	.Y..A...	.D.KY....	E...RR.T.	D.Y.....	I...S....
G_DJ_008_DRB1_2F	....TKK..	R.S.....	F.D..Y...	.Y.....	.Y..A...	RS..Y....	E.P.R..A	D.Y.....	G...T....
H_SA_96_DRB1_1F	....ST...	.....	F.D..Y...	.YV.....	.Y..A...	RS..Y....	E...RR.T.	D.Y.....	...T....
I_DJ_29_DRB1_1F	.....	.....	F.D..Y...	.YV.....	.Y..A...	.D.KY....	EV..RR.T.	D.Y.....	I...S....
J_DJ_53_DRB1_1F	....AK...	R.....	F.....Y...	.YV.....	.Y..A...	.D.KY....	DF..R..N	D.Y.....	...S....
K_DJ_64_DRB1_1F	....HK...	.....	F.D..Y...	.CV.....	.....AG..	QS..Y....	EL..RR.T.	D.Y.....	...S....
L_SA_27A_DRB1_4F	....ST...	.....	F.D..Y...	.YV.....	.Y..A...	.D.KY....	DF..RA..A	D.Y.....	G...T....
M_SA_52_DRB1_1F	....SK...	.....	F.D..YT..	.NV.....	.....	.D.....	DF..R..TA	D.Y.....	I...S....
n_Over-DRB1*0401	....HK...	R.S.....	Y.D..Y...	.YV..N...	.Y..A...	.D.KY....	DF...T...	D.Y.....	I...S....
n_DJ_02_DRB1_1F	....HK...	R.S.....	Y.D..Y...	.YV..N...	.Y..A...	RS..Y....	EL...T...	D.Y.....	I...S....
o_DJ_04_DRB1_2F	....ST...	.....	F.D..Y...	.YV.....	.Y..A...	.D.KY....	DF..RA..A	D.Y.....	I.G.S....
p_DJ_50_DRB1_1F	.....	R.S.....	Y.D..Y...	.YV.....	.Y..A...	.D.KY....	EL..R..N	D.Y.....	I...S....
q_SA_7A_DRB1_1F	....AK...	..S.....	Y.D..Y...	.TL.....	.Y..A...	.D.KY....	E...R..N	D.Y.....	I...S....
R_ovar-DRB1*0501	....AK...	R.....	F...Y...	.TL.....	.Y..A...	.D.KY....	EL..R..N	D.Y.....	G...T....
R_SA_25A_DRB1_2F	....AK...	R.....	F...Y...	.TL.....	.Y..A...	.D.KY....	EL..R..N	D.Y.....	G...T....
ovar-DRB1*0601	....AK...	R.....	F...Y...	.YV.....	.Y..A...	RS..Y....	DF..R..N	D.Y.....	G...T....
S_SA_024_DRB1_1F	....AK...	R.....	F...Y...	.YV..N...	.Y..A...	RS..Y....	EL..R..N	D.Y.....	I...S....
T_Over-DRB1*1402	....TKK..	.....	.....Y...	.YV.....	....A...	.E.KY....	E...SR.TA	D.Y.....	I...S....
DJ_80_DRB1_1F	....TKK..	.....	.....Y...	.YV.....	....A...	.E.KY....	E...SR.TA	D.H.....	I...S....
BTU_BoLA-DRB3	....CKR..	.....	F.D..Y...	.YV.....	.....	.D.KY....	E...RE..Y	D.Y.....	G...T....
U_DJ_8_DRB1_2F	....HK...	.....	F.D..Y...	.YV.....	....A...	RS..Y....	E...R..A	D.Y.....	G.G.T....
v_EU176819.1DRB1	....TKK..	R.S.....	F.D.....	.TL.....	.Y..A...	.D..Y....	DF..RA..A	D.Y.....	G...S....
V_SA_23_DRB1_2F	....TKK..	R.S.....	F.D.....	.TL.....	.Y..A...	.D.KY....	DF..RA..A	D.Y.....	I...T....
W_DJ_53_DRB1_2F	....ST...	.....	F.D..Y...	.YV.....	.Y..A...	.D.KY....	DF..RA..A	D.Y.....	G...T....
X_SA_27A_DRB1_5F	....AK...	R.....	F.D..Y...	.YV.....	....A...	RS..Y....	DF..RA..A	D.Y.....	G...T....
Y_SA_73_DRB1_5F	....AKR..	.....	F.D..Y...	.YV.....S	....A...	RS..Y....	DF..RA..A	D.Y.....	G...T....

### 4.3.3.1 Phylogenetics of the *DRB1*, *DQA2* and *DQA1* alleles in Djallonke and Sahelian sheep populations based on MHC-IPD nomenclature classification

#### 4.3.3.1.1 *OLA-DRB1* coding sequence alignment

One amino acid sequence from each of the 26 classified *DRB1* alleles (A to Z alleles) identified within the two sheep populations was used for the group alignment to estimate the number of polymorphic sites. The product of the codon\_start 1 (used for the GenBank reference) translated *DRB1\*0401* allele group (Prefixed N, O, P & Q) completely misaligned with the remaining *DRB1* allele groups (Table 4.8). However, the codon\_start =3 translation frame for the *DRB1\*0401* allele (Table 4.9) aligned well with the rest of the *DRB1* alleles, and produced an 89-amino acid sequence. For either translation frame, applying the IPD-MHC allele nomenclature yielded an equal number (three) of *DRB1\*0401* allele variants (O, P, and Q). There were 33 polymorphic amino acid sites identified within the CDS sequences of the *DRB1* alleles of the two sheep populations. Analysis of the evolutionary relationships “within allele” of all three loci using three different phylogenetic models (ML, MP & PNJ) showed that the clustering characteristics of allele group members were generally conserved in all three models.

Constructed phylogenetic trees for the 26 classified *DRB1* alleles in the two sheep populations (Figure 4.12 & Figure 4.13) revealed that the codon\_start=1 translated *DRB1\*0401* allele and its variants cluster significantly (100% bootstrap support for 100 replicates) distantly from all the other *DRB1* allele groups. The extent of deviation of the codon\_start=1 translated *DRB1\*0401* allele is evident with respect to the position of the *BOLA-DRB3* (NCBI accession) allele highlighted in orange, which is derived from a different ruminant species, but clusters more closely to the group of remaining *DRB1* alleles. Conversely, the same analysis for a re-constructed tree using the codon\_start=3 translated *DRB1\*0401* allele showed that the group clustered closer to the other *DRB1* alleles (Figure 4.14 & Figure 4.15). Analysis of evolutionary distance using the Poisson correction method revealed that the sum of the length of tree branches of the codon\_start=1 translated *DRB1\*0401* allele was approximately 2.64 evolutionary units, compared to only 1.51 evolutionary units for the codon\_start=3 translated *DRB1\*0401* version.

In the re-constructed *DRB1* allele trees (Figure 4.14 & Figure 4.15), the clustering characteristics of the majority of the 26 *DRB1* allele sequences (prefixed A to Z) show

clustering with like group members, with one minor deviation (Figure 4.14). Three alleles (DJ\_04\_DRB1\_2F, SA\_27A\_DRB1\_4F & DJ\_53\_DRB1\_4F) belonging to three different allele groupings (O, L & W) respectively, clustered on the same internal node with high bootstrap support of 90% (Figure 4.14). The *DRB1\*0901* allele is the most distinct allele, out of the 26 classified *DRB1* alleles and supported by the highest bootstrap value (100%) (Figure 4.14). The tree also revealed two main embranchments at the innermost node, one branch (the smaller one) supporting only six of the *DRB1* alleles (A, B, C, D, M & Z) and the bigger branch supporting all the remaining 20 *DRB1* alleles (Figure 4.14).

#### **4.3.3.1.2 OLA-DQA2 coding sequence alignment**

Multiple sequence alignment of the *OLA-DQA2* alleles in the two sheep populations using one representative per group (A-M alleles) revealed that 42 (greater than half) of the amino acid sites of the 81 possible positions of CDS had polymorphic variants (Table 4.10). The *OLA-DQA2\*1401* allele is 81 amino acids in length, one less than all the other observed *OLA-DQA2* alleles. A phylogenetic tree constructed using the 13 *DQA2* alleles revealed two main internal node embranchments; one node supporting the *DQA2* alleles A, B, E, H, I, J & L, and the other supporting alleles C, D, F, G, K and M. This observation suggests that the *DQA2* alleles in this population may have possibly evolved from two main ancestral alleles. Analysis of the evolutionary relationships among these alleles using the Poisson's correction method revealed that the sum of all branch lengths in the tree is approximately, 1.23 evolutionary units. Phylogenetic analysis using MP, ML & PNJ for *OLA-DQA2* alleles showed that allele group members belonging to *OLA-DQA2\*08011* (C) and *OLA-DQA2\*08012* (M) cluster together on the same node (Figure 4.16). Also, the *OLA-DQA2\*0103* allele (H) and the EU176819.1 allele (I) cluster on the same external node. However, alleles belonging to each of the two groups (C & M and H & I), differ by the number of intronic non-synonymous substitutions. Alleles belonging to groups *OLA-DQA2\*0601* (B) and *OLA-DQA2\*0701* (K) are the most distinctive, sharing the highest bootstrap value of 100 each. The phylogenetic relationship between the different *DQA2* alleles were generally conserved across the three phylogenetic models.

#### **4.3.3.1.3 OLA-DQA1 coding sequence alignment**

Multiple sequence alignment of the 13 groups of *DQA1* alleles (A-M) obtained using the MHC-IPD nomenclature revealed a total of 37 polymorphic amino acid sites, less than half of the total (Table 4.11). Phylogenetic analysis of the *DQA1* alleles revealed two broad internal

node enbranchments; one supporting allele groups A, B, C, D, E, K, L & M and the other supporting allele groups F, G, H, I, & J (Figure 4.17). As with the *DQA2* locus, this analysis also suggests that all the *DQAI* alleles in these two breeds may have evolved from two main ancestral alleles. The phylogeny showed that most of the 13 different *DQAI* allele groups generally clustered by group, except for the M3304.1 group (B, C, & D) and the LN736359.1 group (K). All three M3304.1 allele group variants clustered on different branches. The M3304.1 variant-1 (C), in particular, clustered on the same internal node with the LN827894.1 allele group (M) (Figure 4.17). The LN736359.1 (K) group clustered on the same external node as the LN827892.1 group (L). Bootstrapping analysis with 1000 replicates revealed that the two alleles have an equally high level of bootstrap support (100%). Although the two alleles (L & K) are similar at the CDS, they do differ significantly in the intronic regions by eight nucleotide substitutions. The intronic differences, however, are of lesser relevance in the MHC nomenclature classification. The values obtained for the alleles by bootstrapping with 1000 replicates indicates that none of the *DQAI* alleles can be considered to be distinct from the rest of the members. The sum of the length of the tree branches for all the *DQAI* alleles computed with the Poisson correction method is approximately 0.81 evolutionary units.

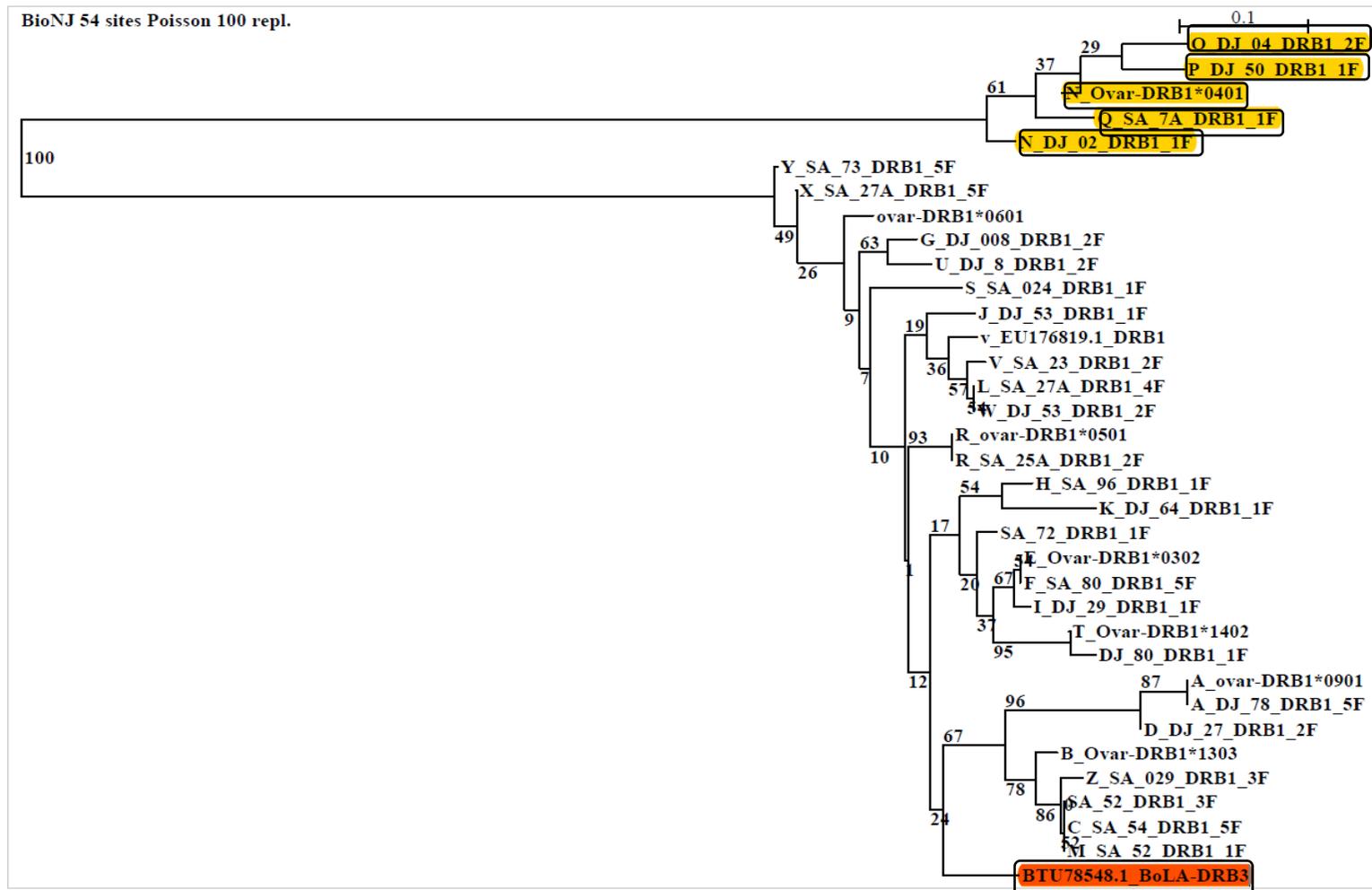


Figure 4.12 Phylogenetic Tree of 26 *DRB1* Alleles (A-Z) in Djallonke and Sahelian sheep using a codon\_start=1 translation frame for the *DRB1\*0401\_1* allele. The numbers at the nodes of each cluster denote the percentage support for that cluster in the constructed evolutionary relationship per 100 bootstrap replicates. The codon\_start=1 translated allele group are highlighted in yellow showing very significant deviation (100% bootstrap support) from all other *DRB1* alleles and the *BoLA-DRB3* (outgroup) is highlighted in orange.



Figure 4.13 Poisson Neighbour joining tree of *DRB1* exon 2 CDS (A to Z) in Djallonke and Sahelian sheep using a codon\_start=1 translation frame for the *DRB1\*0401* allele (highlighted in yellow). The cattle *BoLA-DRB3* (NCBI accession BTU78548.1) as an outgroup (orange), clusters with the *Ovar-DRB1* alleles. The scaled evolutionary distance tree revealed a very prominent deviation of the \*0401 group from remaining *Ovar-DRB1* alleles

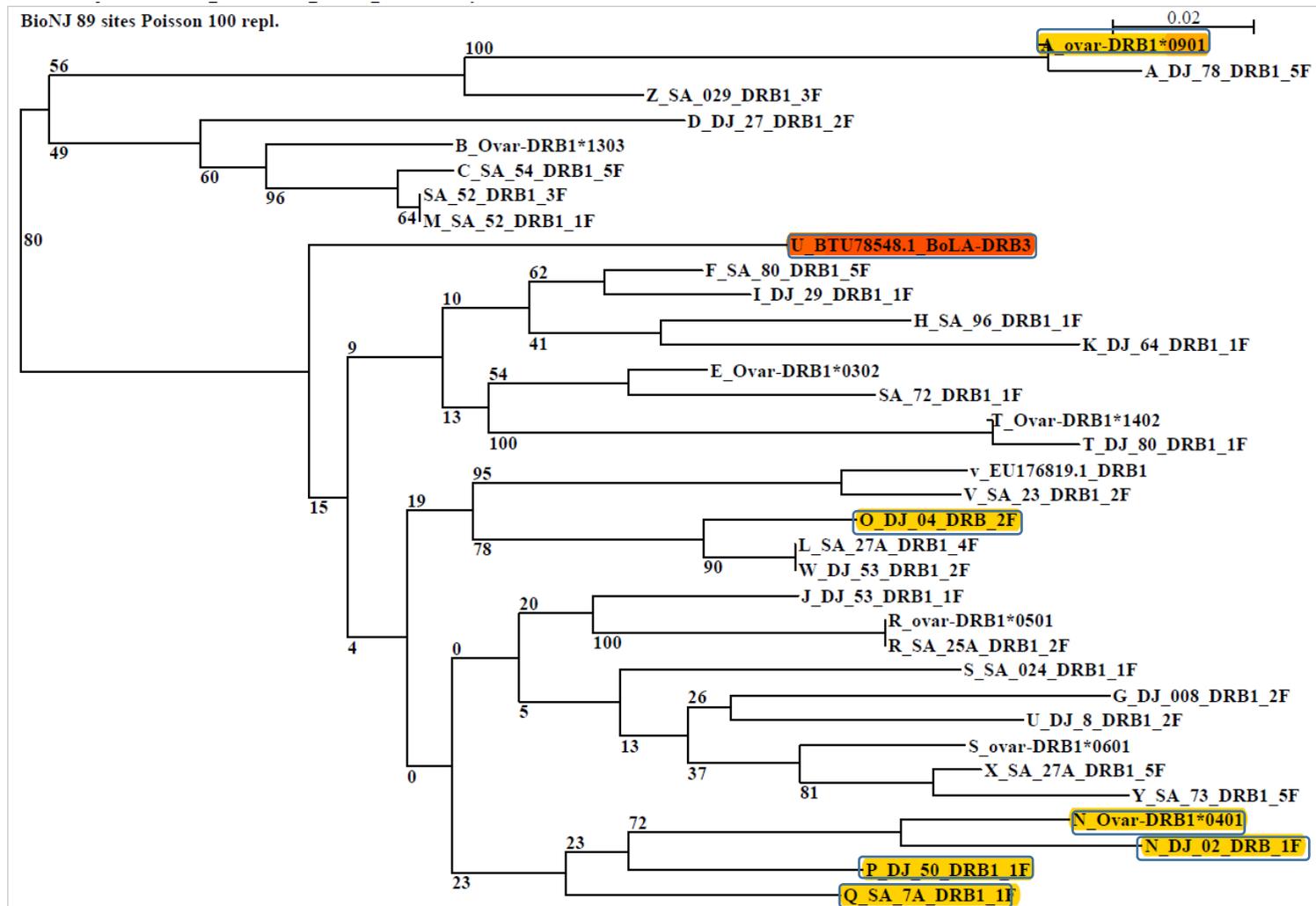


Figure 4.14 Reconstructed phylogenetic Tree of the 26 *DRB1* Alleles (A to Z) in Djallonke and Sahelian sheep showing the codon\_start=3 translation of *DRB1\*0401* alleles (prefixed with letters N, O, P, & Q). The numbers at the nodes of each cluster denote the percentage support for that cluster per 100 bootstrap replicates. The *Ovar-DRB1\*0901* is shown as the most distinctive allele. The outgroup *BoLA-DRB3* is prefixed with the letter U.

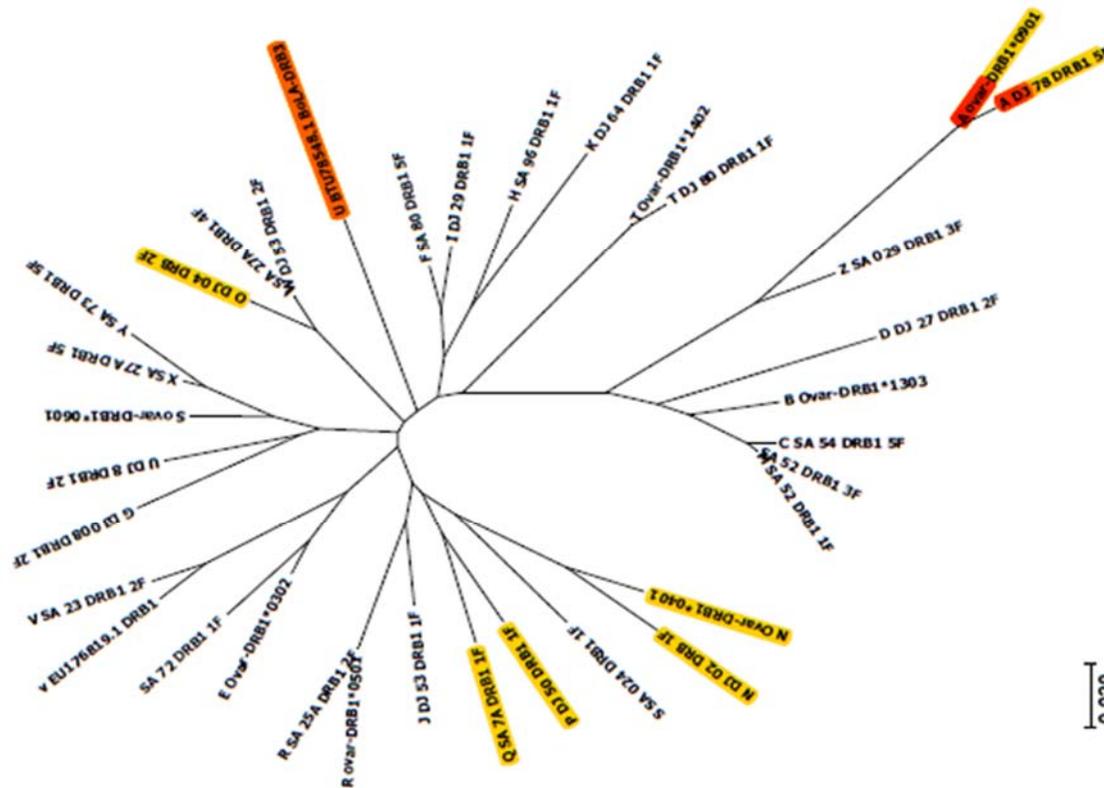
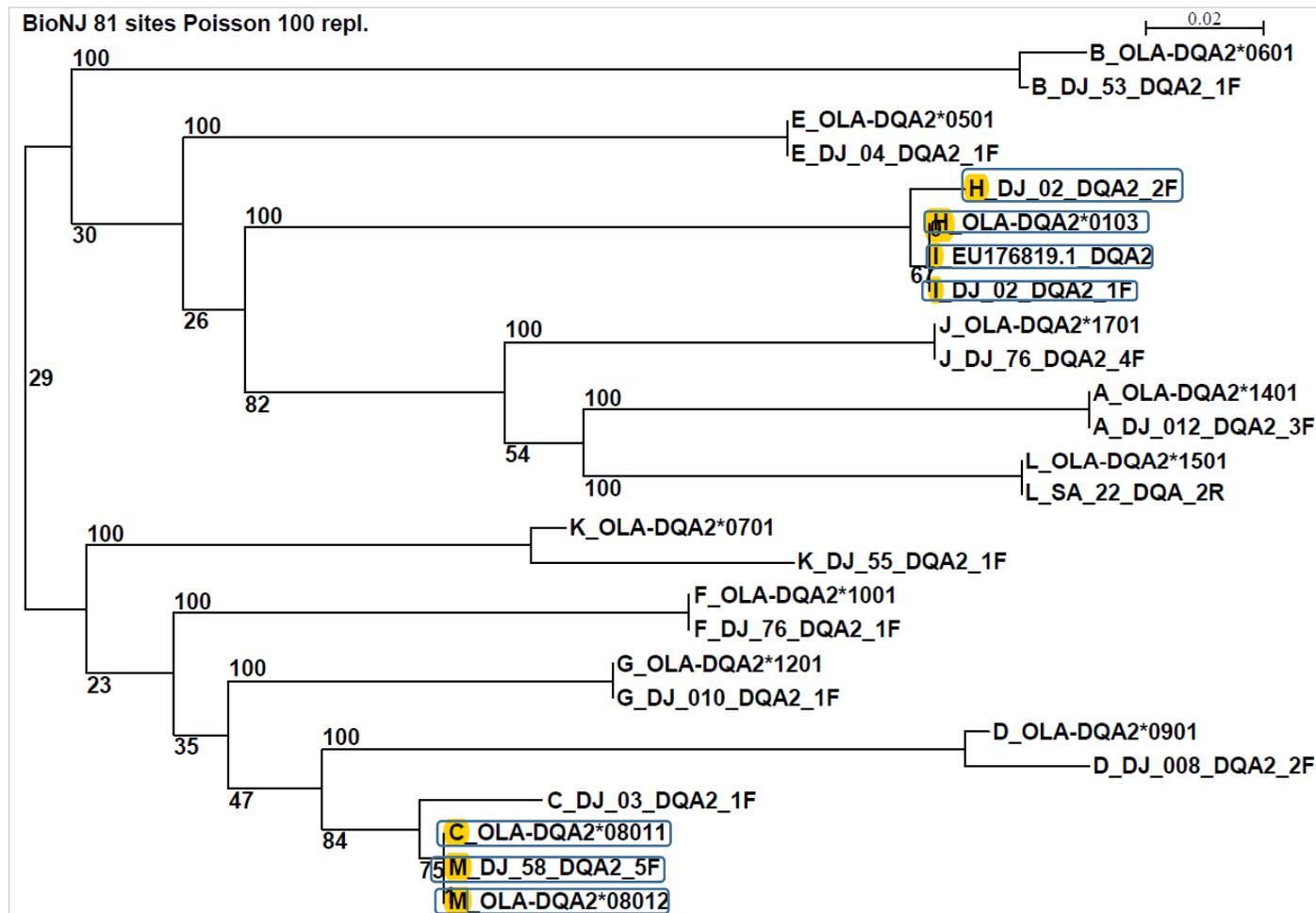


Figure 4.15 Poisson Neighbour joining tree of *DRB1* exon 2 CDS (A to Z) in Djallonke and Sahelian sheep using a codon\_start=3 translation frame for the *DRB1\*0401* allele (highlighted in yellow). The cattle *BoLA-DRB3* (NCBI accession BTU78548.1) as an outgroup (orange), clusters with the *Ovar-DRB1* alleles. The scaled evolutionary distance tree revealed that the \*0401 group now clusters with all other *Ovar-DRB1* alleles. The *ovar-DRB1\*0901* group (highlighted in yellow and orange) is now the most distinct allele in the tree

Table 4.10 Muscle alignment of the 13 *DQA2* Alleles (A to M) in Djallonke and Sahelian sheep populations with respective

GenBank references

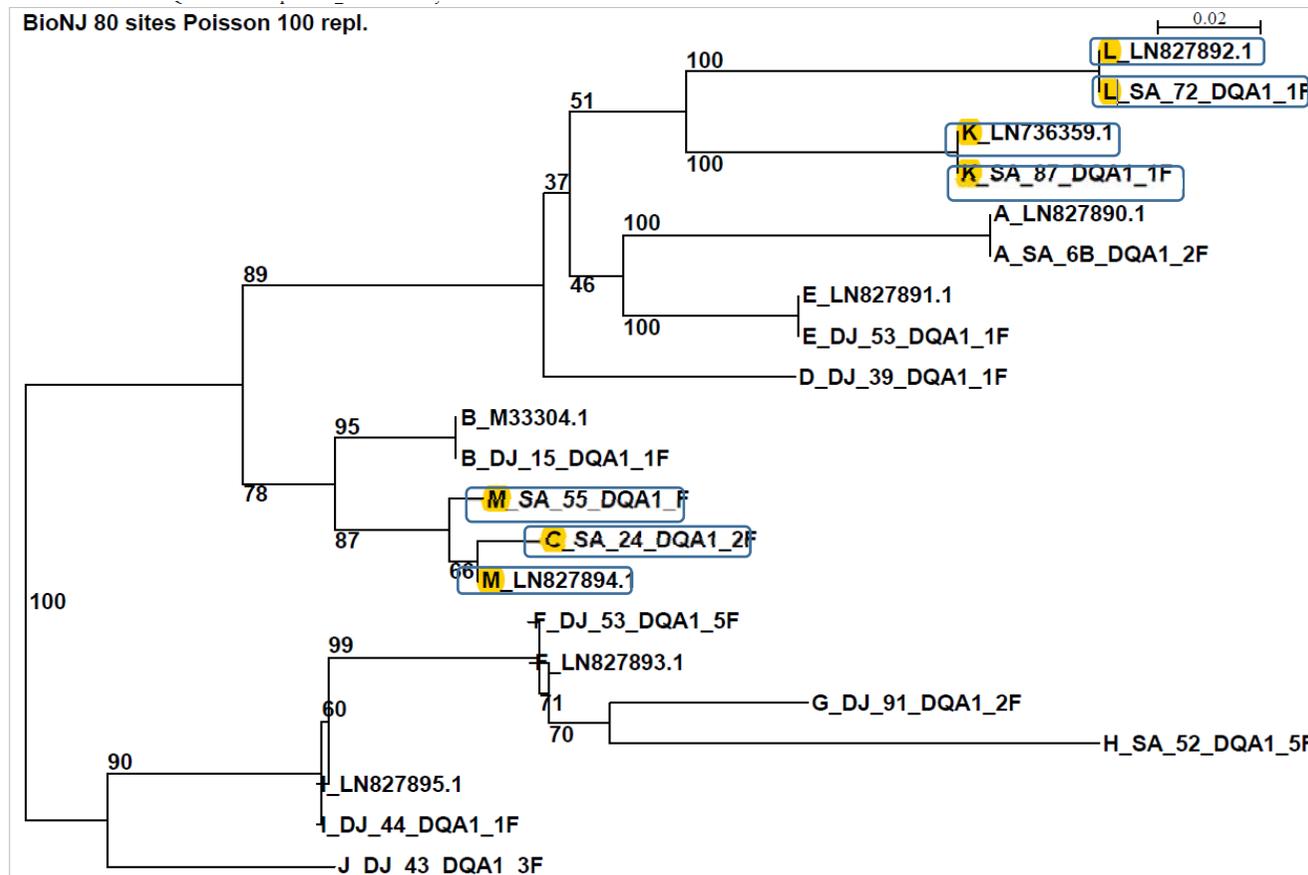
A_OLA-DQA2*1401	DHIGTYGTDF	YQSHGSPSQY	IHEFDGDEQL	YVDLEKKETV	WRLPMFDGL-	SFDPQRALSN	IAIAKHNLDR	LTRWYNFTPV	IN
A_DJ_012_DQA2_3F	.....	.....	.....	.....	.....	.....	.....	.....	..
B_OLA-DQA2*0601	..V...AE.	.....SE.	TQ...E..L.	.....	.....GQFA	G.HI.V....	..T.....V	M.....	..
B_DJ_53_DQA2_1F	..V...AE.	.....SE.	TQ...E..L.	.....	.....GQFA	G.HI.V....	..T.....V	M..R.....	..
C_OLA-DQA2*08011	..F.S...TI	.....F	TQ.....LF	.....	.....SQFA	G...G....	..A.....I	..RS.S...	..
C_DJ_03_DQA2_1F	..F.S...TI	.....F	TQ.....L.	.....	.....SQFA	G...G...S	..A.....I	..RS.S...	..
D_OLA-DQA2*0901	....S...TI	.....S..	TQ.....LF	.....	.....SQFA	G.NI.D..N.	..PA.....GI	..RS.....	..
D_DJ_008_DQA2_2F	....S...TI	.....S..	TQ.....LF	.....	.....SQFA	G.NV.D..N.	..PA.....GI	..RS.....	T.
E_OLA-DQA2*0501	..V.C.S.II	.....F	T.....L.	.....	.....I.GE.T	.....G....	..T.....I	..CS.C...	..
E_DJ_04_DQA2_1F	..V.C.S.II	.....F	T.....L.	.....	.....I.GE.T	.....G....	..T.....I	..CS.C...	..
F_OLA-DQA2*1001	..V.S....	.....F	TQ.....L.	..G....	.....SQFA	D...G..R.	..T..DT..I	..RS.S...	..
F_DJ_76_DQA2_1F	..V.S....	.....F	TQ.....L.	..G....	.....SQFA	D...G..R.	..T..DT..I	..RS.S...	..
G_OLA-DQA2*1201	..F.S...EI	.....	TQ.....LF	..G....	.....SQFA	G...G...E	..T..Q...I	..RS...A..	..
G_DJ_010_DQA2_1F	..F.S...EI	.....	TQ.....LF	..G....	.....SQFA	G...G...E	..T..Q...I	..RS...A..	..
H_OLA-DQA2*0103	..V.I..A..	.....	T.....LF	..G....	.....GEFT	.....G...E	..K..QT..I	MI.RS....	..
H_DJ_02_DQA2_2F	..V.I..A.S	.....	T.....LF	..G....	.....GEFT	.....G...E	..K..QT..I	MI.RS....	..
I_EU176819.1_DQA	..V.I..A..	.....	T.....LF	..G....	.....GEFT	.....G...E	..K..QT..I	MI.RS....	..
I_DJ_02_DQA2_1F	..V.I..A..	.....	T.....LF	..G....	.....GEFT	.....G...E	..K..QT..I	MI.RS....	..
J_OLA-DQA2*1701	..V....N.	.....	..L....RF	.....	.....GE.I	.....G....	V.AS....I	..RS....	..
J_DJ_76_DQA2_4F	..V....N.	.....	..L....RF	.....	.....GE.I	.....G....	V.AS....I	..RS....	..
K_OLA-DQA2*0701	..V.S...VI	.....	T...R..LF	.....	.....SQFA	G...G....	..T.....I	M.Q.H.S...	..
K_DJ_55_DQA2_1F	.....VI	.....	T...R..LF	.....	.....SQFA	GS...G....	..T...D..I	M.Q.H.S...	..
L_OLA-DQA2*1501	..V.I....	.....E.	..L....EF	.....	.....E.R	R...G..N.	.....I	..R.....	..
L_SA_22_DQA_2R	..V.I....	.....E.	..L....EF	.....	.....E.R	R...G..N.	.....I	..R.....	..
M_OLA-DQA2*08012	..F.S...TI	.....F	TQ.....LF	.....	.....SQFA	G...G....	..A.....I	..RS.S...	..
M_DJ_58_DQA2_5F	..F.S...TI	.....F	TQ.....LF	.....	.....SQFA	G...G....	..A.....I	..RS.S...	..



**Figure 4.16** Poisson Phylogenetic Tree of the 13 *DQA2* Alleles (A to M) in Djallonke and Sahelian sheep with respective GenBank references. The numbers at the nodes of each cluster denote the percentage support for that cluster in the reconstructed evolutionary relationship per 100 bootstrapped replicates. The *OLA-DQA2\*08011* (C) and *OLA-DQA2\*08012* (M) alleles cluster on the same node and two other alleles (H & I) also cluster on the same node. Bootstrap support is variable (23 to 100%), but is predominantly very high.

Table 4.11 Muscle alignment of the 13 *DQA1* Alleles (A to M) in Djallonke and Sahelian sheep populations with respective GenBank references

A_LN827890.1	DHIAAYGINV	YHKYGPSGY	THEFDGDEEF	YVDLEKRETV	WHLPMFSKFR	RFDPQGALRN	IAAAKHNVLEV	LIQDSNSTAA	SN
A_SA_6B_DQA1_2F	.....	.....	.....	.....	.....	.....	.....	.....	..
B_M33304.1	...GT..V.I	.QT.....	.....	.....	.R..E...T	S.....	..TV.....I	...R.....	T.
B_DJ_15_DQA1_1F	...GT..V.I	.QT.....	.....	.....	.R..E...T	S.....	..TV.....I	...R.....	T.
C_SA_24_DQA1_2F	...GI..V..	.QT.....	.....	.....	.C..E...T	S.....	..T.....I	...M.....	TD
D_DJ_39_DQA1_1F	.....	.QT.....	.....	.....K...	...V..Q..	.....	..T.....I	T..R.....	T.
E_LN827891.1	.....	..S.....	.....	.....	...V..Q..	S.....	..TV.....	...R.....	T.
E_DJ_53_DQA1_1F	.....	..S.....	.....	.....	...V..Q..	S.....	..TV.....	...R.....	T.
F_LN827893.1	...GT..V..	.QT..S...	.....Q.	.....	.R....E	Y..A.F...	..TF.....L	M..R.....	T.
F_DJ_53_DQA1_5F	...GT..V..	.QT..S...	.....Q.	.....	.R....E	Y..A.F...	..TF.....L	M..R.....	T.
G_DJ_91_DQA1_2F	...GT*.V..	.QT..S...	.....QS	.E.....	.C....E	Y..A.F...	..TF.....L	M..R.....	..
H_SA_52_DQA1_5F	...GT..V..	.QT..S..H.	.....Q.	.....	.R....E	Y..A.F...	..TF.....L	M..R.TLLLL	PI
LN827895.1	...GT..V..	.QT.....	.....	.....	.R....E	Y...F...	..TF.....L	M..R.....	T.
I_DJ_44_DQA1_1F	...GT..V..	.QT.....	.....	.....	.R....E	Y...F...	..TF.....L	M..R.....	T.
J_DJ_43_DQA1_3F	...GT..V..	.QT.....F	.....Q.	.....	.R.....G	D...F...	..TL.....L	I..R.....	T.
K_LN736359.1	.....	..S.....	.....	.....	.R..V..Q..	.....	..VG.QS..I	...R.....	T.
K_SA_87_DQA1_1F	.....	..S.....	.....	.....	.R..V..Q..	.....	..VG.QS..I	...R.....	T.
L_LN827892.1	.....	..S.....	.....Q.	.....K...	.R..E...T	S.....	M.SG.QT..I	M..R.....	T.
L_SA_72_DQA1_1F	.....	..S.....	.....Q.	.....K...	.R..E...T	S.....	M.SG.QT..I	M..R.....	T.
M_LN827894.1	...GI..V..	.QT.....	.....	.....	.R..E...T	S.....	..T.....I	...M.....	T.
M_SA_55_DQA1_F	...GI..V..	.QT.....	.....	.....	.R..E...T	S.....	..TT.....I	...M.....	T.



**Figure 4.17 Neighbour Joining Phylogenetic Tree of *DQA1* Alleles (A to M) in Djallonke and Sahelian sheep with respective GenBank references. The numbers at the nodes of each cluster denote the percentage support for that cluster in the re-constructed evolutionary relationship per 100 bootstrapped replicates. Two alleles (K&L) cluster on the same note with very high (100%) bootstrap support, and two other alleles (M & C) cluster with moderately high (66%) bootstrap support.**

## 4.4 Discussion

The multiple sequence analysis of all the *DRB1* (Table 4.2), *DQA2* (Table 4.3) and the *DQA1* (Table 4.4) putative amino acid sequences derived from the two sheep populations suggests that the *DRB1* locus (83 different sequences) is the most diverse and the *DQA2* locus (41 different sequences) is the least diverse for these two sheep breeds. Furthermore, an evolutionary distance analysis of the unique amino acid sequences for the three loci of the two sheep populations confirmed the *DRB1* locus as being the most diverse, as shown by it having the highest sum of evolutionary distance (1.90 evolutionary units) (Figure 4.3). In contrast, the *DQA1* allele tree showed the lowest diversity (Figure 4.9) with the smallest sum of evolutionary distance (1.26) compared to 1.72 for the *DQA2* allele tree (Figure 4.6). This observation agrees with previous reports suggesting that the *DRB1* locus is the most diverse of all the sheep MHC loci (Ballingall et al., 2011; Charon, 2004). Inter-population comparison of the evolutionary distance showed that the Djallonke sheep overall had a higher diversity at all three loci when compared to the Sahelian population. Greater diversity at the MHC has been suggested as being associated with higher population resistance to diseases and general fitness (Aguilar et al., 2004). The observation in this study supports the reported differential resistance to parasitic diseases between these two sheep breeds. Specifically, the Djallonke sheep generally have higher resistant to parasitic diseases than the Sahelian sheep (Goossens et al., 1999; Traoré et al., 2017).

The classification of all of the alleles identified in this study, using the IPD-MHC nomenclature, revealed that the *DRB1\*0401* allele sequence group in the multiple sequence alignment of the *DRB1* alleles showed a distinctly pronounced frame shift (Table 4.8). The large magnitude of the deviation for this group of alleles (codon\_start=1 translated frame) from the other *Ovar-DRB1* alleles, as determined using Poisson's evolutionary distance method, clearly suggests that the translation frame is incorrect for the *Ovar-DRB1\*0401* alleles (Figure 4.12 & Figure 4.13). A proposed translation frame using codon\_start=3 for *Ovar-DRB1\*0401* alleles appeared to conform with all other *DRB1* alleles as represented in Table 4.9. Comparison of the Poisson's evolutionary distance analyses for these two translations of the *DRB1\*0401* alleles (codon\_start=3 & codon\_start=1) clearly showed that the codon\_start=3 translation is probably the correct translation frame (Figure 4.14 & Figure 4.15). Furthermore, this conclusion is supported by the congruent clustering characteristics observed with the cattle *BoLA-DRB3* allele (GenBank accession) used as an outgroup in relation to the remaining *Ovar-*

*DRB1* alleles in the phylogenetic trees (Figure 4.12, Figure 4.13, Figure 4.14 & Figure 4.15). Although, it is not uncommon for ovine MHC class II genes to share similar polymorphisms that precede speciation with that of the bovine (Bryja et al., 2006; Zhou & Hickford, 2004), the *Bola-DRB3* position on the trees (Figure 4.12 Figure 4.13 Figure 4.14 & Figure 4.15) brings evolutionary perspective to the magnitude of the deviation of the codon\_start=1 translated \*0401 allele group. These analyses have demonstrated that the codon\_start=1 translation frame used for the *Ovar-DRB1\*0401* allele (GenBank Accession AM884914.1) in the NCBI database may be wrong. Therefore, the codon\_start=3 translated *DRB1\*0401* alleles were used for all subsequent analyses in this study. Interestingly, a recent study has associated variations within *BoLA-DRB3* with the infection levels of bovine leukaemia virus in two dairy cattle breeds (Carignano et al., 2017).

The rationale for construction of genotype discovery curves (Figure 4.1 & Figure 4.1 Figure 4.2) was to assess if the sample size of 100 sheep for each of the two breeds was adequate for sampling the majority, if not all, of the possible MHC class IIa genotypes present in the populations. The genotype discovery curves for all three loci plateaued within the range of the sampled population, which implies that most of the alleles within the population of the two breeds have been accounted for in this study. However, the presence of null alleles in all three loci suggests that potentially more alleles could be discovered. Also, the general characteristic of the genotype curve for the Djallonke *DRB1* locus in terms of steepness of slope and plateauing at the 95th sample suggests a possibility of the presence more *DRB1* alleles at this locus in the population.

In terms of allele novelty, based on MHC-IPD allele standard classification guidelines (<http://www.ebi.ac.uk/ipd/mhc/group/OLA/>) a very high percentage (73%) of the *DRB1* alleles are new, compared to 38% new alleles for *DQA1* and none for *DQA2*. A recent study of the same class II region in 235 Texel sheep found only 11% new *DRB1* alleles and none for either the *DQA2* or the *DQA1* loci (Ali et al., 2017). The computed evolutionary relationships based on the MHC-IPD classified allele grouping also showed the same trend in locus diversity; the *DRB1* being the most diverse for the Djallonke and Sahelian populations and the *DQA1* being the least diverse (Figure 4.14, Figure 4.16 & Figure 4.17). The total evolutionary distance obtained for the *DQA1* locus was 0.81 units compared 1.23 units at the *DQA2* locus. The differences in the evolutionary distances obtained for the three loci in this latter classification compared to that obtained for all unique sequences per locus (Figure 4.3, Figure 4.6 & Figure 4.9) were to be expected because only one representative sequence of the re-classified allele group was used to perform this analysis. Although, the use of one sequence from each classified

allele group within each locus might not be an adequate representation of all possible variation, the trees are informative and permitted a standardised comparison with published literature to be performed.

This high number of new alleles for *DRB1* and *DQA1* is not unexpected because most of the MHC alleles within the MHC-IPD and GenBank databases are derived from different domestic sheep breeds to the ones studied here. Furthermore, most of these sheep breeds have evolved in different regions of the world, and have different demographic history of diseases and environmental challenges.

An earlier report of the *DRB1* genotyping of flocks of 105 Texel and 71 Suffolk sheep identified a total of only 8 and 7 *DRB1* alleles, respectively (Sayers et al., 2005). A recent study using the same markers in a flock of 235 British Texel sheep identified fewer *DRB1* (18), *DQA2* (8) and *DQA1* (7) alleles (Ali et al., 2017) than were observed in this study. Although, some individuals in the Texel sheep study were known to be related, probably contributing to a lower observed allelic diversity, the difference in numbers of alleles found between the two studies is quite remarkable. Only three *Ovar-DRB1* alleles were found to be common between the two studies (\*0901, \*0401 & \*0302). These three common *DRB1* alleles were also identified in the Djallonke population, although only two were present in the Sahelian population (\*0401 & \*0302).

The *DRB1\*0901* allele is common in the British Texel and the Djallonke sheep. Given that, unlike the Sahelian, these two breeds are known for their higher resistance to gastrointestinal nematodes. These observations may suggest that the *DRB1\*0901* allele contributes to resistance to nematodes in sheep. However, a previous study (Table 4.12) identified the *DRB1\*0901* allele in flocks of Scottish Blackface, Greyface, and Suffolk sheep breeds (Ballingall & Tassi, 2010). The Suffolk and Greyface sheep, unlike the Texel breed, have been shown to be more susceptible to GIT nematode infections (Good et al., 2006, Sayers et al 2005), thus discounting any proposition of direct association of the *DRB1\*0901* allele with resistance to nematode infections. However, an association of specific mutations derived from the *DRB1\*0901* allele and resistance to nematode infection cannot be ruled out.

All the MHC gene studies in sheep so far have shown that the *DRB1* gene is the most polymorphic (Charon, 2004; Davies et al., 2006; Hassan et al., 2011; Sayers et al., 2005). Similarly, three of the 13 *DQA1* allele types identified in this study (\*0901, \*0401, & \*0302) (Table 4.12), were in common with those identified by Ali et al. (2017). However, all these three *DQA1* alleles are present within both the Djallonke and Sahelian populations.

The *OLA-DQA2\*1401* allele, also referred to as the “*DQA2*-like” sequence (Ali et al., 2017; Ballingall et al., 2015; Hickford et al., 2004), was identified in this study. This “*DQA2*-like” allele has only 81 exon 2 CDS amino acids as compared to 82 in the normal sheep *DQA2* allele, and was reported to have low functionality (Hickford et al., 2004). The “*DQA2*-like” alleles have been the subject of conjecture regarding the continuous evolution of the MHC (Ali et al., 2017; Hickford et al., 2004). In a recent study, orthologues of the “*DQA2*-like” allele were reported to be present in both cattle and goats, and it was suggested that the allele resulted from a possible recombination event that pre-dated speciation, and it was suggested that it is functional in sheep (Ballingall et al., 2015). Interestingly, whereas one report has suggested that “*DQA2*-like” alleles are usually found in a *DQA1* null MHC class II haplotype (Ballingall et al., 2015), a later study has shown an MHC class II haplotype containing both the “*DQA2*-like” allele and *DQA1* allele (Ali et al., 2017). This study has identified both types of haplotypes, and it is clear that, as more breeds are genotyped, our understanding of the diversity of the ovine MHC will be greatly expanded. Interestingly, in a footrot vaccine study in sheep, specific MHC class II haplotypes containing the “*DQA2*-like” alleles have been shown to significantly ameliorate the disease outcomes (Ennen et al., 2009). In a more recent study using dairy sheep, a variety of *DQA2* alleles and haplotypes were shown to be significantly associated with either lower or higher susceptibility to foot rot infection (Gelasakis et al., 2013). However, foot rot is known to be a more significant constraint to production in temperate livestock production systems in comparison to tropical production systems (Raadsma & Egerton, 2013). Another study has indicated a more complex “*DQA2*-like” allele association with gastrointestinal parasitic infections in sheep, affected by both the age of the sheep as well as the species of parasite(s) studied (Hickford et al., 2011). The conclusion of this study suggested doubt as to the potential effectiveness of markers based on a single MHC gene in their use for disease control. This conclusion is not unexpected given that resistance to gastrointestinal parasites is known to be a complex quantitative trait. Such complex traits in livestock are thought to be controlled by a combination of many genomic mutations, each having a small effect, together with the presence of a few mutations with larger effects. It has been reported that this is a major constraint to the widespread application of marker assisted breeding (Georges, 2007; McRae et al., 2014; Meuwissen et al., 2016). Nonetheless, a *DQA2* allele marker based diagnostic test has been used commercially for evaluating and selection of tolerance for foot rot in sheep (Dukkipati et al., 2006). The differences in the *DQA2* alleles between the Djallonke and Sahelian sheep, particularly the four breed-specific *DQA2* alleles found in this study, will provide a valuable resource for studying the differential resistance to

gastrointestinal parasitic infection in these two sheep breeds. Twelve of the 13 *DQA2* alleles identified in this study, including the “*DQA2*-like” allele (Table 4.12), were also reported in six New Zealand sheep breeds (Hickford et al., 2004). All these 12 *DQA2* alleles were found in the Djallonke population, and 8 were found in the Sahelian population (Table 4.12). Another study involving 16 different New Zealand sheep breeds also reported five of the *DQA2* alleles found in this study (Hickford et al., 2007). All five were present in the Djallonke, whereas only four were present in the Sahelian (Table 4.12). Six of the *DQA2* alleles identified in this study have also been reported in the British Texel sheep (Table 4.12). All these six *DQA2* alleles were present in the Djallonke, but only three (\*0901, \*1401 & \*0601) were present in the Sahelian sheep, including the *DQA2*-like (\*1401) allele. Again, this suggests that the Djallonke sheep share three *DQA2* alleles exclusively with the British Texel sheep. However, more meaningful comparisons of these variations, both the shared and unshared MHC class II alleles, between breeds can be made when future disease association studies are undertaken in the Djallonke and Sahelian sheep populations.

**Table 4.12 Comparison of the known MHC Class IIa alleles in the Djallonke and Sahelian with reports from other sheep breeds**

Sheep Breed	MHC class IIa alleles			Reference
	<i>DRB1</i>	<i>DQA2</i>	<i>DQA1</i>	
<b>Djallonke</b>	*0901, *1303, *0302, *0401, *0501, 1402, <i>BoLA DRB3</i>	<b>*1401</b> , *0601, *08011, *0901, *0501, *1001, *1201, *0103, *1701, *0701, *08012, EU176819.1	LN736359, LN827890, LN827891, LN827892, LN823893, LN823895, M33304	This study
<b>Sahelian</b>	*1303, *0302, *0401, *0501, 1402, <i>BoLA DRB3</i>	<b>*1401</b> , *0601, *08011, *0901, *0501, *1201, *1701, *1501	LN736359, LN827890, LN827891, LN827892, LN823894, LN823893, LN823895, M33304	This study

Sheep Breed	MHC class IIa alleles			Reference
	<i>DRB1</i>	<i>DQA2</i>	<i>DQA1</i>	
<b>British Texel</b>	*0901, *0401, *0302,	*08012 *1001, *0103, *0601, *0901, * <b>1401</b>	LN736359, LN827890, M33304	(Ali et al., 2017)
<b>Suffolk</b>	*0501, *0901			(Ballingall & Tassi, 2010)
<b>Scottish Blackface</b>	*0302, *0501, *0901			(Ballingall & Tassi, 2010)
<b>Greyface</b>	*0302, *0501, *0901			(Ballingall & Tassi, 2010)
<b>Six New Zealand sheep breeds</b>		* <b>1401</b> , *0601, *08011, *0901, *0501, *1001, *1201, *0103, *1701, *0701, *08012, *1501		(Hickford et al., 2004)
<b>16 New Zealand sheep breeds</b>		*1201, *0301, *08011, *0901, *0601		(Hickford et al., 2007)

The fundamental mechanism for diversity at the MHC is thought to be a consequence of the accumulation of many point mutations over time (Andersson & Mikko, 1995), together with “non-reciprocal” type intra genic recombination (gene conversion) events occurring between alleles at different loci (Hedrick, 1998). Furthermore, mutations arising at the MHC are known to be maintained through balancing selection via the interplay of maternal-foetal interactions, pathogen-mediated selection and negative assortative mating (Aguilar et al., 2004; Charbonnel & Pemberton, 2005; Edwards & Hedrick, 1998). Historically, due to the agrarian nature of the

production system, livestock species within the SSA region are known to face many challenges, including high temperature, feed scarcity and diseases (Kosgey et al., 2006; Kosgey & Okeyo, 2007). These demographic factors have exerted strong natural selection pressures which have shaped the genomic architecture of these livestock species, and this may be reflected in the MHC genes. For example, African animal trypanosomiasis has remained an endemic disease in SSA for a millennium despite all the remedial efforts directed at its eradication (as reviewed in chapter .1 2). However, the magnitude of MHC allele novelty identified in this study provides a quantifiable measure as to the extent of genetic diversity and uniqueness of these two breeds with respect to reported information on sheep breeds from other regions of the world. These results provide the first molecular evidence of the diversity of these MHC loci in the Djallonke and the Sahelian sheep breeds.

Evidence of breed specificity of some of the alleles was observed for all the three MHC class IIa loci in this study. Equal numbers of breed specific *DRBI* alleles (5) were identified in the Djallonke and Sahelian breeds. Although, the two breeds have 21 *DRBI* alleles in common, the combined differences in terms of breed- specific alleles (total of 10 alleles) is quite significant. In comparison, a total of three breed specific alleles were identified in the Polypay and the Columbia sheep breeds (Herrmann et al., 2005). These observations become more interesting given that specific variations within *DRBI* alleles have been linked with resistance to parasitic nematode infection (Davies et al., 2006; Hassan et al., 2011; Schwaiger et al., 1995). The Djallonke sheep is known to be more resistant to nematode infection in comparison to Sahelian sheep (Goossens 1997, 1999, Alvarez et al., 2017). These differences in allele diversity might hold the key for unravelling the differences in resistance to nematodes between these two sheep breeds. Because this is the first report of the new alleles in these sheep breeds, these novel variations can be referred to as variants of unknown significance (VUS). The VUS will serve as valuable information for future association studies, such as for elucidation of the basis for disease resistance or other important adaptive traits in the two populations.

Interestingly, the *Ovar-DRBI\*1303* allele (GenBank Accession FR751085.1) identified in both the Djallonke and Sahelian in this study had previously been reported only in the Red Maasai sheep from Kenya (East African), suggesting a possible genetic relationship between all three of these indigenous African sheep breeds. A recent mitogenomic study indicated a common maternal co-ancestry for in the Djallonke and Sahelian sheep (Brahi et al., 2015). However, the presence of *DRBI\*1303* could have resulted from convergent evolution in response to a common adaptive selection pressure faced by these three sheep breeds within the SSA region rather than resulting from a common ancestry. This is because the Red Maasai,

unlike the Djallonke and Sahelian sheep (both are thin tailed sheep type), is a fat tailed-sheep type, and a previous report has indicated distinctly different genetic histories for sheep with different tail-types (Muigai & Hanotte, 2013).

#### **4.5 Conclusions**

Analyses of the class II *DRB1*, *DQA2* and *DQA1* loci in Djallonke and Sahelian sheep populations revealed commonality as well as breed specific alleles. The Djallonke population showed more diversity at all three MHC loci than the Sahelian population, suggestive of increased population fitness. This result is consistent with the observed phenotypes in the Djallonke and Sahelian breeds as revealed by previous parasitological and haematological studies. The large number of novel alleles identified at the *DRB1* and *DQA2* alleles in both breeds is suggestive of the unique demographic factors shaping the MHC architecture in these two breeds in comparison to sheep from other parts of the world. This MHC evidence illustrates that the Djallonke and Sahelian sheep breeds are important genetic resources that can contribute to maintaining global sheep diversity.

## **Chapter 5 Genetic diversity of MHC class II region in Djallonke and Sahelian sheep populations**

*This chapter presents a comparative population genetic analysis of 100 Djallonke and 100 Sahelian sheep from Ghana. Allelic frequencies, heterozygosity, linkage disequilibrium, genic differentiation, and departure from Hardy Weinberg equilibrium were determined in the two breeds. To understand the demographic factors shaping the MHC class II region in these two sheep populations, the Nei-Gojobori codon-based selection test was conducted for all three class II loci. This is the first study to provide information using a population genetic approach for the analysis of the MHC class II region in the Djallonke and Sahelian sheep. The results offer some valuable insights as to the type and extent to which demographic factors may have shaped these two sheep populations.*

### **5.1 Introduction**

The Djallonke has been shown to be more resistant to many livestock diseases, including haemonchosis and trypanosomosis, than the Sahelian sheep described in chapter 1. These trypanotolerance and helminth resistance attributes of the Djallonke sheep have enabled the predominantly small-scale farmers to successfully rear, and continue to derive their sustenance from, these animals without recourse to the use of trypanocides and anthelmintics. Although other studies have demonstrated the differences in disease resistance for the two sheep breeds as described in chapter, the genetic basis for the different phenotypes has not been reported so far.

As described in chapter 1, resistance to both nematode infections and trypanosomiasis have been reported to be heritable in livestock (Hanotte et al., 2003; Trail et al., 1991). Furthermore, effective response to trypanosomiasis has been linked to both adaptive and innate components of the immune system of the host (Namangala, 2012). Therefore, the genetic region of the host that is most likely contributing to the acquired immunity is the MHC. In sheep, resistance to infections, including nematodes, has been associated with variation within genes in the MHC class II region (Hickford et al., 2011; Paterson, Wilson, & Pemberton, 1998; Sayers et al., 2005; Schwaiger et al., 1995). The genes in this region encode glycoproteins that present antigens to circulating CD4<sup>+</sup> T cells, and the most polymorphic ones are reported to be *DRB1*, *DQA1* and

*DQA2* (Ballingall & Tassi, 2010). Therefore, contrasting molecular variation at these class II MHC genes is expected for the Djallonke and Sahelian breeds.

As the first investigation of the genetic basis for the difference in resistance between these two sheep, this study focused on sequencing-based genotyping and analysis of the hypervariable MHC class II- *DRB1*, *DQA1* and *DQA2* loci in a population of 200 sheep from the Djallonke and Sahelian breeds from Ghana. The information on the genetic diversity at these loci for the two populations will facilitate disease association studies, and the efficient management, in terms of sustainable improvement programs for the two breeds.

## 5.2 Methods

Blood was collected from 100 Djallonke and 100 Sahelian sheep, and genomic DNA was extracted and genotyped at three loci (*DRB1*, *DQA2* & *DQA1*) as described in chapter 4. Sequence data was analysed for all three loci as described in chapter 4. Population genetic analysis for the identified alleles in the two sheep populations was done using GENEPOP statistical software (Rousset, 2008). Multi-locus frequency statistics for allele and genotype frequencies were computed for the functional exon 2 CDS in the three MHC class II loci in the two sheep populations using the Expectation Maximisation (EM) algorithm in GENEPOP statistical software v4.2 (Rousset, 2008). The EM algorithm is used to compute the maximum likelihood frequency statistics for the two populations (Dempster, Laird, & Rubin, 1977; Kalinowski & Taper, 2006). The average coefficient of inbreeding at each locus for the two populations was estimated as in Weir and Cockerham (1984) as implemented in GENEPOP. The Hardy-Weinberg exact test for the genotypic equilibrium for each of the loci in the two populations was computed with Fisher's probability in GENEPOP. The expected and observed numbers of heterozygotes at each locus for each sheep population was estimated using the Levene's correction method, and Chi square analysis was used to test for significant differences ( $p < 0.05$ ) between observed and expected heterozygosity at all three loci for the two populations. The Markov chain parameters used for all the tests for estimation of the exact probability values (P-values) for all the population genetic indices; include 1000 dememorization, 100 batches and 1000 iterations per batch. Fisher's exact probability test and the exact-G test were used to compute genic differentiation ( $F_{st}$ ) for all the loci genotyped between the Djallonke and Sahelian populations as in Weir and Cockerham (1984). Genotypic

linkage disequilibrium was computed for all three loci in the two populations using the log likelihood ratio statistic (G-test) as implemented in GENEPOP.

### 5.2.1 Test for Signatures of Selection

The two populations were tested for selection signatures at each of the three MHC class II loci using the Codon-based Test of Selection in MEGA (Nei and Gojobori, 1986; Kumar et al., 2016). This method is based on the ratio of non-synonymous (dN) to synonymous (dS) substitution rates at the CDS for each locus. Each locus was tested for three alternative hypotheses;  $dN \neq dS$ ,  $dN < dS$  and  $dN > dS$  with significance level of 5% ( $p < 0.05$ ). The Nei-Gojobori method (Nei and Gojobori 1986) was used to compute the variance of the differences between the dN and dS.

## 5.3 Results

As previously reported in chapter 4, 26 *DRB1*, 13 *DQA2* and 13 *DQA1* alleles were identified in the Djallonke and Sahelian populations (Tables 4.5, 4.6, & 4.7). Samples that did not amplify at a locus were considered as null alleles. The EM maximum likelihood estimated allele frequencies for the *DRB1*, *DQA2* and *DQA1* loci in the two sheep populations are presented in Tables 5.1, 5.2 & 5.3. The two sheep populations showed remarkable variability in frequency of alleles at all three loci genotyped (Figures 5.1, 5.2, & 5.3). The computed allele frequencies obtained for the *DQA1* and *DQA2* loci showed a higher proportion of null alleles for the Sahelian than for the Djallonke sheep. Conversely, the Djallonke had higher *DRB1* null allele frequency than the Sahelian.

At the *DRB1* locus, the *ovar-DRB1\*0901* allele is the most frequently amplified allele (16.67%) in the Djallonke population. Four other alleles (*\*1303* variant-1, *\*0501*, EU176819.1, & *BoLA-DRB3*-variant 1) recorded the lowest frequencies (0.51%) in the Djallonke population. These four *DRB1* alleles are also present in the Sahelian population, but at higher frequencies. The *DRB1\*1303* allele had the highest frequency (19.59%) for the Sahelian population, and eight other alleles recorded the lowest frequency of 0.67% (Table 5.1). Five of these eight *DRB1* alleles were also present in the Djallonke, but at marginally higher frequencies (Table 5.1). Four out of the five *DRB1* alleles were present at only 1.52%, whereas the remaining one (*DRB1\*0401*) was present at 4.55% frequency. Interestingly, while the *DRB1\*0901* allele (the most frequent allele in the Djallonke population) was not detected in the Sahelian sheep

population, a high frequency (12.71%) of *DRBI*\*1303 allele (the most frequent in the Sahelian population) was also found in the Djallonke sheep population (13.64%). The null allele frequency at the *DRBI* locus in the Djallonke population was 18.18%, whereas the Sahelian population recorded 12% null allele frequency.

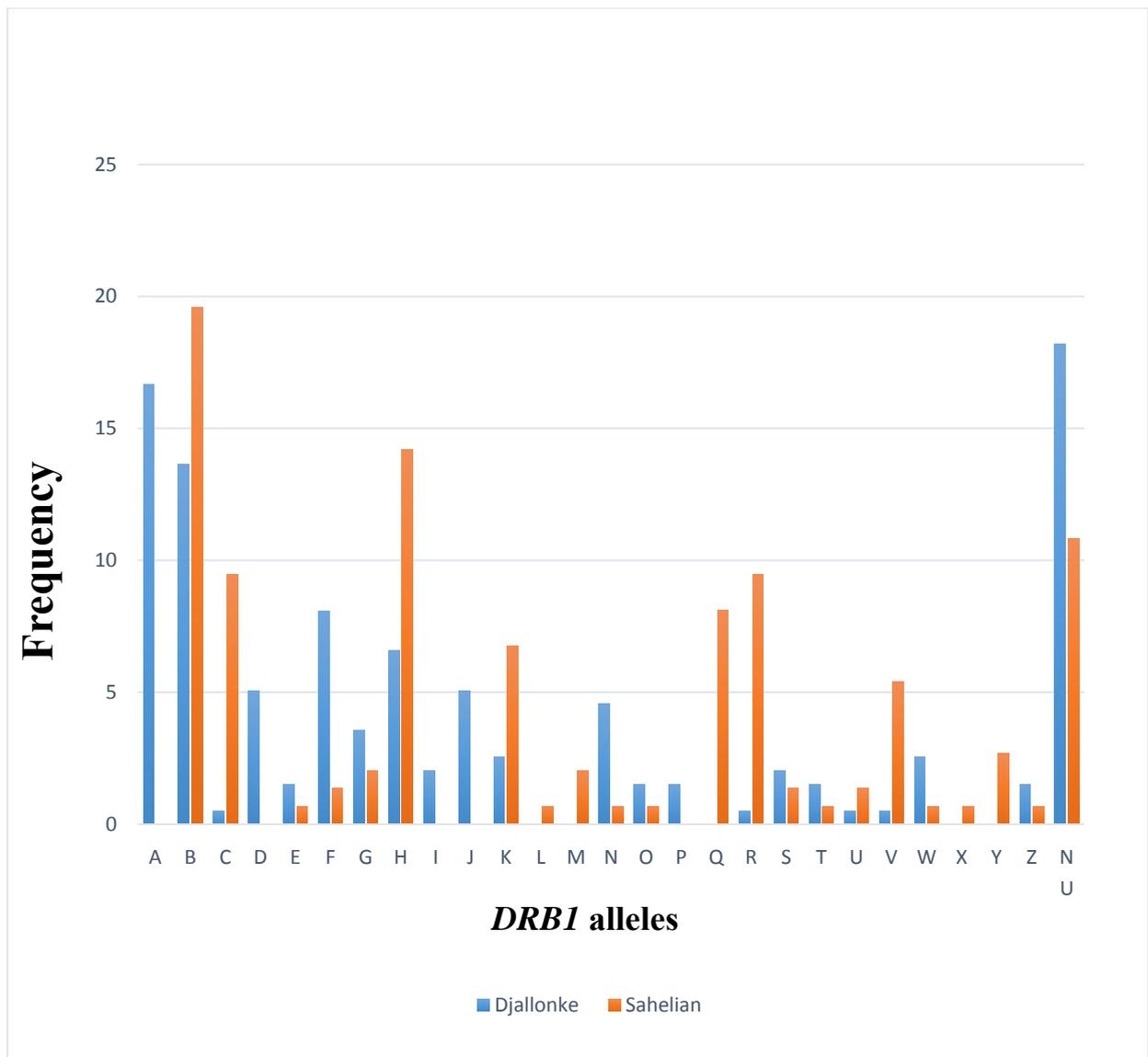
For the *DQA2* locus (Table 5.2), the EM maximum likelihood analysis showed that the *OLA-DQA2* alleles \*0901 and \*0103 had the highest (12.63%) and lowest (0.51%) frequencies respectively, in the Djallonke sheep population. The least frequent allele (\*0103) in the Djallonke population was absent in the Sahelian population. The \*1401 allele is the most frequent (13.33%) in the Sahelian population, and two other *DQA2* alleles (\*1201 & \*1701) had the lowest frequency of 0.67% each. Both of these *DQA2* alleles show contrasting frequencies for the two populations, and at higher frequencies in the Djallonke population (Table 5.2). The frequency of *OLA-DQA2*\*1201 and *OLA-DQA2*\*1701 in the Djallonke population was 8.08 and 9.09% respectively. However, a very high (73.33%) frequency estimate for *DQA2* null alleles was obtained for the Sahelian population compared to the 12.12% for the Djallonke population.

In the Djallonke population, the *DQA1* locus (Table 5.3) M33304.1 allele was the most frequently observed allele (23.23%), whereas the LN827890.1 was the most frequent within the Sahelian population (18%). With a frequency estimate of 0.51%, the LN827892.1 allele was the least frequent *DQA1* allele in the Djallonke, whereas two other alleles (LN827894.1 & LN8278913.1 variant 1) were the least frequent alleles for the Sahelian population (Table 5.3). Two *DQA1* alleles (M33304 variant 2 & LN827895.1 variant 1) in the Djallonke population were not detected in the Sahelian population, and similarly, two other alleles (LN827894.1 & LN827893.1-2) that were present in the Sahelian population were not detected in the Djallonke population. The frequency of null alleles at the *DQA1* locus was 24.24% and 42.67% for the Djallonke and Sahelian populations, respectively.

**Table 5.1 Maximum likelihood estimates of allele frequencies of *DRB1* in 100 Djallonke and 100 Sahelian sheep from Ghana**

<i>DRB1</i> Allele		Allele Frequency (%)	
Official Name/ Assigned ID	Allele code	Djallonke	Sahelian
<i>DRB1*0901</i>	A	16.67	0
<b>*0901 variant 1</b>	Z	1.52	0.67
<i>DRB1*1303</i>	B	13.64	19.59
<b>*1303 variant 1</b>	C	0.51	9.33
<b>*1303 variant 2</b>	D	5.05	0
<i>DRB1*0302</i>	E	1.52	0.67
<b>*0302 variant 1</b>	F	8.08	1.33
<b>*0302 variant 2</b>	G	3.54	2
<b>*0302 variant 3</b>	H	6.57	14
<b>*0302 variant 4</b>	I	2.02	0
<b>*0302 variant 5</b>	J	5.05	0
<b>*0302 variant 6</b>	K	2.53	6.67
<b>*0302 variant 7</b>	L	0	0.67
<b>*0302 variant 8</b>	M	0	2
<i>DRB1*0401</i>	N	4.55	0.67
<b>*0401 variant 1</b>	O	1.52	0.67
<b>*0401 variant 2</b>	P	1.52	0
<b>*0401 variant 3</b>	Q	0	8
<i>DRB1*0501</i>	R	0.51	9.33
<b>*0601 variant 1</b>	S	2.02	1.33

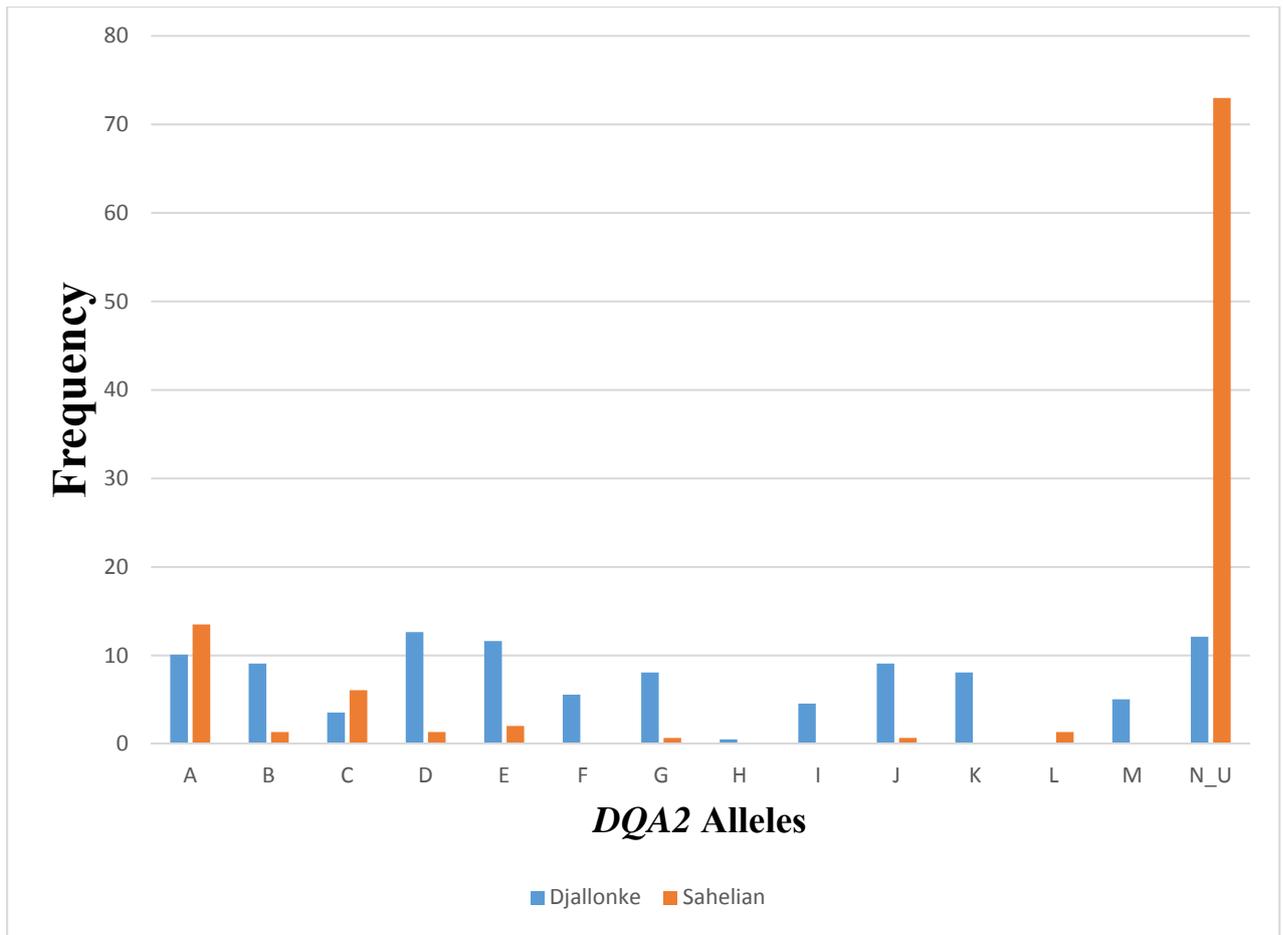
<i>DRBI</i> Allele		Allele Frequency (%)	
<i>DRBI*1402</i>	T	1.52	0.67
<b>BOLA-DRB3 Variant 1</b>	U	0.51	1.33
<b>EU176819.1 <i>DRBI</i></b>	V	0.51	5.33
<b>EU176819.1 variant 1</b>	W	2.53	0.67
<b>EU176819.1 variant 2</b>	X	0	0.67
<b>EU176819.1 variant 3</b>	Y	0	2.67
<b>Null alleles</b>	N_U	18.18	12



**Figure 5.1 Comparison of the *Ovar-DRB1* Allele frequency in Djallonke and Sahelian sheep populations**

**Table 5.2 Maximum likelihood Estimates of Allele frequencies of *DQA2* in 100 Djallonke and 100 Sahelian sheep from Ghana**

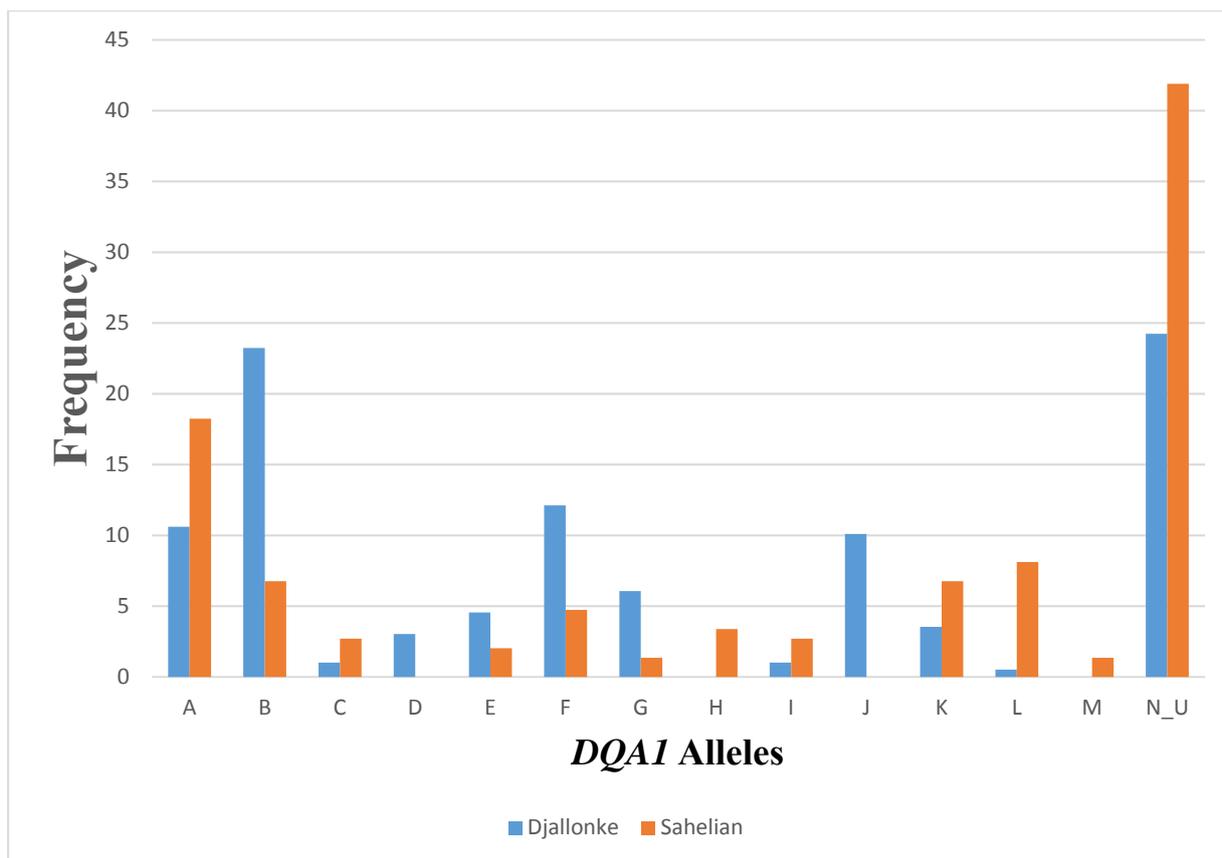
<i>OLA-DQA2</i> Alleles		Allele frequency (%)	
Accession number	Allele code	Djallonke	Sahelian
<i>OLA-DQA2*1401</i>	A	10.1	13.33
<i>OLA-DQA2*0601</i>	B	9.09	1.33
<i>OLA-DQA2*08011</i>	C	3.54	6
<i>OLA-DQA2*0901</i>	D	12.63	1.33
<i>OLA-DQA2*0501</i>	E	11.62	2
<i>OLA-DQA2*1001</i>	F	5.56	0
<i>OLA-DQA2*1201</i>	G	8.08	0.67
<i>OLA-DQA2*0103</i>	H	0.51	0
EU176819.1 <i>DQA2</i>	I	4.55	0
<i>OLA-DQA2*1701</i>	J	9.09	0.67
<i>OLA-DQA2*0701</i>	K	8.08	0
<i>OLA-DQA2*1501</i>	L	0	1.33
<i>OLA-DQA2*08012</i>	M	5.05	0
Null alleles	N_U	12.12	73.33



**Figure 5.2 Comparison of the *OLA-DQA2* Allele frequency in Djallonke and Sahelian sheep populations**

**Table 5.3 Maximum likelihood Estimates of Allele frequencies of *DQA1* in 100 Djallonke and 100 Sahelian sheep from Ghana**

<b>DQA1 Alleles</b>			<b>Allele frequencies (%)</b>	
<b>Accession ID</b>	<b>Number/Assigned</b>	<b>Allele code</b>	<b>Djallonke</b>	<b>Sahelian</b>
LN827890.1		A	10.6	18
M33304.1		B	23.23	6.67
M33304.1 variant 1		C	1.01	2.67
M33304.1 variant 2		D	3.03	0
LN827891.1		E	4.55	2
LN827893.1		F	12.12	4.67
LN827893.1 variant 1		G	6.06	1.33
LN827893.1 variant 2		H	0	3.33
LN827895.1		I	1.01	2.67
LN827895.1 variant 1		J	10.10	0
LN736359.1		K	3.54	6.67
LN827892.1		L	0.51	8
LN827894.1		M	0	1.33
Null alleles		N_U	24.24	42.67



**Figure 5.3 Comparison of the *DQA1* Allele frequency in Djallonke and Sahelian sheep populations**

### **5.3.1 Heterozygosity, Inbreeding, Hardy Weinberg Equilibrium and Linkage disequilibrium in the Djallonke and Sahelian populations**

The Levene's correction estimates for the detected alleles showed that the observed number of heterozygotes was lower than the expected number of heterozygotes at all three loci for both the Djallonke and Sahelian sheep populations (Table 5.4). However, Chi square probability was only significant for the *DQA2* locus ( $p < 0.05$ ) in the two populations.

**Table 5.4 The Levene's correction estimates for number of observed and expected heterozygotes in the Djallonke and Sahelian populations**

Locus	Djallonke Population		Sahelian Population		P-value
	Expected	Observed	Expected	Observed	
<i>DRB1</i>	89.35	69	66.71	64	0.379
<i>DQA2</i>	90.26	66	33.11	10	0.023*
<i>DQA1</i>	83.88	55	57.36	30	0.442

\* denotes significant Chi square test at  $p < 0.05$

**Table 5.5 Weir & Cockerham estimates of population average coefficient of Inbreeding (Fis) at all loci for Djallonke and Sahelian sheep**

Locus	Djallonke Population	P-value	Sahelian Population	P-value
<i>DRB1</i>	0.2287	<0.001**	0.0534	<0.001**
<i>DQA2</i>	0.2698	<0.001**	0.7003	<0.001**
<i>DQA1</i>	0.3454	<0.001**	0.4822	<0.001**

\*\* denoted significant Fisher's statistics test at  $p < 0.001$

The Weir & Cockerham estimated inbreeding coefficients (Fis) were very high for all MHC loci genotyped, with the *DQA2* loci in the Sahelian population showing the highest estimation (Table 5.5). Analysis of these estimates using the Fishers method showed highly significant Fis ( $p < 0.001$ ) at all three loci for both the Djallonke and Sahelian populations. Further analysis using Fisher's method detected highly significant deviations from Hardy-Weinberg equilibrium at all three loci for both sheep populations ( $p < 0.001$ ). Interestingly, pairwise analysis at each locus between the two sheep populations using the Fisher's exact probability test showed a highly significant ( $p < 0.01$ ) genic differentiation. Similarly, a highly significant genotypic differentiation was also obtained using the exact-G test method for the two sheep populations ( $p < 0.01$ ).

Pairwise genotypic LD estimates as per Markov parameters for the three loci showed highly significant levels of LD between the *DRB1* and *DQA2* loci across the two sheep populations ( $p < 0.01$ ). Significant population LD was observed between the *DRB1* and the

*DQA2* loci ( $p < 0.05$ ), but not between the *DQA2* and *DQA1* loci in the two sheep breeds ( $p > 0.05$ ). However, per population analysis showed that genotypic LD between the *DRB1* and *DQA2* loci in the Sahelian population had a higher level of significance ( $p < 0.001$ ) than the corresponding LD ( $p < 0.01$ ) in the Djallonke population.

### **5.3.2 Selection Signature at the MHC class II in Djallonke and Sahelian sheep populations**

At the *DRB1* locus, significant positive selection signatures ( $p < 0.05$ ) were detected at four and five different *DRB1* alleles in the Djallonke and Sahelian populations, respectively (Table 5.6). Three of the *DRB1* alleles appear to be under positive selection (\*0401, \*0501, \*0302) and these were common for the two populations. No significant signatures of purifying or neutral selection ( $p > 0.05$ ) were detected at the *DRB1* locus in either population. Similarly, five and four *DQA1* alleles in the Djallonke and Sahelian populations, respectively, show significant signatures of positive selection ( $p < 0.05$ ). However, only one of the *DQA1* alleles under positive selection (LN827895.1) occurred in both populations. Conversely, at the *DQA2* locus, significant purifying selection was detected for both populations, with five *DQA2* alleles within the Djallonke and only two *DQA2* alleles within the Sahelian populations being under purifying selection. One of the *DQA2* alleles (\*08011) appears to be under purifying selection within both populations. No significant positive or neutral selection was detected at the *DQA2* locus in either population.

**Table 5.6 Class II MHC alleles under selection in Djallonke and Sahelian populations**

Population Locus	Nei-Gojobori codon based Z-test for selection (p < 0.05)	
	Djallonke	Sahelian
<i>DRB1</i>	<i>*0401</i>	<i>*0401</i>
	<i>*0501</i>	<i>*0501</i>
	<i>*0302</i>	<i>*0302</i>
	<i>BoLA-DRB3 variant</i>	<i>*0601</i>
		<i>*1303</i>
<i>DQA2</i>	<b>*08011</b>	<b>*08011</b>
	<b>*08012</b>	<b>*1701</b>
	<b>*1201</b>	
	<b>EU176819.1_</b>	
<i>DQA1</i>	<i>LN827890.1</i>	<i>LN736359.1</i>
	<i>LN827891.1</i>	<i>LN827892.1</i>
	<i>LN827893.1</i>	<i>LN827894.1</i>
	<i>LN827895.1</i>	<i>LN827895.1</i>
	<i>M33304.1</i>	

Alleles in bold text showed significant purifying selection  
Alleles in italics showed significant positive selection

## 5.4 Discussion

The EM maximum likelihood estimates for allele frequency for the two sheep populations revealed the remarkable level of contrasting frequencies at the three class II MHC loci. This observation agrees with the general hypothesis of this study, given the documented differences in the phenotypic characteristics of the two sheep breeds (Geerts et al., 2009; Goossens et al., 1999; Mwai et al., 2015; Okaiyeto et al., 2010; Traoré et al., 2017). The total number of detected class II MHC alleles of 26, 13, and 13 at the *DRB1*, *DQA2*, and *DQA1* loci, respectively, in this study are higher than what was detected in recent study of 235 of Texel sheep (18 *DRB1*, 8 *DQA2*, 7 *DQA1*), using the same primers (Ali et al., 2017). However, while

the Texel population studied was known to contain some related individuals, there is no such information for the Djallonke or Sahelian populations in this study.

At The *DRB1* locus, which is the principally transcribed sheep MHC locus, the numbers of detected alleles per breed (21 *DRB1* alleles each) are comparable to that obtained (18 *DRB1* alleles) in a population of 64 Scottish blackface sheep (Ballingall & Tassi, 2010). There was a higher frequency of *DQA1* null alleles detected for the Sahelian population compared with the Djallonke population at all three loci. In an earlier comparative epigenetic study of nematode resistant and susceptible sheep breeds in New Zealand, higher levels of *DQA1* null alleles were found in the nematode susceptible sheep (Keane et al., 2007). Although, the study did not conclude that the null alleles were the direct cause of the susceptibility, it intimated that the null alleles were most probably in linkage with the causal MHC alleles for the nematode susceptibility.

The consistently lower than expected levels of observed heterozygotes across all three loci in both the Djallonke and Sahelian sheep populations could be attributable to a range of factors. Notably, the high frequencies of null alleles at all loci across the two populations (Figure 5.1 to 5.3), significant levels of inbreeding ( $p < 0.01$ ), and/or pathogen mediated selection in the two sheep breeds. Pinpointing the exact dynamics of the contribution of each of these factors to the observed levels of heterozygote deficiency in these two populations is not possible without full information on the specific demographic histories. Particularly as pedigree records are not available for these sampled populations, and this appears to be a common problem for many indigenous African livestock breeds (Asamoah-Boaheng & Sam, 2016).

It is important to note that the genepop software allows a maximum likelihood estimation of null allele frequency (Brookfield 1996; Kalinowski & Taper 2006). This is because consistent non- amplification of a sample or “apparent null” could either be due to a true null genotype or due to a range of PCR technical failures caused by factors such as the presence of SNPs at one or both priming sites. In addition to these polymorphisms, the MHC diversity is further enhanced by gene duplications, conversions and copy number variation (Ballingall et al., 2011; Ujvari & Belov, 2011; Yang et al., 2005). In the specific case for the ungulate MHC class II genes, codon deletions have often been reported (Cai et al., 2015; Mikko, Lewin, & Andersson, 1997). For example, deleted codons of the *BoLA-DRB3* (Mikko et al., 1997) and the *Mobe-DRB* (Cai et al., 2015) genes were reported in cattle and Musk deer respectively. The extraordinary complexity of the genomic organisation often complicates the MHC genotyping

process with one set of marker. In any case, an estimate of the “apparent null” homozygous frequency in this study is informative and of statistical relevance, even if not all are actually true null genotypes (Kalinowski & Taper 2006). Given that the MHC primers used in this study were designed in different sheep breeds with different demographic histories (Ali et al., 2017; Ballingall & Tassi, 2010; Hickford et al., 2004), ascertainment bias is likely to have contributed to the frequency of null homozygotes observed (Table 5.1, 5.2 & 5.3). Particularly, when there is no prior information on the MHC architecture of these two indigenous African sheep. Therefore, in order avoid skewing of the results, the computations for the other population genetic parameters (inbreeding coefficients, departure from Hardy Weinberg, linkage disequilibrium and population heterozygosity) in this study were performed by excluding null homozygotes from the dataset. However, the presence of null alleles in the heterozygote state could still inflate some of the computed population genetic parameters to some extent. A panel of specifically designed MHC primers Djallonke and Sahelian sheep might help to clarify these issues.

A significant contribution of inbreeding to the low heterozygosities is plausible for these populations, given the documented history of indiscriminate breeding of indigenous African livestock (Alvarez, Traore, et al., 2012; Berthier et al., 2016; Geerts et al., 2009; Kosgey & Okeyo, 2007). Consequences of such panmictic populations include; the reduction of effective population size, high inbreeding coefficients and genetic drift (Barker et al. 2001). In an earlier study in cattle breeds, Wahlund effects have been associated with the occurrence of heterozygote deficiencies (Loftus et al., 1999). Although no evidence of sub population structure was identified for these two sheep populations at the point of sampling, a contribution of Wahlund effect to this observation cannot be completely ruled out because of poor record keeping on farm.

It could be hypothesised that the lower observed heterozygosity for the Djallonke population at the *DRB1* loci and higher for both the *DQA2* and *DQAI* in comparison with the Sahelian sheep could also be ascribed to a direct response to pathogen-mediated selection. As the two sheep breeds are known for their difference in resistance to disease, some differences are expected in their respective immuno-genetics dynamics, which may be evident by the presence of breed specific variants within the MHC genes. This is corroborated by the presence of significant levels of signatures of positive selection, particularly in the breed specific MHC class II alleles, but also in common alleles found in these two populations (Table 5.6).

For classical MHC genes, such high variations at the coding regions are thought to facilitate recognition of a wider range of pathogens (Meyer-Lucht & Sommer, 2005). Furthermore, it has been reported that the presence of positive selection signatures at immune genes are generally a result of host parasite co-evolution (Hurst, 2009). The predominant disease causing pathogens in the host's environment (*Haemonchus contortus* and several species of Trypanosomes) are possibly crucial in shaping the genomic variation observed at the MHC through direct pathogen driven selection mechanisms (Hedrick, 1998, 2001). Two mechanisms reported for pathogen driven selection at the MHC are the rare allele advantage (Takahata & Nei, 1990) and the heterozygous advantage (Doherty & Zinkernagel, 1975). In this study, high levels of new alleles (38% at *DQAI* & 73% at *DRBI*) were found within the two populations and many of these alleles have not been identified in any other sheep population studies. In the largely extensive systems of animal production found within developing countries, including Ghana (Asamoah-Boaheng & Sam, 2016), accumulated favourable rare alleles or mutations within farm animals are more likely to have been favourably selected due to the intense environmental challenges the animals face. Furthermore, it is probable that most of these rare variants would be expected to be associated with beneficial rather than detrimental host phenotypes (Andersson, 2001). This is because, in these types of low input animal production systems, individuals with accumulated deleterious variants are likely to be eliminated naturally over time due to the environmental selective forces they face. Therefore, this probably supports a rare allele advantage MHC mechanism for the differences observed in the *DRBI* and *DQAI* loci at the class II MHC of Djallonke and Sahelian sheep populations. Conversely, for the *DQA2* locus, a combination of the absence of novel alleles and the significant level of purifying selection detected is suggestive of a heterozygote allele advantage driven selection mechanism.

Significant population genotypic differentiation ( $p < 0.01$ ) at all three loci in both sheep breeds, coupled with the significant deviation from Hardy Weinberg equilibrium ( $p < 0.01$ ), suggests a contrasting adaptivity at the coding MHC class IIa within these two sheep populations. This is evident by the presence of 19 breed-specific alleles at three MHC class II loci (10 at *DRBI*, 5 at *DQA2* & 4 at *DQAI*) between the two populations, with some at significantly high frequencies (Table 5.1). The genomic architecture of the MHC class II dictates how a host recognises and responds to disease challenge, and consequently, determines the host's resistance or susceptibility (Ali et al., 2016). These observed differences at the MHC class II loci, for these two populations, were hypothesised given the evidence that the two

breeds differ in disease resistance (Geerts et al., 2009; Osaer et al., 1994). Evidence of generalised resistance to trypanosomiasis and haemonchosis in Djallonke sheep, and the opposite in Sahelian sheep (more susceptibility), has been reported in many studies (Goossens et al., 1997; Goossens et al., 1999; Okaiyeto et al., 2010; Traoré et al., 2017).

While the high LD observed between the three loci in both sheep populations will facilitate future disease association investigations, it also poses a challenge in distinguishing which mutations are actual disease-causing candidates from those which are linked non-causal variants or hitch-hikers. Mapping of complex traits in domestic animals is reported to be particularly problematic when some of the linked mutations have neutral or even opposite effects (Goddard & Hayes, 2009). In domestic animals LD is generally known to extend over wider genomic distance (Goddard & Hayes, 2009), but in comparison, the LD in sheep is reported to be less extensive over short genomic distance (Kijas et al., 2014). The observed higher LD between *DRB1* and *DQA2* in the Sahelian population compared with the Djallonke population is consistent with the expectation that the greater the genetic diversity of a population, the lower the LD within that population (Kijas et al., 2014). This deduction agrees with the higher estimates of genetic diversity of obtained at for the Djallonke population over the Sahelian populations at these MHC class II loci via Poisson's correction algorithm, as described in chapter 4.

While the identified differences in allelic diversity in the two breeds are important for the obvious reasons discussed above, the observed similarities are also of key importance for complete understanding of the complexity of the genetic basis of the differences in resistance to diseases in the two sheep breeds. A total of 33 MHC class II alleles (16 at *DRB1*, 8 at *DQA2* & 9 at *DQA1*) are common between the two sheep populations, with some at similar frequencies (Tables 5.1 to 5.3). These observations could also be ascribed to several potential factors, ranging from admixture to ongoing co-evolution of the MHC class II regions of the two breeds because of shared recent demographic history. Again, it is difficult to pinpoint the contribution of all the genetic factors to this observation without accurate pedigree records of the two populations. However, elucidation of the basis of these substantial variant similarities will provide complementary information that would enhance our understanding of the genetic basis of the differences in disease resistance in the two breeds.

In recent years, a shifting focus from the theory that genetic variation (rare allele mode or common allele mode) is considered causal for a given phenotype, to the overall molecular

mechanism of interplay of many genes that leads to the expressed phenotypes has been suggested (Gibson, 2012). This broader approach encapsulates not only the genetics, but also possible epigenetic mechanisms as an important aspect of a comprehensive elucidation of the occurrence a phenotype, particularly, for complex diseases such as trypanosomiasis and haemonchosis.

The understanding of the overall underlying molecular mechanisms for these breed phenotypes will offer new opportunities for breed improvement strategies such as the application of novel marker assisted selective breeding for the desirable traits in both the Djallonke and the Sahelian sheep. This knowledge will contribute to the sustainable management of these two important genetic resources. Furthermore, identification of genes responsible for conferring resistance (or susceptibility) has applications in other non African breeds.

## **5.5 Conclusions**

The genetic analysis of the class IIa region of the MHC in Djallonke and Sahelian sheep showed high levels of allelic similarities between the two breeds, but also notable differences. Evidence of a rare allele advantage selection mechanism was identified as the basis of the observed phenotypic differences between the two breed populations at the *DRB1* and *DQA2* loci, and of heterozygote advantage mechanism at the *DQAI* locus. This is the first report on the genomic analysis of the MHC in these sheep breeds, and the observed differences support the hypothesis of the existence of molecular evidence for the differences in disease resistance in these two sheep breeds. These differences will provide important resources and new opportunity for disease association studies between the two sheep breeds.

## Chapter 6    General Discussion

The main aim of this study: to characterise the Djallonke sheep breed of Ghana using molecular makers, was accomplished, and additionally the Sahelian sheep was similarly characterised. The genetic information obtained from these two indigenous Ghanaian and African sheep breeds with contrasting phenotypic traits facilitated a wide range of comparative analyses to be performed.

At the whole genome level, described in chapter 2, the two sheep breeds showed substantial similarities in both the numbers and types of variants observed (approximately, 96% of SNPs & indels were in common). However, more importantly, significant differences were also present, particularly with respect to potential breed specific SNP alleles occurring between the Djallonke and Sahelian sheep breeds. These findings are consistent with the general hypothesis of this study; “NGS sequencing of a small representative number of unrelated individuals from a target population can permit the identification of informative breed specific SNPs” is well supported by these findings. The high numbers of similarities observed are also consistent with past reports describing the practice of indiscriminate breeding and genetic introgression between the two sheep breeds (Geerts et al., 2009; Goossens et al., 1999), and of post domestication admixture in these two sheep (Brahi et al., 2015). These potential breed specific variants will provide important targets for the development of breed specific markers that in the future can be applied to assist in the sustainable management of these two sheep breeds. The main advantage of any breed specific SNPs for designing a breed identification panel for the Djallonke and Sahelian breeds is that they are unlikely to suffer from ascertainment bias as might be the case with the use of commercially available SNP panels such as the Illumina Ovine 50K SNP BeadChip (Albrechtsen et al., 2010).

The whole genome analysis of the Djallonke and Sahelian genomes against previously published variants at the Ensembl dbSNP (release 85) also revealed many new variants (approximately, two million in each breed). This finding supports the concept that the Djallonke and Sahelian sheep breeds are a valuable and unique genetic resource and these breeds have many genetic differences that can contribute to global sheep diversity.

## **6.1 Molecular adaptive signatures of Selection in Djallonke and Sahelian Sheep**

Given that the contrasting adaptive phenotypic traits of the Djallonke and Sahelian sheep arose as a consequence of a millennia of natural selection (Brahi et al., 2015; Murray & Black, 1985; Mwai et al., 2015), a key hypothesis of this study was that there should be differential molecular selection signatures within the respective genomes that possibly correspond to these observed differences in the adaptive characteristics of the two sheep breeds (chapter 2.1). To this end, selection signature analysis using a multiplexed-genome sequencing strategy was proposed as a method of identifying selection sweeps (chapter 2). The hallmark of selection signature analysis is that it can be implemented for elucidating evidence of past selection, even without prior information on the pedigree of the target population (Qanbari & Simianer, 2014). Any molecular evidence identified can then be mapped back to the phenotypic traits in the target population. Therefore, the strategy of selection signature analysis of genomes from a small representative group of unrelated trypanotolerant Djallonke and trypanosusceptible Sahelian sheep for the identification of trypanotolerance associated loci was performed (see chapter 2.1). The analyses of the whole genome variants in the Djallonke and Sahelian sheep did reveal genomic regions of extended high homozygosity (or reduced heterozygosity), also referred to as selective sweeps, that were co-localised with adaptive genes, including previously identified QTL regions shown to be associated with trypanotolerance and resistance to Haemonchosis (see chapter 3). Therefore, these findings supported the hypothesis for this study. Although, the molecular determinants of adaptation to various ecological niches and of resistance to gastrointestinal parasites have been reported in other sheep breeds (Kijas et al., 2012; Roffler et al., 2016), this study provides the first report of the molecular support to putative regions harbouring possible QTL for trypanotolerance in any sheep breed. This finding is significant and provides the impetus needed for more focus and support for research programs aimed at sustainable management of these two sheep breeds.

## 6.2 Molecular characterisation of Ovar-MHC class II region reveals novel alleles in the Djallonke and Sahelian sheep breeds of Ghana

The contrasting resistance against parasite infections between the Djallonke and Sahelian sheep breeds has been demonstrated via many parasitological and haematological studies (Goossens et al., 1998; Goossens et al., 1999; Traoré et al., 2017). However, molecular genetic analyses for the determinism of these differential responses has not previously been reported. Resistance to parasite infection in sheep has been predominantly associated with variations within genes in the MHC class II region (Gelasakis et al., 2013; Hickford et al., 2011; Keane et al., 2007). These class II genes encode glycoproteins that present antigens to circulating CD4<sup>+</sup> T cells, and the most polymorphic ones are reported to be the *DRB1*, *DQA1* and *DQA2* (Ballingall et al., 2008; Dukkupati et al., 2006; Herrmann-Hoesing et al., 2008).

Therefore, this study conducted a comparative analysis of the MHC class II region in a population of 100 Djallonke and 100 Sahelian sheep from two sheep breeding stations in Ghana. The allelic variations within the three most polymorphic MHC class II genes (*DRB1*, *DQA1* & *DQA2*) were analysed in the two breeds using sequencing-based genotyping. The classification of sequences, according to the Immuno-polymorphism-MHC nomenclature (Maccari et al., 2017), revealed a high number of novel alleles at the *DRB1* (73%) and the *DQA1* (38%) loci, but none at *DQA2* locus in the two populations. Of the three loci, the *DRB1* locus showed the highest amount of variation, an observation that supports earlier reports that the *DRB1* locus is the most diverse (Ballingall et al., 2011; Charon, 2004). The Djallonke breed showed higher diversity at all the three loci. Higher diversity at the MHC region in a target population connotes higher resistance to infection and population fitness (Aguilar et al., 2004). The higher diversity observed at the class II region of the Djallonke sheep as compared with the Sahelian sheep is consistent with the higher resistance to infection reported for the Djallonke breed. A total of 18 breed specific alleles were identified between the two sheep populations. These breed specific alleles present prime targets for studies to elucidate the causal variants for resistance and susceptibility to the parasitic infections in these two sheep breeds.

Interestingly, the *ovar-DRB1\*1303* allele (GenBank id FR751085.1) that was present in both the Djallonke and Sahelian sheep populations, was previously identified in the red Maasai sheep breed from Kenya (unpublished). This is suggestive of an ancestral relationship between these three indigenous African sheep breeds. However, an earlier report had indicated different

ancestry for the Red Maasai sheep breed (Muigai & Hanotte, 2013). Therefore, the presence of *ovar-DRB1\*1303* allele in these three sheep breeds is probably due to similar adaptive responses to common demographic selection factors within the SSA region in these three breeds.

A potential application of breed-specific MHC alleles is for molecular identification of pure breeds of each of these two sheep populations. A panel of SNP markers based on the highly informative polymorphisms within the breed specific alleles could be used for distinguishing pure individuals of Djallonke and Sahelian sheep from mixed populations. This is particularly important in the context of the mainly extensive systems of sheep production in SSA. Typical features of these extensive systems are that sheep of different breeds flock together on communal grazing field, and breeding is largely uncontrolled (Asamoah-Boaheng & Sam, 2016). These types of management systems have contributed to the high levels of admixture between the Djallonke and Sahelian, making it a significant threat to each of these two genetic resources (Alvarez et al., 2009). Indeed, evidence of introgression-mediated genetic dilution of the unique adaptive traits in other indigenous African ruminants has recently been reported (Mwai et al., 2015), indicating an even more widespread problem. However, the use of breed-specific markers generated from non-neutral loci such as the MHC class IIa alleles are prone to loss of efficiency over time, due to the dynamics of selection, or hitchhiking at such loci. This shortcoming will potentially undermine the long-term use of such markers for conservation programmes. Nonetheless, such markers will be an ideal predictor of the evolutionary potential of a target population (fitness or loss of it, over time). The measure of the evolutionary potential of a target population provides probably the best indicator of the resilience potential or the extinction risk for that population.

Identification of livestock breeds at the molecular level will provide an objective and reliable option for these two sheep breeds. The existing phenotypic methods of identification, as described by Birteeb et al. (2012), are inadequate markers of genetic diversity, but could complement the molecular methods. This is because breed identification at the phenotypic level does not always correspond to the identification at the molecular level (Felius et al., 2014). Reliable identification of livestock breeds will facilitate sustainable livestock utilisation and conservation programmes in SSA.

This work is the first analysis of MHC architecture and allelic diversity in either of these breeds of sheep. This information also offers a new opportunity for the application of molecular

tools for sustainable management aimed at maintaining MHC diversity in these two indigenous African sheep breeds.

### **6.3 Genetic diversity of MHC class II region in Djallonke and Sahelian populations**

A comparative assessment of the levels of diversity in the two sheep populations was conducted primarily to understand the richness of the gene pool and the demographic factors that are shaping the genomic architecture of the MHC class II region in the two sheep populations. To this end, population genetic parameters including; the frequency of alleles and genotypes, departure from Hardy Weinberg equilibrium, coefficient of inbreeding, heterozygosity as well as signatures of selection were analysed for the Djallonke and Sahelian sheep populations. Substantial differences in allelic frequency were identified at all the loci analysed in the two population. Strikingly, the *DRB1\*0901* allele, which had the highest frequency (16.67%) in the Djallonke populations, was completely absent in the Sahelian population. Conversely, the most frequent allele (*ovar-DRB1\*1303*) in the Sahelian population (19.59%), was present at moderately high frequency (13.64%) in the Djallonke. The high frequency of this common allele in both Djallonke and Sahelian breed, which is also shared with the red Maasai suggests an important adaptive role to the indigenous sheep breeds in SSA region. A higher frequency of null alleles was found at *DRB1* for the Djallonke sheep than the Sahelian sheep, but they had a lower null allele frequency for the other two (*DQA2* & *DQA1*) loci. Interestingly, a high frequency of null alleles at the *DQA1* locus has been associated with susceptibility to nematode infection in sheep (Keane et al., 2007). The observed higher frequency of *DQA1* null allele in the Sahelian sheep in this study supports this earlier report. The test of heterozygosity revealed reduced heterozygosity at all the three MHC class II loci, but was only statistically significant ( $p < 0.05$ ) at the *DQA2* locus in the two populations. Furthermore, the fact that significant deviation from Hardy Weinberg equilibrium coupled with significant inbreeding were detected in all three loci for both sheep populations is suggestive of management challenges at the breeding stations.

To understand the types and extent to which demographic factors have influenced the genomic architecture of the MHC class II regions in both sheep populations, the codon-based selection test was conducted for all the three loci in each of the two populations. Significant signatures of positive selection were detected at the *DRB1* and *DQA1* loci, and significant purifying selection was detected at the *DQA2* locus for the two populations. A total of 13 alleles

were under significant selection in the Djallonke population as compared to 11 in the Sahelian sheep population. Interestingly, four of these alleles were under same direction of selection in the two sheep populations, suggestive of a similar response to a common demographic selection pressure.

Evidence of a rare allele advantage mode of balancing selection mechanism was identified at the *DRB1* and *DQAI* loci, and of heterozygote advantage mode of balancing selection mechanism for the *DQAI* locus. This observation represents the first genomic analysis of the MHC in these sheep breeds, and support the hypothesis of the existence of molecular evidence for the contrasting phenotypes in these two sheep breeds. Collectively, these results will provide valuable resources and offer new opportunity for disease mechanism studies for the two sheep breeds.

Overall, this research has identified, for the first time, whole genome variant characteristics of Djallonke and Sahelian sheep from Ghana, and provided a comprehensive comparative analysis of these variant types between the two breeds. This information has increase our understanding of the unique diversity of these two sheep breeds in relation to global sheep diversity. Furthermore, the findings also provide, for the first time, an observation of the genetic basis for phenotypes, in particular resistance to parasitic infection, that are of huge economic importance to the SSA region. This is important not only for the research community and livestock farmers, but also for policy makers in all the host countries for these two sheep breeds.

## **6.4 Future directions**

Immediate future study should focus on constructing a panel of breed-specific markers for the Djallonke and Sahelian breeds, to enable the analysis of different populations of Djallonke and Sahelian sheep from Ghana. Such a panel could be derived from highly informative breed-specific SNPs or indels identified from the considerable number of breed-specific variants found between these two sheep breeds (Accession number PRJEB15642 at EMBL European variant archive & GenBank accession for identified MHC class II alleles see appendix E table 1 for complete list). As a precaution, it is important to select informative variants that are in neutral genomic regions or that are not under any form of adaptive selection. This will guard against the loss of efficiency of the marker panel over a few generations of selection. Variants that are fixed in a given breed have been recommended for the construction of a more robust

breed-specific panel (as described in chapter 2). When developed, a breed specific panel could provide reliable breed identification tool for the accurate assessment of the conservation status of these sheep breeds.

In the past panels of breed-specific SNPs derived from lower density markers such as the ovine SNP50BeadChip (Dodd et al 2014) and bovine 50K SNP BeadChip (Suekana et al 2010; Bertolini et al 2015; Wilkinson et al 2011) have been used for assignment of individuals to respective breeds with some limited success due to ascertainment bias. However, the methodology used in the current study was designed to minimise this drawback for the Djallonke and Sahelian breed-specific markers. Given the general lack of biotechnology infrastructure within SSA, a panel of breed-specific markers that can be used to resolve breed differences using PCR band size (e.g. indels) rather than at the level of sequence (e.g. SNPs) will be more practical. Sustainable management of these sheep breeds is dependent on the ability to reliably and simply differentiate between pure Djallonke and Sahelian animals, as well as composite animals.

Given that the Djallonke and Sahelian sheep are present in at least 14 countries (transboundary sheep breeds per the FAO classification), coupled with the predominantly extensive production systems of management of sheep within the SSA region, future molecular characterisation should be conducted on populations from different countries. Comparisons of multiple genomes from different populations will provide a bigger picture of the overall diversity of these sheep breeds across the region, will increase understanding of the population structure of the breeds, and will inform a regional approach to the management of these important genetic resources. Host countries could then pool resources to set up a regional centre for Djallonke and Sahelian sheep improvement. Such a centre could form collaborations with already existing global livestock management organisations like the ILRI, EFABIS and the DAGRIS to help deliver its mandate.

Equally important for future studies is to investigate further the genomic evidence for resistance to gastrointestinal parasites and trypanosomiasis in larger numbers of both Djallonke and Sahelian sheep. Although, the samples used in the current studies were from infected areas for these diseases, phenotypic data relating to infection and response to infection was not available. The evidence obtained about possible QTL for tolerance/ resistance in this current work, which corresponds with similar regions in other livestock, strongly suggest that they are real, and therefore future studies should include comparison of infected and uninfected animals to

strengthen the data. In addition, future work should integrate targeted sequencing and epigenetics of the candidate regions to facilitate the identification of more refined QTL (pinpointing of causal variants) for these traits. Identification of more refined QTL for these traits will facilitate the use of genomic applications such as marker assisted breeding for rapid genetic gain. Given the absolute reliance of the predominantly smallholder farmers within the SSA region on the disease resistance traits of the Djallonke breed, carrying out these objectives has the potential not only for sustainable management of these two sheep breeds, but also for sustainable livelihoods for the farmers. It is therefore, clear that the scientific and socio-economic benefits expected to be derived will be worth whatever the investments needed for this undertaking.

## References

- Adrian, V. S. H., Simon, N. R. Y., Catherine, E. M. A., Gupta, S., Gilbert, S. C., Lalvani, A., . . . Plebanski, M. (1994). Human Leukocyte Antigens and Natural Selection by Malaria. *Philosophical Transactions: Biological Sciences*, 346(1317), 379-385. Retrieved from <http://www.jstor.org/stable/55971>
- Agaviezor, B. O., Peters, S. O., Adefenwa, M. A., Yakubu, A., Ajayi, O. O., & Amusan, S. A. (2012). Morphological and microsatellite DNA diversity of Nigerian indigenous sheep. *Journal of Animal Science and Biotechnology*, 3(1), 38.
- Aguilar, A., Roemer, G., Debenham, S., Binns, M., Garcelon, D., & Wayne, R. K. (2004). High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proceedings of the National Academy of Sciences of the United States of America*, 101(10), 3490-3494. <http://dx.doi.org/10.1073/pnas.0306582101>
- Ai, H., Fang, X., Yang, B., Huang, Z., Chen, H., Mao, L., . . . Huang, L. (2015). Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat Genet*, 47(3), 217-225. <http://dx.doi.org/10.1038/ng.3199>
- <http://www.nature.com/ng/journal/v47/n3/abs/ng.3199.html#supplementary-information>
- Ajmone-Marsan, P., & The Globaldiv Consortium. (2010). A global view of livestock biodiversity and conservation – GLOBALDIV. *Animal Genetics*, 41, 1-5. <http://dx.doi.org/10.1111/j.1365-2052.2010.02036.x>
- Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., . . . Searle, S. M. J. (2016). The Ensembl gene annotation system. *Database*, 2016, baw093-baw093. <http://dx.doi.org/10.1093/database/baw093>
- Al-Atiyat, R. M., Salameh, N. M., & Tabbaa, M. J. (2014). Analysis of genetic diversity and differentiation of sheep populations in Jordan. *Electronic Journal of Biotechnology*, 17(4), 168-173. <http://dx.doi.org/http://dx.doi.org/10.1016/j.ejbt.2014.04.002>
- Al-Mamun, H. A., A Clark, S., Kwan, P., & Gondro, C. (2015). Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. *Genetics Selection Evolution*, 47(1), 90. <http://dx.doi.org/10.1186/s12711-015-0169-6>
- Albrechtsen, A., Nielsen, F. C., & Nielsen, R. (2010). Ascertainment Biases in SNP Chips Affect Measures of Population Divergence. *Molecular Biology and Evolution*, 27(11), 2534-2547. <http://dx.doi.org/10.1093/molbev/msq148>
- Ali, A. O., Stear, A., Fairlie-Clarke, K., Brujeni, G. N., Isa, N. M., Salisi, M. S., . . . Stear, M. J. (2017). The genetic architecture of the MHC class II region in British Texel sheep. *Immunogenetics*, 69(3), 157-163. <http://dx.doi.org/10.1007/s00251-016-0962-6>
- Ali, B. A. (2003). Genetics similarity among four breeds of sheep in Egypt detected by random amplified polymorphic DNA markers. *African Journal of Biotechnology*, 2(7), 194-197.
- Ali, B. A., Huang, T.-H., Qin, D.-N., & Wang, X.-M. (2004). A review of random amplified polymorphic DNA (RAPD) markers in fish research. *Reviews in Fish Biology and Fisheries*, 14(4), 443-453.
- Allendorf, F. W. (1986). Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology*, 5(2), 181-190. <http://dx.doi.org/10.1002/zoo.1430050212>
- Alsford, S., Kelly, J. M., Baker, N., & Horn, D. (2013). Genetic dissection of drug resistance in trypanosomes. *Parasitology*, 140(Special Issue 12), 1478-1491. <http://dx.doi.org/doi:10.1017/S003118201300022X>
- Alvarez, I., Capote, J., Traoré, A., Fonseca, N., Pérez, K., Cuervo, M., . . . Goyache, F. (2012). Genetic relationships of the Cuban hair sheep inferred from microsatellite polymorphism. *Small Ruminant Research*, 104(1-3), 89-93. <http://dx.doi.org/http://dx.doi.org/10.1016/j.smallrumres.2011.10.025>

- Alvarez, I., Traore, A., Kabore, A., Zare, Y., Fernandez, I., Tamboura, H. H., & Goyache, F. (2012). Microsatellite analysis of the Rousse (Red Sokoto) goat of Burkina Faso. *Small Ruminant Res*, *105*( 1-3), 83–88. <http://dx.doi.org/10.1016/j.smallrumres.2012.01.012>
- Alvarez, I., Traoré, A., Tamboura, H. H., Kabore, A., Royo, L. J., Fernández, I., . . . Goyache, F. (2009). Microsatellite analysis characterizes Burkina Faso as a genetic contact zone between Sahelian and Djallonké sheep. *Animal Biotechnology*, *20*(2), 47-57.
- Amills, M., Ramiya, V., Norimine, J., & Lewin, H. A. (1998). The major histocompatibility complex of ruminants. *Revue scientifique et technique (International Office of Epizootics)*, *17*(1), 108-120. Retrieved from <http://europemc.org/abstract/MED/9638804>
- Anderson, S. (2003). Animal genetic resources and sustainable livelihoods. *Ecological Economics*, *45*(3), 331-339. [http://dx.doi.org/10.1016/S0921-8009\(03\)00088-0](http://dx.doi.org/10.1016/S0921-8009(03)00088-0)
- Anderson, S., & Centonze, R. (2007). Property rights and the management of animal genetic resources. *World Development*, *35*(9), 1529-1541. <http://dx.doi.org/http://dx.doi.org/10.1016/j.worlddev.2006.05.020>
- Andersson, L. (2001). Genetic dissection of phenotypic diversity in farm animals. *Nat Rev Genet*, *2*(2), 130-138. Retrieved from <http://dx.doi.org/10.1038/35052563>
- Andersson, L., & Georges, M. (2004). Domestic-animal genomics: deciphering the genetics of complex traits. *Nature Reviews Genetics*, *5*(3), 202-212.
- Andersson, L., & Mikko, S. (1995). Generation of MHC Class II Diversity by Intra- and Intergenic Recombination. *Immunological Reviews*, *143*(1), 5-12. <http://dx.doi.org/10.1111/j.1600-065X.1995.tb00667.x>
- Andrabi, S. M., & Maxwell, W. M. (2007). A review on reproductive biotechnologies for conservation of endangered mammalian species. *Animal Reproduction Science*, *99*(3-4), 223-243. <http://dx.doi.org/10.1016/j.anireprosci.2006.07.002>
- Asamoah-Boaheng, M., & Sam, E. K. (2016). Morphological characterization of breeds of sheep: a discriminant analysis approach. *SpringerPlus*, *5*, 69. <http://dx.doi.org/10.1186/s40064-016-1669-8>
- Ashley, M., & Dow, B. (1994). The use of microsatellite analysis in population biology: background, methods and potential applications. In B. Schierwater, Streit, B., Wagner, G.P., Desalle, R. (Eds.) (Ed.), *Molecular Ecology and Evolution: Approaches and Applications* (pp. 185-201). Birkhäuser Basel, Switzerland: Springer.
- Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., . . . Saunders, N. C. (1987). Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, *18*, 489-522. <http://dx.doi.org/10.2307/2097141>
- Avise, J. C., & Ellis, D. (1986). Mitochondrial DNA and the evolutionary genetics of higher animals [and discussion]. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *312*(1154), 325-342.
- Babikir, O., Muchina, S., Sebsibe, A., Bika, A., Kwai, A., Agosa, C., . . . Wakhusama, S. (2015). Agricultural Systems in IGAD Region — A Socio-Economic Review. In P. V. Pilipavicius (Ed.), *Agroecology* (pp. 56): Intech Open Science. <http://dx.doi.org/DOI: 10.5772/60012>
- Baker, R. (1995). Genetics of disease resistance in small ruminants in Africa. *ACIAR Monograph Serie*, *34*(34), 119-138.
- Ballingall, K. T., Fardoe, K., & McKeever, D. J. (2008). Genomic organisation and allelic diversity within coding and non-coding regions of the Ovar-DRB1 locus. *Immunogenetics*, *60*(2), 95-103. <http://dx.doi.org/10.1007/s00251-008-0278-2>
- Ballingall, K. T., Herrmann-Hoesing, L., Robinson, J., Marsh, S. G. E., & Stear, M. J. (2011). A single nomenclature and associated database for alleles at the major histocompatibility complex class II DRB1 locus of sheep. *Tissue Antigens*, *77*(6), 546-553. <http://dx.doi.org/10.1111/j.1399-0039.2011.01637.x>
- Ballingall, K. T., Rocchi, M. S., McKeever, D. J., & Wright, F. (2010). Trans-Species Polymorphism and Selection in the MHC Class II DRA Genes of Domestic Sheep. *PLoS ONE*, *5* <http://dx.doi.org/10.1371/journal.pone.0011402>

- Ballingall, K. T., Steele, P., Lantier, I., Cotelli, M., Todd, H., Lopez, G., . . . Lantier, F. (2015). An ancient interlocus recombination increases class II MHC DQA diversity in sheep and other Bovidae. *Animal Genetics*, 46(3), 333-336. <http://dx.doi.org/10.1111/age.12290>
- Ballingall, K. T., & Tassi, R. (2010). Sequence-based genotyping of the sheep MHC class II DRB1 locus. *Immunogenetics*, 62(1), 31-39. <http://dx.doi.org/10.1007/s00251-009-0410-y>
- Banik, S., Pankaj, P., & Naskar, S. (2015). Climate change: Impacts on livestock diversity in tropical countries. *Livestock Production and Climate Change*, 6, 162.
- Barbato, M., Orozco-terWengel, P. A., Tapio, M., & Bruford, M. W. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics*, 6, 109. <http://dx.doi.org/10.3389/fgene.2015.00109>
- Barendse, W., Vaiman, D., Kemp, S., Sugimoto, Y., Armitage, S., Williams, J., . . . Aleyasin, S. (1997). A medium-density genetic linkage map of the bovine genome. *Mammalian Genome*, 8(1), 21-28.
- Barrett, J. C. (1992). *The economic role of cattle in communal farming systems in Zimbabwe*. England: Overseas Development Institute, London.
- Baumung, R., Cubric-Curik, V., Schwend, K., Achmann, R., & Sölkner, J. (2006). Genetic characterisation and breed assignment in Austrian sheep breeds using microsatellite marker information. *Journal of Animal Breeding and Genetics*, 123(4), 265-271. <http://dx.doi.org/10.1111/j.1439-0388.2006.00583.x>
- Bayrakli, F., Poyrazoglu, H. G., Yuksel, S., Yakicier, C., Erguner, B., Sagiroglu, M. S., . . . Ziyal, I. (2015). Hereditary spastic paraplegia with recessive trait caused by mutation in KLC4 gene. *J Hum Genet*, 60(12), 763-768. <http://dx.doi.org/10.1038/jhg.2015.109>
- Beckmann, J., & Soller, M. (1983). Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. *Theoretical and Applied Genetics*, 67(1), 35-43.
- Beckmann, J., & Soller, M. (1986). Restriction fragment length polymorphisms and genetic improvement of agricultural species. *Euphytica*, 35(1), 111-124.
- Benavides, M. V., Sonstegard, T. S., Kemp, S., Mugambi, J. M., Gibson, J. P., Baker, R. L., . . . Van Tassell, C. (2015). Identification of Novel Loci Associated with Gastrointestinal Parasite Resistance in a Red Maasai x Dorper Backcross Population. *PLoS ONE*, 10(4), e0122797. <http://dx.doi.org/10.1371/journal.pone.0122797>
- Benavides, M. V., Sonstegard, T. S., & Van Tassell, C. (2016). Genomic Regions Associated with Sheep Resistance to Gastrointestinal Nematodes. *Trends in Parasitology*, 32(6), 470-480. <http://dx.doi.org/http://dx.doi.org/10.1016/j.pt.2016.03.007>
- Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., & Petrov, D. A. (2014). Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in *Drosophila*. *PLoS Genet*, 10(11), e1004775. <http://dx.doi.org/10.1371/journal.pgen.1004775>
- Berthier, D., Breniere, S. F., Bras-Goncalves, R., Lemesre, J. L., Jamonneau, V., Solano, P., . . . Bucheton, B. (2016). Tolerance to Trypanosomatids: A Threat, or a Key for Disease Elimination? *Trends Parasitol*, 32(2), 157-168. <http://dx.doi.org/10.1016/j.pt.2015.11.001>
- Berthier, D., Quere, R., Thevenon, S., Belemsaga, D., Piquemal, D., Marti, J., & Maillard, J. C. (2003). Serial analysis of gene expression (SAGE) in bovine trypanotolerance: preliminary results. *Genet Sel Evol*, 35 Suppl 1, S35-47. <http://dx.doi.org/10.1051/gse:2003015>
- Berthouly, C., Bed'Hom, B., Tixier-Boichard, M., Chen, C. F., Lee, Y. P., Laloë, D., . . . Rognon, X. (2008). Using molecular markers and multivariate methods to study the genetic diversity of local European and Asian chicken breeds. *Animal Genetics*, 39(2), 121-129. <http://dx.doi.org/10.1111/j.1365-2052.2008.01703.x>
- Bertolini, F., Galimberti, G., Calò, D. G., Schiavo, G., Matassino, D., & Fontanesi, L. (2015). Combined use of principal component analysis and random forests identify population-informative single nucleotide polymorphisms: application in cattle breeds. *Journal of Animal Breeding and Genetics*, 132(5), 346-356. <http://dx.doi.org/10.1111/jbg.12155>
- Bezie, M., Girma, M., Dagnachew, S., Tadesse, & D.Tadesse, G. (2014). African Trypanosomes: Virulence Factors, Pathogenicity And Host Responses. *J Vet Adv*, 4(11), 732-745. Retrieved from <http://www.scopemed.org/?mno=171563>

- Bianco, E., Nevado, B., Ramos-Onsins, S. E., & Pérez-Enciso, M. (2015). A Deep Catalog of Autosomal Single Nucleotide Variation in the Pig. *PLOS ONE*, *10*(3), e0118867. <http://dx.doi.org/10.1371/journal.pone.0118867>
- Birteeb, P. T., Peters, S. O., Yakubu, A., Adeleke, M. A., & Ozoje, M. O. (2012). Multivariate characterisation of the phenotypic traits of Djallonke and Sahel sheep in Northern Ghana. *Tropical animal health and production*, *45*(1), 267-274.
- Biscarini, F., Nicolazzi, E., Stella, A., Boettcher, P., & Gandini, G. (2015). Challenges and opportunities in genetic improvement of local livestock breeds. *Frontiers in Genetics*, *6*, 33. <http://dx.doi.org/10.3389/fgene.2015.00033>
- Bishop, S., & MacKenzie, K. (2003). Genetic management strategies for controlling infectious diseases in livestock populations. *Genet Select Evol*, *35*, S3-S18.
- Bishop, S. C. (2012). Possibilities to breed for resistance to nematode parasite infections in small ruminants in tropical production systems. *Animal*, *6*(5), 741-747. <http://dx.doi.org/10.1017/S1751731111000681>
- Black, S., Sendashonga, C., O'Brien, C., Borowy, N., Naessens, M., Webster, P., & Murray, M. (1985). Regulation of parasitaemia in mice infected with *Trypanosoma brucei* *The Biology of Trypanosomes* (pp. 93-118). Berlin Heidelberg: Springer.
- Black, S. J., & Seed, J. R. (Eds.). (2002). *The African Trypanosomes*. Kluwer academic Publishers, USA.
- Blears, M., De Grandis, S., Lee, H., & Trevors, J. (1998). Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. *Journal of Industrial Microbiology and Biotechnology*, *21*(3), 99-114.
- Boettcher, P. J., Tixier-Boichard, M., Toro, M. A., Simianer, H., Eding, H., Gandini, G., . . . The Globaldiv Consortium. (2010). Objectives, criteria and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. *Animal Genetics*, *41*, 64-77. <http://dx.doi.org/10.1111/j.1365-2052.2010.02050.x>
- Boitard, S., Boussaha, M., Capitan, A., Rocha, D., & Servin, B. (2016). Uncovering Adaptation from Sequence Data: Lessons from Genome Resequencing of Four Cattle Breeds. *Genetics*, *203*(1), 433. Retrieved from <http://www.genetics.org/content/203/1/433.abstract>
- Bollmer, J. L., Vargas, F. H., & Parker, P. G. (2007). Low MHC variation in the endangered Galápagos penguin (*Spheniscus mendiculus*). *Immunogenetics*, *59*(7), 593-602. <http://dx.doi.org/10.1007/s00251-007-0221-y>
- Bowles, D., Carson, A., & Isaac, P. (2014). Genetic distinctiveness of the herdwick sheep breed and two other locally adapted Hill Breeds of the UK. *PLoS ONE*, *9*(1), e87823. <http://dx.doi.org/http://dx.doi.org/10.1371/journal.pone.0087823>
- Bozzi, R., Degl'Innocenti, P., Rivera Diaz, P., Nardi, L., Crovetto, A., Sargentini, C., & Giorgetti, A. (2009). Genetic characterization and breed assignment in five Italian sheep breeds using microsatellite markers. *Small Ruminant Research*, *85*(1), 50-57. <http://dx.doi.org/http://dx.doi.org/10.1016/j.smallrumres.2009.07.005>
- Bradley, D. G., MacHugh, D. E., Loftus, R. T., Sow, R. S., Hoste, C. H., & Cunningham, E. P. (1994). Zebu-taurine variation in Y chromosomal DNA: a sensitive assay for genetic introgression in West African trypanotolerant cattle populations. *Anim Genet*, *25*(S2), 7-12. <http://dx.doi.org/10.1111/j.1365-2052.1994.tb00440.x>
- Brahi, O. H. D., Xiang, H., Chen, X., Farougou, S., & Zhao, X. (2015). Mitogenome revealed multiple postdomestication genetic mixtures of West African sheep. *Journal of Animal Breeding and Genetics*, *132*(5), 399-405. <http://dx.doi.org/10.1111/jbg.12144>
- Brinks, J., Clark, R., Kieffer, N., & Urick, J. (1964). Estimates of genetic, environmental and phenotypic parameters in range Hereford females. *Journal of Animal Science*, *23*(3), 711-716.
- Brown, E. A., Pilkington, J. G., Nussey, D. H., Watt, K. A., Hayward, A. D., Tucker, R., . . . Slate, J. (2013). Detecting genes for variation in parasite burden and immunological traits in a wild population: testing the candidate gene approach. *Molecular Ecology*, *22*(3), 757-773. <http://dx.doi.org/10.1111/j.1365-294X.2012.05757.x>
- Brown, J. L., & Eklund, A. (1994). Kin Recognition and the Major Histocompatibility Complex: An Integrative Review. *The American Naturalist*, *143*(3), 435-461. Retrieved from <http://www.jstor.org/stable/2462739>

- Broxham, E. T. (2015). A case study on strains of Buša cattle structured into a metapopulation to show the potential for use of single-nucleotide polymorphism genotyping in the management of small, cross-border populations of livestock breeds and varieties. *Frontiers in Genetics*, 6, 73. <http://dx.doi.org/10.3389/fgene.2015.00073>
- Bruce, D. (1915). The Croonian lectures on trypanosomes causing disease in man and domestic animals in Central Africa. *Br Med J*, 1(2843), 1073-1078.
- Bruford, M. W., Bradley, D. G., & Luikart, G. (2003). DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics*, 4(11), 900-910.
- Bruford, M. W., Ginja, C., Hoffmann, I., Joost, S., Orozco-terWengel, P., Alberto, F., . . . Zhan, X. (2015). Prospects and Challenges for the Conservation of Farm Animal Genomic Resources, 2015-2025. *Frontiers in Genetics*, 6 [http://www.frontiersin.org/Journal/Abstract.aspx?s=620&name=livestock\\_genomics&ART\\_DOI=10.3389/fgene.2015.00314](http://www.frontiersin.org/Journal/Abstract.aspx?s=620&name=livestock_genomics&ART_DOI=10.3389/fgene.2015.00314), <http://dx.doi.org/10.3389/fgene.2015.00314>
- Bruford, M. W., & Wayne, R. K. (1993). Microsatellites and their application to population genetic studies. *Current Opinion in Genetics & Development*, 3(6), 939-943. [http://dx.doi.org/http://dx.doi.org/10.1016/0959-437X\(93\)90017-J](http://dx.doi.org/http://dx.doi.org/10.1016/0959-437X(93)90017-J)
- Brun, R., Blum, J., Chappuis, F., & Burri, C. (2010). Human African trypanosomiasis. *Lancet*, 375(9709), 148-159. [http://dx.doi.org/10.1016/S0140-6736\(09\)60829-1](http://dx.doi.org/10.1016/S0140-6736(09)60829-1)
- Bruns, E., & Glodek, P. (1999). *Animal genetic biodiversity-methodological considerations*. Paper presented at the Deutscher Tropentag
- Bryja, J., Galan, M., Charbonnel, N., & Cosson, J. F. (2006). Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (Arvicolinae). *Immunogenetics*, 58(2), 191-202. <http://dx.doi.org/10.1007/s00251-006-0085-6>
- Buchanan, F. C., Adams, L. J., Littlejohn, R. P., Maddox, J. F., & Crawford, A. M. (1994). Determination of evolutionary relationships among sheep breeds using microsatellites. *Genomics*, 22(2), 397-403.
- Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide *Current Protocols in Molecular Biology*: John Wiley & Sons, Inc. <http://dx.doi.org/10.1002/0471142727.mb2129s109>
- Bult, C. J., Krupke, D. M., Begley, D. A., Richardson, J. E., Neuhauser, S. B., Sundberg, J. P., & Eppig, J. T. (2015). Mouse Tumor Biology (MTB): a database of mouse models for human cancer. *Nucleic Acids Research*, 43(D1), D818-D824. <http://dx.doi.org/10.1093/nar/gku987>
- Burt, D. W. (2009). The cattle genome reveals its secrets. *Journal of Biology*, 8(4), 36. <http://dx.doi.org/10.1186/jbiol137>
- Büscher, P., Cecchi, G., Jamonneau, V., & Priotto, G. (2017). Human African trypanosomiasis. *The Lancet* [doi.org/10.1016/S0140-6736\(17\)31510-6](http://dx.doi.org/https://doi.org/10.1016/S0140-6736(17)31510-6), [http://dx.doi.org/https://doi.org/10.1016/S0140-6736\(17\)31510-6](http://dx.doi.org/https://doi.org/10.1016/S0140-6736(17)31510-6)
- Cai, R., Shafer, A. B. A., Laguardia, A., Lin, Z., Liu, S., & Hu, D. (2015). Recombination and selection in the major histocompatibility complex of the endangered forest musk deer (*Moschus berezovskii*). *Scientific Reports*, 5, 17285. <http://dx.doi.org/10.1038/srep17285>
- Carignano, H. A., Beribe, M. J., Caffaro, M. E., Amadio, A., Nani, J. P., Gutierrez, G., . . . Poli, M. A. (2017). BOLA-DRB3 gene polymorphisms influence bovine leukaemia virus infection levels in Holstein and Holstein × Jersey crossbreed dairy cattle. *Animal Genetics*, n/a-n/a. <http://dx.doi.org/10.1111/age.12566>
- Casas, E., Shackelford, S., Keele, J., Stone, R., Kappes, S., & Koohmaraie, M. (2000). Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms of myostatin. *Journal of animal science*, 78(3), 560-569.
- Chakraborty, R., & Nei, M. (1977). Bottleneck Effects on Average Heterozygosity and Genetic Distance with the Stepwise Mutation Model. *Evolution*, 31(2), 347-356. <http://dx.doi.org/10.2307/2407757>
- Charbonnel, N., Bryja, J., Galan, M., Deter, J., Tollenaere, C., Chaval, Y., . . . Cosson, J.-F. (2010). Negative relationships between cellular immune response, Mhc class II heterozygosity and secondary sexual trait in the montane water vole. *Evolutionary Applications*, 3(3), 279-290. <http://dx.doi.org/10.1111/j.1752-4571.2009.00108.x>

- Charbonnel, N., & Pemberton, J. (2005). A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. *Heredity*, 95(5), 377-388. Retrieved from <http://dx.doi.org/10.1038/sj.hdy.6800735>
- Charon, K. M. (2004). Genes controlling resistance to gastrointestinal nematodes in ruminants. *Animal Science Papers and Reports*, 22(1), 135-139.
- Chen, A., Zhang, R., & Yu, S. (2008). Comparative results of survival of vitrified biopsied goat embryos and mouse morulae. *Turkish Journal of Veterinary and Animal Sciences*, 32(2), 93-97.
- Chenyambuga, S., Hanotte, O., Hirbo, J., Watts, P., Kemp, S., Kifaro, G., . . . Rege, J. (2004). Genetic characterization of indigenous goats of Sub-Saharan Africa using microsatellite DNA markers. *Australasian Journal of Animal Sciences*, 17, 445-452.
- Chesser, R. K., Smith, M. H., & Brisbin, I. L. (1980). Management and maintenance of genetic variability in endangered species. *International Zoo Yearbook*, 20(1), 146-154. <http://dx.doi.org/10.1111/j.1748-1090.1980.tb00958.x>
- Chistiakov, D. A., Hellemans, B., & Volckaert, F. A. M. (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. *Aquaculture*, 255(1-4), 1-29. <http://dx.doi.org/http://dx.doi.org/10.1016/j.aquaculture.2005.11.031>
- Chitneedi, P. K., Arranz, J. J., Suarez-Vega, A., García-Gómez, E., & Gutiérrez-Gil, B. (2017). Estimations of linkage disequilibrium, effective population size and ROH-based inbreeding coefficients in Spanish Churra sheep using imputed high-density SNP genotypes. *Animal Genetics*, 48(4), 436-446. <http://dx.doi.org/10.1111/age.12564>
- Choi, J.-W., Liao, X., Stothard, P., Chung, W.-H., Jeon, H.-J., Miller, S. P., . . . Lee, S.-J. (2014). Whole-Genome Analyses of Korean Native and Holstein Cattle Breeds by Massively Parallel Sequencing. *PLoS ONE*, 9(7), e101127. <http://dx.doi.org/10.1371/journal.pone.0101127>
- Choi, J. W., Choi, B. H., Lee, S. H., Lee, S. S., Kim, H. C., Yu, D., . . . Lim, D. (2015). Whole-Genome Resequencing Analysis of Hanwoo and Yanbian Cattle to Identify Genome-Wide SNPs and Signatures of Selection. *Mol Cells*, 38(5), 466-473. <http://dx.doi.org/10.14348/molcells.2015.0019>
- Ciani, E., Ciampolini, R., D'Andrea, M., Castellana, E., Cecchi, F., Incoronato, C., . . . Cianci, D. (2013). Analysis of genetic variability within and among Italian sheep breeds reveals population stratification and suggests the presence of a phylogeographic gradient. *Small Ruminant Research*, 112(1-3), 21-27. <http://dx.doi.org/http://dx.doi.org/10.1016/j.smallrumres.2012.12.013>
- Clarke, S. M., Henry, H. M., Dodds, K. G., Jowett, T. W. D., Manley, T. R., Anderson, R. M., & McEwan, J. C. (2014). A high throughput single nucleotide polymorphism multiplex assay for parentage assignment in New Zealand sheep. *PLoS ONE*, 9(4), e93392. <http://dx.doi.org/10.1371/journal.pone.0093392>
- Cnops, J., De Trez, C., Bulte, D., Radwanska, M., Ryffel, B., & Magez, S. (2015). IFN-gamma mediates early B-cell loss in experimental African trypanosomosis. *Parasite Immunol*, 37(9), 479-484. <http://dx.doi.org/10.1111/pim.12208>
- Cnops, J., Magez, S., & De Trez, C. (2015). Escape mechanisms of African trypanosomes: why trypanosomosis is keeping us awake. *Parasitology*, 142(3), 417-427. <http://dx.doi.org/10.1017/S0031182014001838>
- Coltman, D. W., Wilson, K., Pilkington, J. G., Stear, M. J., & Pemberton, J. M. (2001). A microsatellite polymorphism in the gamma interferon gene is associated with resistance to gastrointestinal nematodes in naturally-parasitised population of soay sheep. *Parasitol*, 122 <http://dx.doi.org/10.1017/s0031182001007570>
- Coop, R. L., & Holmes, P. H. (1996). Nutrition and parasite interaction. *Int J Parasitol*, 26(8-9), 951-962. [http://dx.doi.org/http://dx.doi.org/10.1016/S0020-7519\(96\)80070-1](http://dx.doi.org/http://dx.doi.org/10.1016/S0020-7519(96)80070-1)
- Coop, R. L., & Kyriazakis, I. (1999). Nutrition-parasite interaction. *Vet Parasitol*, 84(3-4), 187-204. [http://dx.doi.org/http://dx.doi.org/10.1016/S0304-4017\(99\)00070-9](http://dx.doi.org/http://dx.doi.org/10.1016/S0304-4017(99)00070-9)
- Coop, R. L., & Kyriazakis, I. (2001). Influence of host nutrition on the development and consequences of nematode parasitism in ruminants. *Trends Parasitol*, 17(7), 325-330. [http://dx.doi.org/http://dx.doi.org/10.1016/S1471-4922\(01\)01900-6](http://dx.doi.org/http://dx.doi.org/10.1016/S1471-4922(01)01900-6)

- Cooper, P. J. M., Dimes, J., Rao, K. P. C., Shapiro, B., Shiferaw, B., & Twomlow, S. (2008). Coping better with current climatic variability in the rain-fed farming systems of sub-Saharan Africa: An essential first step in adapting to future climate change? *Agriculture, Ecosystems & Environment*, 126(1-2), 24-35. <http://dx.doi.org/10.1016/j.agee.2008.01.007>
- Courtin, D., Berthier, D., Thevenon, S., Dayo, G. K., Garcia, A., & Bucheton, B. (2008). Host genetics in African trypanosomiasis. *Infect Genet Evol*, 8(3), 229-238. <http://dx.doi.org/10.1016/j.meegid.2008.02.007>
- Cox, F. E. (2004). History of sleeping sickness (African trypanosomiasis). *Infect Dis Clin North Am*, 18(2), 231-245. <http://dx.doi.org/10.1016/j.idc.2004.01.004>
- Crawford, A. M., Paterson, K. A., Dodds, K. G., Diez Tascon, C., Williamson, P. A., Roberts Thomson, M., . . . McEwan, J. C. (2006). Discovery of quantitative trait loci for resistance to parasitic nematode infection in sheep: I. Analysis of outcross pedigrees. *BMC Genomics*, 7(1), 178. <http://dx.doi.org/10.1186/1471-2164-7-178>
- Cunningham-Rundles, S., McNeeley, D. F., & Moon, A. (2005). Mechanisms of nutrient modulation of the immune response. *J Allergy Clin Immunol*, 115(6), 1119-1128; quiz 1129. <http://dx.doi.org/10.1016/j.jaci.2005.04.036>
- Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., . . . Flicek, P. (2015). Ensembl 2015. *Nucleic Acids Research*, 43(D1), D662-D669. <http://dx.doi.org/10.1093/nar/gku1010>
- Curković, M., Ramljak, J., Ivanković, S., Mioč, B., Ivanković, A., Pavić, V., . . . Medugorac, I. (2015). The genetic diversity and structure of 18 sheep breeds exposed to isolation and selection. *Journal of Animal Breeding and Genetics*, ISSN 1439-0388, <http://dx.doi.org/10.1111/jbg.12160>. <http://dx.doi.org/10.1111/jbg.12160>
- Dagnachew, S., & Bezie, M. (2015). Review on *Trypanosoma vivax*. *Afr J Basic App Sci*, 7(1), 41-64.
- Dagnachew, S., Bezie, M., Terefe, G., Abebe, G., Barry, J. D., & Goddeeris, B. M. (2015). Comparative clinico-haematological analysis in young Zebu cattle experimentally infected with *Trypanosoma vivax* isolates from tsetse infested and non-tsetse infested areas of Northwest Ethiopia. *Acta Vet Scand*, 57(1), 24. <http://dx.doi.org/10.1186/s13028-015-0114-2>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Group, G. P. A. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158. <http://dx.doi.org/10.1093/bioinformatics/btr330>
- Davies, G., Stear, M. J., Benothman, M., Abuagob, O., Kerr, A., Mitchell, S., & Bishop, S. C. (2006). Quantitative trait loci associated with parasitic infection in Scottish blackface sheep. *Heredity*, 96 <http://dx.doi.org/10.1038/sj.hdy.6800788>
- Dayo, G. K., Gautier, M., Berthier, D., Poivey, J. P., Sidibe, I., Bengaly, Z., . . . Thevenon, S. (2012). Association studies in QTL regions linked to bovine trypanotolerance in a West African crossbred population. *Anim Genet*, 43(2), 123-132. <http://dx.doi.org/10.1111/j.1365-2052.2011.02227.x>
- Dayo, G. K., Thevenon, S., Berthier, D., Moazami-Goudarzi, K., Denis, C., Cuny, G., . . . Gautier, M. (2009). Detection of selection signatures within candidate regions underlying trypanotolerance in outbred cattle populations. *Mol Ecol*, 18(8), 1801-1813. <http://dx.doi.org/10.1111/j.1365-294X.2009.04141.x>
- de Souza, J. M., Batista, R. I., Melo, L. M., & Freitas, V. J. d. F. (2011). Reproductive biotechnologies applied to the conservation of endangered ruminant-past, present and future *Revista Portuguesa de Ciências Veterinárias*, 106(577-580), 31-38.
- Deichmann, U., Meisner, C., Murray, S., & Wheeler, D. (2011). The economics of renewable energy expansion in rural Sub-Saharan Africa. *Energy Policy*, 39(1), 215-227. <http://dx.doi.org/http://dx.doi.org/10.1016/j.enpol.2010.09.034>
- Delespaux, V., & de Koning, H. P. (2007). Drugs and drug resistance in African trypanosomiasis. *Drug Resist Updat*, 10(1-2), 30-50. <http://dx.doi.org/10.1016/j.drug.2007.02.004>
- Delespaux, V., Geysen, D., Van den Bossche, P., & Geerts, S. (2008). Molecular tools for the rapid detection of drug resistance in animal trypanosomes. *Trends Parasitol*, 24(5), 236-242. <http://dx.doi.org/http://dx.doi.org/10.1016/j.pt.2008.02.006>
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1-38.

- Dib, C., Fauré, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., . . . Seboun, E. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature*, *380*(6570), 152-154.
- Dimauro, C., Nicoloso, L., Cellesi, M., Macciotta, N. P. P., Ciani, E., Moioli, B., . . . Crepaldi, P. (2015). Selection of discriminant SNP markers for breed and geographic assignment of Italian sheep. *Small Ruminant Research*, *128*, 27-33. <http://dx.doi.org/http://dx.doi.org/10.1016/j.smallrumres.2015.05.001>
- Dlugosch, K. M., & Parker, I. M. (2008). Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, *17*(1), 431-449. <http://dx.doi.org/10.1111/j.1365-294X.2007.03538.x>
- Dobney, K., & Larson, G. (2006). Genetics and animal domestication: new windows on an elusive process. *Journal of Zoology*, *269*(2), 261-271. <http://dx.doi.org/10.1111/j.1469-7998.2006.00042.x>
- Dodds, K. G., Auvray, B., Newman, S.-A. N., & McEwan, J. C. (2014). Genomic breed prediction in New Zealand sheep. *BMC Genetics*, *15*(1), 92.
- Dodds, K. G., McEwan, J. C., Brauning, R., Anderson, R. M., van Stijn, T. C., Kristjánsson, T., & Clarke, S. M. (2015). Construction of relatedness matrices using genotyping-by-sequencing data. *BMC Genomics*, *16*(1), 1047. <http://dx.doi.org/10.1186/s12864-015-2252-3>
- Dodgson, J. B., Cheng, H. H., & Okimoto, R. (1997). DNA marker technology: a revolution in animal genetics. *Poultry Science*, *76*(8), 1108-1114.
- Doherty, P. C., & Zinkernagel, R. M. (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature*, *256*(5512), 50-52. Retrieved from <http://dx.doi.org/10.1038/256050a0>
- Dolan, R. B. (1987). Genetics and trypanotolerance. *Parasitol Today*, *3*(5), 137-143. [http://dx.doi.org/http://dx.doi.org/10.1016/0169-4758\(87\)90197-9](http://dx.doi.org/http://dx.doi.org/10.1016/0169-4758(87)90197-9)
- Donelson, J. E. (2003). Antigenic variation and the African trypanosome genome. *Acta Tropica*, *85*(3), 391-404. [http://dx.doi.org/http://dx.doi.org/10.1016/S0001-706X\(02\)00237-1](http://dx.doi.org/http://dx.doi.org/10.1016/S0001-706X(02)00237-1)
- Doyle, J. J., Moloo, S. K., & Borowy, N. K. (1984). Development of Improved Control Methods of Animal Trypanosomiasis - a Review. *Prev Vet Med*, *2*(1-4), 43-52. [http://dx.doi.org/Doi10.1016/0167-5877\(84\)90047-3](http://dx.doi.org/Doi10.1016/0167-5877(84)90047-3)
- Dreszer, T. R., Karolchik, D., Zweig, A. S., Hinrichs, A. S., Raney, B. J., Kuhn, R. M., . . . Rosenbloom, K. R. (2012). The UCSC genome browser database: extensions and updates 2011. *Nucleic Acids Research*, *40*(D1), D918-D923.
- Drucker, A. G., Gomez, V., & Anderson, S. (2001). The economic valuation of farm animal genetic resources: a survey of available methods. *Ecological Economics*, *36*(1), 1-18.
- Dukkipati, V., Blair, H., Garrick, D., & Murray, A. (2006). 'Ovar-Mhc'- Ovine major histocompatibility complex: Role in genetic resistance to diseases. *NZ Vet J*, *54* <http://dx.doi.org/10.1080/00480169.2006.36689>
- Dzingirai, V., Bukachi, S., Leach, M., Mangwanya, L., Scoones, I., & Wilkinson, A. (2017). Structural drivers of vulnerability to zoonotic disease in Africa. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1725) Retrieved from <http://rstb.royalsocietypublishing.org/content/372/1725/20160169.abstract>
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792-1797. <http://dx.doi.org/10.1093/nar/gkh340>
- Eding, H., & Meuwissen, T. (2001). Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *Journal of Animal Breeding and Genetics*, *118*(3), 141-159.
- Edwards, S. V., Chesnut, K., Satta, Y., & Wakeland, E. K. (1997). Ancestral polymorphism of Mhc class II genes in mice: implications for balancing selection and the mammalian molecular clock. *Genetics*, *146*(2), 655-668.
- Edwards, S. V., & Hedrick, P. W. (1998). Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology & Evolution*, *13*(8), 305-311. [http://dx.doi.org/http://dx.doi.org/10.1016/S0169-5347\(98\)01416-5](http://dx.doi.org/http://dx.doi.org/10.1016/S0169-5347(98)01416-5)

- Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., & Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nature reviews. Genetics*, *11*(6), 446-450. <http://dx.doi.org/10.1038/nrg2809>
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*, *6*(5), e19379. <http://dx.doi.org/10.1371/journal.pone.0019379>
- Engel, S. R., Linn, R. A., Taylor, J. F., & Davis, S. K. (1996). Conservation of microsatellite loci across species of artiodactyls: implications for population studies. *Journal of Mammalogy*, *77*, 504-518.
- Ennen, S., Hamann, H., Distl, O., Hickford, J., Zhou, H., & Ganter, M. (2009). A field trial to control ovine footrot via vaccination and genetic markers. *Small Ruminant Research*, *86*(1-3), 22-25. <http://dx.doi.org/https://doi.org/10.1016/j.smallrumres.2009.09.011>
- Eppig, J. T., Blake, J. A., Bult, C. J., Kadin, J. A., Richardson, J. E., & Group, T. M. G. D. (2015). The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic Acids Research*, *43*(D1), D726-D736. <http://dx.doi.org/10.1093/nar/gku967>
- Escayg, A. P., Hickford, J. G. H., & Bullock, D. W. (1997). Association between alleles of the ovine major histocompatibility complex and resistance to footrot. *Research in Veterinary Science*, *63*(3), 283-287. [http://dx.doi.org/http://dx.doi.org/10.1016/S0034-5288\(97\)90035-7](http://dx.doi.org/http://dx.doi.org/10.1016/S0034-5288(97)90035-7)
- Fabuel, E., Barragan, C., Silio, L., Rodriguez, M. C., & Toro, M. A. (2004). Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers. *Heredity*, *93*(1), 104-113. Retrieved from <http://dx.doi.org/10.1038/sj.hdy.6800488>
- Fakae, B. B., Harrison, L. J., Ross, C. A., & Sewell, M. M. (1997). Expression of acquired immunity to *Heligmosomoides polygyrus* in mice concurrently infected with *Trypanosoma congolense*. *Int J Parasitol*, *27*(9), 1107-1114. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9363495>
- Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, *7*(4), 574-578.
- Fan, H., & Chu, J.-Y. (2007). A brief review of short tandem repeat mutation. *Genomics, Proteomics & Bioinformatics*, *5*(1), 7-14. [http://dx.doi.org/http://dx.doi.org/10.1016/S1672-0229\(07\)60009-6](http://dx.doi.org/http://dx.doi.org/10.1016/S1672-0229(07)60009-6)
- FAO. (2000). *Watch list of domestic animal diversity*. Rome, Italy: Retrieved from <http://www.fao.org/News/2000/001201-e.htm>
- FAO. (2007). *Global plan of action for animal genetic resources and the Interlaken Declaration: Adopted by the international technical conference on animal genetic resources for food and agriculture, Interlaken, Switzerland, 3-7 September 2007*. Rome, Italy: Commission On Genetic Resources for Food and Agriculture, Food and Agriculture Organization of the United Nations.
- FAO. (2013). *Status and trends of animal genetic resources 2012*. Rome, Italy Retrieved from <http://www.fao.org/docrep/meeting/027/mg046e.pdf>
- Felius, M., Theunissen, B., & Lenstra, J. A. (2014). Conservation of cattle genetic resources: the role of breeds. *The Journal of Agricultural Science*, *153*, 152-162. <http://dx.doi.org/doi:10.1017/S0021859614000124>
- Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*, *39*(4), 783-791. <http://dx.doi.org/10.2307/2408678>
- Ferguson, A., Taggart, J., Prodöhl, P., McMeel, O., Thompson, C., Stone, C., . . . Hynes, R. (1995). The application of molecular markers to the study and conservation of fish populations, with special reference to *Salmo*. *Journal of Fish Biology*, *47*(sA), 103-126.
- Fisher, P., Hedeler, C., Wolstencroft, K., Hulme, H., Noyes, H., Kemp, S., . . . Brass, A. (2007). A systematic strategy for large-scale analysis of genotype phenotype correlations: identification of candidate genes involved in African trypanosomiasis. *Nucleic Acids Res*, *35*(16), 5625-5633. <http://dx.doi.org/10.1093/nar/gkm623>
- Foll, M., Beaumont, M. A., & Gaggiotti, O. (2008). An approximate Bayesian computation approach to overcome biases that arise when using amplified fragment length polymorphism markers to study population structure. *Genetics*, *179*(2), 927-939.

- Ford, L. B. (2007). Civil conflict and sleeping sickness in Africa in general and Uganda in particular. *Conflict and health*, 1(6), 1-10.
- Francois, O., Martins, H., Caye, K., & Schoville, S. D. (2016). Controlling false discoveries in genome scans for selection. *Mol Ecol*, 25(2), 454-469. <http://dx.doi.org/10.1111/mec.13513>
- Frankham, R. (1995). Conservation genetics. *Annual Review of Genetics*, 29(1), 305-327.
- Fremont, D. H., Henderickson, W. A., Marrack, P., & Kappler, J. (1996). Structures of an MHC class II molecule with covalently bound single peptides. *Science*, 272(5264), 1001.
- Frisch, J. E., O'Neill, C. J., & Kelly, M. J. (2000). Using genetics to control cattle parasites - the Rockhampton experience. *International J Parasitol*, 30(3), 253-264. [http://dx.doi.org/Doi10.1016/S0020-7519\(00\)00010-2](http://dx.doi.org/Doi10.1016/S0020-7519(00)00010-2)
- Frkonja, A., Gredler, B., Schnyder, U., Curik, I., & Sölkner, J. (2012). Prediction of breed composition in an admixed cattle population. *Animal Genetics*, 43(6), 696-703. <http://dx.doi.org/10.1111/j.1365-2052.2012.02345.x>
- Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: advantages, limitations, and practical recommendations. *Molecular Ecology*, n/a-n/a. <http://dx.doi.org/10.1111/mec.14264>
- Futschik, A., & Schlötterer, C. (2010). The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics*, 186(1), 207-218. <http://dx.doi.org/10.1534/genetics.110.114397>
- Gao, J., Liu, K., Liu, H., Blair, H. T., Li, G., & Chen, C. (2010). A complete DNA sequence map of the ovine major histocompatibility complex. *BMC Genomics*, 11 <http://dx.doi.org/10.1186/1471-2164-11-466>
- García-Gámez, E., Sahana, G., Gutiérrez-Gil, B., & Arranz, J.-J. (2012). Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genetics*, 13(1), 43. <http://dx.doi.org/10.1186/1471-2156-13-43>
- Garnett, T. (2009). Livestock-related greenhouse gas emissions: impacts and options for policy makers. *Environmental Science & Policy*, 12(4), 491-503. <http://dx.doi.org/http://dx.doi.org/10.1016/j.envsci.2009.01.006>
- Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., . . . Lathrop, G. M. (2007). Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics*, 177(2), 1059-1070.
- Gautier, M., Flori, L., Riebler, A., Jaffrezic, F., Laloe, D., Gut, I., . . . Foulley, J. L. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics*, 10(1), 550. <http://dx.doi.org/10.1186/1471-2164-10-550>
- Gautier, M., Foucaud, J., Gharbi, K., Cezard, T., Galan, M., Loiseau, A., . . . Estoup, A. (2013). Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. *Mol Ecol*, 22(14), 3766-3779. <http://dx.doi.org/10.1111/mec.12360>
- Gautier, M., & Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Molecular Ecology*, 20(15), 3128-3143. <http://dx.doi.org/10.1111/j.1365-294X.2011.05163.x>
- Gavrilova, O., Jonas, M., Erb, K., & Haberl, H. (2010). International trade and Austria's livestock system: Direct and hidden carbon emission flows associated with production and consumption of products. *Ecological Economics*, 69(4), 920-929. <http://dx.doi.org/http://dx.doi.org/10.1016/j.ecolecon.2009.11.015>
- Geerts, S., Holmes, P. H., Eisler, M. C., & Diall, O. (2001). African bovine trypanosomiasis: the problem of drug resistance. *Trends Parasitol*, 17(1), 25-28. [http://dx.doi.org/http://dx.doi.org/10.1016/S1471-4922\(00\)01827-4](http://dx.doi.org/http://dx.doi.org/10.1016/S1471-4922(00)01827-4)
- Geerts, S., Osaer, S., Goossens, B., & Faye, D. (2009). Trypanotolerance in small ruminants of sub-Saharan Africa. *Trends Parasitol*, 25(3), 132-138. <http://dx.doi.org/10.1016/j.pt.2008.12.004>
- Gelasakis, A. I., Arsenos, G., Hickford, J., Zhou, H., Psifidi, A., Valergakis, G. E., & Banos, G. (2013). Polymorphism of the MHC-DQA2 gene in the Chios dairy sheep population and its association with footrot. *Livestock Science*, 153(1-3), 56-59. <http://dx.doi.org/https://doi.org/10.1016/j.livsci.2013.02.011>

- Georges, M. (2007). Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals. *Annu Rev Genomics Hum Genet*, 8, 131-162. <http://dx.doi.org/10.1146/annurev.genom.8.080706.092408>
- Gibson, G. (2012). Rare and common variants: twenty arguments. *Nature reviews. Genetics*, 13(2), 135-145. <http://dx.doi.org/10.1038/nrg3118>
- Gilbert, K. J., Andrew, R. L., Bock, D. G., Franklin, M. T., Kane, N. C., Moore, J. S., . . . Veen, T. (2012). Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program structure. *Molecular Ecology*, 21(20), 4925-4930.
- Gilly, A., Ritchie, G. R., Southam, L., Farmaki, A.-E., Tsafantakis, E., Dedoussis, G., & Zeggini, E. (2016). Very low-depth sequencing in a founder population identifies a cardioprotective APOC3 signal missed by genome-wide imputation. *Human Molecular Genetics*, 25(11), 2360-2365. <http://dx.doi.org/10.1093/hmg/ddw088>
- Globaldiv consortium. (2010). *Current strategies to characterize farm animal genetic resources*. Retrieved from [http://www.globaldiv.eu/NL/GlobalDiv\\_Newsletter%20no%2014.pdf](http://www.globaldiv.eu/NL/GlobalDiv_Newsletter%20no%2014.pdf)
- Goddard, M. E., & Hayes, B. J. (2009). Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics*, 10(6), 381-391.
- Goe, M., & Stranzinger, G. (2002). Livestock as a component of disaster risk management: quantifying the effect of emergency relief operations on animal genetic resources. *Concept Note prepared for FAO Animal Genetic Resources Group, Rome. Breeding Biology Group, Institute of Animal Sciences, Swiss Federal Institute of Technology, Zurich*.
- Goossens, B., Osaer, S., Kora, S., Jaitner, J., Ndao, M., & Geerts, S. (1997). The interaction of *Trypanosoma congolense* and *Haemonchus contortus* in Djallonke sheep. *Int J Parasitol*, 27(12), 1579-1584. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9467745>
- Goossens, B., Osaer, S., Kora, S., & Ndao, M. (1998). Haematological changes and antibody response in trypanotolerant sheep and goats following experimental *Trypanosoma congolense* infection. *Vet Parasitol*, 79(4), 283-297. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9831952>
- Goossens, B., Osaer, S., Ndao, M., Van Wingham, J., & Geerts, S. (1999). The susceptibility of Djallonke and Djallonke-Sahelian crossbred sheep to *Trypanosoma congolense* and helminth infection under different diet levels. *Vet Parasitol*, 85(1), 25-41. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10447190>
- Gormez, Z., Bakir-Gungor, B., & Sagiroglu, M. S. (2014). HomSI: a homozygous stretch identifier from next-generation sequencing data. *Bioinformatics*, 30(3), 445-447. <http://dx.doi.org/10.1093/bioinformatics/btt686>
- Gouy, M., Guindon, S., & Gascuel, O. (2010). SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution*, 27(2), 221-224. <http://dx.doi.org/10.1093/molbev/msp259>
- Groenen, M. A., Megens, H.-J., Zare, Y., Warren, W. C., Hillier, L. W., Crooijmans, R. P., . . . Cheng, H. H. (2011). The development and characterization of a 60K SNP chip for chicken. *BMC Genomics*, 12, 274.
- Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., . . . The Globaldiv Consortium. (2010). Genetic diversity in farm animals – a review. *Animal Genetics*, 41, 6-31. <http://dx.doi.org/10.1111/j.1365-2052.2010.02038.x>
- Gu, W., Gurguis, C. I., Zhou, J. J., Zhu, Y., Ko, E.-A., Ko, J.-H., . . . Zhou, T. (2015). Functional and Structural Consequence of Rare Exonic Single Nucleotide Polymorphisms: One Story, Two Tales. *Genome Biology and Evolution*, 7(10), 2929-2940. <http://dx.doi.org/10.1093/gbe/evv191>
- Gudbjartsson, D. F., Sulem, P., Helgason, H., Gylfason, A., Gudjonsson, S. A., Zink, F., . . . Stefansson, K. (2015). Sequence variants from whole genome sequencing a large group of Icelanders. *Scientific Data*, 2, 150011. <http://dx.doi.org/10.1038/sdata.2015.11>
- Gudmundsson, J., Sulem, P., Gudbjartsson, D. F., Masson, G., Agnarsson, B. A., Benediksdottir, K. R., . . . Stefansson, K. (2012). A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. *Nat Genet*, 44(12), 1326-1329. <http://dx.doi.org/http://www.nature.com/ng/journal/v44/n12/abs/ng.2437.html#supplementary-information>

- Gulati, S. (2008). Technology-Enhanced Learning in Developing Nations: A review (developing countries; technology; e-learning; access; rhetoric; reality) Retrieved from <http://www.irrodl.org/index.php/irrodl/article/view/477>. (1492-3831). Retrieved 2008-02-05, from The International Review of Research in Open and Distributed Learning. [S.l.], v. 9, n. 1, February 2008. ISSN 1492-3831 <http://www.irrodl.org/index.php/irrodl/article/view/477>
- Guo, S., Savolainen, P., Su, J., Zhang, Q., Qi, D., Zhou, J., . . . Liu, J. (2006). Origin of mitochondrial DNA diversity of domestic yaks. *BMC Evolutionary Biology*, 6(1), 73.
- Guo, Y., Long, J., He, J., Li, C.-I., Cai, Q., Shu, X.-O., . . . Li, C. (2012). Exome sequencing generates high quality data in non-target regions. *BMC Genomics*, 13(1), 1-10. <http://dx.doi.org/10.1186/1471-2164-13-194>
- Guo, Z., González, J. F., Hernandez, J. N., McNeilly, T. N., Corripio-Miyar, Y., Frew, D., . . . Li, R. W. (2016). Possible mechanisms of host resistance to *Haemonchus contortus* infection in sheep breeds native to the Canary Islands. *Scientific Reports*, 6, 26200. <http://dx.doi.org/10.1038/srep26200>
- Gutierrez-Espeleta, G. A., Hedrick, P. W., Kalinowski, S. T., Garrigan, D., & Boyce, W. M. (2001). Is the decline of desert bighorn sheep from infectious disease the result of low MHC variation? *Heredity*, 86(4), 439-450. Retrieved from <http://dx.doi.org/10.1046/j.1365-2540.2001.00853.x>
- Gwakisa, P., Kemp, S., & Teale, A. (1994). Characterization of Zebu cattle breeds in Tanzania using random amplified polymorphic DNA markers. *Animal Genetics*, 25(S1), 89-94.
- Habila, N., Inuwa, M. H., Aimola, I. A., Udeh, M. U., & Haruna, E. (2012). Pathogenic mechanisms of *Trypanosoma evansi* infections. *Research in Veterinary Science*, 93(1), 13-17. <http://dx.doi.org/http://dx.doi.org/10.1016/j.rvsc.2011.08.011>
- Hamblin, M. T., Warburton, M. L., & Buckler, E. S. (2007). Empirical comparison of simple sequence repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. *PloS one*, 2(12), e1367.
- Hammond, K. (1994). *Conservation of domestic animal diversity: Global overview*. Paper presented at the Proceedings of the World Congress on Genetics Applied to Livestock Production, Department of Animal and Poultry Science, University of Guelph, Ontario, Canada
- Hammond, K. (1999). *Management of farm animal genetic resources*. Paper presented at the Economic valuation of animal genetic resources, Proceedings of FAO/ILRI workshop, 15-17 March, 1999, Fao Headquarters, Rome Italy
- Hanks, J. (2001). Conservation strategies for Africa's large mammals. *Reproduction, Fertility and Development.*, 13, 459-468. <http://dx.doi.org/10.1071/RD01070>
- Hanotte, O., Ronin, Y., Agaba, M., Nilsson, P., Gelhaus, A., Horstmann, R., . . . Teale, A. (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(13), 7443-7448. <http://dx.doi.org/10.1073/pnas.1232392100>
- Harrison, R. G. (1989). Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology & Evolution*, 4(1), 6-11. [http://dx.doi.org/http://dx.doi.org/10.1016/0169-5347\(89\)90006-2](http://dx.doi.org/http://dx.doi.org/10.1016/0169-5347(89)90006-2)
- Hassan, M., Good, B., Hanrahan, J. P., Champion, D., Sayers, G., Mulcahy, G., & Sweeney, T. (2011). The dynamic influence of the DRB1\*1101 allele on the resistance of sheep to experimental *Teladorsagia circumcincta* infection. *Veterinary Research*, 42(1), 46. <http://dx.doi.org/10.1186/1297-9716-42-46>
- Hayes, B., & Goddard, M. (2010). Genome-wide association and genomic selection in animal breeding. *Genome*, 53(11), 876-883. Retrieved from <http://www.ingentaconnect.com/content/nrc/gen/2010/00000053/00000011/art00006>
- Hedrick, P. W. (1998). Balancing selection and MHC. *Genetica*, 104(3), 207-214. <http://dx.doi.org/10.1023/a:1026494212540>
- Hedrick, P. W. (2001). Conservation genetics: where are we now? *Trends in Ecology & Evolution*, 16(11), 629-636. [http://dx.doi.org/http://dx.doi.org/10.1016/S0169-5347\(01\)02282-0](http://dx.doi.org/http://dx.doi.org/10.1016/S0169-5347(01)02282-0)
- Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M. I., Ogden, R., Limborg, M. T., . . . Nielsen, E. E. (2011). Application of SNPs for population genetics of nonmodel organisms: new

- opportunities and challenges. *Molecular Ecology Resources*, 11, 123-136. <http://dx.doi.org/10.1111/j.1755-0998.2010.02943.x>
- Hendrickx, G., de La Rocque, S., & Mattioli, R. C. (2004). *Long-term tsetse and trypanosomiasis management options in West Africa* (Vol. 6). Rome, Italy: Food & Agriculture Organisation.
- Henson, E. L. (1992). *In situ conservation of livestock and poultry*. Rome: Food and Agriculture Organization of the United Nations.
- Herrero-Medrano, J. M., Megens, H.-J., Groenen, M. A., Ramis, G., Bosse, M., Pérez-Enciso, M., & Crooijmans, R. P. (2013). Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. *BMC genetics*, 14(1), 106.
- Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., . . . Flicek, P. (2016). Ensembl comparative genomics resources. *Database*, 2016 <http://dx.doi.org/10.1093/database/bav096>
- Herrero, M., Thornton, P. K., Gerber, P., & Reid, R. S. (2009). Livestock, livelihoods and the environment: understanding the trade-offs. *Current Opinion in Environmental Sustainability*, 1(2), 111-120. <http://dx.doi.org/http://dx.doi.org/10.1016/j.cosust.2009.10.003>
- Herrmann-Hoesing, L. M., White, S. N., Kappmeyer, L. S., Herndon, D. R., & Knowles, D. P. (2008). Genomic analysis of *Ovis aries* (Ovar) MHC class IIa loci. *Immunogenetics*, 60(3), 167-176. <http://dx.doi.org/10.1007/s00251-008-0275-5>
- Herrmann, L. M., Brown, W. C., Lewis, G. S., & Knowles, D. P. (2005). Identification and phylogenetic analysis of 15 MHC class II DRB1  $\beta$ 1 expressed alleles in a ewe-lamb flock. *Immunogenetics*, 57(11), 855-863. <http://dx.doi.org/10.1007/s00251-005-0050-9>
- Hickford, J. G. H., Forrest, R. H. J., Zhou, H., Fang, Q., & Frampton, C. M. (2011). Association between variation in faecal egg count for a mixed field-challenge of nematode parasites and ovine MHC-DQA2 polymorphism. *Veterinary Immunology and Immunopathology*, 144(3-4), 312-320. <http://dx.doi.org/https://doi.org/10.1016/j.vetimm.2011.08.014>
- Hickford, J. G. H., Zhou, H., & Fang, Q. (2007). Haplotype analysis of the DQA genes in sheep: Evidence supporting recombination between the loci1. *Journal of Animal Science*, 85(3), 577-582. <http://dx.doi.org/10.2527/jas.2006-217>
- Hickford, J. G. H., Zhou, H., Slow, S., & Fang, Q. (2004). Diversity of the ovine DQA2 gene1. *Journal of Animal Science*, 82(6), 1553-1563. <http://dx.doi.org/10.2527/2004.8261553x>
- Hiemstra, S. J., Drucker, A. G., Tvedt, M. W., Louwaars, N., Oldenbroek, J., Awgichew, K., . . . da Silva Mariante, A. (2006a). Exchange, use and conservation of animal genetic resources. *Food and Agriculture Organization of the United Nations, Rome*,
- Hiemstra, S. J., van der Lende, T., & Woelders, H. (2006b). The potential of cryopreservation and reproductive technologies for animal genetic resources conservation strategies. In John Ruane & A. Sonnino (Eds.), *The role of biotechnology in exploring and protecting agricultural genetic resources*. FAO, Rome (pp. 45-59).
- Hill, E. W., O'Gorman, G. M., Agaba, M., Gibson, J. P., Hanotte, O., Kemp, S. J., . . . MacHugh, D. E. (2005). Understanding bovine trypanosomiasis and trypanotolerance: the promise of functional genomics. *Vet Immunol Immunopathol*, 105(3-4), 247-258. <http://dx.doi.org/10.1016/j.vetimm.2005.02.004>
- Hoffmann, I. (2010). Climate change and the characterization, breeding and conservation of animal genetic resources. *Animal Genetics*, 41, 32-46. <http://dx.doi.org/10.1111/j.1365-2052.2010.02043.x>
- Hoffmann, I., & Scherf, B. (2005). *Animal genetic resources-Time to worry? in Livestock report 2006 global perspective*. Rome Italy: Information division, Food and Agriculture Organisation, Viale delle Terme di Caracalla. A. Mcleod (Ed.) (pp. 57-76). Retrieved from <http://www.fao.org/docrep/009/a0255e/a0255e08.htm>
- Holmes, P. H. (1997). New approaches to the integrated control of trypanosomiasis. *Vet. Parasitol.*, 71(2-3), 121-135. [http://dx.doi.org/http://dx.doi.org/10.1016/S0304-4017\(97\)00026-5](http://dx.doi.org/http://dx.doi.org/10.1016/S0304-4017(97)00026-5)
- Holt, W. V., & Pickard, A. R. (1999). Role of reproductive technologies and genetic resource banks in animal conservation. *Reviews of Reproduction*, 4(3), 143-150.
- Horn, D. (2014). Antigenic variation in African trypanosomes. *Mol Biochem Parasitol*, 195(2), 123-129. <http://dx.doi.org/10.1016/j.molbiopara.2014.05.001>

- Horton, R., Wilming, L., Rand, V., Lovering, R. C., Bruford, E. A., Khodiyar, V. K., . . . Beck, S. (2004). Gene map of the extended human MHC. *Nat Rev Genet*, 5(12), 889-899. Retrieved from <http://dx.doi.org/10.1038/nrg1489>
- Hu, Z. L., Park, C. A., & Reecy, J. M. (2016). Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Res*, 44(D1), D827-833. <http://dx.doi.org/10.1093/nar/gkv1233>
- Hu, Z. L., Park, C. A., Wu, X. L., & Reecy, J. M. (2013). Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res*, 41(Database issue), D871-879. <http://dx.doi.org/10.1093/nar/gks1150>
- Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 4(1), 44-57. <http://dx.doi.org/10.1038/nprot.2008.211>
- Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9(5), 1322-1332. <http://dx.doi.org/10.1111/j.1755-0998.2009.02591.x>
- Hughes, A. L., & Yeager, M. (1998). Natural selection at major histocompatibility complex loci of vertebrates. *Annual review of genetics*, 32(1), 415-435.
- Hundertmark, K., & Van Daele, L. (2010). Founder effect and bottleneck signatures in an introduced, insular population of elk. *Conservation Genetics*, 11(1), 139-147. <http://dx.doi.org/10.1007/s10592-009-0013-z>
- Huo, J., Wu, G., Chen, T., Huo, H., Yuan, F., Liu, L., . . . Miao, Y. (2014). Genetic diversity of local Yunnan chicken breeds and their relationships with Red Junglefowl. *Genetics and Molecular Research*, 13(2), 3371.
- Hurst, L. D. (2009). Genetics and the understanding of selection. *Nat Rev Genet*, 10(2), 83-93. Retrieved from <http://dx.doi.org/10.1038/nrg2506>
- IAEA. (2002). Campaign launched to eliminate tsetse fly, Pesticide Outlook. *The Royal Society of Chemistry*, 13(2), 77-78. <http://dx.doi.org/DOI: 10.1039/b203222h>
- Inoue, H., Nojima, H., & Okayama, H. (1990). High efficiency transformation of Escherichia coli with plasmids. *Gene*, 96(1), 23-28. [http://dx.doi.org/http://dx.doi.org/10.1016/0378-1119\(90\)90336-P](http://dx.doi.org/http://dx.doi.org/10.1016/0378-1119(90)90336-P)
- International Livestock Research Institute. (2005). *Djallonke Origin and Distribution*. Retrieved from [http://agtr.ilri.cgiar.org/index.php?option=com\\_content&view=article&id=134&Itemid=148](http://agtr.ilri.cgiar.org/index.php?option=com_content&view=article&id=134&Itemid=148) Accessed 13.04.2014
- Iraqi, F., Clapcott, S. J., Kumari, P., Haley, C. S., Kemp, S. J., & Teale, A. J. (2000). Fine mapping of trypanosomiasis resistance loci in murine advanced intercross lines. *Mamm Genome*, 11(8), 645-648. <http://dx.doi.org/10.1007/s003350010133>
- Jahnke, H., Tacher, G., Kiel, P., & Rojat, D. (1988). *Livestock production in tropical Africa, with special reference to the tsetse-affected zone*. Nairobi, Kenya: International Livestock research Center for Africa and International center for research on Animal diseases.
- Jahnke, H. E., & Jahnke, H. E. (1982). *Livestock production systems and livestock development in tropical Africa* (Vol. 35): Kieler Wissenschaftsverlag Vauk Kiel.
- Jensen, P. (2006). Domestication—From behaviour to genes and back again. *Applied Animal Behaviour Science*, 97(1), 3-15. <http://dx.doi.org/http://dx.doi.org/10.1016/j.applanim.2005.11.015>
- Jha, B. A., Gazestani, V. H., Yip, C. W., & Salavati, R. (2015). The DRBD13 RNA binding protein is involved in the insect-stage differentiation process of Trypanosoma brucei. *FEBS Lett*, 589(15), 1966-1974. <http://dx.doi.org/10.1016/j.febslet.2015.05.036>
- Jin, K., Speed, T. P., & Thomson, G. (1995). Tests of random mating for a highly polymorphic locus: application to HLA data. *Biometrics*, 51(3), 1064-1076. <http://dx.doi.org/10.2307/2533005>
- Johansson, B. M., & Wiles, M. V. (1995). Evidence for involvement of activin A and bone morphogenetic protein 4 in mammalian mesoderm and hematopoietic development. *Molecular and Cellular Biology*, 15(1), 141-151. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC231923/>
- Jolles, A. E., Beechler, B. R., & Dolan, B. P. (2015). Beyond mice and men: environmental change, immunity and infections in wild ungulates. *Parasite Immunol*, 37(5), 255-266. <http://dx.doi.org/10.1111/pim.12153>

- Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*, 8(3), 275-282. <http://dx.doi.org/10.1093/bioinformatics/8.3.275>
- Kabayo, J. P. (2002). Aiming to eliminate tsetse from Africa. *Trends Parasitol*, 18(11), 473-475. [http://dx.doi.org/http://dx.doi.org/10.1016/S1471-4922\(02\)02371-1](http://dx.doi.org/http://dx.doi.org/10.1016/S1471-4922(02)02371-1)
- Kadarmideen, H. N., Watson-Haigh, N. S., & Andronicos, N. M. (2011). Systems biology of ovine intestinal parasite resistance: disease gene modules and biomarkers. *Molecular BioSystems*, 7(1), 235-246. <http://dx.doi.org/10.1039/C0MB00190B>
- Kadwell, M., Fernandez, M., Stanley, H. F., Baldi, R., Wheeler, J. C., Rosadio, R., & Bruford, M. W. (2001). Genetic analysis reveals the wild ancestors of the llama and the alpaca. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1485), 2575. Retrieved from <http://rspb.royalsocietypublishing.org/content/268/1485/2575.abstract>
- Kahn, L. H. (2006). Confronting zoonoses, linking human and veterinary medicine. *Emerging Infectious Diseases*, 12(4), 556.
- Kalinowski, S. T., & Taper, M. L. (2006). Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. *Conservation Genetics*, 7(6), 991-995.
- Kancheva, D., Atkinson, D., De Rijk, P., Zimon, M., Chamova, T., Mitev, V., . . . Jordanova, A. (2016). Novel mutations in genes causing hereditary spastic paraplegia and Charcot-Marie-Tooth neuropathy identified by an optimized protocol for homozygosity mapping based on whole-exome sequencing. *Genet Med*, 18(6), 600-607. <http://dx.doi.org/10.1038/gim.2015.139>
- Kantanen, J., Løvendahl, P., Strandberg, E., Eythorsdottir, E., Li, M.-H., Kettunen-Praebel, A., . . . Meuwissen, T. (2015). Utilization of farm animal genetic resources in a changing agro-ecological environment in the Nordic countries. *Frontiers in Genetics*, 6, 52. <http://dx.doi.org/10.3389/fgene.2015.00052>
- Kardos, M., Luikart, G., Bunch, R., Dewey, S., Edwards, W., McWilliam, S., . . . Kijas, J. (2015). Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. *Mol Ecol*, 24(22), 5616-5632. <http://dx.doi.org/10.1111/mec.13415>
- Kato, C. D., Nanteza, A., Mugasa, C., Edyelu, A., Matovu, E., & Alibu, V. P. (2015). Clinical profiles, disease outcome and co-morbidities among T. b. rhodesiense sleeping sickness patients in Uganda. *PLoS One*, 10(2), e0118370. <http://dx.doi.org/10.1371/journal.pone.0118370>
- Kaufmann, J., Dwinger, R. H., Hallebeek, A., van Dijk, B., & Pfister, K. (1992). The interaction of *Trypanosoma congolense* and *Haemonchus contortus* infections in trypanotolerant N'Dama cattle. *Vet Parasitol*, 43(3-4), 157-170. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/1413448>
- Kawahara-Miki, R., Tsuda, K., Shiwa, Y., Arai-Kichise, Y., Matsumoto, T., Kanesaki, Y., . . . Yoshikawa, H. (2011). Whole-genome resequencing shows numerous genes with nonsynonymous SNPs in the Japanese native cattle Kuchinoshima-Ushi. *Bmc Genomics*, 12(1), 103.
- Keane, O. M., Dodds, K. G., Crawford, A. M., & McEwan, J. C. (2007). Transcriptional profiling of *Ovis aries* identifies Ovar-DQA1 allele frequency differences between nematode-resistant and susceptible selection lines. *Physiological Genomics*, 30(3), 253. Retrieved from <http://physiolgenomics.physiology.org/content/30/3/253.abstract>
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649. <http://dx.doi.org/10.1093/bioinformatics/bts199>
- Kelly, T. N., Li, C., Hixson, J. E., Gu, D., Rao, D. C., Huang, J., . . . He, J. (2017). Resequencing Study Identifies Rare Renin–Angiotensin–Aldosterone System Variants Associated With Blood Pressure Salt-Sensitivity: The GenSalt Study. *American Journal of Hypertension*, 30(5), 495-501. <http://dx.doi.org/10.1093/ajh/hpx004>
- Kemp, S. J., & Teale, A. J. (1998). Genetic basis of trypanotolerance in cattle and mice. *Parasitol Today*, 14(11), 450-454. [http://dx.doi.org/http://dx.doi.org/10.1016/S0169-4758\(98\)01334-9](http://dx.doi.org/http://dx.doi.org/10.1016/S0169-4758(98)01334-9)
- Kijas, J. W., Lenstra, J. A., Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., . . . International Sheep Genomics Consortium, M. (2012). Genome-wide analysis of the world's sheep breeds

- reveals high levels of historic mixture and strong recent selection. *PLoS Biol*, 10(2), e1001258. <http://dx.doi.org/10.1371/journal.pbio.1001258>
- Kijas, J. W., Porto-Neto, L., Dominik, S., Reverter, A., Bunch, R., McCulloch, R., . . . McEwan, J. (2014). Linkage disequilibrium over short physical distances measured in sheep using a high-density SNP chip. *Animal Genetics*, 45(5), 754-757.
- Kijas, J. W., Townley, D., Dalrymple, B. P., Heaton, M. P., Maddox, J. F., McGrath, A., . . . McWilliam, S. (2009). A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS one*, 4(3), e4668.
- Kim, E. S., Elbeltagy, A. R., Aboul-Naga, A. M., Rischkowsky, B., Sayre, B., Mwacharo, J. M., & Rothschild, M. F. (2016). Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*, 116(3), 255-264. <http://dx.doi.org/10.1038/hdy.2015.94>
- Kim, K. I., Lee, J. H., Li, K., Zhang, Y. P., Lee, S. S., Gongora, J., & Moran, C. (2002). Phylogenetic relationships of Asian and European pig breeds determined by mitochondrial DNA D-loop sequence polymorphism. *Animal Genetics*, 33(1), 19-25. <http://dx.doi.org/10.1046/j.1365-2052.2002.00784.x>
- Kim, S. Y., Li, Y., Guo, Y., Li, R., Holmkvist, J., Hansen, T., . . . Nielsen, R. (2010). Design of association studies with pooled or un-pooled next-generation sequencing data. *Genetic Epidemiology*, 34(5), 479-491.
- Klein, J., Ono, H., Klein, D., & O'hUigin, C. (1993). The Accordion Model of Mhc Evolution. In J. Gergely, M. Benczúr, A. Erdei, A. Falus, G. Füst, G. Medgyesi, G. Petrányi & É. Rajnavölgyi (Eds.), *Progress in Immunology Vol. VIII* (pp. 137-143). Berlin, Heidelberg: Springer Berlin Heidelberg. [http://dx.doi.org/10.1007/978-3-642-51479-1\\_18](http://dx.doi.org/10.1007/978-3-642-51479-1_18)
- Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Pääbo, S., Villablanca, F. X., & Wilson, A. C. (1989). Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences*, 86(16), 6196-6200.
- Koehler-Rollefson, I., & Meyer, H. (2014). Access and Benefit-sharing of Animal Genetic Resources Using the Nagoya Protocol as a Framework for the Conservation and Sustainable Use of Locally. Retrieved from [http://www.fao.org/WAICENT/faoINFO/AGRICULT/AGAInfo/programmes/en/genetics/documents/ITWG\\_AnGR\\_8/side-event/01\\_Invitation-ABS\\_for\\_AnGR\\_GIZ\\_LPP.pdf](http://www.fao.org/WAICENT/faoINFO/AGRICULT/AGAInfo/programmes/en/genetics/documents/ITWG_AnGR_8/side-event/01_Invitation-ABS_for_AnGR_GIZ_LPP.pdf)
- Kofler, R., Nolte, V., & Schlotterer, C. (2016). The impact of library preparation protocols on the consistency of allele frequency estimates in Pool-Seq data. *Mol Ecol Resour*, 16(1), 118-122. <http://dx.doi.org/10.1111/1755-0998.12432>
- Kofler, R., Pandey, R. V., & Schlotterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, 27(24), 3435-3436. <http://dx.doi.org/10.1093/bioinformatics/btr589>
- Koh, M. C., Lim, C. H., Chua, S. B., Chew, S. T., & Phang, S. T. W. (1998). Random amplified polymorphic DNA (RAPD) fingerprints for identification of red meat animal species. *Meat Science*, 48(3-4), 275-285. [http://dx.doi.org/http://dx.doi.org/10.1016/S0309-1740\(97\)00104-6](http://dx.doi.org/http://dx.doi.org/10.1016/S0309-1740(97)00104-6)
- Köhler-Rollefson, I. (1997). Indigenous practices of animal genetic resource management and their relevance for the conservation of domestic animal diversity in developing countries. *Journal of Animal Breeding and Genetics*, 114(1-6), 231-238.
- Kohn, M. H., Murphy, W. J., Ostrander, E. A., & Wayne, R. K. (2006). Genomics and conservation genetics. *Trends in Ecology & Evolution*, 21(11), 629-637. <http://dx.doi.org/http://dx.doi.org/10.1016/j.tree.2006.08.001>
- Konnai, S., Nagaoka, Y., Takesima, S., Onuma, M., & Aida, Y. (2003). Technical Note: DNA Typing for Ovine MHC DRB1 Using Polymerase Chain Reaction-Restriction Fragment Length Polymorphism (PCR-RFLP). *Journal of Dairy Science*, 86(10), 3362-3365. [http://dx.doi.org/https://doi.org/10.3168/jds.S0022-0302\(03\)73939-3](http://dx.doi.org/https://doi.org/10.3168/jds.S0022-0302(03)73939-3)
- Konnai, S., Takeshima, S.-n., Tajima, S., Yin, S. A., Okada, K., Onuma, M., & Aida, Y. (2003). The Influence of Ovine MHC Class II DRB1 Alleles on Immune Response in Bovine Leukemia

- Virus Infection. *Microbiology and Immunology*, 47(3), 223-232. <http://dx.doi.org/10.1111/j.1348-0421.2003.tb03391.x>
- Kosgey, I. S., Baker, R. L., Udo, H. M. J., & Van Arendonk, J. A. M. (2006). Successes and failures of small ruminant breeding programmes in the tropics: a review. *Small Ruminant Res*, 61(1), 13-28. <http://dx.doi.org/http://dx.doi.org/10.1016/j.smallrumres.2005.01.003>
- Kosgey, I. S., & Okeyo, A. M. (2007). Genetic improvement of small ruminants in low-input, smallholder production systems: Technical and infrastructural issues. *Small Ruminant Res*, 70(1), 76-88. <http://dx.doi.org/10.1016/j.smallrumres.2007.01.007>
- Kuehn, L. A., Keele, J. W., Bennett, G. L., McDanel, T. G., Smith, T. P. L., Snelling, W. M., . . . Thallman, R. M. (2011). Predicting breed composition using breed frequencies of 50,000 markers from the US Meat Animal Research Center 2,000 Bull Project. *Journal of Animal Science*, 89(6), 1742-1750. <http://dx.doi.org/10.2527/jas.2010-3530>
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7), 1870-1874. <http://dx.doi.org/10.1093/molbev/msw054>
- Lacy, R. C. (1997). Importance of genetic variation to the viability of mammalian populations. *Journal of Mammalogy*, 78, 320-335.
- Lawrence, P. R., & Pearson, R. A. (2002). Use of draught animal power on small mixed farms in Asia. *Agricultural Systems*, 71(1-2), 99-110. [http://dx.doi.org/http://dx.doi.org/10.1016/S0308-521X\(01\)00038-5](http://dx.doi.org/http://dx.doi.org/10.1016/S0308-521X(01)00038-5)
- Le, S. Q., & Durbin, R. (2011). SNP detection and genotyping from low-coverage sequencing data on multiple diploid samples. *Genome Research*, 21(6), 952-960. <http://dx.doi.org/10.1101/gr.113084.110>
- Lee, C. Y., Qin, J., Munyard, K., Subramaniam, N., Wetherall, J., Stear, M., & Groth, D. (2012). Conserved haplotype blocks within the sheep MHC and low SNP heterozygosity in the Class IIa subregion.
- Lee, C. Y., Qin, J., Munyard, K. A., Siva Subramaniam, N., Wetherall, J. D., & Stear, M. J. (2011). Conserved haplotype blocks within the sheep MHC and low SNP heterozygosity in the Class IIa subregion. *Anim Genet*, 43 <http://dx.doi.org/10.1111/j.1365-2052.2011.02268.x>
- Lee, K. T., Chung, W. H., Lee, S. Y., Choi, J. W., Kim, J., Lim, D., . . . Choy, Y. H. (2013). Whole-genome resequencing of Hanwoo (Korean cattle) and insight into regions of homozygosity. *BMC Genomics*, 14(1), 519.
- Lee, S., Abecasis, Gonçalo R., Boehnke, M., & Lin, X. (2014). Rare-Variant Association Analysis: Study Designs and Statistical Tests. *The American Journal of Human Genetics*, 95(1), 5-23. <http://dx.doi.org/http://dx.doi.org/10.1016/j.ajhg.2014.06.009>
- Leigh, O., Emikpe, B., & Ogunsola, J. (2014). Histopathological changes in some reproductive and endocrine organs of Trypanosoma brucei infected West African Dwarf goat does. *Bulgarian J Vet Med*, 18(1), 31-39.
- Levin, I., Crittenden, L. B., & Dodgson, J. B. (1993). Genetic map of the chicken Z chromosome using random amplified polymorphic DNA (RAPD) markers. *Genomics*, 16(1), 224-230.
- Lewis, J., Abas, Z., Dadousis, C., Lykidis, D., Paschou, P., & Drineas, P. (2011). Tracing Cattle Breeds with Principal Components Analysis Ancestry Informative SNPs. *PLoS ONE*, 6(4), e18007. <http://dx.doi.org/10.1371/journal.pone.0018007>
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, 26(5), 589-595. <http://dx.doi.org/10.1093/bioinformatics/btp698>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079. <http://dx.doi.org/10.1093/bioinformatics/btp352>
- Li, Y., Sidore, C., Kang, H. M., Boehnke, M., & Abecasis, G. R. (2011). Low-coverage sequencing: Implications for design of complex trait association studies. *Genome Research*, <http://dx.doi.org/10.1101/gr.117259.110>
- Liu, G., Sun, D., Wu, H., Zhang, M., Huan, H., Xu, J., . . . Shi, M. (2015). Distinct Contributions of CD4+ and CD8+ T Cells to Pathogenesis of Trypanosoma brucei Infection in the Context of Gamma Interferon and Interleukin-10. *Infect Immun*, 83(7), 2785-2795. <http://dx.doi.org/10.1128/IAI.00357-15>

- Liu, Z., & Cordes, J. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*, 238(1), 1-37.
- Loftus, R. T., Ertugrul, O., Harba, A. H., El-Barody, M. A. A., MacHugh, D. E., Park, S. D. E., & Bradley, D. G. (1999). A microsatellite survey of cattle from a centre of origin: the Near East. *Molecular Ecology*, 8(12), 2015-2022. <http://dx.doi.org/10.1046/j.1365-294x.1999.00805.x>
- Lopes, M. S., Mendonça, D., Rojer, H., Cabral, V., Bettencourt, S. X., & da Câmara Machado, A. (2015). Morphological and genetic characterization of an emerging Azorean horse breed: the Terceira Pony. *Frontiers in Genetics*, 6, 62. <http://dx.doi.org/10.3389/fgene.2015.00062>
- Luikart, G., & Cornuet, J.-M. (1998). Empirical Evaluation of a Test for Identifying Recently Bottlenecked Populations from Allele Frequency Data. *Conservation Biology*, 12(1), 228-237. <http://dx.doi.org/10.1111/j.1523-1739.1998.96388.x>
- Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The Power and Promise of Population Genomics: from Genotyping to Genome Typing. *Nature Reviews: Genetics*, 4(12), 981-994. <http://dx.doi.org/http://dx.doi.org/10.1038/nrg1226>
- Luikart, G., Fernández, H., Mashkour, M., England, P., Taberlet, P., Zeder, M., . . . Bradley, D. (2006). Origins and diffusion of domestic goats inferred from DNA markers *Documenting domestication: New genetic and archaeological paradigms* (pp. 294-305). Los Angeles: University of California Press.
- Maccari, G., Robinson, J., Ballingall, K., Guethlein, L. A., Grimholt, U., Kaufman, J., . . . Marsh, Steven G. E. (2017). IPD-MHC 2.0: an improved inter-species database for the study of the major histocompatibility complex. *Nucleic Acids Research*, 45(D1), D860-D864. <http://dx.doi.org/10.1093/nar/gkw1050>
- Maddox, J. F., Davies, K. P., Crawford, A. M., Hulme, D. J., Vaiman, D., Cribiu, E. P., . . . Kang, N. (2001). An enhanced linkage map of the sheep genome comprising more than 1000 loci. *Genome Research*, 11(7), 1275-1289.
- Magez, S., & Radwanska, M. (2014). Adaptive Immunity and Trypanosomiasis-Driven B-Cell Destruction. In S. Magez & M. Radwanska (Eds.), *Trypanosomes and Trypanosomiasis* (pp. 115-138): Springer Vienna. [http://dx.doi.org/10.1007/978-3-7091-1556-5\\_5](http://dx.doi.org/10.1007/978-3-7091-1556-5_5)
- Maillard, J. C., Berthier, D., Thevenon, S., Piquemal, D., Chantal, I., & Marti, J. (2005). Efficiency and limits of the Serial Analysis of Gene Expression (SAGE) method: discussions based on first results in bovine trypanotolerance. *Vet Immunol Immunopathol*, 108(1-2), 59-69. <http://dx.doi.org/10.1016/j.vetimm.2005.08.005>
- MÄKinen, E. A. M. A., Chowdhary, B. P., Andersson, L., & Gustavsson, I. (1989). Chromosomal localization of the ovine major histocompatibility complex (OLA) by in situ hybridization. *Hereditas*, 111(1), 87-90. <http://dx.doi.org/10.1111/j.1601-5223.1989.tb00381.x>
- Mamoudou, A., Delespaux, V., Chepnda, V., Hachimou, Z., Andrikaye, J. P., Zoli, A., & Geerts, S. (2008). Assessment of the occurrence of trypanocidal drug resistance in trypanosomes of naturally infected cattle in the Adamaoua region of Cameroon using the standard mouse test and molecular tools. *Acta Trop*, 106(2), 115-118. <http://dx.doi.org/10.1016/j.actatropica.2008.02.003>
- Manna, P. T., Boehm, C., Leung, K. F., Natesan, S. K., & Field, M. C. (2014). Life and times: synthesis, trafficking, and evolution of VSG. *Trends Parasitol*, 30(5), 251-258. <http://dx.doi.org/10.1016/j.pt.2014.03.004>
- Mansfield, J. M., Paulnock, D. M., & Hedberg, G. M. (2014). Bridging Innate and Adaptive Immunity in African Trypanosomiasis *Trypanosomes and Trypanosomiasis* (pp. 89-114, DOI 10.1007/978-1003-7091-1556-1005\_1004). Verlag Wien 2014: Springer. [http://dx.doi.org/DOI 10.1007/978-3-7091-1556-5\\_4](http://dx.doi.org/DOI 10.1007/978-3-7091-1556-5_4)
- Mara, L., Casu, S., Carta, A., & Dattena, M. (2013). Cryobanking of farm animal gametes and embryos as a means of conserving livestock genetics. *Animal Reproduction Science*, 138(1), 25-38.
- Marletta, D., Tupac-Yupanqui, I., Bordonaro, S., Garcia, D., Guastella, A., Criscione, A., . . . Dunner, S. (2006). Analysis of genetic diversity and the determination of relationships among western Mediterranean horse breeds using microsatellite markers. *Journal of Animal Breeding and Genetics*, 123(5), 315-325.

- Marmesat, E., Schmidt, K., Saveljev, A. P., Seryodkin, I. V., & Godoy, J. A. (2017). Retention of functional variation despite extreme genomic erosion: MHC allelic repertoires in the Lynx genus. *BMC Evolutionary Biology*, *17*(1), 158. <http://dx.doi.org/10.1186/s12862-017-1006-z>
- Marshall, K., Mugambi, J. M., Nagda, S., Sonstegard, T. S., Van Tassell, C. P., Baker, R. L., & Gibson, J. P. (2013). Quantitative trait loci for resistance to *Haemonchus contortus* artificial challenge in Red Maasai and Dorper sheep of East Africa. *Animal Genetics*, *44*(3), 285-295. <http://dx.doi.org/10.1111/j.1365-2052.2012.02401.x>
- Marshall, K., Quiros-Campos, C., van der Werf, J. H. J., & Kinghorn, B. (2011). Marker-based selection within smallholder production systems in developing countries. *Livestock Science*, *136*(1), 45-54. <http://dx.doi.org/http://dx.doi.org/10.1016/j.livsci.2010.09.006>
- Martín-Burriel, I., García-Muro, E., & Zaragoza, P. (1999). Genetic diversity analysis of six Spanish native cattle breeds using microsatellites. *Animal Genetics*, *30*(3), 177-182. <http://dx.doi.org/10.1046/j.1365-2052.1999.00437.x>
- Matovu, E., Seebeck, T., Enyaru, J. C., & Kaminsky, R. (2001). Drug resistance in *Trypanosoma brucei* spp., the causative agents of sleeping sickness in man and nagana in cattle. *Microbes Infect*, *3*(9), 763-770. [http://dx.doi.org/http://dx.doi.org/10.1016/S1286-4579\(01\)01432-0](http://dx.doi.org/http://dx.doi.org/10.1016/S1286-4579(01)01432-0)
- Matthews, K. R., McCulloch, R., & Morrison, L. J. (2015). The within-host dynamics of African trypanosome infections. *Philos Trans R Soc Lond B Biol Sci*, *370*(1675) <http://dx.doi.org/10.1098/rstb.2014.0288>
- Mattioli, R. C., Pandey, V. S., Murray, M., & Fitzpatrick, J. L. (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 Trop*, *75*(3), 263-277. [http://dx.doi.org/http://dx.doi.org/10.1016/S0001-706X\(00\)00063-2](http://dx.doi.org/http://dx.doi.org/10.1016/S0001-706X(00)00063-2)
- Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., Taylor, J. F., Allan, M. F., Heaton, M. P., . . . Sonstegard, T. S. (2009). Development and characterization of a high density SNP genotyping assay for cattle. *PLoS one*, *4*(4), e5350.
- Maudlin, I. (2006). African trypanosomiasis. *Ann Trop Med Parasitol*, *100*(8), 679-701. <http://dx.doi.org/10.1179/136485906X112211>
- McCorkle, C. M., & James, V. (1996). The roles of animals in cultural, social, and agro-economic systems. In V. U. James (Ed.), *Sustainable development in Third World countries: Applied and Theoretical Perspectives*. (ISBN 0275953076 ed., pp. pp 25-43). Westport, Connecticut, USA.: Praeger Publishers, Westport, Connecticut, USA.
- McCulloch, R., & Field, M. C. (2015). Quantitative sequencing confirms VSG diversity as central to immune evasion by *Trypanosoma brucei*. *Trends Parasitol*, *31*(8), 346-349. <http://dx.doi.org/10.1016/j.pt.2015.05.001>
- McCulloch, R., Rudenko, G., & Borst, P. (1997). Gene conversions mediating antigenic variation in *Trypanosoma brucei* can occur in variant surface glycoprotein expression sites lacking 70-base-pair repeat sequences. *Mol Cell Biol*, *17*(2), 833-843. Retrieved from <Go to ISI>://WOS:A1997WC87900033
- McMahon, B. J., Teeling, E. C., & Höglund, J. (2014). How and why should we implement genomics into conservation? *Evolutionary Applications*, *7*(9), 999-1007.
- McMichael, A. J., Powles, J. W., Butler, C. D., & Uauy, R. (2007). Food, livestock production, energy, climate change, and health. *The Lancet*, *370*(9594), 1253-1263. [http://dx.doi.org/http://dx.doi.org/10.1016/S0140-6736\(07\)61256-2](http://dx.doi.org/http://dx.doi.org/10.1016/S0140-6736(07)61256-2)
- McRae, K. M., McEwan, J. C., Dodds, K. G., & Gemmell, N. J. (2014). Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. *BMC Genomics*, *15*(1), 1-13. <http://dx.doi.org/10.1186/1471-2164-15-637>
- McTavish, E. J., & Hillis, D. M. (2015). How do SNP ascertainment schemes and population demographics affect inferences about population history? *BMC Genomics*, *16*(1), 266.
- Medugorac, I., Veit-Kensch, C. E., Ramljak, J., Brka, M., Marković, B., Stojanović, S., . . . Grünfelder, H. P. (2011). Conservation priorities of genetic diversity in domesticated metapopulations: a study in taurine cattle breeds. *Ecology and Evolution*, *1*(3), 408-420.
- Mendelsohn, R. (2003). The challenge of conserving indigenous domesticated animals. *Ecological Economics*, *45*(3), 501-510.

- Meuwissen, T., Hayes, B., & Goddard, M. (2016). Genomic selection: A paradigm shift in animal breeding. *Anim Front*, 6(1), DOI: 10.2527/af.2016-0002. <http://dx.doi.org/10.2527/af.2016-0002>
- Meyer-Lucht, Y., & Sommer, S. (2005). MHC diversity and the association to nematode parasitism in the yellow-necked mouse (*Apodemus flavicollis*). *Molecular Ecology*, 14(7), 2233-2243. <http://dx.doi.org/10.1111/j.1365-294X.2005.02557.x>
- Michel, A. L., & Bengis, R. G. (2012). The African buffalo: a villain for inter-species spread of infectious diseases in southern Africa. *Onderstepoort J Vet Res*, 79(2), 453. <http://dx.doi.org/10.4102/ojvr.v79i2.453>
- Mignon-Grasteau, S., Boissy, A., Bouix, J., Faure, J.-M., Fisher, A. D., Hinch, G. N., . . . Beaumont, C. (2005). Genetics of adaptation and domestication in livestock. *Livestock Production Science*, 93(1), 3-14. <http://dx.doi.org/http://dx.doi.org/10.1016/j.livprodsci.2004.11.001>
- Mikko, S., Lewin, H. A., & Andersson, L. (1997). A phylogenetic analysis of cattle DRB3 alleles with a deletion of codon 65. *Immunogenetics*, 47(1), 23-29. <http://dx.doi.org/10.1007/s002510050322>
- Millot, P. (1978). The major histocompatibility complex of sheep (OLA) and two minor loci. *Animal Blood Groups and Biochemical Genetics*, 9(2), 115-121. <http://dx.doi.org/10.1111/j.1365-2052.1978.tb01421.x>
- Mony, B. M., & Matthews, K. R. (2015). Assembling the components of the quorum sensing pathway in African trypanosomes. *Mol Microbiol*, 96(2), 220-232. <http://dx.doi.org/10.1111/mmi.12949>
- Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R., & Schäffer, A. A. (2008). Database indexing for production MegaBLAST searches. *Bioinformatics*, 24(16), 1757-1764. <http://dx.doi.org/10.1093/bioinformatics/btn322>
- Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, 19(4), 208-216.
- Morris, C. A. (2007). A review of genetic resistance to disease in *Bos taurus* cattle. *Vet J*, 174(3), 481-491. <http://dx.doi.org/10.1016/j.tvjl.2006.09.006>
- Morris, J. P. (2006). *The ethics of biotechnology*. Philadelphia: Chelsea House Publishers.
- Muigai, A. W. T., & Hanotte, O. (2013). The Origin of African Sheep: Archaeological and Genetic Perspectives. *Afr Archaeol Rev*, 30(1), 39-50. <http://dx.doi.org/10.1007/s10437-013-9129-0>
- Mukesh, M., Sodhi, M., Bhatia, S., & Mishra, B. P. (2004). Genetic diversity of Indian native cattle breeds as analysed with 20 microsatellite loci. *Journal of Animal Breeding and Genetics*, 121(6), 416-424. <http://dx.doi.org/10.1111/j.1439-0388.2004.00468.x>
- Munday, J. C., Settimo, L., & de Koning, H. P. (2015). Transport proteins determine drug sensitivity and resistance in a protozoan parasite, *Trypanosoma brucei*. *Front Pharmacol*, 6, 32. <http://dx.doi.org/10.3389/fphar.2015.00032>
- Munro, R., & Adams, D. (1991). Genetic diversity and sustainable agriculture—Implications for animal production systems. *Buffalo and Goats in Asia: Genetic Diversity and its Application*, Tulloh, NM (Ed.). Australian Centre for International Agricultural Research (ACIAR), Australia, ISBN-13: 9781863200349, pp 123-128.
- Murray, M., & Black, S. J. (1985). African trypanosomiasis in cattle: working with nature's solution. *Vet Parasitol*, 18(2), 167-182. [http://dx.doi.org/http://dx.doi.org/10.1016/0304-4017\(85\)90065-2](http://dx.doi.org/http://dx.doi.org/10.1016/0304-4017(85)90065-2)
- Murray, M., Morrison, W. I., & Whitelaw, D. D. (1982). Host susceptibility to African trypanosomiasis: trypanotolerance. *Adv Parasitol*, 21, 1-68. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/6762064>
- Murray, M., Trail, J. C., Davis, C. E., & Black, S. J. (1984). Genetic resistance to African Trypanosomiasis. *J Infect Dis*, 149(3), 311-319. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/6371153>
- Murray, M., & Trail, J. C. M. (1984). Genetic resistance to animal trypanosomiasis in Africa. *Prev Vet Med*, 2(1-4), 541-551. [http://dx.doi.org/http://dx.doi.org/10.1016/0167-5877\(84\)90100-4](http://dx.doi.org/http://dx.doi.org/10.1016/0167-5877(84)90100-4)
- Mwacharo, J. M., Elbeltagy, A. R., Kim, E. S., Haile, A., Rischkowsky, B., & Rothschild, M. F. (2016). S0124 Indigenous stocks as treasure troves for sustainable livestock production in the 21st

- century: Insights from small ruminant genomics. *Journal of Animal Science*, 94(7supplement4), 12-13. <http://dx.doi.org/10.2134/jas2016.947supplement412a>
- Mwai, O., Hanotte, O., Kwon, Y. J., & Cho, S. (2015). African Indigenous Cattle: Unique Genetic Resources in a Rapidly Changing World. *Asian-Australas J Anim Sci*, 28(7), 911-921. <http://dx.doi.org/10.5713/ajas.15.0002R>
- Naessens, J. (2006). Bovine trypanotolerance: A natural ability to prevent severe anaemia and haemophagocytic syndrome? *Int J Parasitol*, 36(5), 521-528. <http://dx.doi.org/10.1016/j.ijpara.2006.02.012>
- Naessens, J., Teale, A. J., & Sileghem, M. (2002). Identification of mechanisms of natural resistance to African trypanosomiasis in cattle. *Vet Immunol Immunopathol*, 87(3-4), 187-194. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/12072233>
- Nagaoka, Y., Kabeya, H., Onuma, M., Kasai, N., Okada, K., & Aida, Y. (1999). Ovine MHC Class II & DRB1 Alleles Associated with Resistance or Susceptibility to Development of Bovine Leukemia Virus-induced Ovine Lymphoma. *Cancer Research*, 59(4), 975. Retrieved from <http://cancerres.aacrjournals.org/content/59/4/975.abstract>
- Nakayima, J., Nakao, R., Alhassan, A., Mahama, C., Afakye, K., & Sugimoto, C. (2012). Molecular epidemiological studies on animal trypanosomiasis in Ghana. *Parasit Vectors*, 5, 217. <http://dx.doi.org/10.1186/1756-3305-5-217>
- Namangala, B. (2012). Contribution of innate immune responses towards resistance to African trypanosome infections. *Scand J Immunol*, 75(1), 5-15. <http://dx.doi.org/10.1111/j.1365-3083.2011.02619.x>
- Nardone, A., Ronchi, B., Lacetera, N., Ranieri, M. S., & Bernabucci, U. (2010). Effects of climate changes on animal production and sustainability of livestock systems. *Livestock Science*, 130(1-3), 57-69. <http://dx.doi.org/http://dx.doi.org/10.1016/j.livsci.2010.02.011>
- Naskar, S., Gowane, G., & Chopra, A. (2015). Strategies to improve livestock genetic resources to counter climate change impact. In V. Sejian, J. Gaughan, L. Baumgard & C. Prasad (Eds.), *Climate Change Impact on Livestock: Adaptation and Mitigation* (pp. 441-475): Springer India. [http://dx.doi.org/10.1007/978-81-322-2265-1\\_25](http://dx.doi.org/10.1007/978-81-322-2265-1_25)
- Nei, M., Maruyama, T., & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, 29, 1-10.
- Newman, D., & Pilson, D. (1997). Increased probability of extinction due to decreased genetic effective population size: Experimental populations of *Clarkia pulchella*. *Evolution*, 51(2), 354-362. <http://dx.doi.org/10.2307/2411107>
- Notter, D., Mariante, A. d. S., & Sheng, Z. (1994). *Modern approaches to active conservation of domestic animal diversity*. Paper presented at the Proc. 5th World Congress on Genetic Applied to Livestock Production. Dept. Animal and Poultry Science, University of Guelph. Ontario, Canada
- Noyes, H., Brass, A., Obara, I., Anderson, S., Archibald, A. L., Bradley, D. G., . . . Kemp, S. J. (2011). Genetic and expression analysis of cattle identifies candidate genes in pathways responding to *Trypanosoma congolense* infection. *Proc Natl Acad Sci U S A*, 108(22), 9304-9309. <http://dx.doi.org/10.1073/pnas.1013486108>
- Nyimba, P. H., Komba, E. V., Sugimoto, C., & Namangala, B. (2015). Prevalence and species distribution of caprine trypanosomosis in Sinazongwe and Kalomo districts of Zambia. *Vet Parasitol*, 210(3-4), 125-130. <http://dx.doi.org/10.1016/j.vetpar.2015.04.003>
- O'Brien, J. K., Steinman, K. J., & Robeck, T. R. (2009). Application of sperm sorting and associated reproductive technology for wildlife management and conservation. *Theriogenology*, 71(1), 98-107. <http://dx.doi.org/10.1016/j.theriogenology.2008.09.052>
- O'Gorman, G. M., Park, S. D., Hill, E. W., Meade, K. G., Coussens, P. M., Agaba, M., . . . MacHugh, D. E. (2009). Transcriptional profiling of cattle infected with *Trypanosoma congolense* highlights gene expression signatures underlying trypanotolerance and trypanosusceptibility. *BMC Genomics*, 10(1), 207. <http://dx.doi.org/10.1186/1471-2164-10-207>
- Ogunremi, O., & Tabel, H. (1995). Genetics of resistance to *Trypanosoma congolense* in inbred mice: efficiency of apparent clearance of parasites correlates with long-term survival. *J Parasitol*, 81(6), 876-881. <http://dx.doi.org/10.2307/3284033>

- Okaiyeto, S. O., Ajanusi, O. J., Sackey, A. K., & Tekdek, L. B. (2010). Changes in haematological Values Associated with Mixed Trypanosoma congolense and Haemonchus contortus infection in Yankassa Sheep. *Veterinary Research*, 3, 9-13.
- Olson, H., Betton, G., Robinson, D., Thomas, K., Monro, A., Kolaja, G., . . . Bracken, W. (2000). Concordance of the toxicity of pharmaceuticals in humans and in animals. *Regulatory Toxicology and Pharmacology*, 32(1), 56-67.
- Orenge, C. O., Munga, L., Kimwele, C. N., Kemp, S., Korol, A., Gibson, J. P., . . . Soller, M. (2012). Trypanotolerance in N'Dama x Boran crosses under natural trypanosome challenge: effect of test-year environment, gender, and breed composition. *BMC Genet*, 13(1), 87. <http://dx.doi.org/10.1186/1471-2156-13-87>
- Osaer, S., Goossens, B., Clifford, D. J., Kora, S., & Kassama, M. (1994). A comparison of the susceptibility of djallonke sheep and west-african dwarf goats to experimental-infection with 2 different strains of trypanosoma-congolense. *Vet Parasitol*, 51(3-4), 191-204.
- Oyelaran-Oyeyinka, B., & Lal, K. (2005). Internet diffusion in sub-Saharan Africa: A cross-country analysis. *Telecommunications Policy*, 29(7), 507-527. <http://dx.doi.org/http://dx.doi.org/10.1016/j.telpol.2005.05.002>
- Paiva, S., Faria, D., Silvério, V., McManus, C., Egito, A., Dergam, J., . . . Mariante, A. (2005). Genetic variability among Brazilian sheep using microsatellites. *Proc. The Role of Biotechnology, Turin, March*, 5-7.
- Parham, P., Adams, E. J., & Arnett, K. L. (1995). The Origins of HLA-A,B,C Polymorphism. *Immunological Reviews*, 143(1), 141-180. <http://dx.doi.org/10.1111/j.1600-065X.1995.tb00674.x>
- Parham, P., & Ohta, T. (1996). Population biology of antigen presentation by MHC class I molecules. *SCIENCE-NEW YORK THEN WASHINGTON-*, 67-73.
- Parsons, Y. M., & Shaw, K. (2001). Species boundaries and genetic diversity among Hawaiian crickets of the genus *Laupala* identified using amplified fragment length polymorphism. *Molecular Ecology*, 10(7), 1765-1772.
- Pasaniuc, B., Rohland, N., McLaren, P. J., Garimella, K., Zaitlen, N., Li, H., . . . Price, A. L. (2012). Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nat Genet*, 44(6), 631-635. <http://dx.doi.org/http://www.nature.com/ng/journal/v44/n6/abs/ng.2283.html#supplementary-information>
- Paterson, S., Wilson, K., & Pemberton, J. M. (1998). Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries* L.). *Proceedings of the National Academy of Sciences*, 95 <http://dx.doi.org/10.1073/pnas.95.7.3714>
- Pépin, L., & Nguyen, T. C. (1994). Blood groups and protein polymorphisms in five goat breeds (*Capra hircus*). *Animal Genetics*, 25(5), 333-336. <http://dx.doi.org/10.1111/j.1365-2052.1994.tb00367.x>
- Peregrine, A. S. (1994). Chemotherapy and delivery systems: haemoparasites. *Vet Parasitol*, 54(1-3), 223-248. [http://dx.doi.org/http://dx.doi.org/10.1016/0304-4017\(94\)90092-2](http://dx.doi.org/http://dx.doi.org/10.1016/0304-4017(94)90092-2)
- Peregrine, A. S., & Mamman, M. (1993). Pharmacology of diminazene: a review. *Acta Trop*, 54(3-4), 185-203. [http://dx.doi.org/http://dx.doi.org/10.1016/0001-706X\(93\)90092-P](http://dx.doi.org/http://dx.doi.org/10.1016/0001-706X(93)90092-P)
- Periasamy, K., Pichler, R., Poli, M., Cristel, S., Cetrá, B., Medus, D., . . . Diallo, A. (2014). Candidate Gene Approach for Parasite Resistance in Sheep – Variation in Immune Pathway Genes and Association with Fecal Egg Count. *PLoS ONE*, 9(2), e88337. <http://dx.doi.org/10.1371/journal.pone.0088337>
- Philipsson, J., Rege, J., & Okeyo, A. (2006). Sustainable breeding programmes for tropical farming systems. In J. M. Ojango, Malmfors, B. and Okeyo, A.M. (Ed.), *Animal genetics training resource, version 2, training module 3*. Nairobi, Kenya. <https://cgspace.cgiar.org/handle/10568/3662>: International Livestock Research Institute.
- Philipsson, J., Zonabend, E., Bett, R., & Okeyo, A. (2011). Swedish university of agricultural sciences (SLU), department of animal breeding and genetics, PO Box 7023, S-750 07 Uppsala, Sweden. 2 International Livestock Research Institute (ILRI), PO Box 30709, Nairobi 00100, Kenya.

- Piedrafita, D., Raadsma, H. W., Gonzalez, J., & Meeusen, E. (2010). Increased production through parasite control: can ancient breeds of sheep teach us new lessons? *Trends Parasitol*, 26(12), 568-573. <http://dx.doi.org/http://dx.doi.org/10.1016/j.pt.2010.08.002>
- Pintado, B., & Hourcade, J. (2011). Cryopreservation *Advanced protocols for animal transgenesis* (pp. 577-599). Berlin Heidelberg: Springer.
- Plachy, J., Pink, J. R., & Hála, K. (1992). Biology of the chicken MHC (B complex). *Critical reviews in immunology*, 12(1-2), 47-79. Retrieved from <http://europepmc.org/abstract/MED/1358107>
- Povolotskaya, I. S., Kondrashov, F. A., Ledda, A., & Vlasov, P. K. (2012). Stop codons in bacteria are not selectively equivalent. *Biology Direct*, 7, 30-30. <http://dx.doi.org/10.1186/1745-6150-7-30>
- Prentice, J. R., & Anzar, M. (2011). Cryopreservation of mammalian oocyte for conservation of animal genetics. *Strategies*, 3, 8. <http://dx.doi.org/10.4061/2011/146405>
- Preston, J. M., & Allonby, E. W. (1979). The influence of breed on the susceptibility of sheep to *Haemonchus contortus* infection in Kenya. *Res. Vet. Sci*, 26, 134-139.
- Price, E. O. (1999). Behavioral development in animals undergoing domestication. *Applied Animal Behaviour Science*, 65(3), 245-271. [http://dx.doi.org/http://dx.doi.org/10.1016/S0168-1591\(99\)00087-8](http://dx.doi.org/http://dx.doi.org/10.1016/S0168-1591(99)00087-8)
- Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T. M., Fries, R., . . . Simianer, H. (2014). Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet*, 10(2), e1004148. <http://dx.doi.org/10.1371/journal.pgen.1004148>
- Qanbari, S., & Simianer, H. (2014). Mapping signatures of positive selection in the genome of livestock. *Livestock Science*, 166, 133-143. <http://dx.doi.org/http://dx.doi.org/10.1016/j.livsci.2014.05.003>
- Qin, J., Mamotte, C., Cockett, N. E., Wetherall, J. D., & Groth, D. M. (2008). A map of the class III region of the sheep major histocompatibility complex. *BMC Genomics*, 9 <http://dx.doi.org/10.1186/1471-2164-9-409>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841-842. <http://dx.doi.org/10.1093/bioinformatics/btq033>
- Raadsma, H., & Fullard, K. (2006, 13-18 August, 2006). *QTL mapping and gene markers for resistance to infectious diseases in sheep and cattle*. Paper presented at the Proceedings of the 8th World Congress on Genetics Applied to Livestock Production. CDRom Communication No. 0-1023-1760, Belo Horizonte, Minas Gerais, Brazil
- Raadsma, H. W., & Egerton, J. R. (2013). A review of footrot in sheep: Aetiology, risk factors and control methods. *Livestock Science*, 156(1), 106-114. <http://dx.doi.org/http://dx.doi.org/10.1016/j.livsci.2013.06.009>
- Radwanska, M., Guirnalda, P., De Trez, C., Ryffel, B., Black, S., & Magez, S. (2008). Trypanosomiasis-induced B cell apoptosis results in loss of protective anti-parasite antibody responses and abolishment of vaccine-induced memory responses. *PLoS Pathog*, 4(5), e1000078. <http://dx.doi.org/10.1371/journal.ppat.1000078>
- Ramstad, K. M., Woody, C. A., Sage, G. K., & Allendorf, F. W. (2004). Founding events influence genetic population structure of sockeye salmon (*Oncorhynchus nerka*) in Lake Clark, Alaska. *Molecular Ecology*, 13(2), 277-290. <http://dx.doi.org/10.1046/j.1365-294X.2003.2062.x>
- Rege, J. E. O. (2003). *Defining livestock breeds in the context of community-based management of farm animal genetic resources*. Paper presented at the Community-based management of animal genetic resources, Proceedings of workshop held in Mbabane, Swaziland, 7-11 May, 2011
- Rege, J. E. O., & Gibson, J. P. (2003). Animal genetic resources and economic development: issues in relation to economic valuation. *Ecological Economics*, 45(3), 319-330. [http://dx.doi.org/10.1016/S0921-8009\(03\)00087-9](http://dx.doi.org/10.1016/S0921-8009(03)00087-9)
- Rege, J. E. O., Marshall, K., Notenbaert, A., Ojango, J. M. K., & Okeyo, A. M. (2011). Pro-poor animal improvement and breeding — What can science do? *Livestock Science*, 136(1), 15-28. <http://dx.doi.org/http://dx.doi.org/10.1016/j.livsci.2010.09.003>
- Reverter, A., Johnston, D., Ferguson, D., Perry, D., Goddard, M., Burrow, H., . . . Bindon, B. (2003). Genetic and phenotypic characterisation of animal, carcass, and meat quality traits from

- temperate and tropically adapted beef breeds. 4. Correlations among animal, carcass, and meat quality traits. *Crop and Pasture Science*, 54(2), 149-158.
- Richman, A. D., Herrera, L. G., & Nash, D. (2001). MHC class II beta sequence diversity in the deer mouse (*Peromyscus maniculatus*): implications for models of balancing selection. *Molecular Ecology*, 10(12), 2765-2773. <http://dx.doi.org/10.1046/j.0962-1083.2001.01402.x>
- Rischkowsky, B., Pilling, D., & Commission on Genetic Resources for Food Agriculture. (2007). *The state of the world's animal genetic resources for food and agriculture*. Rome, Italy: Commission on Genetic Resources for Food and Agriculture, Food and Agriculture Organization of the United Nations.
- Rivas, M. A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C. K., . . . Daly, M. J. (2011). Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet*, 43(11), 1066-1073. <http://dx.doi.org/http://www.nature.com/ng/journal/v43/n11/abs/ng.952.html#supplementary-information>
- Robinson, J., Halliwell, J. A., Hayhurst, J. D., Flicek, P., Parham, P., & Marsh, Steven G E. (2015). The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Research*, 43(Database issue), D423-D431. <http://dx.doi.org/10.1093/nar/gku1161>
- Robinson, J., Waller, M. J., Parham, P., Groot, N. d., Bontrop, R., Kennedy, L. J., . . . Marsh, S. G. E. (2003). IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Research*, 31(1), 311-314. <http://dx.doi.org/10.1093/nar/gkg070>
- Roeber, F., Jex, A. R., & Gasser, R. B. (2013). Impact of gastrointestinal parasitic nematodes of sheep, and the role of advanced molecular tools for exploring epidemiology and drug resistance - an Australian perspective. *Parasites & vectors*, 6, 153. <http://dx.doi.org/10.1186/1756-3305-6-153>
- Roffler, G. H., Amish, S. J., Smith, S., Cosart, T., Kardos, M., Schwartz, M. K., & Luikart, G. (2016). SNP discovery in candidate adaptive genes using exon capture in a free-ranging alpine ungulate. *Molecular Ecology Resources*, 16(5), 1147-1164. <http://dx.doi.org/10.1111/1755-0998.12560>
- Rohrer, G. A., Alexander, L. J., Keele, J. W., Smith, T. P., & Beattie, C. W. (1994). A microsatellite linkage map of the porcine genome. *Genetics*, 136(1), 231-245.
- Rohrer, G. A., & Keele, J. (1998). Identification of quantitative trait loci affecting carcass composition in swine: I. Fat deposition traits. *Journal of Animal Science*, 76(9), 2247-2254.
- Rolf, M. M., Decker, J. E., McKay, S. D., Tizito, P. C., Branham, K. A., Whitacre, L. K., . . . Taylor, J. F. (2014). Genomics in the United States beef industry. *Livestock Science*, 166, 84-93. <http://dx.doi.org/http://dx.doi.org/10.1016/j.livsci.2014.06.005>
- Rosenberg, L., Cooperman, D., & Payne, R. (1983). HLA and mate selection. *Immunogenetics*, 17(1), 89.
- Rousset, F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, 8(1), 103-106. <http://dx.doi.org/10.1111/j.1471-8286.2007.01931.x>
- Ruane, J., & Sonnino, A. (2011). Agricultural biotechnologies in developing countries and their possible contribution to food security. *Journal of Biotechnology*, 156(4), 356-363.
- Rubin, C. J., Zody, M. C., Eriksson, J., Meadows, J. R., Sherwood, E., Webster, M. T., . . . Andersson, L. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, 464(7288), 587-591. <http://dx.doi.org/10.1038/nature08832>
- Russo, V., Martelli, M., Mauro, A., Di Giacinto, O., Nardinocchi, D., & Berardinelli, P. (2007). Nuclear remodelling in growing oocytes of sheep. *Veterinary Research Communications*, 31, 201-204.
- Sachs, M. M. (2009). Cereal germplasm resources. *Plant Physiology*, 149(1), 148-151.
- Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406-425. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a040454>
- Sanni, M. T., Gbolabo, O. O., Mufliat, A. A., Abdulmojeed, Y., Christian, O. N. I., Olufunmilayo, A. A., . . . Ikhide, G. I. (2013). Molecular Diagnosis of Subclinical African Trypanosoma vivax Infection and Association with Physiological Indices and Serum Metabolites in Extensively

- Managed Goats in the Tropics. *Open J Vet Med*, 03, 39-39. <http://dx.doi.org/10.4236/ojvm.2013.31007>
- Santos, R., Affonso, P., Diniz, D., Medrado, A., Silva, K., & Carneiro, P. (2013). Identification of naturalized goat breeds under conservation from northeastern Brazil using chromosomal markers. *Genetics and Molecular Research*, 12(3), 3193-3200.
- Sasazaki, S., Mutoh, H., Tsurifune, K., & Mannen, H. (2007). Development of DNA markers for discrimination between domestic and imported beef. *Meat Science*, 77(2), 161-166. <http://dx.doi.org/http://dx.doi.org/10.1016/j.meatsci.2007.02.024>
- Savas, S., Tuzmen, S., & Ozcelik, H. (2006). Human SNPs resulting in premature stop codons and protein truncation. *Human Genomics*, 2(5), 274. <http://dx.doi.org/10.1186/1479-7364-2-5-274>
- Savelkoul, P., Aarts, H., De Haas, J., Dijkshoorn, L., Duim, B., Otsen, M., . . . Lenstra, J. (1999). Amplified-fragment length polymorphism analysis: the state of an art. *Journal of Clinical Microbiology*, 37(10), 3083-3091.
- Sayers, E. W., Barrett, T., Benson, D. A., Bolton, E., Bryant, S. H., Canese, K., . . . Federhen, S. (2011). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 39(suppl 1), D38-D51.
- Sayers, G., Good, B., Hanrahan, J. P., Ryan, M., Angles, J. M., & Sweeney, T. (2005). Major Histocompatibility Complex DRB1 gene: its role in nematode resistance in Suffolk and Texel sheep breeds. *Parasitology*, 131(03), 403-409. <http://dx.doi.org/doi:10.1017/S0031182005007778>
- Schierup, M. H., Mikkelsen, A. M., & Hein, J. (2001). Recombination, balancing selection and phylogenies in MHC and self-incompatibility genes. *Genetics*, 159(4), 1833-1844.
- Schlötterer, C., Kofler, R., Versace, E., Tobler, R., & Franssen, S. (2014). Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity*, 114, 431-440.
- Schlotterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. *Nat Rev Genet*, 15(11), 749-763. <http://dx.doi.org/10.1038/nrg3803>
- Schwaiger, F. W., Gostomski, D., Stear, M. J., Duncan, J. L., McKellar, Q. A., Epplen, J. T., & Buitkamp, J. (1995). An ovine Major histocompatibility complex DRB1 allele is associated with low faecal egg counts following natural, predominantly *Ostertagia circumcincta* infection. *International Journal for Parasitology*, 25(7), 815-822. [http://dx.doi.org/http://dx.doi.org/10.1016/0020-7519\(94\)00216-B](http://dx.doi.org/http://dx.doi.org/10.1016/0020-7519(94)00216-B)
- Schwaiger, F. W., Maddox, J. F., Ballingall, K. T., & Buitkamp, J. (1996). The ovine major histocompatibility complex. In L. B. Schook & S. J. Lamont (Eds.), *The major histocompatibility complex region of domestic animal species*. Boca Raton, FL, USA: CRC Press, Inc.
- Schwartz, M. K., & McKelvey, K. S. (2009). Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, 10(2), 441-452.
- Seck, M. T., Bouyer, J., Sall, B., Bengaly, Z., & Vreysen, M. J. B. (2010). The prevalence of African animal trypanosomoses and tsetse presence in Western Senegal. *Parasite*, 17(3), 257-265. Retrieved from <http://dx.doi.org/10.1051/parasite/2010173257>
- Shand, H. (1997). *Human nature : Agricultural biodiversity and farm based food security*. Ottawa : Rural Advancement Foundation International, ISBN-10: 0968311202: RAFI.
- Shaw, A. P. M. (2009). Assessing the economics of animal trypanosomosis in Africa - History and current perspectives. *Onderstepoort J Vet Res*, Volume 76(1), 27-32.
- Shields, S., & Orme-Evans, G. (2015). The impacts of climate change mitigation strategies on animal welfare. *Animals*, 5(2), 361-394.
- Shiomi, M., Ito, T., Yamada, S., Kawashima, S., & Fan, J. (2003). Development of an animal model for spontaneous myocardial infarction (WHHLMI rabbit). *Arteriosclerosis, Thrombosis, and Vascular Biology*, 23(7), 1239-1244.
- Signorello, G., & Pappalardo, G. (2003). Domestic animal biodiversity conservation: a case study of rural development plans in the European Union. *Ecological Economics*, 45(3), 487-499. [http://dx.doi.org/http://dx.doi.org/10.1016/S0921-8009\(03\)00099-5](http://dx.doi.org/http://dx.doi.org/10.1016/S0921-8009(03)00099-5)

- Sima, M., Havelkova, H., Quan, L., Svobodova, M., Jarosikova, T., Vojtiskova, J., . . . Lipoldova, M. (2011). Genetic control of resistance to *Trypanosoma brucei brucei* infection in mice. *PLoS Negl Trop Dis*, 5(6), e1173. <http://dx.doi.org/10.1371/journal.pntd.0001173>
- Simon, D. (1999). European approaches to conservation of farm animal genetic resources. *Animal Genetic Resources Information*, 25, 77-97.
- Siva Subramaniam, N., Morgan, E., Wetherall, J., Stear, M., & Groth, D. (2015). A comprehensive mapping of the structure and gene organisation in the sheep MHC class I region. *BMC Genomics*, 16(1), 810. <http://dx.doi.org/10.1186/s12864-015-1992-4>
- Smetko, A., Soudre, A., Silbermayr, K., Muller, S., Brem, G., Hanotte, O., . . . Solkner, J. (2015). Trypanosomosis: potential driver of selection in African cattle. *Front Genet*, 6, 137. <http://dx.doi.org/10.3389/fgene.2015.00137>
- Smith, C. M., Finger, J. H., Hayamizu, T. F., McCright, I. J., Xu, J., Berghout, J., . . . Ringwald, M. (2014). The mouse Gene Expression Database (GXD): 2014 update. *Nucleic Acids Research*, 42(D1), D818-D824. <http://dx.doi.org/10.1093/nar/gkt954>
- Smith, J., Taylor, E. M., & Kingsley, P. (2015). One World-One Health and neglected zoonotic disease: elimination, emergence and emergency in Uganda. *Soc Sci Med*, 129(0), 12-19. <http://dx.doi.org/10.1016/j.socscimed.2014.06.044>
- Spangler, G. L., Rosen, B. D., Ilori, M. B., Hanotte, O., Kim, E.-S., Sonstegard, T. S., . . . Van Tassell, C. P. (2017). Whole genome structural analysis of Caribbean hair sheep reveals quantitative link to West African ancestry. *PLOS ONE*, 12(6), e0179021. <http://dx.doi.org/10.1371/journal.pone.0179021>
- Stear, M. J., Bairden, K., Bishop, S. C., Buitkamp, J., Duncan, J. L., Gettinby, G., . . . Murray, M. (1997). The genetic basis of resistance to *Ostertagia circumcincta* in lambs. *Vet J*, 154(2), 111-119. [http://dx.doi.org/10.1016/s1090-0233\(97\)80049-4](http://dx.doi.org/10.1016/s1090-0233(97)80049-4)
- Steverding, D. (2008). The history of African trypanosomiasis. *Parasit Vectors*, 1(3), 411.
- Steverding, D. (2015). Evaluation of trypanocidal activity of combinations of anti-sleeping sickness drugs with cysteine protease inhibitors. *Exp Parasitol*, 151-152(0), 28-33. <http://dx.doi.org/http://dx.doi.org/10.1016/j.exppara.2015.01.016>
- Stijlemans, B., Cnops, J., Naniima, P., Vaast, A., Bockstal, V., De Baetselier, P., & Magez, S. (2015). Development of a pHrodo-based assay for the assessment of in vitro and in vivo erythrophagocytosis during experimental trypanosomosis. *PLoS Negl Trop Dis*, 9(3), e0003561  
DOI: 0003510.0001371/journal.pntd.0003561. <http://dx.doi.org/10.1371/journal.pntd.0003561>
- Stijlemans, B., De Baetselier, P., Caljon, G., Van Den Abbeele, J., Van Ginderachter, J. A., & Magez, S. (2017). Nanobodies As Tools to Understand, Diagnose, and Treat African Trypanosomiasis. *Frontiers in Immunology*, 8(724) <http://dx.doi.org/10.3389/fimmu.2017.00724>
- Stijlemans, B., Radwanska, M., De Trez, C., & Magez, S. (2017). African Trypanosomes Undermine Humoral Responses and Vaccine Development: Link with Inflammatory Responses? *Frontiers in Immunology*, 8(582) <http://dx.doi.org/10.3389/fimmu.2017.00582>
- Suekawa, Y., Aihara, H., Araki, M., Hosokawa, D., Mannen, H., & Sasazaki, S. (2010). Development of breed identification markers based on a bovine 50K SNP array. *Meat Science*, 85(2), 285-288. <http://dx.doi.org/http://dx.doi.org/10.1016/j.meatsci.2010.01.015>
- Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology & Evolution*, 15(5), 199-203. [http://dx.doi.org/http://dx.doi.org/10.1016/S0169-5347\(00\)01825-5](http://dx.doi.org/http://dx.doi.org/10.1016/S0169-5347(00)01825-5)
- Taberlet, P., Coissac, E., Pansu, J., & Pompanon, F. (2011). Conservation genetics of cattle, sheep, and goats. *Comptes Rendus Biologies*, 334(3), 247-254. <http://dx.doi.org/http://dx.doi.org/10.1016/j.crv.2010.12.007>
- Takahata, N., & Nei, M. (1990). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*, 124(4), 967-978.
- Takezaki, N., & Nei, M. (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, 144(1), 389-399.
- Takezaki, N., & Nei, M. (2008). Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. *Genetics*, 178(1), 385-392.

- Tano, K., Kamuanga, M., Faminow, M. D., & Swallow, B. (2003). Using conjoint analysis to estimate farmer's preferences for cattle traits in West Africa. *Ecol Econ*, 45(3), 393-407. [http://dx.doi.org/10.1016/S0921-8009\(03\)00093-4](http://dx.doi.org/10.1016/S0921-8009(03)00093-4)
- Taylor, J. E., & Rudenko, G. (2006). Switching trypanosome coats: what's in the wardrobe? *Trends Genet*, 22(11), 614-620. <http://dx.doi.org/http://dx.doi.org/10.1016/j.tig.2006.08.003>
- Taylor, K. A. (1998). Immune responses of cattle to African trypanosomes: protective or pathogenic? *Int J Parasitol*, 28(2), 219-240. [http://dx.doi.org/10.1016/S0020-7519\(97\)00154-9](http://dx.doi.org/10.1016/S0020-7519(97)00154-9)
- Templeton, A. R. (1980). The theory of speciation via the founder principle. *Genetics*, 94(4), 1011-1038.
- Teweldmehidin, M. Y., & Conroy, A. (2010). The economic importance of draught oxen on small farms in Namibia's Eastern Caprivi Region. *African Journal of Agricultural Research*, 5(9), 928-934.
- The Bovine HapMap Consortium. (2009). Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science*, 324(5926), 528-532. <http://dx.doi.org/10.1126/science.1167936>
- The International Sheep Genomics, C., Archibald, A. L., Cockett, N. E., Dalrymple, B. P., Faraut, T., Kijas, J. W., . . . Xun, X. (2010). The sheep genome reference sequence: a work in progress. *Animal Genetics*, 41(5), 449-453. <http://dx.doi.org/10.1111/j.1365-2052.2010.02100.x>
- Thomas, M. L., Harger, J. H., Wagener, D. E., Rabin, B. S., & Gill, T. J. (1986). HLA Sharing and Spontaneous Abortion in Humans. *Obstetrical & Gynecological Survey*, 41(2), 105-107.
- Thornton, P. K., Boone, R. B., Galvin, K. A., BurnSilver, S. B., Waithaka, M. M., Kuyiah, J., . . . Herrero, M. (2007). Coping strategies in livestock-dependent households in east and southern Africa: a synthesis of four case studies. *Human Ecology*, 35(4), 461-476.
- Thornton, P. K., van de Steeg, J., Notenbaert, A., & Herrero, M. (2009). The impacts of climate change on livestock and livestock systems in developing countries: A review of what we know and what we need to know. *Agricultural Systems*, 101(3), 113-127. <http://dx.doi.org/http://dx.doi.org/10.1016/j.agsy.2009.05.002>
- Thurston, L. M., Siggins, K., Mileham, A. J., Watson, P. F., & Holt, W. V. (2002). Identification of amplified restriction fragment length polymorphism markers linked to genes controlling boar sperm viability following cryopreservation. *Biology of Reproduction*, 66(3), 545-554.
- Tihon, E., Imamura, H., Dujardin, J.-C., Van Den Abbeele, J., & Van den Broeck, F. (2017). Discovery and genomic analyses of hybridization between divergent lineages of *Trypanosoma congolense*, causative agent of Animal African Trypanosomiasis. *Molecular Ecology*, n/a-n/a. <http://dx.doi.org/10.1111/mec.14271>
- Tisdell, C. (2003). Socioeconomic causes of loss of animal genetic diversity: analysis and assessment. *Ecological Economics*, 45(3), 365-376. [http://dx.doi.org/http://dx.doi.org/10.1016/S0921-8009\(03\)00091-0](http://dx.doi.org/http://dx.doi.org/10.1016/S0921-8009(03)00091-0)
- Tixier-Boichard, M. (2014). *Status and gaps in characterization of animal genetic resources*. Paper presented at the 10th World Congress on Genetics Applied to Livestock Production, Vancouver, Canada. August 17-22, 2014
- Tollenaere, C., Bryja, J., Galan, M., Cadet, P., Deter, J., Chaval, Y., . . . Charbonnel, N. (2008). Multiple parasites mediate balancing selection at two MHC class II genes in the fossorial water vole: insights from multivariate analyses and population genetics. *Journal of Evolutionary Biology*, 21(5), 1307-1320. <http://dx.doi.org/10.1111/j.1420-9101.2008.01563.x>
- Toro, M. A., Fernández, J., & Caballero, A. (2009). Molecular characterization of breeds and its use in conservation. *Livestock Science*, 120(3), 174-195. <http://dx.doi.org/http://dx.doi.org/10.1016/j.livsci.2008.07.003>
- Torr, S. J., & Vale, G. A. (2015). Know your foe: lessons from the analysis of tsetse fly behaviour. *Trends Parasitol.*, 31(3), 95-99. <http://dx.doi.org/http://dx.doi.org/10.1016/j.pt.2014.12.010>
- Trail, J. C., d'Ieteren, G. D., Colardelle, C., Maille, J. C., Ordner, G., Sauveroche, B., & Yangari, G. (1990). Evaluation of a field test for trypanotolerance in young N'Dama cattle. *Acta Trop*, 48(1), 47-57. [http://dx.doi.org/http://dx.doi.org/10.1016/0001-706X\(90\)90064-7](http://dx.doi.org/http://dx.doi.org/10.1016/0001-706X(90)90064-7)
- Trail, J. C., d'Ieteren, G. D., Maille, J. C., & Yangari, G. (1991). Genetic aspects of control of anaemia development in trypanotolerant N'Dama cattle. *Acta Trop*, 48(4), 285-291. [http://dx.doi.org/http://dx.doi.org/10.1016/0001-706X\(91\)90016-D](http://dx.doi.org/http://dx.doi.org/10.1016/0001-706X(91)90016-D)

- Traoré, A., Álvarez, I., Fernández, I., Pérez-Pardal, L., Kaboré, A., Ouédraogo-Sanou, G. M., . . . Goyache, F. (2012). Ascertaining gene flow patterns in livestock populations of developing countries: a case study in Burkina Faso goat. *BMC Genetics*, *13*(1), 35. <http://dx.doi.org/10.1186/1471-2156-13-35>
- Traoré, A., Notter, D. R., Soudre, A., Kaboré, A., Álvarez, I., Fernández, I., . . . Goyache, F. (2017). Resistance to gastrointestinal parasite infection in Djallonké sheep. *animal*, 1-9. <http://dx.doi.org/10.1017/S1751731116002640>
- Tsegaye, B., Dagnachew, S., & Terefe, G. (2015). Review on Drug Resistant Animal Trypanosomes in Africa and Overseas. *Afr J Basic App Sci*, *7*(2), 73-83.
- Tuggle, C. K., Giuffra, E., White, S. N., Clarke, L., Zhou, H., Ross, P. J., . . . Watson, M. (2016). GO-FAANG meeting: a Gathering On Functional Annotation of Animal Genomes. *Animal Genetics*, *47*(5), 528-533. <http://dx.doi.org/10.1111/age.12466>
- Tuncer, F. N., Gormez, Z., Calik, M., Altioikka Uzun, G., Sagiroglu, M. S., Yuceturk, B., . . . Ozbek, U. (2015). A clinical variant in SCN1A inherited from a mosaic father cosegregates with a novel variant to cause Dravet syndrome in a consanguineous family. *Epilepsy Research*, *113*, 5-10. <http://dx.doi.org/http://dx.doi.org/10.1016/j.epilepsyres.2015.02.020>
- Ujvari, B., & Belov, K. (2011). Major Histocompatibility Complex (MHC) Markers in Conservation Biology. *International Journal of Molecular Sciences*, *12*(8), 5168. Retrieved from <http://www.mdpi.com/1422-0067/12/8/5168>
- United Nations. (2000). *Millennium Development Goals and Beyond 2015*. <http://www.un.org/millenniumgoals/>; Retrieved from <http://www.un.org/millenniumgoals/>
- Van der Werf, J. H. (2007). Marker assisted selection in sheep and goats *Marker-assisted selection: current status and future perspectives in crops, livestock, forestry and fish* (pp. 229-247): FAO, Rome, Italy.
- Van Houtert, M. F. J., & Sykes, A. R. (1996). Implications of nutrition for the ability of ruminants to withstand gastrointestinal nematode infections. *Int J Parasitol*, *26*(11), 1151-1167. [http://dx.doi.org/http://dx.doi.org/10.1016/S0020-7519\(96\)00120-8](http://dx.doi.org/http://dx.doi.org/10.1016/S0020-7519(96)00120-8)
- Vanhala, T., Tuiskula-Haavisto, M., Elo, K., Vilkki, J., & Maki-Tanila, A. (1998). Evaluation of genetic variability and genetic distances between eight chicken lines using microsatellite markers. *Poultry Science*, *77*(6), 783-790.
- Vasconcellos, L. P. d. M. K., Tambasco-Talhari, D., Pereira, A. P., Coutinho, L. L., & Regitano, L. C. d. A. (2003). Genetic characterization of Aberdeen Angus cattle using molecular markers. *Genetics and Molecular Biology*, *26*, 133-137. Retrieved from [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1415-47572003000200005&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572003000200005&nrm=iso)
- Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, *34*(3), 275-306.
- Vigne, J.-D. (2011). The origins of animal domestication and husbandry: A major change in the history of humanity and the biosphere. *Comptes Rendus Biologies*, *334*(3), 171-181. <http://dx.doi.org/http://dx.doi.org/10.1016/j.crv.2010.12.009>
- Visscher, P., Woolliams, J., Smith, D., & Williams, J. (2002). Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. *Journal of Dairy Science*, *85*(9), 2368-2375.
- Vodička, P., Smetana, K., Dvořánková, B., Emerick, T., Xu, Y. Z., Ourednik, J., . . . Motlík, J. A. N. (2005). The miniature pig as an animal model in biomedical research. *Annals of the New York Academy of Sciences*, *1049*(1), 161-171. <http://dx.doi.org/10.1196/annals.1334.015>
- Wafula, P., Jianlin, H., Sangare, N., Sowe, J., Coly, R., Diallo, B., & Hanotte, O. (2005). *Genetic characterization of West African Djallonké sheep using microsatellite markers*. Paper presented at the International workshop "The role of biotechnology for the characterization and conservation of crop, forestry, animal and fishery genetic resources", Villa Gualino, Turin
- Wallace, D. S., Bairden, K., Duncan, J. L., Fishwick, G., Gill, M., Holmes, P. h., . . . Stear, M. J. (1995). Influence of supplementation with dietary soyabean meal on resistance to haemonchosis in Hampshire down lambs. *Res Vet Sci*, *58*(3), 232-237. [http://dx.doi.org/http://dx.doi.org/10.1016/0034-5288\(95\)90108-6](http://dx.doi.org/http://dx.doi.org/10.1016/0034-5288(95)90108-6)

- Wasko, A. P., & Galetti, P. M. (2003). PCR primed with minisatellite core sequences yields species-specific patterns and assessment of population variability in fishes of the genus *Brycon*. *Journal of Applied Ichthyology*, *19*(2), 109-113. <http://dx.doi.org/10.1046/j.1439-0426.2003.00443.x>
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *evolution*, 1358-1370.
- Welburn, S. C., Fèvre, E. M., Coleman, P. G., Odiit, M., & Maudlin, I. (2001). Sleeping sickness: a tale of two diseases. *Trends Parasitol*, *17*(1), 19-24. [http://dx.doi.org/http://dx.doi.org/10.1016/S1471-4922\(00\)01839-0](http://dx.doi.org/http://dx.doi.org/10.1016/S1471-4922(00)01839-0)
- Welburn, S. C., Maudlin, I., & Simarro, P. P. (2009). Controlling sleeping sickness – a review. *Parasitology*, *136*(Special Issue 14), 1943-1949. <http://dx.doi.org/doi:10.1017/S0031182009006416>
- WHO. (2002). World Health Organisation programme to eliminate sleeping sickness : Building a global alliance (Publication no. [http://apps.who.int/iris/bitstream/10665/67352/1/WHO\\_CDS\\_CSR\\_EPH\\_2002.13.pdf?ua=1](http://apps.who.int/iris/bitstream/10665/67352/1/WHO_CDS_CSR_EPH_2002.13.pdf?ua=1)). Institutional Repository for Information Sharing (IRIS) Retrieved May 29, 2014, from Institutional Repository for Information Sharing (IRIS), WHO publications [http://apps.who.int/iris/bitstream/10665/67352/1/WHO\\_CDS\\_CSR\\_EPH\\_2002.13.pdf?ua=1](http://apps.who.int/iris/bitstream/10665/67352/1/WHO_CDS_CSR_EPH_2002.13.pdf?ua=1)
- Wilkinson, S., Lu, Z. H., Megens, H.-J., Archibald, A. L., Haley, C., Jackson, I. J., . . . Wiener, P. (2013). Signatures of Diversifying Selection in European Pig Breeds. *PLOS Genetics*, *9*(4), e1003453. <http://dx.doi.org/10.1371/journal.pgen.1003453>
- Wilkinson, S., Wiener, P., Archibald, A. L., Law, A., Schnabel, R. D., McKay, S. D., . . . Ogden, R. (2011). Evaluation of approaches for identifying population informative markers from high density SNP Chips. *BMC Genetics*, *12*(1), 45.
- Witko-Sarsat, V., Friedlander, M., Capeillère-Blandin, C., Nguyen-Khoa, T., Nguyen, A. T., Zingraff, J., . . . Descamps-Latscha, B. (1996). Advanced oxidation protein products as a novel marker of oxidative stress in uremia. *Kidney International*, *49*(5), 1304-1313. <http://dx.doi.org/http://dx.doi.org/10.1038/ki.1996.186>
- Wolde-Rufael, Y. (2006). Electricity consumption and economic growth: a time series experience for 17 African countries. *Energy Policy*, *34*(10), 1106-1114. <http://dx.doi.org/http://dx.doi.org/10.1016/j.enpol.2004.10.008>
- Wollny, C. B. A. (2003). The need to conserve farm animal genetic resources in Africa: should policy makers be concerned? *Ecological Economics*, *45*(3), 341-351. [http://dx.doi.org/http://dx.doi.org/10.1016/S0921-8009\(03\)00089-2](http://dx.doi.org/http://dx.doi.org/10.1016/S0921-8009(03)00089-2)
- Wong, T.-Y., Fernandes, S., Sankhon, N., Leong, P. P., Kuo, J., & Liu, J.-K. (2008). Role of Premature Stop Codons in Bacterial Evolution. *Journal of Bacteriology*, *190*(20), 6718-6725. <http://dx.doi.org/10.1128/jb.00682-08>
- Worley, K., Collet, J., Spurgin, L. G., Cornwallis, C., Pizzari, T., & Richardson, D. S. (2010). MHC heterozygosity and survival in red junglefowl. *Molecular Ecology*, *19*(15), 3064-3075.
- Xiao-Yong, C., Shu-Jun, T., Run-Zi, S., Shu-Chun, S., Zhu-Jun, Z., Jun-Jie, L., & Wei-Tao, D. (2008). Induction of lamb follicular development and embryo production in vitro. *Chinese Journal of Agricultural Biotechnology*, *5*(03), 257-262.
- Yamazaki, K., Beauchamp, G., Kupniewski, D., Bard, J., Thomas, L., & Boyse, E. (1988). FAMILIAL IMPRINTING DETERMINES H-2 SELECTIVE MATING PREFERENCES. *SCIENCE*, *240*(4857), 1331-1332.
- Yang, G., Yan, J., Zhou, K., & Wei, F. (2005). Sequence Variation and Gene Duplication at MHC DQB Loci of Baiji (*Lipotes vexillifer*), a Chinese River Dolphin. *Journal of Heredity*, *96*(4), 310-317. <http://dx.doi.org/10.1093/jhered/esi055>
- Yang, Y., Zhou, Q.-J., Chen, X.-Q., Yan, B.-L., Guo, X.-L., Zhang, H.-L., & Du, A.-F. (2015). Profiling of differentially expressed genes in sheep T lymphocytes response to an artificial primary *Haemonchus contortus* infection. *Parasites & Vectors*, *8*(1), 235. <http://dx.doi.org/10.1186/s13071-015-0844-z>
- Yapi-Gnaoré, C. V., Dagnogo, B., & Oya, B. (2003). *Community-based livestock improvement and conservation: experiences from open-nucleus breeding programmes in West Africa*. Paper presented at the COMMUNITY-BASED MANAGEMENT OF ANIMAL GENETIC RESOURCES, Proceedings of the workshop held in Mbabane, Swaziland 7-11 May, 2001

- Yaro, M., Ayizanga, R. A., & Abdul-Rahman, I. I. (2012). Growth Performance of Zero Grazed Sahelian x Djallonke Crossbreeds Under Probiotic Strategic Supplementation. *J. Anim. Sci. Adv.*, 2(5), 487-491. Retrieved from <http://www.scopemed.org/?mno=21699>
- Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., . . . Flicek, P. (2016). Ensembl 2016. *Nucleic Acids Research*, 44(D1), D710-D716. <http://dx.doi.org/10.1093/nar/gkv1157>
- York, R., & Mancus, P. (2013). The invisible animal anthrozoology and macrosociology. *Sociological Theory*, 31(1), 75-91.
- Yuan, Y., Xu, H., & Leung, R. K.-K. (2016). An optimized protocol for generation and analysis of Ion Proton sequencing reads for RNA-Seq. *BMC Genomics*, 17(1), 403. <http://dx.doi.org/10.1186/s12864-016-2745-8>
- Zane, L., Bargelloni, L., & Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology*, 11(1), 1-16.
- Zeder, M. A., Emshwiller, E., Smith, B. D., & Bradley, D. G. (2006). Documenting domestication: the intersection of genetics and archaeology. *Trends in Genetics*, 22(3), 139-155. <http://dx.doi.org/http://dx.doi.org/10.1016/j.tig.2006.01.007>
- Zhan, B., Fadista, J., Thomsen, B., Hedegaard, J., Panitz, F., & Bendixen, C. (2011). Global assessment of genomic variation in cattle by genome resequencing and high-throughput genotyping. *BMC Genomics*, 12 <http://dx.doi.org/10.1186/1471-2164-12-557>
- Zhan, X., Dixon, A., Batbayar, N., Bragin, E., Ayas, Z., Deuschova, L., . . . Bruford, M. W. (2015). Exonic versus intronic SNPs: contrasting roles in revealing the population genetic differentiation of a widespread bird species. *Heredity*, 114(1), 1-9. <http://dx.doi.org/10.1038/hdy.2014.59>
- Zhang, D.-X., & Hewitt, G. M. (1996). Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution*, 11(6), 247-251. [http://dx.doi.org/http://dx.doi.org/10.1016/0169-5347\(96\)10031-8](http://dx.doi.org/http://dx.doi.org/10.1016/0169-5347(96)10031-8)
- Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A Greedy Algorithm for Aligning DNA Sequences. *Journal of Computational Biology*, 7(1-2), 203-214. <http://dx.doi.org/10.1089/10665270050081478>
- Zhao, S., Zheng, P., Dong, S., Zhan, X., Wu, Q., Guo, X., . . . Fan, W. (2013). Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nature Genetics*, 45(1), 67-71.
- Zhou, H., & Hickford, J. G. H. (2004). Allelic polymorphism in the ovine DQA1 gene1. *Journal of Animal Science*, 82(1), 8-16. <http://dx.doi.org/10.2527/2004.8218>
- Zougmore, R., Traoré, A. S., & Mbodj, Y. (2015). Overview of the Scientific, Political and Financial Landscape of Climate-Smart Agriculture in West Africa. In A. S. T. Robert Zougmore, Yamar Mbodj (Ed.), *CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS)* (Vol. Working Paper No 118, pp. 27-35). Denmark: CCAFS Coordination Unit - Faculty of Science, Department of Plant and Environmental Sciences, University of Copenhagen, Rolighedsvej 21, DK-1958 Frederiksberg C. Tel: +45 35331046; Email: [ccafs@cgiar.org](mailto:ccafs@cgiar.org). Retrieved from [https://cgispace.cgiar.org/bitstream/handle/10568/67103/CCAFS\\_WP118\\_English\\_web.pdf?squence=1](https://cgispace.cgiar.org/bitstream/handle/10568/67103/CCAFS_WP118_English_web.pdf?squence=1)
- Zuckerkandl, E. (1965). Evolutionary divergence and convergence in proteins. *Evolving genes and proteins*, Retrieved from <http://ci.nii.ac.jp/naid/10015598121/en/>
- Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences*, 109(4), 1193-1198. <http://dx.doi.org/10.1073/pnas.1119675109>

***"Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged."***

---

## APPENDICES

---

### Appendix A Images of sampling regions and animals



**Appendix Figure 1 Sampling location for the Djallonke (Ejura) and Sahelian(Pong Tamale) sheep cohorts**

(adapted from <http://www.lib.utexas.edu/maps/ghana.html> University of Texas, Austin)



**Appendix Figure 2 Adult female (left) and male (right) Djallonke Sheep breed (<http://www.fao.org/wairdocs/ilri/x5468e/x5468e04.htm> accessed on 21.10.2014)**



**Appendix Figure 3 Adult Male (left) and female (right) Sahelian sheep breed (<http://www.ansi.okstate.edu/breeds/sheep/saheltype>)**

## Appendix B Protocols for preparing media, buffers and solutions

**Appendix Table B.1 Recipes and protocols for reagents used for ligation and transformation**

Reagent name	Preparation Protocol
IPTG 0.1 M Sigma-Aldrich	IPTG was prepared as a 50 mg/ml stock solution by dissolving 2.38 g of IPTG powder (catalogue number 367-93-1) in 10 ml of ultra-pure water, and stored at -20°C.
X-GAL Fisher-Biotech	X-GAL was prepared as a 20 mg/ml stock solution by dissolving 0.4 g of X-GAL (catalogue number 1758-0300-1) in 20 ml DMSO (catalogue number 9326410026999, CHEM-SUPPLY, Australia), and stored at -20°C, protected from light.
Ampicillin sodium salt Amresco	Ampicillin was prepared as a 50 mg/ml stock solution by dissolving 2 g of Ampicillin (catalogue number 3590C218, Ohio, USA) in 5 ml of absolute ethanol (catalogue number 1407175240, Thermo Fisher Scientific) and 5 ml of ultra-pure water, and stored at -20°C.
SOC medium Sigma-Aldrich	SOC medium was commercially prepared and purchased from Sigma-Aldrich, Australia (catalogue number S1797-10X5ML) and stored at 4°C.
50X TAE buffer	For 1 L of 50X TAE buffer, 242 g of Tris Base ultrapure grade (catalogue number 77-86-1, Astral, Australia), 57.1 mL acetic acid (catalogue number: 64-19-7, Sigma-Aldrich), and 100 ml of 0.5 M EDTA at pH 8 (catalogue number: 60-00-4, Sigma-Aldrich) were added to per 1 L of double deionized water (ddH <sub>2</sub> O), and stored at room temperature. To prepare working solutions of TAE, the 50X stock was diluted to 1X using ddH <sub>2</sub> O.
LB agar medium	Per 2 L of LB agar medium, 20 g of Bacto-Tryptone (catalogue number 211705, Becton, Dickson and Co. USA), 10 g of Bacto-yeast extract (catalogue number 212750, Becton, Dickson and Co.) and 20 g of NaCl (catalogue number 7647-14-5, Ajax Finechem, Australia) were dissolved in 1700 ml ddH <sub>2</sub> O and the pH adjusted to 7.5, followed by addition of 30g of

	Bacto Agar (catalogue number 214010 Becton, Dickson and Co.) and the final volume made up to 2 litres with ddH <sub>2</sub> O. The solution was sterilized by, then allowed to cool to 55-60°C before pouring into sterile petri dishes in a Class II biosafety cabinet.
LB broth medium	Commercially prepared LB broth (Catalogue number 10855-001, GIBCO, Life technologies) and in-house prepared LB broth were used. The in-house preparation was like that of LB agar described above, except that no Bacto Agar medium was added and LB broth was cooled to room temperature and stored at 4°C.

**Appendix Table B.2 Optimised PCR master mix for *DRB1*, *DQA1* and *DQA2* for HS MyTaq polymerase 50 µl PCR reaction**

MyTaq components	Volume used
5x MyTaq Reaction Buffer	10 µl
Template	1-5 µl (depending on template concentration)
Primers (Forward and Reverse) 20µM each	1 µl
MyTaq DNA Polymerase	0.5 µl
Molecular grade water	from 33.5 µl -37.5 µl (depending on template volume used)

**Appendix Table B.3 Optimised *DRB1* PCR cycling conditions for MyTaq HS**

Step	Temperature (°C)	Time	Number of Cycles
Initial denaturation	95	1 minute	1
Denaturation	95	15 seconds	36
Annealing	58	15 seconds	
Extension	72	10 seconds	
Final extension	72	7 minutes	1
Hold	4	Hold indefinite	Nil

## **B.1 NanoDrop DNA analysis**

The NanoDrop analyser was decontaminated by wiping the upper and lower pedestal with 0.5% sodium hypochlorite solution (Catalogue number SH-5L; Hurst Scientific PTY Ltd, Australia) before each use to ensure that there were no active biological agents on the pedestals. To measure the concentration of DNA, the software was set at DNA50. Measurement was initialized using 1  $\mu\text{L}$  of ultra-pure water, then another 1  $\mu\text{L}$  of ultra-pure water was used to record a “blank” measurement. DNA samples (1  $\mu\text{L}$ ) were then measured. After each measurement, sample pedestals were wiped with sterile blotting paper to prevent residual sample carryover. To ensure accuracy of measured DNA concentrations, measurements were performed in duplicate.

## **B.2 ExoSAP PCR clean up protocol**

For the ExoSAP protocol, each PCR amplicon required 10  $\mu\text{L}$  of master mix containing 0.025  $\mu\text{L}$  of Exonuclease I, 0.250  $\mu\text{L}$  of Shrimp Alkaline Phosphatase (rSAP) and 9.725  $\mu\text{L}$  of Ultrapure water was prepared. For each approximately 46  $\mu\text{L}$  of PCR amplicon (representing original 50  $\mu\text{L}$  minus 4 $\mu\text{L}$  used for gel electrophoresis, 10  $\mu\text{L}$  of 1x ExoSAP mix was added. Samples were incubated at 37°C for 30 minutes, then 80°C for 20 minutes on a thermocycler to de-activate the enzyme. 4  $\mu\text{L}$  of cleaned PCR amplicons were analysed using 1% agarose gel electrophoresis to check for success of the process before cloning. A clean DNA band of the expected size of amplicon indicated a successful clean-up.

## **B.3 Favorgen PCR clean up Protocol**

At room, temperature, FAPC wash buffer was activated by adding 180 ml of absolute ethanol (Fisher Biotec, Australia). 230  $\mu\text{L}$  of FAPC buffer was added to each 46  $\mu\text{L}$  of PCR amplicon, vortexed, then transferred to an FAPC column, centrifuged at 11,000 rcf for 30 seconds and the flow-through discarded. 600  $\mu\text{L}$  of FAPC wash buffer was added, centrifuged at 11,000 rcf for 30 seconds, the flow-through discarded, then the sample centrifuged at 18,000 rcf for 3 minutes (aka dry spin) to remove residual FAPC wash buffer. 40  $\mu\text{L}$  of pre-heated ultra-pure water at 60°C was added, incubated for 60 seconds, centrifuged at 18,000 rcf for 60 seconds, and the flow analysed by agarose gel electrophoresis.

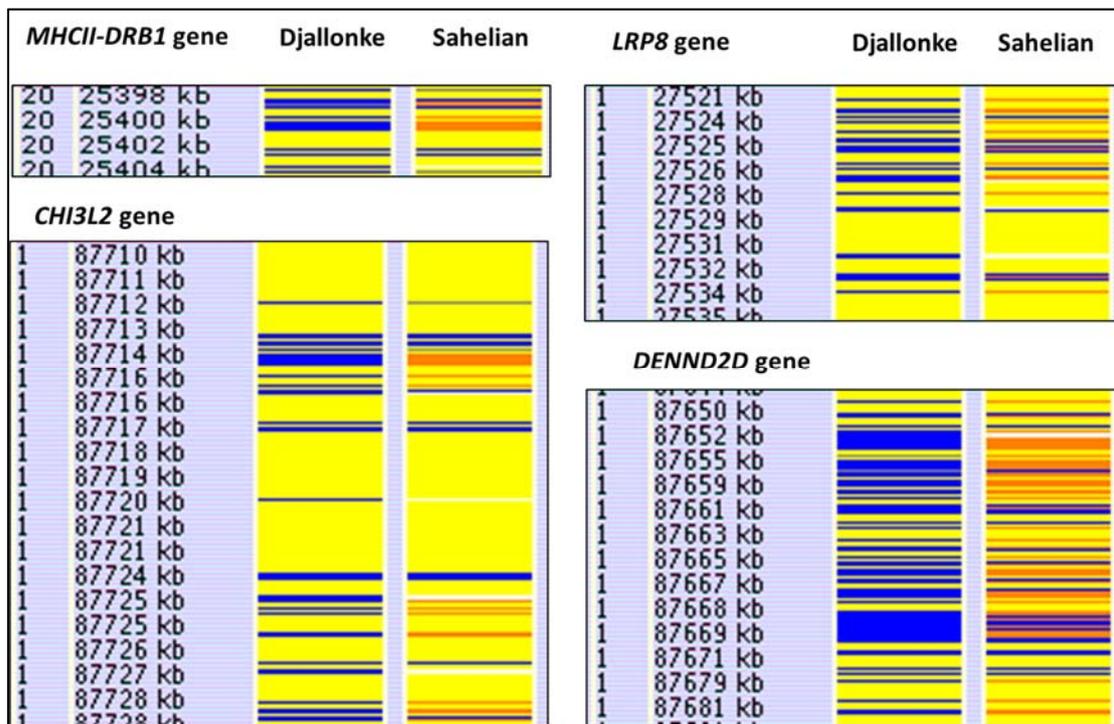
**Appendix Table B.4 Template requirement for Macrogen Incorporation standard sequencing**

Template (size)	Required concentration ng/ $\mu$ L	Required Sample Volume	Primer Volume + Concentration
Plasmid (below 4 kb)	100-150	5 $\mu$ L per reaction	5 $\mu$ L + 5 pmol/ $\mu$ L
Plasmid (over 4 kb)	200	5 $\mu$ L per reaction	5 $\mu$ L + 5 pmol/ $\mu$ L
PCR product (below 200 bp)	10	5 $\mu$ L per reaction	5 $\mu$ L + 5 pmol/ $\mu$ L
PCR product (201 bp-500 bp)	20	5 $\mu$ L per reaction	5 $\mu$ L + 5 pmol/ $\mu$ L
PCR product (over 500 bp)	Over 25	5 $\mu$ L per reaction	5 $\mu$ L + 5 pmol/ $\mu$ L
PCR product (over 1000 bp)	30-60	5 $\mu$ L per reaction	5 $\mu$ L + 5 pmol/ $\mu$ L
Bacterial Artificial Chromosome	Over 500	5 $\mu$ L per reaction	5 $\mu$ L + 10 pmol/ $\mu$ L

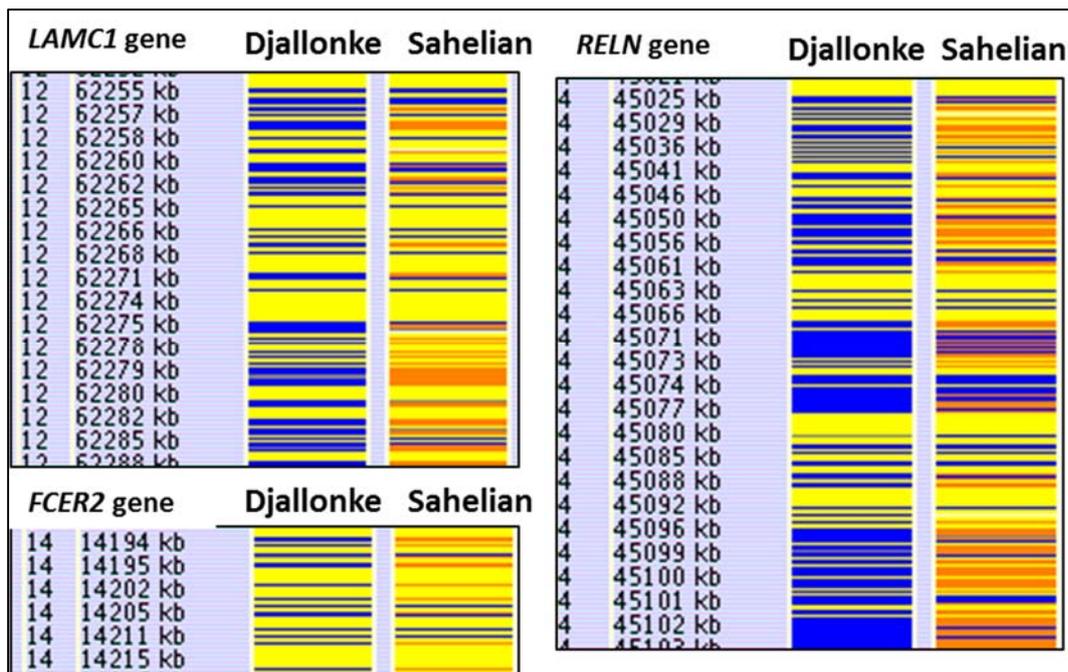
## Appendix C Additional results for chapter 3

### C.1 HomSI images of genomic regions associated with resistance to *H. contortus* and other GIT parasites in Djallonke and Sahelian sheep breeds

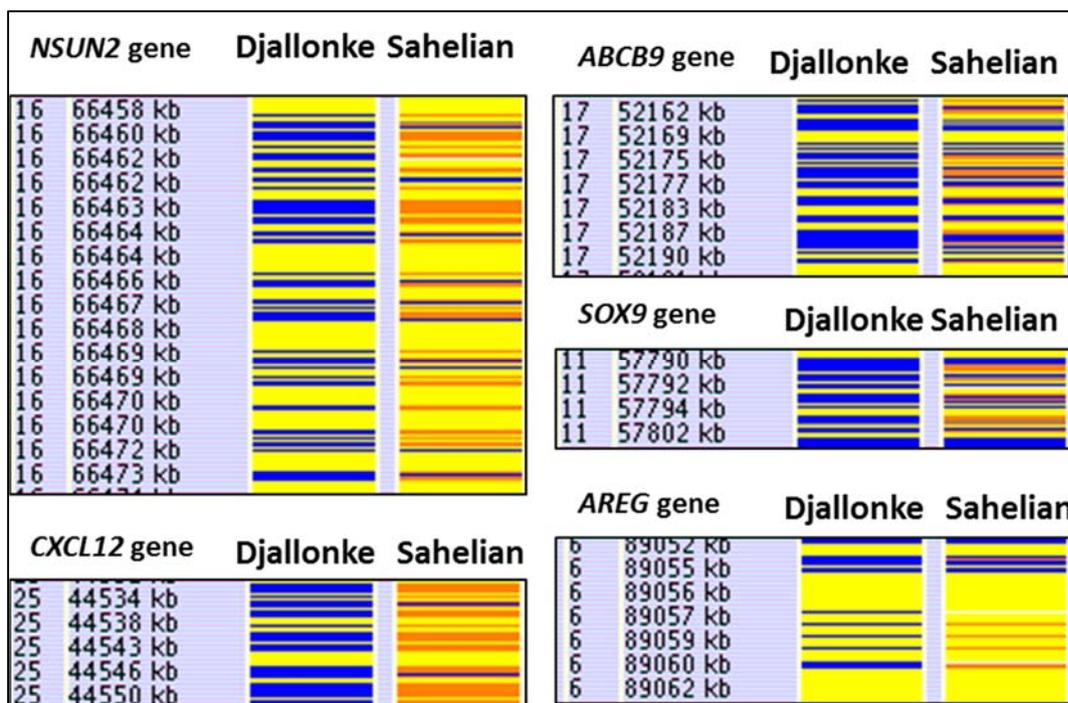
#### C.1.1 Contrasting signatures of selection in genetic regions associated with resistance to Haemonchosis and other GIT parasite the Djallonke (high ROH) and Sahelian (low ROH) sheep breeds



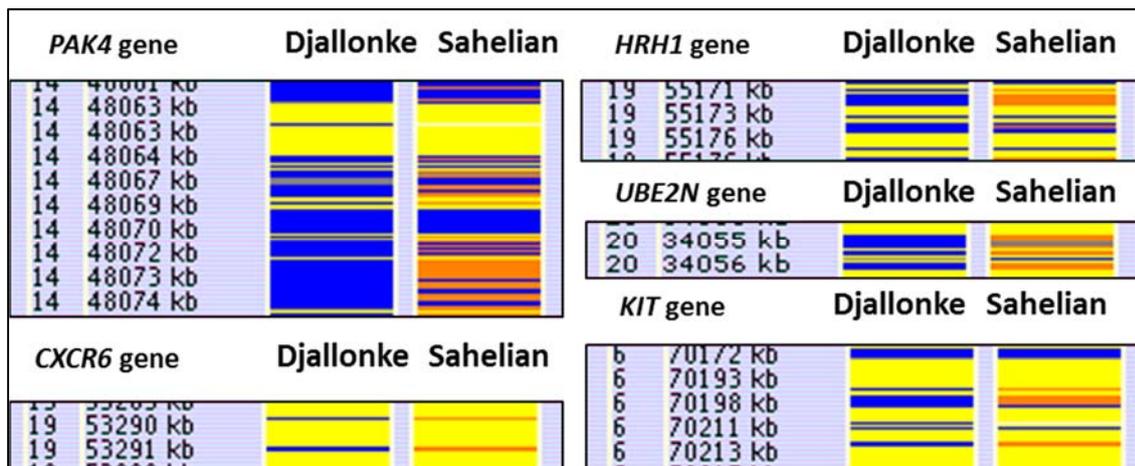
Appendix Figure 4 HomSI analysis adjacent to *MHC II-DRB1*, *CHI3L2*, *LRP8* and *DENND2D* genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH)



Appendix Figure 5 HomSI analysis adjacent to *LAMC1*, *RELN* and *FCER2* genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH)

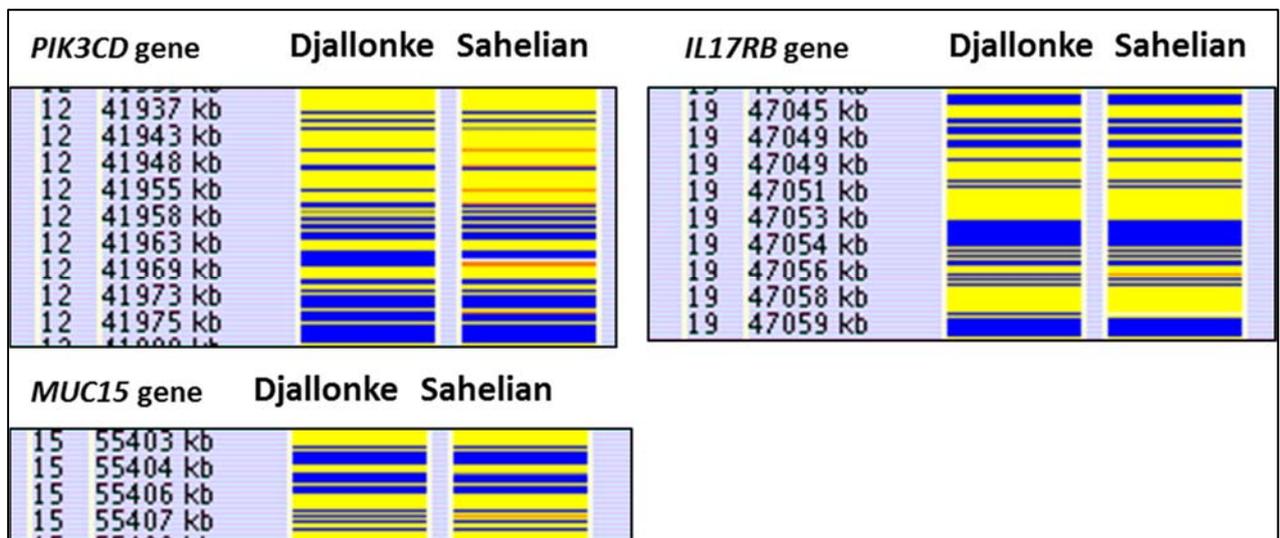


Appendix Figure 6. HomSI analysis adjacent to *NSUN2*, *CXCL12*, *ABCB9*, *SOX9* and *AREG* genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH)



Appendix Figure 7 HomSI region adjacent to *PAK4*, *CXCR6*, *HRH1*, *UBE2N* and *KIT* genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH)

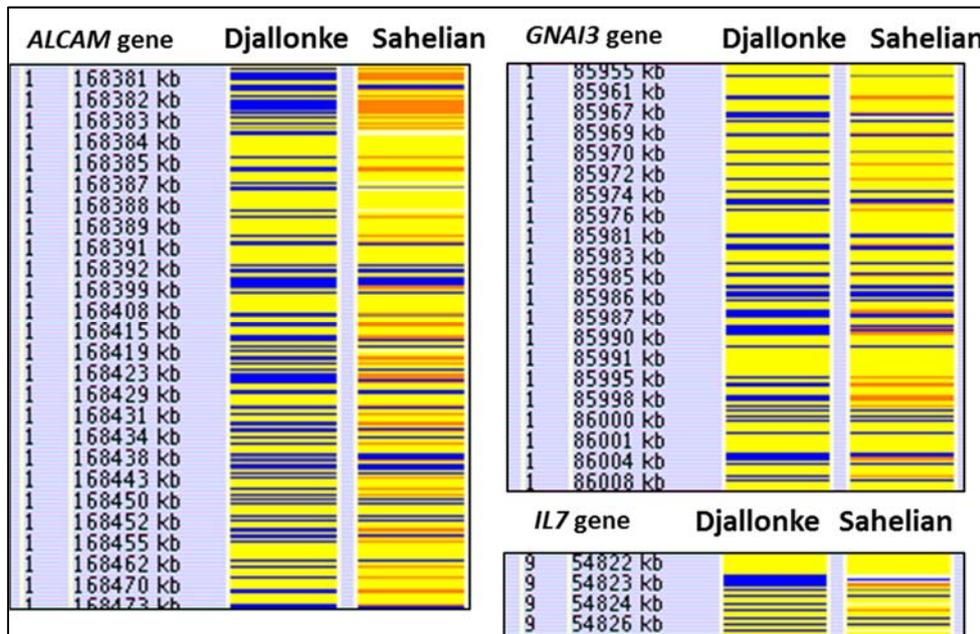
C.1.2 Similar signatures of selection in genetic regions associated with resistance to Haemonchosis and other GIT parasite the Djallonke (high ROH) and Sahelian (high ROH) sheep breeds



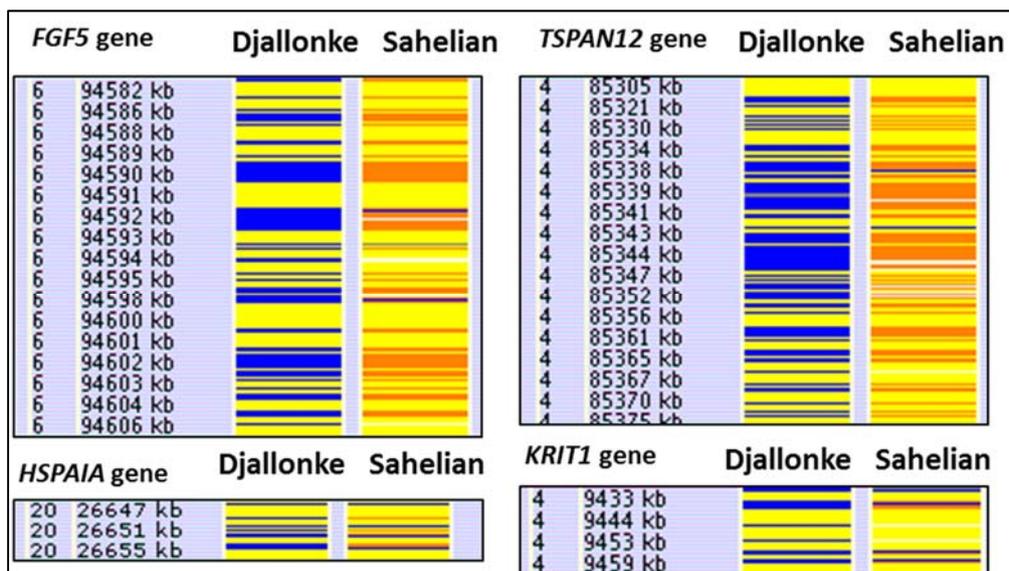
Appendix Figure 8 HomSI analysis adjacent to *PIK3CD*, *IL17RB* & *MUC15* genes showing similar sweeps in Djallonke (high ROH) and Sahelian (high ROH).

## C.2 HomSI analyses of genomic regions associated with adaptive selection in Djallonke and Sahelian sheep breeds

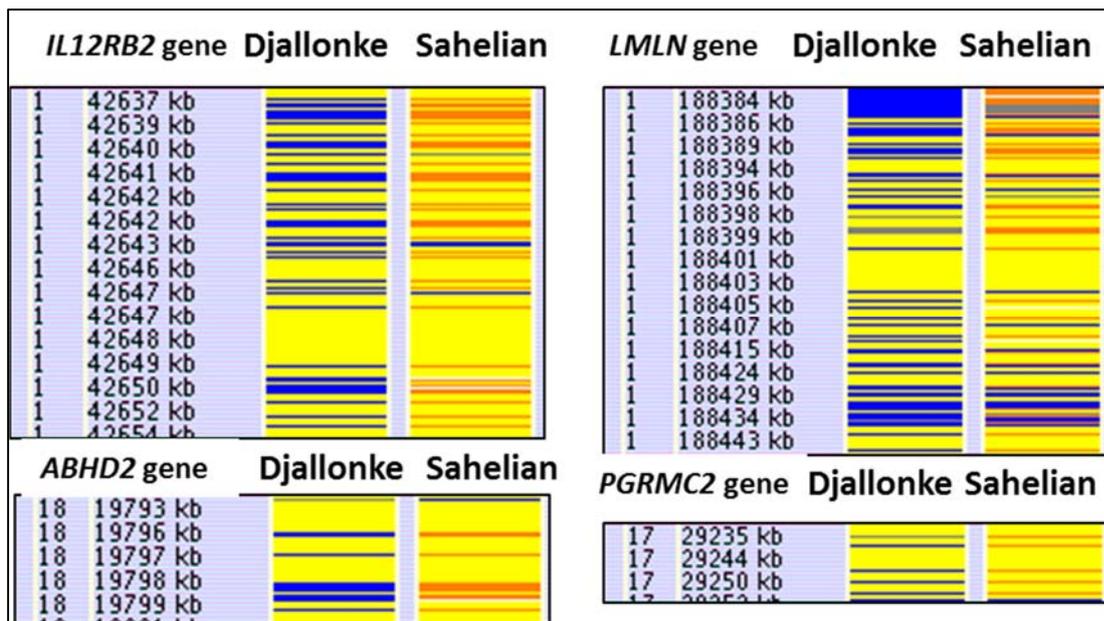
### C.2.1 Contrasting signatures of adaptive selection in genomic regions associated with the Djallonke (high ROH) and Sahelian (low ROH) sheep breeds



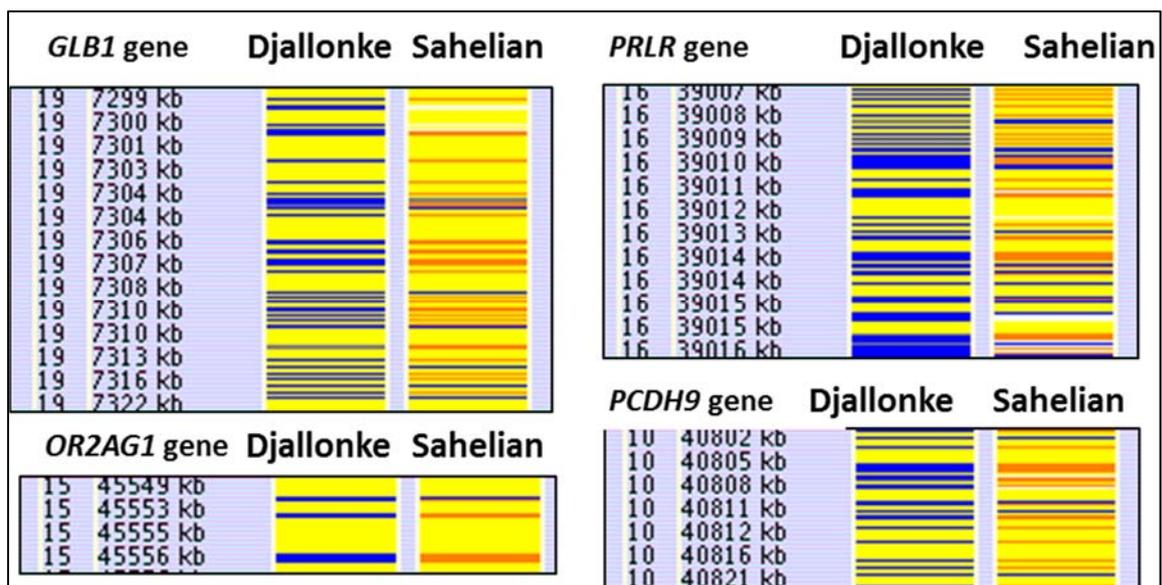
Appendix Figure 9 HomSI analysis adjacent to *ALCAM*, *GNAI3* & *IL7* genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH).



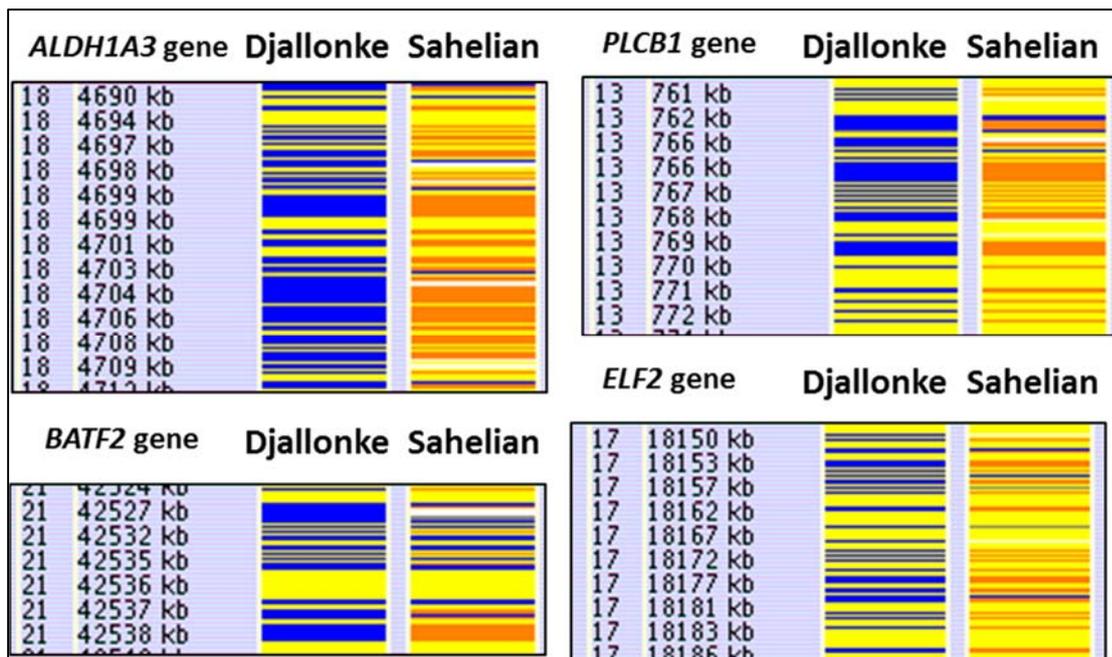
Appendix Figure 10 HomSI analysis adjacent to *FGF5*, *TSAPAN12*, *HSPAIA* & *KRIT1* genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH).



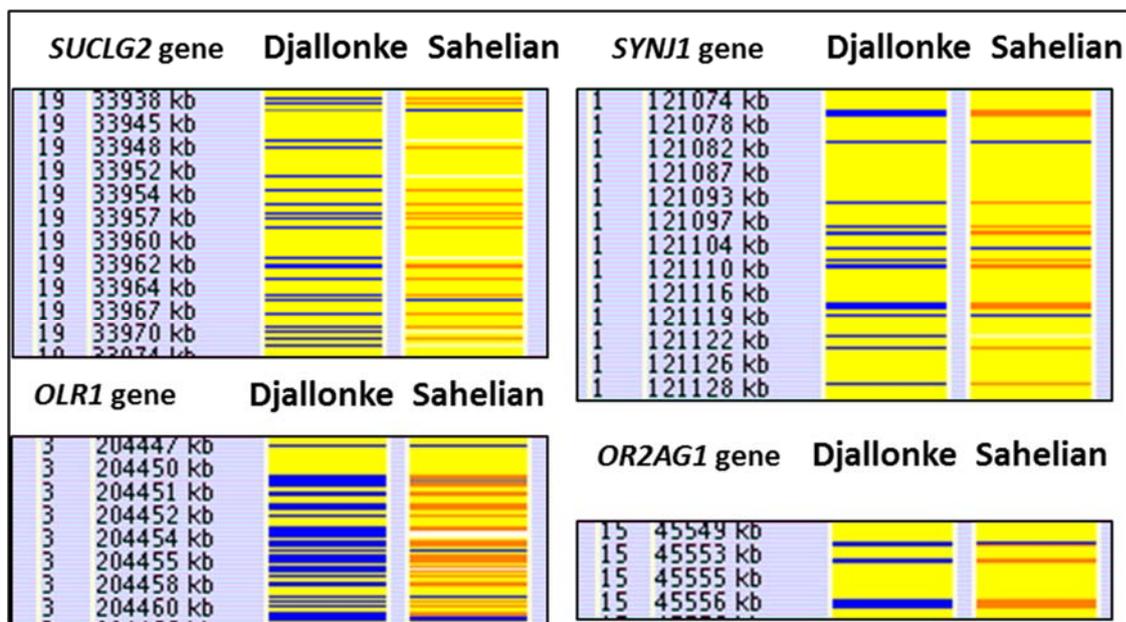
Appendix Figure 11 HomSI analysis adjacent to *IL12RB2*, *LMLN*, *ABHD2* & *PGRMC2* genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH).



Appendix Figure 12 HomSI analysis adjacent to *GLB1*, *OR2AG1*, *PRLR* & *PCDH9* genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH).

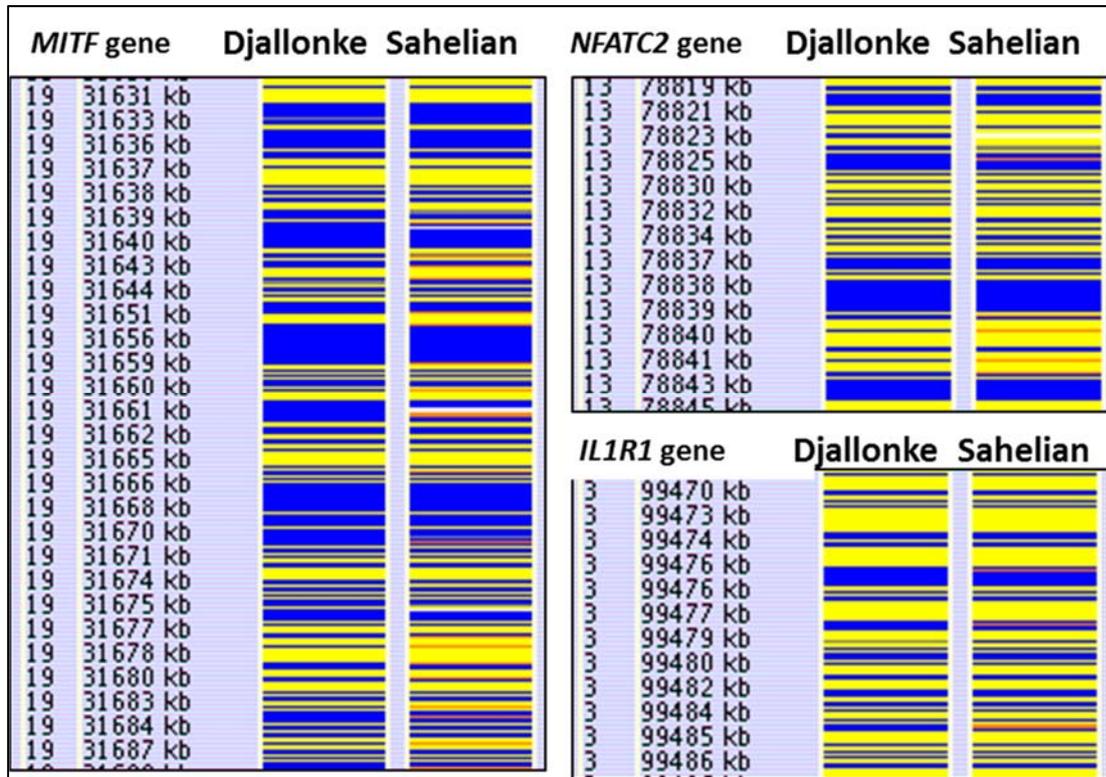


Appendix Figure 13 HomSI analysis adjacent to *ALDH1A3*, *PLCB1*, *BATF2* & *ELF2* genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH).

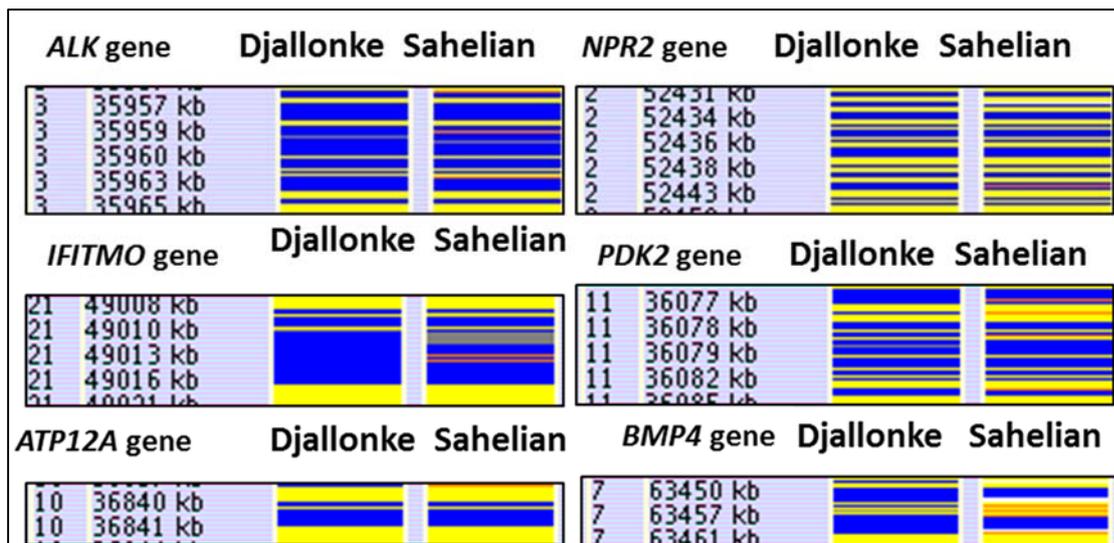


Appendix Figure 14 HomSI analysis adjacent to *SUCLG2*, *OLR1*, *SYNJ1* & *OR2AG1* genes showing contrasting sweeps in Djallonke (high ROH) and Sahelian (low ROH).

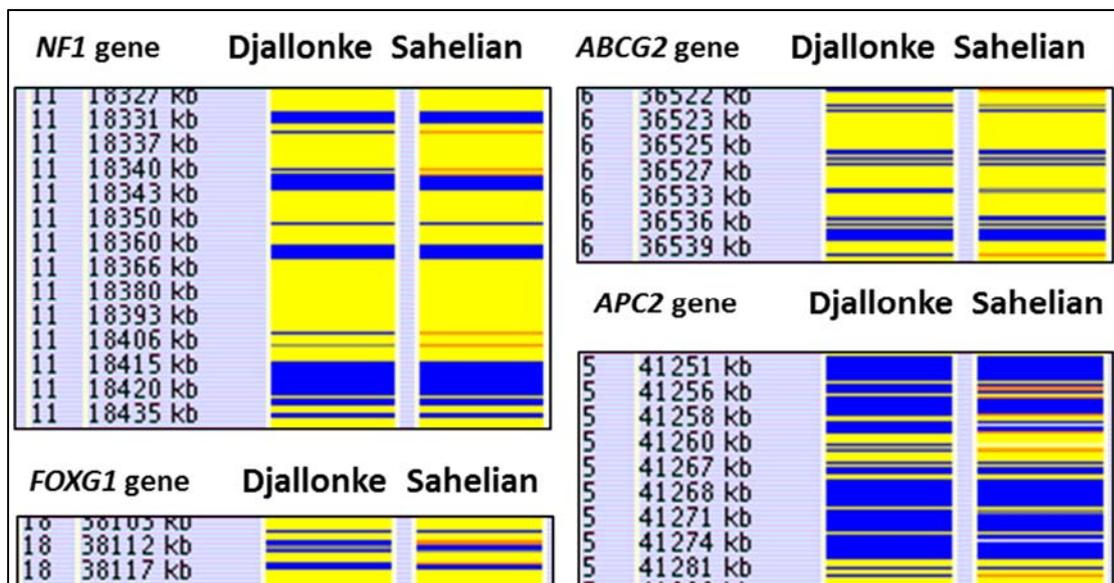
**C.2.2 Similar signatures of selection in genomic regions associated with adaptation in both the Djallonke (high ROH) and Sahelian (high ROH) sheep breeds**



**Appendix Figure 15 HomSI analysis adjacent to *MITF*, *NFATC2* & *IL1R1* genes showing similar sweeps in Djallonke (high ROH) and Sahelian (low ROH).**



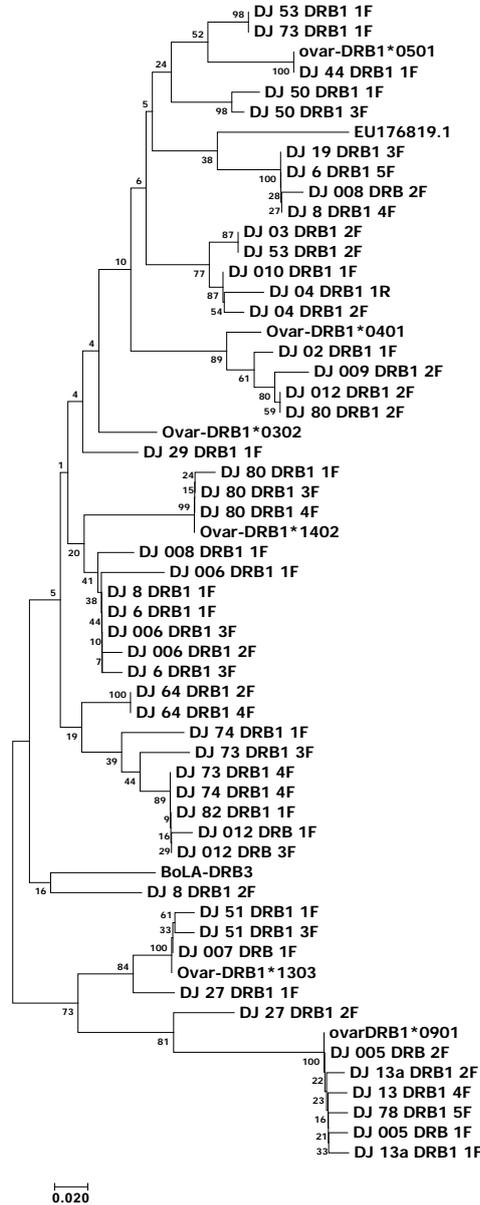
Appendix Figure 16 HomSI analysis adjacent to *ALK*, *NPR2*, *IFITMO*, *PDK2*, *ATP12A* & *BMP4* genes showing similar sweeps in Djallonke (high ROH) and Sahelian (low ROH).



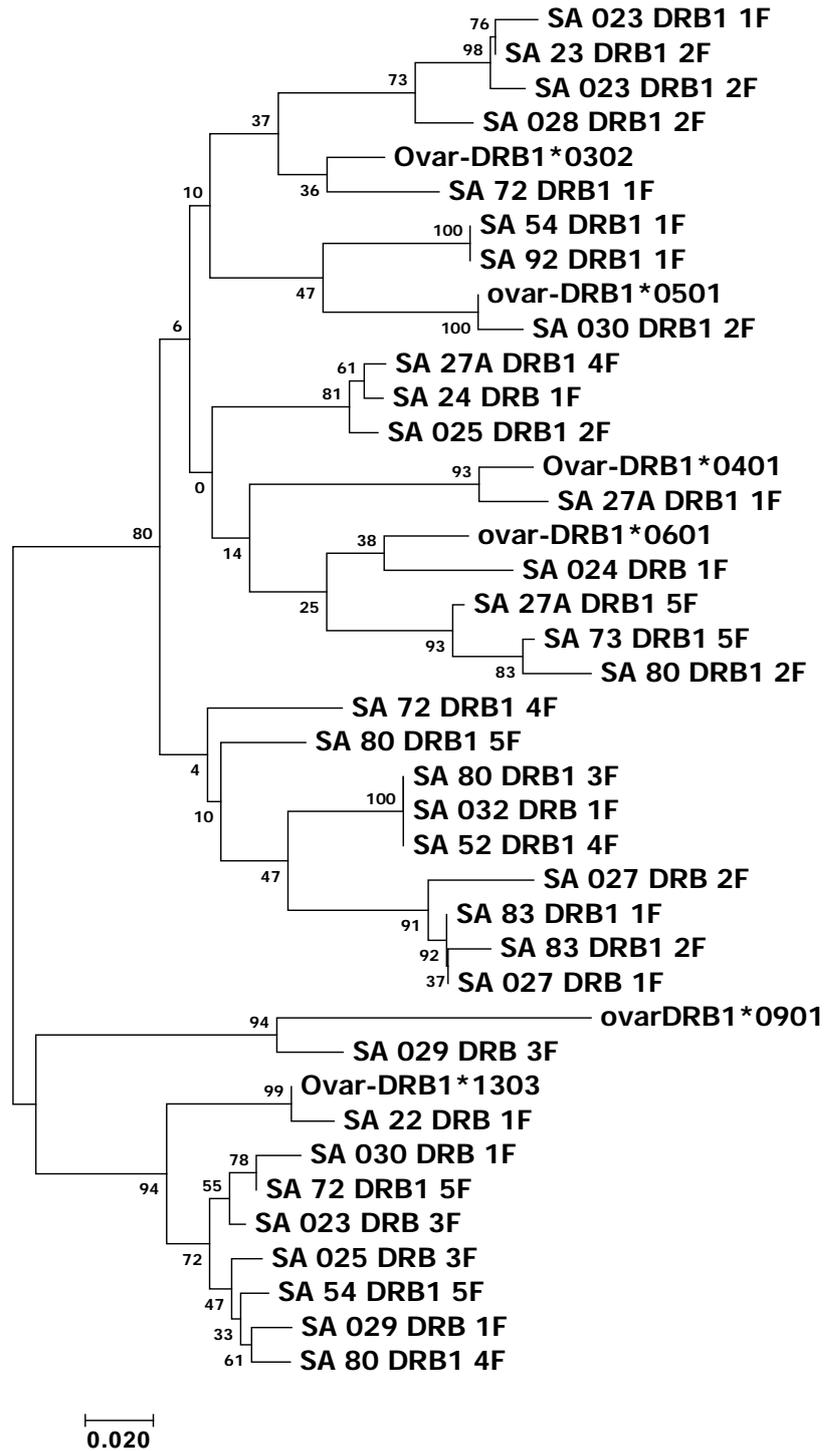
Appendix Figure 17 HomSI analysis adjacent to *NF1*, *ABCG2*, *FOXG1* & *APC2* genes showing similar sweeps in Djallonke (high ROH) and Sahelian (low ROH).

# Appendix D Additional results for Chapter 4

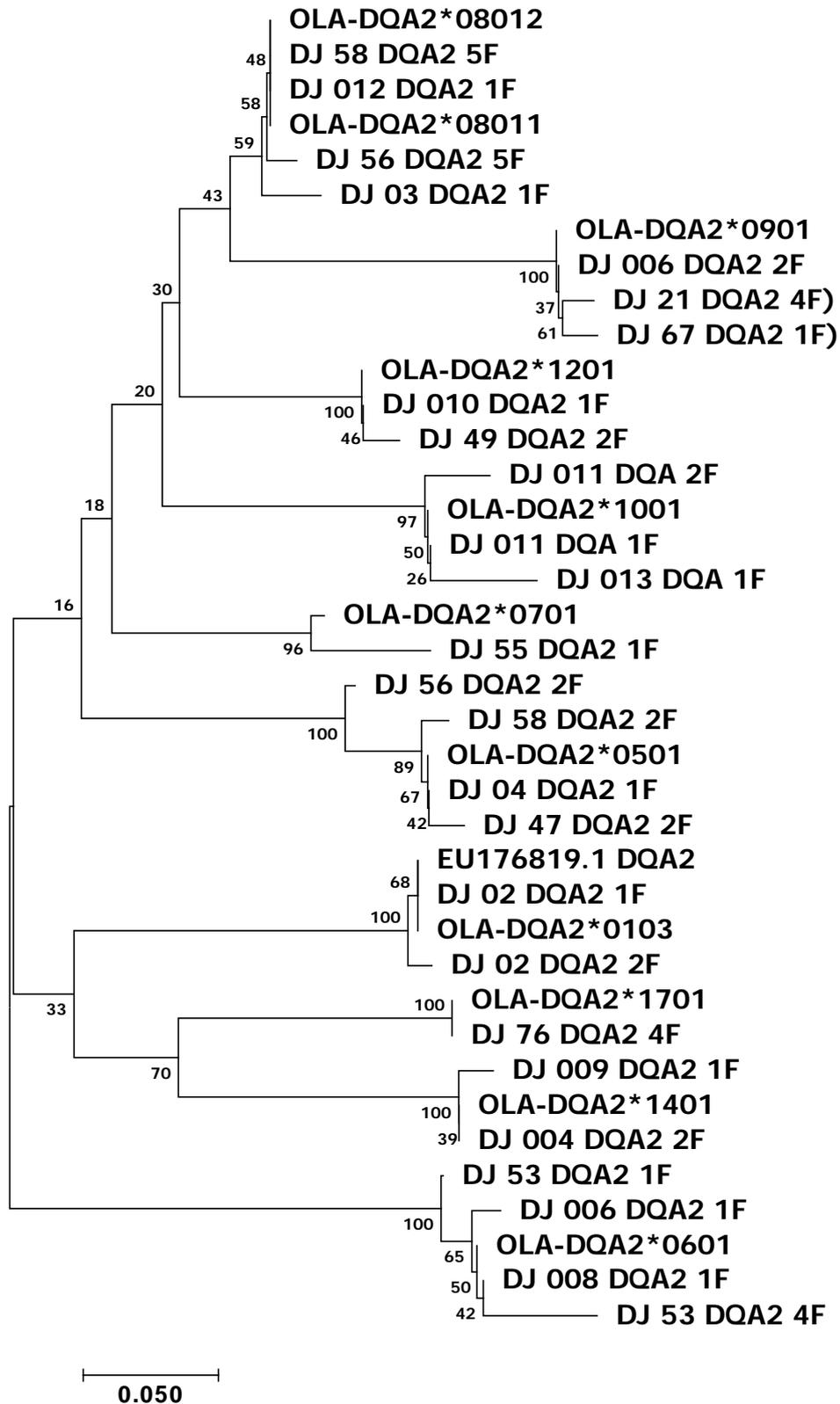
## D.1 Maximum Likelihood Neighbour Joining Tree (ML-NJT) Model for Djallonke the sheep



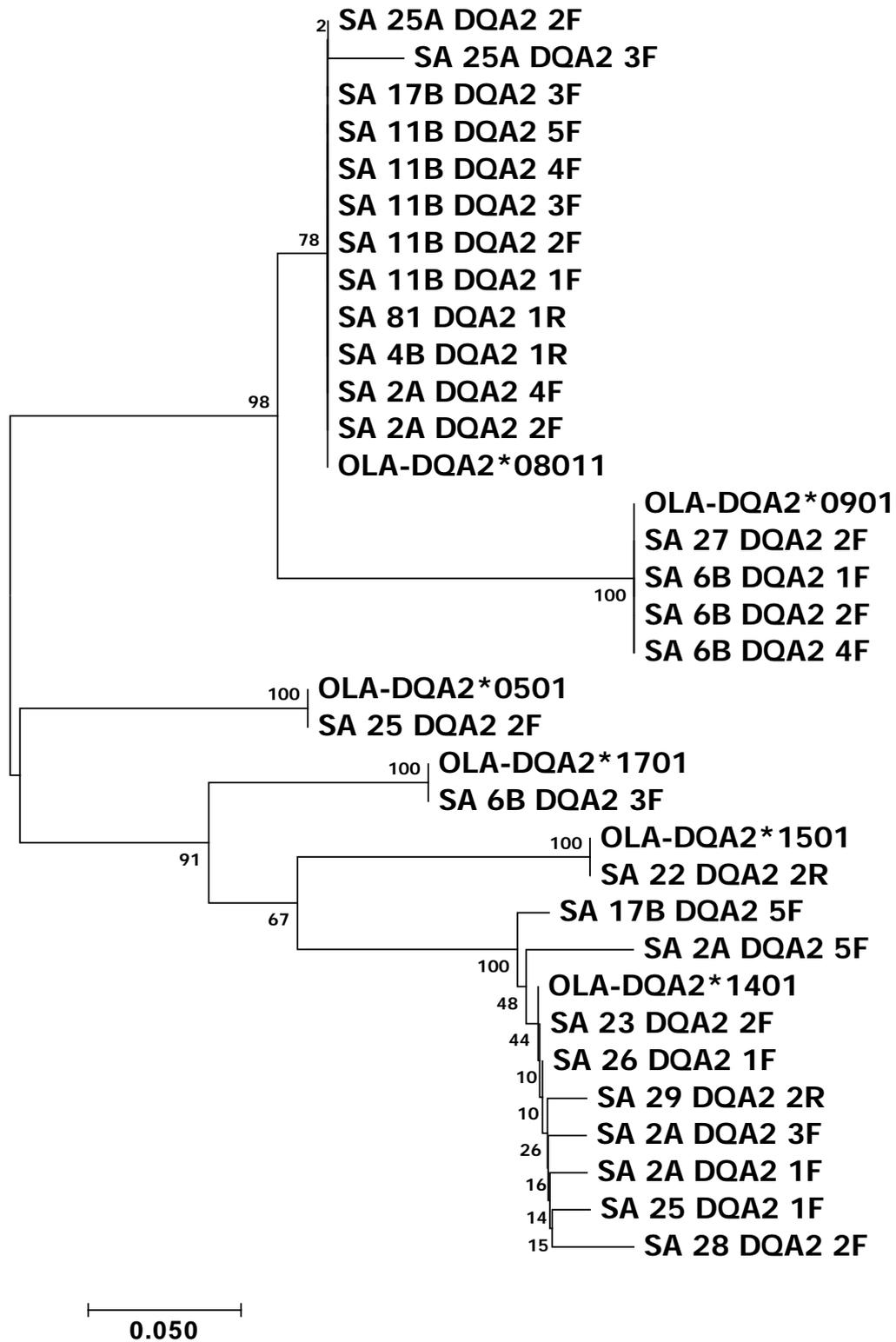
Appendix Figure 18 Re-constructed Neighbour Joining Tree (NJT) Model for the functional and unique *DRB1* exon 2 coding sequences in the Djallonke sheep population showing conserved clustering characteristic as the ML JJT tree Model (chapter 4 Figure 4.2)



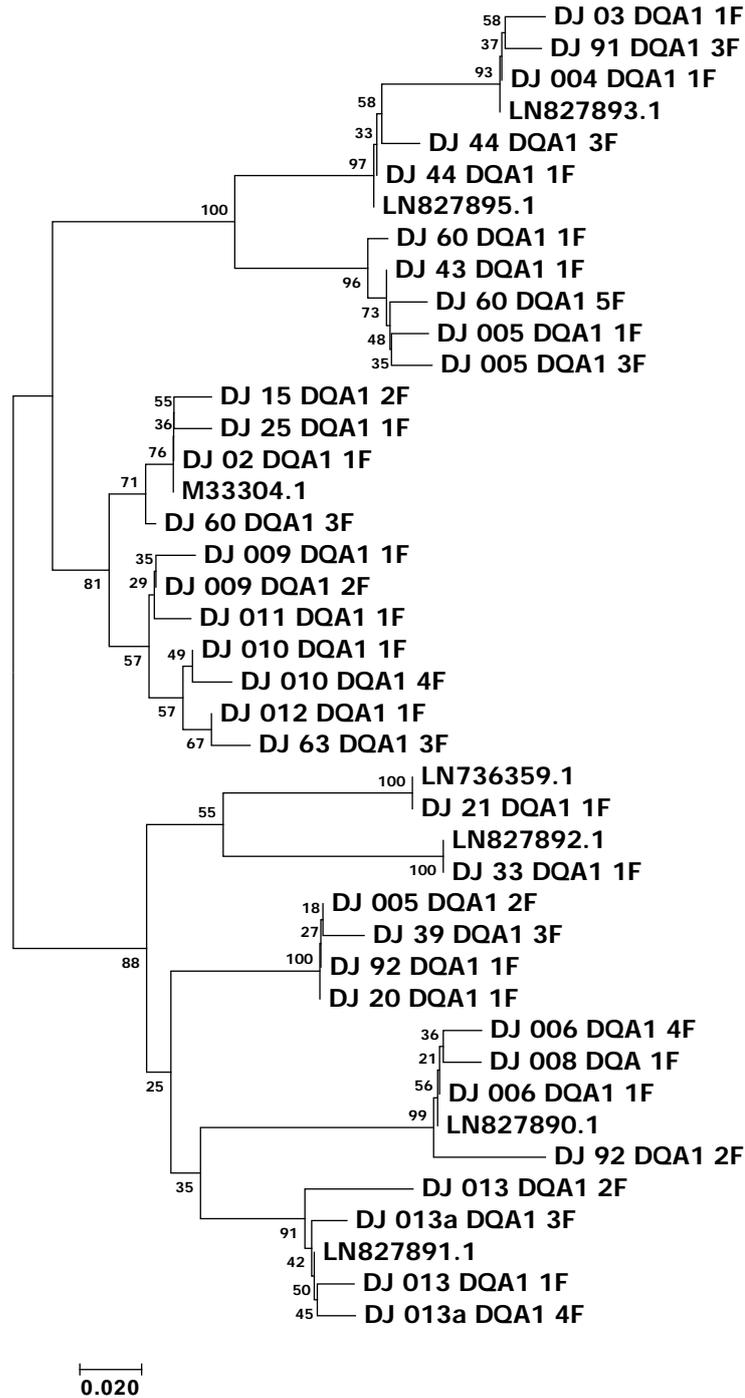
**Appendix Figure 19 Re-constructed Neighbour Joining Tree (NJT) Model for the functional and unique *DRB1* exon 2 coding sequences in the Sahelian sheep population showing conserved clustering characteristic as the ML JJT tree Model (chapter 4 Figure 4.3)**



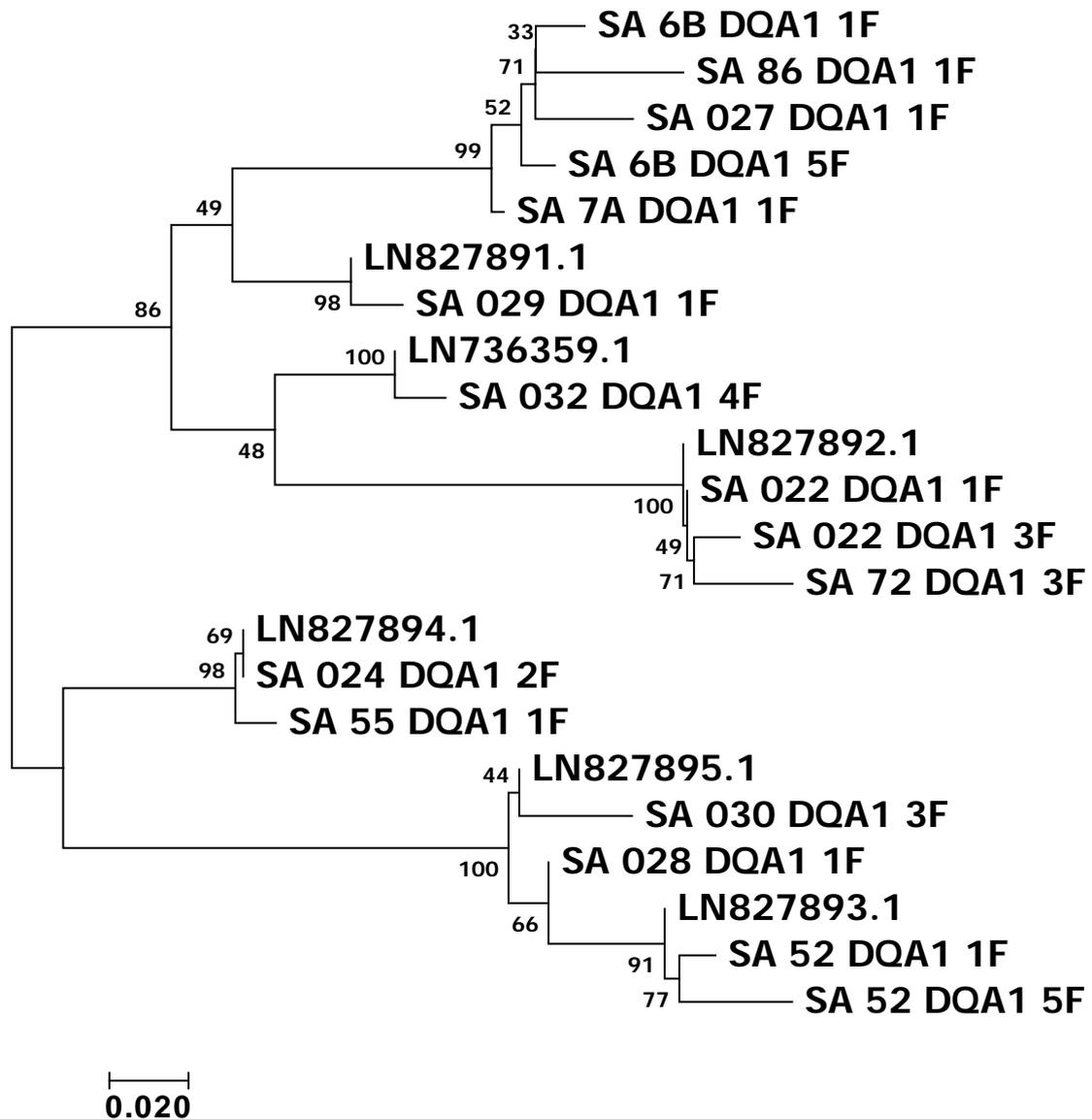
**Appendix Figure 20 Re-constructed Neighbour Joining Tree (NJT) Model for the functional and unique *DQA2* exon 2 coding sequences in the Djallonke sheep population showing conserved clustering characteristic as the ML JJT tree Model (chapter 4 Figure 4.5)**



**Appendix Figure 21** Re-constructed Neighbour Joining Tree (NJT) Model for the functional and unique *DQA2* exon 2 coding sequences in the Sahelian sheep population showing conserved clustering characteristic as the ML JJT tree Model (chapter 4 Figure 4.6)



**Appendix Figure 22 Re-constructed Neighbour Joining Tree (NJT) Model for the functional and unique *DQA1* exon 2 coding sequences in the Djallonke sheep population showing conserved clustering characteristic as the ML JTT tree Model (chapter 4 Figure 4.8)**



**Appendix Figure 23 Re-constructed Neighbour Joining Tree (NJT) Model for the functional and unique *DQA2* exon 2 coding sequences in the Sahelian sheep population showing conserved clustering characteristic as the ML JJT tree Model (chapter 4 Figure 4.9)**

## Appendix E    GenBank accessions for MHC class II genes generated in this study

**Appendix Table E 1 GenBank accession for first batch of 72 *DRB1* alleles for Djallonke and Sahelian sheep breeds**

	Accession		Accession		Accession
<b>1</b>	<b>MF682546</b>	25	<b>MF682570</b>	49	<b>MF682594</b>
<b>2</b>	<b>MF682547</b>	26	<b>MF682571</b>	50	<b>MF682595</b>
<b>3</b>	<b>MF682548</b>	27	<b>MF682572</b>	51	<b>MF682596</b>
<b>4</b>	<b>MF682549</b>	28	<b>MF682573</b>	52	<b>MF682597</b>
<b>5</b>	<b>MF682550</b>	29	<b>MF682574</b>	53	<b>MF682598</b>
<b>6</b>	<b>MF682551</b>	30	<b>MF682575</b>	54	<b>MF682599</b>
<b>7</b>	<b>MF682552</b>	31	<b>MF682576</b>	55	<b>MF682600</b>
<b>8</b>	<b>MF682553</b>	32	<b>MF682577</b>	56	<b>MF682601</b>
<b>9</b>	<b>MF682554</b>	33	<b>MF682578</b>	57	<b>MF682602</b>
<b>10</b>	<b>MF682555</b>	34	<b>MF682579</b>	58	<b>MF682603</b>
<b>11</b>	<b>MF682556</b>	35	<b>MF682580</b>	59	<b>MF682604</b>

---

<b>12</b>	<b>MF682557</b>	36	<b>MF682581</b>	60	<b>MF682605</b>
<hr/>					
<b>13</b>	<b>MF682558</b>	37	<b>MF682582</b>	61	<b>MF682606</b>
<hr/>					
<b>14</b>	<b>MF682559</b>	38	<b>MF682583</b>	62	<b>MF682607</b>
<hr/>					
<b>15</b>	<b>MF682560</b>	39	<b>MF682584</b>	63	<b>MF682608</b>
<hr/>					
<b>16</b>	<b>MF682561</b>	40	<b>MF682585</b>	64	<b>MF682609</b>
<hr/>					
<b>17</b>	<b>MF682562</b>	41	<b>MF682586</b>	65	<b>MF682610</b>
<hr/>					
<b>18</b>	<b>MF682563</b>	42	<b>MF682587</b>	66	<b>MF682611</b>
<hr/>					
<b>19</b>	<b>MF682564</b>	43	<b>MF682588</b>	67	<b>MF682612</b>
<hr/>					
<b>20</b>	<b>MF682565</b>	44	<b>MF682589</b>	68	<b>MF682613</b>
<hr/>					
<b>21</b>	<b>MF682566</b>	45	<b>MF682590</b>	69	<b>MF682614</b>
<hr/>					
<b>22</b>	<b>MF682567</b>	46	<b>MF682591</b>	70	<b>MF682615</b>
<hr/>					
<b>23</b>	<b>MF682568</b>	47	<b>MF682592</b>	71	<b>MF682616</b>
<hr/>					
<b>24</b>	<b>MF682569</b>	48	<b>MF682593</b>	72	<b>MF682617</b>

---

**Appendix Table E 2 GenBank accession for second batch of 81 *DRBI* alleles for Djallonke and Sahelian sheep breeds**

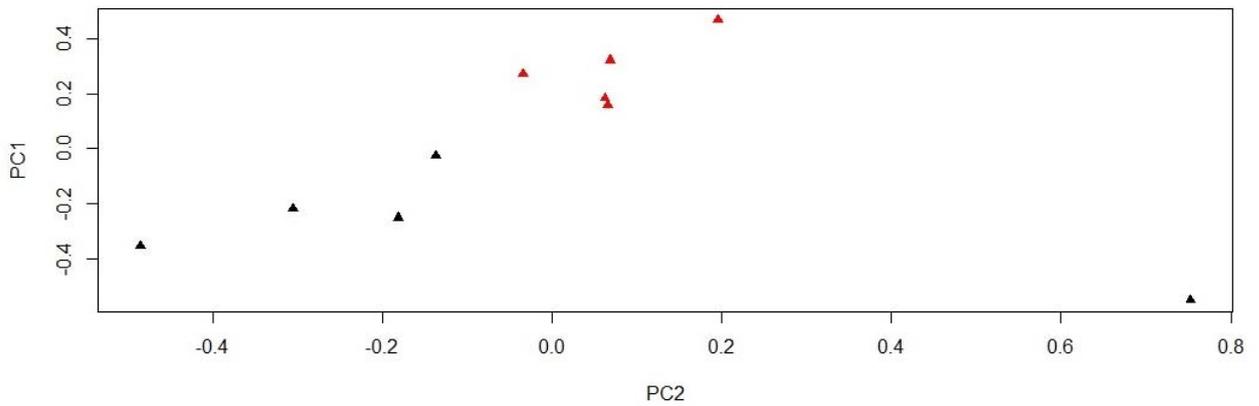
	<b>Accession</b>		<b>Accession</b>		<b>Accession</b>
<b>1</b>	<b>MG000506</b>	28	<b>MG000533</b>	55	<b>MG000560</b>
<b>2</b>	<b>MG000507</b>	29	<b>MG000534</b>	56	<b>MG000561</b>
<b>3</b>	<b>MG000508</b>	30	<b>MG000535</b>	57	<b>MG000562</b>
<b>4</b>	<b>MG000509</b>	31	<b>MG000536</b>	58	<b>MG000563</b>
<b>5</b>	<b>MG000510</b>	32	<b>MG000537</b>	59	<b>MG000564</b>
<b>6</b>	<b>MG000511</b>	33	<b>MG000538</b>	60	<b>MG000565</b>
<b>7</b>	<b>MG000512</b>	34	<b>MG000539</b>	61	<b>MG000566</b>
<b>8</b>	<b>MG000513</b>	35	<b>MG000540</b>	62	<b>MG000567</b>
<b>9</b>	<b>MG000514</b>	36	<b>MG000541</b>	63	<b>MG000568</b>
<b>10</b>	<b>MG000515</b>	37	<b>MG000542</b>	64	<b>MG000569</b>
<b>11</b>	<b>MG000516</b>	38	<b>MG000543</b>	65	<b>MG000570</b>
<b>12</b>	<b>MG000517</b>	39	<b>MG000544</b>	66	<b>MG000571</b>
<b>13</b>	<b>MG000518</b>	40	<b>MG000545</b>	67	<b>MG000572</b>
<b>14</b>	<b>MG000519</b>	41	<b>MG000546</b>	68	<b>MG000573</b>

---

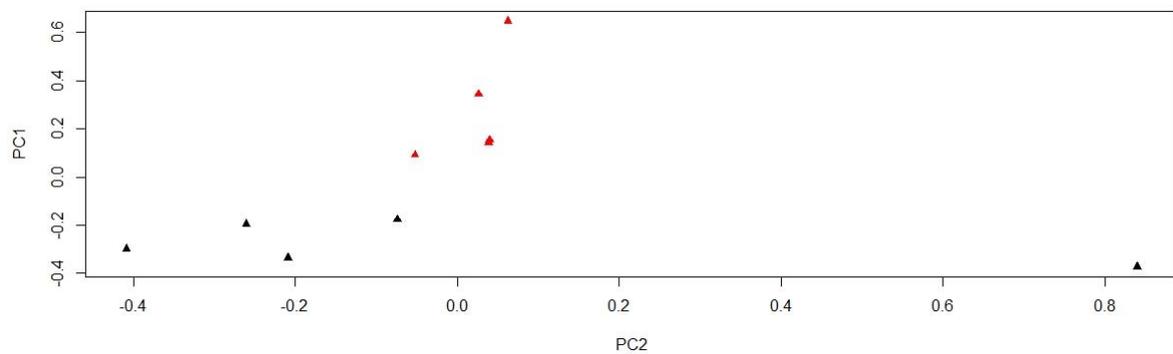
<b>15</b>	<b>MG000520</b>	42	<b>MG000547</b>	69	<b>MG000574</b>
<hr/>					
<b>16</b>	<b>MG000521</b>	43	<b>MG000548</b>	70	<b>MG000575</b>
<hr/>					
<b>17</b>	<b>MG000522</b>	44	<b>MG000549</b>	71	<b>MG000576</b>
<hr/>					
<b>18</b>	<b>MG000523</b>	45	<b>MG000550</b>	72	<b>MG000577</b>
<hr/>					
<b>19</b>	<b>MG000524</b>	46	<b>MG000551</b>	73	<b>MG000578</b>
<hr/>					
<b>20</b>	<b>MG000525</b>	47	<b>MG000552</b>	74	<b>MG000579</b>
<hr/>					
<b>21</b>	<b>MG000526</b>	48	<b>MG000553</b>	75	<b>MG000580</b>
<hr/>					
<b>22</b>	<b>MG000527</b>	49	<b>MG000554</b>	76	<b>MG000581</b>
<hr/>					
<b>23</b>	<b>MG000528</b>	50	<b>MG000555</b>	77	<b>MG000582</b>
<hr/>					
<b>24</b>	<b>MG000529</b>	51	<b>MG000556</b>	78	<b>MG000583</b>
<hr/>					
<b>25</b>	<b>MG000530</b>	52	<b>MG000557</b>	79	<b>MG000584</b>
<hr/>					
<b>26</b>	<b>MG000531</b>	53	<b>MG000558</b>	80	<b>MG000585</b>
<hr/>					
<b>27</b>	<b>MG000532</b>	54	<b>MG000559</b>	81	<b>MG000586</b>

---

## Appendix F Principal Component analysis for ten individuals of Djallonke and Sahelian sheep breeds



**Appendix Figure F 1 PCA of chromosome 20 for 5 Djallonke (black) vrs 5 Sahelian sheep (red) shows clear distinct clustering of the two breeds**



**Appendix Figure F 2 PCA of chromosome 23 for 5 Djallonke (black) vrs 5 Sahelian (red) sheep shows clear distinct clustering of the two breeds**

## Appendix G Genomic relationship matrix for all ten individuals of Djallonke and Sahelian sheep (1 to 10)

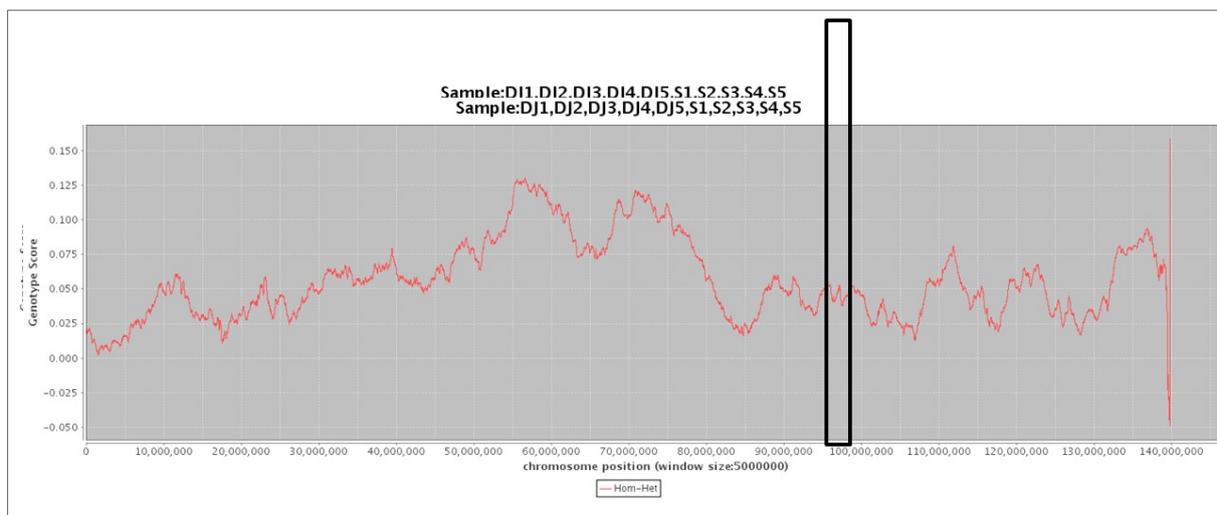
**Appendix Table G 1 GRM conforming that the ten individual animals sampled (1 to 10) for this study were unrelated**

Sample ID	No. of non-missing SNPs	Genetic relationship matrix	
1	1	9.30E+06	1.61E+00
2	1	8.37E+06	-1.99E-01
2	2	9.64E+06	1.65E+00
3	1	7.48E+06	-2.49E-01
3	2	7.72E+06	-2.51E-01
3	3	8.52E+06	1.55E+00
4	1	8.04E+06	-2.05E-01
4	2	8.28E+06	-2.23E-01
4	3	7.41E+06	-2.51E-01
4	4	9.19E+06	1.60E+00
5	1	7.87E+06	-2.35E-01
5	2	8.12E+06	-2.41E-01
5	3	7.28E+06	-1.31E-01
5	4	7.78E+06	-2.23E-01
5	5	8.99E+06	1.54E+00
6	1	5.27E+06	-2.71E-01
6	2	5.39E+06	-2.79E-01
6	3	4.95E+06	-2.72E-01
6	4	5.22E+06	-2.76E-01
6	5	5.14E+06	-2.83E-01
6	6	7.79E+06	1.68E+00
7	1	5.26E+06	-3.23E-01
7	2	5.38E+06	-3.16E-01
7	3	4.94E+06	-3.10E-01

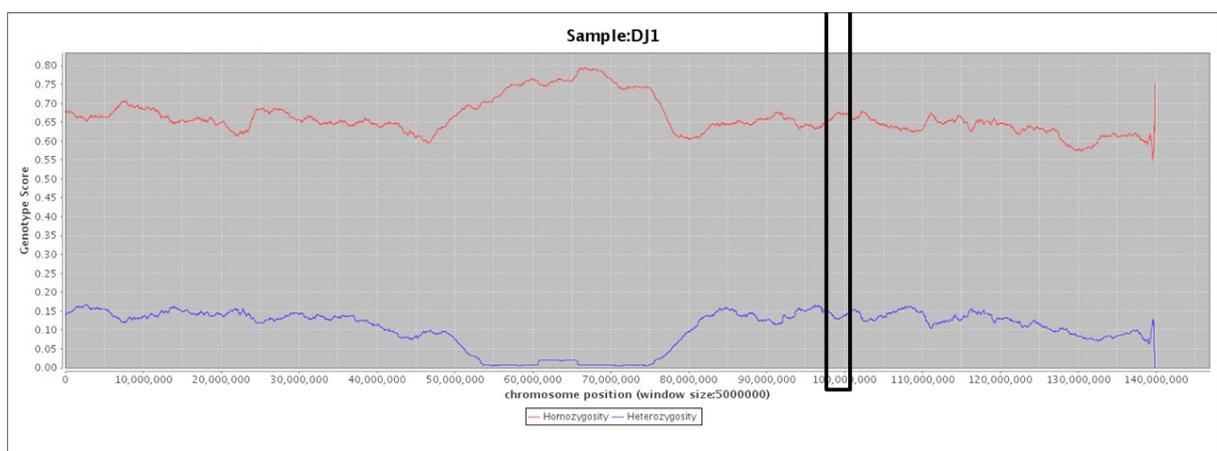
7	4	5.21E+06	-3.17E-01
7	5	5.13E+06	-3.14E-01
7	6	6.86E+06	-2.55E-01
7	7	7.75E+06	1.53E+00
8	1	4.20E+06	-3.58E-01
8	2	4.25E+06	-4.07E-01
8	3	3.98E+06	-2.59E-01
8	4	4.17E+06	-3.31E-01
8	5	4.09E+06	-3.14E-01
8	6	5.17E+06	-2.72E-01
8	7	5.13E+06	-1.57E-01
8	8	5.71E+06	1.68E+00
9	1	4.79E+06	-3.44E-01
9	2	4.88E+06	-3.62E-01
9	3	4.52E+06	-2.95E-01
9	4	4.75E+06	-3.32E-01
9	5	4.67E+06	-3.22E-01
9	6	6.10E+06	-1.75E-01
9	7	6.07E+06	-7.65E-02
9	8	4.72E+06	-4.83E-02
9	9	6.83E+06	1.52E+00
10	1	4.91E+06	-3.35E-01
10	2	5.01E+06	-3.47E-01
10	3	4.63E+06	-2.92E-01
10	4	4.86E+06	-3.20E-01
10	5	4.80E+06	-3.11E-01
10	6	6.31E+06	-2.68E-01
10	7	6.29E+06	-1.09E-01
10	8	4.80E+06	-4.30E-02
10	9	5.63E+06	-1.40E-01
10	10	7.09E+06	1.57E+00

# Appendix H Detailed chromosomal homozygosity profile map for the Djallonke and Sahelian sheep

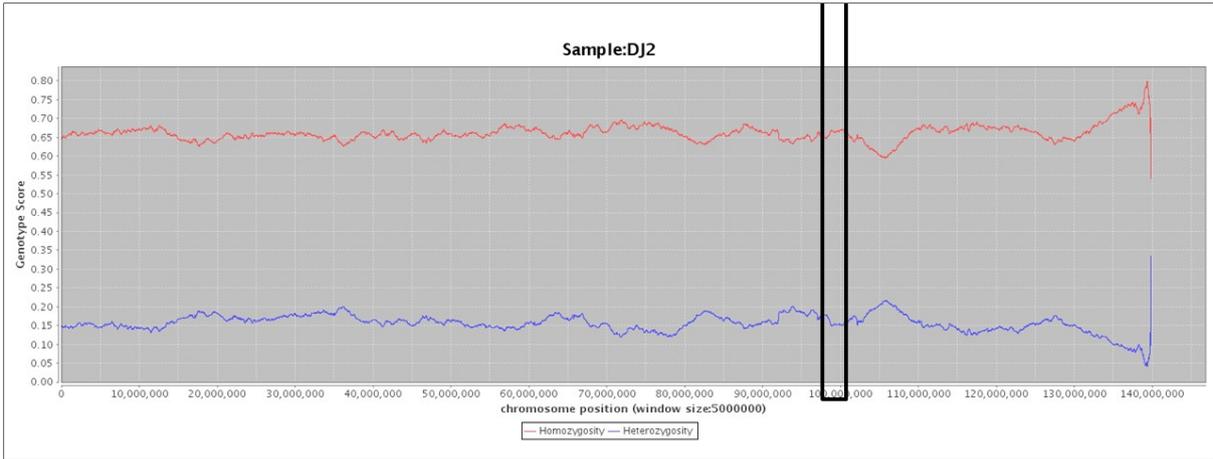
HomHet predictive signal for the Trypanotolerance *CTSS* candidate loci (Djallonke (high ROH) and Sahelian (low ROH)) and HomHet profile of the *CTSS* loci for each of the 10 individual samples. The homhet predictive signal maps is the normalised homozygous to heterozygous variants ratio across the genome for the Djallonke relative to the Sahelian samples.



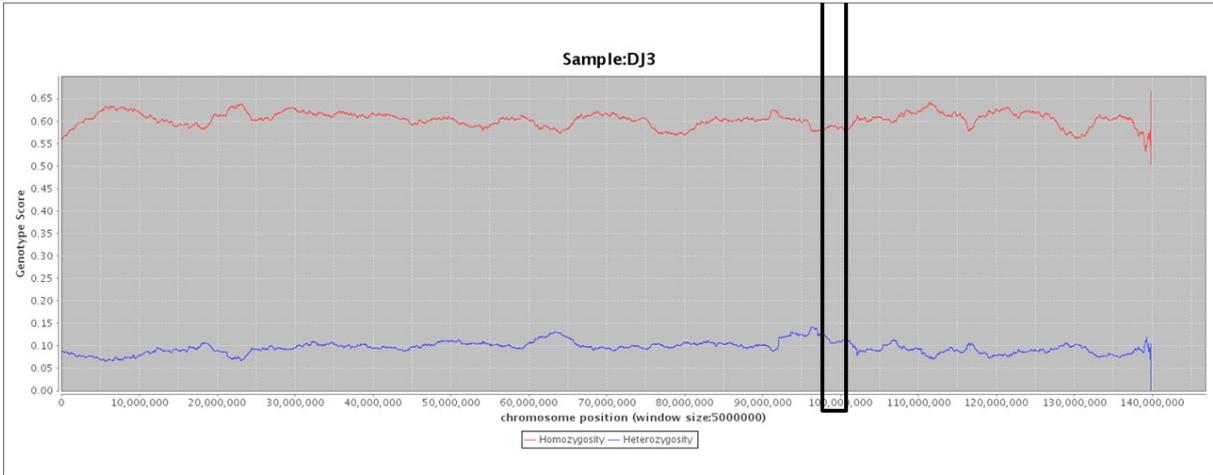
Appendix Figure H 1 HomHet predictive signal for chromosome 1 showing the Trypanotolerance *CTSS* candidate loci



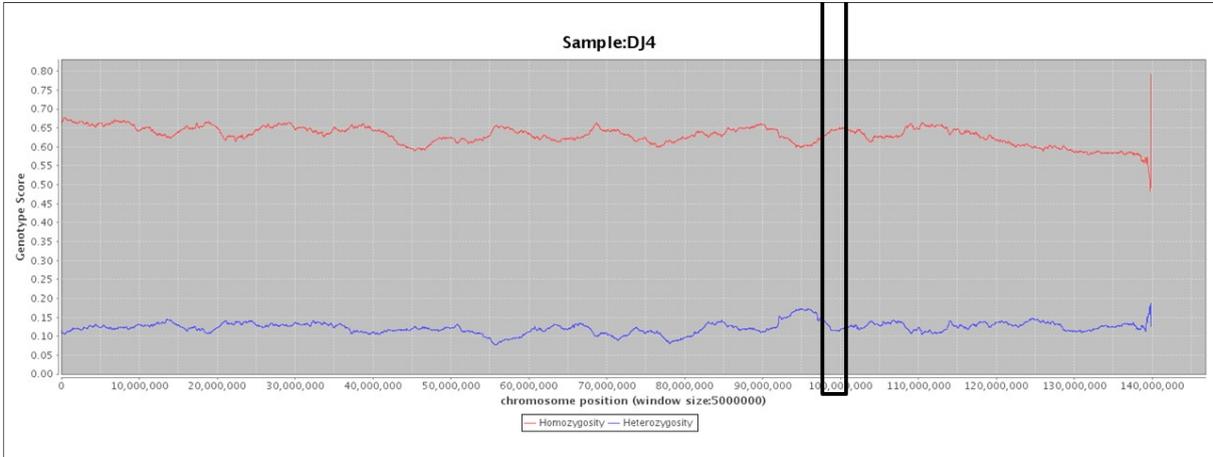
Appendix Figure H 2 Homozygosity profile signal for chromosome 1 of DJ1 showing the Trypanotolerance *CTSS* candidate loci



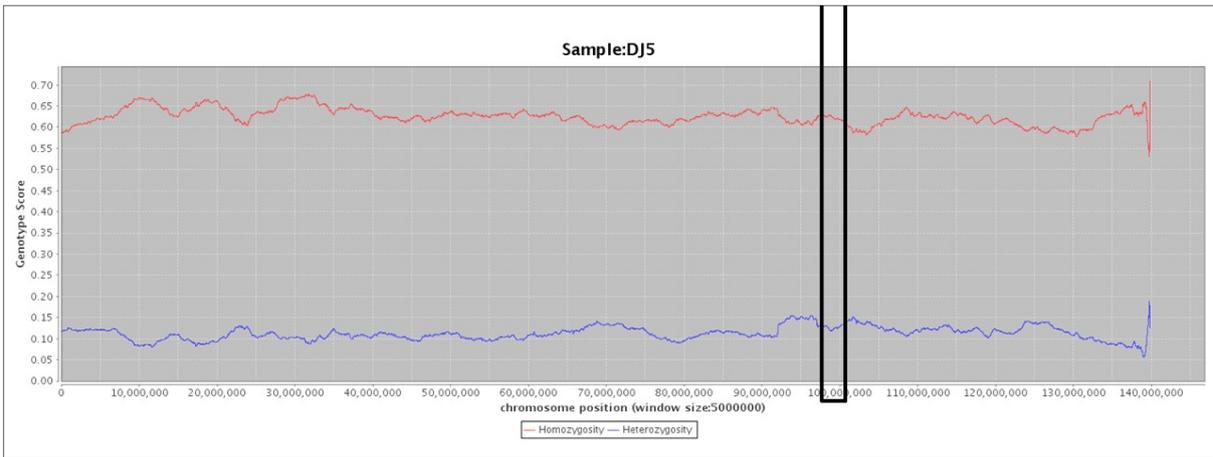
**Appendix Figure H 3 Homozygosity profile signal for chromosome 1 of DJ2 showing the Trypanotolerance *CTSS* candidate loci**



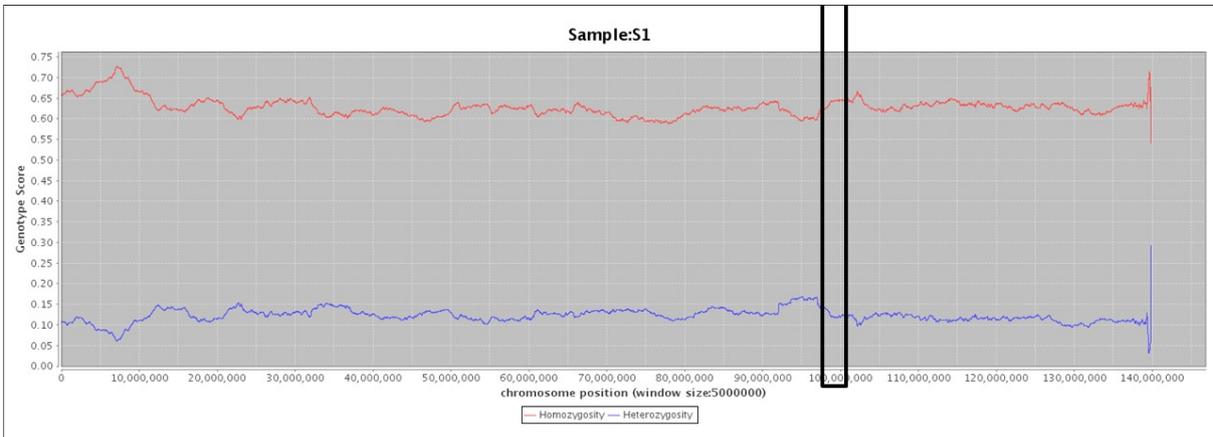
**Appendix Figure H 4 Homozygosity profile signal for chromosome 1 of DJ3 showing the Trypanotolerance *CTSS* candidate loci**



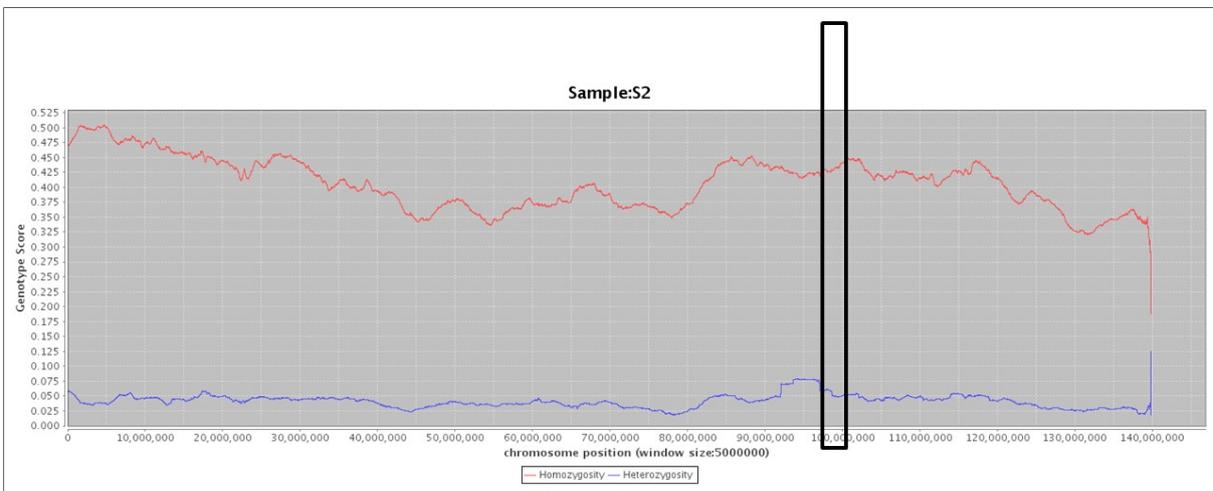
**Appendix Figure H 5 Homozygosity profile signal for chromosome 1 of DJ4 showing the Trypanotolerance *CTSS* candidate loci**



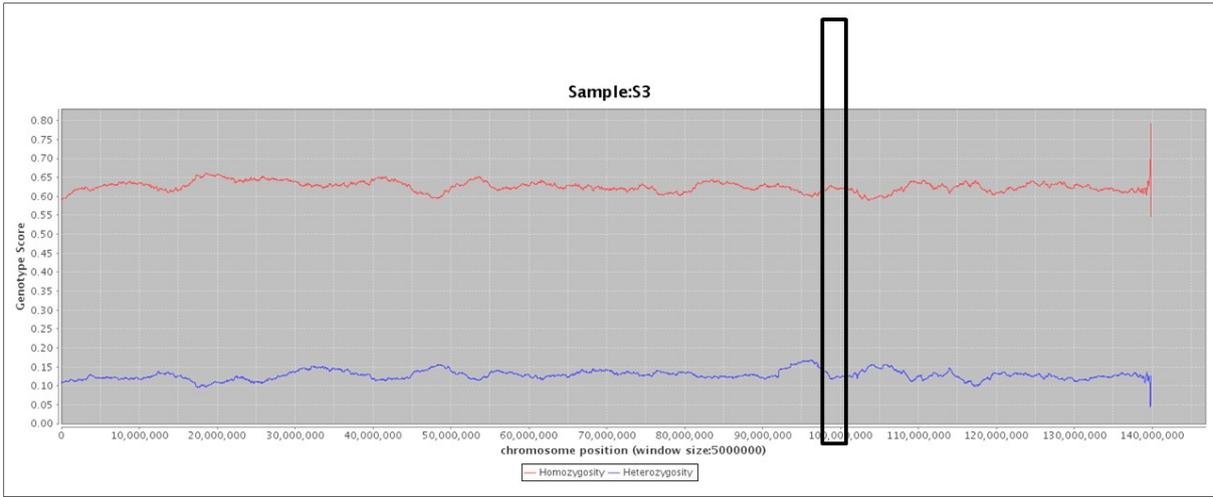
**Appendix Figure H 6 Homozygosity profile signal for chromosome 1 of DJ5 showing the Trypanotolerance *CTSS* candidate loci**



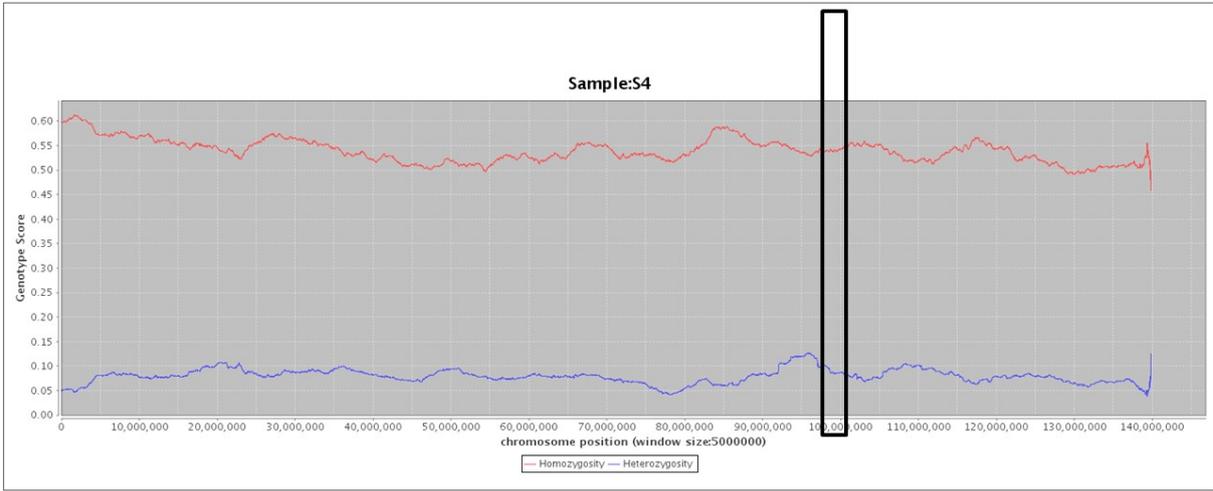
**Appendix Figure H 7 Homozygosity profile signal for chromosome 1 of S1 showing the Trypanotolerance *CTSS* candidate loci**



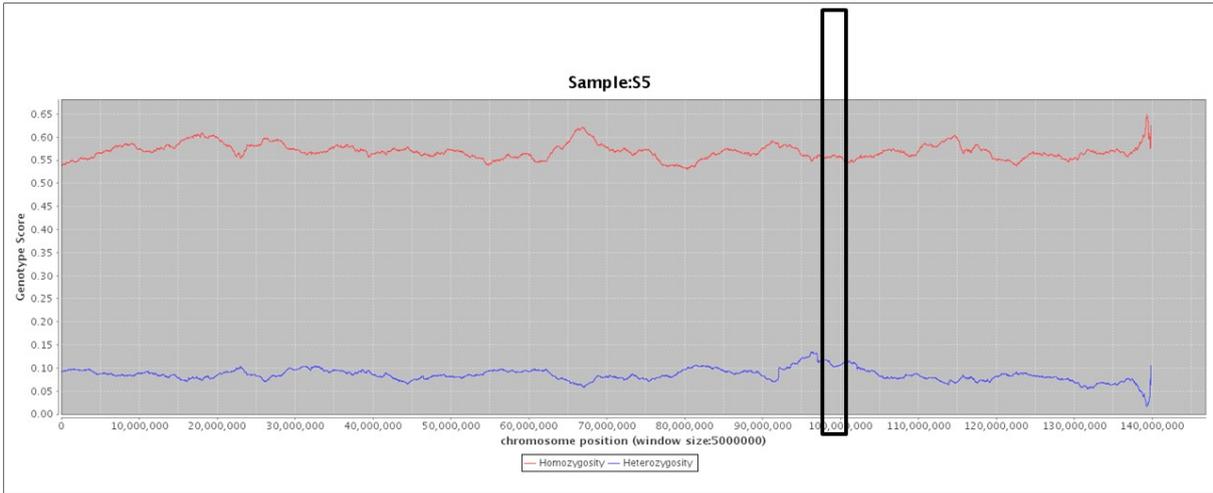
**Appendix Figure H 8 Homozygosity profile signal for chromosome 1 of S2 showing the Trypanotolerance *CTSS* candidate loci**



**Appendix Figure H 9 Homozygosity profile signal for chromosome 1 of S3 showing the Trypanotolerance *CTSS* candidate loci**

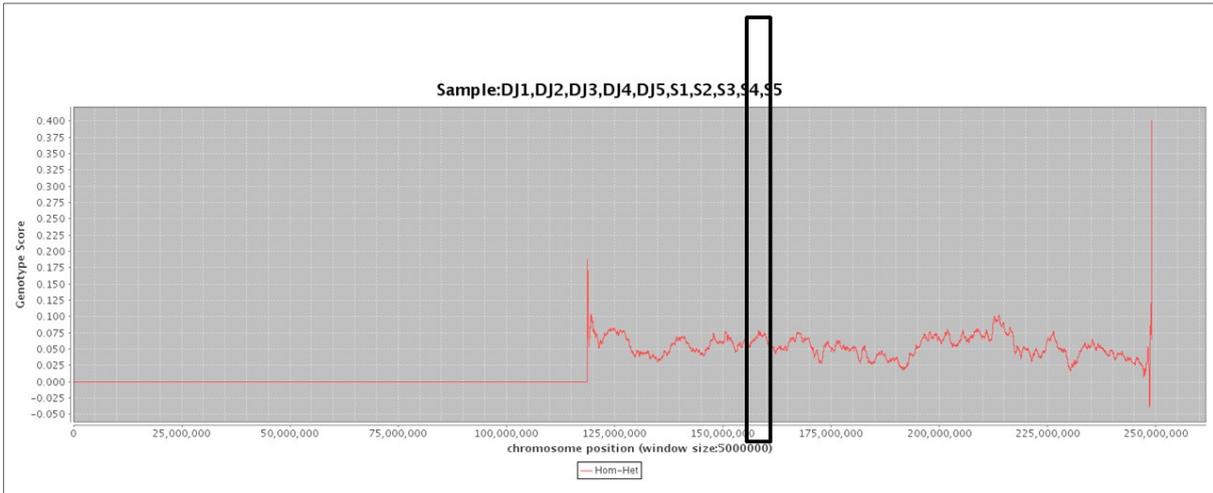


**Appendix Figure H 10 Homozygosity profile signal for chromosome 1 of S4 showing the Trypanotolerance *CTSS* candidate loci**

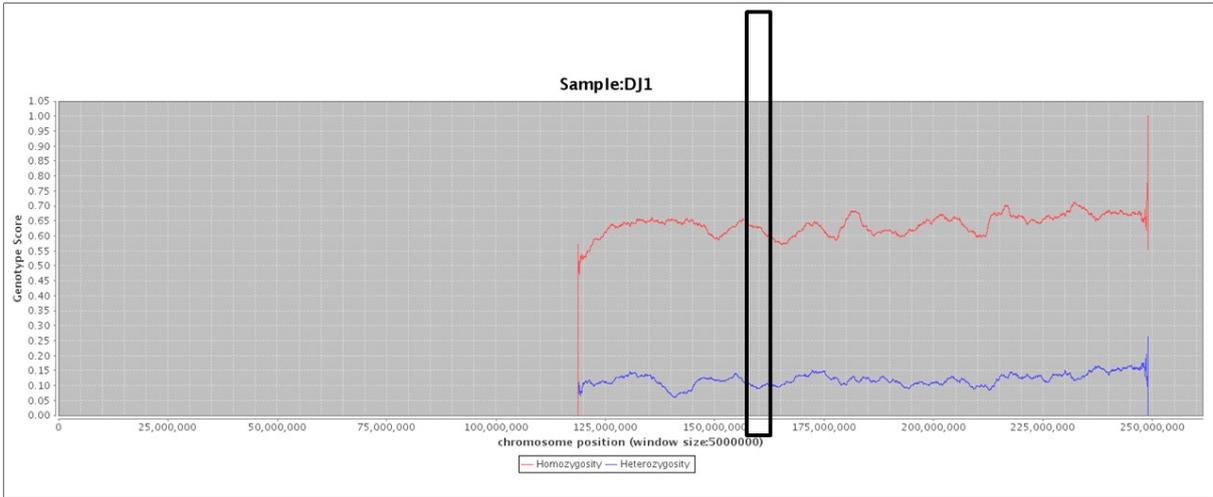


**Appendix Figure H 11 Homozygosity profile signal for chromosome 1 of S5 showing the Trypanotolerance *CTSS* candidate loci**

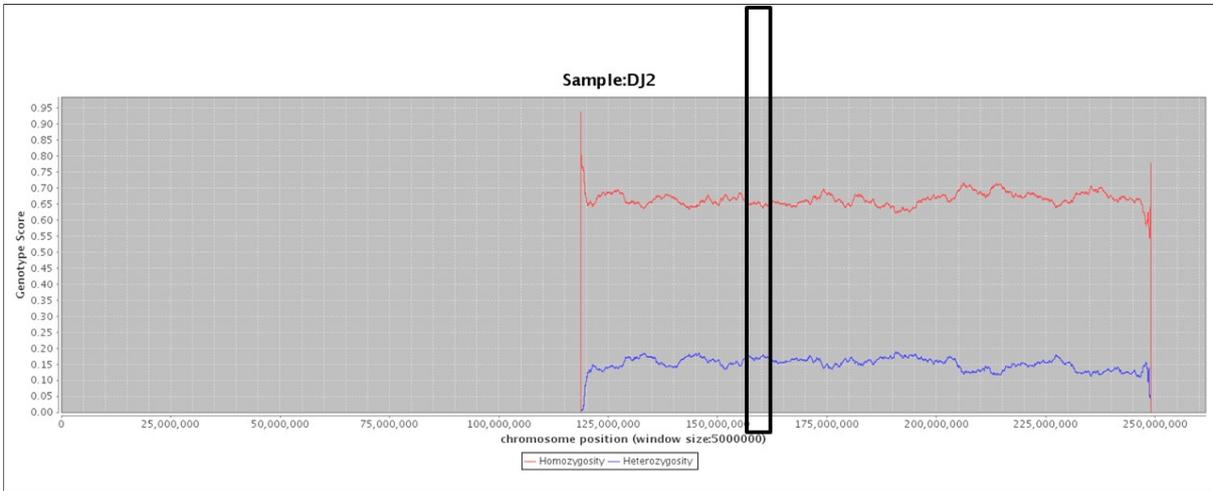
HomHet predictive signal for the Trypanotolerance *ARHGAP15* candidate loci (Djallonke (high ROH) and Sahelian (low ROH)) and HomHet analysis of the loci for all 10 individual samples



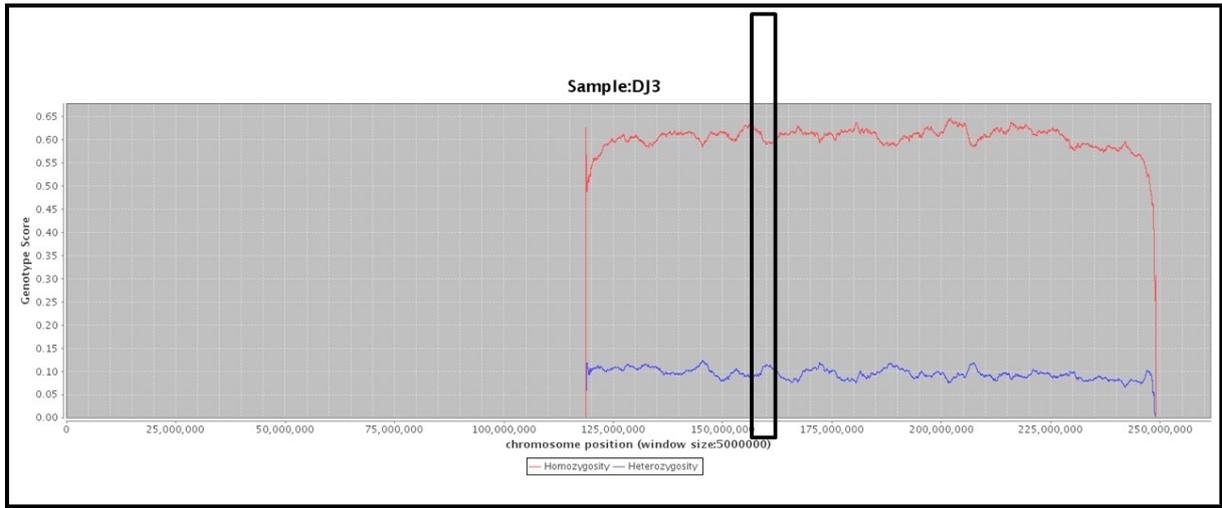
**Appendix Figure H 12 HomHet predictive signal for chromosome 2 showing the Trypanotolerance *ARHGAP15* candidate loci**



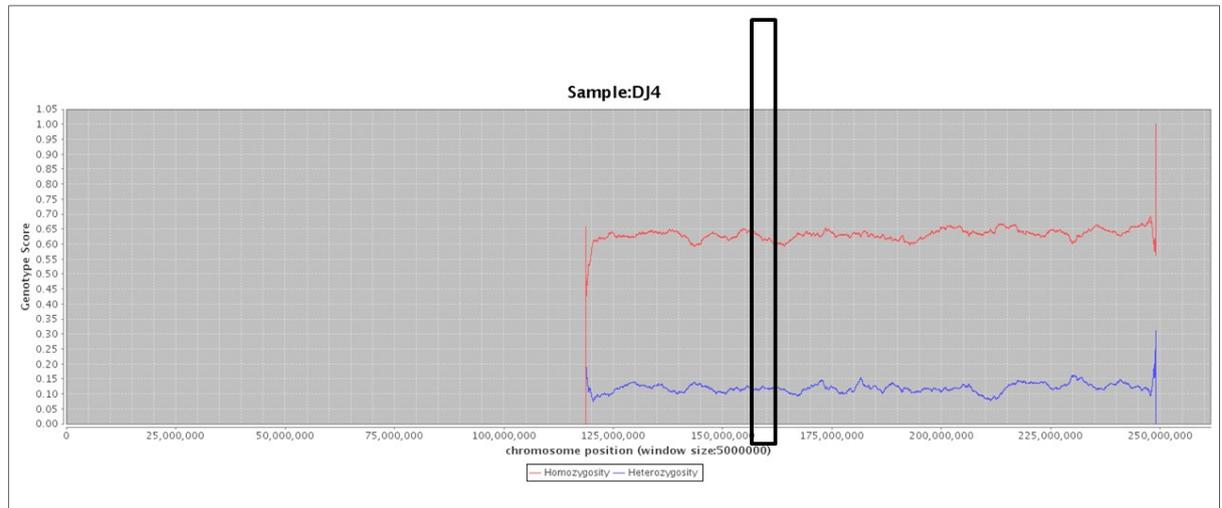
**Appendix Figure H 13 Homozygosity profile signal for chromosome 2 of DJ1 showing the Trypanotolerance *ARHGAP15* candidate loci**



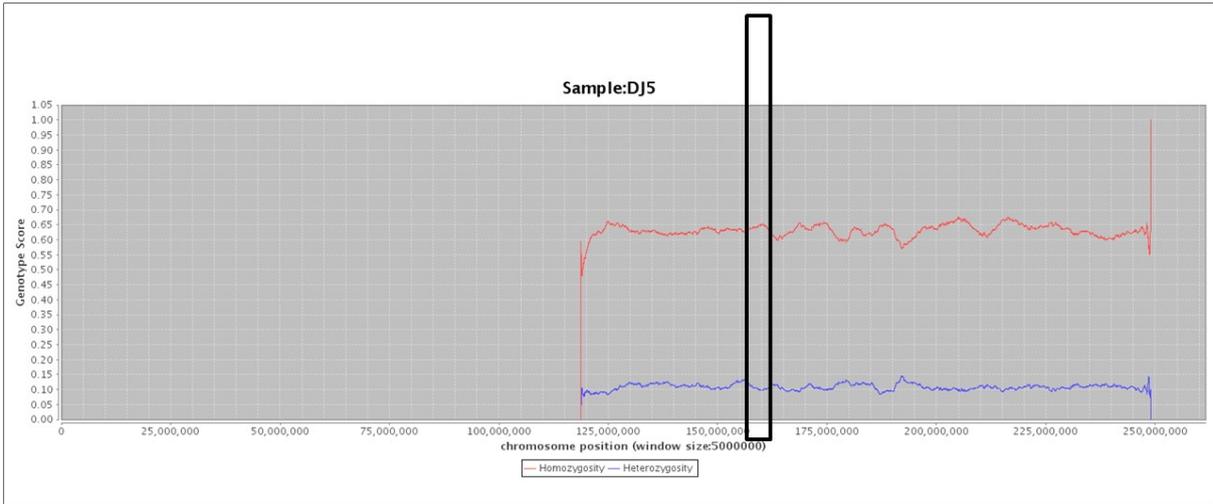
**Appendix Figure H 14 Homozygosity profile signal for chromosome 2 of DJ2 showing the Trypanotolerance *ARHGAP15* candidate loci**



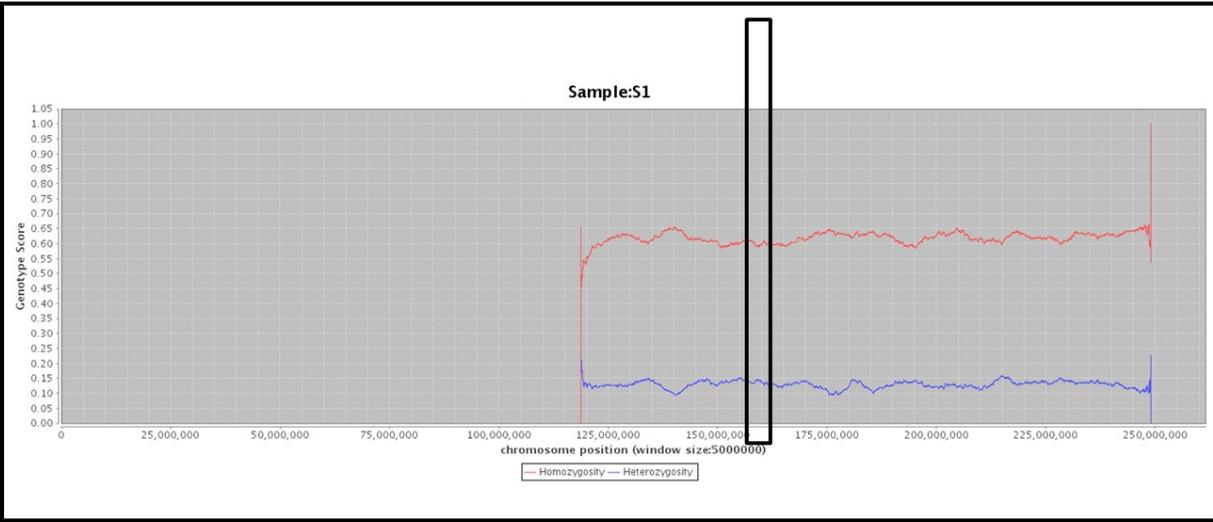
**Appendix Figure H 15 Homozygosity profile signal for chromosome 2 of DJ3 showing the Trypanotolerance *ARHGAP15* candidate loci**



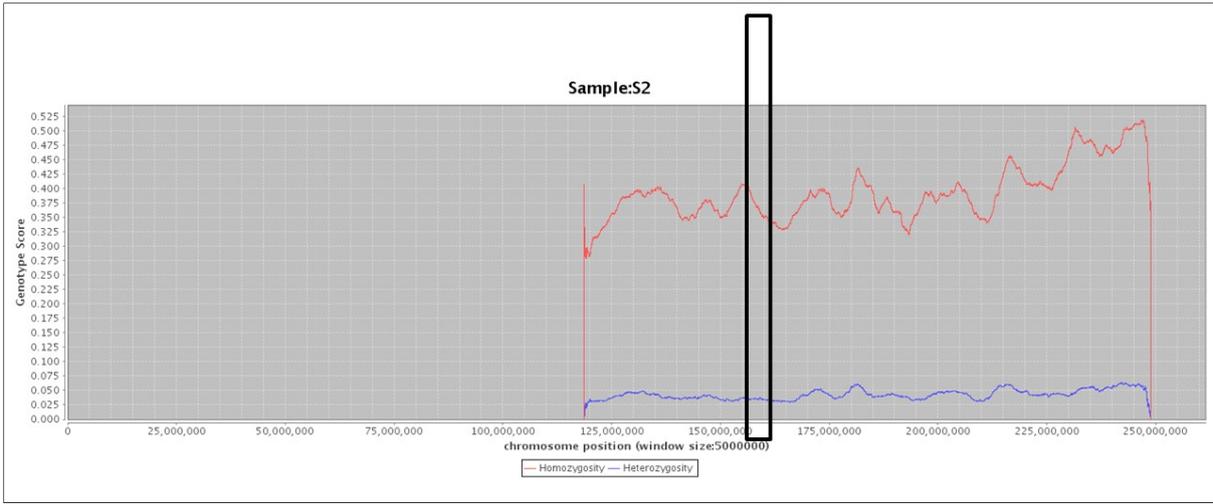
**Appendix Figure H 16 Homozygosity profile signal for chromosome 2 of DJ4 showing the Trypanotolerance *ARHGAP15* candidate loci**



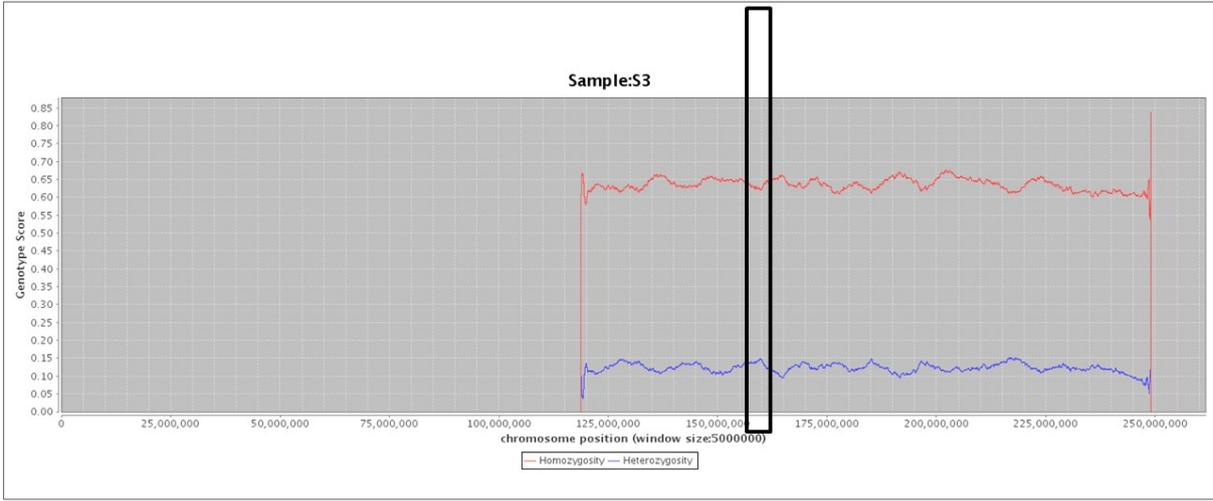
**Appendix Figure H 17 Homozygosity profile signal for chromosome 2 of DJ5 showing the Trypanotolerance *ARHGAP15* candidate loci**



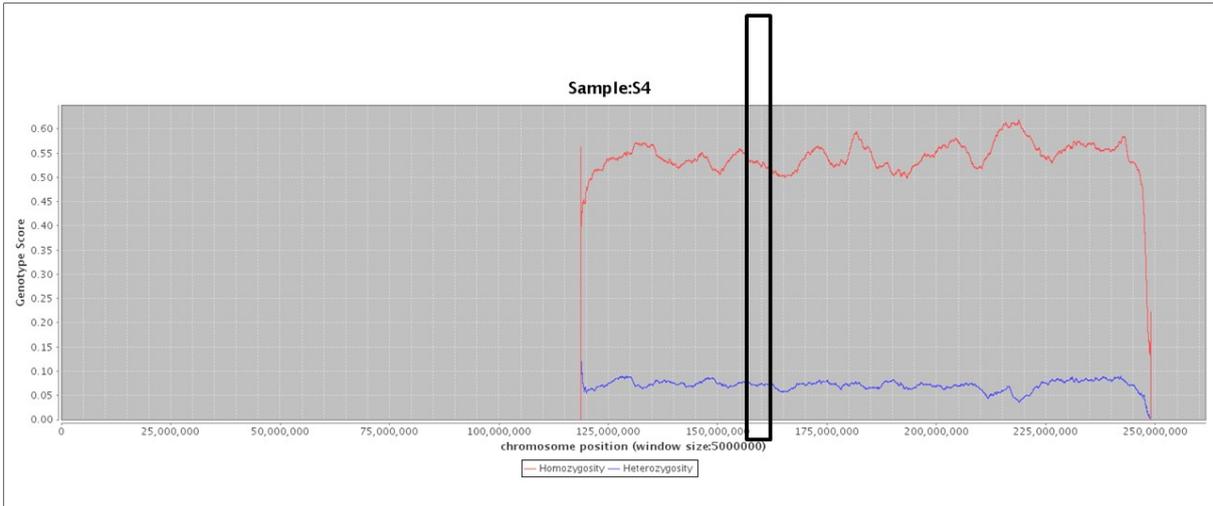
**Appendix Figure H 18 Homozygosity profile signal for chromosome 2 of S1 showing the Trypanotolerance *ARHGAP15* candidate loci**



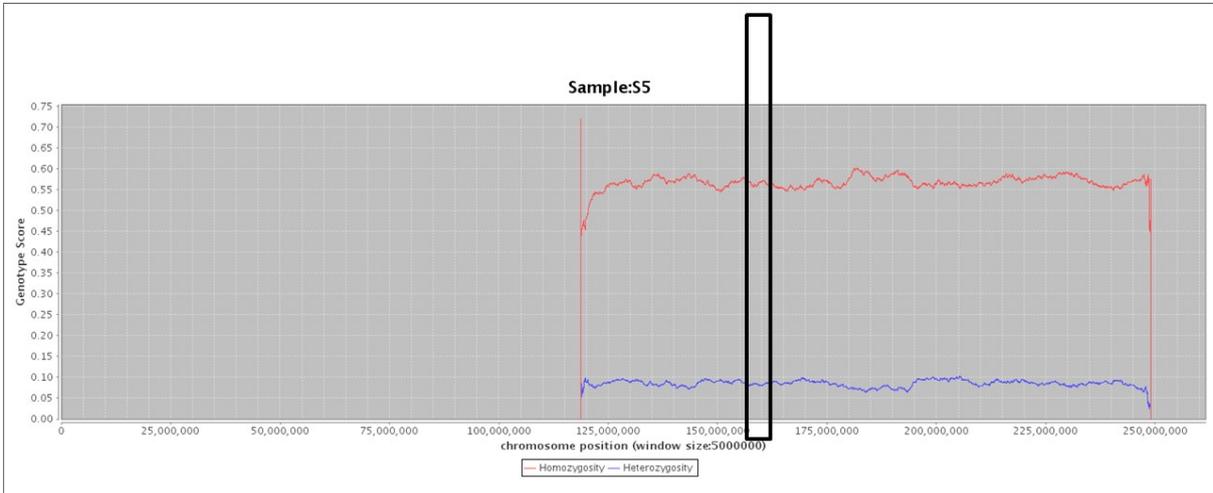
**Appendix Figure H 19 Homozygosity profile signal for chromosome 2 of S2 showing the Trypanotolerance *ARHGAP15* candidate loci**



**Appendix Figure H 20 Homozygosity profile signal for chromosome 2 of S3 showing the Trypanotolerance *ARHGAP15* candidate loci**

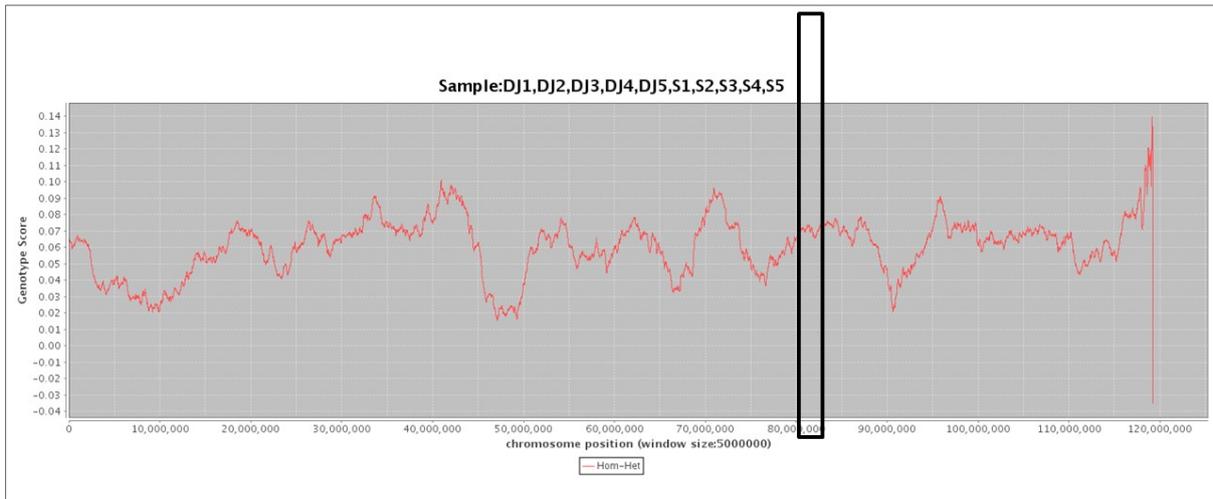


**Appendix Figure H 21 Homozygosity profile signal for chromosome 2 of S4 showing the Trypanotolerance *ARHGAP15* candidate loci**

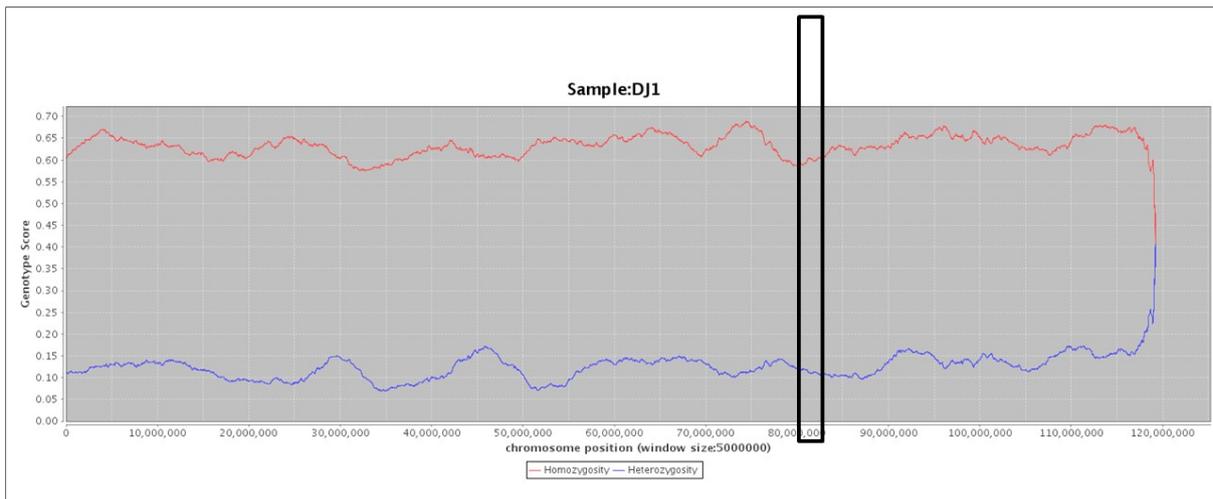


**Appendix Figure H 22 Homozygosity profile signal for chromosome 2 of S5 showing the Trypanotolerance *ARHGAP15* candidate loci**

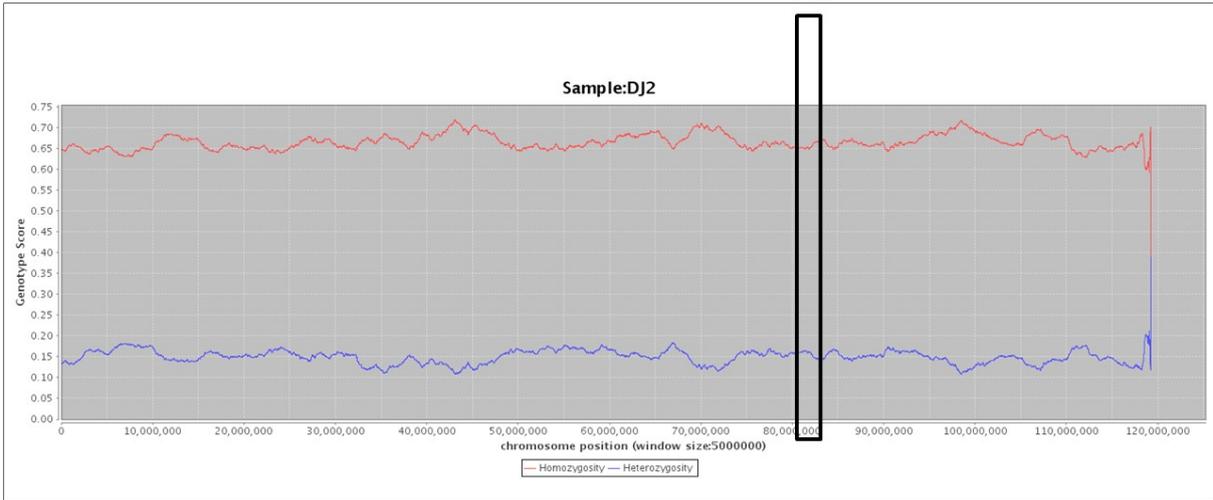
**HomHet predictive signal for the Trypanotolerance *INHBA* candidate loci (Djallonke (high ROH) and Sahelian (low ROH)) and HomHet analysis of the loci for all 10 individual samples**



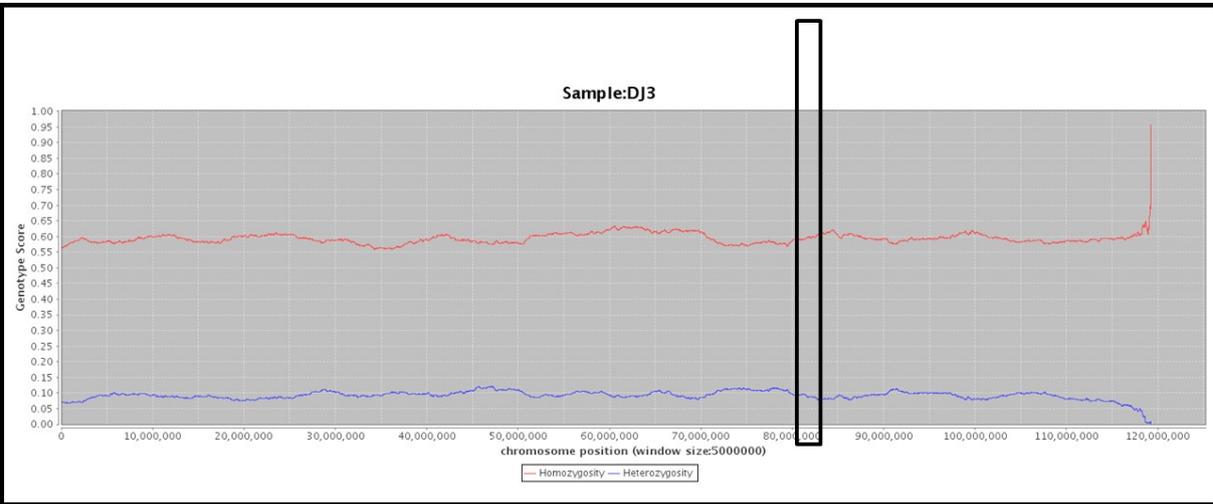
**Appendix Figure H 23 HomHet predictive signal for chromosome 4 showing the Trypanotolerance *INHBA* candidate loci**



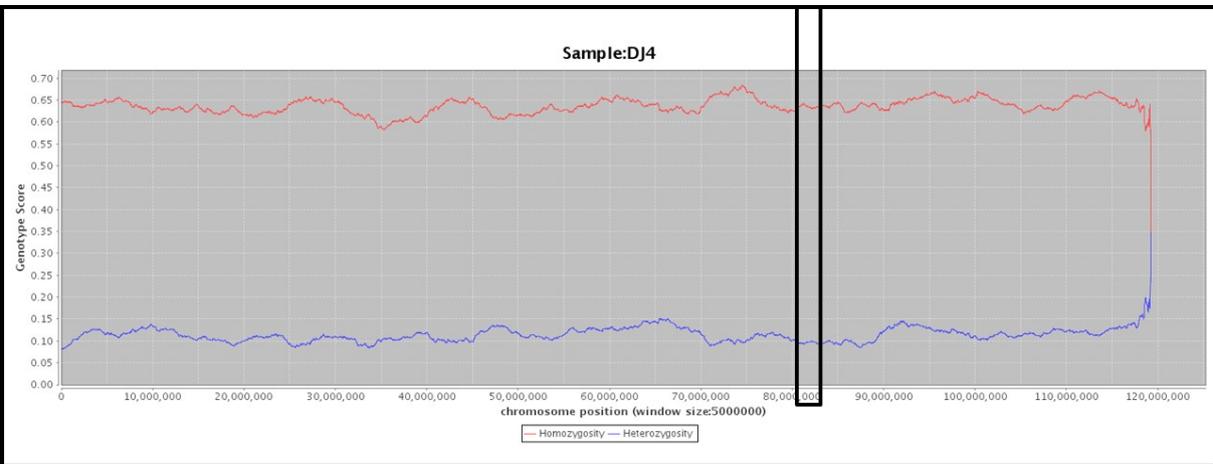
**Appendix Figure H 24 Homozygosity profile signal for chromosome 4 of DJ1 showing the Trypanotolerance *INHBA* candidate loci**



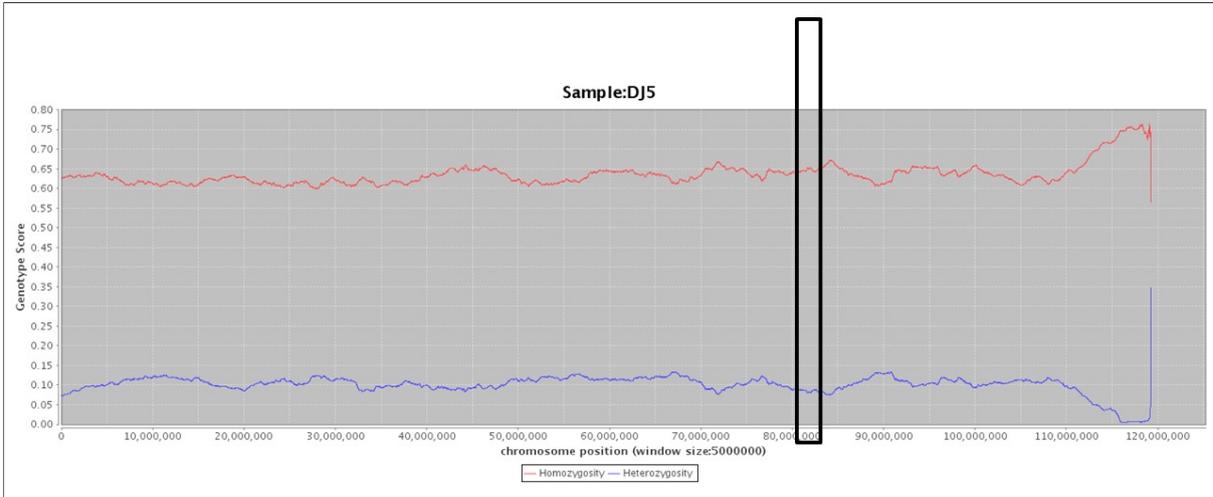
**Appendix Figure H 25 Homozygosity profile signal for chromosome 4 of DJ2 showing the Trypanotolerance *INHBA* candidate loci**



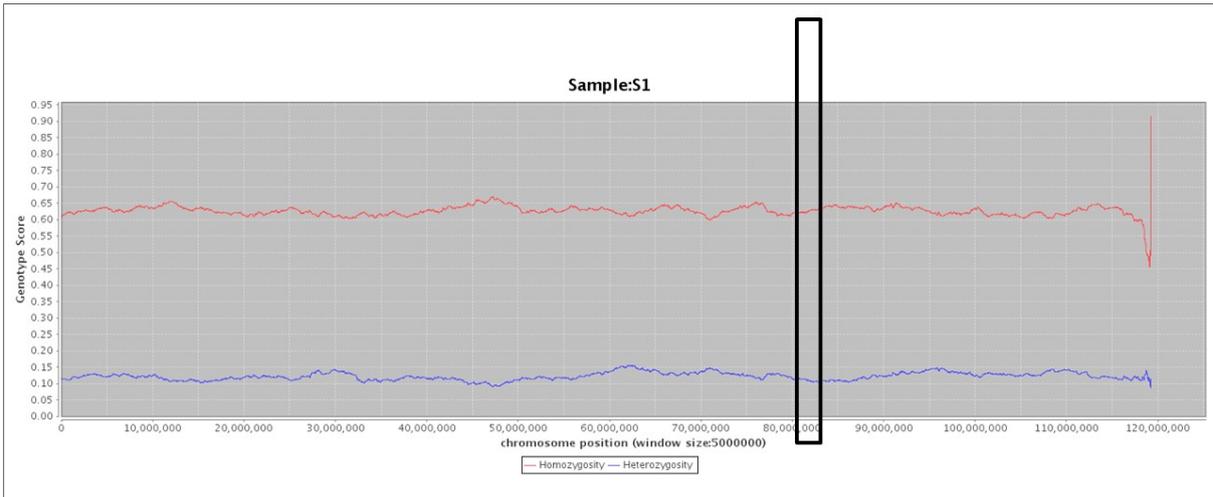
**Appendix Figure H 26 Homozygosity profile signal for chromosome 4 of DJ3 showing the Trypanotolerance *INHBA* candidate loci**



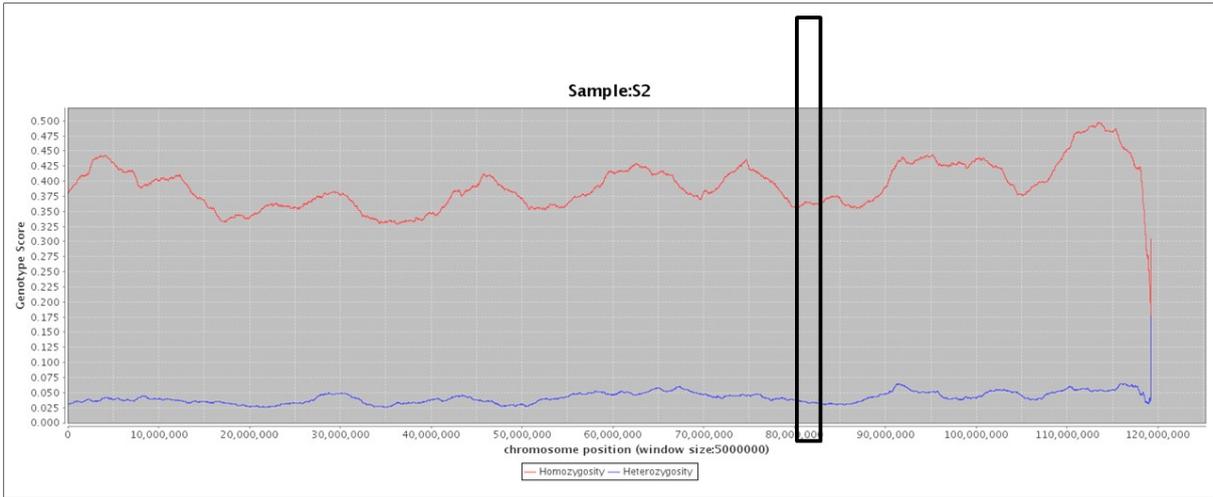
**Appendix Figure H 27 Homozygosity profile signal for chromosome 4 of DJ4 showing the Trypanotolerance *INHBA* candidate loci**



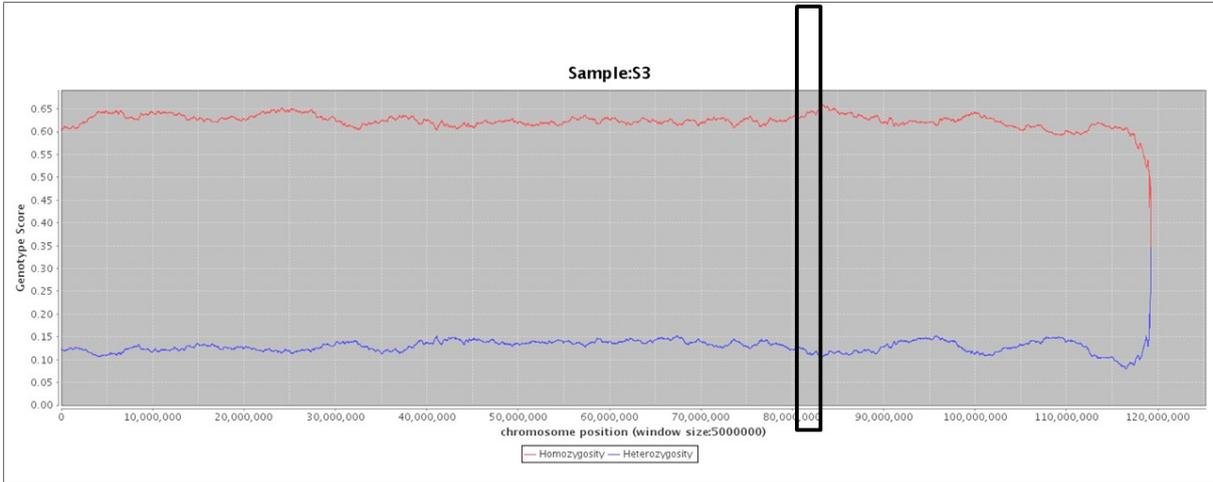
**Appendix Figure H 28 Homozygosity profile signal for chromosome 4 of DJ5 showing the Trypanotolerance *INHBA* candidate loci**



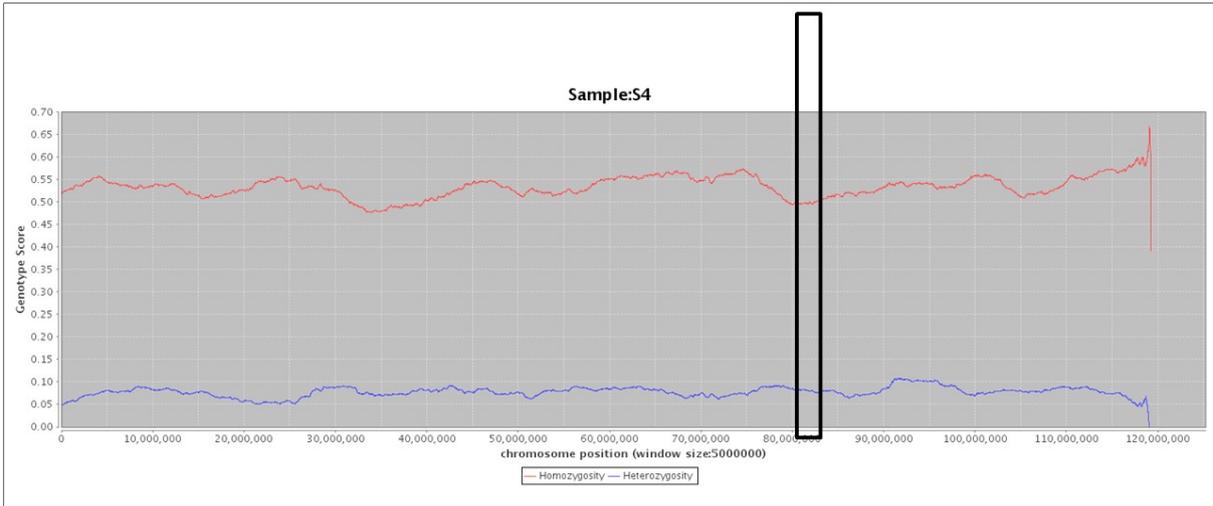
**Appendix Figure H 29 Homozygosity profile signal for chromosome 4 of S1 showing the Trypanotolerance *INHBA* candidate loci**



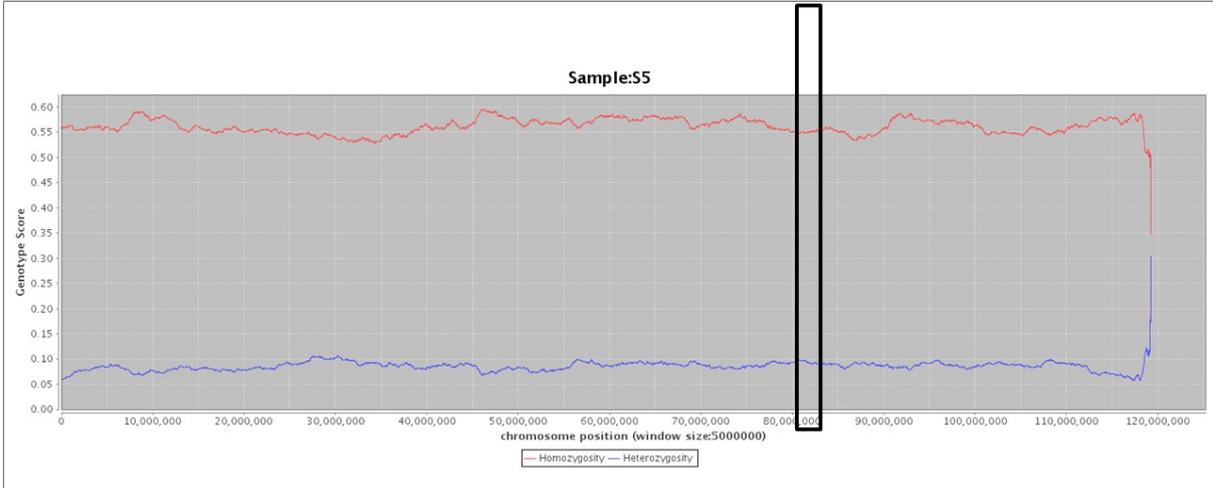
**Appendix Figure H 30 Homozygosity profile signal for chromosome 4 of S2 showing the Trypanotolerance *INHBA* candidate loci**



**Appendix Figure H 31 Homozygosity profile signal for chromosome 4 of S3 showing the Trypanotolerance *INHBA* candidate loci**

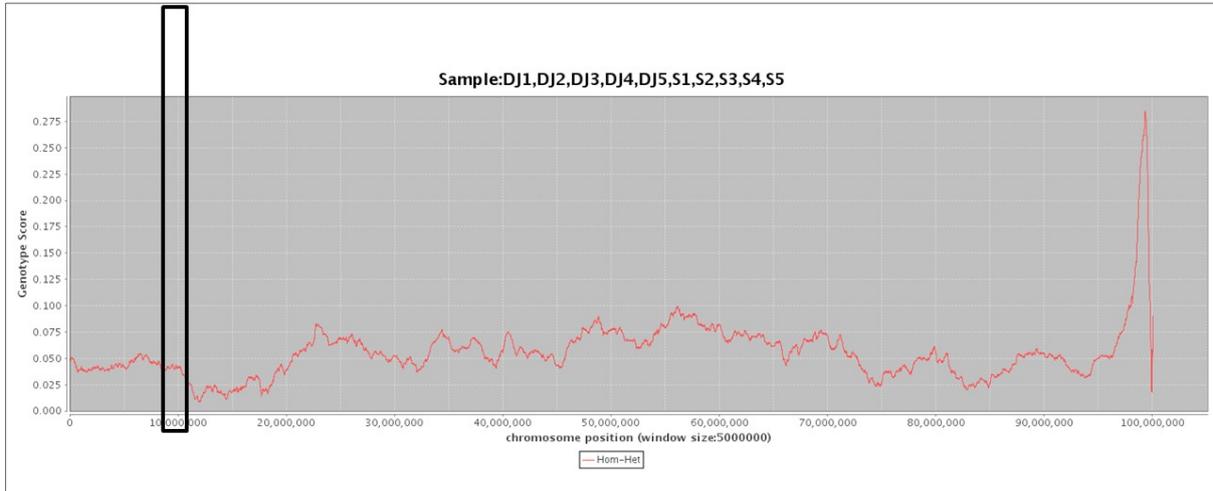


**Appendix Figure H 32 Homozygosity profile signal for chromosome 4 of S4 showing the Trypanotolerance *INHBA* candidate loci**

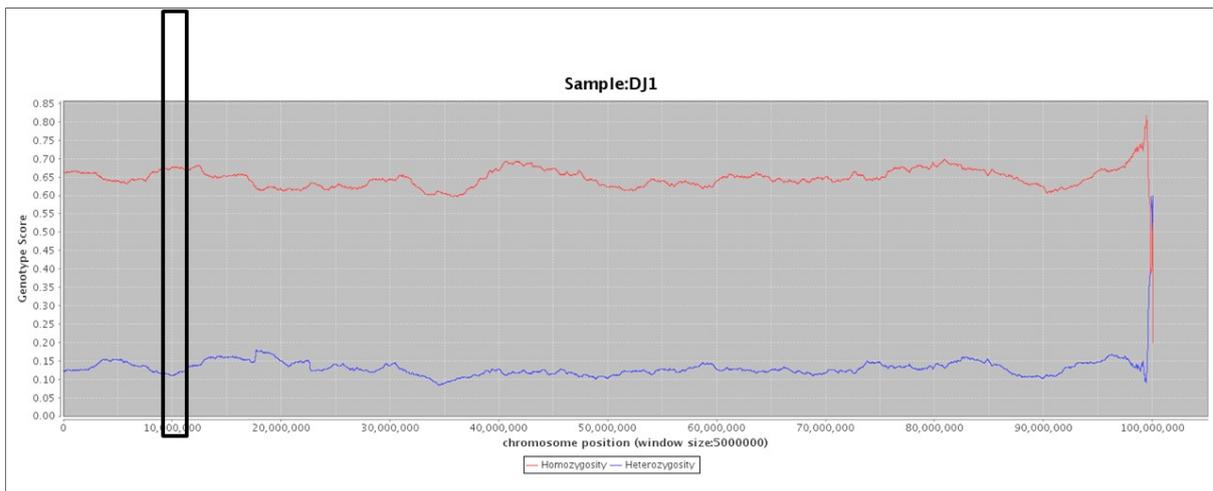


**Appendix Figure H 33 Homozygosity profile signal for chromosome 4 of S5 showing the Trypanotolerance *INHBA* candidate loci**

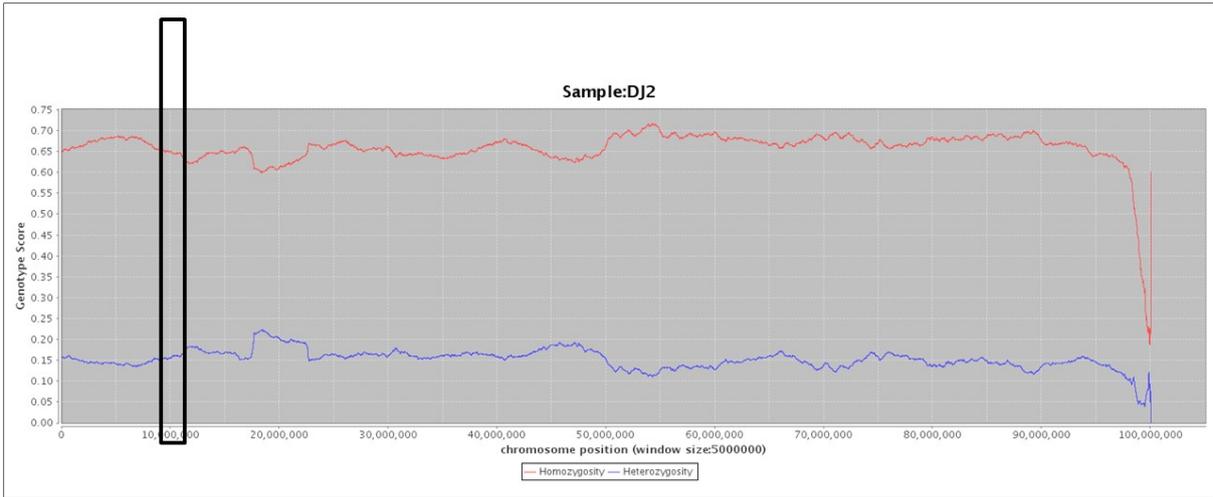
**HomHet predictive signal for the Trypanotolerance *SCAMP1* candidate loci (Djallonke (high ROH) and Sahelian (high ROH)) and HomHet analysis of the loci for all 10 individual samples**



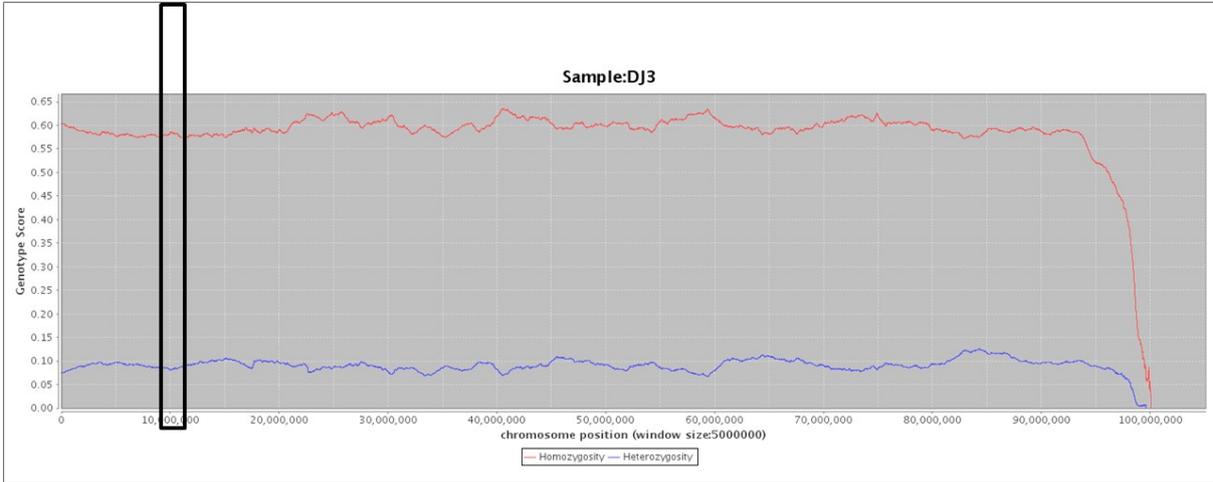
**Appendix Figure H 34 HomHet predictive signal for chromosome 7 showing the Trypanotolerance *SCAMP1* candidate loci**



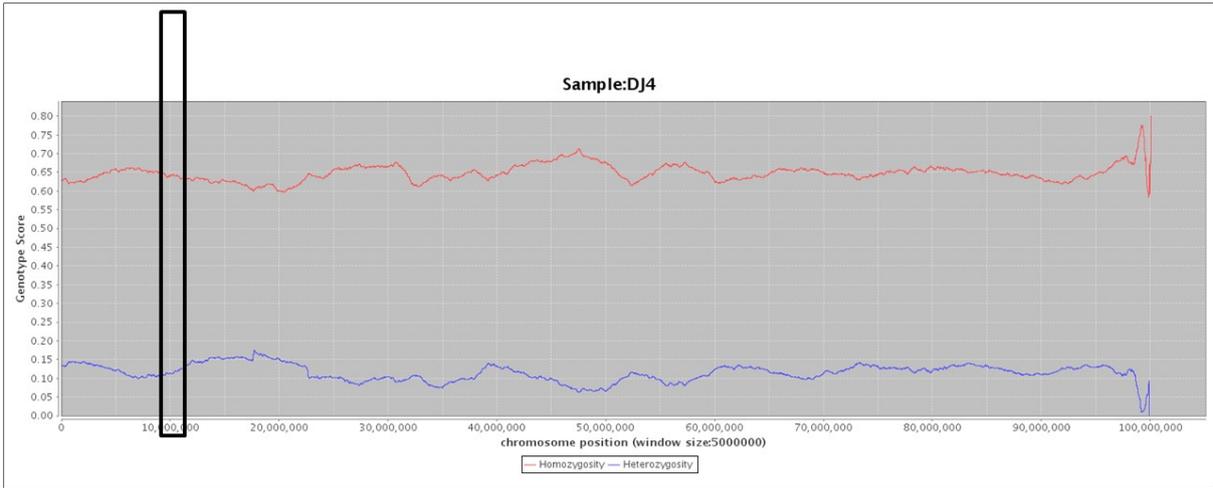
**Appendix Figure H 35 Homozygosity profile signal for chromosome 7 of DJ1 showing the Trypanotolerance *SCAMP1* candidate loci**



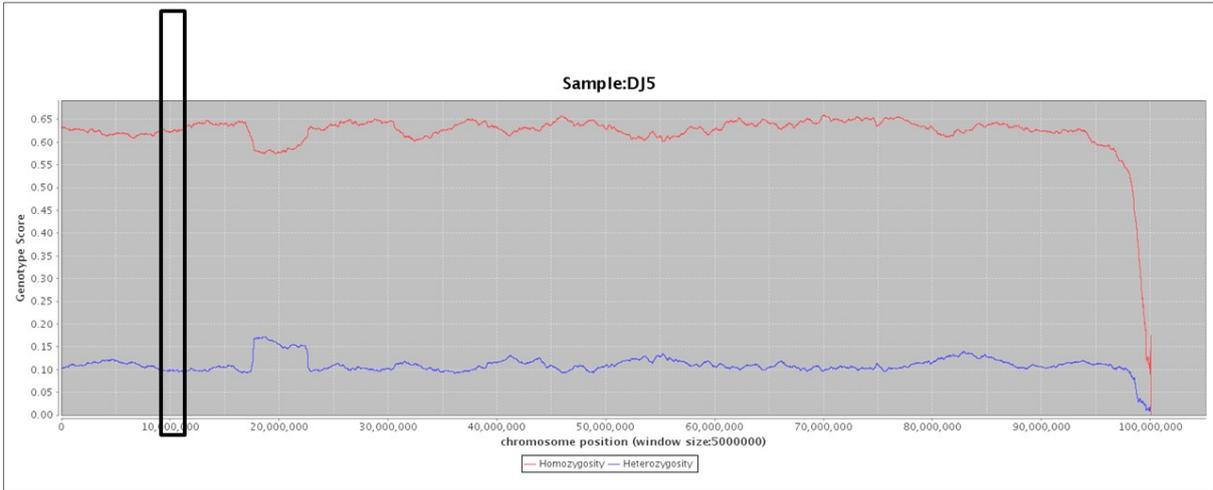
**Appendix Figure H 36 Homozygosity profile signal for chromosome 7 of DJ2 showing the Trypanotolerance *SCAMP1* candidate loci**



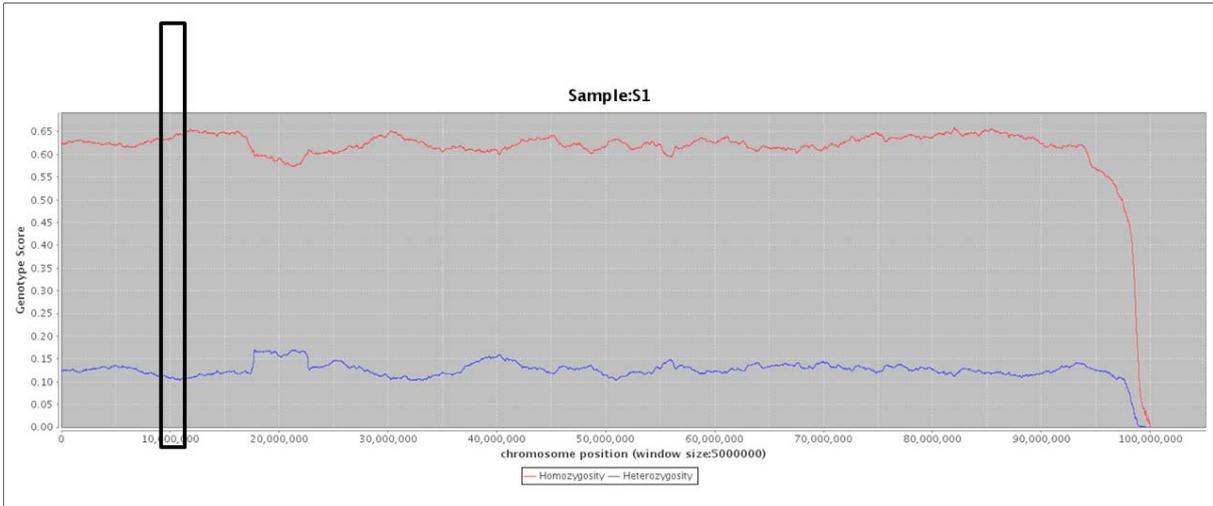
**Appendix Figure H 37 Homozygosity profile signal for chromosome 7 of DJ3 showing the Trypanotolerance *SCAMP1* candidate loci**



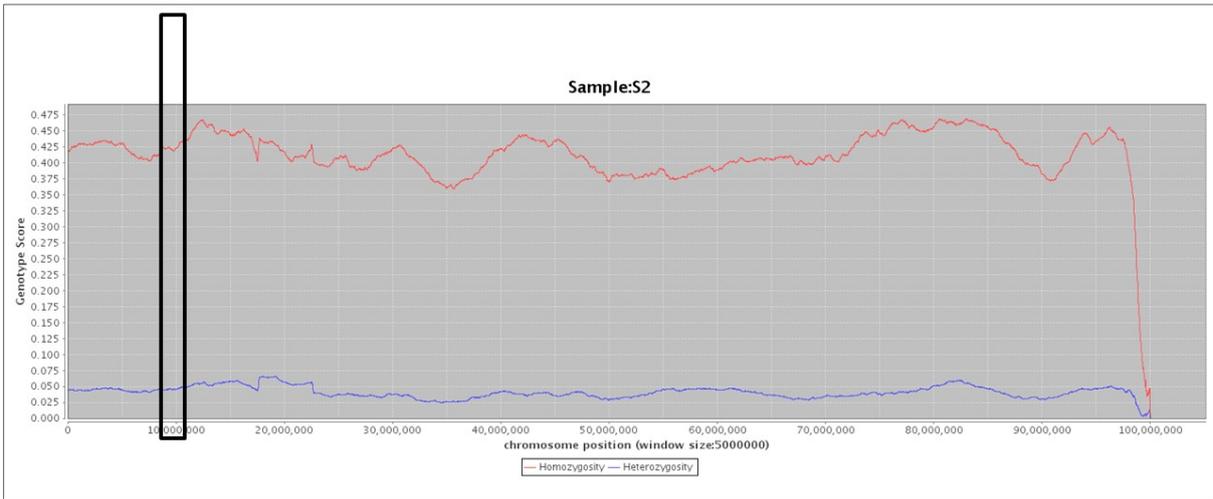
**Appendix Figure H 38 Homozygosity profile signal for chromosome 7 of DJ4 showing the Trypanotolerance *SCAMP1* candidate loci**



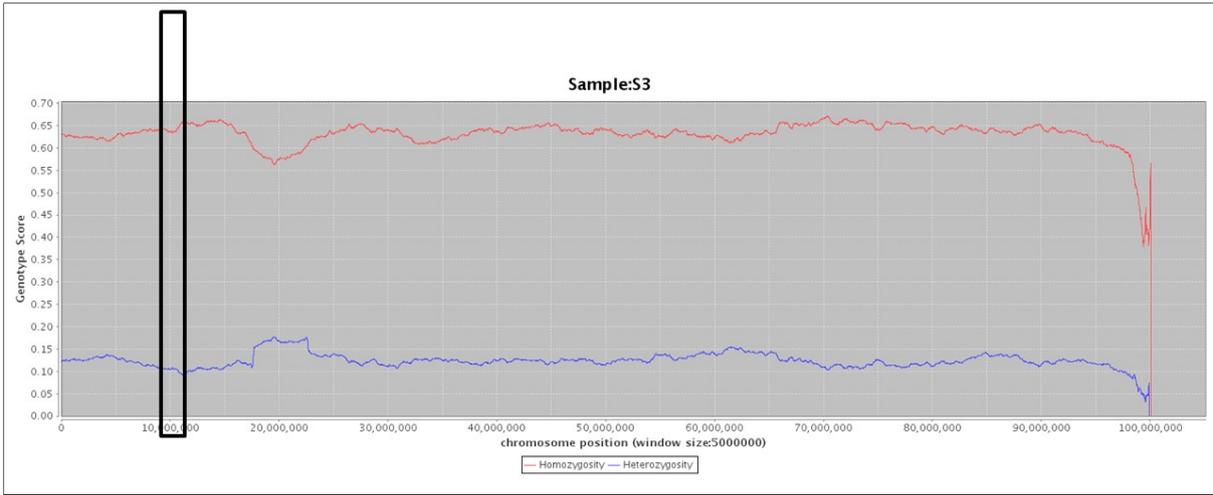
**Appendix Figure H 39 Homozygosity profile signal for chromosome 7 of DJ5 showing the Trypanotolerance *SCAMP1* candidate loci**



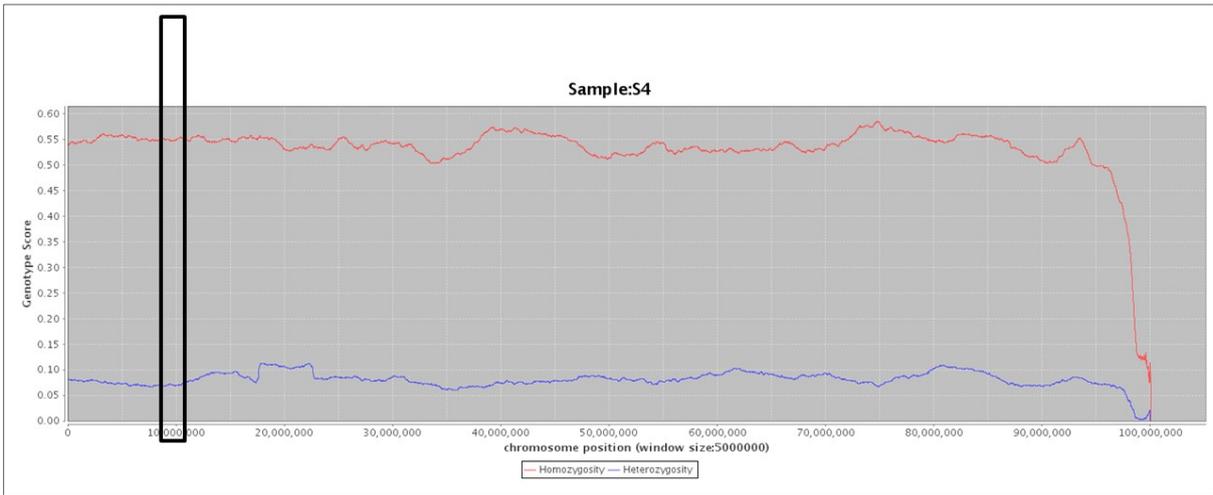
**Appendix Figure H 40 Homozygosity profile signal for chromosome 7 of S1 showing the Trypanotolerance *SCAMP1* candidate loci**



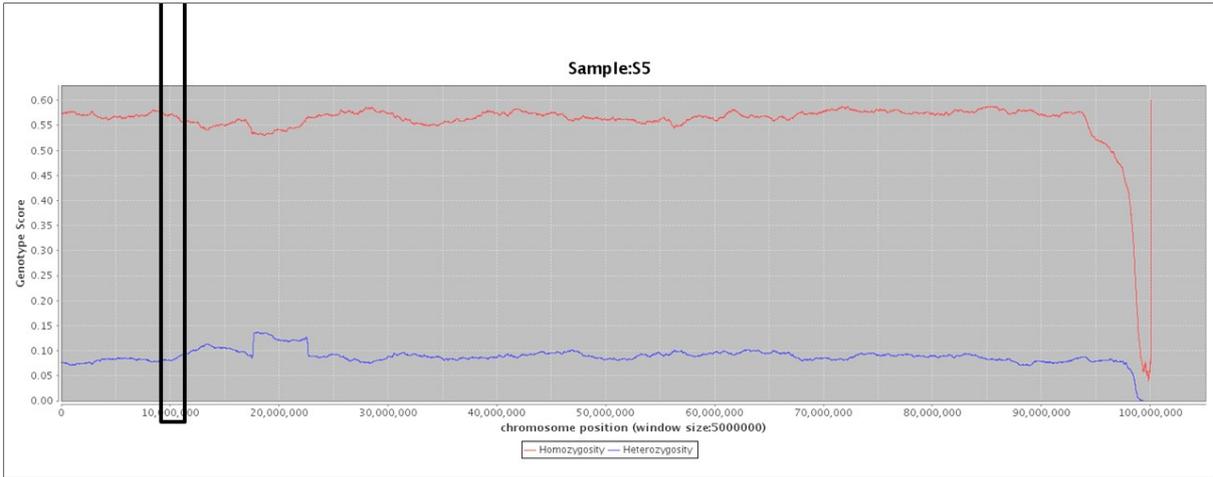
**Appendix Figure H 41 Homozygosity profile signal for chromosome 7 of S2 showing the Trypanotolerance *SCAMP1* candidate loci**



**Appendix Figure H 42 Homozygosity profile signal for chromosome 7 of S3 showing the Trypanotolerance *SCAMP1* candidate loci**

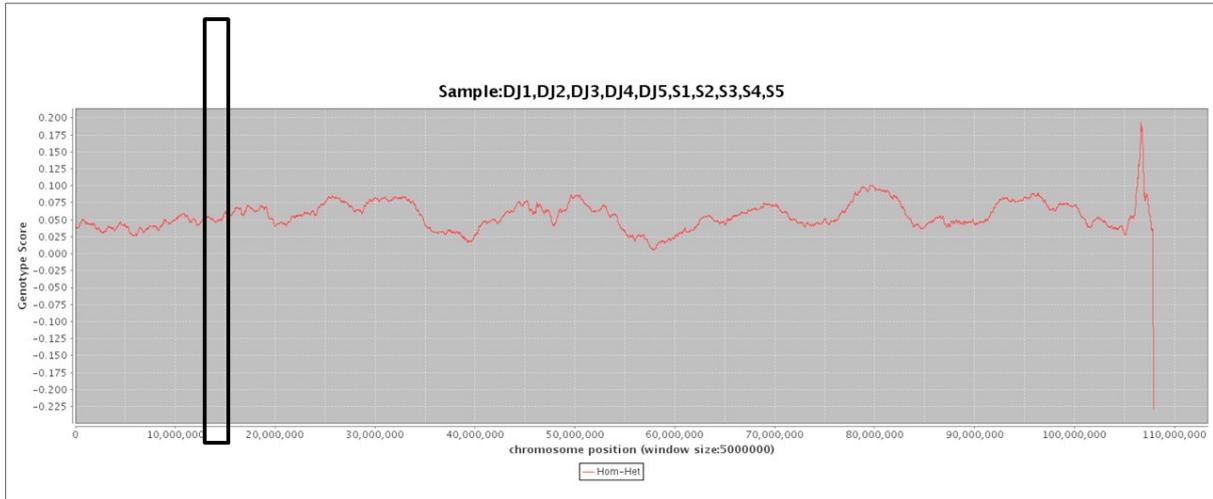


**Appendix Figure H 43 Homozygosity profile signal for chromosome 7 of S4 showing the Trypanotolerance *SCAMP1* candidate loci**

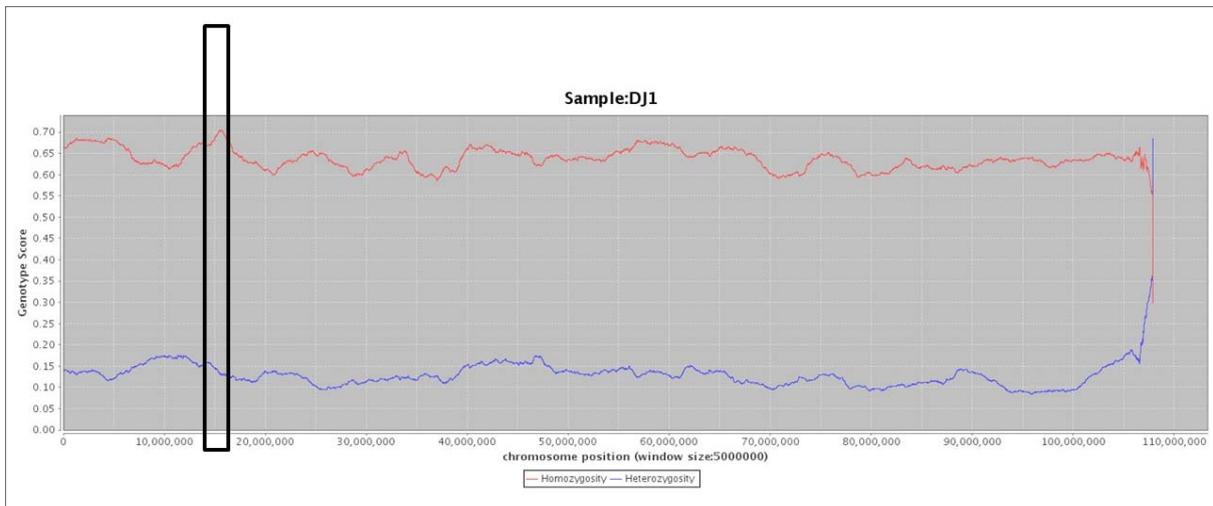


**Appendix Figure H 44 Homozygosity profile signal for chromosome 7 of S5 showing the Trypanotolerance *SCAMP1* candidate loci**

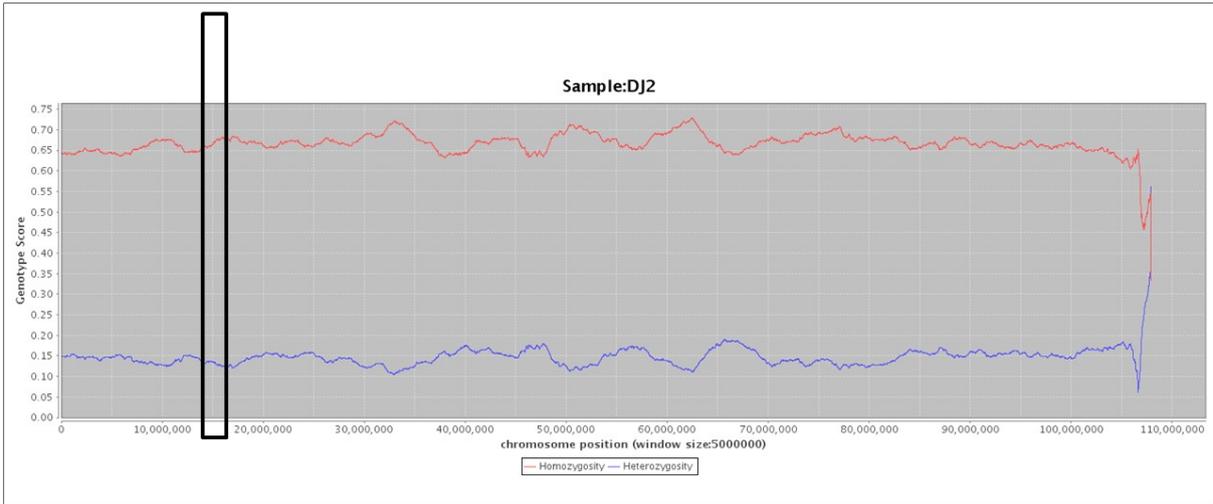
**HomHet predictive signal for the Trypanotolerance *TICAM1* candidate loci (Djallonke (high ROH) and Sahelian (high ROH)) and HomHet analysis of the loci for all 10 individual samples**



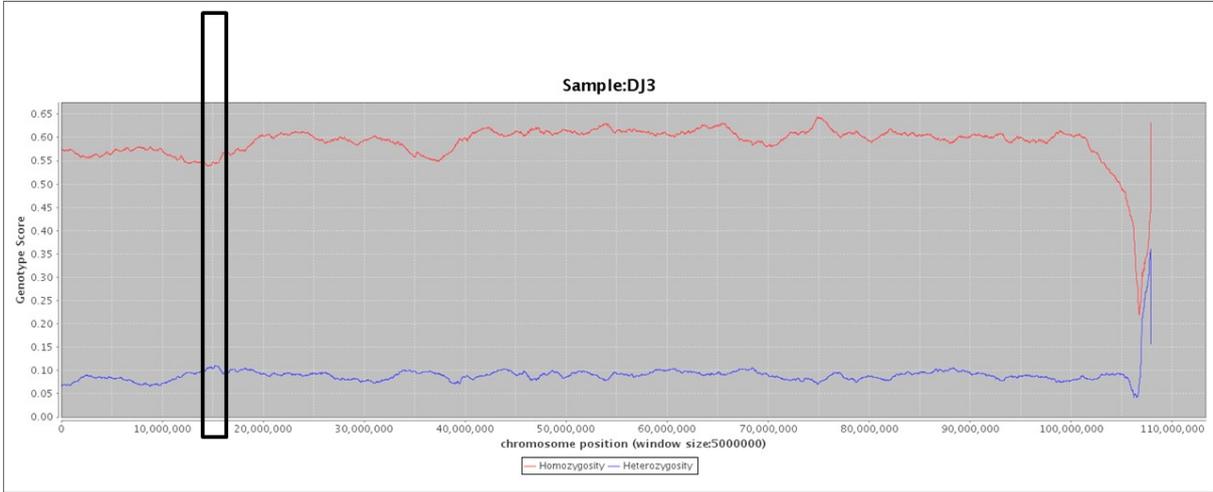
**Appendix Figure H 45 HomHet predictive signal for chromosome 5 showing the Trypanotolerance *TICAM1* candidate loci**



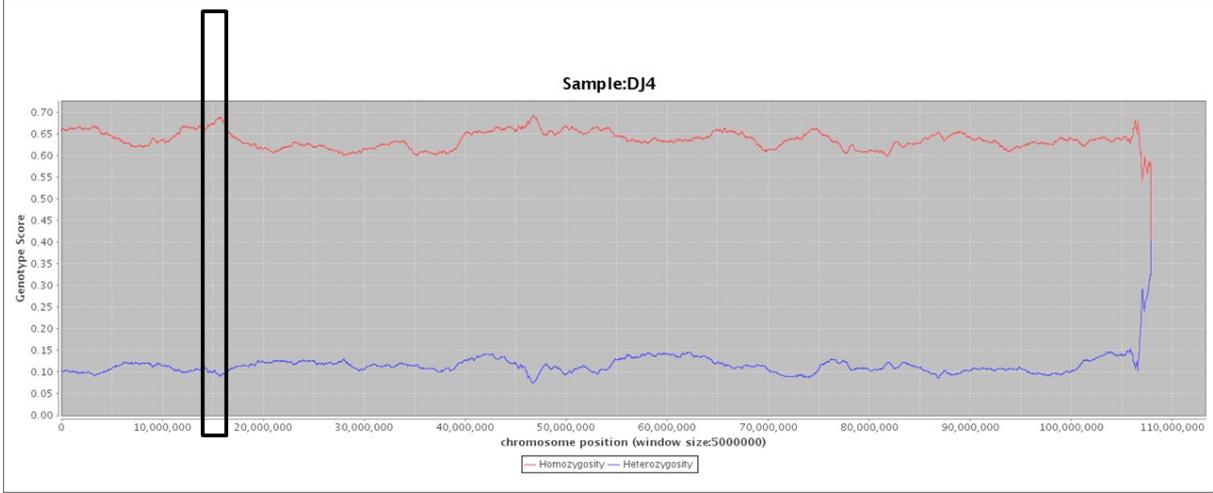
**Appendix Figure H 46 Homozygosity profile signal for chromosome 5 of DJ1 showing the Trypanotolerance *TICAM1* candidate loci**



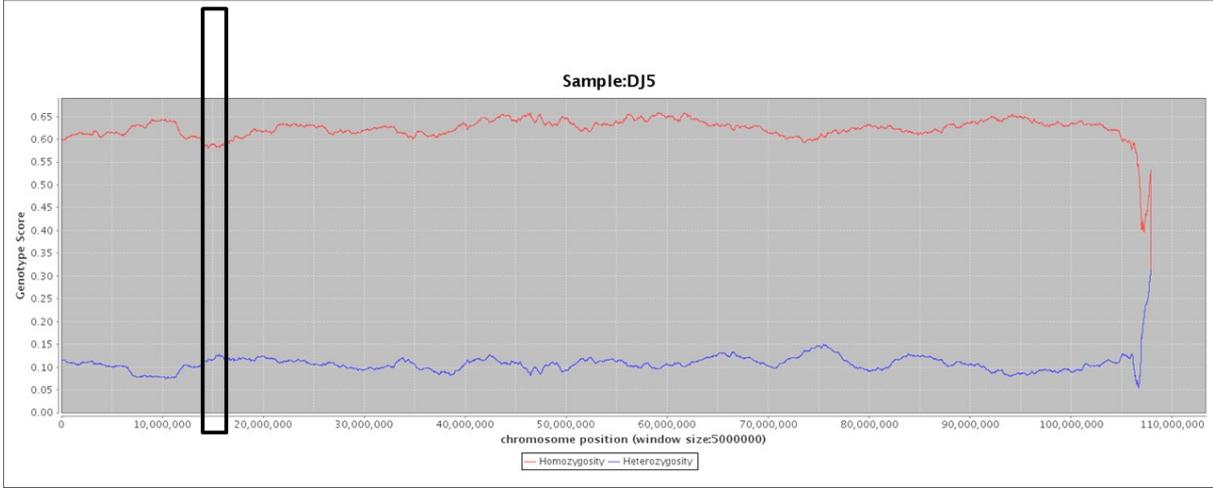
**Appendix Figure H 47 Homozygosity profile signal for chromosome 5 of DJ2 showing the Trypanotolerance *TICAM1* candidate loci**



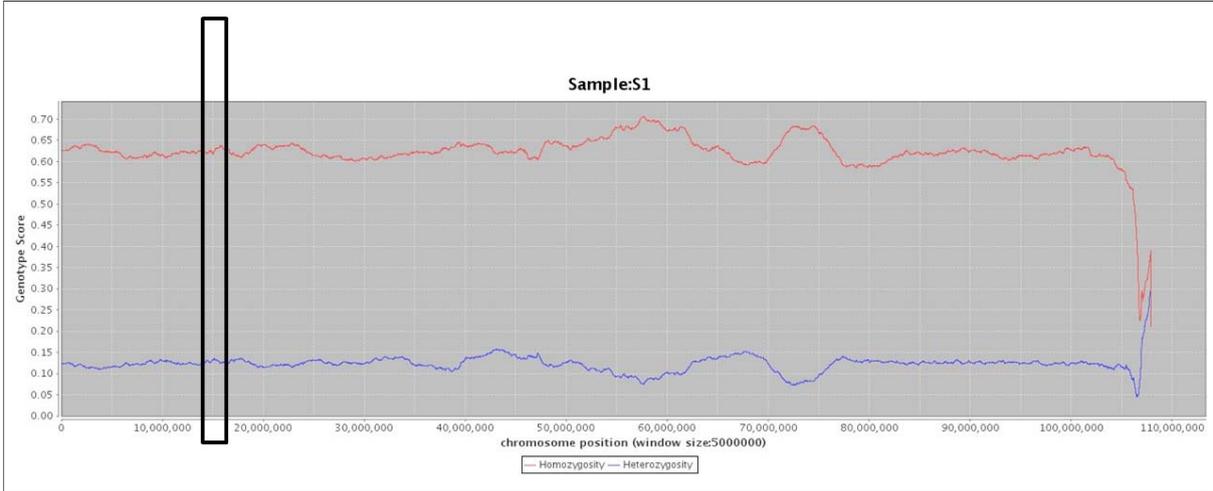
**Appendix Figure H 48 Homozygosity profile signal for chromosome 5 of DJ3 showing the Trypanotolerance *TICAM1* candidate loci**



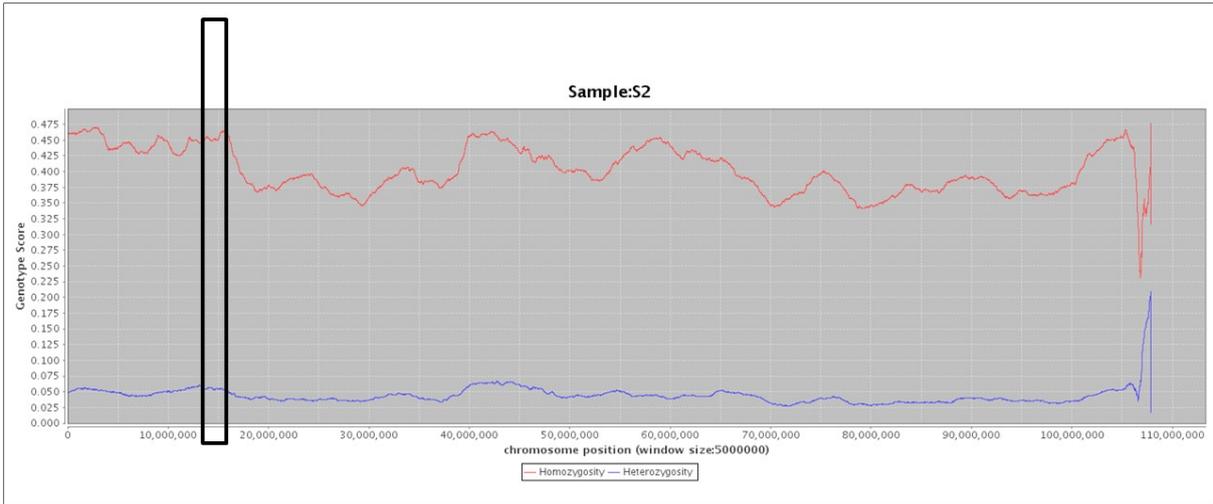
**Appendix Figure H 49 Homozygosity profile signal for chromosome 5 of DJ4 showing the Trypanotolerance *TICAM1* candidate loci**



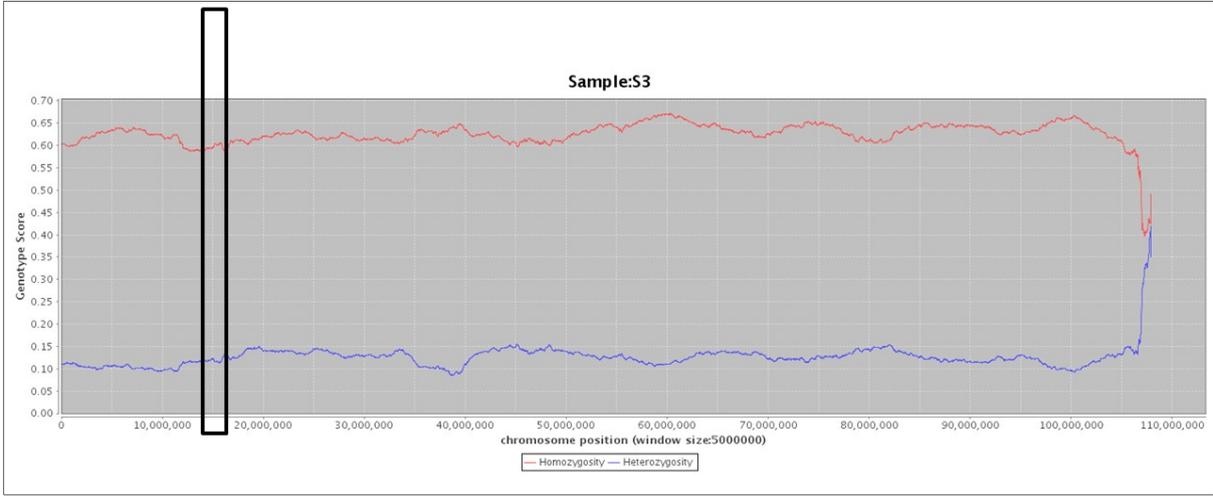
**Appendix Figure H 50 Homozygosity profile signal for chromosome 5 of DJ5 showing the Trypanotolerance *TICAM1* candidate loci**



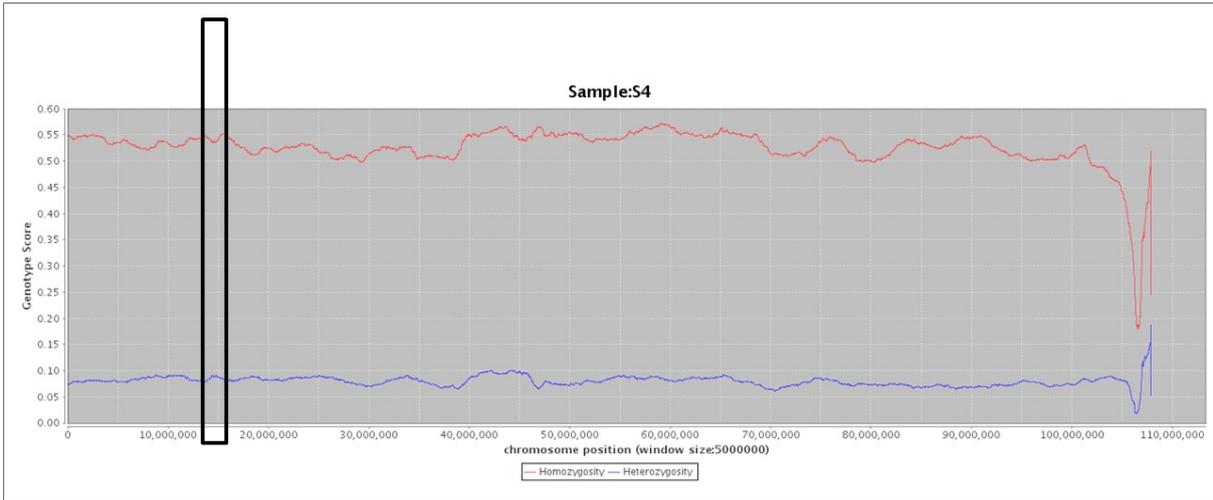
**Appendix Figure H 51 Homozygosity profile signal for chromosome 5 of S1 showing the Trypanotolerance *TICAM1* candidate loci**



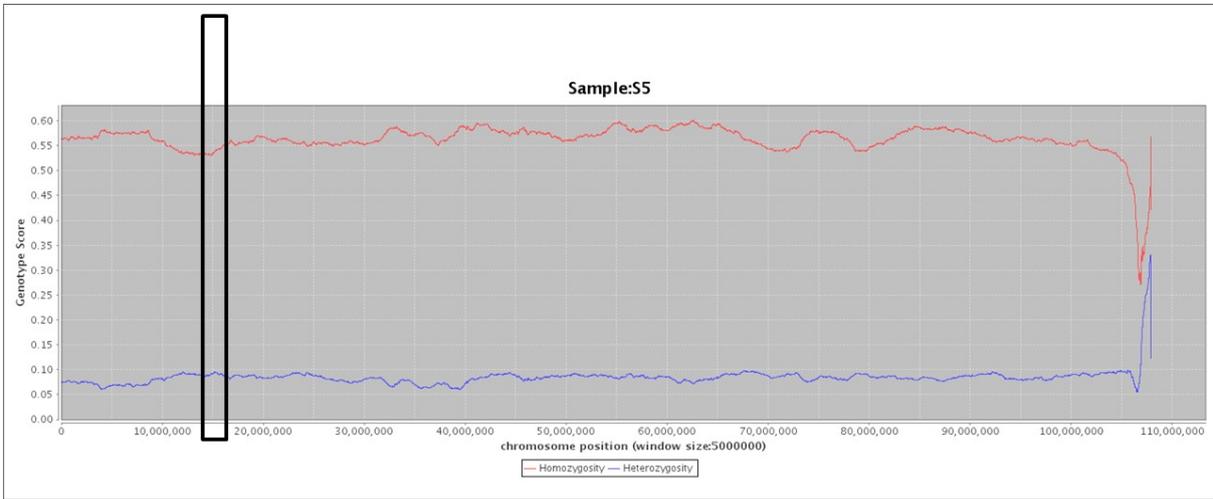
**Appendix Figure H 52 Homozygosity profile signal for chromosome 5 of S2 showing the Trypanotolerance *TICAM1* candidate loci**



**Appendix Figure H 53 Homozygosity profile signal for chromosome 5 of S3 showing the Trypanotolerance *TICAM1* candidate loci**

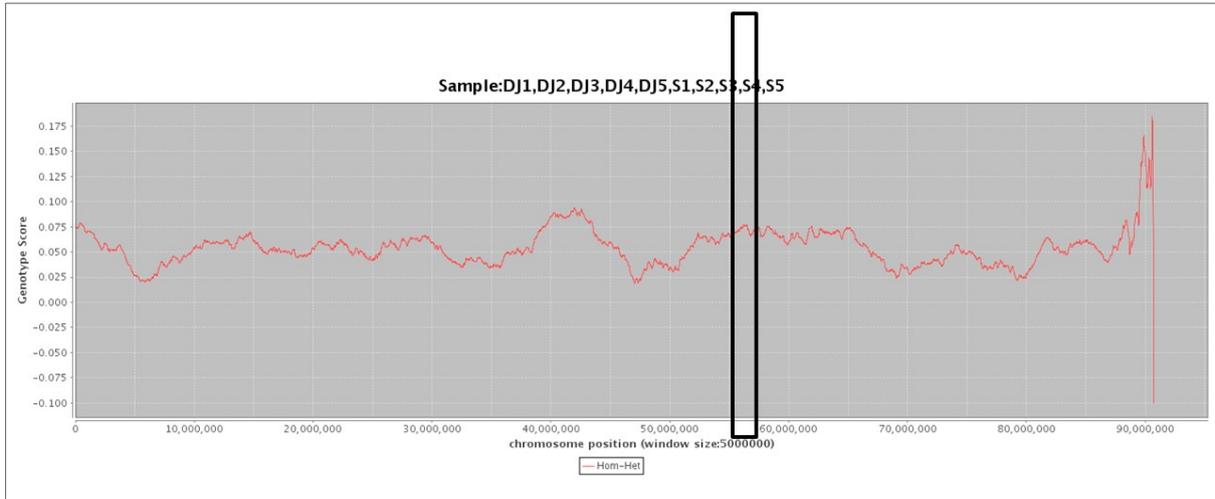


**Appendix Figure H 54 Homozygosity profile signal for chromosome 5 of S4 showing the Trypanotolerance *TICAM1* candidate loci**

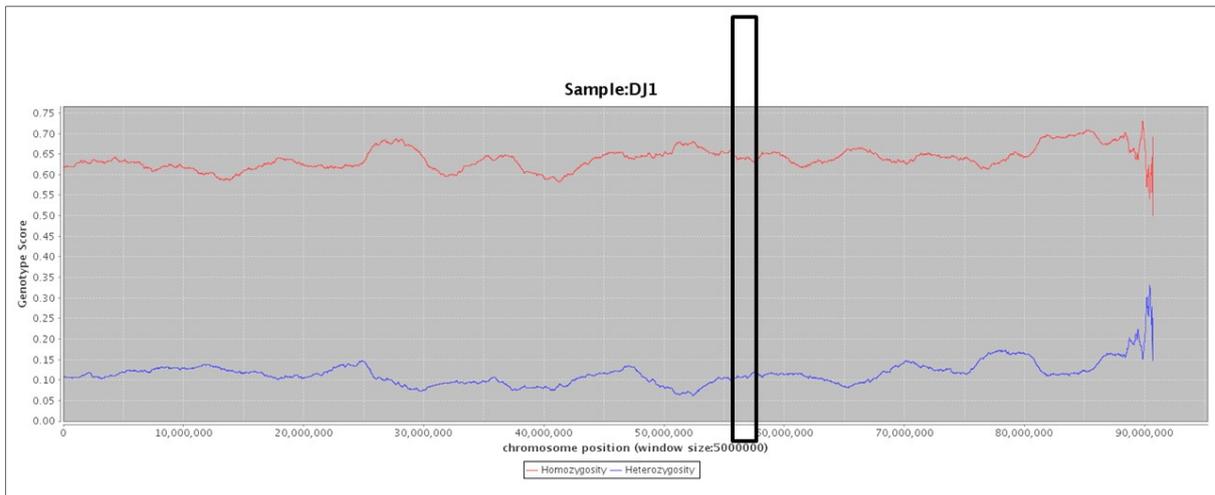


**Appendix Figure H 55 Homozygosity profile signal for chromosome 5 of S5 showing the Trypanotolerance *TICAM1* candidate loci**

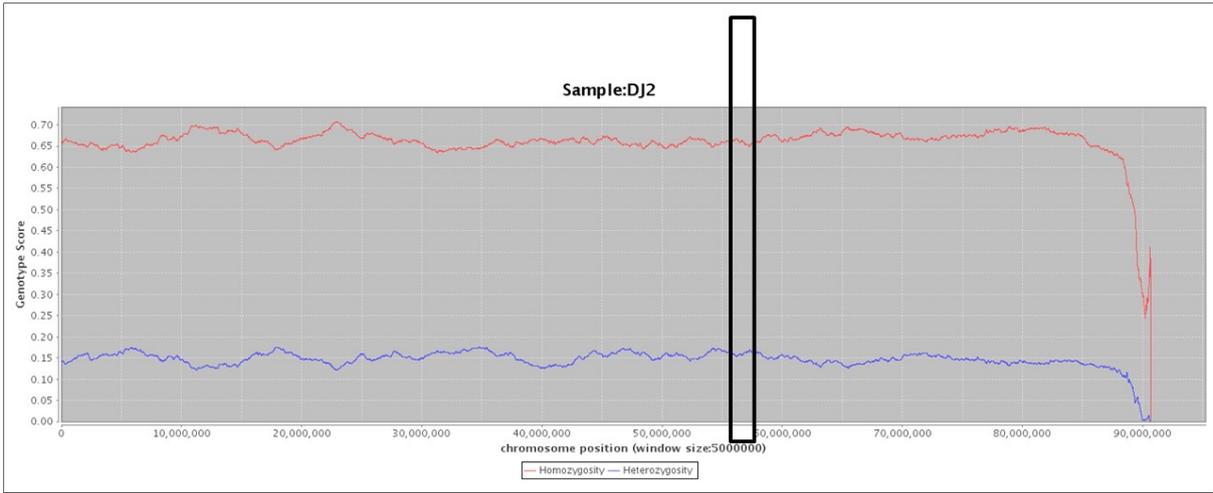
**HomHet predictive signal for the Trypanotolerance *STX7* candidate loci (Djallonke (high ROH) and Sahelian (low ROH)) and HomHet analysis of the loci for all 10 individual samples**



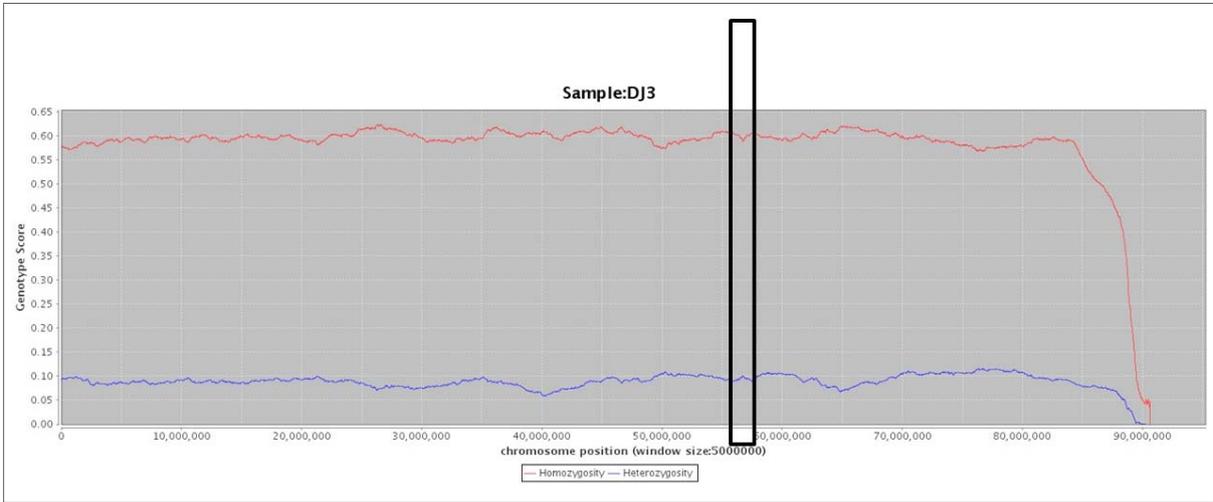
**Appendix Figure H 56 HomHet predictive signal for chromosome 8 showing the Trypanotolerance *STX7* candidate loci**



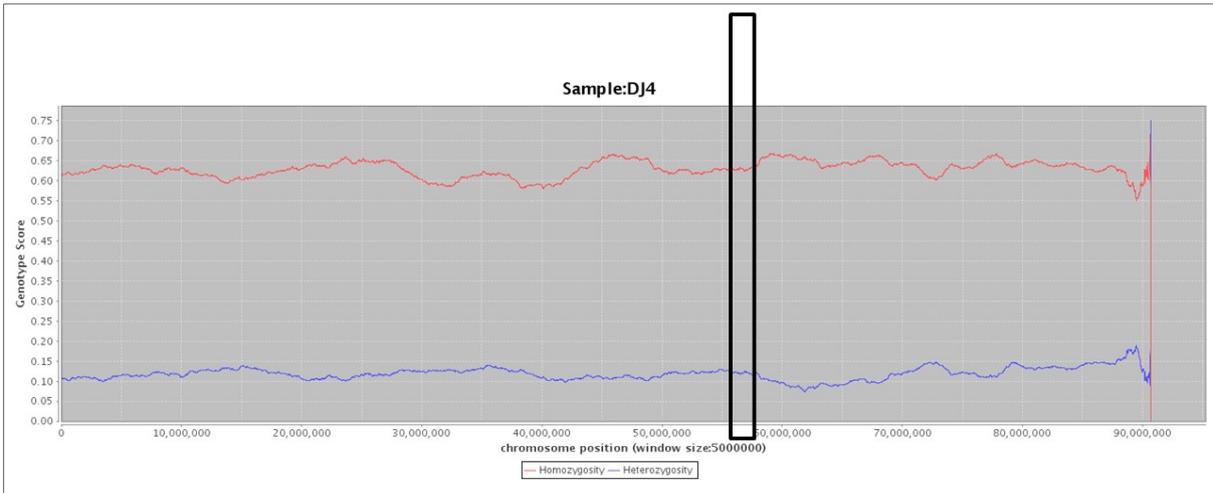
**Appendix Figure H 57 Homozygosity profile signal for chromosome 8 of DJ1 showing the Trypanotolerance *STX7* candidate loci**



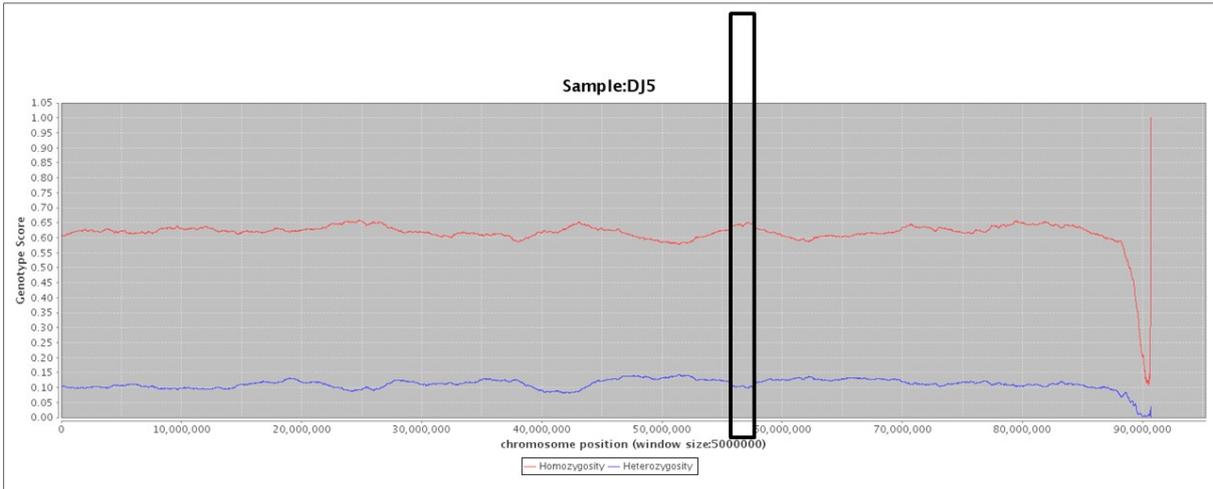
**Appendix Figure H 58 Homozygosity profile signal for chromosome 8 of DJ2 showing the Trypanotolerance *STX7* candidate loci**



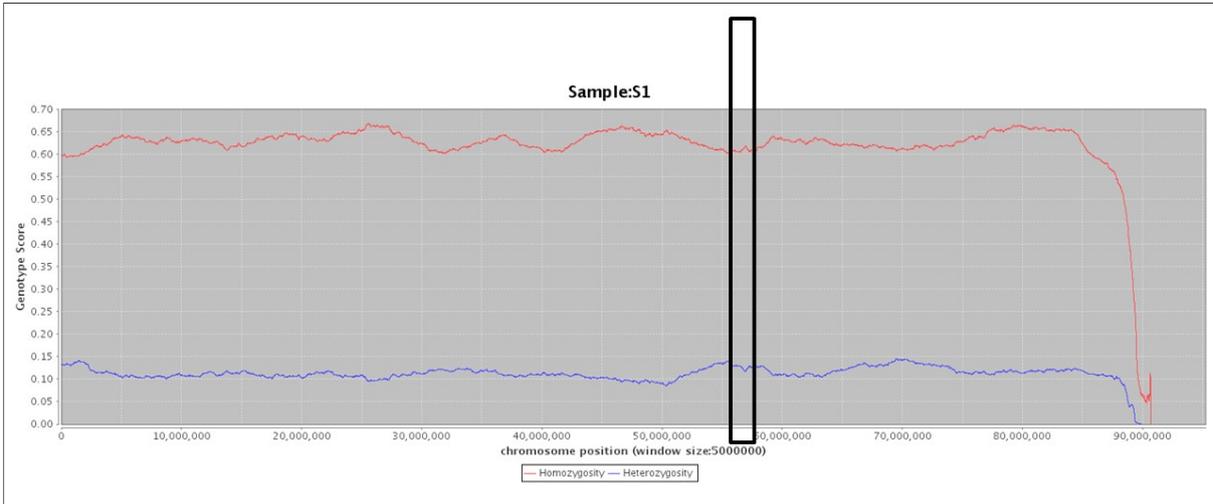
**Appendix Figure H 59 Homozygosity profile signal for chromosome 8 of DJ3 showing the Trypanotolerance *STX7* candidate loci**



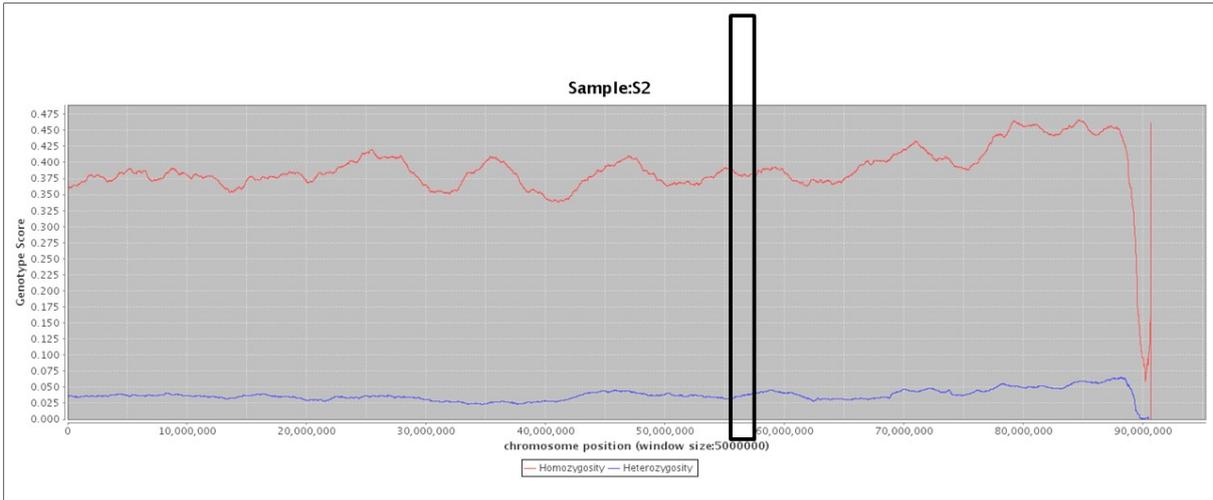
**Appendix Figure H 60 Homozygosity profile signal for chromosome 8 of DJ4 showing the Trypanotolerance *STX7* candidate loci**



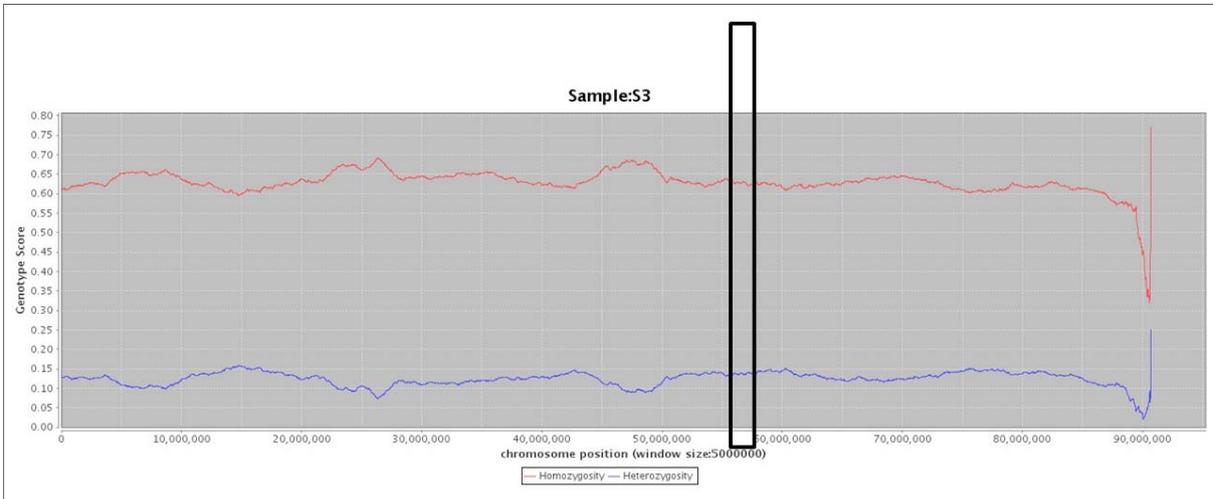
**Appendix Figure H 61 Homozygosity profile signal for chromosome 8 of DJ5 showing the Trypanotolerance *STX7* candidate loci**



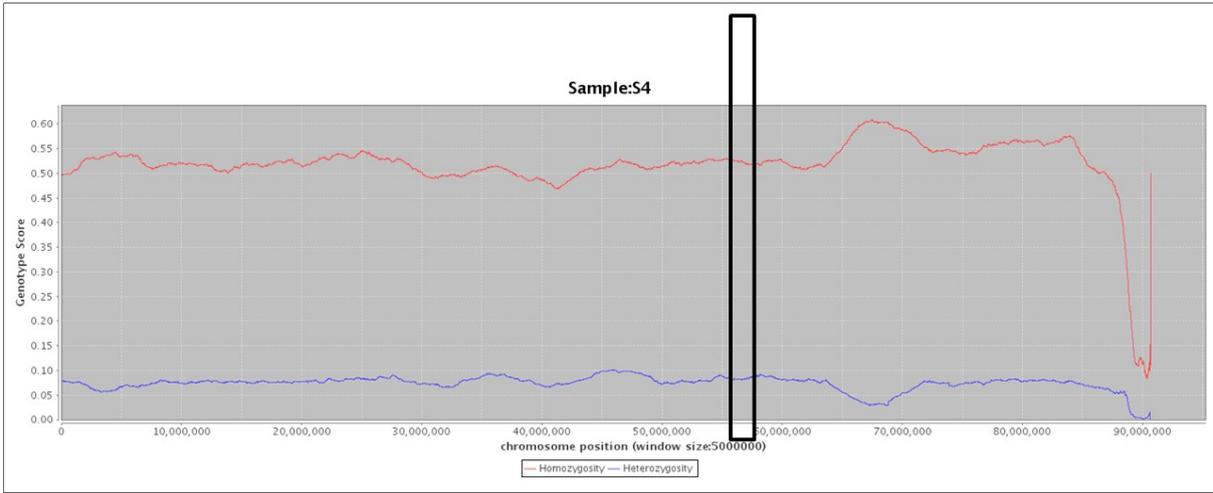
**Appendix Figure H 62 Homozygosity profile signal for chromosome 8 of S1 showing the Trypanotolerance *STX7* candidate loci**



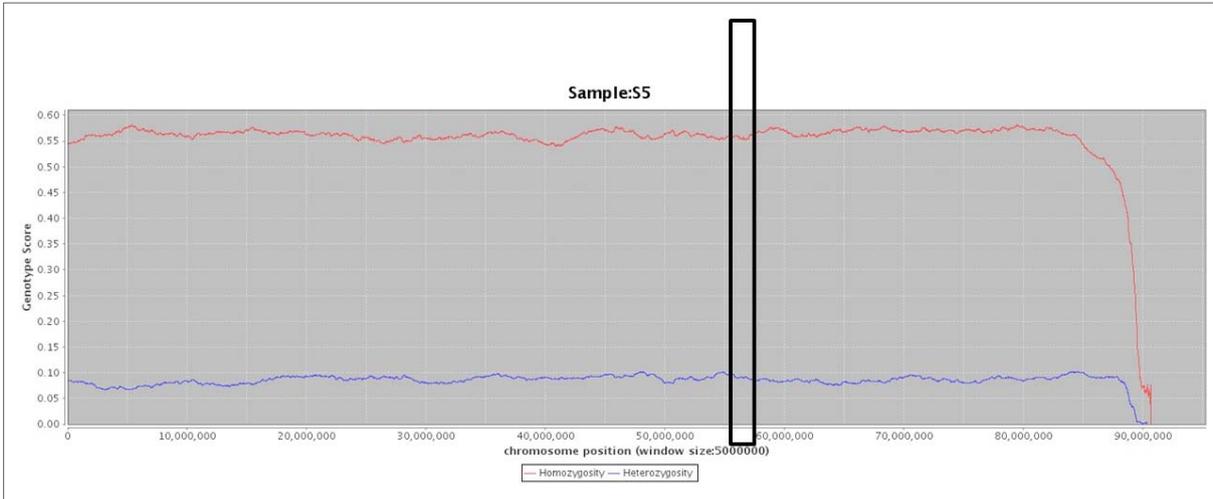
**Appendix Figure H 63 Homozygosity profile signal for chromosome 8 of S2 showing the Trypanotolerance *STX7* candidate loci**



**Appendix Figure H 64 Homozygosity profile signal for chromosome 8 of S3 showing the Trypanotolerance *STX7* candidate loci**

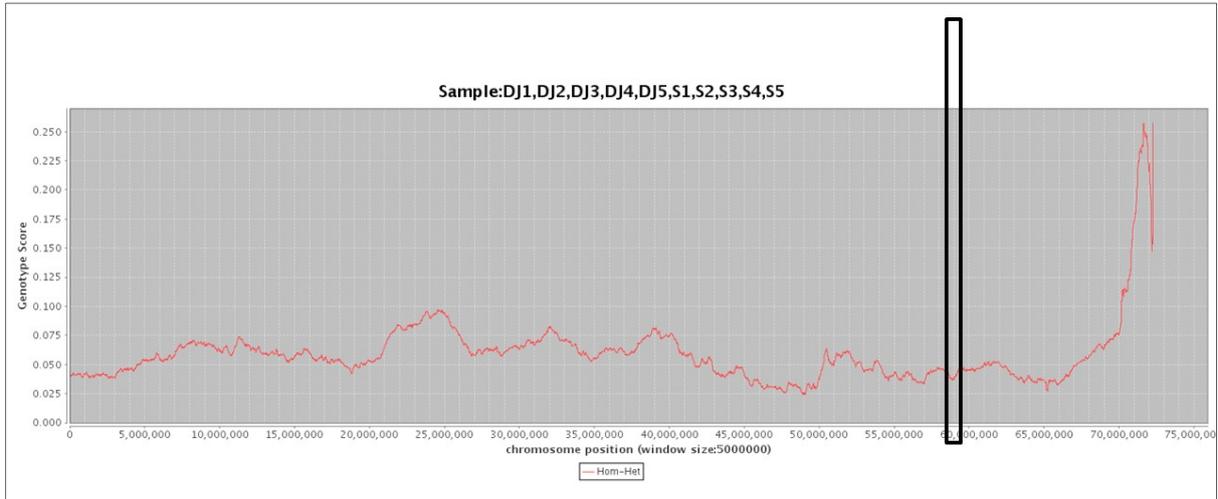


**Appendix Figure H 65 Homozygosity profile signal for chromosome 8 of S4 showing the Trypanotolerance *STX7* candidate loci**

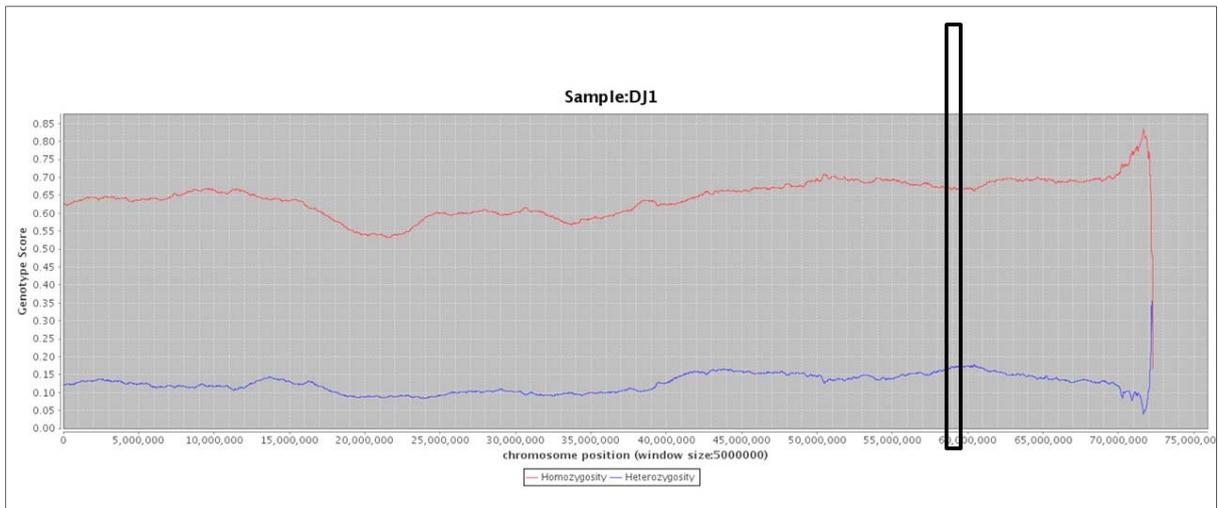


**Appendix Figure H 66 Homozygosity profile signal for chromosome 8 of S5 showing the Trypanotolerance *STX7* candidate loci**

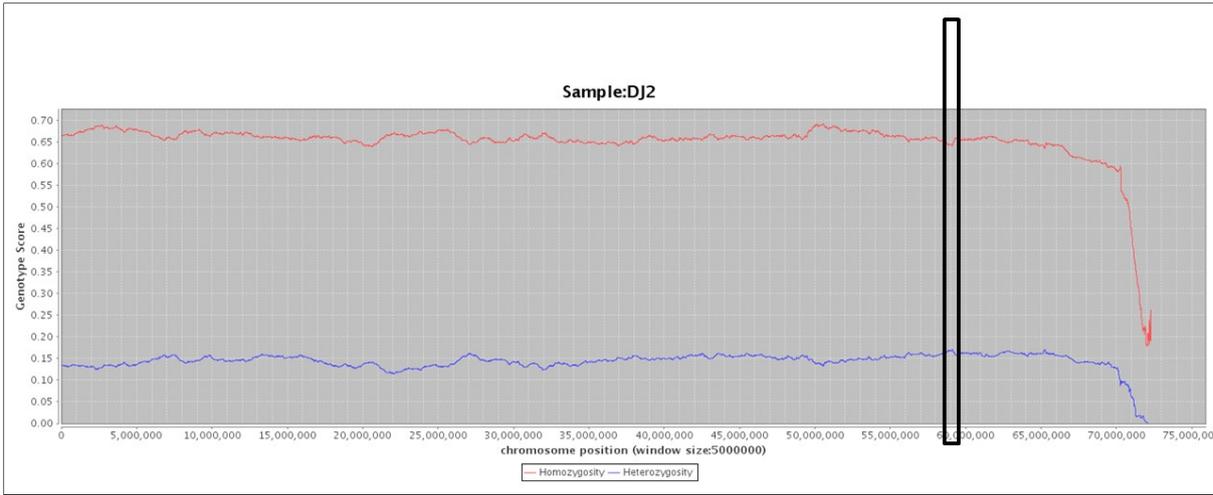
**HomHet predictive signal for the Trypanotolerance *RAB35* candidate loci (Djallonke (high ROH) and Sahelian (low ROH)) and HomHet analysis of the loci for all 10 individual samples**



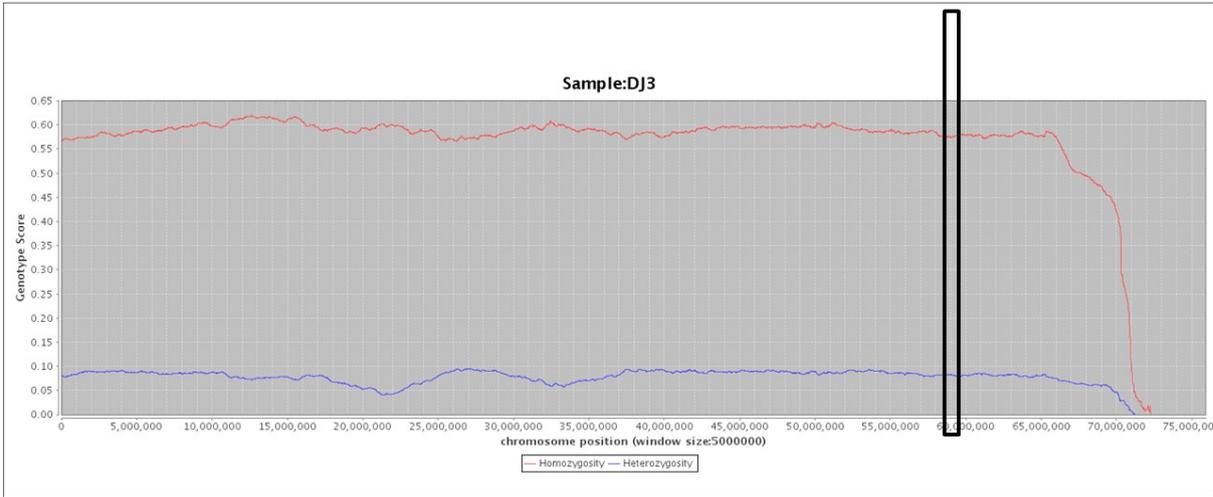
**Appendix Figure H 67 HomHet predictive signal for chromosome 17 showing the Trypanotolerance *RAB35* candidate loci**



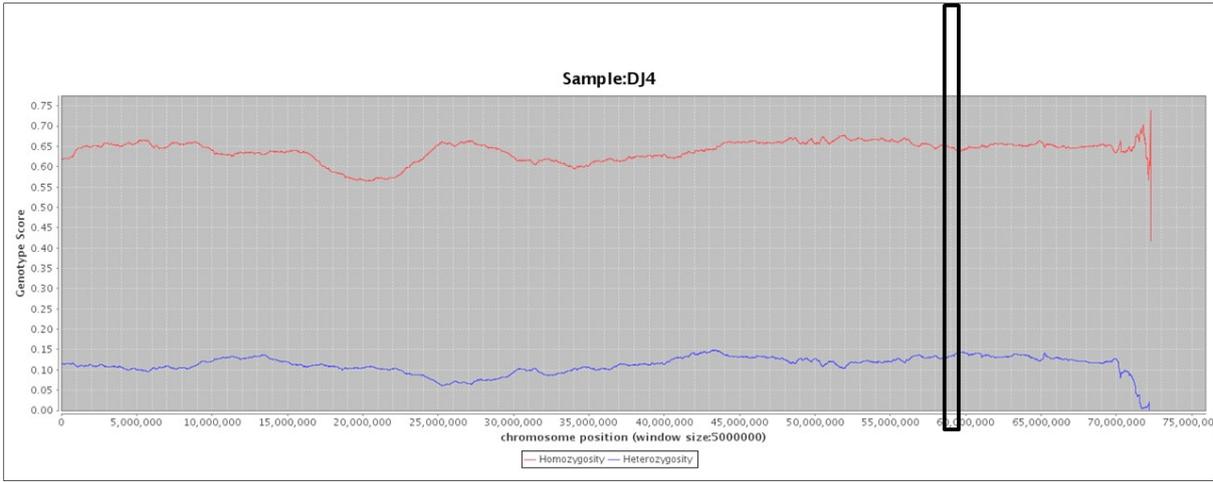
**Appendix Figure H 68 Homozygosity profile signal for chromosome 17 of DJ1 showing the Trypanotolerance *RAB35* candidate loci**



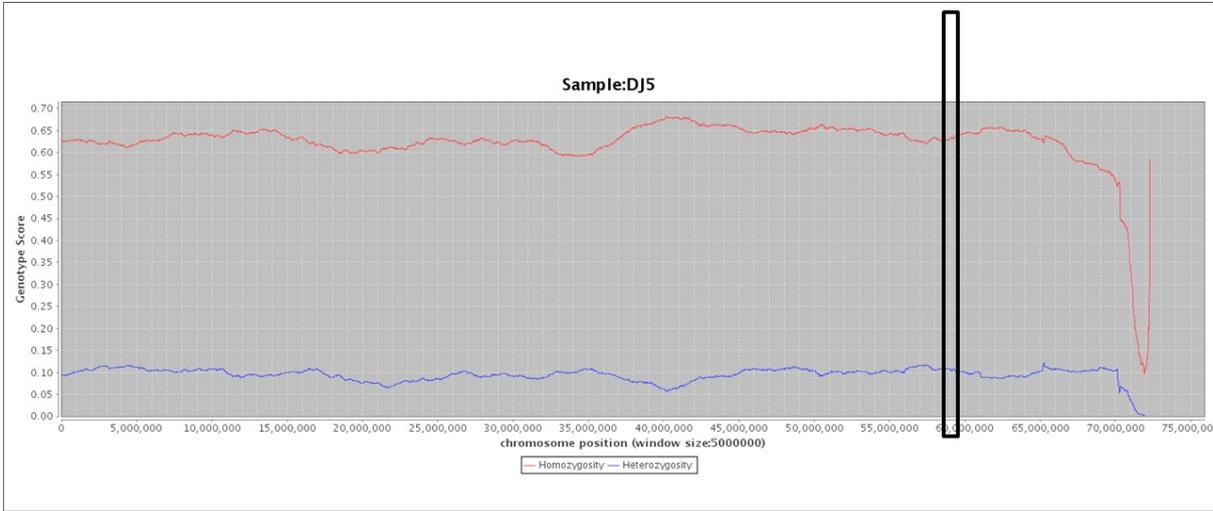
**Appendix Figure H 69 Homozygosity profile signal for chromosome 17 of DJ2 showing the Trypanotolerance *RAB35* candidate loci**



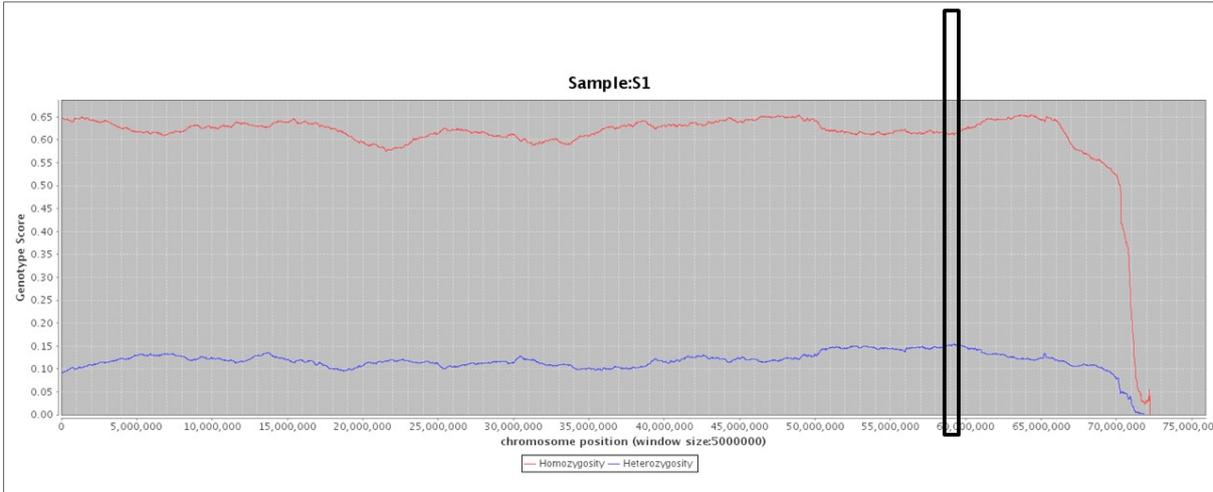
**Appendix Figure H 70 Homozygosity profile signal for chromosome 17 of DJ3 showing the Trypanotolerance *RAB35* candidate loci**



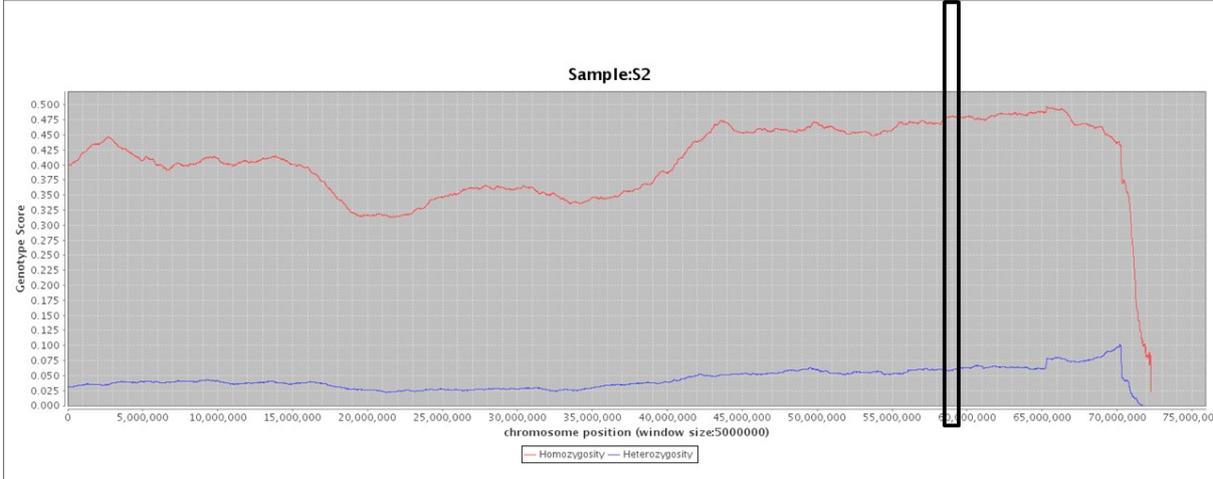
**Appendix Figure H 71 Homozygosity profile signal for chromosome 17 of DJ4 showing the Trypanotolerance *RAB35* candidate loci**



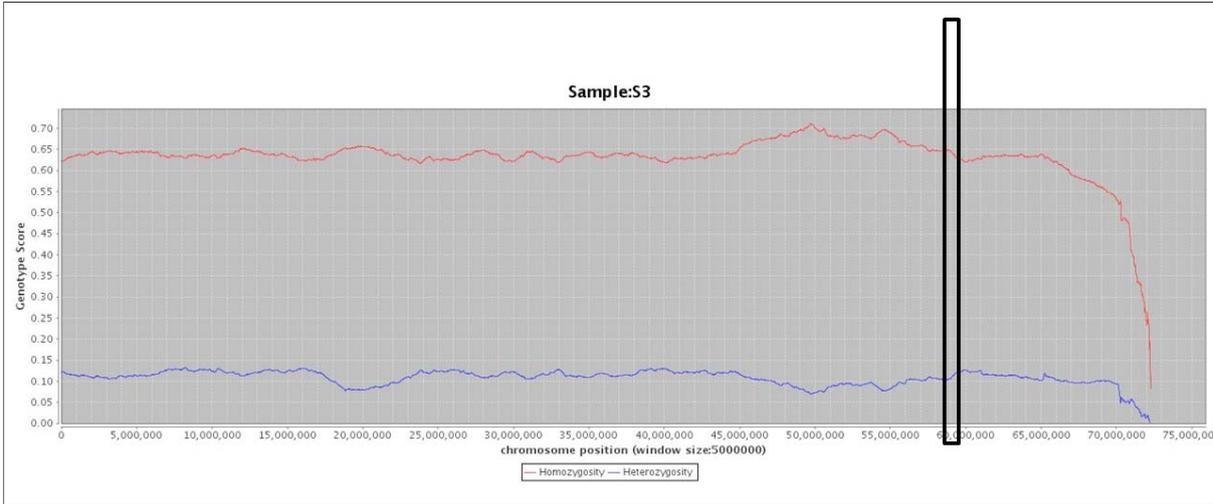
**Appendix Figure H 72 Homozygosity profile signal for chromosome 17 of DJ5 showing the Trypanotolerance *RAB35* candidate loci**



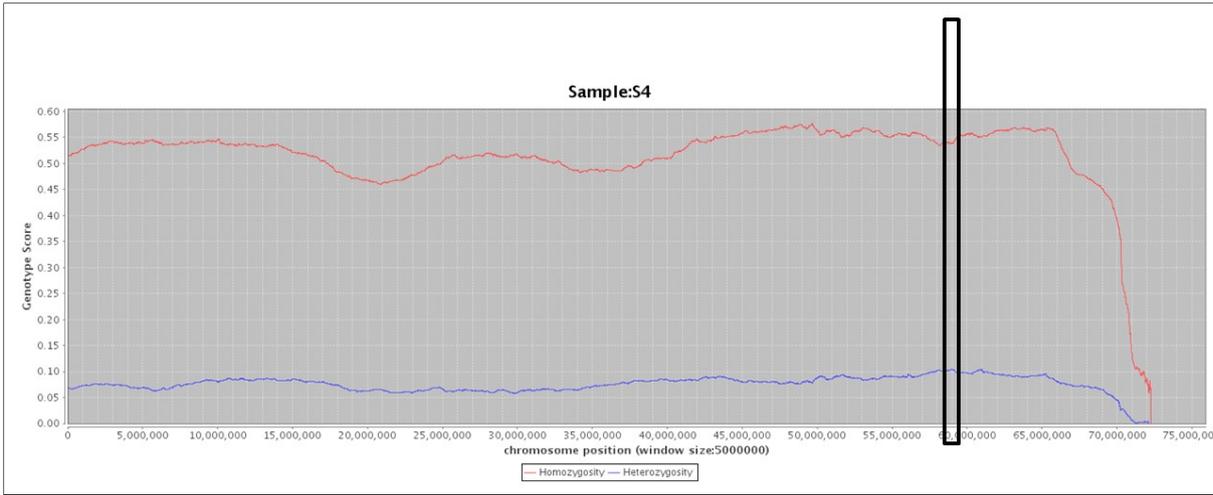
**Appendix Figure H 73 Homozygosity profile signal for chromosome 17 of S1 showing the Trypanotolerance *RAB35* candidate loci**



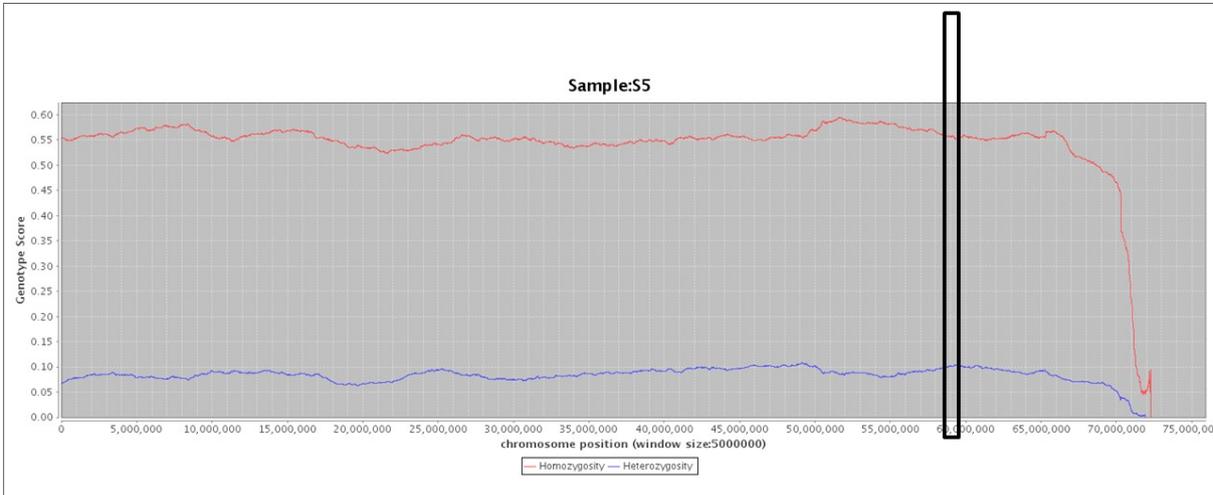
**Appendix Figure H 74 Homozygosity profile signal for chromosome 17 of S2 showing the Trypanotolerance *RAB35* candidate loci**



**Appendix Figure H 75 Homozygosity profile signal for chromosome 17 of S3 showing the Trypanotolerance *RAB35* candidate loci**

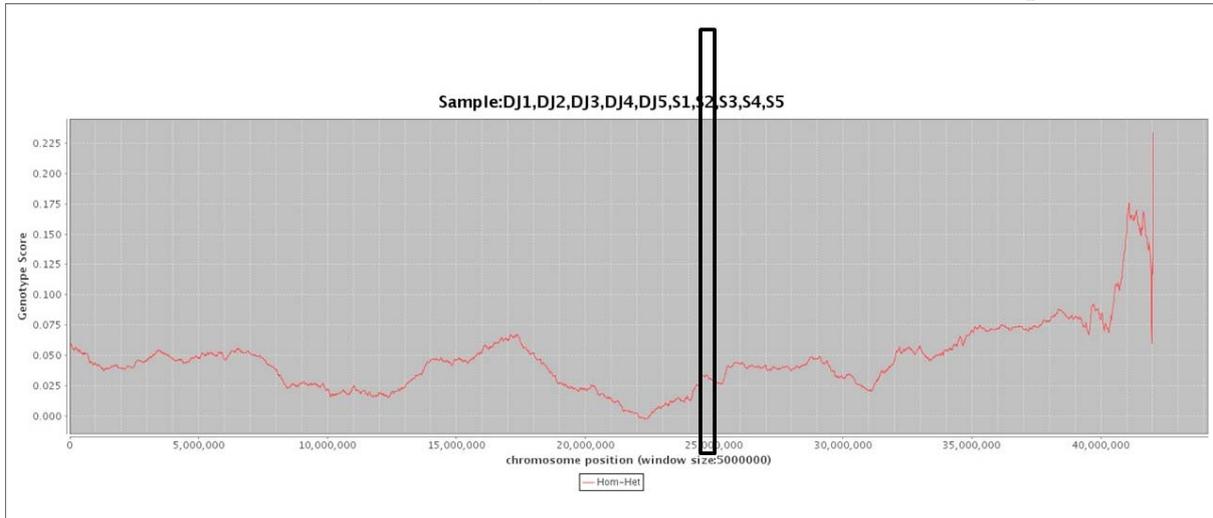


**Appendix Figure H 76 Homozygosity profile signal for chromosome 17 of S4 showing the Trypanotolerance *RAB35* candidate loci**

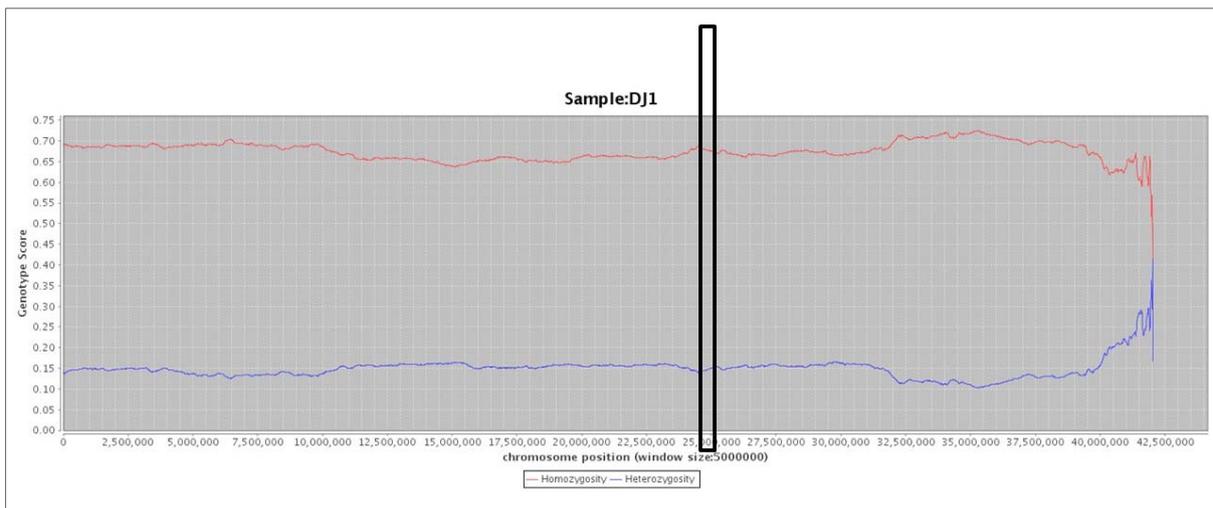


**Appendix Figure H 77 Homozygosity profile signal for chromosome 17 of S5 showing the Trypanotolerance *RAB35* candidate loci**

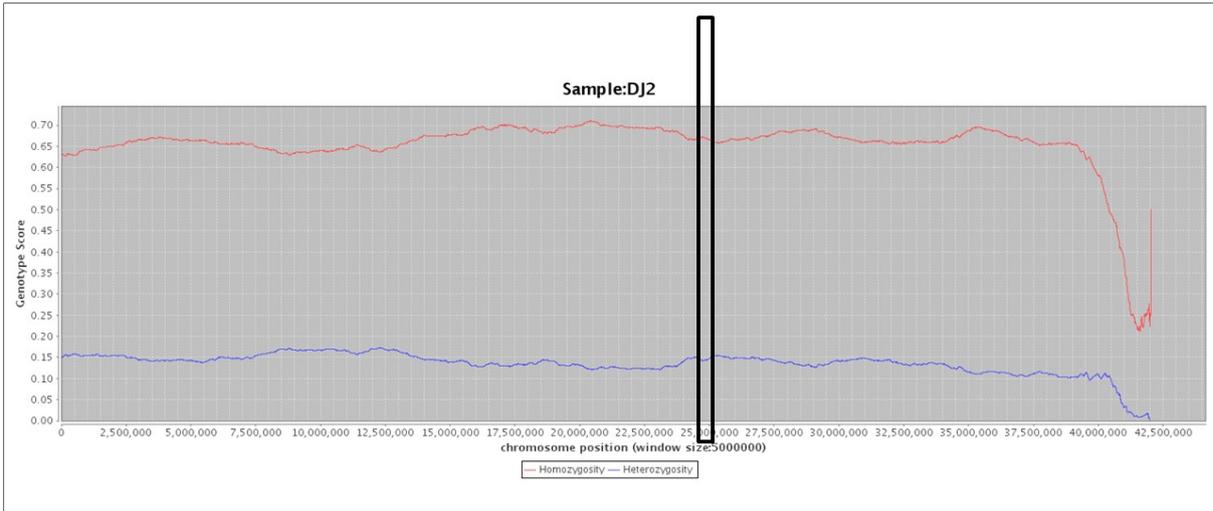
**HomHet predictive signal for the Trypanotolerance *CD19* candidate loci (Djallonke (high ROH) and Sahelian (low ROH)) and HomHet analysis of the loci for all 10 individual samples**



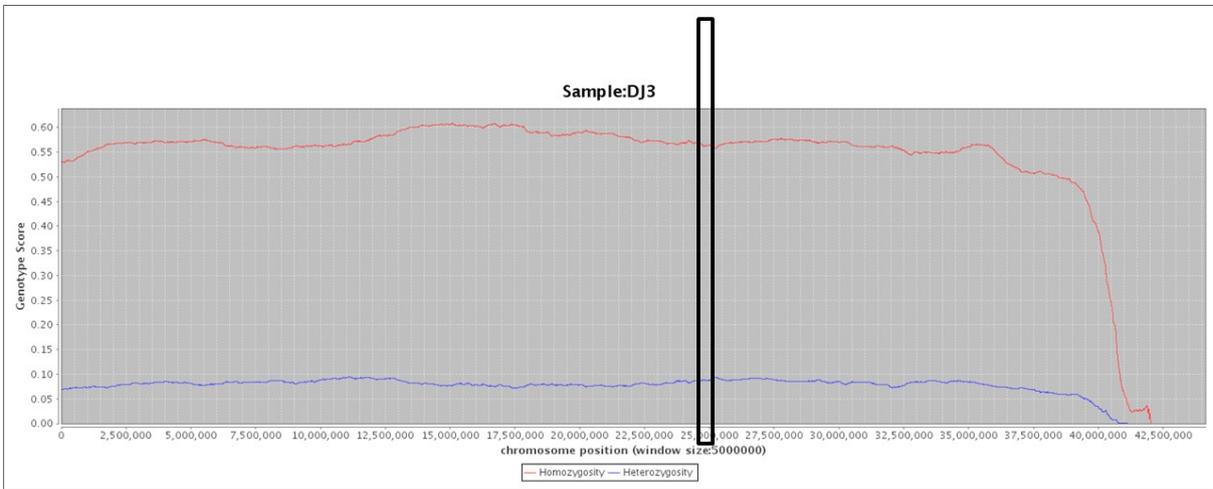
**Appendix Figure H 78 HomHet predictive signal for chromosome 24 showing the Trypanotolerance *CD19* candidate loci**



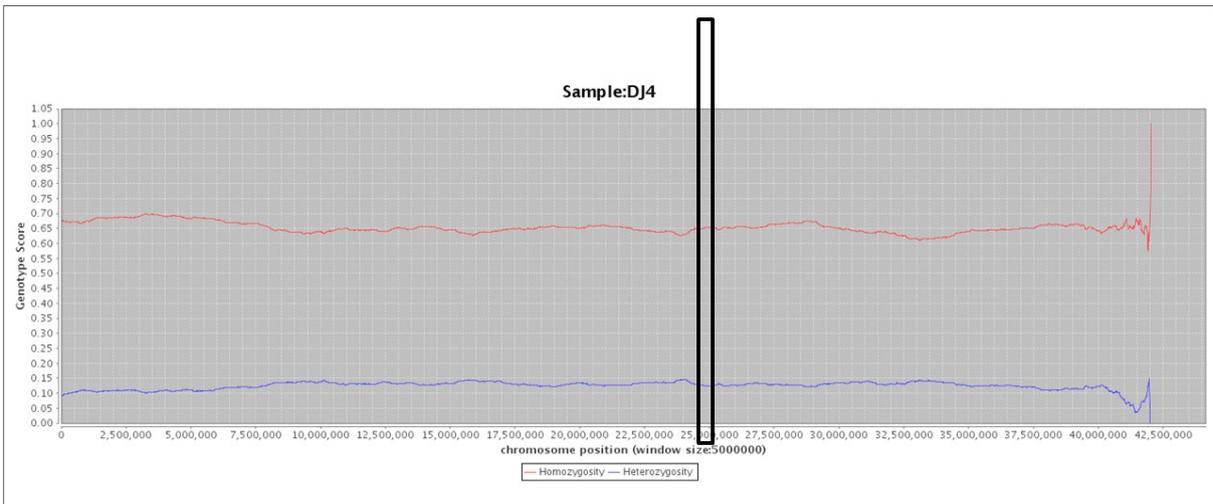
**Appendix Figure H 79 Homozygosity profile signal for chromosome 24 of DJ1 showing the Trypanotolerance *CD19* candidate loci**



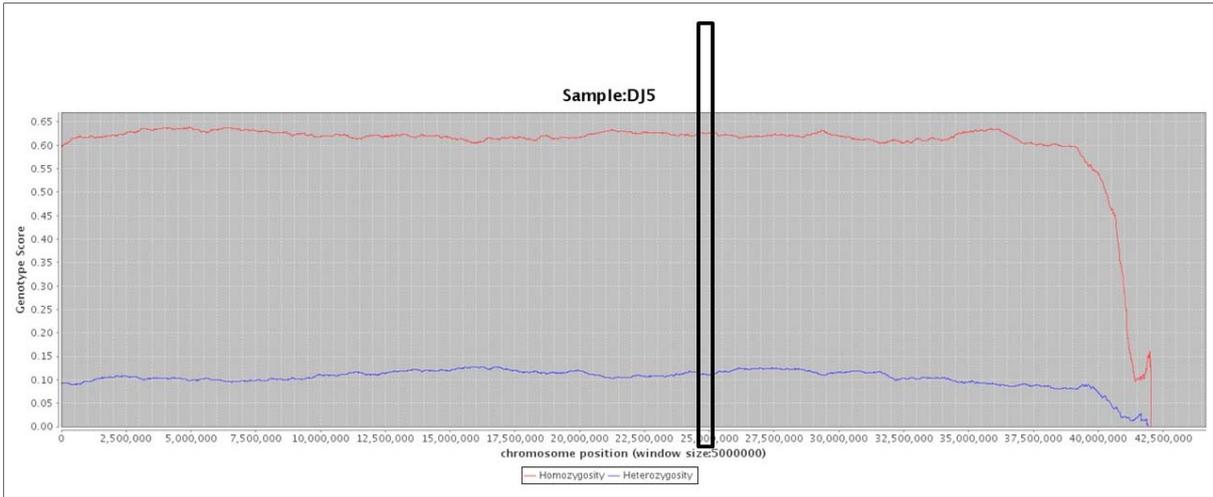
**Appendix Figure H 80 Homozygosity profile signal for chromosome 24 of DJ2 showing the Trypanotolerance *CD19* candidate loci**



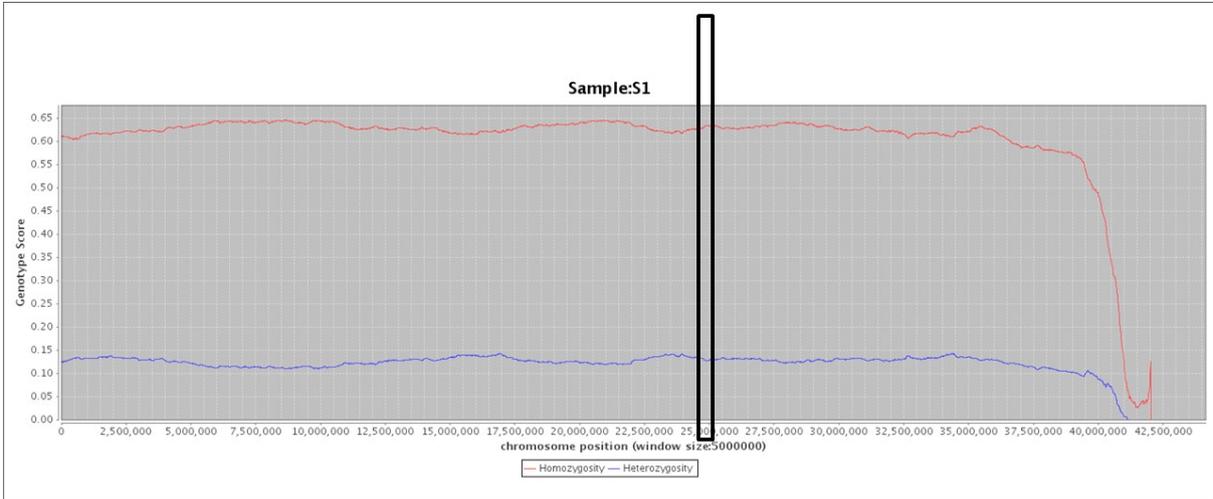
**Appendix Figure H 81 Homozygosity profile signal for chromosome 24 of DJ3 showing the Trypanotolerance *CD19* candidate loci**



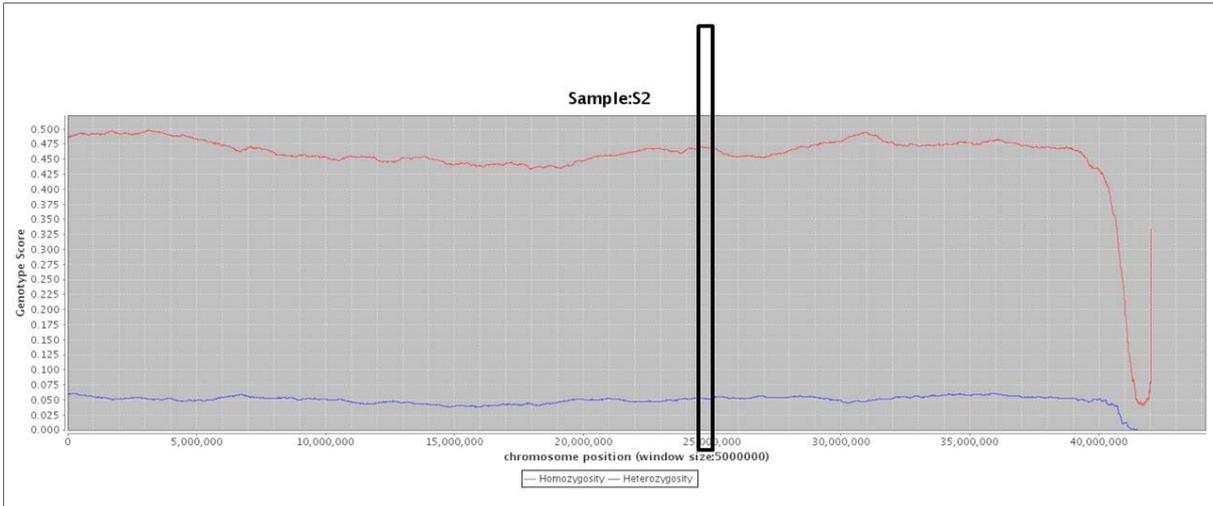
**Appendix Figure H 82 Homozygosity profile signal for chromosome 24 of DJ4 showing the Trypanotolerance *CD19* candidate loci**



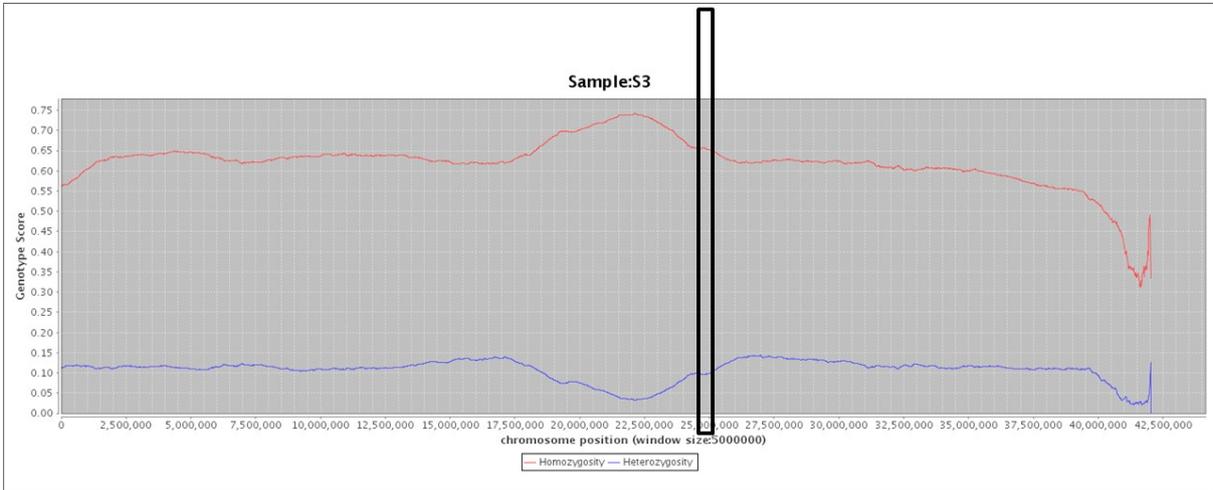
**Appendix Figure H 83 Homozygosity profile signal for chromosome 24 of DJ5 showing the Trypanotolerance *CD19* candidate loci**



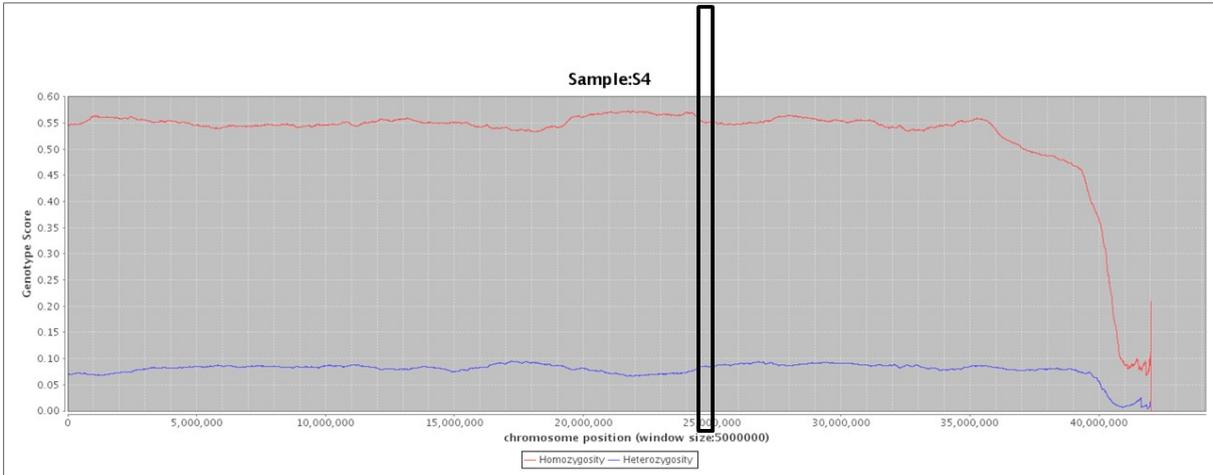
**Appendix Figure H 84 Homozygosity profile signal for chromosome 24 of S1 showing the Trypanotolerance *CD19* candidate loci**



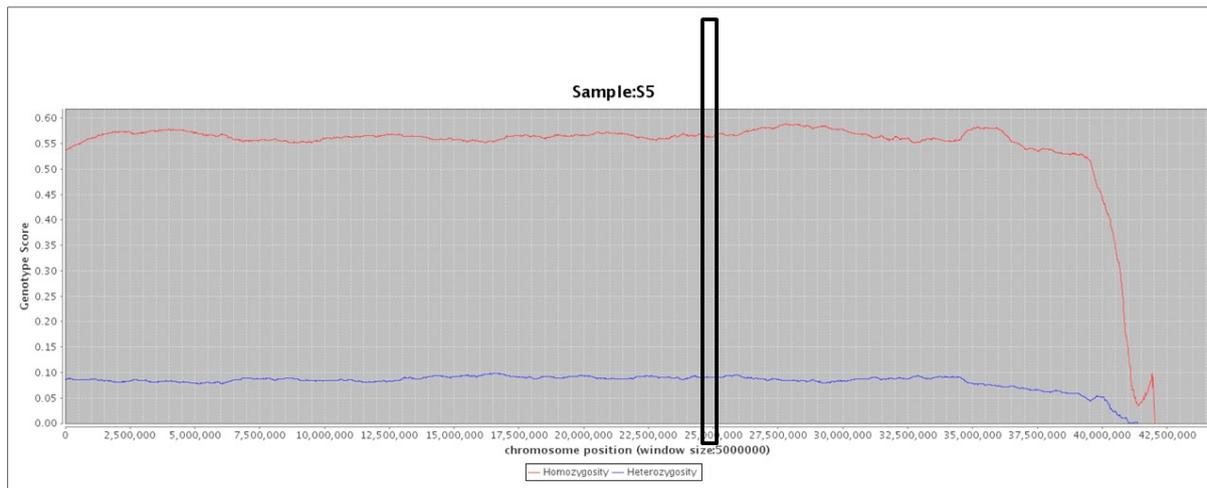
**Appendix Figure H 85 Homozygosity profile signal for chromosome 24 of S2 showing the Trypanotolerance *CD19* candidate loci**



**Appendix Figure H 86 Homozygosity profile signal for chromosome 24 of S3 showing the Trypanotolerance *CD19* candidate loci**

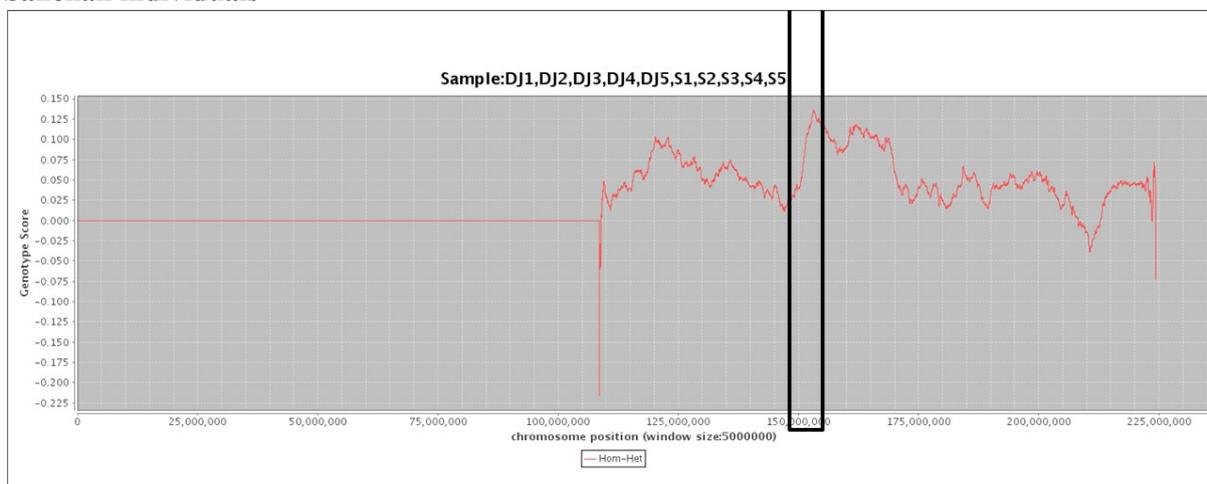


**Appendix Figure H 87 Homozygosity profile signal for chromosome 24 of S4 showing the Trypanotolerance *CD19* candidate loci**



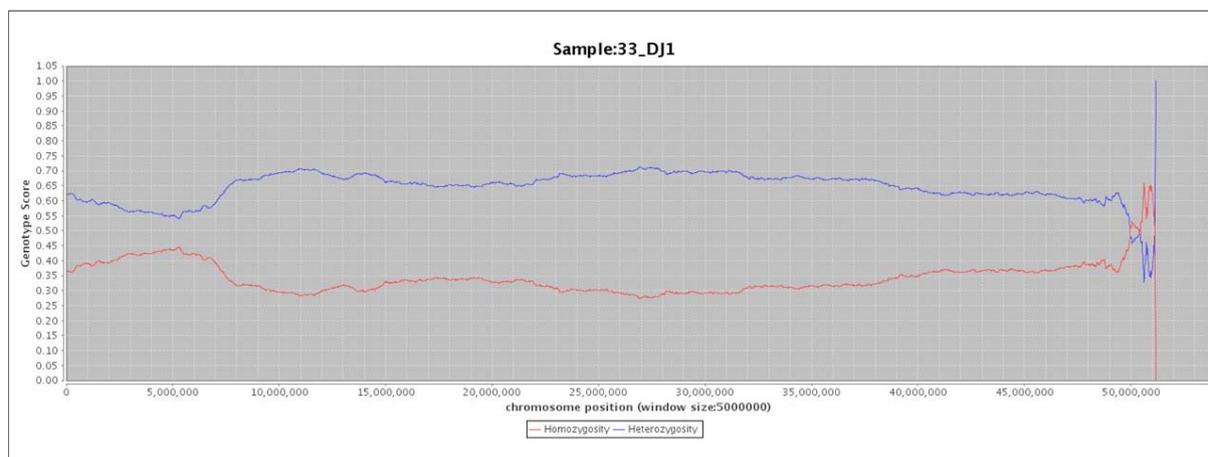
**Appendix Figure H 88 Homozygosity profile signal for chromosome 24 of S5 showing the Trypanotolerance *CD19* candidate loci**

**HomHet predictive signal for the *MSRB3* candidate gene (Djallonke (high ROH) and Sahelian (low ROH)) associated with adaptive selection for all the Djallonke individuals relative to all the Sahelian individuals**

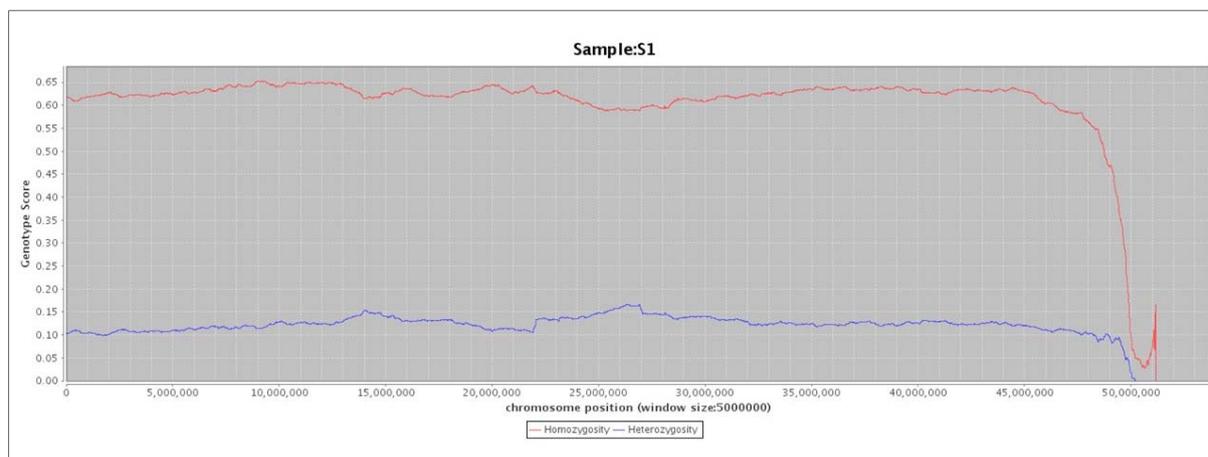


**Appendix Figure H 89 HomHet predictive signal for chromosome 3 showing peak signal at the *MSRB3* candidate loci**

HomHet analysis for chromosome 20 showing homozygosity profile for pooled Djallonke sheep and pooled Sahelian sheep



**Appendix Figure H 90 Homozygosity profile signal for chromosome 20 of pooled Djallonke samples showing the decreased homozygosity over the entire region (containing the MHC)**



**Appendix Figure H 91 Homozygosity profile signal for chromosome 20 of pooled Sahelian samples showing the increase homozygosity over the entire region (containing the MHC)**

**Appendix I    Co-Authors consent for thesis  
associated publications and  
copyright agreements for  
publications**

To Whom It May Concern,

I, Mohammed Yaro, as the first author of the publication entitled “Whole genome analysis of pooled sequences from Djallonke and Sahelian Sheep of Ghana reveals co-localisation of regions of reduced heterozygosity with candidate genes for disease resistance and adaptation to a tropical environment”, declare that this work was primarily designed, experimentally executed, interpreted, and written by the first author of this manuscript.

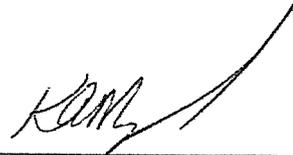


First author signature

I, as a Co-author, endorse that this level of contribution by the first author indicated above is appropriate.

Kylie A. Munyard

Co-Author 1 printed name



Co-Author 1 signature

Eleanor Morgan

Co-Author 2 printed name



Co-Author 2 signature

Richard N.J. Allcock

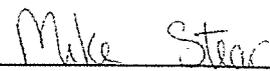
Co-Author 3 printed name



Co-Author 3 signature

Michael J. Stear

Co-Author 4 printed name



Co-Author 4 signature

David M. Groth

Co-Author 5 printed name



Co-Author 5 signature

To Whom It May Concern,

I, Mohammed Yaro, as the first author of the publication entitled "P4041 Pooled whole-genome sequencing reveals molecular signatures of natural adaptive selection in Djallonke sheep of Ghana. Journal of Animal Science, 94(7supplement4), 98-99. <http://dx.doi.org/10.2134/jas2016.947supplement498a>", declare that this work was primarily designed, experimentally executed, interpreted, and written by the first author of this manuscript.



---

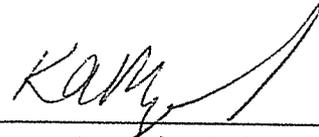
First author signature

I, as a Co-author, endorse that this level of contribution by the first author indicated above is appropriate.

Kylie A. Munyard

---

Co-Author 1 printed name



---

Co-Author 1 signature

Eleanor Morgan

---

Co-Author 2 printed name



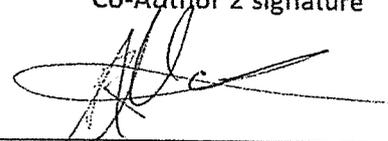
---

Co-Author 2 signature

Richard N.J. Allcock

---

Co-Author 3 printed name



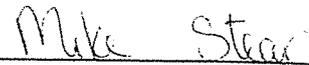
---

Co-Author 3 signature

Michael J. Stear

---

Co-Author 4 printed name



---

Co-Author 4 signature

David M. Groth

---

Co-Author 5 printed name



---

Co-Author 5 signature

To Whom It May Concern,

I, Mohammed Yaro, as the first author of the publication entitled "**Combatting African Animal trypanosomiasis: the potential role of trypanotolerance, Veterinary Parasitology, 225, 43-52.** <http://dx.doi.org/http://dx.doi.org/10.1016/j.vetpar.2016.05.003>", declare that this work was primarily designed, experimentally executed, interpreted, and written by the first author of this manuscript.



---

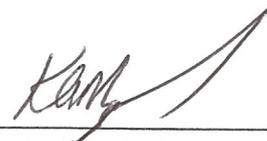
**First author signature**

I, as a Co-author, endorse that this level of contribution by the first author indicated above is appropriate.

Kylie A. Munyard

---

**Co-Author 1 printed name**



---

**Co-Author 1 signature**

Michael J. Stear

---

**Co-Author 2 printed name**



---

**Co-Author 2 signature**

David M. Groth

---

**Co-Author 3 printed name**



---

**Co-Author 3 signature**

To Whom It May Concern,

I, Mohammed Yaro, as the first author of the publication entitled "*Molecular identification of livestock breeds: a tool for modern conservation biology. Biological Reviews, Cambridge Philosophical Society, 92(2), 993-1010. <http://dx.doi.org/10.1111/brv.12265>*", declare that this work was primarily designed, experimentally executed, interpreted, and written by the first author of this manuscript.



---

**First author signature**

I, as a Co-author, endorse that this level of contribution by the first author indicated above is appropriate.

Kylie A. Munyard

---

**Co-Author 1 printed name**



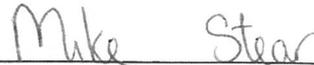
---

**Co-Author 1 signature**

Michael J. Stear

---

**Co-Author 2 printed name**



---

**Co-Author 2 signature**

David M. Groth

---

**Co-Author 3 printed name**



---

**Co-Author 3 signature**

**JOHN WILEY AND SONS LICENSE  
TERMS AND CONDITIONS**

Aug 28, 2017

---

This Agreement between Mr. Mohammed Yaro ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4177941271316
License date	Aug 28, 2017
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Biological Reviews
Licensed Content Title	Molecular identification of livestock breeds: a tool for modern conservation biology
Licensed Content Author	Mohammed Yaro,Kylie A. Munyard,Michael J. Stear,David M. Groth
Licensed Content Date	Mar 22, 2016
Licensed Content Pages	18
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Title of your thesis / dissertation	Characterisation of the Djallonke sheep in Ghana using Molecular markers
Expected completion date	Aug 2017
Expected size (number of pages)	250
Requestor Location	Mr. Mohammed Yaro Curtin University B305 Kent street, Bentley Perth, WA 6102 Australia Attn: Mr. Mohammed Yaro

Publisher Tax ID	EU826007151
Billing Type	Invoice
Billing Address	Mr. Mohammed Yaro Curtin University B305 Kent street, Bentley Perth, Australia 6102 Attn: Mr. Mohammed Yaro
Total	0.00 USD

#### Terms and Conditions

### TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a "Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work (collectively "WILEY"). By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the billing and payment terms and conditions established by the Copyright Clearance Center Inc., ("CCC's Billing and Payment terms and conditions"), at the time that you opened your RightsLink account (these are available at any time at <http://myaccount.copyright.com>).

#### Terms and Conditions

- The materials you have requested permission to reproduce or reuse (the "Wiley Materials") are protected by copyright.
- You are hereby granted a personal, non-exclusive, non-sub licensable (on a stand-alone basis), non-transferable, worldwide, limited license to reproduce the Wiley Materials for the purpose specified in the licensing process. This license, **and any CONTENT (PDF or image file) purchased as part of your order**, is for a one-time use only and limited to any maximum distribution number specified in the license. The first instance of republication or reuse granted by this license must be completed within two years of the date of the grant of this license (although copies prepared before the end date may be distributed thereafter). The Wiley Materials shall not be used in any other manner or for any other purpose, beyond what is granted in the license. Permission is granted subject to an appropriate acknowledgement given to the author, title of the material/book/journal and the publisher. You shall also duplicate the copyright notice that appears in the Wiley publication in your use of the Wiley

Material. Permission is also granted on the understanding that nowhere in the text is a previously published source acknowledged for all or part of this Wiley Material. Any third party content is expressly excluded from this permission.

- With respect to the Wiley Materials, all rights are reserved. Except as expressly granted by the terms of the license, no part of the Wiley Materials may be copied, modified, adapted (except for minor reformatting required by the new Publication), translated, reproduced, transferred or distributed, in any form or by any means, and no derivative works may be made based on the Wiley Materials without the prior permission of the respective copyright owner. **For STM Signatory Publishers clearing permission under the terms of the [STM Permissions Guidelines](#) only, the terms of the license are extended to include subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and does not involve the separate exploitation of the permitted figures or extracts,** You may not alter, remove or suppress in any manner any copyright, trademark or other notices displayed by the Wiley Materials. You may not license, rent, sell, loan, lease, pledge, offer as security, transfer or assign the Wiley Materials on a stand-alone basis, or any of the rights granted to you hereunder to any other person.
- The Wiley Materials and all of the intellectual property rights therein shall at all times remain the exclusive property of John Wiley & Sons Inc, the Wiley Companies, or their respective licensors, and your interest therein is only that of having possession of and the right to reproduce the Wiley Materials pursuant to Section 2 herein during the continuance of this Agreement. You agree that you own no right, title or interest in or to the Wiley Materials or any of the intellectual property rights therein. You shall have no rights hereunder other than the license as provided for above in Section 2. No right, license or interest to any trademark, trade name, service mark or other branding ("Marks") of WILEY or its licensors is granted hereunder, and you agree that you shall not assert any such right, license or interest with respect thereto
- NEITHER WILEY NOR ITS LICENSORS MAKES ANY WARRANTY OR REPRESENTATION OF ANY KIND TO YOU OR ANY THIRD PARTY, EXPRESS, IMPLIED OR STATUTORY, WITH RESPECT TO THE MATERIALS OR THE ACCURACY OF ANY INFORMATION CONTAINED IN THE MATERIALS, INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTY OF MERCHANTABILITY, ACCURACY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, USABILITY, INTEGRATION OR NON-INFRINGEMENT AND ALL SUCH WARRANTIES ARE HEREBY EXCLUDED BY WILEY AND ITS LICENSORS AND WAIVED

## BY YOU.

- WILEY shall have the right to terminate this Agreement immediately upon breach of this Agreement by you.
- You shall indemnify, defend and hold harmless WILEY, its Licensors and their respective directors, officers, agents and employees, from and against any actual or threatened claims, demands, causes of action or proceedings arising from any breach of this Agreement by you.
- IN NO EVENT SHALL WILEY OR ITS LICENSORS BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR ENTITY FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, EXEMPLARY OR PUNITIVE DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, PROVISIONING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.
- Should any provision of this Agreement be held by a court of competent jurisdiction to be illegal, invalid, or unenforceable, that provision shall be deemed amended to achieve as nearly as possible the same economic effect as the original provision, and the legality, validity and enforceability of the remaining provisions of this Agreement shall not be affected or impaired thereby.
- The failure of either party to enforce any term or condition of this Agreement shall not constitute a waiver of either party's right to enforce each and every term and condition of this Agreement. No breach under this agreement shall be deemed waived or excused by either party unless such waiver or consent is in writing signed by the party granting such waiver or consent. The waiver by or consent of a party to a breach of any provision of this Agreement shall not operate or be construed as a waiver of or consent to any other or subsequent breach by such other party.

- This Agreement may not be assigned (including by operation of law or otherwise) by you without WILEY's prior written consent.
- Any fee required for this permission shall be non-refundable after thirty (30) days from receipt by the CCC.
- These terms and conditions together with CCC's Billing and Payment terms and conditions (which are incorporated herein) form the entire agreement between you and WILEY concerning this licensing transaction and (in the absence of fraud) supersedes all prior agreements and representations of the parties, oral or written. This Agreement may not be amended except in writing signed by both parties. This Agreement shall be binding upon and inure to the benefit of the parties' successors, legal representatives, and authorized assigns.
- In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall prevail.
- WILEY expressly reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
- This Agreement will be void if the Type of Use, Format, Circulation, or Requestor Type was misrepresented during the licensing process.
- This Agreement shall be governed by and construed in accordance with the laws of the State of New York, USA, without regards to such state's conflict of law rules. Any legal action, suit or proceeding arising out of or relating to these Terms and Conditions or the breach thereof shall be instituted in a court of competent jurisdiction in New York County in the State of New York in the United States of America and each party hereby consents and submits to the personal jurisdiction of such court, waives any objection to venue in such court and consents to service of process by registered or certified mail, return receipt requested, at the last known address of such party.

### **WILEY OPEN ACCESS TERMS AND CONDITIONS**

Wiley Publishes Open Access Articles in fully Open Access Journals and in Subscription journals offering Online Open. Although most of the fully Open Access journals publish

open access articles under the terms of the Creative Commons Attribution (CC BY) License only, the subscription journals and a few of the Open Access Journals offer a choice of Creative Commons Licenses. The license type is clearly identified on the article.

#### **The Creative Commons Attribution License**

The [Creative Commons Attribution License \(CC-BY\)](#) allows users to copy, distribute and transmit an article, adapt the article and make commercial use of the article. The CC-BY license permits commercial and non-

#### **Creative Commons Attribution Non-Commercial License**

The [Creative Commons Attribution Non-Commercial \(CC-BY-NC\) License](#) permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.(see below)

#### **Creative Commons Attribution-Non-Commercial-NoDerivs License**

The [Creative Commons Attribution Non-Commercial-NoDerivs License](#) (CC-BY-NC-ND) permits use, distribution and reproduction in any medium, provided the original work is properly cited, is not used for commercial purposes and no modifications or adaptations are made. (see below)

#### **Use by commercial "for-profit" organizations**

Use of Wiley Open Access articles for commercial, promotional, or marketing purposes requires further explicit permission from Wiley and will be subject to a fee.

Further details can be found on Wiley Online Library

<http://olabout.wiley.com/WileyCDA/Section/id-410895.html>

#### **Other Terms and Conditions:**

**v1.10 Last updated September 2015**

**Questions? [customercare@copyright.com](mailto:customercare@copyright.com) or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.**

**ELSEVIER LICENSE  
TERMS AND CONDITIONS**

Aug 28, 2017

---

This Agreement between Mr. Mohammed Yaro ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4177940126518
License date	Aug 28, 2017
Licensed Content Publisher	Elsevier
Licensed Content Publication	Veterinary Parasitology
Licensed Content Title	Combatting African Animal Trypanosomiasis (AAT) in livestock: The potential role of trypanotolerance
Licensed Content Author	M. Yaro,K.A. Munyard,M.J. Stear,D.M. Groth
Licensed Content Date	Jul 30, 2016
Licensed Content Volume	225
Licensed Content Issue	n/a
Licensed Content Pages	10
Start Page	43
End Page	52
Type of Use	reuse in a thesis/dissertation
Portion	full article
Format	both print and electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Order reference number	1
Title of your thesis/dissertation	Characterisation of the Djallonke sheep in Ghana using Molecular markers
Expected completion date	Aug 2017
Estimated size (number of pages)	250
Requestor Location	Mr. Mohammed Yaro Curtin University B305 Kent street, Bentley Perth, WA 6102 Australia Attn: Mr. Mohammed Yaro
Total	0.00 USD
Terms and Conditions	

**INTRODUCTION**

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and

conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

### GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at [permissions@elsevier.com](mailto:permissions@elsevier.com)). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. **Objection to Contrary Terms:** Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. **Revocation:** Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

### LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation:** This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. **Posting licensed content on any Website:** The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx> or the Elsevier homepage for books at <http://www.elsevier.com>; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at <http://www.elsevier.com>. All content posted to the web site must maintain the copyright information line on the bottom of each image.

**Posting licensed content on Electronic reserve:** In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

#### **Preprints:**

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available

version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

**Accepted Author Manuscripts:** An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- immediately
  - via their non-commercial person homepage or blog
  - by updating a preprint in arXiv or RePEc with the accepted manuscript
  - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
  - directly by providing copies to their students or to research collaborators for their personal use
  - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- After the embargo period
  - via non-commercial hosting platforms such as their institutional repository
  - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

**Published journal article (JPA):** A published journal article (JPA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

**Subscription Articles:** If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

**Gold Open Access Articles:** May be shared according to the author-selected end-user license and should contain a [CrossMark logo](#), the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's [posting policy](#) for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you

scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation:** If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

### **Elsevier Open Access Terms and Conditions**

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our [open access license policy](#) for more information.

#### **Terms & Conditions applicable to all Open Access articles published with Elsevier:**

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

#### **Additional Terms & Conditions applicable to each Creative Commons user license:**

**CC BY:** The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at <http://creativecommons.org/licenses/by/4.0>.

**CC BY NC SA:** The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-sa/4.0>.

**CC BY NC ND:** The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-nd/4.0>. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access

- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

## 20. Other Conditions:

v1.9

Questions? [customercare@copyright.com](mailto:customercare@copyright.com) or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

---

---