Analysis of relatedness and genetic diversity in Swedish sheep breeds using medium density (50K) SNP data

Krishani Devika Wijesinghe Sinhalage

Independent project • 30 credits
Swedish University of Agricultural Sciences, SLU
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics
European Master in Animal Breeding and Genetics (EMABG)
Uppsala 2023

Analysis of relatedness and genetic diversity in Swedish sheep breeds using medium density (50K) SNP data

Krishani Devika Wijesinghe Sinhalage

Supervisor:	**Dr. Anna Maria Johansson, SLU, Animal Breeding and Genetics**
Assistant supervisor:	Dr. Steffen Weigend, UGOE, Animal Breeding and Genetics
Assistant supervisor:	Dr. Christian Reimer, Friedrich-Loeffler-Institute (FLI), Germany
Examiner:	Prof. Erling Strandberg, SLU, Animal Breeding and Genetics

Credits:	30 credits
Level:	Advanced, A2E
Course title:	Independent Project in Animal Science
Course code:	EX0870
Programme/education:	European Master in Animal Breeding and Genetics (EMABG)
Course coordinating dept:	Animal Breeding and Genetics
Place of publication:	Uppsala
Year of publication:	2023

Keywords:	sheep, genetic diversity, population structure, inbreeding, runs of homozygosity, coat colour, MC1R

Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics

Abstract

The local Swedish sheep breeds belong to a group of sheep breeds known as North European short-tailed sheep breeds. Evaluation of genetic diversity, relatedness, and population structure of local breeds offers crucial metrics to prevent inbreeding depression, genetic erosion, and crossbreeding between locally adapted and exotic sheep breeds since these occurrences can have detrimental effects and ultimately lead to extinction. The aim of this thesis was to study genetic diversity and the population structure of 191 sheep belonging to fourteen Swedish native sheep breeds using medium density (50K) SNP genotype data. The population structure was studied with principal coordinate analysis (PCoA), neighbour-joining (NJ) trees (a, b), and Structure analysis. The genetic diversity was assessed based on the inbreeding coefficients, average observed heterozygosity, minor allele frequency (MAF), polymorphic information content (PIC), and pairwise F_{ST}. The level of inbreeding was estimated using two measures: inbreeding coefficient (F_{IS}) and runs of homozygosity-based inbreeding (FROH). Moreover, genetic variants in the melanocortin 1 receptor (MC1R) gene that is responsible for coat colour in sheep were examined using seven Swedish sheep breeds. The population structure analyses using Structure v.2.3.4 showed twelve distinct genetic populations among fourteen breeds. Rya sheep showed high genetic admixture within the population. The Roslag sheep were highly varied from the other breeds (PC1). The Tabacktorp sheep had the longest branch length in the NJ tree (b). The highest average inbreeding level (F_{IS}) was reported in Roslag sheep (0.25), while the highest average FROH was observed in Tabacktorp sheep (0.2). The highest F_{ST} value (0.52) was observed between Roslag and Tabacktorp populations. Moreover, two previously discovered mutations in the MC1R gene in sheep were found in this thesis: c.-31 G>A and c.218 T>A. The c.218 T>A mutation was associated with black/brown coat colour in Finull sheep. In conclusion, Roslag was highly varied from other breeds and appeared to be genetically isolated with no genetic admixture. High genetic drift was observed within the Tabacktorp sheep population. A high F_{ST} value indicates high genetic differentiation between Roslag and Tabacktorp breeds. Furthermore, relatively high inbreeding levels in Roslag and Tabacktorp sheep indicate low genetic diversity within populations.

Keywords: sheep, genetic diversity, population structure, inbreeding, runs of homozygosity, coat colour, MC1R

Table of contents

List of tables **6**

List of figures **7**

Abbreviations **12**

1. Introduction **13**

1.1 Background 13

1.2 Statement of Problem 14

1.3 Objectives 15

1.3.1 Main Objective 15

1.3.2 Specific Objectives 15

2. Literature Review **17**

2.1 Sheep Domestication 17

2.2 North European Short-tailed Sheep Breeds and Swedish Sheep Breeds 17

2.3 Genetic Diversity 18

2.4 Sheep Coat Colour Genes 19

2.5 Melanocortin 1 Receptor (MC1R) Gene 20

2.6 Genetic Diversity Measures 21

2.7 Population Structure 23

3. Materials and Methods **25**

3.1 Data Source 25

3.2 Sheep Sample 25

3.3 Genotyping 26

3.4 Data Merging, Removal of Related Individuals, and SNP Quality Control 26

3.5 Genetic Diversity Analysis 28

3.5.1 Observed Heterozygosity 28

3.5.2 Inbreeding Coefficient (F_{IS}) 29

3.5.3 Minor Allele Frequency (MAF) 29

3.5.4 Polymorphic Information Content (PIC) 30

3.5.5 Pairwise F_{ST} (Wright's F_{ST} Estimate / Fixation Index) 30

3.5.6 Runs of Homozygosity based Inbreeding Coefficient (FROH) 30

3.6 Genetic Population Structure 32

3.6.1 Principal Coordinate Analysis (PCoA) 32

3.6.2 The Neighbour-joining (NJ) Method 32
3.6.3 Population Admixture Analysis using Structure Software v.2.3.4 33
3.7 DNA Sequencing, Multiple Sequencing Alignment and Determination of Variants in MC1R that Responsible for Black Coat Colour 33
3.7.1 DNA Sequencing 34
3.7.2 Sequencing Analysis 34

4. Results 35
4.1 Population Structure 35
4.1.1 Principal Coordinate Analysis (PCoA) 35
4.1.2 Neighbour-joining Trees 36
4.1.3 Population Admixture Analysis 39
4.2 Genetic Diversity 40
4.2.1 Observed Heterozygosity 43
4.2.2 Inbreeding Coefficient (F_{IS}) 44
4.2.3 Inbreeding Coefficient based on Runs of Homozygosity (FROH) 45
4.3 Genetic Variants in MC1R Gene that Responsible for Coat Colour in Sheep 47

5. Discussion 49
5.1 Population Structure 49
5.2 Genetic Diversity 51
5.3 Genetic Variants in the MC1R Gene that Responsible for Coat Colour in Sheep 52

6. Conclusions 55

7. References 57

Popular science summary 65

Acknowledgements 67

Appendix 1 - Structure Analysis 69

Appendix 2 – ROH Plots per Chromosome 83

Appendix 3 – Genetic Variants in MC1R 97

List of tables

Table 1 Information on the number of individuals from each breed and the geographical origin 25

Table 2 SNP quality control parameters 28

Table 3 List of parameters for the ROH analysis with the detectRUNS package in R software 31

Table 4 Information on breeds and the number of samples from each breed 33

Table 5 Summary of Genetic Diversity estimates of fourteen native Swedish sheep breeds 41

Table 6 Pairwise F_{ST} (Wright's Fixation index) estimates of twelve native Swedish sheep breeds 42

Table 7 The average lengths of ROH and the number of ROH segments observed within each breed 46

Table 8 Detected phenotypes of c.218 T>A in Melanocortin receptor 1 (MC1R) by coat colour and breed 47

Table 9 Summary of genetic variants in the Melanocortin 1 receptor (MC1R) gene and associated phenotypic variability in five native Swedish sheep breeds 97

List of figures

Figure 1 Distribution of missing loci in individuals. Log10(-1.5) was used for the threshold determination. Accordingly, we excluded SNPs with a missing rate of 0.03 (3%) or higher (geno). 27

Figure 2 Distribution of missing values in individuals. Log10(-1.5) was used for the threshold determination. Accordingly, we excluded individuals with a missing rate of 0.03 (3%) or higher (mind)..... 27

Figure 3 Genetic relatedness of fourteen Swedish sheep breeds using principal coordinate analyses (PCoA) (multidimensional scaling plots). The first two coordinates (PC1 and PC2) (left) and the third coordinate (PC3) (right) explained 9.15%, 6.51%, and 5.93%, respectively, of the total variation. 36

Figure 4 Population relationship and structure of native Swedish sheep breeds using the Neighbour-joining Trees based on the genetic distance: (a) Fan type NJ tree: within-breed genetic relatedness of 191 sheep from 14 Swedish breeds. (b) Unrooted NJ tree (b): genetic relatedness between 12 Swedish sheep breeds. 38

Figure 5 Genome-wide admixture proportions at K=12 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 12 assumed genetic populations. The numbers denoted in the x-axis show different populations, and here 14 Swedish sheep breeds used as different populations: 1 Roslag; 2 Brännö; 3 Gammelrya; 4 Tabacktorp; 5 Rya; 6 Gotland; 7 Gute; 8 Åsen; 9 Finull; 10 Värmland; 11 Svärdsjö; 12 Helsinge; 13 Gestrike; 14 Fjällnäs..... 39

Figure 6 The observed heterozygosity of 189 individuals belonging to 13 Swedish sheep breeds. 43

Figure 7 The inbreeding coefficients of 189 individuals belonging to 13 native Swedish sheep breeds. The colours in the plot match the breed colours in the previous figure. 44

Figure 8 The inbreeding coefficient based on ROH for 14 native Swedish sheep breeds. The minimum length of a ROH was defined as 15 SNPs. For other parameters, see the materials and method section. 45

Figure 9 Genome-wide admixture proportions at K=2 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 2 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs. 69

Figure 10 Genome-wide admixture proportions at K=3 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 3 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs. 70

Figure 11 Genome-wide admixture proportions at K=4 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 4 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs. 71

Figure 12 Genome-wide admixture proportions at K=5 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 5 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs. 72

Figure 13 Genome-wide admixture proportions at K=6 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 6 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs. 73

Figure 14 Genome-wide admixture proportions at K=7 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 7 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs. 74

Figure 15 Genome-wide admixture proportions at K=8 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 8 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs. 75

Figure 16 Genome-wide admixture proportions at K=9 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 9 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs. 76

Figure 17 Genome-wide admixture proportions at K=10 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 10 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs. 77

Figure 18 Genome-wide admixture proportions at K=11 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 11 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs. 78

Figure 19 Genome-wide admixture proportions at K=12 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 12 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3,

Gammelrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs. 79

Figure 20 Genome-wide admixture proportions at K=13 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 13 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs. 80

Figure 21 Genome-wide admixture proportions at K=14 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 14 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs. 81

Figure 22 ROH segments on Chromosome 1. 83

Figure 23 ROH segments on Chromosome 2. 84

Figure 24 ROH segments on Chromosome 3. 84

Figure 25 ROH segments on Chromosome 4. 85

Figure 26 ROH segments on Chromosome 5. 85

Figure 27 ROH segments on Chromosome 6. 86

Figure 28 ROH segments on Chromosome 7. 86

Figure 29 ROH segments on Chromosome 8. 87

Figure 30 ROH segments on Chromosome 9. 87

Figure 31 ROH segments on Chromosome 10. 88

Figure 32 ROH segments on Chromosome 11. 88

Figure 33 ROH segments on Chromosome 12. 89

Figure 34 ROH segments on Chromosome 13. 89

Figure 35 ROH segments on Chromosome 14. 90

Figure 36 ROH segments on Chromosome 15. 90

Figure 37 ROH segments on Chromosome 16. 91

Figure 38 ROH segments on Chromosome 17. 91

Figure 39 ROH segments on Chromosome 18. 92

Figure 40 ROH segments on Chromosome 19.....92
Figure 41 ROH segments on Chromosome 20.....93
Figure 42 ROH segments on Chromosome 21.....93
Figure 43 ROH segments on Chromosome 22.....94
Figure 44 ROH segments on Chromosome 23.....94
Figure 45 ROH segments on Chromosome 24.....95
Figure 46 ROH segments on Chromosome 25.....95
Figure 47 ROH segments on Chromosome 26.....96

Abbreviations

F_{IS}	Inbreeding Coefficient
FROH	Inbreeding Coefficient based on Runs of Homozygosity
F_{ST}	Pairwise F_{ST}/ Fixation Index
HWE	Hardy Weinberg Equillibrium
IBD	Identical by Descent
IBS	Identical by State
LD	Linkage Disequilibrium
MAF	Minor Allele Frequency
MC1R	Melanocortin 1 Receptor
NJ	Neighbor-joining
PC	Principal Coordinate
PCoA	Principal Coordinate Analysis
PIC	Polymorphic Information Content
ROH	Runs of Homozygosity
SNP	Single Nucleotide Polymorphism

1. Introduction

1.1 Background

The sheep was one of the first grazing animals to be domesticated in part due to its ability to adapt to diverse climates and diets with low-quality nutritious value (Zeder, 2008). Later, a large number of breeds with different morphological characteristics (absence of horns), coat colour patterns, or specialized production (milk, meat, or wool) were shaped by artificial selection by farmers and breeders (Fariello et al., 2014). A study by Kijas et al. (2012) showed that the sheep breeds had preserved significant levels of genetic variation, unlike other domestic species such as dogs. Furthermore, sheep showed a number of genomic regions with high differentiation between breeds, including candidate genes for growth, body size, reproduction and coat colour (Kijas et al., 2012). Swedish sheep breeds belong to a group of sheep referred to as North European short-tailed sheep breeds, which are primitive and genetically unique, display various phenotypes (particularly coat colour patterns), and are one of the major contributors to genetic variation in domestic sheep (Tapio et al., 2005).

However, genetic diversity within and between breeds of current global livestock populations is declining. Plenty of factors have influenced the genetic dilution and loss of native breeds. This has been significantly influenced by the competition for highly specialized breeds with high economic value in the current production background (Boettcher et al., 2010). Furthermore, local breeds tend to decline as a result of introduction of non-native breeds. According to the FAO Commission on Genetic Resources for Food and Agriculture 2022 report, cattle, sheep and horses have the largest numbers of breeds at risk among mammalian species. It was shown that there are 660 sheep breeds out of 1491 (about 44%) at unknown risk status.

Genetic diversity is essential for genetic variation in a biological population. Therefore, the genetic resources of livestock for food and agriculture are the fundamental biological capital for livestock development and are vital for combating food insecurity and malnutrition (Hoffmann, 2010). In addition, the diversity of livestock genetic resources is important for sustainable and increased productivity and adaptation to diverse environmental conditions, including climate-

associated conditions, changes in husbandry practices, management and market strategies and disease and parasitic challenges (Hanotte O. et al., 2005). The evaluation of genetic variability, relatedness and the structure of the populations is a vital task, not only to determine the evolutionary history of breed origins but also to provide crucial information for managing and conserving local biodiversity (Williams et al., 2016). Furthermore, this information offers essential metrics to prevent inbreeding depression, genetic erosion and crossbreeding between locally adapted and exotic breeds, as these events have detrimental effects that ultimately lead to breed extinction (Senczuk et al., 2020).

Genetic diversity can be assessed using both pedigree and genomic-based approaches (Biscarini et al., 2020). It can now be accurately estimated on large datasets in livestock species without pedigree information with the advancement of next-generation high-throughput sequencing and genotyping techniques, such as high-density SNP arrays and microsatellite genotypes (Li et al., 2011). This thesis study assessed genetic diversity, population structure and genetic admixture in native Swedish breeds using medium-density Illumina Ovine SNP50v3 BeadChip data and further investigated genetic variants in the melanocortin 1 receptor (MC1R) gene using sequencing data that contribute to distinctive coat colour patterns.

1.2 Statement of Problem

Information about the history of domestication and how livestock were exported and bred by farmers is encoded in the DNA of domesticated animals. Next-generation high-throughput genotyping techniques have enabled population geneticists to use genome-wide markers to examine the histories of many species, including sheep, humans, cattle, dogs, mice, horses, yeast, grapes, wheat and maize (Decker et al., 2014). Post-domestication, the evolution of phenotypically different sheep breeds has generated substantial biodiversity due to genomic variations (Kalds et al., 2022). Domestic sheep can be found worldwide in various environmental conditions and are raised for the production of milk, meat, and wool.

Native sheep breeds need special attention as they are valuable reservoirs of genetic diversity (Schmidtmann et al., 2021). Their conservation requires proper breeding decisions and sufficient genetic management of small populations to preserve genetic variability and minimize inbreeding. Studies of genetic diversity and population structure shed light on the genetic relatedness and differences between breeds, a source of information for implementing successful breeding programs, future breed development and future scientific research (Abebe et al., 2015). Some studies have shown the genetic diversity of a few Swedish sheep breeds using pedigree data, high-density SNP chip data and microsatellite genotype data (Mukiibi et al., 2015; Rochus et al., 2020; Rochus et al., 2019; Tapio et al.,

2005). The genetic relationship of Swedish sheep breeds has been assessed using endogenous retroviruses as molecular markers by Mukiibi et al. (2015). Moreover, Rochus et al. (2019) investigated the genetic variants in two coat colour candidate genes: MC1R and agouti signaling protein (ASIP), and their association with coat colour, which revealed different selection histories in these Swedish sheep breeds. The high-density (600K) SNP array study showed that Fjällnäs, Klövsjö, Dalapäls, Gute and Gotland were distinct and unique from each other and individuals within each breed were clustered together (Rochus et al., 2020).

However, this genetic diversity information of some local Swedish sheep breeds has not been fully studied and documented. Hence, it is fascinating to uncover the underlying genetic diversity between breeds and the population structure. This thesis aims to determine the genetic diversity and population structure of fourteen native Swedish sheep breeds, including some that have not been previously studied, using medium-density Illumina Ovine SNP50v3 Beadchip data. Firstly, I will investigate the inbreeding level in sheep breeds based on the F_{IS} and runs of homozygosity (ROH). In addition, genetic diversity measures such as average observed heterozygosity, minor allele frequency (MAF), polymorphic information content (PIC), and pairwise F_{ST} will be measured. Secondly, population structure, genetic relatedness, and genetic admixture (structure) will be examined using principal coordinate analysis (PCoA), Neighbour-joining trees (NJ tree) and Structure software version 2.3.4. Thirdly, I will study the genetic variants in MC1R gene and associated coat colour phenotype variability in Swedish sheep breeds.

1.3 Objectives

1.3.1 Main Objective

- To study the genetic diversity and population structure of native Swedish sheep breeds using medium-density Illumina Ovine SNP50v3 BeadChip data

1.3.2 Specific Objectives

- To determine the genetic diversity based on the level of inbreeding within breeds (F_{IS}), ROH-based genome-wide inbreeding, observed heterozygosity, MAF, PIC and pairwise F_{ST}
- To study the population structure of native Swedish sheep breeds
- To study the genetic variants in the coat colour gene, MC1R and the associated phenotypic variation in native Swedish sheep breeds

2. Literature Review

2.1 Sheep Domestication

Sheep *(Ovis aries*) and goats (*Capra hircus*) were among the earliest animal species to be domesticated approximately 11,000 years before present (BP) in the eastern Anatolia region and North-west Iran regions (Taberlet et al., 2011). After domestication, sheep spread around the globe, including expansion into Europe via two major routes, north and south, following the Neolithic agricultural revolution, which was the primary food production system in prehistoric Europe (Zeder, 2008). For instance, the secondary distribution throughout northern Europe involves a group of sheep breeds referred to as the North European short-tailed sheep. The Asian Mouflon (*Ovis Orientalis*) is regarded to be the ancestor of all domestic sheep breeds, and the European Mouflon (*Ovis orientalis musimon*) is in the clade of Asiatic Mouflon (Su et al., 2020).

Moreover, domesticated sheep today are a distinct species from their progenitor, the Asian Mouflon, which still exists in Southwest Asia, because of genetic, morphological and behavioural changes (Hiendleder et al., 2002). The great majority of modern breeds were influenced by a later phase of migration involving sheep with improved productivity qualities. Thus, the ability to distinguish modern breeds from genetically primitive sheep offers essential insights into sheep domestication history. Initially, these domesticated sheep were raised primarily for meat purposes, but by the fifth millennium BP, a specialization of secondary products such as wool became evident (Chessa et al., 2009). Today, sheep are bred and raised for various purposes, including wool, milk, meat production, conservation, or a combination of purposes (Rochus et al., 2020). They are also used for land management, as they graze vegetation.

2.2 North European Short-tailed Sheep Breeds and Swedish Sheep Breeds

The Swedish sheep breeds belong to a group of sheep breeds known as North European short-tailed sheep breeds. Study on genetic diversity estimation using microsatellite markers has reported that the North European short-tailed sheep breeds, including some Swedish sheep breeds, are unique, and they are highly contributing to domestic sheep diversity, although many of these breeds have low within-breed diversity and small effective population sizes (Tapio et al., 2005). The North European short-tailed sheep, native to a region from Russia to Iceland, are considered a primitive type that was spread to several countries by Norse Vikings

from the late 8th to mid-11th centuries (Dýrmundsson & Niżnikowski, 2010). The Soay sheep is the most primitive breed in the Northern short-tailed breeds and resembles the Asian Mouflon sheep (Ryder et al., 1974). Similar to their ancestors, the small, fluke-shaped tails, which vary slightly in length, are their common features. Usually, short-tailed sheep have 8 to 10 vertebrae in their tail, while long-tailed sheep have 16 to 18 vertebrae (Frandson, 1986). In addition to the short tail, this group of sheep breeds possesses several characteristics, such as a wide range of coat colour patterns, robustness, dual-coated wool, and prolificacy. Moreover, both horned and polled sheep are found in these breeds, varying in size and productive performance and capable of thriving in harsh environmental conditions (Ryder et al., 1974). The Finnsheep and Romanov are the two best-known breeds of this group outside of Northern Europe and have been exported to other countries around the world where their genetics have been used through crossbreeding programs (Thomas, 2010).

Many of these breeds, or their genetic material, should play an integral role in sustainable grassland-based production systems in the future. Furthermore, North European short-tailed sheep breeds are genetically more diverse than other sheep breeds, utilize a variety of pastures from steep mountain slopes to lowlands, and offer a range of high-quality products (Dýrmundsson & Niżnikowski, 2010). Hence, it is important to develop methods to preserve the genetic resources of North European short-tailed sheep breeds for future purposes while utilizing them sustainably today, both in pure and crossbreeding programs.

2.3 Genetic Diversity

The variety of alleles and genotypes present in a population is referred to as genetic diversity (Frankham et al., 2002). Conservation biology emerged in the seventies intending to preserve genes, species and ecosystems. Conservation of genetic diversity in farm animals is crucial as future needs are unpredictable. Furthermore, the conservation of genetic resources is vital for improved and sustainable production and to counter future challenges, including diseases and environmental stressors, increased demand for livestock products and food security (Groeneveld et al., 2010).

Natural selection plays a vital role in determining what animals are best adapted to distinct environments. Humans brought wild animals into captivity as part of the domestication process, altering their behaviour, morphology, and genetics to breed animals with specific characteristics (Mignon-Grasteau et al., 2005). Adapting livestock breeds to the local environment, either hot, cold or other climatic conditions, is an important factor for contemporary agriculture as it lowers environmental stressors on animals and leads to increased and sustainable production (Abied et al., 2020). In sheep, domestication, subsequent artificial

selection over thousands of years and natural selection have produced a vast spectrum of breeds to achieve more desirable and profitable traits and are capable of thriving in diverse environmental conditions (Brito et al., 2017).

How a genome can produce such a wonderful diversity of phenotypes is one of the fascinating problems in life sciences. Different genotypes have been generated during and after the domestication of sheep (Kalds et al., 2022) due to the main evolutionary forces of genetic drift, bottlenecks, mutation, adaptation to climate, selective breeding, isolation, diseases and parasites and available nutrition (Barker, 2001). Besides, specific genomic patterns have been artificially or naturally selected to produce diverse phenotypes. The evolution of phenotypically distinct sheep breeds has produced rich biodiversity (Alberto et al., 2018). This extensive phenotypic heterogeneity occurs as a consequence of genomic changes ranging from a single nucleotide to several thousands of nucleotides. In order to encourage selection for economically significant traits, it is, therefore, interesting and important to uncover and understand the genetic changes underlying the phenotypic diversity of sheep breeds (Kalds et al., 2022).

Over the past centuries, artificial selection has produced many well-defined breeds that are used for various tasks with varying levels of performance. Nevertheless, these high-yielding breeds have been replacing local breeds, and as a result, the unique genomic makeup of some local breeds is being lost (Groeneveld et al., 2010). Hence, it is essential to conserve farm animal genetics of low-producing breeds and traditional indigenous breeds for future breeding options while maintaining genetic diversity in a population or species (Bruford et al., 2003).

Livestock genetic resources management involves comprehensive knowledge of breed-specific traits, information on population structure and size, geographical distribution, within and between breed genetic diversity and production environments (Groeneveld et al., 2010). In this thesis, the genetic diversity and relatedness of fourteen Swedish sheep breeds: Åsen, Brännö, Finull (Swedish Finewool), Fjällnäs, Gammelrya, Gestrike, Gotland, Gute, Helsinge, Rya, Svärdsjö, Tabacktorp, Roslag, and Värmland, will be studied using methods to estimate inbreeding, observed heterozygosity, MAF, PIC, pairwise F_{ST} and population structure using medium density (50K) SNP chip data and sequencing of coat colour gene MC1R.

2.4 Sheep Coat Colour Genes

Coat colour in domestic animals is one of the most visible and remarkably varying traits. It is one of the economically most important traits in livestock husbandry, particularly for animals that are bred for skin and wool production such as sheep, which affects behaviour of animals and helps animals to survive in different environmental conditions (to cope with environmental and biological factors) and

avoid predators (Koseniuk et al., 2018). Moreover, coat colour trait has been widely used as a distinctive phenotype in the morphological selection for breed identification and characterization (Gebreselassie et al., 2021).

The coat colour diversity in a wide number of mammalian species is primarily determined by the availability and distribution of two basic melanin pigments produced in melanocytes: eumelanin and phaeomelanin. Depending on the presence or absence of these two melanocytes there are two categories of coat colour phenotypes: patterned and non-patterned (Hubbard et al., 2010). Eumelanin produces black and brown colour, whereas phaeomelanin produces red and yellow colour phenotypes, which are genetically regulated by the Extension (E) and Agouti (A) loci, respectively (Klungland et al., 1999). The E locus encodes for the melanocortin 1 receptor (MC1R) gene, a protein with seven transmembrane domains that is part of the G-protein coupled receptor found on the surface of the melanocyte membrane (Bultman et al., 1992). In contrast, the A locus encodes for the agouti signaling protein (ASIP), a tiny paracrine signaling molecule that interacts with the E locus's products (Robbins et al., 1993). Apart from ASIP and MC1R genes, other genes, such as melanocyte-inducing transcription factor (MITF), v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homologue (KIT), and tyrosinase-related protein-1 (TYRP1) are also implicated in the coat colour variation (Gebreselassie et al., 2021).

The complexity of the sheep coat color inheritance process was initially brought to light by Ryder et al. (1974). They found that multiple alleles within a locus corresponded to coat colour variation, and the white allele (A1 allele) was dominant over the grey allele (A2 allele). Additionally, they showed that the locus A is epistatic to locus B which is associated with pigment formation and black coat colour (B1), and is dominant over the brown (B2) allele (Ryder et al., 1974). Another study conducted by Renieri et al. (2008) to examine the inheritance of coat color pattern in sheep using lambs also demonstrated that fully white phenotype is dominant over brown and black and the black coat colour is dominant over brown phenotype.

2.5 Melanocortin 1 Receptor (MC1R) Gene

The MC1R gene is reported as a candidate gene that contributes significantly to melanogenesis, wool pigmentation and determining the black coat colour in mammalian species located on sheep chromosome 14 (Gebreselassie et al., 2021; Våge et al., 1999). Conventional genetic studies have shown two alleles (the dominant black, E^D and wild type, E^+) at the E locus affecting coat colour phenotypes in sheep (Yang et al., 2013). The dominant E^D allele results in the black coat pattern in a few breeds, and the wild-type E^+ allele results in the white coat pattern in most breeds in which the segregation of Agouti alleles. In contrast to

other species, the E^+ allele of the E locus in sheep has not yet been precisely identified (Fontanesi et al., 2010).

A study by Våge et al. (1999) examined the MC1R coat colour gene in a completely black or white Norwegian Dala sheep breed and identified two missense mutations, c.218 T>A (p.73Met>Lys) and c.361 G>A (p.121Asp>Asn) responsible for the dominant black (E^D) allele. Moreover, the co-segregation of these two mutations (dominant E^D allele) with black coat colour was confirmed in Darma, Karakul, Black Merino, and Black Corriedale sheep breeds (Royo et al., 2008; Våge et al., 2003). According to Mahmoud et al. (2017), there were five SNPs in the MC1R gene. Two of these variants were nonsynonymous (c.218 T>A and c.361 G>A), and the other three were synonymous (c.429 C>T, c.600 T>G, and c.735 C>T). These two independent nonsynonymous mutations were associated with black coat colour sheep breeds.

In addition, a study conducted to investigate MC1R and ASIP gene mutations causing black coat colour in five Swedish sheep breeds also confirmed the causal variants previously found in MC1R and ASIP genes. They showed that the presence of c.218 T>A and c.361 G>A variants in the coding region of the MC1R gene was associated with black coat colour in Swedish Finewool sheep (Rochus et al., 2019). Moreover, Rochus et al. (2019) found a novel mutation in MC1R (c.452 G>A) in Gute, Gotland and Värmland sheep and evidence of MC1R duplication in Gotland sheep.

A GWAS study was performed to determine the significant coat colour pigmentation gene using SNP50 BeadChip data from Rasa Aragonesa and Manchega sheep, and it was identified that SNP s26449 is strongly associated with coat colour pigmentation and is closest to the MC1R gene (Kijas et al., 2012). In addition, a GWAS performed with the Ovine SNP600K BeadChip to determine coat colour genes identified two significant SNPs (rs409651063 and rs408511664) in the MC1R genomic region were associated with coat colour development in Chinese Tan sheep (Gebreselassie et al., 2021). Consequently, the MC1R gene and its variants play a significant role in the development of coat colour in sheep.

2.6 Genetic Diversity Measures

Genetic variation is the foundation of genetic improvement of animals, and it includes diversity both within and between species. Genetic variation provides the basis for biological evolution and biological diversity. Ecosystems and species depend on genetic diversity for long-term survival, resilience or ability to adapt to diverse environmental conditions and evolutionary potential (Rochus et al., 2020). Evaluation of genetic diversity, relatedness and population structure offers crucial control metrics to prevent inbreeding depression, genetic erosion, and crossbreeding between locally adapted and exotic breeds since these occurrences

have detrimental effects and ultimately lead to breed extinction (Senczuk et al., 2020). Genetic diversity can be estimated using both pedigree and genomic-based approaches (Biscarini et al., 2020), and there are three main parameters to evaluate genetic variation in a population: level and rate of inbreeding, genetic drift and effective population size. These measures provide insight into the current state of population diversity and potential future risks (Falconer & MCKAY, 1996). In addition, population structure is assessed to determine the diversity within and between populations.

Inbreeding occurs due to the mating of related individuals, leading to an increased number of homozygotes and fewer heterozygotes in a population and is inevitable in finite populations. Nevertheless, the level of heterozygosity reflects genetic variability within and between species (Melo et al., 2017). In this thesis, two measures are used to assess the inbreeding of Swedish sheep breeds: observed heterozygosity and the genome-wide inbreeding coefficient based on the runs of homozygosity (ROH). The average expected and observed heterozygosity can vary due to evolutionary forces, such as mutation, genetic drift, migration and selection that cause random changes in allele frequency (Nei, 1978). In addition to evolutionary forces, small effective population size and management and breeding practices can cause a reduction in within-breed genetic diversity (Nadia Aubin-Horth et al., 2005).

Over the past decade, runs of homozygosity (ROH) analyses have become the state-of-the-art method for determining inbreeding in livestock populations (Meyermans et al., 2020a). When the identical haplotypes are inherited from both parents, resulting in a long stretch of the genotype being homozygous, this is known as long runs of homozygosity. Moreover, shorter haplotypes inherited from distant ancestors, whereas longer haplotypes from recent ones. An increased rate of inbreeding in a population gives rise to autozygosity: homozygosity where the two alleles are identical by descent (F. C. Ceballos et al., 2018). This method is predicated on the hypothesis that longer homozygous SNP sequences are more likely to have been inherited from a common ancestor than individual homozygous loci. The ROH-based inbreeding coefficient is calculated by dividing the summed length of the ROH by the total length of the genome. One method of finding the ROH lengths needed to calculate the inbreeding coefficient depends on the sliding window size that scans through the SNPs to find long stretches. In addition, the size of the scanning window, the minimum length in basepairs needed for ROH, the minimum number of homozygous loci needed for ROH, the maximum possible gap between two homozygous SNPs, and the minimum SNP density in the scanning region are important parameters to determine ROH-based inbreeding (Meyermans et al., 2020a). ROH analysis is common in livestock populations using --homozyg flag in PLINK v.1.9 tool (Peripolli et al., 2017). However, in this thesis ROH analysis was performed in R software using detectRUNS package.

Moreover, polymorphic information content (PIC) is an important indicator of quality of genetic markers which helps identifying regions in the genome that are most variable. The PIC is a measure of the amount of genetic variation present at a particular locus (Botstein et al., 1980). In the linkage studied, PIC is often used to assess the informativeness of a genetic marker (Guo & Elston, 1999). The usefulness of a marker for linkage analysis relies on how often marker is found to be polymorphic. This is measured using heterozygosity or PIC (Buchanan & Thue, 1998). Based on the number of markers and frequency of the alleles, the PIC illustrates how the marker can reveal population polymorphism. The PIC value ranges between 0 to 1, where a value of 1 indicates maximum diversity and 0 represents no diversity (Botstein et al., 1980). According to Dakin and Avise (2004), PIC > 0.5 is more informative, 0.25 <PIC <0.5 is moderately informative and PIC < 0.25 is slightly informative. Furthermore, genetic differentiation between each population can be measured using Wright's F_{ST} estimate (Pairwise F_{ST} / Fixation Index). The F_{ST} values range from 0 to 1, where higher values indicate a greater differentiation between the two populations, whereas lower values signify the lower divergence between two populations (Wright, 1984).

2.7 Population Structure

Exploring population structure provides a great deal of insight into the origin of populations and history. In order to visualize and explore genetic relatedness among breeds, principal component analysis (PCA), principal coordinate analysis (PCoA), Neighbor-joining network, calculation of F_{ST} and Bayesian model-based clustering algorithm can be used. Moreover, past admixture events of ancient populations can be visualized using software such as ADMIXTURE, STRUCTURE and R (Mastrangelo et al., 2014). In the current study, we used PCoA, Neighbor-joining tree and STRUCTURE software to examine population structure and admixture proportions.

The PCoA is a model-free method for examining ancestry and population structure, and it is one of the most popular methods in population genomics research as it is relatively simple and easy to interpret. PCoA analysis is computed using multi-dimensional scaling (MDS) plots, and it aims to determine the main axes of variation based on the genetic distance (main principal coordinate) in a dataset with no correlation between each axis. The first coordinate summarizes the larger proportion of variation (major axis variation), followed by the second coordinate and so on. The first two to three principal coordinates are commonly used to visualize genetic distance among populations (Mohammadi & Prasanna, 2003). Generally, PCoA is performed at the beginning of a population genetics study to visualize population stratifications and make follow-up decisions on further analyses (Ganteil et al., 2020).

The neighbour-joining (NJ) method is used to construct phylogenetic (evolutionary) and phenetic (trait-based similarity) trees from evolutionary distance data. The principle of maximum parsimony or minimum evolution is used in constructing phylogenetic trees. The standard NJ algorithm based on this principle is to determine all potential topologies (branching patterns) or a specific number of topologies that are possible to be close to the true tree and then select one that exhibits the least amount of overall evolutionary change as the final tree (Saitou & Nei, 1987).

Moreover, STRUCTURE is a free software package that can be used to investigate the population admixture of multi-locus genotype data. It can be used to deduce the presence of diverse populations, study hybrid zones, assign individuals to populations, estimate population allele frequencies, and identify admixed individuals and migrants. The structure software analysis can be applied to most of the commonly used genetic markers: microsatellites, SNPs, amplified fragment length polymorphism (AFLP) and restriction fragment length polymorphism (RFLP) (Pritchard et al., 2000).

3. Materials and Methods

In this section, firstly, the data used in the thesis will be presented and then the methods used to perform different analyses will be described.

3.1 Data Source

All the data used in this thesis were obtained from my supervisor Dr. Anna Maria Johansson, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences (SLU), Sweden. Sample collection, DNA extraction, sequencing and genotyping were performed by a previous researcher at SLU.

3.2 Sheep Sample

Biological samples: blood, tissue and nasal swabs had been collected from 207 individuals belonging to 14 different Swedish short-tailed sheep breeds. These 14 breeds represented different population sizes and different geographical origins within Sweden. Breeds (populations), the number of individuals in each breed, and their geographical origins are shown in Table 1.

Table 1 Information on the number of individuals from each breed and the geographical origin

Breeds	Number of Individuals	Geographical Origin
Åsen	15	Mainland Sweden
Brännö	17	Mainland Sweden
Finull	15	Mainland Sweden
Fjällnäs	2	Mainland Sweden
Gammelrya	3	Mainland Sweden
Gestrike	10	Mainland Sweden
Gotland	54	Gotland Island
Gute	1	Gotland Island
Helsinge	15	Mainland Sweden
Roslag	22	Roslagen Archipelage
Rya	16	Mainland Sweden
Svärdsjö	15	Mainland Sweden
Tabacktorp	7	Mainland Sweden
Värmland	15	Mainland Sweden
	Total: 207	

3.3 Genotyping

Sequenced genomic data of Swedish sheep breeds were genotyped with Illumina Ovine SNP50v3 BeadChip array. This SNP chip includes 26 autosomes.

3.4 Data Merging, Removal of Related Individuals, and SNP Quality Control

Data merging, removal of related individuals, and SNP quality control on raw data were performed using PLINK v.1.9 program (Purcell et al., 2007). Several quality control measures were applied, as shown in Table 2, and are described below.

Initially, there were two datasets: one consisting of 185 individuals with 64,734 SNP markers, and the other one consisting of 22 individuals with 49,396 SNP markers. Then, two datasets were merged based on the common variants, and the merged dataset consisted of 207 individuals belonging to 14 native Swedish sheep breeds with 49,363 common variants.

Removal of related individuals was done with the --genome flag in the plink tool, which invokes identical by descent (IBD)/ identical by state (IBS) computation. Subsequently, five pairs of individuals were identified as related to each other based on the PI_HAT value: the estimated proportion of IBDs shared. According to Purcell et al. (2007), five related individuals were removed as PI_HAT values were close to 1. Accordingly, as stated below, 202 samples with 49,363 variants were subjected to SNP quality control and LD-pruning.

Next step was performed to remove animals with too much missing genotype data (low call rates), and individuals with a genotype call rate <97% were excluded with the --mind option in plink. SNPs filtering was done if the call rate was <97% based on the missing genotype rate and SNP call rate <95% based on minor allele frequency (MAF) with the --geno option and the --maf option in plink, respectively. The individual and SNP missing call rate thresholds (mind and geno) were calculated in R based on the plink output derived with --missing flag (Figures 1 and 2). Moreover, SNPs were discarded if the Hardy Weinberg Equilibrium exact test (HWE) p-value was less than 0.000001 (--hwe). With this, 40241 variants and 191 samples were passed the filters.

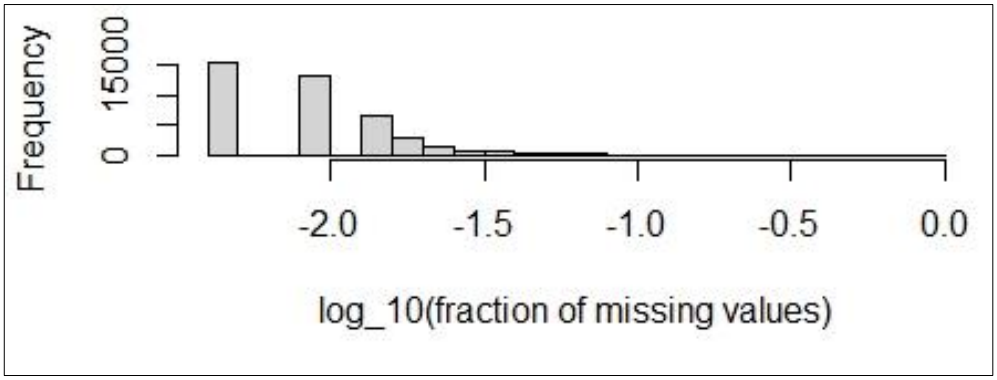


Figure 1 Distribution of missing loci in individuals. Log10(-1.5) was used for the threshold determination. Accordingly, we excluded SNPs with a missing rate of 0.03 (3%) or higher (geno).

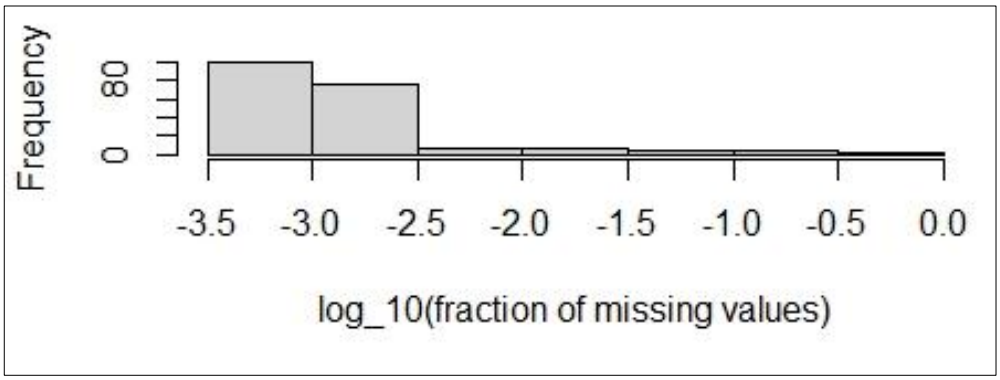


Figure 2 Distribution of missing values in individuals. Log10(-1.5) was used for the threshold determination. Accordingly, we excluded individuals with a missing rate of 0.03 (3%) or higher (mind).

After this quality control (QC1), linkage disequilibrium-based SNP pruning (LD-pruning) was done with the --indep-pairwise flag in plink v1.9. Groups of SNPs mutually in LD will have approximately the same association strength with the phenotype. Hence, SNP pruning is performed to choose representative SNP for each of these groups. There are three parameters to be considered in LD pruning: window size, window step size and LD (r^2) threshold (Purcell et al., 2007). In this experiment, we set the window size to 50 SNPs and the number of SNPs to shift the window at each step, window step size to 5 SNPs. The LD threshold was 0.2, and the SNP pairs were removed if the LD (r^2) was greater than 0.2. Following QC1, the dataset was used to calculate the genetic diversity measures: observed heterozygosity, inbreeding coefficient (F_{IS}), MAF, PIC, and pairwise F_{ST}. Following both QC1 and LD-pruning, data was used to assess the population structure using principal coordinate analysis, neighbor-joining trees and admixture analysis using Structure software version 2.3.4.

The second quality control was performed to calculate runs of homozygosity (ROH). According to Meyermans et al. (2020a), MAF-based SNP filtering and LD-

pruning before ROH analysis affect ROH detection and genome coverage. Further, it showed that employing either LD-pruning or MAF filtering could hamper ROHs detection by drastically lowering the number of SNPs included in the study. Hence, none of these SNP filtering methods were used in the second quality control.

Table 2 SNP quality control parameters

	geno	**Mind**	**maf**	**HWE**	**Indep-pairwise**	**No. SNPs**	**No. Individuals**
After removing related individuals						49363	202
QC1	0.03	0.03	0.05	0.000001		40241	191
LD-pruning for QC1 data					50 5 0.2	23292	191
QC2	0.03	0.03	-	0.000001	-	43518	191

The first five columns refer to the commands in PLINK. The last column states the number of SNPs that remained after the quality control.

geno: maximum missingness per SNP

mind: maximum missingness per individual

maf: number of SNPs removed due to minor allele frequency threshold.

hwe: number of SNPs removed due to HWE exact test p-value threshold.

indep-pairwise: prune for linkage disequilibrium, there are three parameters: window size in variant counts or kilobase units, a variant count of shifting the window at the end of each step, and pairwise LD (r^2) threshold, respectively.

3.5 Genetic Diversity Analysis

Before computing the genetic diversity measures: observed heterozygosity, F_{IS}, MAF, PIC, and pairwise F_{ST}, outliers were removed from the dataset. In this thesis, outlier individuals determination was done based on the PCoA. If an individual was not clustered within its own group and highly varied from the other individuals within the population, that individual was removed.

3.5.1 Observed Heterozygosity

The observed heterozygosity and inbreeding coefficient (F_{IS}) were computed for each breed separately with the --het flag in PLINK v.1.9 program (Purcell et al., 2007). After removing outliers, 190 samples and 40241 SNPs were subjected to the analyses. The observed heterozygosity was calculated as the difference between the

non-missing genotypes counts and the number of observed homozygous genotypes (Equation 1).

$$O(HET) = N(NM) - O(HOM) \quad [1]$$

Where, O(HET) is the observed heterozygous count: the number of observed heterozygous genotypes found in an individual. O(HOM) is the observed homozygous count: the number of homozygous genotypes found in an individual. N(NM) is the number of non-missing counts: the total number of SNPs in an individual after quality control (the summation of observed homozygous and observed heterozygous genotype counts). The average O(HET) and O(HOM) were presented as proportions in the results section.

3.5.2 Inbreeding Coefficient (F_{IS})

Wright's F statistic (F_{IS}) is a measure of the degree of inbreeding within a population. The inbreeding coefficient (F) assessed by plink v.1.9 corresponds to F_{IS} estimate. Wright's F statistic or F_{IS} was introduced to quantify heterozygote deficiency or excess homozygotes (Zhivotovsky, 2015). The inbreeding coefficient (F_{IS}) was estimated based on the observed: O(HET) and expected: E(HET) heterozygous genotype proportions assuming the population was under HWE with the following equation (Equation 2) (Wright, 1984).

$$F_{IS} = [\mathrm{E(HET)} - \mathrm{O(HET)}] / \mathrm{E(HET)} \quad [2]$$

Where, E(HET) is an individual's expected heterozygosity and O(HET) is an individual's observed heterozygosity under HWE. Observed and expected heterozygosity proportions were used in this calculation. A value of $F_{IS} = 0$ indicates no inbreeding, while a value of $F_{IS} > 0$ indicates the inbreeding within the population. A value of $F_{IS} < 0$ denotes the excess heterozygosity within population.

3.5.3 Minor Allele Frequency (MAF)

The rare allele (the second most common allele) frequency in a specific population is known as the MAF. MAF is usually used in population genetics studies as it provides information to distinguish between rare and common variants in the population (Hernandez et al., 2019). In this thesis, in each breed, MAFs of each polymorphic sites were computed using the --freq flag in plink v.1.9. The average MAF was calculated separately for each population.

3.5.4 Polymorphic Information Content (PIC)

According to (Botstein et al., 1980), the PIC is calculated from MAF (maf) using the equation mentioned below (Equation 3). PIC is calculated using p and $q = 1 - p$ allelic frequencies (each SNP is biallelic). Thus, according to HWE, the probability of an individual being heterozygous is 2pq. One of the allelic frequencies (p and q) replaced by the maf, then q would be 1-maf. In order to calculate PIC, first heterozygosity was calculated ($1 - [maf^2 + (1 - maf^2)]$) and then identical heterozygous trios ($2maf^2 (1 - maf)^2$) were subtracted from it (Botstein et al., 1980). This thesis, MAF and PIC values were calculated for QC1 data using the --freq flag in plink v.1.9.

$$PIC = 1 - [maf^2 + (1 - maf)^2] - [2maf^2 * (1 - maf)^2] \quad [3]$$

Accordingly, average observed heterozygosity, average observed homozygosity, average F_{IS}, MAF and PIC values were calculated in each population.

3.5.5 Pairwise F_{ST} (Wright's F_{ST} Estimate / Fixation Index)

Pairwise F_{ST} values between each breed were calculated using the StAMPP package in R software. The stamppConvert() function was used to convert genomic data into allele frequencies and followed by stamppFst() function was used to compute pairwise F_{ST} values (Pembleton et al., 2013; Wright, 1984).

3.5.6 Runs of Homozygosity based Inbreeding Coefficient (FROH)

ROH was computed using the detectRUNS package in R software with the function slidingRUNS (), a sliding-window-based runs detection (Biscarini et al., 2018). The parameters used for the analysis are shown in Table 3, and the values of the parameters were based on Meyermans et al. (2020a) study with slight modifications. There are two types of parameters: sliding-window parameters: windowSize, threshold, minSNP (minimum number of homozygous/heterozygous SNPs in the window), maxMissWindow (maximum number of missing genotypes), maxOppWindow (maximum number of SNPs with opposite genotypes) and run-related parameters: minLengthBps (minimum length of run (base pairs; bps)), maxGap (maximum gap between consecutive homozygous SNPs (bps)), minDensity (number of SNPs in every kbps), maxMissRun (maximum number of missing genotypes in the run), maxOppRun (maximum number of opposite genotypes in the run) (Biscarini et al., 2018). The detectRUNS uses 1SNP/10kb as their scale compared to plink, which uses the distance between SNPs.

Table 3 List of parameters for the ROH analysis with the detectRUNS package in R software

Parameter	**Parameter Name in detectRUNS**	**Value**
Scanning window size	windowSize	15 (default)
Scanning window threshold	threshold	0.05
Minimal number of SNPs	minSNP	15 (default)
Minimal density	minDensity (SNP/kbps)	1/250
Maximal gap	maxGap	100000
Minimum length	minLengthBps	1000
Other		Default values

Increasing the scanning window size leads to reduced detection of ROH and estimation of FROH, and particularly, short ROHs are no longer detected. Similarly, an increased threshold reduces FROH detection (Meyermans et al., 2020a). Moreover, that research article mentioned the importance of controlling false positive ROH in the analysis and considering how long ROHs are expected by chance. In order to compute the minimum number of SNPs (L) in a ROH, equation 4 can be used as described in (Purfield et al., 2012).

$$L = log_e (alpha/n_s n_i)/log_e(1 - het) \quad [4]$$

$L = \log_e (0.05/43518*191)/\log_e(1-0.28) = 57.62 = 58$

Where, alpha (α) is the percentatge of false-positive ROH, n_s is the number of genotyped SNPs, n_i is the number of genotyped individuals, and het is the average observed heterozygosity across all SNPs.

Accordingly, this study's minimum number of SNPs was 58 (rounded 57.62 into 0 decimals). However, with this value, ROH detection was unsuccessful and only detected ROH on four chromosomes (chromosomes 3, 4, 5, and 9). A previous study on genetic diversity of five Swedish sheep breeds using high density SNP array by Rochus et al. (2020) had been used 20 as minimum SNPs count in the ROH analysis in plink. However, 20 minimum SNPs count was not suffiecient to have a complete genome coverage in this thesis (only 24 chromosomes were covered). Thus, the default value, 15 SNPs, was used as it had a complete genome coverage. Furthermore, Meyermans et al. (2020a) recommended setting the scan window size equal to the minimum number of SNPs.

The minimum ROH length, either by minimal kb length or minimal number of SNPs, has been studied by Purfield et al. (2012) and Ferenčaković et al. (2013) and they showed that medium density SNP array is suitable for detecting ROHs longer

than 5Mb, and the minimal ROH length should be adjusted to the SNP density, respectively. The study by Meyermans et al. (2020a) showed that a low minimal density setting might lead to incomplete genome coverage and the plink default minimal density value, 50kb/SNP, is often not suitable. The minimal density value of the current thesis was determined based on Meyermans et al. (2020b). Nevertheless, Meyermans et al. (2020a) showed a substantial variation in parameter settings considering scanning window size, minimal density and maximal gap. Consequently, results will be biased and hinder comparing results with other studies.

The identified ROH was then used to calculate the ROH-based inbreeding coefficient (FROH). Finally, total length of the ROH segments is divided by the total length of the genome to calculate the inbreeding coefficient based on ROH (Equation 5) (F. C. Ceballos et al., 2018).

$$FROH = sum\ of\ ROH/genome\ length \quad [5]$$

3.6 Genetic Population Structure

3.6.1 Principal Coordinate Analysis (PCoA)

After QC1, LD-pruned data was used for PCoA analysis. The PCoA was performed with the --distance-matrix flag in plink to assess the genetic structure of the fourteen sheep breeds (Purcell et al., 2007), which created the distance matrix as an input to the PCoA, and the cmdscale() (classical multi-dimensional scaling) function in R software. The ggplot() function under the tidyverse package in R was used to visualize data in a PCo-biplot. For the plots shown in the results section, the first two principal coordinates and the first and third principal coordinates were demonstrated.

3.6.2 The Neighbour-joining (NJ) Method

Two NJ trees (a and b) were constructed to illustrate the population structure. The NJ tree (a) was constructed based on the pairwise genetic distances between individuals using the dist() function, and followed by the NJ tree estimation was performed using the nj() fuction in the “Ape” package in the R software. The NJ tree (b) was constructed based on the pairwise Nei’s genetic distances between populations computed using the nei.dist() function in the “poppr” package in R software and the NJ tree estimation was perofrmed as the explanation for NJ tree (a) (Nei, 1978; Saitou & Nei, 1987). A unrooted tree (NJ (b)) was constructed using the plot.phylo() function in R. The NJ tree (b) was constructed for 12 breeds: Fjällnäs and Gute sheep were removed due to smaller sample sizes.

3.6.3 Population Admixture Analysis using Structure Software v.2.3.4

The genetic structure of the populations and their admixture proportions were assessed and visualized in plots using the model-based software Structure version 2.3.4, which includes a Bayesian method (Pritchard et al., 2000). LD-pruned data were used for the admixture analysis (Table 2). Model-based clustering was done using admixture model with correlated allele frequencies, 10,000 MCMC (Markov chain Monte Carlo) repetitions, and 10,000 burn-in periods. Five independent runs were performed for each K (number of different clusters) value which was ranged from 2 to 14. The remaining parameter settings were left to default values. In the analysis, the breed names were used as sampling location indicators. Parameter settings were based on the Pritchard et al. (2000). STRUCTURE HARVESTER, a web-based program, was used to collate results derived from Structure 2.3.4 as it is offered the very fast and easiest way to determine and visualize likelihood values across multiple K values. Furthermore, it detects the number of genetic groups (K) among populations that best fit the data based on the Evanno method (Earl & vonHoldt, 2012).

3.7 DNA Sequencing, Multiple Sequencing Alignment and Determination of Variants in MC1R that Responsible for Black Coat Colour

Genomic DNA samples from 24 individuals belonging to five native Swedish sheep breeds were used for the sequencing analysis. The genomic DNA samples were kept at -20°C until analyses started. In addition, three sequences from three breeds were taken from Rochus et al. (2019). Breeds and the number of samples from each breed were shown in Table 4.

Table 4 Information on breeds and the number of samples from each breed

Breed	**No. of Individuals**
Gammelrya	2
Rya	11
Finull	3
Gestrike	2
Svärdsjö	6
Previous Study data (Rochus et al., 2019)	
Gotland	1
Gute	1
Finull	1
	Total: 27

3.7.1 DNA Sequencing

The BigDye® Direct Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) was used for sequencing Exon 1 region of MC1R gene. The PCR primers had been selected based on the literature (Fontanesi et al., 2010). In order to use the BigDye® Direct kit, MC13 universal primer sequences (forward primer sequence: 5'-TGTAAAACGACGGCCAGT-3') were added. Subsequently, genomic DNA amplification was done by performing PCR and cycle sequencing on Applied Biosystems® Gene Amp® PCR System 9700 (Applied Biosystems). The BigDye XTerminator® Purification Kit (Applied Biosystems) was used to purify sequencing products. Finally, DNA sequencing was performed using capillary electrophoresis on the Applied Biosystems® 3500xL Dx Genetic Analyzer (Life Technologies, Foster City, CA, USA).

3.7.2 Sequencing Analysis

DNA sequence alignments, editing and visual inspections were performed using the BioEdit Sequence Alignment Editor tool.

4. Results

In this section, the results of the thesis will be presented. First, the results of the population structure will be described, and then, the genetic diversity measures: the average observed heterozygosity, average observed homozygosity, average inbreeding coefficient (F_{IS}), FROH, average MAF, average PIC, and pairwise F_{ST} values of Swedish sheep breeds will be presented. Lastly, genetic variants in sheep coat colour gene, MC1R, and associated phenotypic variability will be described.

4.1 Population Structure

4.1.1 Principal Coordinate Analysis (PCoA)

Figure 3 illustrates the population structure of 191 individuals belonging to 14 different short-tailed Swedish sheep breeds based on the principal coordinates analysis (PCoA). In this thesis, the first two (Figure 3: left) and first and third (Figure 3: right) principal coordinates analyses are shown. Based on PCoA, the first three principal coordinates (PCs) explain 21.59% of the interbreed variation: PC1, PC2 and PC3 explain 9.15%, 6.51% and 5.93% variation separately. According to the first three PC analyses, four breeds: Gotland, Roslag, Svärdsjö, and Tabacktorp, were distinguishable and unique from each other, and individuals within each breed clustered together. Moreover, Fjällnäs, Gammelrya, Gestrike, Rya, Finull, and Värmland (except one individual) sheep were clustered closer to each other. As per PC1 explanation, Roslag sheep was highly differentiated from all the other breeds. According to PC2, Brännö and Gute and according to PC3, Helsinge, Åsen, and Gute sheep breeds were also distinguishable from other breeds. Although the Brännö sheep were distinguishable according to PC2, it was not separable in PC3. Furthermore, PC2 showed a relatively less genetic variance between Gotland and Brännö sheep, whereas PC3 showed a relatively less genetic variation between Roslag and Tabacktorp sheep. One Värmland sheep (individual ID; OAR898) is genetically distinct from other Värmland sheep, and it appears to be genetically close to Roslag sheep, according to the first three PCs.

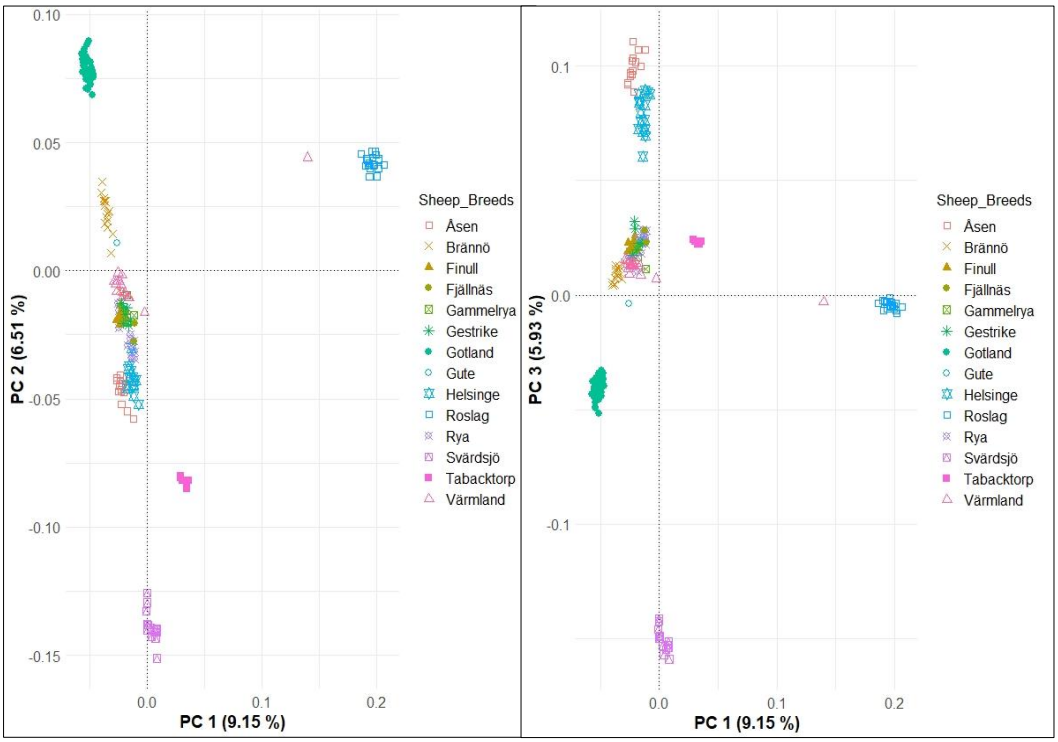


Figure 3 Genetic relatedness of fourteen Swedish sheep breeds using principal coordinate analyses (PCoA) (multidimensional scaling plots). The first two coordinates (PC1 and PC2) (left) and the third coordinate (PC3) (right) explained 9.15%, 6.51%, and 5.93%, respectively, of the total variation.

4.1.2 Neighbour-joining Trees

Apart from PCoA, the population structure was studied using two neighbour-joining (NJ) phylogenetic trees (a and b) shown in Figure 4. The NJ trees indicate the genetic relationship within (a) and between breeds (b). The NJ tree (a) is in agreement with the PCoA analysis. Värmland OAR 898 individual was clustered close to Roslag sheep. Moreover, Gammelrya (Old Rya) OAR 1004 individual was clustered together with Rya sheep. According to the farmer's records, the OAR1004 individual is crossbred from Rya and Gammelrya sheep. Gammelrya is still not an officially recognized breed. Samples had been collected from a farm in northern Sweden, where only a small group of ewes remained on the farm with phenotypes similar to Rya sheep. Hence, a Rya ram was used for breeding purposes. In the current study, we wanted to study the genetic background of these Gammelrya ewes to see if they were genetically similar to Rya sheep. All the other individuals were clustered within their own breed, as we saw in PCoA. Nevertheless, we can see several subgroups within a breed that may be due to individuals from different families.

The NJ tree (b) demonstrates the population structure and interbreed genetic relationship of 12 native Swedish sheep breeds. We removed Fjällnäs and Gute sheep from this analysis due to a low number of individuals within each breed. Consistent with this phylogenetic tree, 12 breeds were distinct and had different branch lengths. The Tabacktorp sheep showed the longest branch length compared to other breeds indicating high amounts of genetic drift within the population. Similarly, in PCoA (PC3) (Figure 3), the Roslag and Tabacktorp sheep showed a close relationship being on the same branch, although they had different branch lengths. Roslag sheep originated from the Roslag archipelago, whereas Tabacktorp sheep originated from Tabacktorp in Västansjö outside Ekshärad in Värmland. Therefore, this genetic relationship cannot be explained by geographical proximity. Furthermore, Gotland and Brännö sheep showed a close relationship being on the same branch, and both breeds had relatively smaller branch lengths indicating less genetic drift within populations. These results were in line with PCoA (PC2) (Figure 3). In addition, Rya and Gammelrya sheep were also on the same branch, demonstrating a common lineage between them. Nonetheless, Gemmelrya had a longer branch length denoting greater genetic changes than Rya sheep. The Gotland, Värmland, Finull, and Rya sheep showed relatively short branch lengths, while Roslag, Gestrike, Helsinge, Svärdsjö, and Åsen breeds showed relatively longer branch lengths, indicating higher genetic drift within populations.

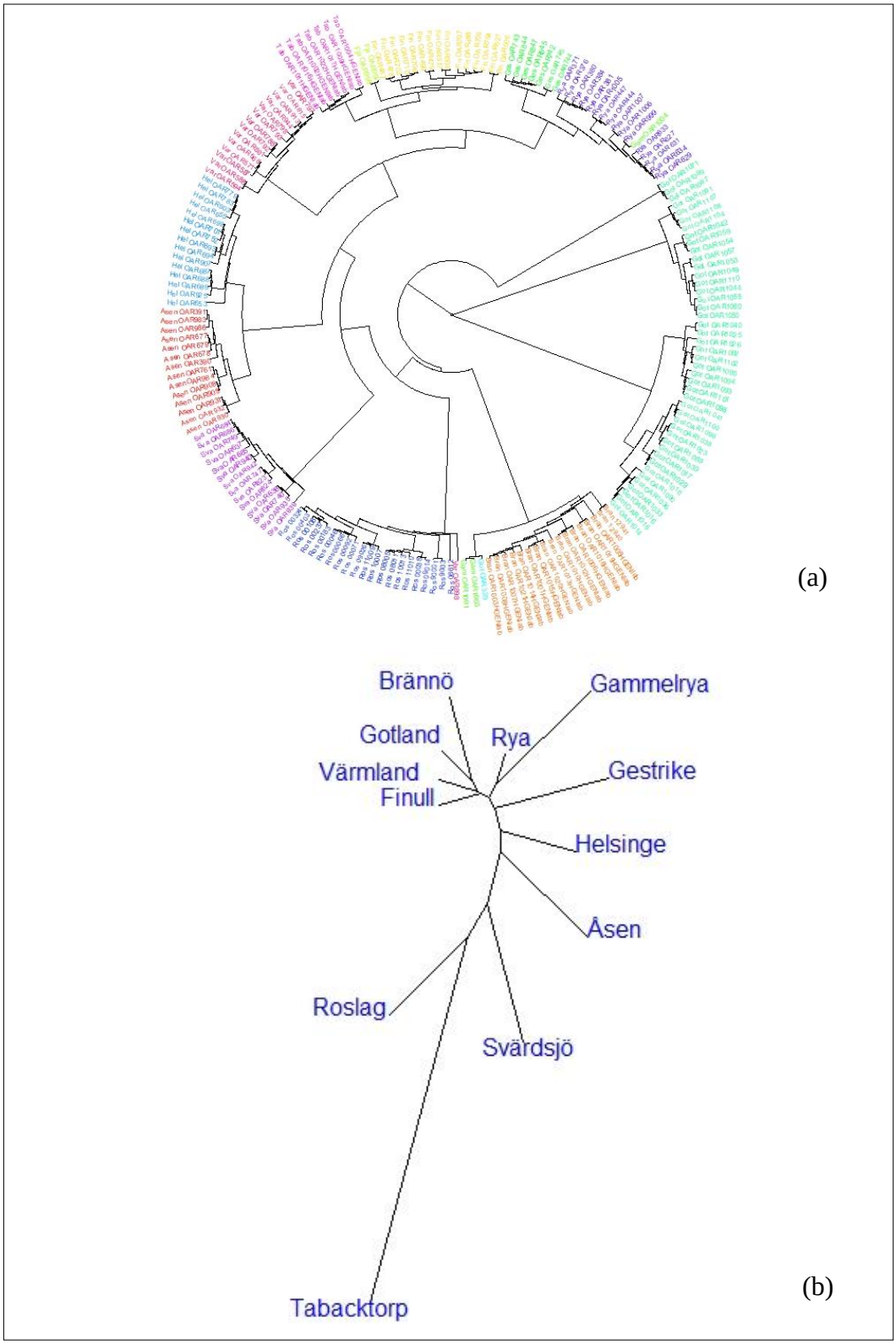


Figure 4 Population relationship and structure of native Swedish sheep breeds using the Neighbour-joining Trees based on the genetic distance: (a) Fan type NJ tree: within-breed genetic relatedness of 191 sheep from 14 Swedish breeds. (b) Unrooted NJ tree (b): genetic relatedness between 12 Swedish sheep breeds.

4.1.3 Population Admixture Analysis

To further investigate the population structure, genetic admixture was analyzed based on the model-based clustering analysis using Structure v.2.3.4, with K values (assumed number of genetic groups) ranging from 2 to 14. The STRUCTURE HARVESTER, a web-based program, was used to collate results derived from Structure 2.3.4. to determine and visualize likelihood values across multiple K values. According to the STRUCTURE HARVESTER output, the number of genetic groups among populations that best fit the data was twelve (K=12) (Appendix 1). The Figure 5 illustrates the genome-wide admixture proportions at K=12 among 191 sheep of 14 breeds. Genome-wide admixture proportions with K values from 2 to 14 were shown in different plots found in Appendix 1: plot (a) showed the admixture proportions among populations, whereas plot (b) demonstrated the admixture proportions of each individual within the population.

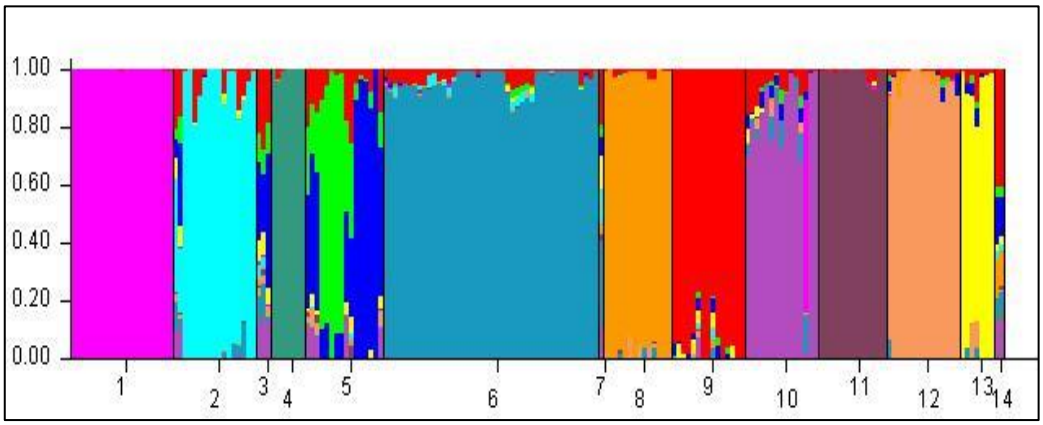


Figure 5 Genome-wide admixture proportions at K=12 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 12 assumed genetic populations. The numbers denoted in the x-axis show different populations, and here 14 Swedish sheep breeds used as different populations: 1 Roslag; 2 Brännö; 3 Gammelrya; 4 Tabacktorp; 5 Rya; 6 Gotland; 7 Gute; 8 Åsen; 9 Finull; 10 Värmland; 11 Svärdsjö; 12 Helsinge; 13 Gestrike; 14 Fjällnäs.

Admixture analysis showed that one genomic cluster was predominant in Roslag, Brännö (except 12940 and 12941; Appendix 1), Tabacktorp, Gotland, Åsen, Finull, Värmland (except OAR 898), Svärdsjö, Helsinge and Gestrike breeds and are distinct from other breeds indicating that they have less shared population ancestries. The Brännö 12940 and 12941 individuals were different from other Brännö sheep, and they appeared to be crossbred and genetically more admixed than others within the breed (Appendix1: Figure 19). Furthermore, the genetic clustering patterns of Gammelrya and Rya appeared to be the same, with slight differences. The genetic structure of Gammelrya OAR 1004 was closer to the Rya sheep and slightly different from the other two Gammelrya sheep, which had a small proportion of shared ancestry with Brännö sheep that is not present in the OAR 1004 individual (Appendix 1: Figure 19). As shown in PCoA (Figure 3), NJ

tree (a) (Figure 4), Värmland OAR 898 individual was different from other Värmland sheep, and the genetic structure of it was more similar to the Roslag genetic structure (Appendix 1: Figure 19).

In addition, Gute sheep appeared to be highly admixed with Gotland sheep, which is in agreement with the known shared ancestry between Gotland and Gute sheep breeds. The Fjällnäs sheep were also a highly genetically admixed breed with a low admixture proportion with Gotland sheep and a high proportion with Finull sheep. Moreover, Gotland and Brännö breeds showed a genomic admixture with Finull, Rya, and Gestrike sheep. The Gestrike, Värmland, Gammelrya, and Rya sheep showed a shared ancestry with Finull sheep (Appendix 1: Figure 19). No genetically shared ancestry with other breeds was observed in the Roslag sheep, and it appeared to be a genetically isolated breed. Nevertheless, Rya sheep showed the highest genetic admixture (structuring) within the population compared to other breeds and showed two different genetic groups (indicated in green and dark blue colours).

4.2 Genetic Diversity

Genetic diversity within each of the 14 breeds was assessed by estimating the proportion of average observed homozygosity (O(HOM)), average observed heterozygosity (O(HET)), inbreeding coefficient (F_{IS}), average polymorphic information content (SNPs), average MAF, and pairwise F_{ST} values. A summary of the genetic diversity estimates of 14 Swedish sheep breeds and pairwise F_{ST} estimates of twelve Swedish sheep breeds are shown in Tables 5 and 6, respectively.

Table 5 Summary of Genetic Diversity estimates of fourteen native Swedish sheep breeds

Breed	Sample Size	O(HOM)	O(HET)	F_{IS}	MAF	PIC
Åsen	14	0.64	0.36	0.21	0.20	0.22
Brännö	17	0.62	0.38	0.20	0.24	0.25
Finull	15	0.61	0.39	0.18	0.28	0.28
Fjällnäs	2	0.43	0.57	-0.18	0.17	0.24
Gammelrya	3	0.53	0.47	0.06	0.23	0.23
Gestrike	7	0.60	0.40	0.17	0.23	0.24
Gotland	44	0.64	0.36	0.22	0.25	0.27
Gute	1	0.00	1.00	-Inf	0.14	0.22
Helsinge	15	0.63	0.37	0.20	0.23	0.24
Roslag	21	0.67	0.33	0.25	0.16	0.17
Rya	16	0.62	0.38	0.20	0.27	0.28
Svärdsjö	14	0.65	0.35	0.23	0.17	0.19
Tabacktorp	7	0.61	0.39	0.17	0.09	0.10
Värmland	14	0.64	0.36	0.22	0.27	0.28

The average estimate of parameters are shown: O(HOM), Observed homozygosity; O(HET), Observed heterozygosity; F_{IS}, Inbreeding coefficient; MAF, Minor allele frequency; PIC, Polymorphic information content.

The estimated average observed homozygosity (O(HOM)) of 14 breeds varied from 0 to 0.67, and the highest was observed in Roslag sheep. The average minor allele frequency (MAF) varied from 0.09 in Tabacktorp to 0.26 in Finull breeds. The proportion of average polymorphic information content (PIC/ polymorphic SNPs) ranged from 0.09 in Tabacktorp sheep to 0.28 in Finull, Rya, and Värmland sheep. The PIC of each SNP of a particular breed was less than 0.5, and the values ranged around 0.25 or below, indicating the presence of moderate to slightly informative polymorphic sites. This may be related to low MAFs, linkage disequilibrium as we did not perform LD-based SNP filtering before calculations, non-random mating, and demographic factors such as migration and admixture.

Table 6 Pairwise F_{ST} (Wright's Fixation index) estimates of twelve native Swedish sheep breeds

	Ros	**Bran**	**Gam**	**Tab**	**Rya**	**Got**	**As**	**Fin**	**Var**	**Sva**	**Hel**	**Ges**
Ros												
Bran	0.32											
Gam	0.33	0.17										
Tab	0.52	0.39	0.46									
Rya	0.26	0.15	0.07	0.33								
Got	0.27	0.15	0.14	0.33	0.12							
As	0.37	0.25	0.24	0.45	0.20	0.20						
Fin	0.26	0.14	0.09	0.32	0.08	0.12	0.19					
Var	0.27	0.14	0.09	0.33	0.08	0.11	0.19	0.08				
Sva	0.41	0.29	0.29	0.50	0.23	0.24	0.34	0.23	0.24			
Hel	0.32	0.20	0.17	0.40	0.14	0.17	0.22	0.14	0.14	0.30		
Ges	0.34	0.20	0.16	0.43	0.13	0.17	0.26	0.12	0.13	0.31	0.20	

Pairwise F_{ST} between breeds: Ros, Roslag; Bran, Brännö; Gam, Gammelrya; Tab, Tabacktorp; Rya, Rya; Got, Gotland; As, Åsen; Fin, Finull; Var, Värmland; Sva, Svärdsjö; Hel, Helsinge; Ges, Gestrike

The genetic differentiation between 12 native Swedish sheep breeds was assessed using the pairwise F_{ST} value (Table 6). The higher value signifies a greater genetic divergence between the two breeds. The F_{ST} values ranged from 0.07 between Rya and Gammelrya sheep to 0.52 between Tabacktorp and Roslag sheep. Low F_{ST} value between Rya and Gammelrya breeds attributed to the high level of gene flow and low levels of genetic differentiation between the two populations, as we observed in genomic admixture analysis (Figure 5). According to PCoA (PC1) (Figure 3), and admixture analysis, Tabacktorp and Roslag are two distinct sheep breeds, although there was a common lineage between them according to NJ phylogenetic tree (b) (Figure 4); thus, the highest F_{ST} value between these two populations indicated the low level of gene flow and high level of genetic distance between Roslag and Tabackorp sheep. Moreover, the high F_{ST} value (0.5) between Svärdsjö and Tabacktorp sheep breeds also showed a relatively high genetic differentiation between the two populations while having less genetic admixture, as shown in the admixture analysis and PCoA.

4.2.1 Observed Heterozygosity

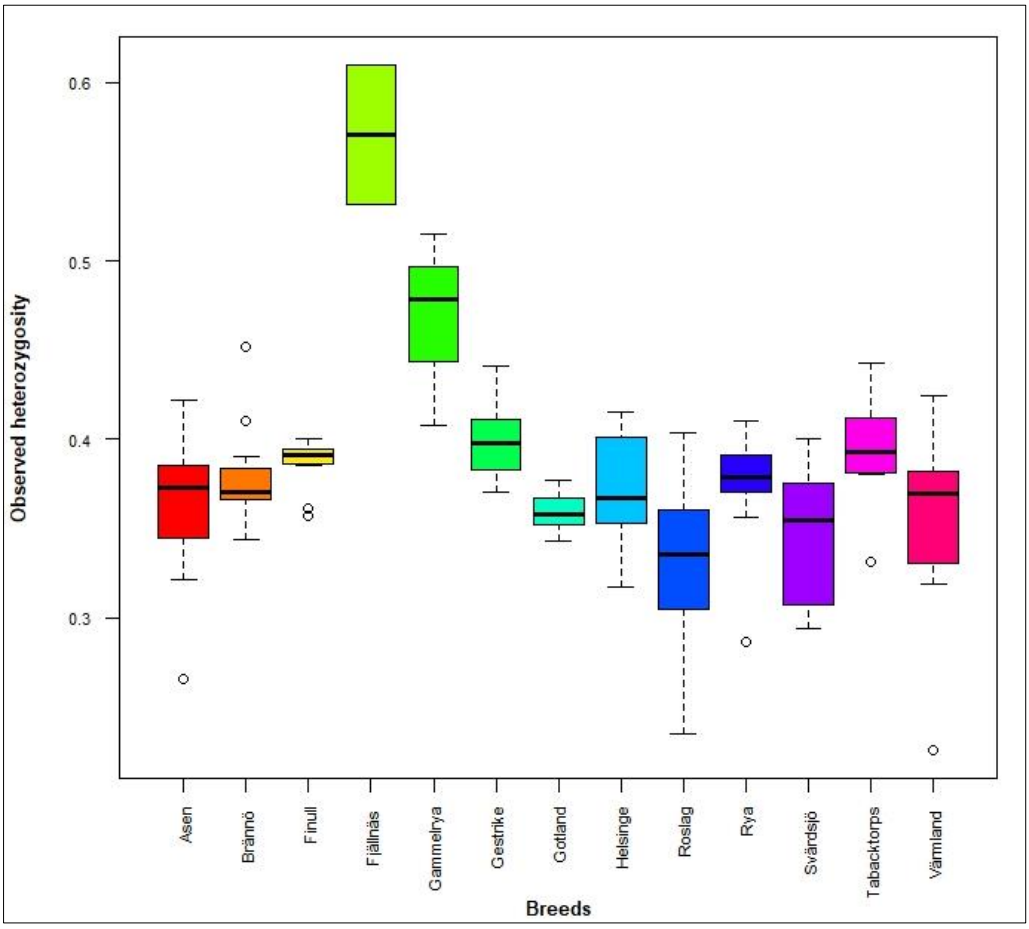


Figure 6 The observed heterozygosity of 189 individuals belonging to 13 Swedish sheep breeds.

The estimated average observed heterozygosity (O(HET)) values of 13 native Swedish sheep breeds are illustrated in Figure 6. The observed heterozygosity values ranged from 0.33 to 0.57. Roslag sheep had the lowest observed heterozygosity, whereas the highest was observed in Fjällnäs sheep. According to Table 5, the highest O(HET) value (O(HET)=1.0) was observed in Gute sheep, which was not illustrated in Figure 6. In this thesis, we had only one Gute sheep; thus, the values of all the diversity parameters can be unreliable due to low number of individuals considered in the analysis. Moreover, when we perform the observed heterozygosity calculation using --het flag in plink v.1.9, plink removes all loci that are not polymorphic within the breed (i.e. where all individuals are homozygous for the same allele). Since we had only one Gute sheep in the analysis, all homozygous loci were removed and only heterozygous loci remained in the calculation, thus the observed heterozygosity of Gute sheep was 1. Hence, it is better to exclude Gute sheep from the analyses.

4.2.2 Inbreeding Coefficient (F_{IS})

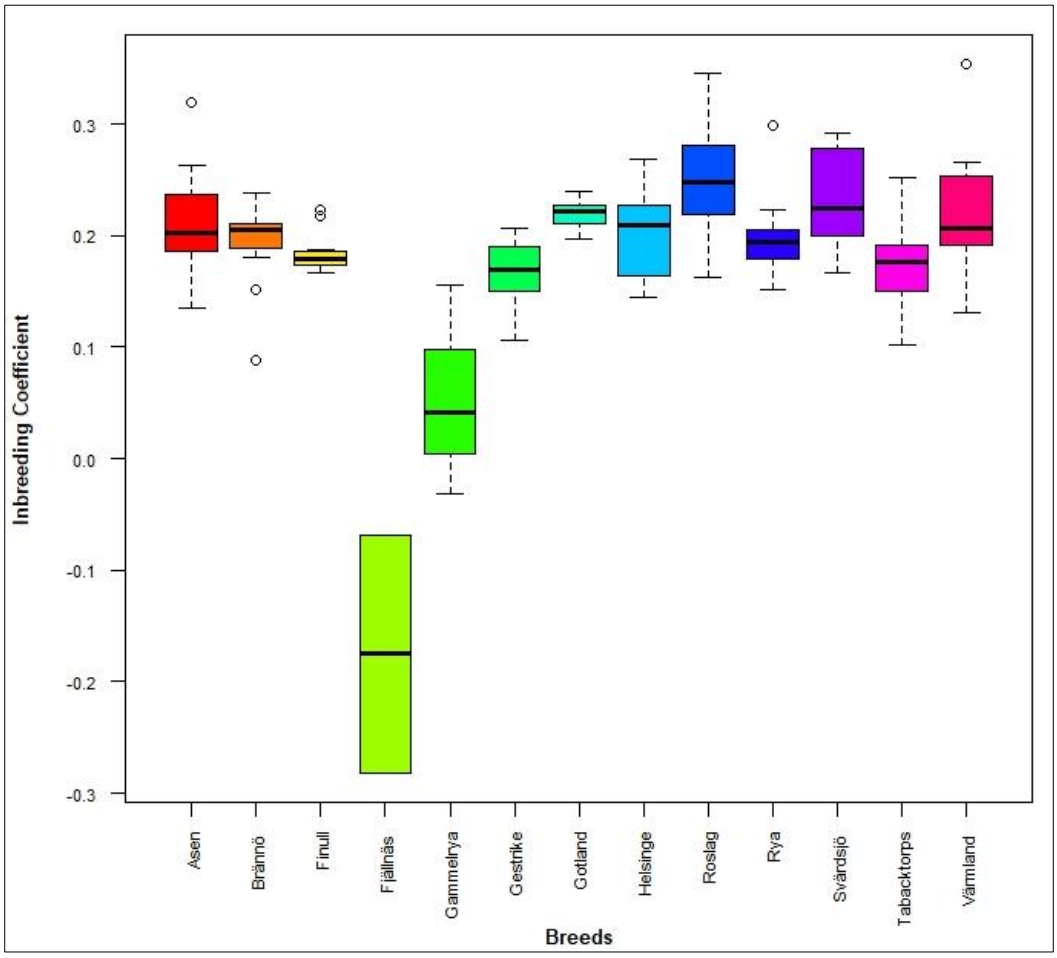


Figure 7 The inbreeding coefficients of 189 individuals belonging to 13 native Swedish sheep breeds. The colours in the plot match the breed colours in the previous figure.

The estimated average inbreeding coefficient (F_{IS}) values of 13 native Swedish sheep breeds are illustrated in Figure 7. The highest average F_{IS} was observed in Roslag sheep, and the lowest value was observed in Fjällnäs sheep. Under HWE and based on breed-specific allele frequencies, the excess homozygosity was measured using F_{IS}, and values ranged from -0.18 to 0.25. The negative average F_{IS} value in Fjällnäs sheep demonstrated the excess heterozygosity within the population. In this thesis, there were only two Fjällnäs sheep; thus, the value can be highly unreliable due to low number of individuals considered in the analysis. As shown in the admixture analysis (Figure 5), Roslag sheep appeared to be isolated: thus, the high inbreeding value within the population is possible.

4.2.3 Inbreeding Coefficient based on Runs of Homozygosity (FROH)

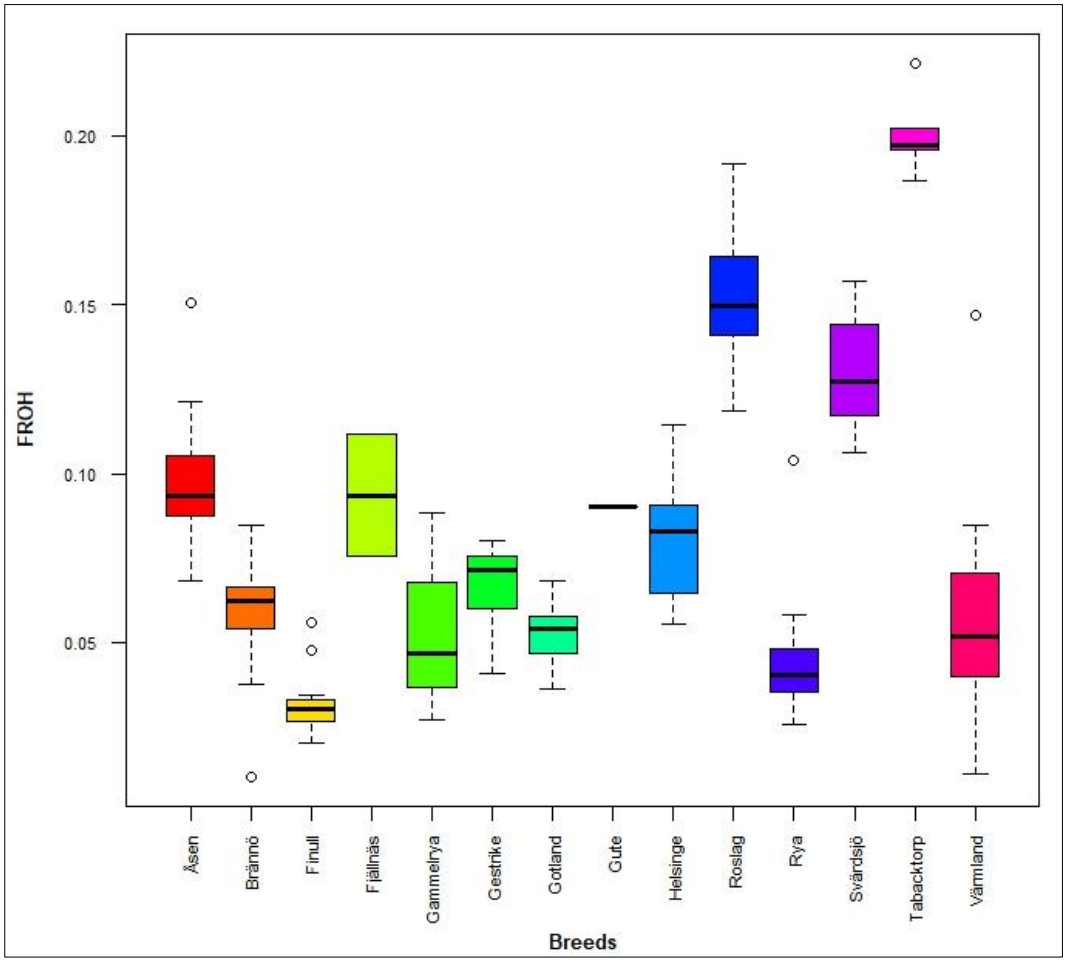


Figure 8 The inbreeding coefficient based on ROH for 14 native Swedish sheep breeds. The minimum length of a ROH was defined as 15 SNPs. For other parameters, see the materials and method section.

The inbreeding coefficients based on ROH (FROH) for 14 native Swedish sheep breeds are illustrated in Figure 8. The highest ROH-based average inbreeding coefficient was observed in the Tabacktorp, while the lowest was observed in the Finull breed. The average FROH values ranged from 0.03 to 0.20. The higher FROH values are associated with longer ROH lengths or a higher number of ROH within an individual's genome. These results were not fully in line with the average observed heterozygosity estimations shown in Figure 6. The higher values in Figure 6 indicate higher diversity, while higher values in Figure 8 indicate less diversity. The reason behind the disparity between Figure 7 and FROH values is that all homozygous positions contribute to observed homozygosity proportions in Figure 7, while only longer homozygous segments contribute to FROH. The F_{IS} calculates

the inbreeding level at each locus, whereas the FROH calculates inbreeding at the genome-wide level in each individual separately.

The average lengths of ROH and the number of ROH segments observed within each breed are shown in Table 7**.** The total ROH lengths were summed within population and then the average ROH per population was calculated to illustrate in the Table 7. All the individuals in each breed showed ROH segments under the given parameters shown in Table 3. The total genome length was 2651.46 Mbps. The average longest ROH length was observed in Tabacktorp sheep (530.56 Mbps), while the average shortest ROH length was observed in Finull sheep (83.82 Mbps). Nevertheless, the highest number of ROH segments was observed in Roslag sheep; 7470, while the lowest number of ROH segments was observed in Gute sheep; 231.

Overall, it is difficult to compare results with literature based on identified ROH islands as the genotyping methods are different, and all the input parameters are not often published.

Table 7 The average lengths of ROH and the number of ROH segments observed within each breed

Breed	**Average ROH length (Mbps)**	**Number of ROH segments**
Åsen	258.73	3257
Brännö	157.96	2516
Finull	83.82	1380
Fjällnäs	248.20	455
Gammelrya	143.47	398
Gestrike	175.56	1126
Gotland	139.54	6350
Gute	239.66	231
Helsinge	210.21	2845
Roslag	405.75	7470
Rya	118.13	1871
Svärdsjö	347.79	4179
Tabacktorp	530.56	3125
Värmland	154.07	2012

In addition, the distribution of ROH segments on each autosomal chromosome was illustrated in graphs. All 26 plots can be found in Appendix 2. In these figures, the y-axis only shows individuals that have a ROH on the corresponding chromosome. In other words, if an individual is missing from a graph, it indicates that the individual did not have a ROH segment on that chromosome under defined parameters. In some chromosomes, ROH segments were distributed only in a few regions: Chromosomes 14, 19, 21, and 24. Overall, relatively longer ROH segments were observed in Roslag and Tabacktorp sheep breeds on almost all the chromosomes: this may be an indication of relatively high inbreeding (as shown in

Figure 8) and recent common ancestors in each population as long ROH are more likely to occur when two copies of a haplotypes are inherited from a recent common ancestor (Francisco C. Ceballos et al., 2018).

4.3 Genetic Variants in MC1R Gene that Responsible for Coat Colour in Sheep

Genetic variants in sheep coat colour gene, MC1R and associated phenotypic variability were examined in 27 individuals of 7 native Swedish sheep breeds. In this thesis, two genetic variants were identified in MC1R exon1 region: one mutation is a known missense mutation on the E^D allele, c.218 T>A in the coding region of the gene, the other mutation, c.-31 G>A in the non-coding 5'-UTR region (Appendix 3). As a summary, detected phenotypes of c.218 T>A in the MC1R gene by coat colour and breed were illustrated in Table 8. At the c.218 T>A polymorphic position, 3 Finull sheep were heterozygous, and all the other individuals were homozygous and carried the recessive genotype exhibiting different phenotypes, as shown in Table 8. Accordingly, Finull sheep were the only breed with a variation at c.218 T>A, and white Finull individual was homozygous for c.218T, while other Finull sheep with black or brown coat colour were heterozygous at c.218 T>A. The genotype frequencies were 0, 0.11, and 0.89 for AA, TA, and TT, respectively; thus, the allele frequencies were 0.056 and 0.944 for the A, and T alleles, respectively, at the c.218 T>A position. There was no clear evidence of a c.-31 G>A-based effect on the coat colour of sheep. According to (Fontanesi et al., 2010), there is another polymorphic site at c.199 C>T in the MC1R exon 1 region. However, in the present study, we did not observe a polymorphism at this site with considered breeds, and all the animals were homozygous for c.199C.

Table 8 Detected phenotypes of c.218 T>A in Melanocortin receptor 1 (MC1R) by coat colour and breed

Coat colour	**c.218 T>A**	**Breed**							**Genotype Frequencies**
		Fin	**Ges**	**Sva**	**Gam**	**Rya**	**Got**	**Gut**	
Black	AA	0							0
Black	TA	1							
Brown		2							0.11
White	TT	1		4	1	4			0.89
Black			2	2		5			
Grey					1		1	1	
Brown						2			

Fin, Finull; Ges, Gestrike; Sva, Svärdsjö; Gam, Gammelrya; Rya, Rya; Got, Gotland; Gut, Gute

5. Discussion

In this study, the genetic diversity and population structure were measured in 14 Swedish sheep breeds using medium-density SNP (50K) data. Moreover, genetic variants in the melanocortin 1 receptor (MC1R) gene and associated coat colour phenotypes variability in 7 Swedish sheep breeds were studied by sequencing the MC1R exon1 region.

5.1 Population Structure

In the present study, population structure was examined using PCoA (Figure 3), NJ phylogenetic trees (a, b) (Figure 4) and admixture analysis using Structure v.2.3.4 software (Figure 5). The PCoA results indicated low to moderate genetic variation between breeds (the first three PCs explained 21.6% variation). The NJ phylogenetic tree (a) and the Structure analysis confirmed the results of PCoA. The NJ phylogenetic tree (b) showed the genetic distance between 12 native Swedish sheep breeds (except Fjällnäs and Gute sheep).

In the population structure analysis, the PCoA analysis revealed 4 distinct clusters: Gotland, Roslag, Svärdsjö, and Tabacktorp based on the genetic relationship. However, admixture analysis using Structure demonstrated 12 different genetic groups among these 191 individuals belonging to 14 Swedish sheep breeds. According to PCoA, Roslag sheep was highly distinguishable and unique from other breeds, and no shared ancestry with other populations (no structuring within population) was observed with the admixture analysis using Structure. The Roslag sheep originates from Roslagen on the island of Raggaron in Uppland county, Sweden. According to owners, Roslag sheep had always been resided on the island (Dahlberg, 2012): thus, this is related to the genetic isolation of the Roslag breed.

Moreover, as shown in PCoA and Structure plot, Tabacktorp, Gotland, and Svärdsjö sheep were distinct and less admixed with other populations. As evidenced by the admixture results, Fjällnäs, Gammelrya, Gute, and Rya breeds, Brännö 12940, 12941, and Värmland OAR 898 individuals showed a relatively high genetic admixture. In contrast, Roslag, Tabacktorp, Gotland, Åsen, Finull, Värmland (except OAR 898), Svärdsjö, Helsinge and Gestrike breeds were not or less admixed or interbred. The Värmland OAR 898 sheep was distinct from other Värmland sheep, and clustered close to Roslag in PCoA and NJ phylogenetic tree (a) and a higher proportion of Roslag admixture was observed in it (Appendix 1: Figure 19). Although the probability is very low, Structure plots suggest that this sheep may be a crossbred of Gotland and Roslag or maybe a new breed.

Additionally, one Gammelrya sheep (OAR 1004) was different from the other animals in the breed, which was clustered with Rya sheep in NJ phylogenetic tree (a). According to the Structure based admixture analysis, genetic structure of OAR 1004 sheep was similar to the genetic structure of Rya sheep. These results confirmed the known history of the OAR 1004 as it was a crossbred of Rya and Gammelrya breeds.

Gotland or Gotland pelt sheep have been bred for wool and skin production in Gotland Island, Sweden. Both male and female animals are hornless. In the first half of the 20th century, Gotland was bred with Rya sheep to improve the wool quality (Westberg Sunesson, 2015) which is in agreement with the population admixture shown in Figure 5. The Gute is a horned, wild sheep found on Gotland Island with a low reproduction ability (Maijala et al., 1990) and bred for conservation purposes. Moreover, according to Rochus et al. (2020), Gotland sheep is genetically more closed to Gute sheep. The Fjällnäs sheep breed is the latest addition to short-tailed Swedish sheep breeds and was discovered in Fjällnäs in North Sweden. They had been isolated with small population sizes (Rochus et al., 2020). According to the PCA analysis and the population tree in (Rochus et al., 2020): individuals genotyped using Ovine Infinium HD SNP BeadChip; Gotland, Gute and Fjällnäs sheep were distinguishable and unique from one another. Moreover, the individual co-ancestry coefficient analysis showed that Gotland and Gute sheep were more related, and there was some shared ancestry between Gotland and Fjällnäs sheep (Rochus et al., 2020). This was not evident in PCoA in this thesis which may be due to the low number of individuals in each population. However, in this thesis, the Structure plots reported evidence of shared ancestry between Gotland and Gute and a relatively lower proportion of admixture between Gotland and Fjällnäs breeds, which may be due to the crossbreeding between breeds in the population history.

Furthermore, the genetic admixture study using Structure revealed that the Rya breed was highly genetically admixed, and there were two distinct genetic groups within this breed. According to farmers' information, some farmers tend to breed Rya sheep within the same colour; thus, this could be a fact of appearing two distinct populations within Rya sheep. Moreover, this may associate with isolation of some Rya sheep farms from other Rya sheep farms. Rya sheep breed is a breed with glossy wool that nearly became extinct in the past. However, it was crossed with meat breeds to improve slaughter weight in the 1970s and is no longer endangered (Maijala et al., 1990). Nevertheless, according to Tapio et al. (2005), there was a sign of reduction in effective population size of Rya sheep.

5.2 Genetic Diversity

The genetic diversity of 14 Swedish sheep breeds was evaluated based on the average inbreeding coefficients (F_{IS}), observed heterozygosity, observed homozygosity, MAF, PIC, F_{ST}, and FROH. The highest average F_{IS} was observed in Roslag sheep (Figure 7), whereas the highest FROH was observed in Tabacktorp sheep (Figure 8). This indicates the relatively high amount of inbreeding within the Roslag and Tabacktorp populations. A study by (Szmatoła et al., 2019) showed the advantages of calculating ROH-based inbreeding coefficient over classical pedigree based inbreeding calculation, as it accurately accounts for autozygosity of the genome and it can be assessed in any animal even in the absence of pedigree information. Furthermore, Gute and Fjällnäs sheep showed high observed heterozygosity (Table 5): this might be an artefact of having very few samples of these two breeds. The PIC values were high in Finull, Värmland, and Rya sheep. The PIC measures the amount of genetic variation present at a particular locus (Botstein et al., 1980). None of the breeds showed PIC greater than 0.5, and the values ranged around 0.25 or below, indicating the presence of moderate to slightly informative polymorphic sites.

Furthermore, higher F_{ST} values indicate a more significant divergence between two populations. This thesis showed the highest F_{ST} value between Tabacktorp and Roslag sheep. In the conservation perspective, high F_{ST} is an indication of high inbreeding within each population. Since, these breeds are under conservation, it is important to monitor and assess the genetic information of breeds and take necessary actions to adjust breeding programs in order to preserve genetic materials. Alternatively, the lowest F_{ST} value between Gammelrya and Rya sheep demonstrated the lower differentiation between the two populations confirming high genetic admixture within Rya and Gammelrya sheep populations (Figure 5).

Data from previous studies are not fully comparable with the results of the present study because the estimation techniques (microsatellite markers and high-density SNP data) and the number of individuals within populations were different. According to Tapio et al. (2005), inbreeding values of Finull, Gotland, Gute, Roslag, and Rya breeds were 0.07, 0.15, 0.21, 0.41, and 0.12, respectively, estimated using 22 microsatellite markers. These values were slightly higher than in the present study average inbreeding coefficients of those breeds except in Gute and Roslag sheep. A study by Dahlberg (2012) showed the average observed heterozygosity of Fjällnäs, Gestrike, Helsinge, Roslag, Värmland, Svärdsjö, and Åsen using 18 microsatellite markers, and values were 0.43, 0.46, 0.61, 0.36, 0.6, 0.47 and 0.45, respectively. Nevertheless, this thesis showed relatively low observed heterozygosity values for those breeds except in Fjällnäs sheep. This may be associated with the high number of alleles present in the microsatellite markers-based analyses. Moreover, according to Rochus et al. (2020) with high density SNP array study, the average observed heterozygosity values of Gotland, Gute and

Fjällnäs sheep were 33.45, 31.44 and 29.98 percent, respectively, where Gotland sheep average observed heterozygosity value was in agreement with the present study output. Consequently, relatively low observed heterozygosity values were observed in the current study.

Furthermore, a study to evaluate the molecular variation in northern European sheep breeds, including Gotland, Finull, Gute, Rya and Roslag breeds, showed a sign of a reduction in effective population sizes using microsatellite data. Moreover, the study demonstrated the importance of maintaining allelic variation to the conservation of breeds and supporting the evolution of populations (Tapio et al., 2005). Hence, taking precautionary measures to minimize within-breed inbreeding is crucial while maintaining genetic diversity between breeds and preserving local genetic materials.

5.3 Genetic Variants in the MC1R Gene that Responsible for Coat Colour in Sheep

Genetic variants in the MC1R gene responsible for coat colour phenotypes in sheep were studied using 27 sheep belonging to seven different breeds: Finull, Gammelrya, Gestrike, Svärdsjö and Rya. The results compared with a previous study (Rochus et al., 2019) taking three sequences from three different breeds (Gotland, Gute, Finull). In this thesis, we sequenced only the exon 1 region in MC1R gene at the E locus. Conventional genetic studies have shown two alleles: dominant black, E^D and wild type, E^+ at the E locus affecting coat colour trait in sheep (Yang et al., 2013). The dominant E^D allele results in the black coat pattern in a few breeds, and the wild-type E^+ allele results in the white coat pattern in most breeds. In contrast to other species, the E^+ allele of the E locus in sheep has not yet been precisely identified (Fontanesi et al., 2010).

There are several mutations in the MC1R gene that affect coat colour in sheep: c.-31G>A, c.199C>T, c.218T>A, c.361G>A, c.429C>T, c.600T>G, and c.735C>T (Fontanesi et al., 2010; Våge et al., 1999). In the present study, two genetic variants were identified in the MC1R exon1 region: one known missense mutation on E^D allele, c.218 T>A (p.73Met>Lys) in the coding region of the gene, the other mutation, c.-31 G>A in the non-coding 5'-UTR region. These results were in agreement with previous studies that assessed genetic variants of the MC1R gene associated with sheep coat colour phenotype (Fontanesi et al., 2010; Rochus et al., 2019; Våge et al., 1999). Moreover, a study showed a missense mutation at c.199 C>T polymorphic position in first intracellular loop of the MC1R gene protein in Sicilian breed (Fontanesi et al., 2010). However, in this thesis, all the individuals were homozygous for c.199C. Furthermore, the Finull sheep were the only breed with a variation at c.218 T>A. The white Finull individual was homozygous for

wildtype (E^{+}), and other Finull sheep were either homozygous for c.218A or heterozygous while carrying black/ brown coat colour phenotype. These results confirmed the study by Rochus et al. (2019). It was not possible to determine the coat color of other breeds: Gammelrya, Svärdsjö, Gotland, Gute and Rya, based on MC1R c.218 T>A mutation. Accordingly, this study reveals that E^{D} is not the only allele that contribute to black coat colour in Swedish sheep as shown in Rochus et al. (2019). There was no clear evidence of a c.-31 G>A-based effect on the coat colour of sheep. However, this position was the first nucleotide of four haplotypes present on MC1R gene in sheep (Fontanesi et al., 2010). Overall, since, we had a limited number of individuals and only sequenced exon 1 region we were unable to draw a proper conclusion on determining black coat colour of Swedish sheep breeds based on the present genetic variants.

6. Conclusions

This study is the first to describe the genetic diversity and population structure of native Swedish sheep breeds using medium density SNP genotypes data. In the present study, genetic diversity and population structure analyses were performed using 50K SNP genotypes of fourteen native Swedish sheep breeds originated from different geographical locations in Sweden. The population structure was studied using the PCoA, NJ phylogenetic trees and Structure analysis. The genetic diversity was assessed based on the inbreeding coefficients, observed heterozygosity, MAF, PIC, and F_{ST}. The level of inbreeding was assessed using two measures: F_{IS} and FROH. Moreover, genetic variants in the melanocortin 1 receptor (MC1R) gene that responsible for coat colour in sheep was examined using seven Swedish sheep breeds. According to the population structure analysis, twelve distinct genetic populations were identified among fourteen breeds. Rya sheep showed high genetic admixture within the population. The Roslag sheep were highly varied from the others and appeared to be genetically isolated with no genetic admixture within population. The highest inbreeding level (F_{IS}) was reported in Roslag sheep, while the highest FROH was observed in Tabacktorp sheep. High F_{ST} value between Roslag and Tabacktorp populations is an indication of highly genetic differentiation between these two breeds and high level of inbreeding and low level of genetic admixture within each populations showing low genetic diversity. Furthermore, two previously discovered mutations in the MC1R gene in sheep were found in this thesis: c.-31 G>A and c.218 T>A. The c.218 T>A mutation was associated with black/brown coat color in Finull sheep, but in this thesis, it was not able to explain all the variations in coat color phenotypes in Swedish sheep breeds.

Future research is needed with increased number of individuals of a particular breed to study the population structure, genetic diversity and genetic variants in coat colour genes. It is important to compute rate of inbreeding and effective population size in order to draw proper conclusions. Furthermore, it is essential to focus on other candidate genes that are responsible for black and white coat colour of sheep. In addition, it is interesting to compare Swedish sheep breeds with other north European short-tailed sheep breeds and sheep breeds of different geographical origins in the world (HapMap data).

7. References

Abebe, A. S., Mikko, S [Sofia], & Johansson, A. M [Anna M.] (2015). Genetic diversity of five local Swedish chicken breeds detected by microsatellite markers. *PLOS ONE*, *10*(4), e0120580. https://doi.org/10.1371/journal.pone.0120580

Abied, A., Bagadi, A., Bordbar, F., Pu, Y., Augustino, S. M. A., Xue, X., Xing, F., Gebreselassie, G., Mwacharo, J.-L. H. J. M., Ma, Y., & Zhao, Q. (2020). Genomic Diversity, Population Structure, and Signature of Selection in Five Chinese Native Sheep Breeds Adapted to Extreme Environments. *Genes*, *11*(5), 494. https://doi.org/10.3390/genes11050494

Alberto, F. J., Boyer, F., Orozco-terWengel, P., Streeter, I., Servin, B., Villemereuil, P. de, Benjelloun, B., Librado, P., Biscarini, F [Filippo], Colli, L [Licia], Barbato, M., Zamani, W., Alberti, A [Adriana], Engelen, S., Stella, A., Joost, S [Stéphane], Ajmone-Marsan, P [Paolo], Negrini, R [Riccardo], Orlando, L., . . . Pompanon, F [François] (2018). Convergent genomic signatures of domestication in sheep and goats. *Nature Communications*, *9*(1), 813. https://doi.org/10.1038/s41467-018-03206-y

Barker, J. S. F. (2001). Conservation and management of genetic diversity: A domestic animal perspective. *Canadian Journal of Forest Research*, *31*(4), 588–595. https://doi.org/10.1139/x00-180

Biscarini, F [Filippo], Cozzi, P., Gaspa, G., & Marras, G. (2018). *Detectruns: Detect runs of homozygosity and runs of heterozygosity in diploid genomes*. CRAN (The Comprehensive R Archive Network). https://orca.cardiff.ac.uk/id/eprint/108906/

Biscarini, F [Filippo], Mastrangelo, S [Salvatore], Catillo, G., Senczuk, G [Gabriele], & Ciampolini, R [Roberta] (2020). Insights into Genetic Diversity, Runs of Homozygosity and Heterozygosity-Rich Regions in Maremmana Semi-Feral Cattle Using Pedigree and Genomic Data. *Animals*, *10*(12), 2285. https://doi.org/10.3390/ani10122285

Boettcher, P. J., Tixier-Boichard, M., Toro, M. A., Simianer, H., Eding, H., Gandini, G., Joost, S [S.], Garcia, D., Colli, L [L.], & Ajmone-Marsan, P [P.] (2010). Objectives, criteria and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. *Animal Genetics*, *41 Suppl 1*, 64–77. https://doi.org/10.1111/j.1365-2052.2010.02050.x

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). *Construction of a genetic linkage map in man using restriction fragment length polymorphisms* (3rd ed.). American journal of human genetics. https://www.ncbi.nlm.nih.gov/pmc/articles/pmc1686077/

Brito, L. F., Kijas, J. W., Ventura, R. V., Sargolzaei, M., Porto-Neto, L. R., Cánovas, A., Feng, Z., Jafarikia, M., & Schenkel, F. S. (2017). Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genomics*, *18*(1), 229. https://doi.org/10.1186/s12864-017-3610-0

Bruford, M. W., Bradley, D. G., & Luikart, G. (2003). Dna markers reveal the complexity of livestock domestication. *Nature Reviews Genetics*, *4*(11), 900–910. https://doi.org/10.1038/nrg1203

Buchanan, F. C., & Thue, T. D. (1998). Intrabreed polymorphic information content of microsatellites in cattle and sheep. *Canadian Journal of Animal Science*, *78*(3), 425–428. https://doi.org/10.4141/A98-002

Bultman, S. J., Michaud, E. J., & Woychik, R. P. (1992). Molecular characterization of the mouse agouti locus. *Cell*, *71*(7), 1195–1204. https://doi.org/10.1016/S0092-8674(05)80067-4

Ceballos, F. C [F. C.], Joshi, P. K., Clark, D. W., Ramsay, M [M.], & Wilson, J. F. (2018). Runs of homozygosity: Windows into population history and trait architecture. *Nature Reviews Genetics*, *19*(4), 220–234. https://doi.org/10.1038/nrg.2017.109

Ceballos, F. C [Francisco C.], Hazelhurst, S., & Ramsay, M [Michèle] (2018). Assessing runs of Homozygosity: A comparison of SNP Array and whole genome sequence low coverage data. *BMC Genomics*, *19*(1), 106. https://doi.org/10.1186/s12864-018-4489-0

Chessa, B., Pereira, F., Arnaud, F., Amorim, A., Goyache, F [Félix], Mainland, I., Kao, R. R., Pemberton, J. M., Beraldi, D., Stear, M. J., Alberti, A [Alberto], Pittau, M., Iannuzzi, L., Banabazi, M. H., Kazwala, R. R., Zhang, Y.-P., Arranz, J. J [Juan J.], Ali, B. A., Wang, Z [Zhiliang], . . . Palmarini, M. (2009). Revealing the history of sheep domestication using retrovirus integrations. *Science*, *324*(5926), 532–536. https://doi.org/10.1126/science.1170587

Dahlberg, L. (2012). *Genetic variation in local Swedish sheep breeds*. https://stud.epsilon.slu.se/4140/1/dahlberg_l_120420.pdf

Dakin, E. E., & Avise, J. C. (2004). Microsatellite null alleles in parentage analysis. *Heredity*, *93*(5), 504–509. https://doi.org/10.1038/sj.hdy.6800545

Decker, J. E., McKay, S. D., Rolf, M. M., Kim, J., Molina Alcalá, A., Sonstegard, T. S., Hanotte, O., Götherström, A., Seabury, C. M., Praharani, L., Babar, M. E., Correia de Almeida Regitano, L., Yildiz, M. A., Heaton, M. P., Liu, W.-S., Lei, C.-Z., Reecy, J. M., Saif-Ur-Rehman, M., Schnabel, R. D., & Taylor, J. F. (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLOS Genetics*, *10*(3), e1004254. https://doi.org/10.1371/journal.pgen.1004254

Dýrmundsson, O. R., & Niżnikowski, R. (2010). North European short-tailed breeds of sheep: A review. *Animal*, *4*(8), 1275–1282. https://doi.org/10.1017/s175173110999156x

Earl, D. A., & vonHoldt, B. M. (2012). Structure HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359–361. https://doi.org/10.1007/s12686-011-9548-7

Falconer, D. S., & MCKAY, T. (1996). *Quantitative Genetics.* Longman Group Ltd.

Fariello, M.-I., Servin, B., Tosser-Klopp, G., Rupp, R., Moreno, C., San Cristobal, M., & Boitard, S. (2014). Selection signatures in worldwide sheep populations. *PLOS ONE*, *9*(8), e103813. https://doi.org/10.1371/journal.pone.0103813

Ferenčaković, M., Sölkner, J., & Curik, I. (2013). Estimating autozygosity from high-throughput information: Effects of SNP density and genotyping errors. *Genetics Selection Evolution*, *45*(1), 42. https://doi.org/10.1186/1297-9686-45-42

Fontanesi, L., Beretti, F., Riggio, V., Dall'Olio, S., Calascibetta, D., Russo, V., & Portolano, B [B.] (2010). Sequence characterization of the melanocortin 1 receptor (MC1R) gene in sheep with different coat colours and identification of the putative e allele at the ovine Extension locus. *Small Ruminant Research*, *91*(2-3), 200–207. https://doi.org/10.1016/j.smallrumres.2010.03.015

Frandson, R. D. (1986). *Anatomy and physiology of farm animals* (No. Ed. 4). Lea & Febiger.

Frankham, R., Briscoe, D. A., & Ballou, J. D. (2002). *Introduction to conservation genetics*. Cambridge University Press.

Ganteil, A., Rodriguez-Ramilo, S. T., Ligonesche, B., & Larzul, C. (2020). Characterization of Autozygosity in Pigs in Three-Way Crossbreeding. *Frontiers in Genetics*, *11*, 584556. https://doi.org/10.3389/fgene.2020.584556

Gebreselassie, G., Liang, B., Berihulay, H., Islam, R., Abied, A., Jiang, L., Zhao, Z., & Ma, Y. (2021). Correction: Genomic mapping identifies two genetic variants in the MC1R gene for coat colour variation in Chinese Tan sheep. *PLOS ONE*, *16*(1), e0245674. https://doi.org/10.1371/journal.pone.0245674

Groeneveld, L. F., Lenstra, J. A [J. A.], Eding, H., Toro, M. A., Scherf, B., Pilling, D., Negrini, R [R.], Finlay, E. K., Jianlin, H., Groeneveld, E., & Weigend, S. (2010). Genetic diversity in farm animals--a review. *Animal Genetics*, *41 Suppl 1*, 6–31. https://doi.org/10.1111/j.1365-2052.2010.02038.x

Guo, X., & Elston, R. C. (1999). Linkage information content of polymorphic genetic markers. *Human Heredity*, *49*(2), 112–118. https://doi.org/10.1159/000022855

Hanotte O., Toll J., In˜iguez L., & Rege E. (2005). *Farm animal genetic resources: why and what do we need to conserve*. https://scholar.google.com/citations?user=qz7ws2uaaaaj&hl=en&oi=sra

Hernandez, R. D., Uricchio, L. H., Hartman, K., Ye, C., Dahl, A., & Zaitlen, N. (2019). Ultrarare variants drive substantial cis heritability of human gene expression. *Nature Genetics*, *51*(9), 1349–1355. https://doi.org/10.1038/s41588-019-0487-7

Hiendleder, S., Kaupe, B., Wassmuth, R., & Janke, A. (2002). Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proceedings. Biological Sciences*, *269*(1494), 893–904. https://doi.org/10.1098/rspb.2002.1975

Hoffmann, I. (2010). Climate change and the characterization, breeding and conservation of animal genetic resources. *Animal Genetics*, *41*, 32–46. https://doi.org/10.1111/j.1365-2052.2010.02043.x

Hubbard, J. K., Uy, J. A. C., Hauber, M. E., Hoekstra, H. E., & Safran, R. J. (2010). Vertebrate pigmentation: From underlying genes to adaptive function. *Trends in Genetics*, *26*(5), 231–239. https://doi.org/10.1016/j.tig.2010.02.002

Kalds, P., Zhou, S., Gao, Y [Yawei], Cai, B., Huang, S., Chen, Y., & Wang, X. (2022). Genetics of the phenotypic evolution in sheep: A molecular look at diversity-driving genes. *Genetics Selection Evolution*, *54*(1), 61. https://doi.org/10.1186/s12711-022-00753-3

Kijas, J. W., Lenstra, J. A [Johannes A.], Hayes, B., Boitard, S., Porto Neto, L. R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E [Elena], Raadsma, H., McEwan, J., & Dalrymple, B. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLOS Biology*, *10*(2), e1001258. https://doi.org/10.1371/journal.pbio.1001258

Klungland, H., Røed, K. H., Nesbø, C. L., Jakobsen, K. S., & Våge, D. I. (1999). The melanocyte-stimulating hormone receptor (MC1-R) gene as a tool in evolutionary studies of artiodactyles. *Hereditas*, *131*(1), 39–46. https://doi.org/10.1111/j.1601-5223.1999.00039.x

Koseniuk, A., Ropka-Molik, K., Rubiś, D., & Smołucha, G. (2018). Genetic background of coat colour in sheep. *Archives Animal Breeding*, *61*(2), 173–178. https://doi.org/10.5194/aab-61-173-2018

Li, M.-H., Strandén, I., Tiirikka, T., Sevón-Aimonen, M.-L., & Kantanen, J [Juha] (2011). A comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in a sheep population. *PLOS ONE*, *6*(11), e26256. https://doi.org/10.1371/journal.pone.0026256

Mahmoud, A. H., Mashaly, A. M., Rady, A. M., Al-Anazi, K. M., & Saleh, A. A. (2017). Allelic variation of melanocortin-1 receptor locus in Saudi indigenous sheep exhibiting different color coats. *Asian-Australasian Journal of Animal Sciences*, *30*(2), 154–159. https://doi.org/10.5713/ajas.16.0138

Maijala, k., Neimann-Sorensen, A., Adalsteinsson, S., Kolstad, N., Danell, B., & Gjelstad, B. (1990). Conservation OF GENE RESOURCES OF FARM ANIMALS IN THE NORDIC COUNTRIES. *Animal Genetic Resources Information*, *7*, 45–53. https://doi.org/10.1017/s101423390000300x

Mastrangelo, S [Salvatore], Di Gerlando, R., Tolone, M., Tortorici, L., Sardina, M. T., & Portolano, B [Baldassare] (2014). Genome wide linkage disequilibrium and genetic structure in Sicilian dairy sheep breeds. *BMC Genetics*, *15*(1), 108. https://doi.org/10.1186/s12863-014-0108-5

Melo, R. C. D., Trevisani, N., Pereira, T. C. V., Guidolin, A. F., & Coimbra, J. L. M. (2017). *Heterozygosity level and its relationship with genetic variability mechanisms in beans1*. Revista Ciência Agronômica. https://www.scielo.br/j/rca/a/3pthcfy7jtgtq54rtsryfbv/?format=html&lang=en

Meyermans, R., Gorssen, W., Buys, N., & Janssens, S. (2020a). How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics*, *21*(1), 94. https://doi.org/10.1186/s12864-020-6463-x

Meyermans, R., Gorssen, W., Wijnrocx, K., Lenstra, J. A [J. A.], Vellema, P., Buys, N., & Janssens, S. (2020b). Unraveling the genetic diversity of Belgian Milk Sheep using medium-density SNP genotypes. *Animal Genetics*, *51*(2), 258–265. https://doi.org/10.1111/age.12891

Mignon-Grasteau, S., Boissy, A., Bouix, J., Faure, J.-M., Fisher, A. D., Hinch, G. N., Jensen, P., Le Neindre, P., Mormède, P., Prunet, P., Vandeputte, M., & Beaumont, C. (2005). Genetics of adaptation and domestication in livestock. *Livestock Production Science*, *93*(1), 3–14. https://doi.org/10.1016/j.livprodsci.2004.11.001

Mohammadi, S. A., & Prasanna, B. M. (2003). Analysis of Genetic Diversity in Crop Plants—Salient Statistical Tools and Considerations. *Crop Science*, *43*(4), 1235–1248. https://doi.org/10.2135/cropsci2003.1235

Mukiibi, R., Rochus, C. M., Andersson, G., & Johansson, A. M [A. M.] (2015). The use of endogenous retroviruses as markers to describe the genetic relationships among local Swedish sheep breeds. *Animal Genetics*, *46*(2), 220–223. https://doi.org/10.1111/age.12264

Nadia Aubin-Horth, ‡ Daniel A.J. Ryan, Shawn P. Good§, & Julian J. Dodson (2005). Balancing selection on size: Effects on the incidence of an alternative reproductive tactic. *Evolutionary Ecology Research*, *7*(8), 1171–1182.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, *89*(3), 583–590. https://doi.org/10.1093/genetics/89.3.583

Pembleton, L. W., Cogan, N. O. I., & Forster, J. W. (2013). Stampp: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, *13*(5), 946–952. https://doi.org/10.1111/1755-0998.12129

Peripolli, E., Munari, D. P., Silva, M. V. G. B., Lima, A. L. F., Irgang, R., & Baldi, F. (2017). Runs of homozygosity: Current knowledge and applications in livestock. *Animal Genetics*, *48*(3), 255–271. https://doi.org/10.1111/age.12526

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. https://doi.org/10.1093/genetics/155.2.945

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., Bakker, P. I. W. de, Daly, M. J., & Sham, P. C. (2007). Plink: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Purfield, D. C., Berry, D. P., McParland, S., & Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genetics*, *13*(1), 70. https://doi.org/10.1186/1471-2156-13-70

Renieri, C., Valbonesi, A., La Manna, V., Antonini, M., & Lauvergne, J. J. (2008). Inheritance of coat colour in Merino sheep. *Small Ruminant Research*, *74*(1-3), 23–29. https://doi.org/10.1016/j.smallrumres.2007.03.001

Robbins, L. S., Nadeau, J. H., Johnson, K. R., Kelly, M. A., Roselli-Rehfuss, L., Baack, E., Mountjoy, K. G., & Cone, R. D. (1993). Pigmentation phenotypes of variant extension locus alleles result from point mutations that alter MSH receptor function. *Cell*, *72*(6), 827–834. https://doi.org/10.1016/0092-8674(93)90572-8

Rochus, C. M., Jonas, E., & Johansson, A. M [A. M.] (2020). Population structure of five native sheep breeds of Sweden estimated with high density SNP genotypes. *BMC Genetics*, *21*(1), 27. https://doi.org/10.1186/s12863-020-0827-8

Rochus, C. M., Westberg S., K., Jonas, E., Mikko, S [S.], & Johansson, A. M [A. M.] (2019). Mutations in ASIP and MC1R: Dominant black and recessive black alleles segregate in native Swedish sheep populations. *Animal Genetics*, *50*(6), 712–717. https://doi.org/10.1111/age.12837

Royo, L. J., Alvarez, I., Arranz, J. J [J. J.], Fernández, I., Rodríguez, A., Pérez-Pardal, L., & Goyache, F [F.] (2008). Differences in the expression of the ASIP gene are involved in the recessive black coat colour pattern in sheep: Evidence from the rare Xalda sheep breed. *Animal Genetics*, *39*(3), 290–293. https://doi.org/10.1111/j.1365-2052.2008.01712.x

Ryder, M. L., Land, R. B., & Ditchburn, R. (1974). Coat colour inheritance in Soay, Orkney and Shetland sheep. *Journal of Zoology*, *173*(4), 477–485. https://doi.org/10.1111/j.1469-7998.1974.tb04128.x

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Schmidtmann, C., Schönherz, A., Guldbrandtsen, B., Marjanovic, J., Calus, M., Hinrichs, D., & Thaller, G. (2021). Assessing the genetic background and genomic relatedness of red cattle populations originating from Northern Europe. *Genetics Selection Evolution*, *53*(1), 23. https://doi.org/10.1186/s12711-021-00613-6

Senczuk, G [G.], Mastrangelo, S [S.], Ciani, E [E.], Battaglini, L., Cendron, F., Ciampolini, R [R.], Crepaldi, P., Mantovani, R., Bongioni, G., Pagnacco, G., Portolano, B [B.], Rossoni, A., Pilla, F., & Cassandro, M. (2020). The genetic heritage of Alpine local cattle breeds using genomic SNP data. *Genetics Selection Evolution*, *52*(1), 40. https://doi.org/10.1186/s12711-020-00559-1

Su, R., Qiao, X., Gao, Y [Yun], Li, X., Jiang, W., Chen, W., Fan, Y., Zheng, B., Zhang, Y., Liu, Z., Wang, R., Wang, Z [Zhiying], Wang, Z [Zhixin], Wan, W., Dong, Y., & Li, J. (2020). Draft Genome of the European Mouflon (Ovis orientalis musimon). *Frontiers in Genetics*, *11*, 533611. https://doi.org/10.3389/fgene.2020.533611

Szmatoła, T., Gurgul, A., Jasielczuk, I., Ząbek, T., Ropka-Molik, K., Litwińczuk, Z., & Bugno-Poniewierska, M. (2019). A Comprehensive Analysis of Runs of Homozygosity of Eleven Cattle Breeds Representing Different Production Types. *Animals*, *9*(12), 1024. https://doi.org/10.3390/ani9121024

Taberlet, P [P.], Coissac, E [E.], Pansu, J., & Pompanon, F [F.] (2011). Conservation genetics of cattle, sheep, and goats. *Comptes Rendus Biologies*, *334*(3), 247–254. https://doi.org/10.1016/j.crvi.2010.12.007

Tapio, M., Tapio, I., Grislis, Z., Holm, L. E., Jeppsson, S., Kantanen, J [J.], Miceikiene, I., Olsaker, I [I.], Viinalass, H., & Eythorsdottir, E [E.] (2005). Native breeds demonstrate high contributions to the molecular variation in northern European sheep. *Molecular Ecology*, *14*(13), 3951–3963. https://doi.org/10.1111/j.1365-294X.2005.02727.x

Thomas, D. L. (2010). Performance and utilization of Northern European short-tailed breeds of sheep and their crosses in North America: A review. *Animal*, *4*(8), 1283–1296. https://doi.org/10.1017/s1751731110000856

Våge, D. I., Fleet, M. R., Ponz, R., Olsen, R., T., Monteagudo, L. V., Tejedor, M. T., Arruga, M. V., Gagliardi, R., Postiglioni, A., Nattrass, G. S., & Klungland, H. (2003). Mapping and characterization of the dominant black colour locus in sheep. *Pigment Cell Research*, *16*(6), 693–697. https://doi.org/10.1046/j.1600-0749.2003.00090.x

Våge, D. I., Klungland, H., Lu, D., & Cone, R. D. (1999). Molecular and pharmacological characterization of dominant black coat color in sheep. *Mammalian Genome : Official Journal of the International Mammalian Genome Society*, *10*(1), 39–43. https://doi.org/10.1007/s003359900939

Westberg Sunesson, K. (2015). *Genetic variation of ASIP and MC1R in past and present sheep of Gotland. Examensarbete / Sveriges lantbruksuniversitet, Fakulteten för veterinärmedicin och husdjursvetenskap, Veterinärprogrammet: 2015:03*. SLU, Dept. of Animal Breeding and Genetics. https://stud.epsilon.slu.se/7613/

Williams, J. L., Hall, S. J. G., Del Corvo, M., Ballingall, K. T., Colli, L [L.], Ajmone Marsan, P., & Biscarini, F [F.] (2016). Inbreeding and purging at the genomic

Level: The Chillingham cattle reveal extensive, non-random SNP heterozygosity. *Animal Genetics*, *47*(1), 19–27. https://doi.org/10.1111/age.12376

Wright, S. (1984). *Evolution and the genetics of populations: A treatise in four volumes*. the University of Chicago Press.

Yang, G.-L., Fu, D.-L., Lang, X., Wang, Y.-T., Cheng, S.-R., Fang, S.-L., & Luo, Y.-Z. (2013). Mutations in MC1R gene determine black coat color phenotype in Chinese sheep. *TheScientificWorldJournal*, *2013*, 675382. https://doi.org/10.1155/2013/675382

Zeder, M. A. (2008). Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(33), 11597–11604. https://doi.org/10.1073/pnas.0801317105

Zhivotovsky, L. A. (2015). Relationships Between Wright's F ST and F IS Statistics in a Context of Wahlund Effect. *The Journal of Heredity*, *106*(3), 306–309. https://doi.org/10.1093/jhered/esv019

Popular science summary

The sheep was one of the first grazing animals to be domesticated and have been shown to be genetically diverse livestock species. The Swedish sheep breeds belong to a group of sheep breeds known as North European short-tailed sheep breeds, which are genetically unique, display various phenotypes (particularly coat colour patterns), and are one of the major contributors to genetic variation in domestic sheep. Native sheep breeds conservation requires proper breeding decisions and sufficient genetic management of small populations to preserve genetic variability and minimize inbreeding. Studies of genetic diversity and population structure shed light on the genetic relatedness and differences between breeds, a source of information for implementing successful breeding programs, future breed development and future scientific research. The thesis aimed to study genetic diversity and the population structure of 191 sheep belonging to fourteen Swedish native sheep breeds using medium density (50K) SNP genotype data.

The population structure was studied with principal coordinate analysis (PCoA), neighbour-joining (NJ) trees (a, b), and admixture analysis using Structure v.2.3.4. The genetic diversity was assessed based on the inbreeding coefficients, average observed heterozygosity, minor allele frequency (MAF), polymorphic information content (PIC), and pairwise F_{ST}. The level of inbreeding was estimated using two measures: F_{IS} and runs of homozygosity-based inbreeding (FROH). Moreover, genetic variants in the melanocortin 1 receptor (MC1R) gene that is responsible for coat colour in sheep were examined using seven Swedish sheep breeds.

According to the population structure analyses, sheep breeds were genetically different from each other. Genetic diversity analysis showed slightly higher inbreeding levels within Roslag and Tabacktorp sheep. According to the MC1R gene sequencing analysis, two previously discovered mutations in the MC1R gene in sheep were found: c.-31 G>A and c.218 T>A. The c.218 T>A mutation was associated with black/brown coat colour in Finull sheep. There was no evidence of the c.-31 G>A polymorphic site based effects on the coat colour of sheep; hence, further research studies are needed.

Moreover, this thesis suggesting the requirement of future research on computing the rate of inbreeding and effective population size in Swedish sheep breeds and compare the genetic diversity and population structure of them with other North European short-tailed sheep breeds of different geographical origins in

the world. In addition, it is essential to focus on other candidate genes that are responsible for black and white coat colour of sheep.

Acknowledgements

My first and foremost gratitude is dedicated to my supervisor, Dr. Anna Maria Johansson, Department of Animal Breeding and Genetics, Faculty of Veterinary Medicine and Animal Science, Swedish University of Agricultural Sciences, Sweden, for all her excellent guidance throughout the research study and kind supervision for the successful completion of my master thesis.

I would like to express my sincere gratitude to my examiner, Prof. Erling Strandberg, Department of Animal Breeding and Genetics, Faculty of Veterinary Medicine and Animal Science, Swedish University of Agricultural Sciences, Sweden, for all his kind guidance, valuable comments and support in correcting my thesis work.

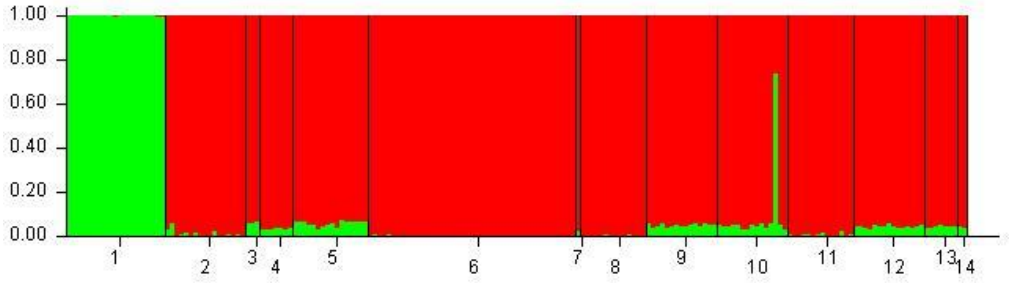
I would like to express my sincere gratitude to my assistant supervisor, Dr. Steffen Weigend, Head of Research Division Breeding and Genetic Resources, Friedrich-Loeffler-Institut, Institute of Farm Animal Genetics (FLI), Germany, for all his kind guidance and valuable support throughout the research study.

I would like to express my sincere gratitude to my assistant supervisor, Dr. Christian Reimer, Research Scientist, Research Division Breeding and Genetic Resources, Friedrich-Loeffler-Institut, Institute of Farm Animal Genetics (FLI), Germany, for all his inspiring guidance, constant encouragement and valuable support throughout the research study.

I am also grateful to Erasmus Mundus Association and EMABG Program Organizing Committee for giving me this valuable opportunity to study in two renowned European universities (University of Goettingen and Swedish University of Agricultural Sciences) including financial support.

Appendix 1 - Structure Analysis

(a)



(b)

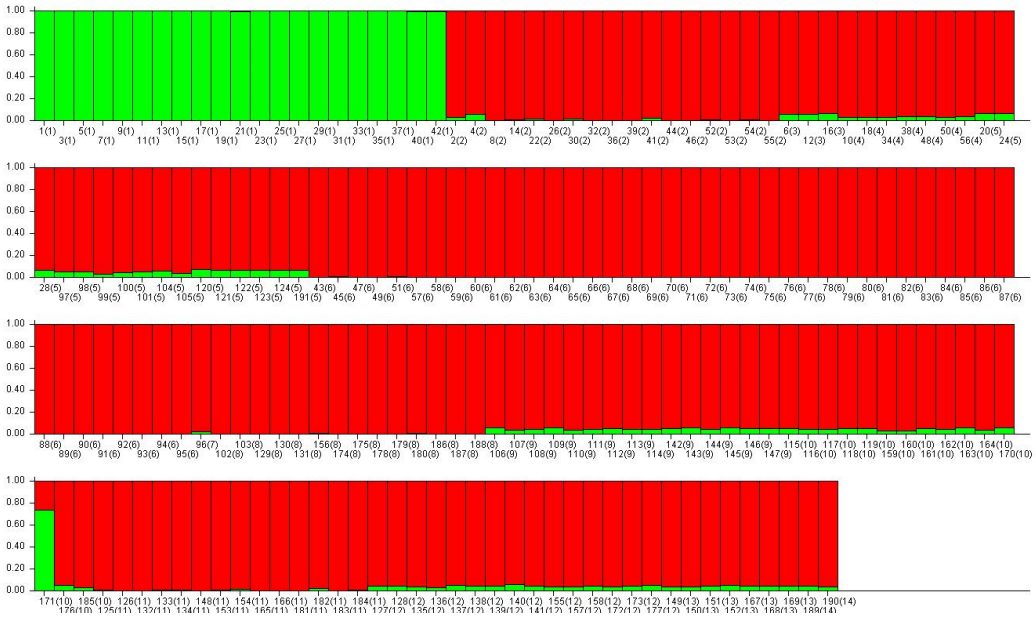
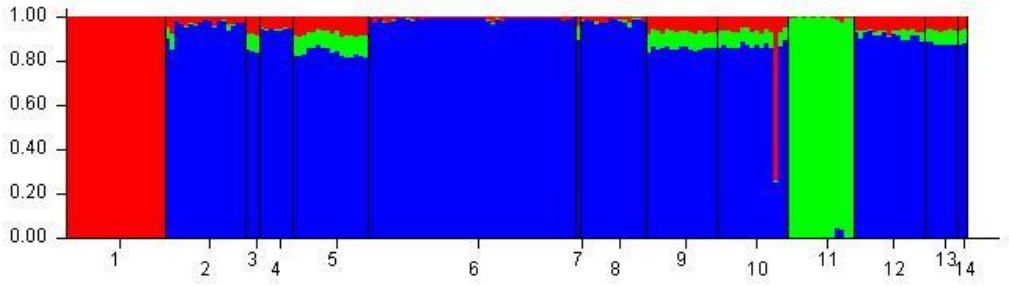


Figure 9 Genome-wide admixture proportions at K=2 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 2 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs.

(a)



(b)

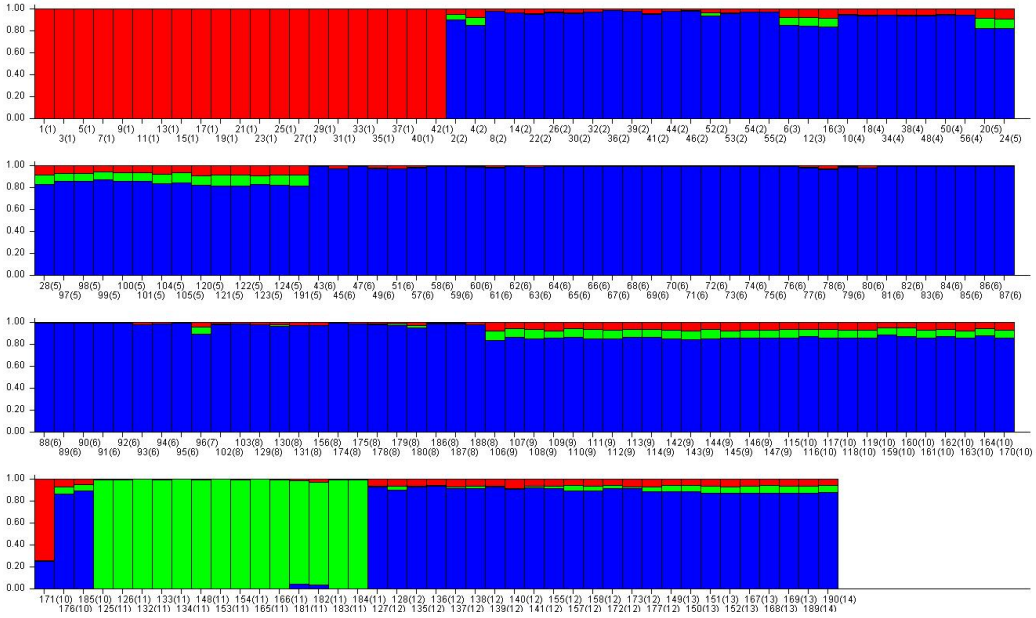
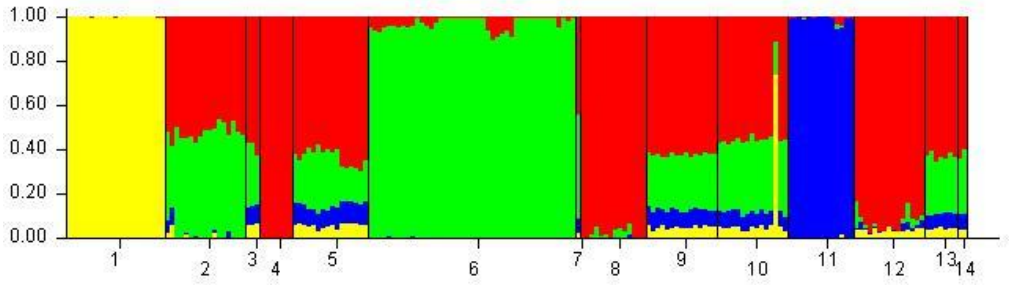


Figure 10 Genome-wide admixture proportions at K=3 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 3 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs.

(a)



(b)

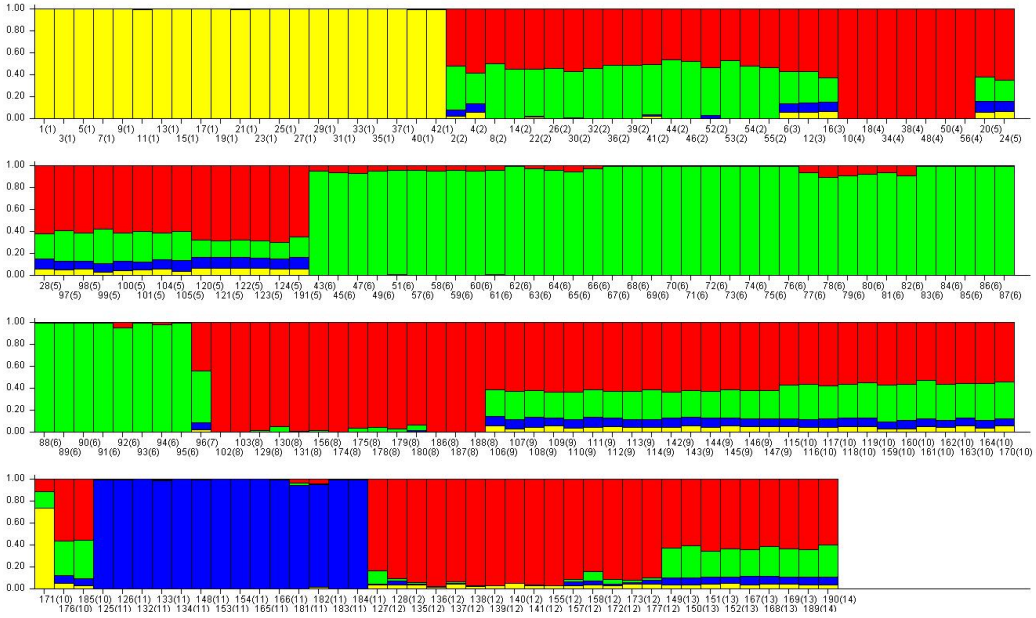
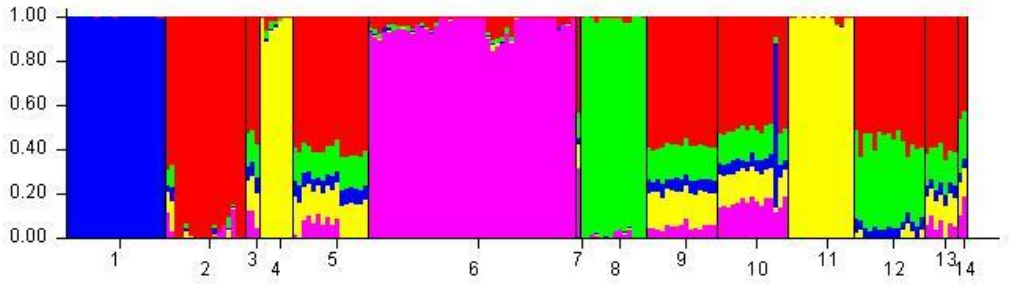


Figure 11 Genome-wide admixture proportions at K=4 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 4 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs.

(a)



(b)

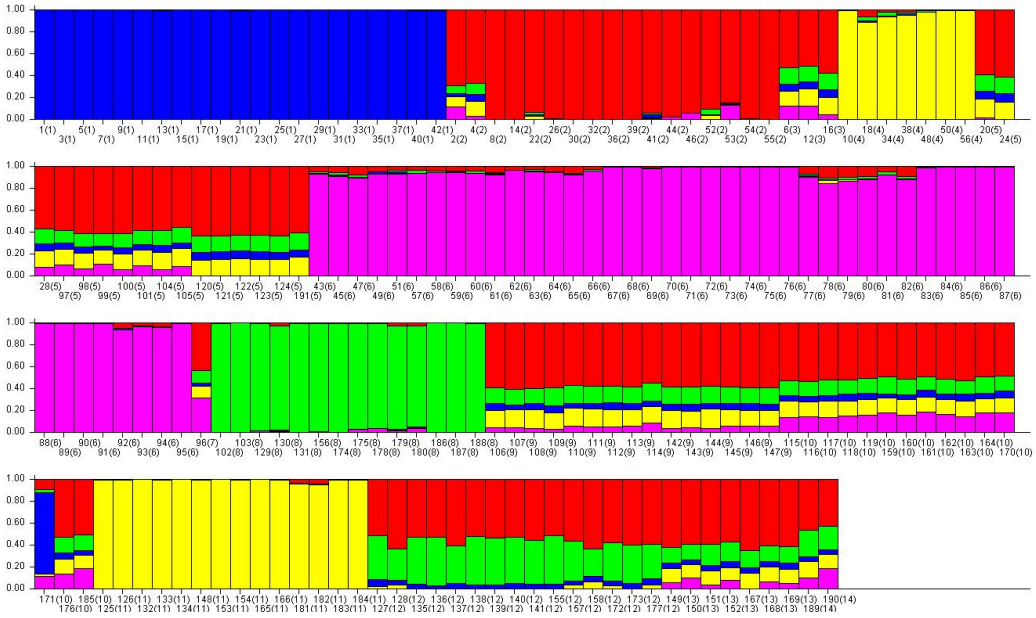
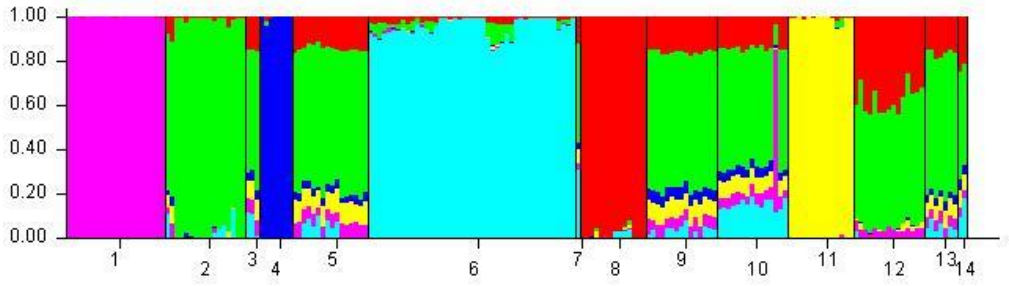


Figure 12 Genome-wide admixture proportions at K=5 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 5 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs.

(a)



(b)

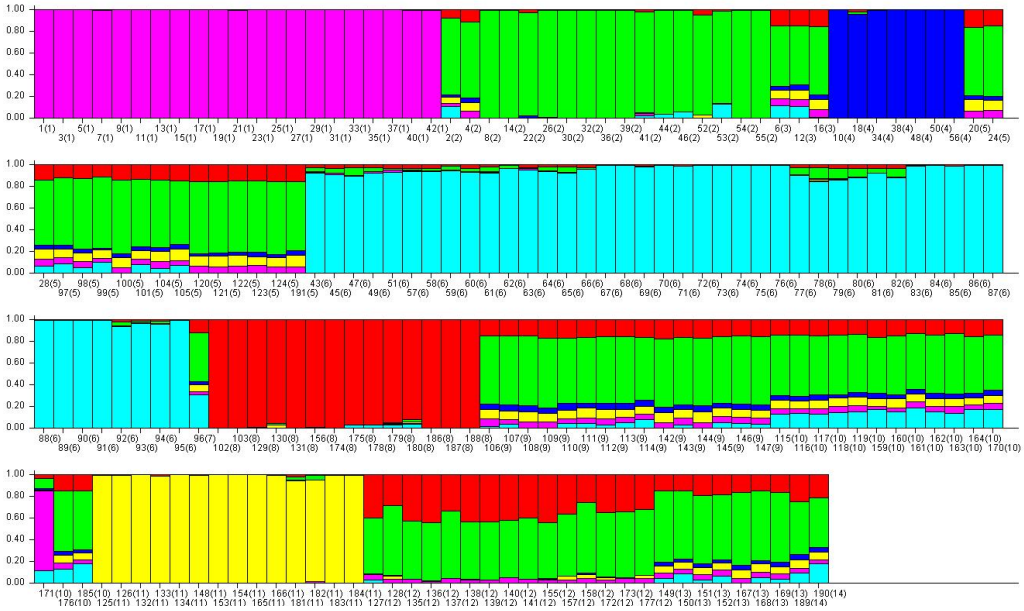
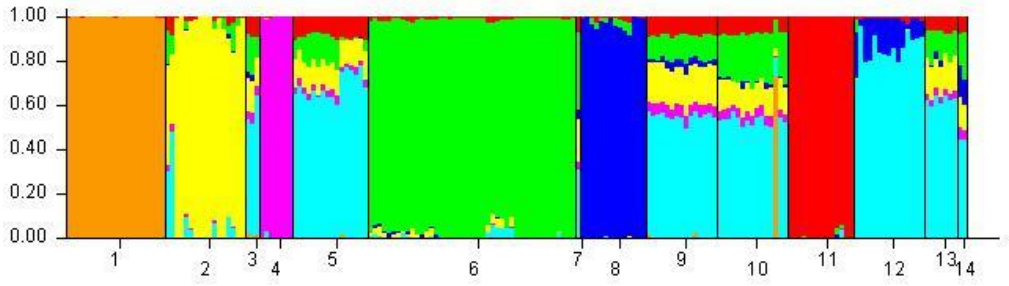


Figure 13 Genome-wide admixture proportions at K=6 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 6 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs.

(a)



(b)

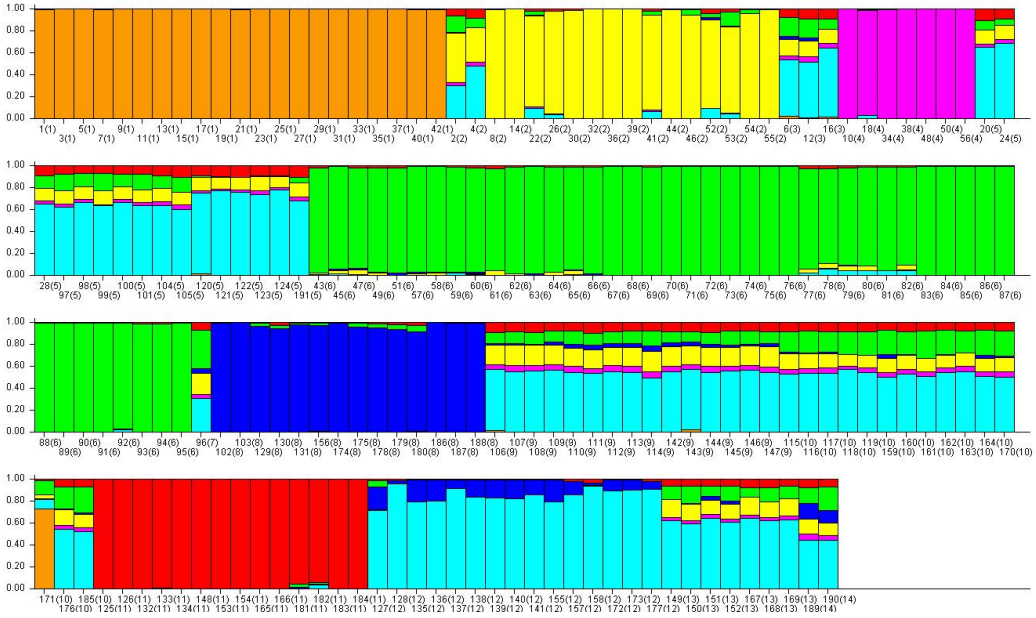
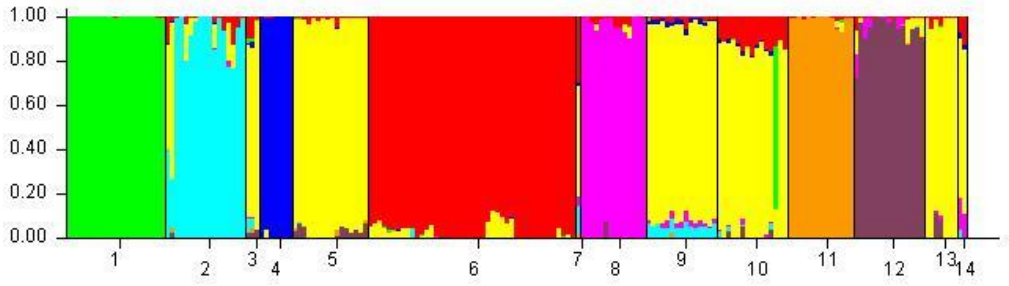


Figure 14 Genome-wide admixture proportions at K=7 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 7 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs.

(a)



(b)

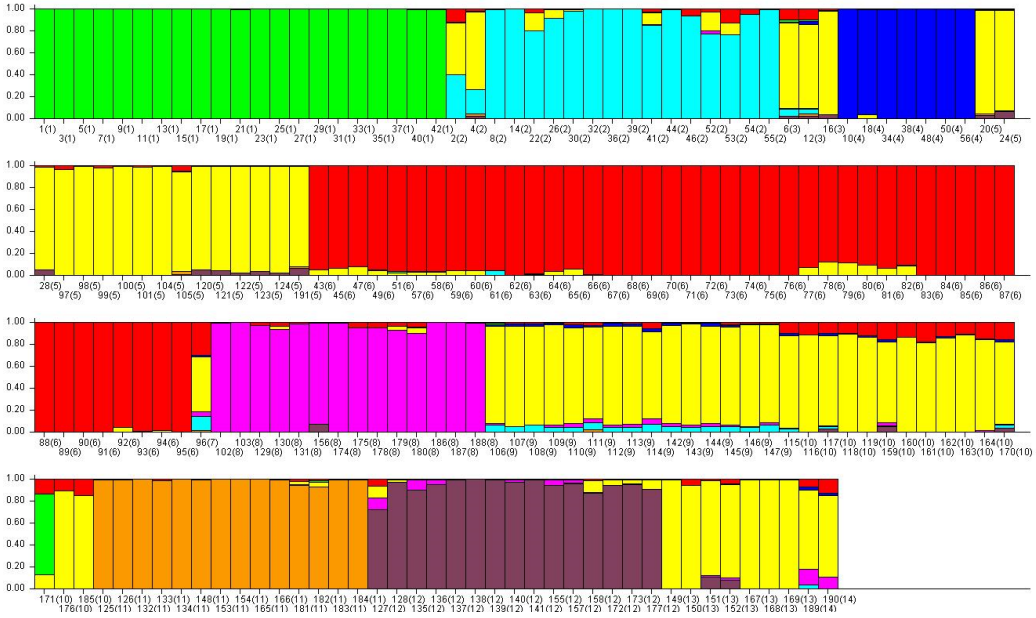
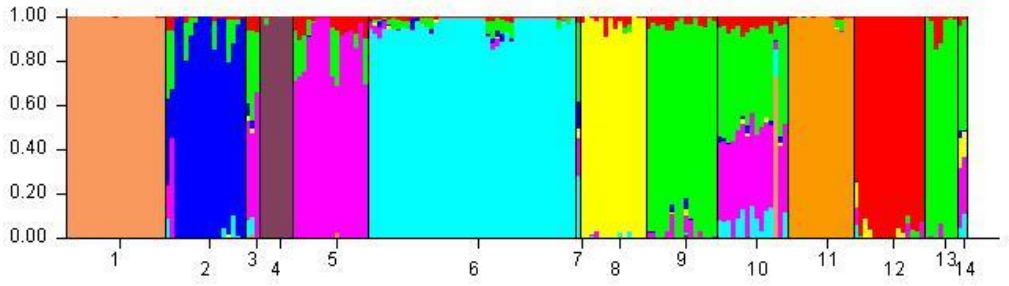


Figure 15 Genome-wide admixture proportions at K=8 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 8 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs.

(a)



(b)

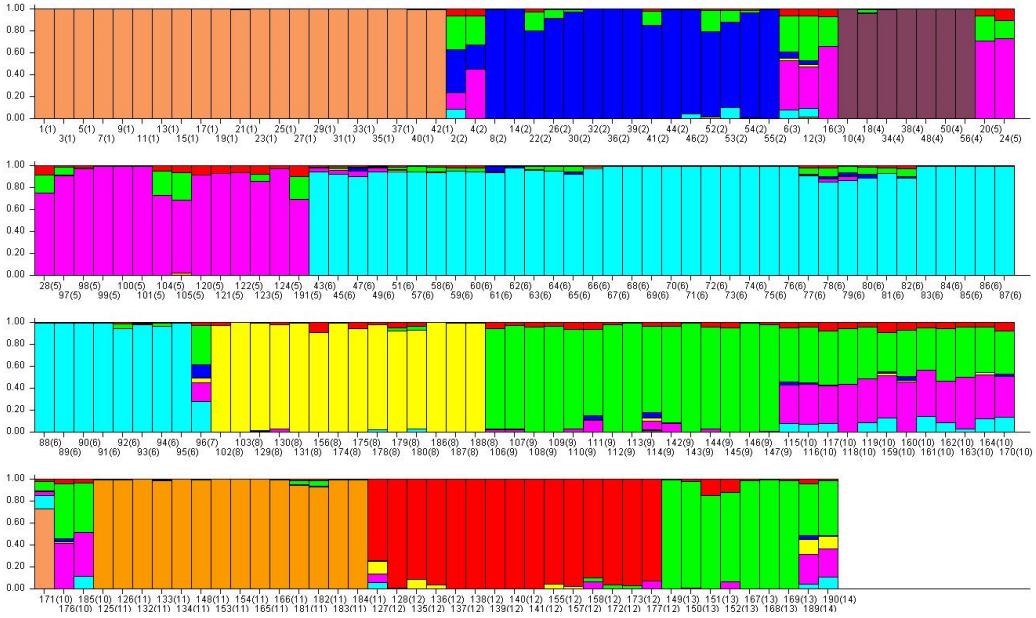
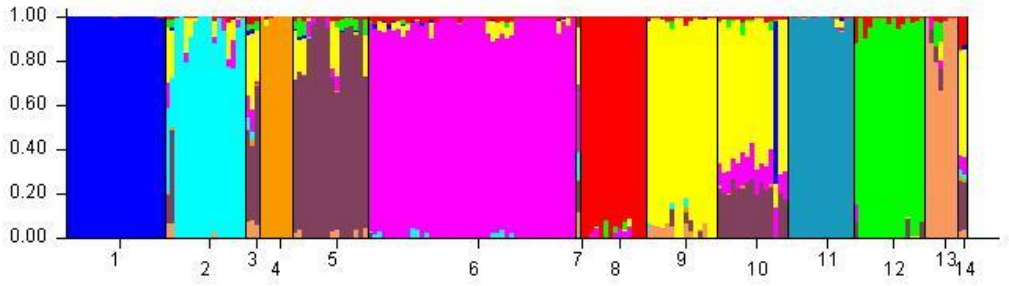


Figure 16 Genome-wide admixture proportions at K=9 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 9 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs.

(a)



(b)

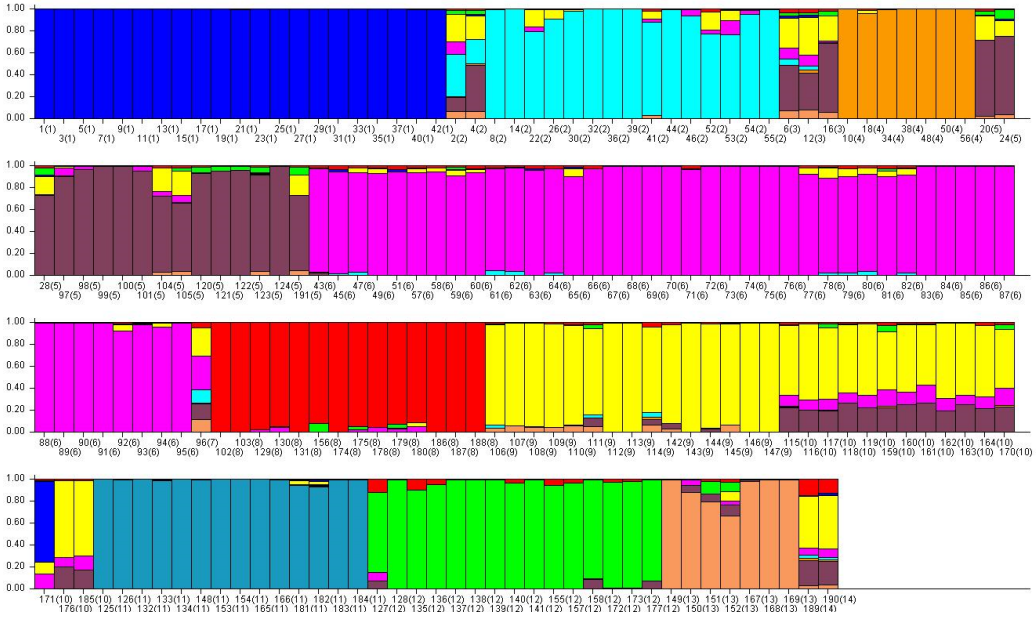
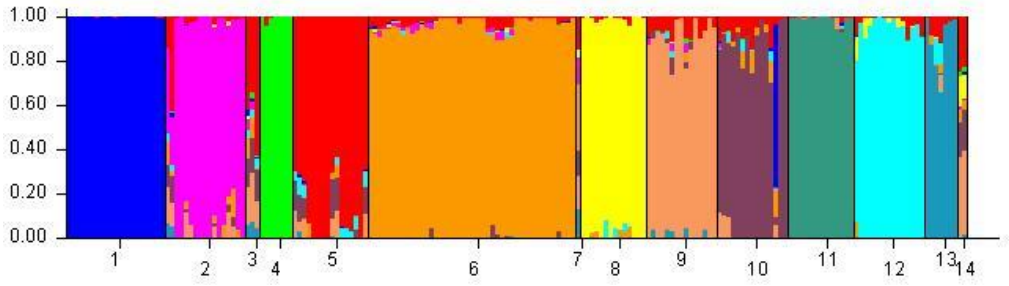


Figure 17 Genome-wide admixture proportions at K=10 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 10 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs.

(a)



(b)

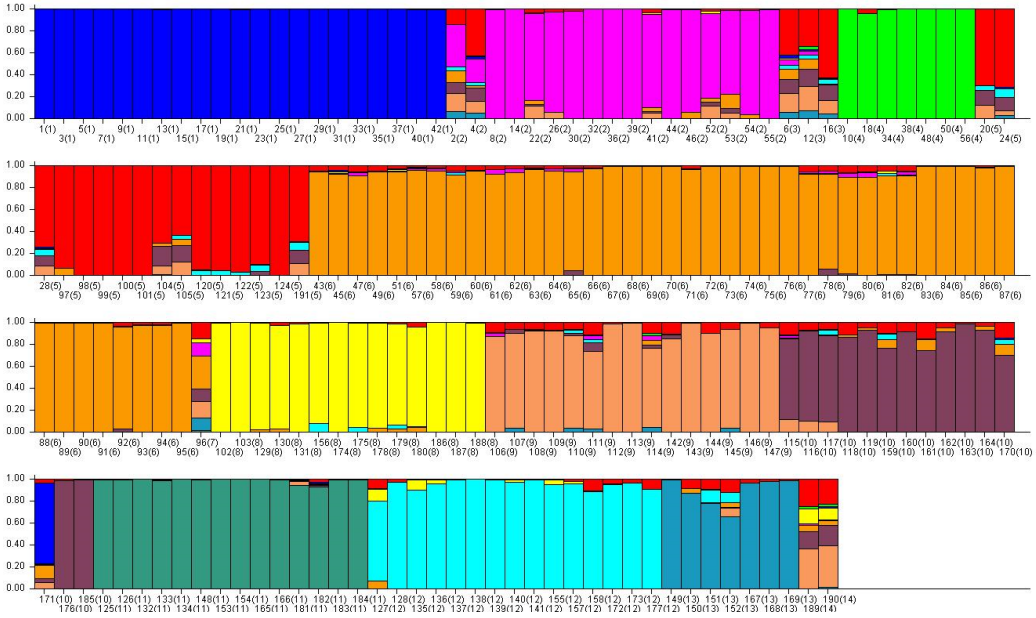
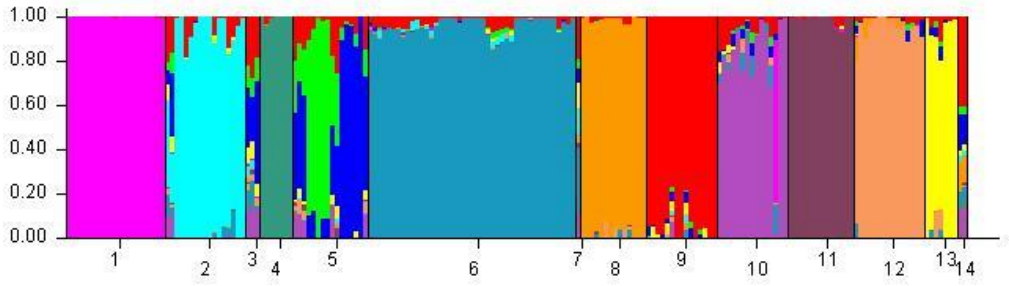


Figure 18 Genome-wide admixture proportions at K=11 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 11 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs.

(a)



(b)

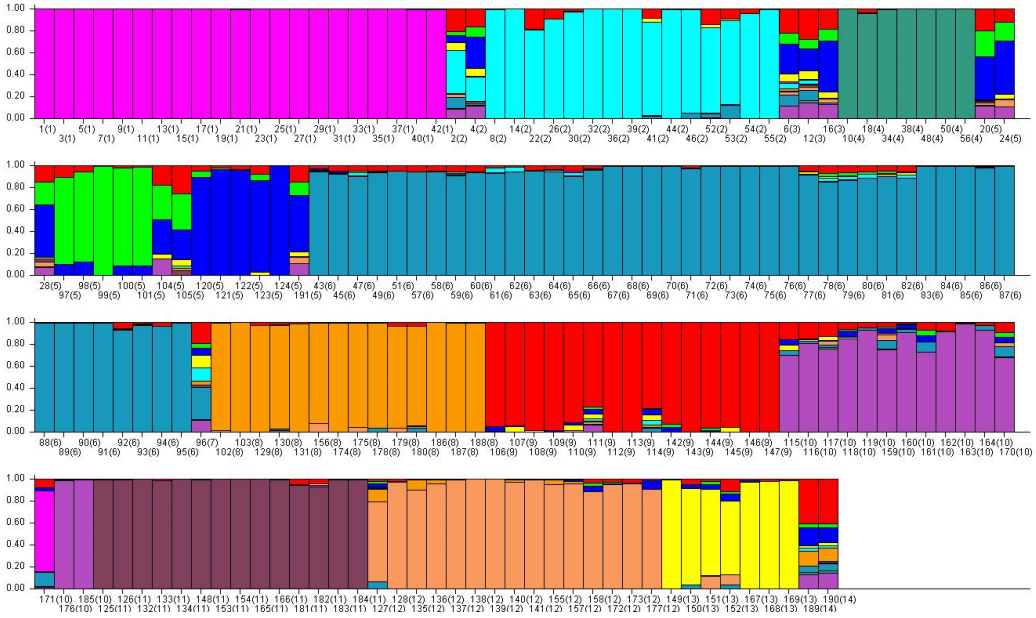
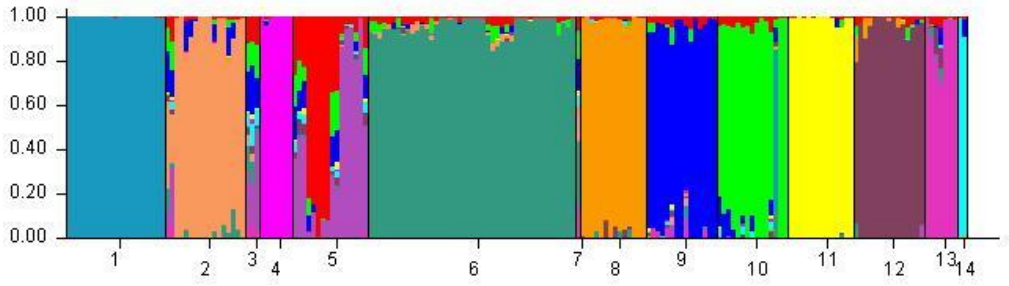


Figure 19 Genome-wide admixture proportions at K=12 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 12 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammelrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs.

(a)



(b)

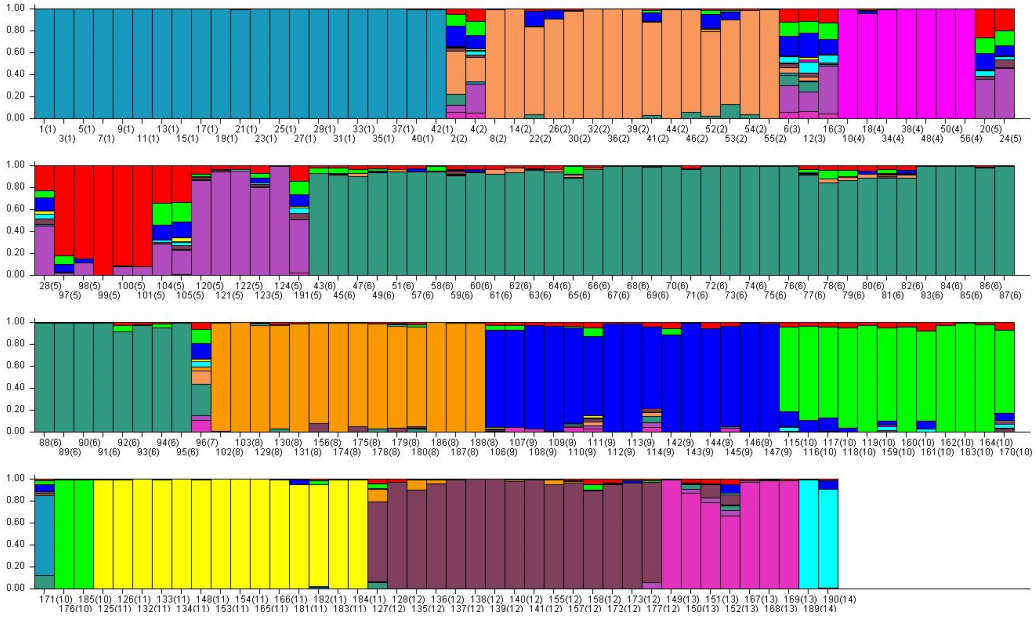
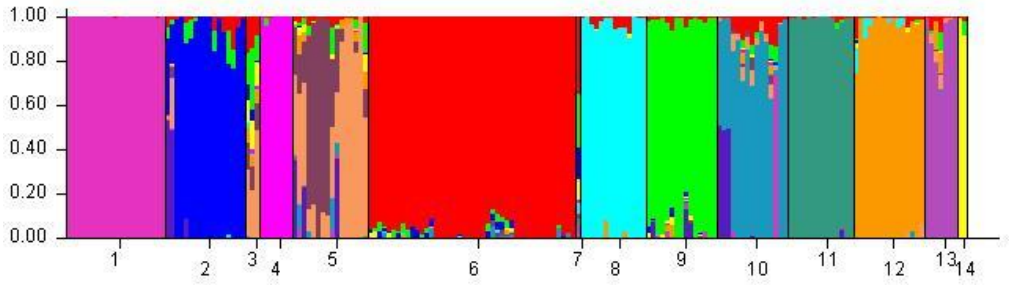


Figure 20 Genome-wide admixture proportions at K=13 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 13 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs.

(a)



(b)

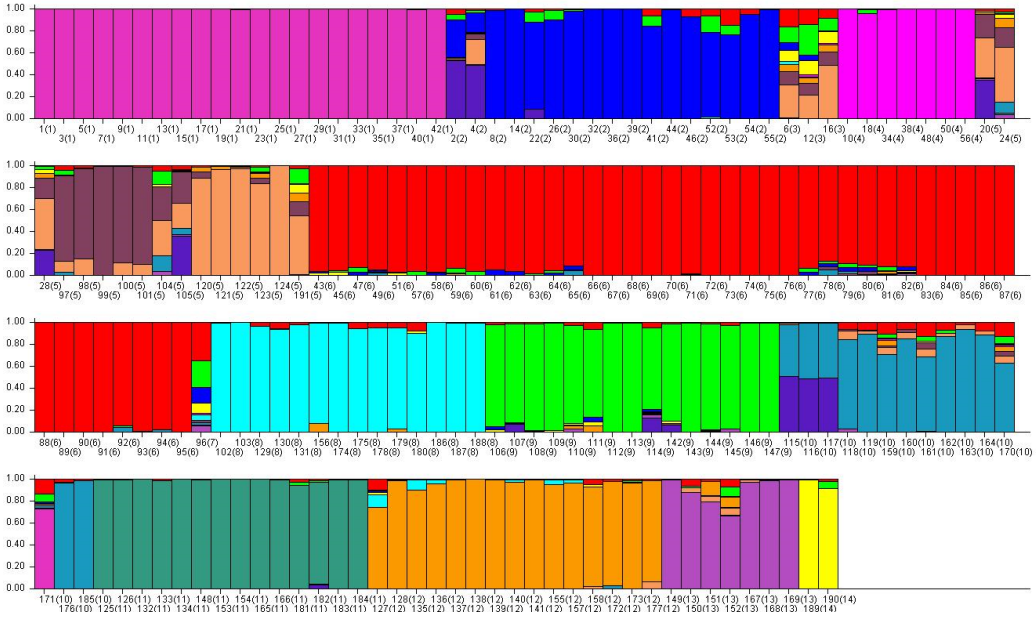


Figure 21 Genome-wide admixture proportions at K=14 among the 191 sheep of the 14 native Swedish sheep breeds. The colours in the plot refer to the 14 assumed genetic populations. The numbers denoted in the x-axis show: 1 to 14, different populations (a); 1 to 191, individuals within each population (b). Here 14 Swedish sheep breeds used as different populations: 1, Roslag; 2, Brännö; 3, Gammerlrya; 4, Tabacktorp; 5, Rya; 6, Gotland; 7, Gute; 8, Åsen; 9, Finull; 10, Värmland; 11, Svärdsjö; 12, Helsinge; 13, Gestrike; 14, Fjällnäs.

Appendix 2 – ROH Plots per Chromosome

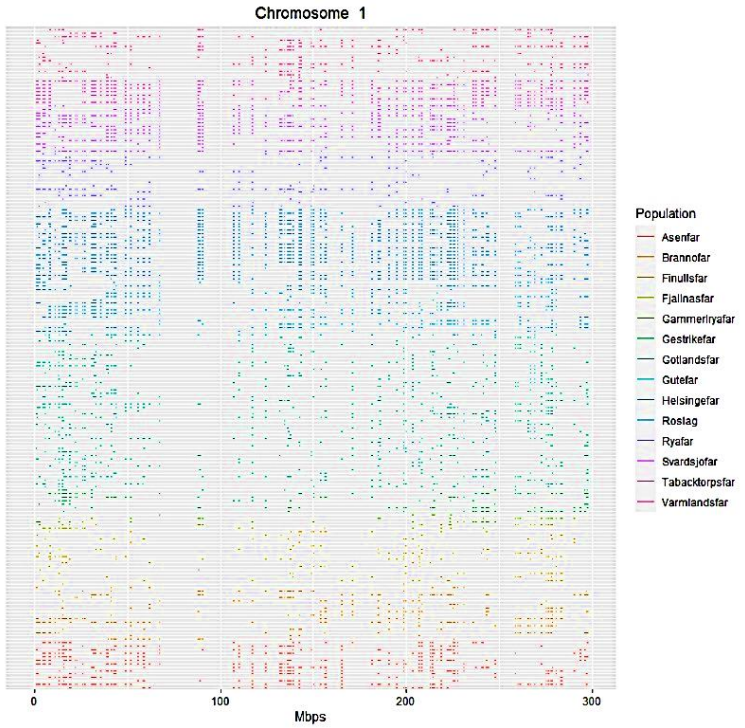


Figure 22 ROH segments on Chromosome 1.

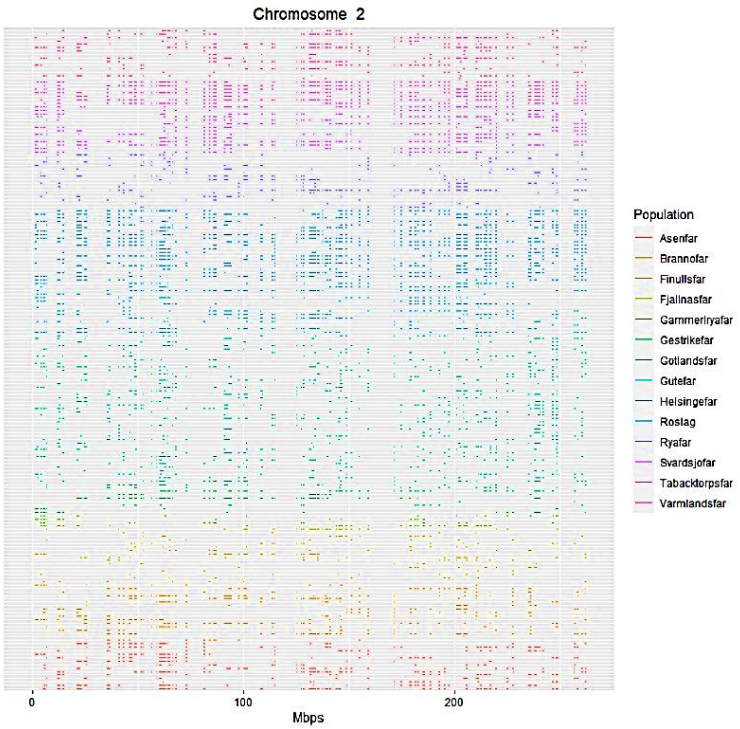


Figure 23 ROH segments on Chromosome 2.

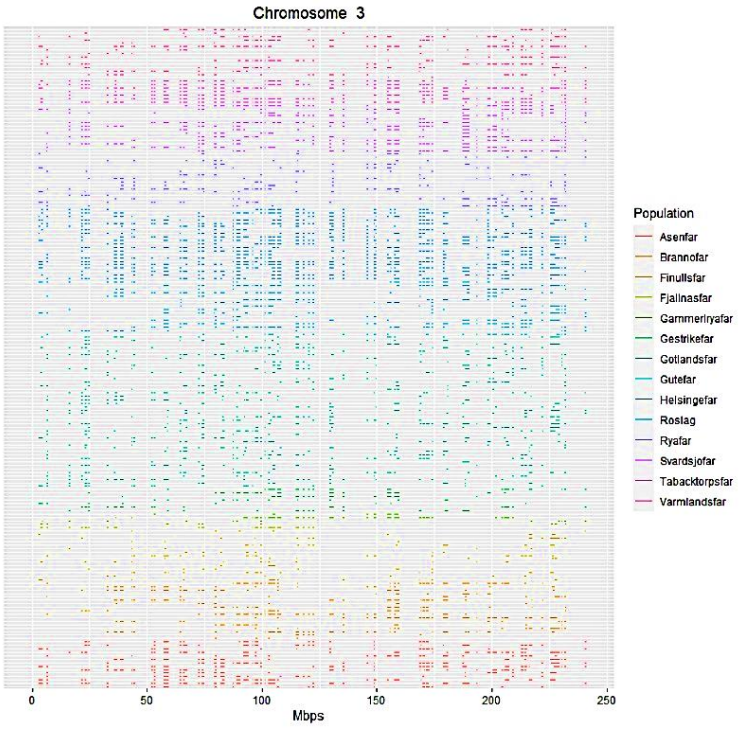


Figure 24 ROH segments on Chromosome 3.

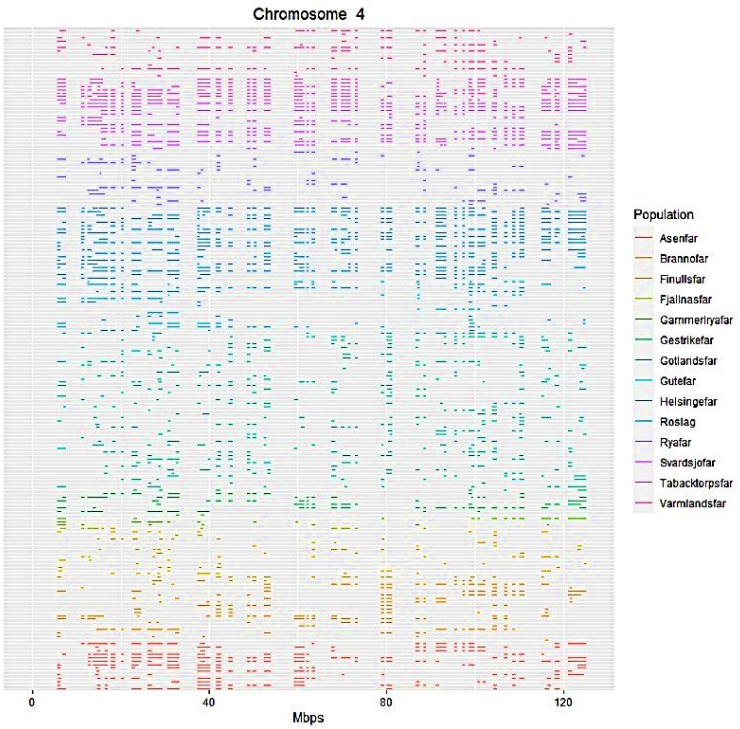


Figure 25 ROH segments on Chromosome 4.

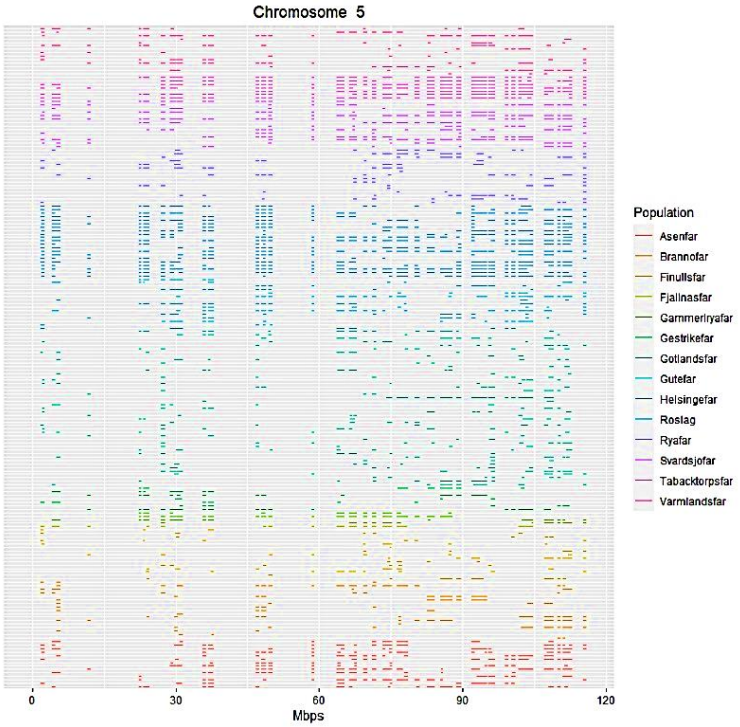


Figure 26 ROH segments on Chromosome 5.

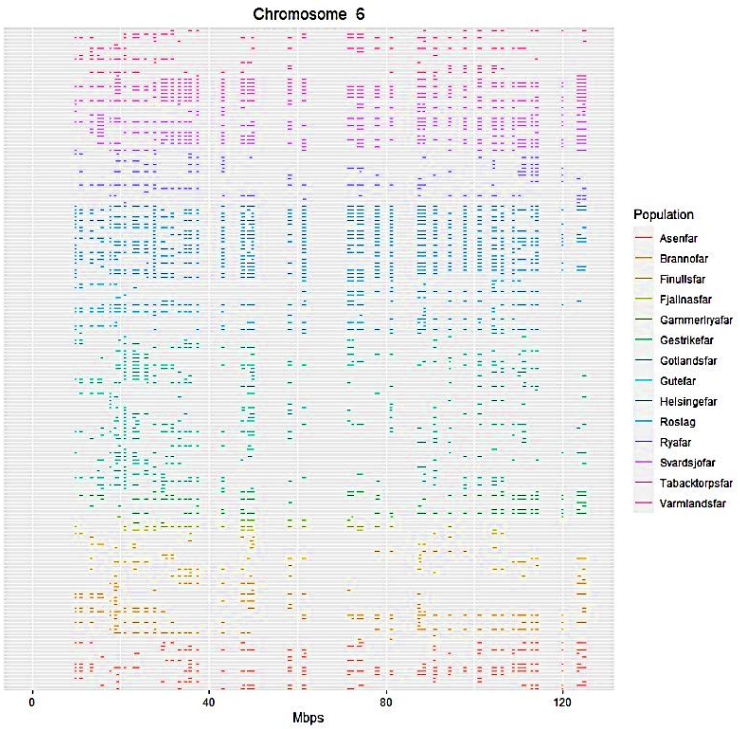


Figure 27 ROH segments on Chromosome 6.

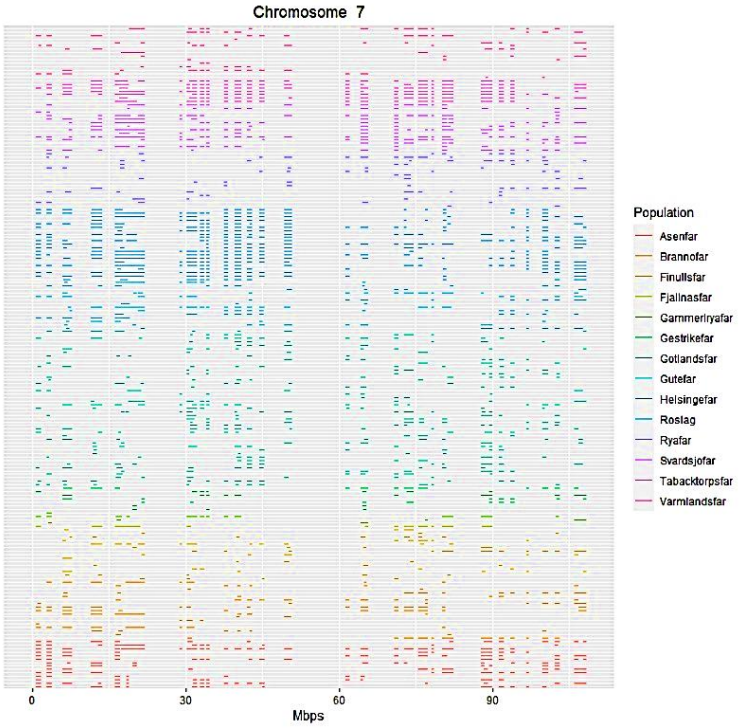


Figure 28 ROH segments on Chromosome 7.

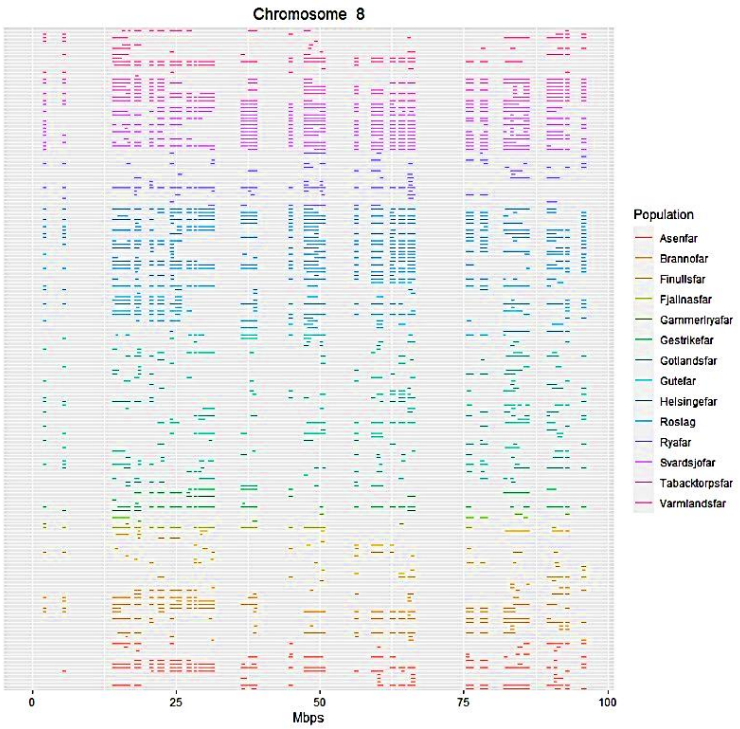


Figure 29 ROH segments on Chromosome 8.

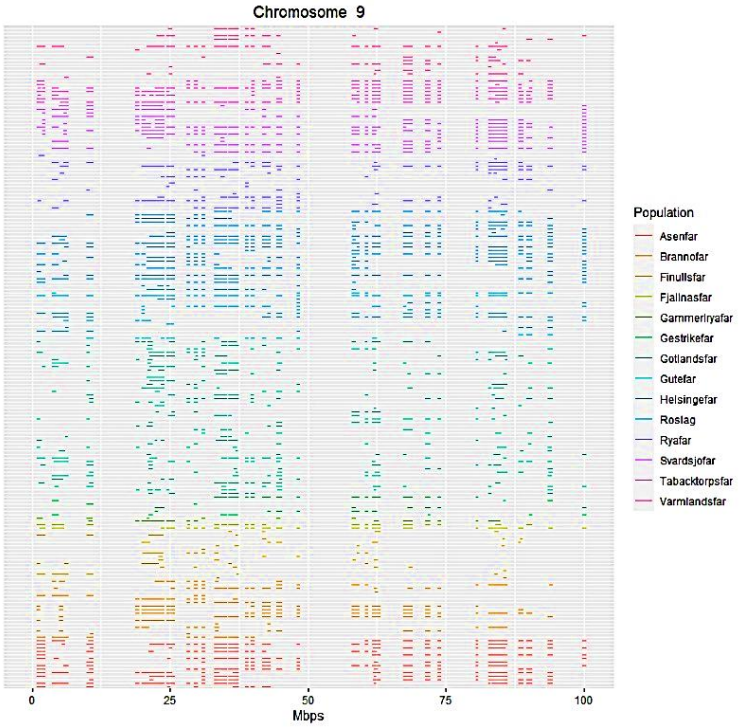


Figure 30 ROH segments on Chromosome 9.

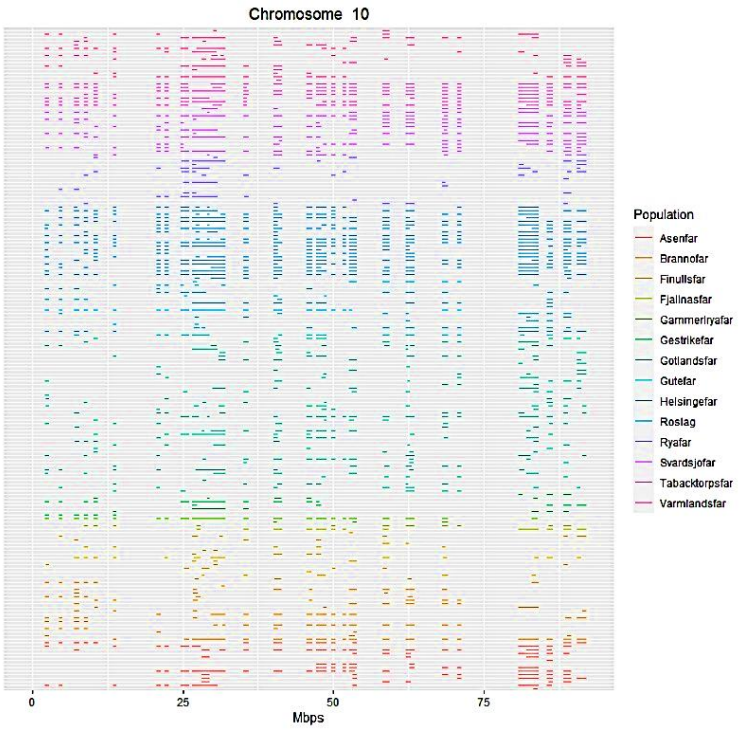


Figure 31 ROH segments on Chromosome 10.

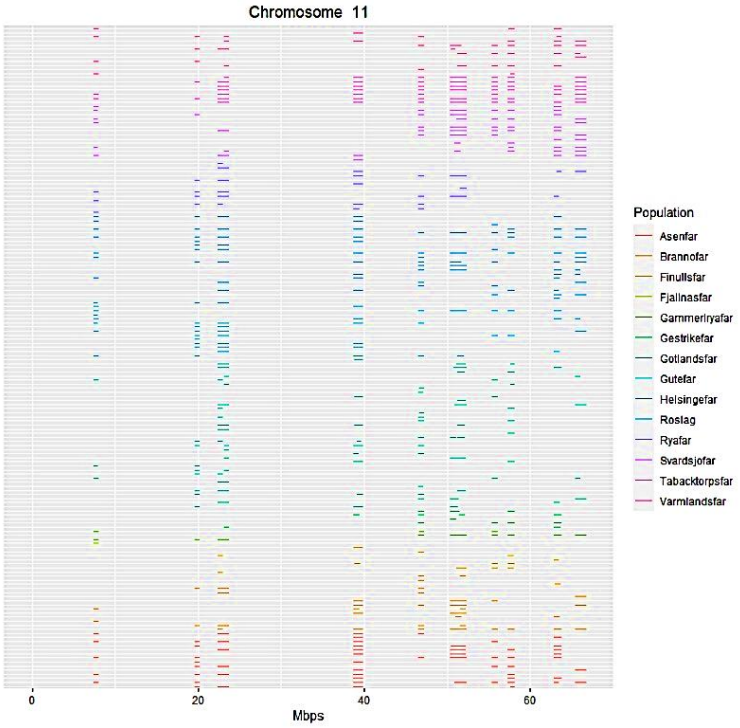


Figure 32 ROH segments on Chromosome 11.

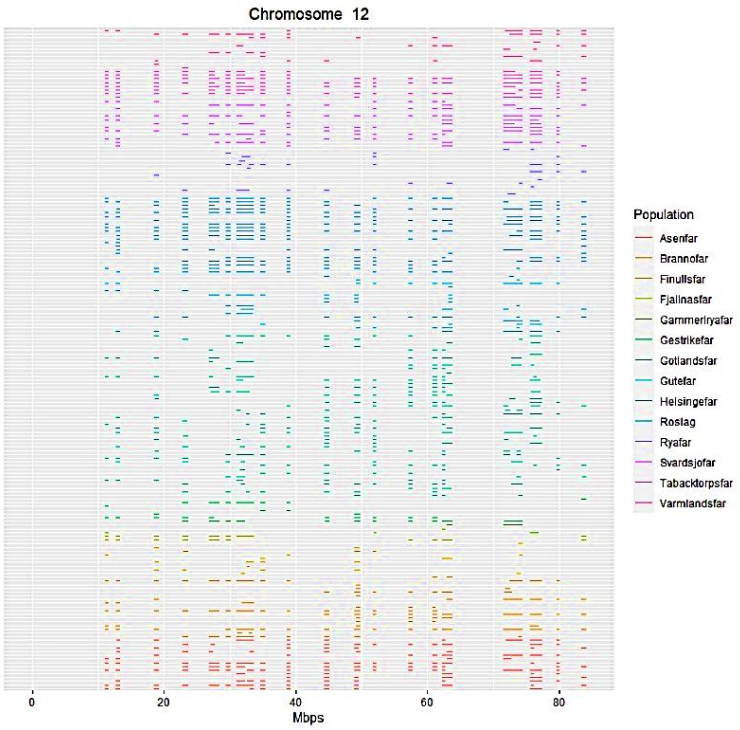


Figure 33 ROH segments on Chromosome 12.

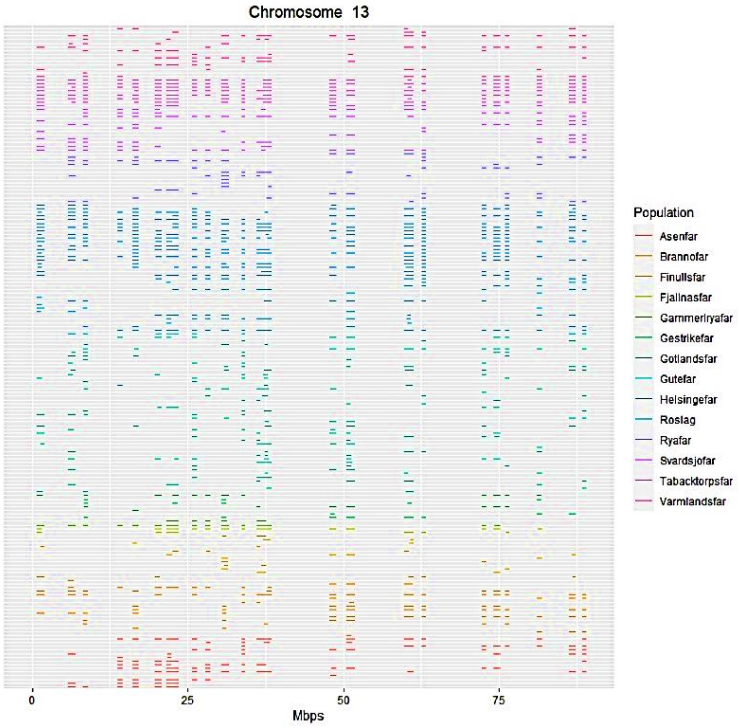


Figure 34 ROH segments on Chromosome 13.

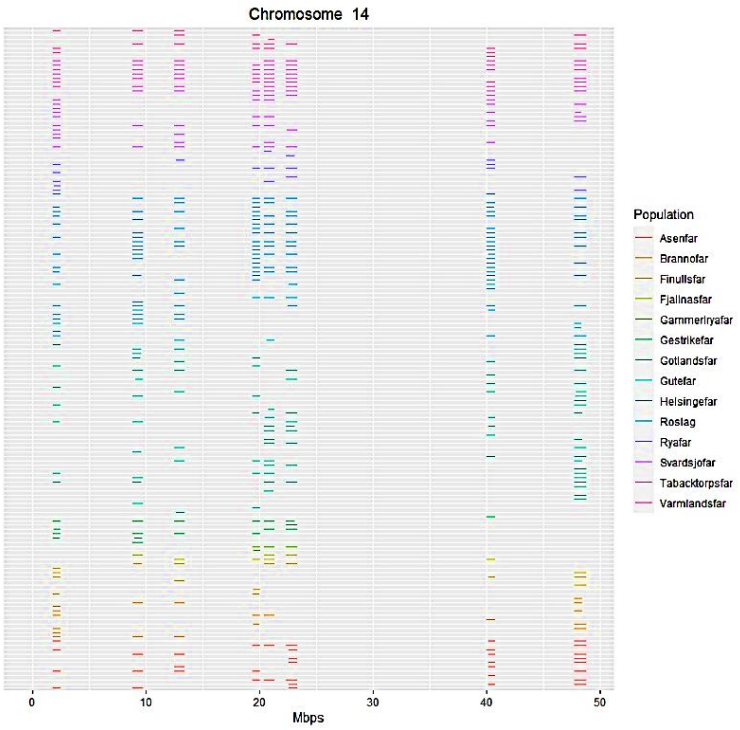


Figure 35 ROH segments on Chromosome 14.

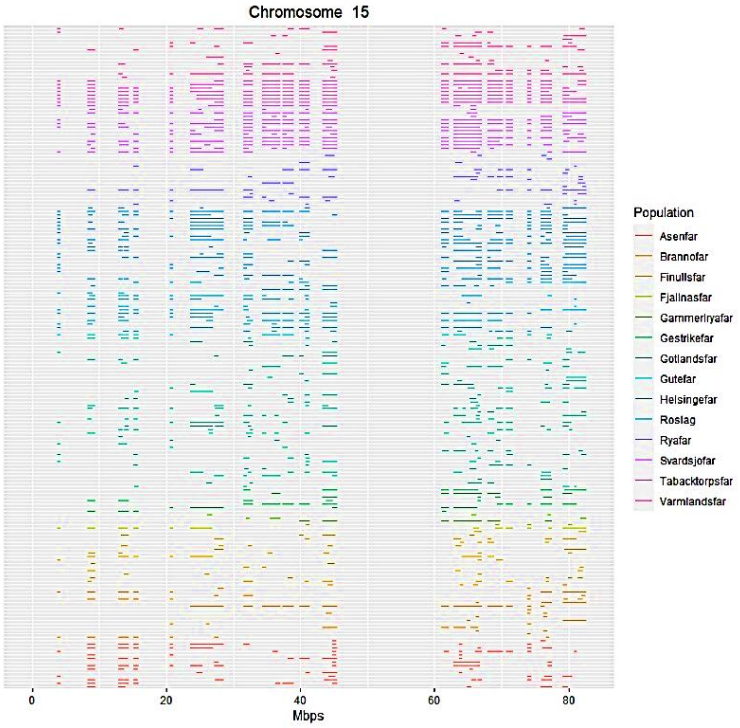


Figure 36 ROH segments on Chromosome 15.

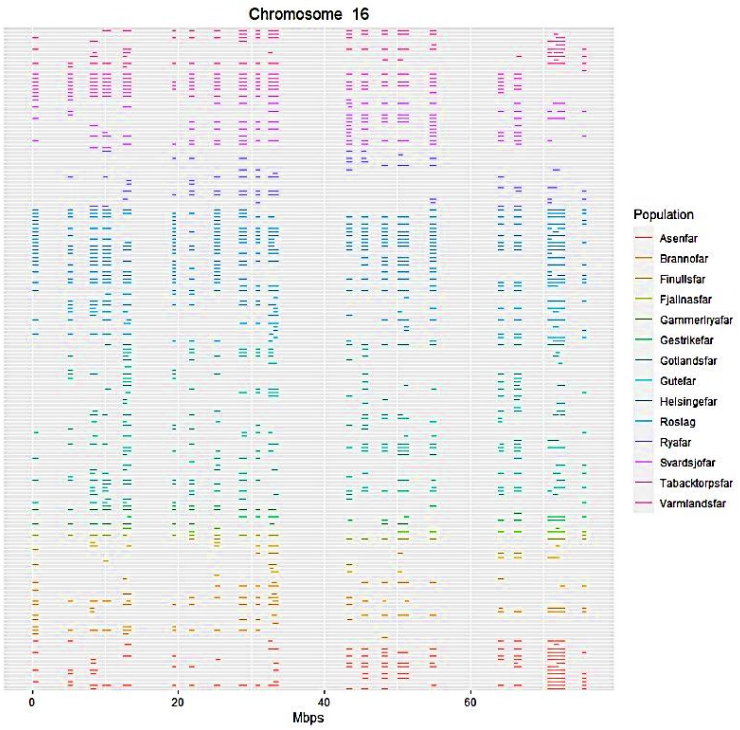


Figure 37 ROH segments on Chromosome 16.

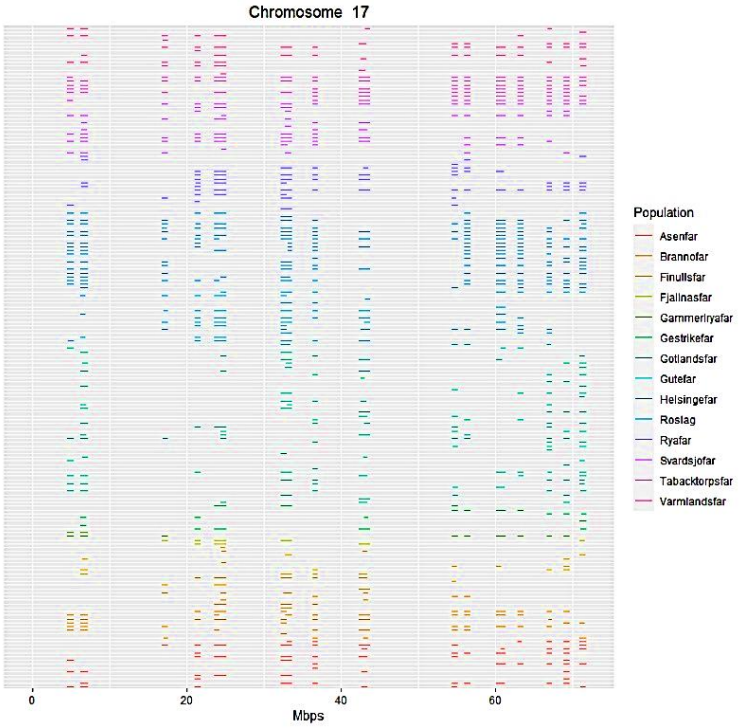


Figure 38 ROH segments on Chromosome 17.

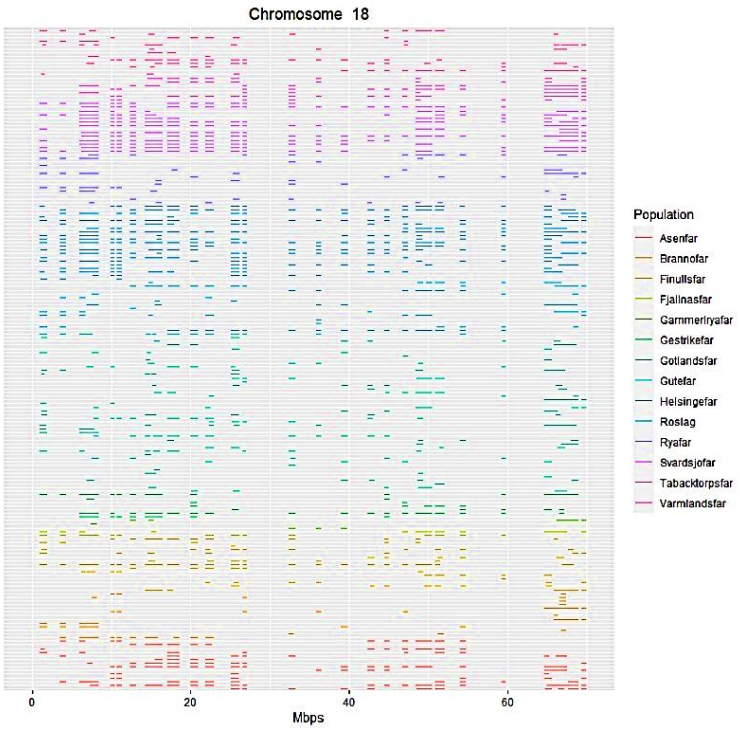


Figure 39 ROH segments on Chromosome 18.

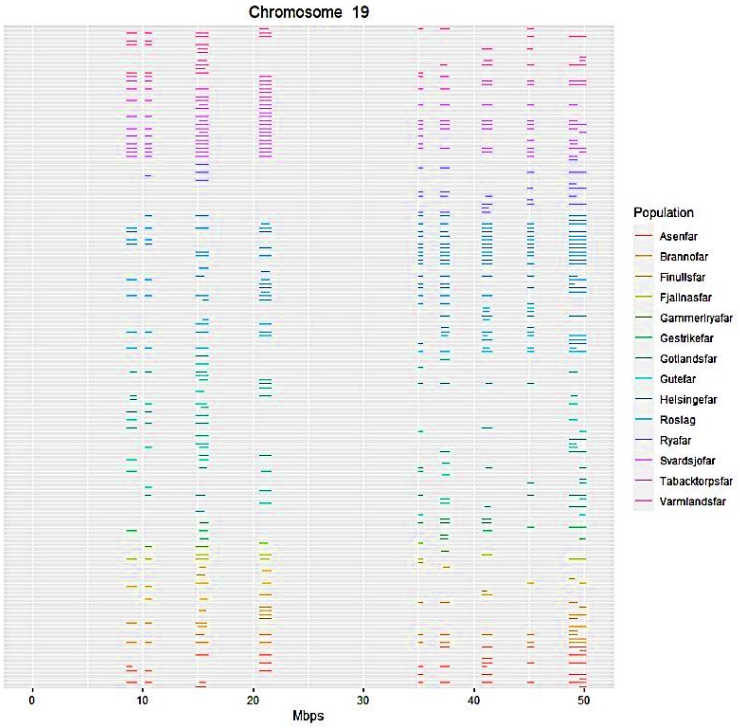


Figure 40 ROH segments on Chromosome 19.

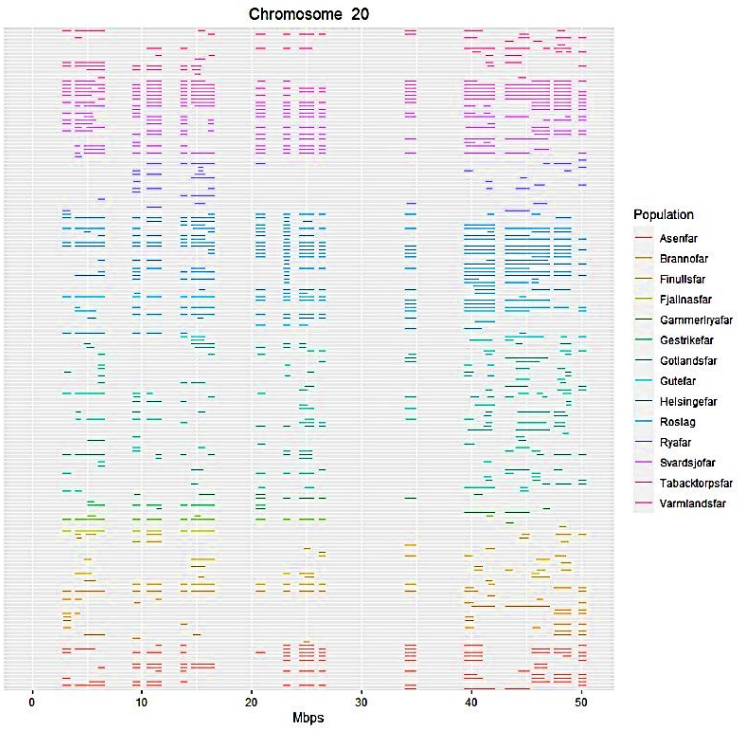


Figure 41 ROH segments on Chromosome 20.

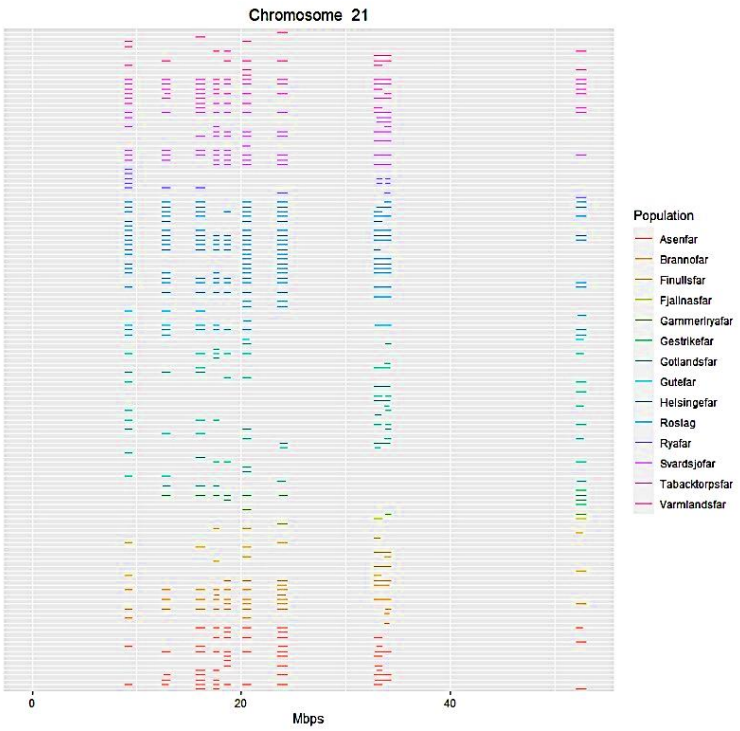


Figure 42 ROH segments on Chromosome 21.

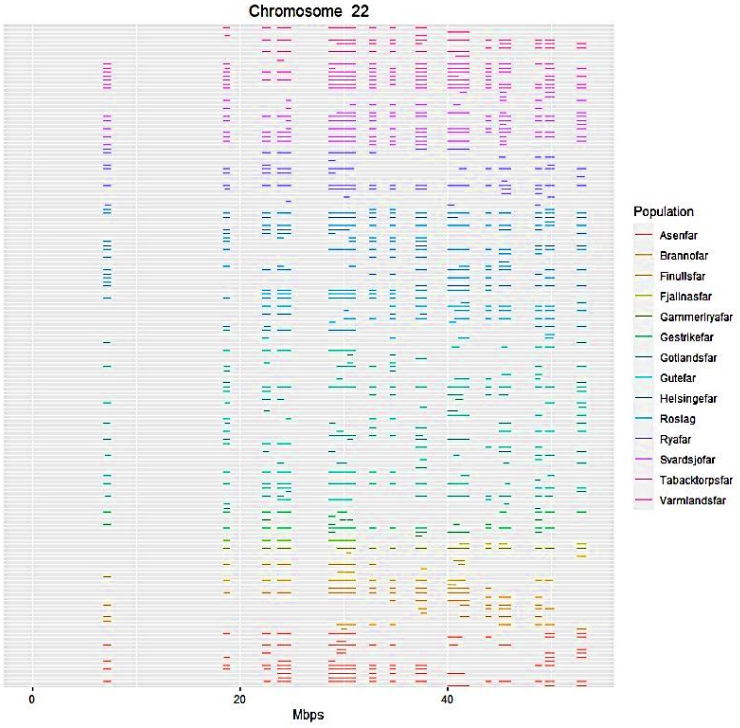


Figure 43 ROH segments on Chromosome 22.

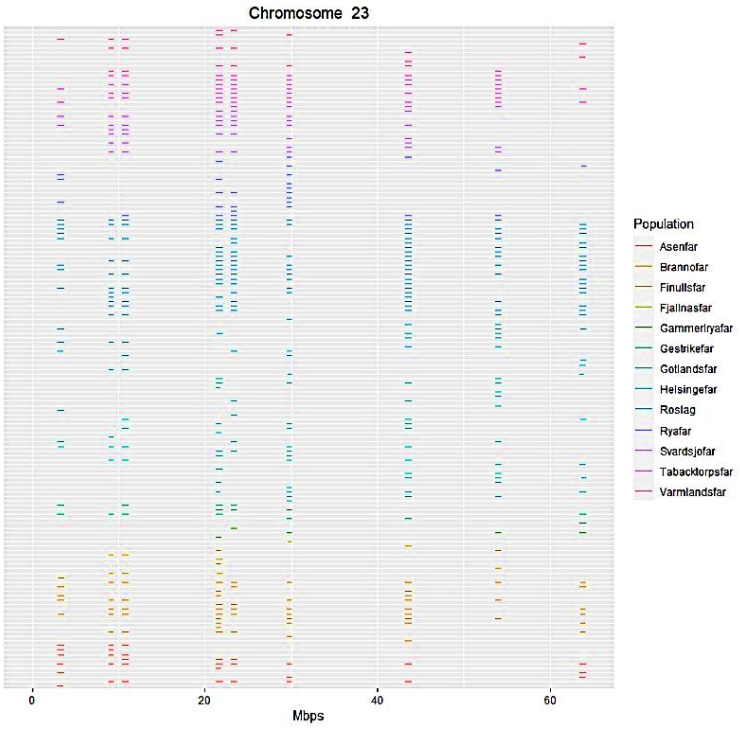


Figure 44 ROH segments on Chromosome 23.

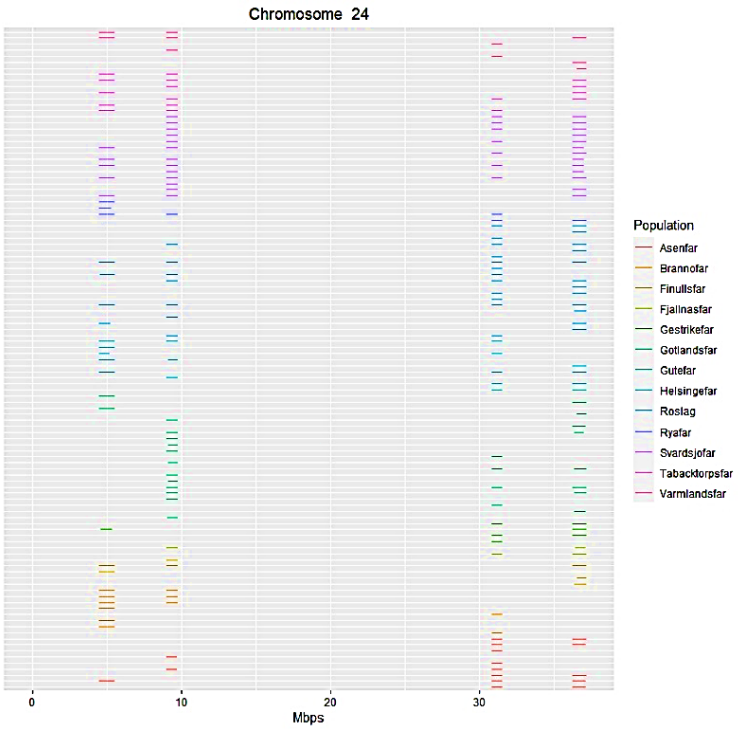


Figure 45 ROH segments on Chromosome 24.

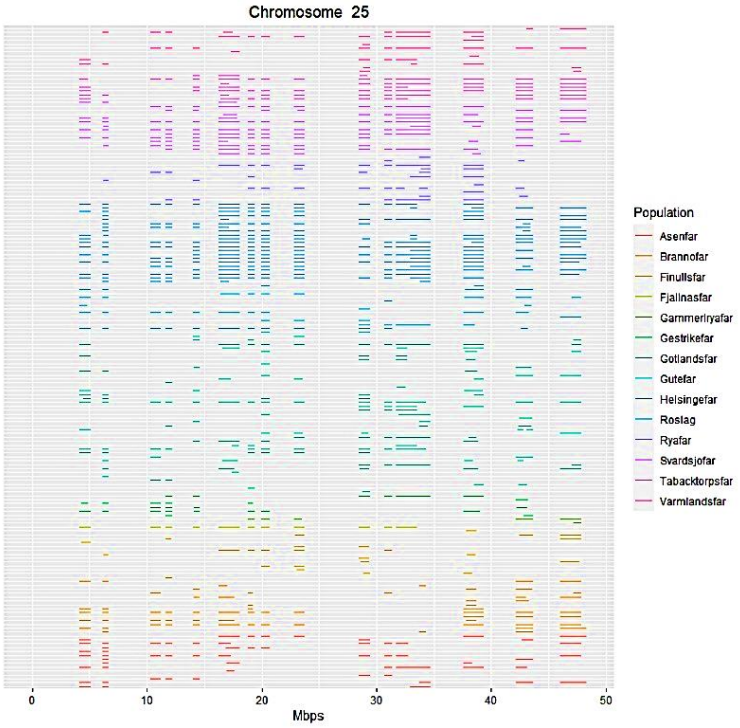


Figure 46 ROH segments on Chromosome 25.

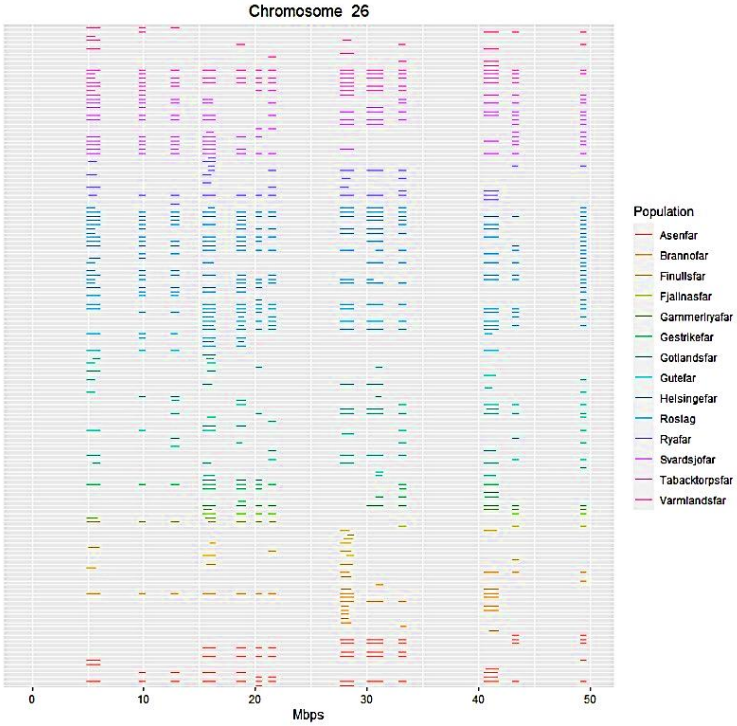


Figure 47 ROH segments on Chromosome 26.

Appendix 3 – Genetic Variants in MC1R

Table 9 Summary of genetic variants in the Melanocortin 1 receptor (MC1R) gene and associated phenotypic variability in five native Swedish sheep breeds

Sample	Breed	Colour	MC1R_C.-31 G>A	MC1R_C.218 T>A
OAR521	Finull	Brown	GA	TA
OAR703	Finull	Brown	AA	TA
OAR487	Finull	Black	AA	TA
OAR645	Gestrike	Black	AA	TT
OAR646	Gestrike	Grey	GA	TT
OAR636	Svärdsjö	Black	AA	TT
OAR637	Svärdsjö	White	GG	TT
OAR638	Svärdsjö	White	AA	TT
OAR684	Svärdsjö	White	GA	TT
OAR685	Svärdsjö	White	GG	TT
OAR686	Svärdsjö	Black	GA	TT
OAR1001	Gammelrya	white	AA	TT
OAR1003	Gammelrya	Grey	GG	TT
OAR376	Rya	Black	GA	TT
OAR999	Rya	White	AA	TT
OAR1005	Rya	Black	GA	TT
OAR1006	Rya	White	AA	TT
OAR380	Rya	Black	AA	TT
OAR381	Rya	Black	AA	TT
OAR384	Rya	Black	GA	TT
OAR444	Rya	Brown	GA	TT
OAR447	Rya	Brown	AA	TT
OAR627	Rya	White	AA	TT
OAR634	Rya	White	GG	TT
Previous Data (Rochus et al., 2019)				
OAR244	Gotland	Dark grey	AA	TT
OAR336	Gute	Grey	GA	TT
OAR491	Finull	White	GG	TT

Publishing and archiving

Approved students' theses at SLU are published electronically. As a student, you have the copyright to your own work and need to approve the electronic publishing. If you check the box for **YES**, the full text (pdf file) and metadata will be visible and searchable online. If you check the box for **NO**, only the metadata and the abstract will be visible and searchable online. Nevertheless, when the document is uploaded it will still be archived as a digital file. If you are more than one author, the checked box will be applied to all authors. You will find a link to SLU's publishing agreement here:

- https://libanswers.slu.se/en/faq/228318.

☒ YES, I/we hereby give permission to publish the present thesis in accordance with the SLU agreement regarding the transfer of the right to publish a work.

☐ NO, I/we do not give permission to publish the present work. The work will still be archived and its metadata and abstract will be visible and searchable.