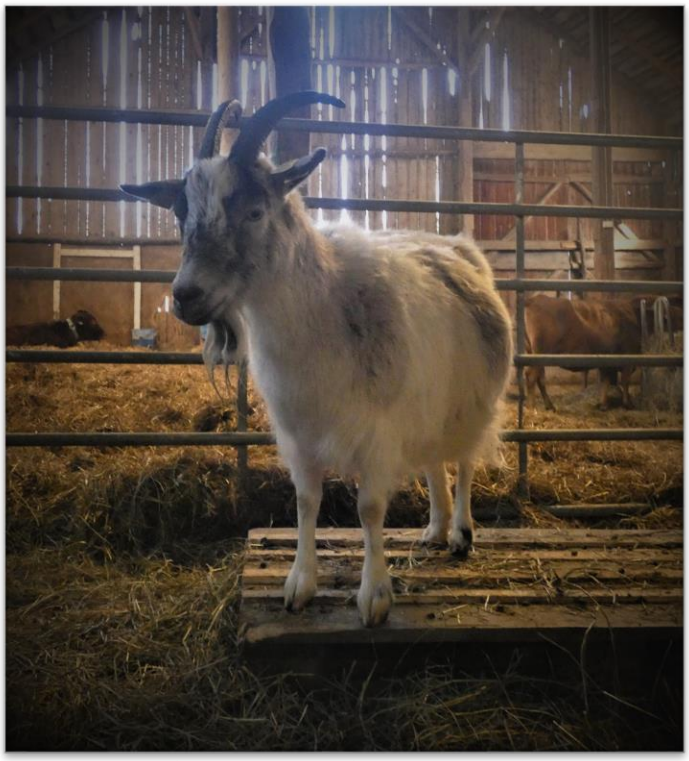
Genomic diversity analysis in the Swedish Landrace Goat

Bernadett Hegedűs



Independent project • 30 credits
Swedish University of Agricultural Sciences, SLU
Department of Animal Breeding and Genetics
European Master in Animal Breeding and Genetics
Uppsala 2022

Genomic diversity analysis in the Swedish Landrace Goat.

Analys av genetisk variation i svensk lantrasget

Bernadett Hegedűs

Supervisor:	**Anna Maria Johansson, SLU, Animal Breeding and Genetics**
Assistant supervisor:	Piter Bijma, Wageningen University and Research, Animal Breeding and Genomics
Examiner:	Martin Johnsson, SLU, Animal Breeding and Genetics

Credits:	30 credits
Level:	A2E
Course title:	**Independent Project in Animal Science**
Course code:	EX0870
Programme/education:	**European Master in Animal Breeding and Genetics**
Course coordinating dept:	**Animal Breeding and Genetics**
Place of publication:	**Uppsala**
Year of publication:	**2022**
Cover picture:	own picture at Gränby 4H, 2022
Keywords:	Swedish Landrace Goat, genetic diversity, inbreeding, runs of homozygosity, population structure

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

Abstract

There are four goat breeds present in Sweden. However, no genetic studies were dedicated to investigating the genetic diversity in these breeds. This thesis therefore aims to describe the population structure and the level of inbreeding in the goat breed with the largest population number, the Swedish Landrace goat. Forty-eight (48) samples from eight farms were genotyped with a medium density SNP chip. To study the population structure, a principal coordinate analysis and a Structure analysis were conducted. The level of inbreeding was investigated with three measures; observed heterozygosity, F_{ROH} and approximated coancestry. The results show that there is some structuring in the population and this structure is not solely due to the geographic location of these farms. The inbreeding level of the farms is comparable to other European non-island goat populations. A potential pattern of selection was identified on Chromosome 6 with ROH in the region of the Casein genes. Further research is needed to calculate the effective population size and the rate of inbreeding in this breed.

Table of contents

List of tables **6**

List of figures **7**

Abbreviations **11**

1. Introduction **12**

1.1 Study animal: The Swedish Landrace Goat 12

1.2 Diversity measures 13

1.2.1 Population structure 14

1.2.2 Inbreeding 16

1.3 Aim of the thesis 17

2. Material and Methods **18**

2.1 Data 18

2.2 Methods 19

2.2.1 Quality control and LD pruning 19

2.2.2 Observed Heterozygosity 20

2.2.3 PCoA 20

2.2.4 Approximated coancestry calculated from genomic relationship matrix 20

2.2.5 Runs of homozygosity 21

2.2.6 Structure 22

2.2.7 Driving distances between farms 23

2.2.8 R scripts 23

3. Results **24**

3.1 Farm locations 24

3.2 Population structure 26

3.2.1 Principal Coordinate Analysis 26

3.2.2 Admixture analysis with Structure 27

3.3 Inbreeding 29

3.3.1 Observed Heterozygosity 29

3.3.2 Inbreeding coefficient based on ROH 30

3.3.3 Coancestry 31

3.4 ROH on Chromosome 6 31

4. Discussion **34**

4.1 Discussion of results 34
4.2 Future risks and threats to the Swedish Landrace Goat 36
4.3 Studies for the future 36

5. Conclusion 38

6. References 39

Popular science summary 43

Appendix 1 - Simulation input for Structure 44

Appendix 2 - Structure Plots 45

Appendix 3 - ROH plots per Chromosome 48

Appendix 4 – Data Management Plan 63

List of tables

Table 1 Information on the number of samples from each herd with the corresponding animal IDs. 18

Table 2 The parameters of the different quality controls. The first five column names refer to the commands in Plink. The last column states the number of SNPs left after the quality control. 19

Table 3 List of parameters for the ROH analysis with the package detectRUNS. 21

Table 4 Calculated driving distance in km between the herds. The colours correspond to the colours used later for the farms. Herd 4 and 5 were given the colour of Herd 4 as both are located in the same town. 25

List of figures

Figure 1 Locations of farms on the map of Sweden. The colours correspond to the colours used later for the farms. Herds 4 and 5 were given the colour of Herd 4 as both are located in the same town. The coordinate system used for this plot was WGS84. 25

Figure 2 Results of principal coordinate analysis on the 48 genotyped individuals. 48111 SNPs were used for the analysis presented here. 27

Figure 3 Results of Structure analysis with K=4. The proportion of the 4 assumed ancestral population in each individual is shown for all the 48 genotyped individuals. The colours in the plot refer to the 4 assumed ancestral populations and should not be confused with the colours used to indicate herds in other plots. 28

Figure 4 Observed heterozygosities for herds of the Swedish Landrace goat (Herds 1 to 8) and for nine European goat breeds. The abbreviations B_MG, B_Se and B_Sk stand for the Norwegian Milk goat, Norwegian coastal goat Selje and the Norwegian coastal goat Skorpa (data from Berg et al. (2020)). The abbreviations M_Bermeya, M_Florida and M_Palmera stand for Spanish goat breeds investigated by Manunza et al. (2016). The abbreviations N_ASP, N_BIO and N_SAM stand for the Italian goat breeds Dell'Aspromonte , Bionda dell'Adamello, and Maltese sampled in Sardinia, respectively (data from Nicoloso et al. (2015)).The number of SNPs left after the quality control in the mentioned papers was 45772, 51136 and 39257 for Berg et al. (2020), Nicoloso et al. (2015) and Manunza et al. (2016), respectively. Note that both B_sk and M_Palmera are island populations. 29

Figure 5 Inbreeding coefficient based on the ROH in each herd. The minimum length of a ROH was defined at 36 SNP, for further parameters see the methods section. The colours in the plot match the herd colours in previous plots. 30

Figure 6 Estimated mean coancestries for 7 herds. Note that the means were calculated after removing the diagonals, therefore there is no value for Herd 5 where there was only one sample available. "All" stands for the mean estimated coancestry of all the sampled individuals, including the individual from Herd 5. The colours in the plot match the herd colours in previous plots. 31

Figure 7 ROH on Chromosome 6. The black rectangle marks the region where 40 of the 48 genotyped individuals have a ROH. Figure 8 zooms in on this region. Individual 39 does not have a ROH on Chromosome 6. 32

Figure 8 ROH at 85.25-86.5 Mbps on Chromosome 6. Note that for illustration in this figure all individuals are present on the y axis, even if they do not have a ROH in this region..... 32

Figure 9 Heatmap of SNPs at the region of 85.75Mbps to 86.25Mbps on Chromosome 6. The colours indicate the different alleles, white stands for missing data. The SNP names are on the X axis. Note that there is not SNP for exon 12 of the CSN1S1 in this SNP chip. 33

Figure 10 Results of structure analysis with K=2. The proportion of the 2 assumed ancestral populations in each individual is shown for all the 48 genotyped individuals. 45

Figure 11 Results of structure analysis with K=3. The proportion of the 3 assumed ancestral populations in each individual is shown for all the 48 genotyped individuals. 45

Figure 12 Results of structure analysis with K=5. The proportion of the 5 assumed ancestral populations in each individual is shown for all the 48 genotyped individuals. 46

Figure 13 Results of structure analysis with K=6. The proportion of the 6 assumed ancestral populations in each individual is shown for all the 48 genotyped individuals. 46

Figure 14 Results of structure analysis with K=6. The proportion of the 6 assumed ancestral populations in each individual is shown for all the 48 genotyped individuals. 47

Figure 15 Results of structure analysis with K=6. The proportion of the 6 assumed ancestral populations in each individual is shown for all the 48 genotyped individuals. 47

Figure 16 ROH on Chromosome 1. The individuals that do not show up in the plot did not have a ROH on this chromosome..... 48

Figure 17 ROH on Chromosome 2. The individuals that do not show up in the plot did not have a ROH on this chromosome..... 48

Figure 18 ROH on Chromosome 3. The individuals that do not show up in the plot did not have a ROH on this chromosome..... 49

Figure 19 ROH on Chromosome 4. The individuals that do not show up in the plot did not have a ROH on this chromosome..... 49

Figure 20 ROH on Chromosome 5. The individuals that do not show up in the plot did not have a ROH on this chromosome.....50

Figure 21 ROH on Chromosome 6. The individuals that do not show up in the plot did not have a ROH on this chromosome.....50

Figure 22 ROH on Chromosome 7. The individuals that do not show up in the plot did not have a ROH on this chromosome.....51

Figure 23 ROH on Chromosome 8. The individuals that do not show up in the plot did not have a ROH on this chromosome.....51

Figure 24 ROH on Chromosome 9. The individuals that do not show up in the plot did not have a ROH on this chromosome.....52

Figure 25 ROH on Chromosome 10. The individuals that do not show up in the plot did not have a ROH on this chromosome.....52

Figure 26 ROH on Chromosome 11. The individuals that do not show up in the plot did not have a ROH on this chromosome.....53

Figure 27 ROH on Chromosome 12. The individuals that do not show up in the plot did not have a ROH on this chromosome.....53

Figure 28 ROH on Chromosome 13. The individuals that do not show up in the plot did not have a ROH on this chromosome.....54

Figure 29 ROH on Chromosome 14. The individuals that do not show up in the plot did not have a ROH on this chromosome.....54

Figure 30 ROH on Chromosome 15. The individuals that do not show up in the plot did not have a ROH on this chromosome.....55

Figure 31 ROH on Chromosome 16. The individuals that do not show up in the plot did not have a ROH on this chromosome.....55

Figure 32 ROH on Chromosome 7. The individuals that do not show up in the plot did not have a ROH on this chromosome.....56

Figure 33 ROH on Chromosome 18. The individuals that do not show up in the plot did not have a ROH on this chromosome.....56

Figure 34 ROH on Chromosome 19. The individuals that do not show up in the plot did not have a ROH on this chromosome.....57

Figure 35 ROH on Chromosome 20. The individuals that do not show up in the plot did not have a ROH on this chromosome.....57

Figure 36 ROH on Chromosome 21. The individuals that do not show up in the plot did not have a ROH on this chromosome.....58

Figure 37 ROH on Chromosome 22. The individuals that do not show up in the plot did not have a ROH on this chromosome.....58

Figure 38 ROH on Chromosome 23. The individuals that do not show up in the plot did not have a ROH on this chromosome.....59

Figure 39 ROH on Chromosome 24. The individuals that do not show up in the plot did not have a ROH on this chromosome.....59

Figure 40 ROH on Chromosome 25. The individuals that do not show up in the plot did not have a ROH on this chromosome.....60

Figure 41 ROH on Chromosome 26. The individuals that do not show up in the plot did not have a ROH on this chromosome.....60

Figure 42 ROH on Chromosome 27. The individuals that do not show up in the plot did not have a ROH on this chromosome.....61

Figure 43 ROH on Chromosome 28. The individuals that do not show up in the plot did not have a ROH on this chromosome.....61

Figure 44 ROH on Chromosome 29. The individuals that do not show up in the plot did not have a ROH on this chromosome.....62

Abbreviations

F_{ROH}	Inbreeding coefficient based on runs of homozygosity
LD	Linkage disequilibrium
MAF	Minor allele frequency
PCoA	Principal coordinate analysis
ROH	Runs of homozygosity
SNP	Single-nucleotide polymorphism

1. Introduction

This part of the thesis will describe the Swedish Landrace Goat and shed light on some of the diversity and inbreeding parameters that will be investigated. In the last chapter of the introduction the aim of the thesis and the research questions will be introduced.

1.1 Study animal: The Swedish Landrace Goat

There are four goat breeds in Sweden: the Göinge Goat, the Jämt Goat, the Lapp Goat and the Swedish Landrace Goat. All of these breeds are categorized as endangered by the FAO (FAO 2021). However, until now no single-nucleotide polymorphism (SNP) based genetic study has been conducted on these breeds. Therefore, this thesis aims to give some information on the genetic diversity present in the Swedish goat breed with the largest population number, the Swedish Landrace Goat.

The Swedish Landrace goat is a dairy breed that is used in Sweden for milk and cheese production. This goat breed has a high average milk yield which is about 700kg per goat per year. However, there is also quite some variation regarding milk yield in this breed with some animals producing 2000kg of milk per year (Svenska Getavelsförbundet 2021).

The Swedish Landrace Goat is a breed without a uniform phenotype. This breed is known for its different colours and patterns. There is also a variety when it comes to hair length as no selection has been conducted on this trait. Furthermore, this breed holds diversity regarding the presence, size and shape of horns (Svenska Getavelsförbundet 2021).

The Swedish organization responsible for goats is called Svenska Getavelsförbundet. However, this is a voluntary organisation, and the actual number of goats in Sweden is likely higher than the number of registered animals. The organization uses the Elitlamm software which was originally developed for sheep breeders. The lack of a software specifically for goats sometimes causes difficulty when handling and collecting data.

When it comes to the history of this breed, it is said to be closely related to the Norwegian and the Icelandic goat breeds. In the 1980s and 1990s semen from Norwegian goats was used to crossbreed with the Swedish breed to battle

inbreeding. According to the Svenska Getavelsförbundet the breed underwent a bottleneck; in the time from 1970 to 2003 the breed had an estimated population size of around 6000 individuals. The population size increased to 20,000 goats by the year 2018 (Svenska Getavelsförbundet 2021). There is no reliable information on the number of breeding individuals currently or the sex distribution among the breeding individuals (FAO 2021). Furthermore, there is no reliable information in the exact geographical distribution of this breed.

A recent Master thesis by Agnes Björk looked into the prevalence of a deletion in the casein alpha s1 (*CSN1S1*) gene in Swedish goat populations. The samples included all four breeds present in Sweden. The results of that thesis show that the deletion that contributes to both an increased milk yield and the reduced Casein content in the milk is frequent in Swedish populations. The estimated allele frequency of this deletion in the Swedish Landrace goat was 91.67% (44 out of 48 alleles). Furthermore, there were two other alleles found with a low frequency (allele counts three and one). The number of successfully sequenced individuals of Swedish Landrace Goat was 24. (Björk 2019).

There is a lack of studies concerning the Swedish Landrace goat (and Swedish goat breeds in general). The studies available mostly describe the milk quality, milk yield and milk compositions of this breed (Högberg 2011; Yurchenko et al. 2018).

1.2 Diversity measures

In population diversity studies there are three closely related measures that are usually assessed: drift, level and rate of inbreeding and effective population size. These measures shed light on the current status of the population and also on potential future risks. These phenomena will be introduced with the help of a well-known quantitative genetics textbook (Falconer & Mackay 1996).

Genetic drift is a phenomenon that has already been described for almost a hundred years and is linked to Sewall Wright. It describes the stochastic sampling of gametes from one generation to the next which leads to allele frequency differences between generations in a finite population. The allele frequencies are changed in the direction of more extreme allele frequencies. This also leads to an increased number of homozygotes, as the number of heterozygotes is the highest at intermediate allele frequencies. In absence of other events (such as migration, mutation, or selection) with time one of the alleles at a given locus will reach fixation in finite populations as a result of drift. The effective population size is closely related to the time it takes for an allele to be fixed (Falconer & Mackay 1996, 48-57).

Secondly when studying populations, it is informative to calculate the inbreeding coefficient or the rate of inbreeding. Inbreeding is the result of mating of related individuals and is unavoidable in finite populations. Inbreeding leads to an

increased number of homozygotes and to fewer heterozygotes. The inbreeding coefficient describes the probability that two alleles at any locus are identical by descent. It can also be seen as the proportion of loci in an individual that are identical by decent. The inbreeding coefficient in the current generation is the sum of current and past inbreeding. The rate of inbreeding is the increase of inbreeding from one generation to the next and is measured in proportion to the distance left to reach complete inbreeding (Falconer & Mackay 1996, 57-63). The FAO recommends keeping the rate of inbreeding under 0.5-1% to manage risks regarding the viability of a population (Tave 1999).

The effective population size is closely related to drift and inbreeding and describes the number of individuals in an idealized population that would experience the same amount of drift and inbreeding as the population in question (Falconer & Mackay 1996, 65-67). The effective population size (N_e) can serve as a measure of drift that the population is under. There are also molecular methods to estimate the historical effective population size with the help of SNP data like the tool SNeP (Barbato et al. 2015) used in a study about Swedish local sheep breeds by Rochus et al. (2020).

In the following part I would like to briefly introduce the methods that I used to assess the genetic diversity of the Swedish Landrace goat. The aim of this part is to make the Material and Methods part of the report easier to understand. The used methods are clustered in two chapters, first, the methods used to assess the structure of the population and second, the methods to assess the inbreeding level in the population will be discussed.

1.2.1 Population structure

To assess the structure of a population and to visualize past admixture of ancient populations the Principal Coordinate Analysis (PCoA) and Admixture analysis like the software Structure offers, are often used. In the following two paragraphs these two methods will be explained.

The Principal Coordinate Analysis is a method that helps to reduce multidimensionality of data and thereby make it easier to visualise and interpret. This is especially useful when we are dealing with thousands of SNPs in a dataset. The input of this analysis is a matrix of similarities or dissimilarities, such a distance matrix expressed as genomic proportions in this study. The output is a list of principal coordinates with their explained variance. The principal coordinates are ordered in a way that the first one explains the most variance in the data, the second principal coordinate explains most of the variance left and so on. The first two to three principal coordinates are often used to visualize data. This method is frequently used in genetic studies as it gives a main picture of the population structure. (Mohammadi & Prasanna 2003). Many studies that compare populations start with a PCoA to get an overview of the data to decide on further analyses and

to spot clearly admixed individuals (Johansson & Nelson 2015; Nicoloso et al. 2015; Upadhyay et al. 2019; Ganteil et al. 2021; Signer-Hasler et al. 2022). The PCoA is also often referred to as the metric multidimensional scaling (MDS) analysis (Zuur et al. 2007). One of the advantages of the PCoA over Structure (see next paragraph) is that it does not try to assign individuals to a certain number of discrete populations. The PCoA analysis is also considerably faster (Patterson et al. 2006).

Structure is a software that assigns individuals in a given sample to assumed populations based on allele frequency differences in the assumed populations (Pritchard et al. 2000; Falush et al. 2003, 2007; Hubisz et al. 2009). This assignment of individuals to different assumed populations allows geneticists to visualize population structure and study demographical events like migration and hybridization. The process of assigning the individuals to assumed populations is a long one with many (>10,000) iterations. In order to gain reasonable results, the user must specify a large enough burn-in length and enough Markov Chain Monte Carlo (MCMC) iterations. However, before one can start the analysis there is quite a number of parameter settings to navigate. As a result there are many resources devoted to helping researchers in navigating the wide range of parameter combinations (Pritchard et al. 2010; Porras Hurtado et al. 2013; Wang 2017). The first thing one needs to decide with regard to the analysis parameters is which admixture model to use. In case there is reason to assume that the sampled individuals are from different populations then the *no admixture* model should be used. In case admixture among the populations is assumed then the *admixture* model is more useful. Another parameter that needs the user's consideration is whether to use correlated allele frequencies. The correlated allele frequency model assumes that the allele frequencies in the populations are likely to be similar, whereas in the uncorrelated model each population has a randomly drawn allele frequency from a given distribution. Furthermore, to aid the analysis the user can chose to use prior population information. This means that the model first starts to group individuals according to, for example, sampling locations. This method is especially useful when the population structure is weak. This parameter is used to help the analysis but will not force structure on the data set if there is no structure present. The list of parameters presented here is just a sample of the parameters that can be used. For an extensive list please refer to the Structure documentation (Pritchard et al. 2010). It is also important to mention that the choice of parameters should be given more attention in special cases, for example when the sampling is unbalanced or there is a large number of populations (K>20) investigated (Wang 2017). Lastly, one should take into account that Structure cannot handle SNPs with linkage disequilibrium so the data filtering before the analysis need to take this into account (Pritchard et al. 2010).

1.2.2 Inbreeding

In this thesis three measures related to inbreeding are used: observed heterozygosity, approximated coancestry and the inbreeding coefficient based on the runs of homozygosity (ROH).

The inbreeding coefficient quantifies the difference between expected (so with random mating) and observed heterozygosity. As the observed fraction of heterozygous positions is more straight-forward, when one wants to compare results to previous studies, this measure was chosen. Furthermore, the observed number of heterozygous loci also serves as an indicator of drift, since alleles especially in smaller populations get fixed and therefore, will not show up in the observed heterozygous count. The comparison of the own results with other studies is possible because all of them used the same SNP chip. Three papers studying European goat breeds were chosen as a reference (Nicoloso et al. 2015; Manunza et al. 2016; Berg et al. 2020).

Coancestry is the relationship by descent between two individuals. It equals the inbreeding coefficient of their progeny (Falconer & Mackay 1996, 85-88). In this thesis, the approximated coancestry was used to investigate the relatedness among the individuals within the herd. This was done to see how related the individuals within each herd are compared to individuals of other herds. The calculation, which can be found in section 2.2.4, compares the similarity of the states of the loci (whether they are homozygous or heterozygous) in different individuals and averages these differences of similarities across all the SNPs.

Lastly the inbreeding coefficient based on the runs of homozygosity was calculated. This method can better differentiate between two different types of homozygotes; when alleles are identical by descent and when they are identical by state. This is done by identifying stretches of homozygous loci. This approach is based on the idea that longer stretches of homozygous SNPs are more likely to be inherited from a common ancestor than individual homozygous loci. Therefore, this method looks for the true autozygous positions. To calculate the inbreeding coefficient based on ROH the length of ROH segments is divided by the total length of the genome (Ceballos et al. 2018). See the formula below:

$$F_{ROH} = \frac{sum\ of\ ROH}{genome\ length}$$

The method to find the stretches of ROH needed to calculate the inbreeding coefficient relies on a sliding window that scans through the SNPs in search of longer stretches. The parameters used determine whether a set of homozygous loci will be regarded as a ROH. Among the most important parameters for this analysis are the size of the scanning window, the minimum number of homozygous loci needed for a ROH, the minimum length in bps needed for a ROH, minimum SNP density in the scanning region and the maximum possible gap between two homozygous SNPs in order to be considered in the same ROH. The scanning

window threshold is a parameter that quantifies the number of outer SNPs that can be included in a ROH. There is no clear consensus in the literature on what the best values for these parameters are in the ROH analysis (Meyermans et al. 2020).

1.3 Aim of the thesis

The FAO classifies the Swedish landrace goat breed as endangered. In order to preserve the breed semen from 8 bucks is stored as part of the cryopreservation programme. No female genetic material is stored (FAO 2021). However, the current genetic diversity has not been investigated and studies like the ones in the Swedish sheep breeds (Rochus & Johansson 2017; Rochus et al. 2020; Ghoreishifar et al. 2021) are needed in order to make educated decisions on the conservation of this breed. This is especially needed as the Svenska Getavelförbundet suggests selecting against the highly prevalent deletion in the *CSN1S1* gene. Strong selection for the absence of this deletion puts the population at risk of inbreeding. Thanks to the availability of a medium density SNP chip for goats (Tosser-Klopp et al. 2014) the genomic density of this livestock species can be characterised like other more popular species.

Therefore, the aim of this thesis is to study the genetic diversity in the Swedish Landrace Goat. Firstly, it will be investigated whether there is a clear structure in this breed, relative to the herds and geographical locations. This question will be answered with the help of an admixture analysis and a principal coordinate analysis. Secondly, the inbreeding level will be investigated. For this the observed heterozygosity of The Swedish Landrace will be compared to other European breeds. Furthermore, the relatedness among herds will be estimated. Lastly, the inbreeding coefficient based on the runs of homozygosity will be calculated.

2. Material and Methods

In this section first the data and the sample IDs used in this thesis will be presented. The following chapters will concentrate on the analyses performed on the data.

2.1 Data

48 Swedish Landrace goats were genotyped with the Goat SNP50 Bead Chip (Tosser-Klopp et al. 2014). This SNP chip includes the 29 autosomes and has an average spacing of markers of around 60 kbps.
The 48 genotyped individuals were from 8 herds situated in Sweden. The samples in this study are a subset of samples used by a previous master thesis student, who looked at the prevalence of caseous lymphadenitis and caprine arthritis encephalitis (CAE) in Sweden (Andersson 2019). The choice of the selected 48 individuals was based on the consent from the owners for genotyping and pedigree data to eliminate closely related individuals. For privacy reasons both the herd and the animal IDs were anonymized. Table 1 shows the number of individuals from each herd as well as the anonymized animals IDS. The anonymized animal IDs follow the order of the herd IDs; animals 1-9 are from Herd 1, animals 10-17 are from Herd 2, animals 18-24 are from Herd 3 and so on.

Table 1 Information on the number of samples from each herd with the corresponding animal IDs.

Herd ID	**Number of samples**	**Animal IDs**
Herd 1	9	Animals 1-9
Herd 2	8	Animals 10-17
Herd 3	7	Animals 18-24
Herd 4	4	Animals 25-28
Herd 5	1	Animal 29
Herd 6	7	Animals 30-36
Herd 7	5	Animals 37-41
Herd 8	7	Animals 41-48

2.2 Methods

In the following part the methods used will be presented. This will include descriptions of programme settings but also motivation for own R functions.

2.2.1 Quality control and LD pruning

Plink (Purcell et al. 2007) was used for the quality control on the raw data. There were three sets of quality controls. The details of each quality control can be found in Table 2 and will be explained below.

A general quality control was used to filter out variants with a lower than 0.9 call rate and a minor allele frequency of 0.02 or lower (allele counts of 0 and 1). Furthermore, SNPs, where the p value of the Hardy Weinberg equilibrium exact test was lower than 0.0001 were discarded. A 0.9 genotype call rate was also set on an individual level, but all 48 individuals passed this quality control. The dataset after this quality control was used for the calculation of observed heterozygosity and the PCoA. This quality control was followed up by LD- pruning for the Structure analysis.

The second quality control differed from the first one in the level of filtering on SNPs with missing values as the own function for calculating the approximated coancestry (see section 2.2.4) cannot handle missing values. This extra filtering meant that SNPs that were not successfully genotyped in all the 48 individuals were removed from the dataset. The dataset after the second quality control was only used for the calculation of the approximated coancestries.

A third quality control was needed for the calculation of the runs of homozygosity (ROH). There is no agreement in literature about the use of MAF filtering and LD pruning before ROH analysis. Meyermans et al. (2020) suggests using neither LD- pruning nor MAF filtering, as these might hinder the ROH detection by reducing the number of SNPs in the analysis too much. Therefore, in the third quality control neither of these filtering methods were used.

Table 2 The parameters of the different quality controls. The first five column names refer to the commands in Plink. The last column states the number of SNPs left after the quality control.

	Geno[1]	Mind[2]	Maf[3]	Hwe[4]	Indep-pairwise[5]	Nr SNP
QC 1	0.1	0.1	0.02	0.0001	-	48111
QC 1 (Structure)	0.1	0.1	0.02	0.0001	50 5 0.2	9839
QC 2	10^{-16}	0.1	0.02	0.0001	-	44744
QC 3	0.1	0.1	-	0.0001	-	49057

[1] Maximum missingness per SNP
[2] Maximum missingness per individual
[3] Minor allele frequency
[4] Hardy Weinberg equilibrium exact test p value

[5] Prune for linkage disequilibrium, requires a window size in variant count, a variant count to shift the window at the end of each step and pairwise r^2 threshold, respectively.

2.2.2 Observed Heterozygosity

The observed heterozygosity was calculated with Plink with the `--het` command (Purcell et al. 2007). In order to compare these results with published results of different goat populations in Europe the absolute number of heterozygous SNPs per individual was scaled by the total number of SNPs left after the quality control (see Table 2 QC 1).

2.2.3 PCoA

The data used for the PCoA analysis was from QC1. The PCoA was performed with the help of the `--distance-matrix` command in Plink (Purcell et al. 2007), that created the similarity matrix as an input for the PCoA, and the function `cmdscale()` in R. For the plots presented in the Results section the first two principal coordinates are shown.

2.2.4 Approximated coancestry calculated from genomic relationship matrix

The genomic relationship matrix of the individuals with an allele frequency of 0.5 (p_i=0.5) was calculated with an own function as an approximation for coancestry. In this relationship matrix the diagonals represent the proportion of homozygous SNPs per individuals. The possible values for the relationship matrix range from -2 to 2, while negative values are not possible for the diagonals, as negative elements indicate opposite homozygotes. The used function was based on method 1 from VanRaden (2008) but with notation in M like in Hayes et al. (2009):

$$W = M - P$$

$$G = \frac{WW'}{2\sum_{i=1}^{nloc} p_i(1-p_i)}$$

M… matrix with the minor allele counts per individual with notation of 2,1,0 for 2, 1 and 0 copies of the minor allele, respectively

P… matrix with $2p_i$, as p=0.5 the P matrix is an all-ones matrix

The M and P matrices had the dimensions of n*m, n being the number of individuals and m the number of markers. G had the dimensions n*n. The reason to choose p=0.5 was that this way the elements of the G matrix are proportional to the expected heterozygosity of the offspring of any two individuals.

The population and herd specific coancestries were calculated as the average relatedness based on the relationship matrix mentioned above. These coancestries

were calculated without the diagonals, as they would affect the results differently in each population given the small sample sizes.

2.2.5 Runs of homozygosity

The runs of homozygosity were calculated with the R package detectRUNS (Biscarini et al. 2019). The function `sldidingRUNS()` was used, which resembles the method used by Plink. I chose to use the R package instead of Plink out of a personal preference for the R environment. Table 3 shows the parameters used for `slidingRUNS` to run the analysis. The motivation for these parameters can be found under the table.

Table 3 List of parameters for the ROH analysis with the package detectRUNS.

Parameter	**Parameter name in detectRUNS**	**Value**
Scanning window size	WindowSize	36
Scanning window threshold	Threshold	0.05
Minimal number of SNPs	minSNP	36
Minimal density	minDensity	1/70*
Maximal gap	maxGap	200kb
Minimum length	minLengthBps	1000

*detectRUNS uses 1SNP/10 kb as their scale compared to Plink that uses distance btw SNPs

To choose the parameters for the ROH analysis the paper of Meyermans et al. (2020) was used as a guideline. This paper aimed to help researchers choose parameters for the ROH analysis with medium density SNP chips in livestock species and had a section devoted to goats.

It is important to control for false positive ROH in the analysis. It needs to be considered how long ROH are expected by chance, given the number of SNPs studied and the observed heterozygosity, without being identical by descent. In order to account for this, for the minimum number of SNP in a ROH Meyermans et al. (2020) uses the formula proposed by Lencz et al. (2007) and adapted by Purfield et al. (2012):

$$L = \frac{log_e \frac{\alpha}{n_s n_i}}{log_e\,(1-het)} = \frac{log_e \frac{0.05}{49057*48}}{log_e\,(1-0.392)} = 35.47$$

n_s… number of genotyped SNPs per individual, 49057 in this case (see Table 2 last column)

n_i… number of genotyped individuals, 48 in this case

α… the percentage of false positive ROH (0.05)

het… the mean heterozygosity across all SNPs, 0.392 in this case

This value (35.47) was rounded up to 36 as only integers are used.

Meyermans et al. (2020) recommends setting the scanning window size the same as the minimum number of SNPs as it does not make sense to scan in smaller windows. Furthermore, according to Meyermans et al. (2020), Additional file 5 S4 for the goat breed SAA (Saanen), a scanning window size of 36, would still be a good value.

For the scanning window threshold Meyermans et al. (2020) recommends using the following formula:

$$t = floor(\frac{N_{out} + 1}{L}, 3)$$

N_{out}…the desired number of final outer SNPs on either side of the homozygous segment

L…the scanning window size

With N_{out}=1 this results in t=0.05. This value seems reasonable when looking at the corresponding figure in the paper; Additional file 6 S5 for the goat breed SAA (Meyermans et al. 2020).

Meyermans et al. (2020) suggests minimizing the maximal gap parameter while still maximizing the genome coverage. Additional file 4 S3 for the goat breed SAA suggests that a gap of 200 kb should be appropriate.

When it comes to the minimal density Meyermans et al. (2020) showed that 50kb/SNP was appropriate for the breed SAA (see additional file 3 Figure S2 in Meyermans et al. (2020)). I chose to use 70kb/SNP as that includes more of the SNPs with the used SNP chip. I based this choice on Figure 2 from Tosser-Klopp et al. (2014) where the SNP chip is described.

The identified ROH were then used to calculate the inbreeding coefficient based on ROH. To calculate the inbreeding coefficient based on ROH the length of ROH segments is divided by the total length of the genome (Ceballos et al. 2018). See the formula below:

$$F_{ROH} = \frac{sum\ of\ ROH}{genome\ length}$$

2.2.6 Structure

For the admixture analysis the software Structure version 2.3.4 was used (Pritchard et al. 2000; Falush et al. 2003, 2007; Hubisz et al. 2009). For the input data for Structure the QC 1 (Structure) (see Table 2) was used. The run length of the burn-in period was 10,000 iterations and the number of MCMC reps after burn-in was also 10,000. The *admixture* model was used with the assumption of correlated allele frequencies between the populations. In the analysis the population IDs were used as sampling location indicators. The remaining settings were left to default. The choice of settings was based on the Structure documentation itself (Pritchard et al. 2010) and Wang (2017), who discusses common mistakes when choosing Structure parameters.

The simulations were run for K 2 to 8 with 5 replicates for each K. For transparency reasons the exact input settings can be found in Appendix 1.

To find the K that best describes the data the replicates from each K with the highest likelihoods were compared. For these seven chosen replicates the posterior probabilities were calculated, as suggested by the Structure documentation (Pritchard et al. 2010).

2.2.7 Driving distances between farms

For the calculation of the driving distance between the herds the closest towns were used for data privacy reasons. The package used to calculate them was gmapsdistance (Zarruk 2018).

2.2.8 R scripts

The R scripts are described in more detail in the data management plan (see Appendix 4) and are also commented in a way that the reader can follow the steps.

3. Results

In this section the results of the thesis will be described. First the geographical distance of the farms will be described. Afterwards, the results regarding the population structure will be presented. Then, the inbreeding of the population will be addressed and compared to other European goat breeds. Lastly, a region in chromosome 6 will be discussed shortly with regard to ROH.

3.1 Farm locations

To put the results into perspective the geographical locations of the herds are presented first.

In Figure 1 the farms are shown on the map of Sweden. In this figure it is visible that 3 of the farms (the green colour stands for both farms 5 and 8) are located in the Southern part of Sweden and the remaining five farms are located in Northern parts of Sweden.

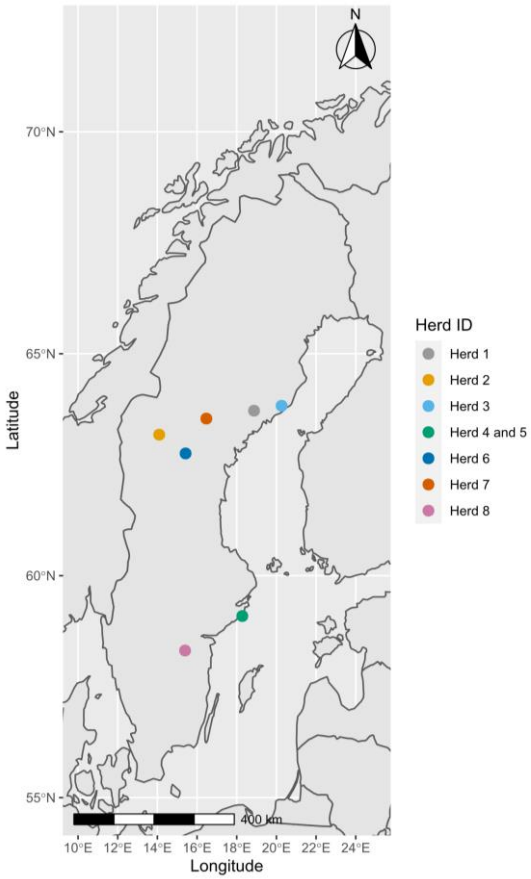


Figure 1 Locations of farms on the map of Sweden. The colours correspond to the colours used later for the farms. Herds 4 and 5 were given the colour of Herd 4 as both are located in the same town. The coordinate system used for this plot was WGS84.

Table 4 shows the driving distance calculated with the help of GoogleMaps between the farm locations.

Table 4 Calculated driving distance in km between the herds. The colours correspond to the colours used later for the farms. Herd 4 and 5 were given the colour of Herd 4 as both are located in the same town.

Driving distances between farm locations in km							
	Herd 1	Herd 2	Herd 3	Herd 4 and 5	Herd 6	Herd 7	Herd 8
Herd 1	0	339	96	627	294	170	812
Herd 2	339	0	431	578	112	175	673
Herd 3	96	431	0	672	338	264	857
Herd 4 and 5	627	578	672	0	521	566	251
Herd 6	294	112	338	521	0	176	690
Herd 7	170	175	264	566	176	0	735
Herd 8	812	673	857	251	690	735	0

3.2 Population structure

In this section the general structure of the Swedish Landrace goat will be discussed. The population structure of the Swedish Landrace goat was investigated with 2 methods: principal coordinate analysis and admixture analysis with the software Structure.

3.2.1 Principal Coordinate Analysis

Figure 2 shows the grouping of the 48 genotyped individuals regarding the first two principal coordinates. The first two principal coordinates together explain almost 15 percent of the variance seen in these 48 individuals. Figure 2 shows some grouping of individuals; some difference is visible between Herds 1, 2 and 7 (colours grey, orange and red in the figure). Herds 4, 5, 6 and 8 do not show a clear distinction (colours green, yellow, dark blue and pink). Furthermore, there is some distinction between individuals of herd 2 (orange in the figure). One individual of Herd 3 (light blue in the figure) is grouped closer with individuals of Herd 1(grey in the figure) than its own herd. One individual of Herd 1 also groups closer to Herds 4 and 6 than to its own herd.

When comparing Figure 2 of the PCoA and Figure 1 of the samples plotted on the map of Sweden then it is visible that the three herds that group more separately (Herds 1, 2 and 7) are herds that are located in the North of Sweden. Furthermore, the individual from Herd 3, that groups closer together with Herd 1 than the individuals from its own herd, could be explained by the geographical proximity of these 2 farms (96 km, the shortest possible driving distance between herds in this dataset, see Table 4). The close grouping of Herds 4, 5, and 8, and Herds 3 and 6 cannot be explained by geographical proximity as Herds 4, 5 and 8 are located in the Southern part of Sweden while Herds 3 and 6 are in the Northern part.

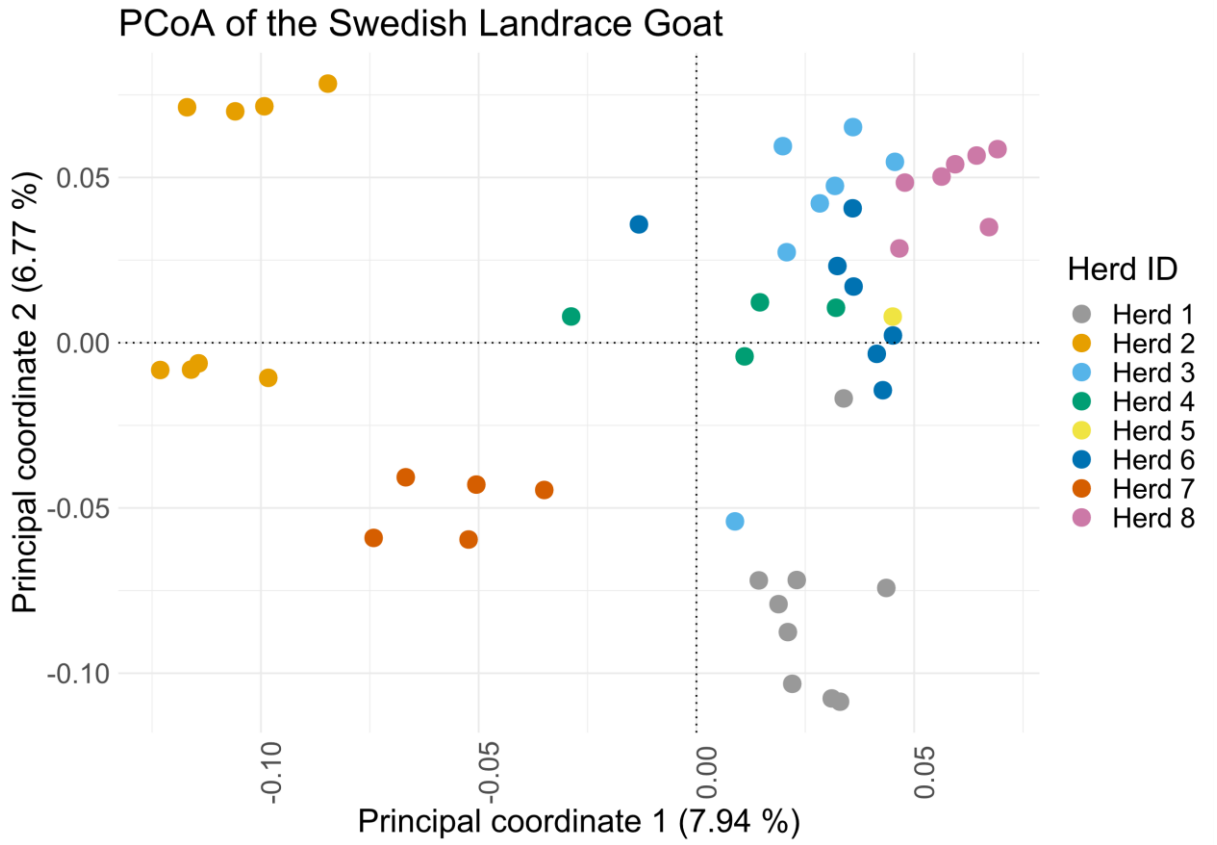


Figure 2 *Results of principal coordinate analysis on the 48 genotyped individuals. 48111 SNPs were used for the analysis presented here.*

3.2.2 Admixture analysis with Structure

The admixture analysis run with the help of Structure also shows some distinction but also similarities between the herds.

Already the analysis with K=2 showed some distinction. Herd 2 grouped mostly to a different ancestral population than the rest of the herds. As K was increased to 3 there still seems to be a differentiation of Herd 2 from the other herds. Furthermore Herd 8 and Herd 3 show a large proportion of shared assumed ancestral population. For plots for K={1,2,3,5,6,7,8} see Appendix 2.

Figure 3 shows the proportions of ancestry in each individual for 4 assumed ancestral populations (K=4). K=4 was the number of ancestral populations with the highest probability and therefore it will be explained in more detail in the main text. The first thing that one might spot when looking at this plot is the uniqueness of Herd 2 when regarding proportions of assumed ancestral populations (see green colour in the plot for Goats 10 to 17). However, Herd 2 also shows some structuring within the herd; four individuals are also grouped close to Herd 7 (see grey colour both in Herd 2 and Herd 7). The 4 individuals from Herd 2 that group differently in the Structure analysis correspond to the grouping seen in the PCoA (see Figure 2 colour orange). Herds 1 and 8 are to a large proportion assigned to just one inferred

population. Herd 1 also grouped separately in the PCoA; however, it is close to Herds 3, 4 and 5 when regarding the first principal coordinate (see Figure 2). This close grouping is also visible in the proportion of the population indicated with blue in these populations (see the colour blue in Herds 1, 4, 5 and 6, Figure 3). The shared ancestry described before (at K=3) for the Herds 3 and 8, are also visible in this plot (see the colour orange in Herds 3 and 8).

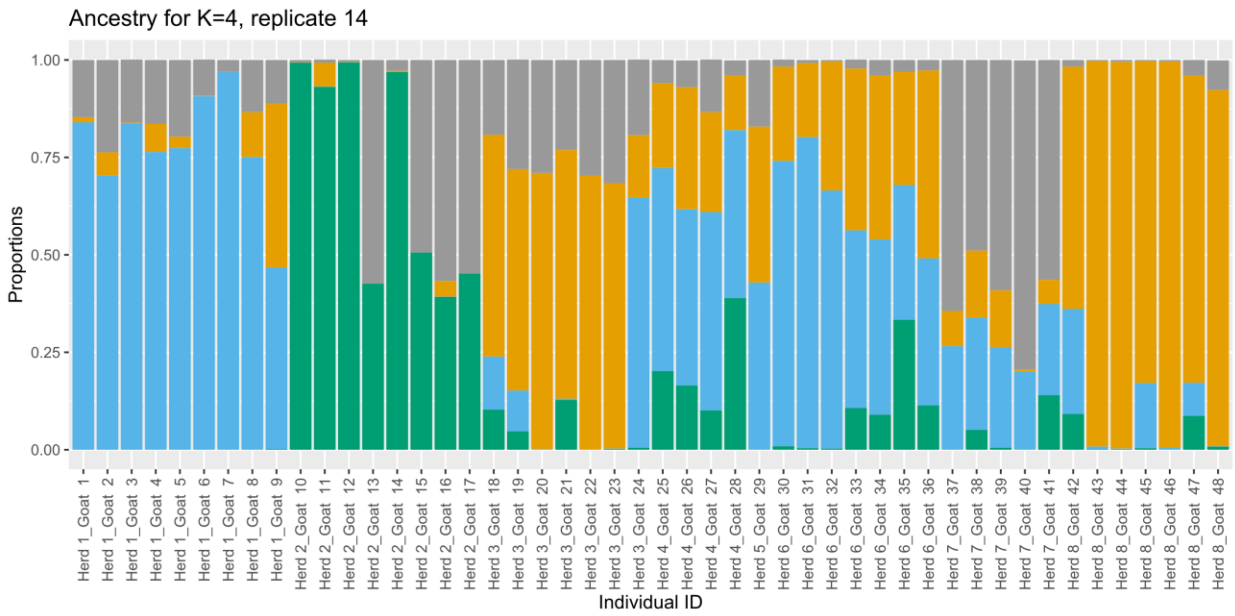


Figure 3 Results of Structure analysis with K=4. The proportion of the 4 assumed ancestral population in each individual is shown for all the 48 genotyped individuals. The colours in the plot refer to the 4 assumed ancestral populations and should not be confused with the colours used to indicate herds in other plots.

As K was increased during the analysis the rising number of assumed ancestral populations complicate the plots and make them more difficult to interpret. However, some characteristics that were mentioned with smaller K values are still visible in the plots with K=8. The uniqueness of Herd 2 is one example. The connection of Herd 2 and Herd 6 with regard to the 4 individuals in Herd 2 is also visible at K=8. The shared ancestry of Herds 4 and 6 seems constant. For the plot regarding K=8 see Appendix 2.

The fact that individuals are asymmetrically assigned to the inferred populations indicates that the structure detected is real and not an artefact. In case of an artefact, one would expect the individuals to be assigned to the inferred populations to equal proportions. As in the case of Figure 3 there are 4 inferred population, so one would expect each individual to be assigned to one fourth to every inferred population (Pritchard et al. 2010).

3.3 Inbreeding

3.3.1 Observed Heterozygosity

The observed heterozygosities for the 8 herds are illustrated in Figure 4. This figure also contains mean observed heterozygosities of nine other European goat breeds as a reference. The nine additional goat breeds are Italian, Spanish and Norwegian goat breeds. Figure 4 shows that the observed heterozygosities for the Swedish Landrace Goat range from 0.33 to 0.41. This is comparable to the European goat breeds, which are used as reference. The two goat breeds with a lower mean observed heterozygosity (B_Sk and M_Palmerain Figure 4) are both island populations.

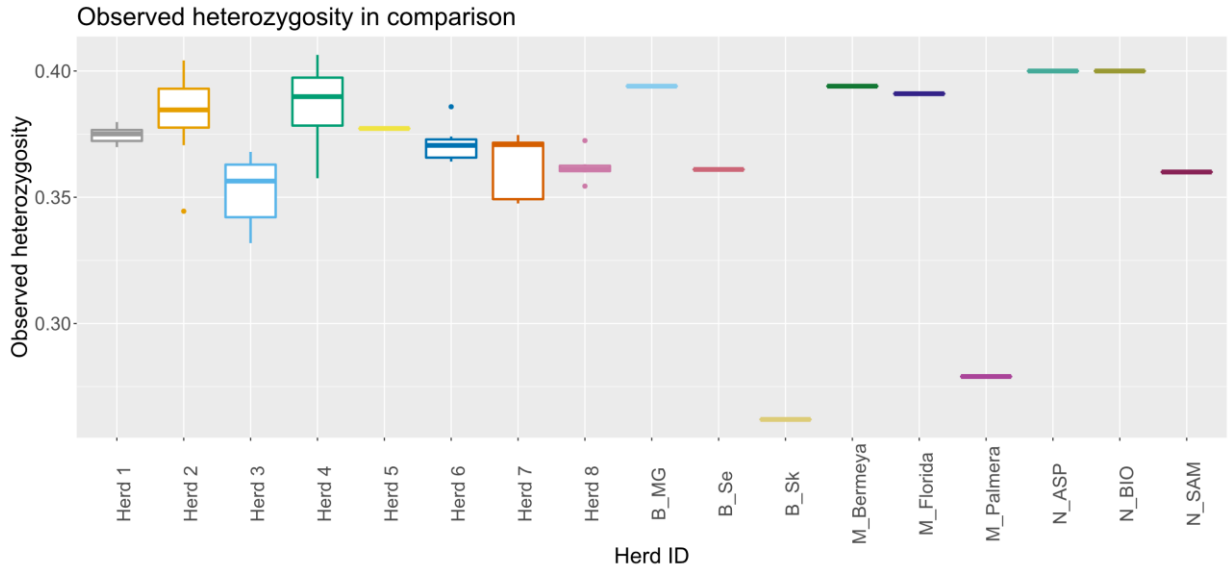


Figure 4 Observed heterozygosities for herds of the Swedish Landrace goat (Herds 1 to 8) and for nine European goat breeds. The abbreviations B_MG, B_Se and B_Sk stand for the Norwegian Milk goat, Norwegian coastal goat Selje and the Norwegian coastal goat Skorpa (data from Berg et al. (2020)). The abbreviations M_Bermeya, M_Florida and M_Palmera stand for Spanish goat breeds investigated by Manunza et al. (2016). The abbreviations N_ASP, N_BIO and N_SAM stand for the Italian goat breeds Dell'Aspromonte , Bionda dell'Adamello, and Maltese sampled in Sardinia, respectively (data from Nicoloso et al. (2015)).The number of SNPs left after the quality control in the mentioned papers was 45772, 51136 and 39257 for Berg et al. (2020), Nicoloso et al. (2015) and Manunza et al. (2016), respectively. Note that both B_sk and M_Palmera are island populations.

3.3.2 Inbreeding coefficient based on ROH

The inbreeding coefficients based on ROH for the 8 herds are illustrated in Figure 5. The inbreeding coefficients based on ROH are the highest in Herd 3. This is in line with what was visible in Figure 4 above as well, as Herd 3 had the lowest observed heterozygosity. Note that whereas high values in Figure 4 indicate higher diversity, high values in Figure 5 indicate less diversity. A further difference between the Figures 4 and 5 are, that whereas in Figure 4 all homozygous positions contribute to the observed homozygosity, not all homozygous positions contribute to the F_{ROH}, only the ones that are in longer segments.

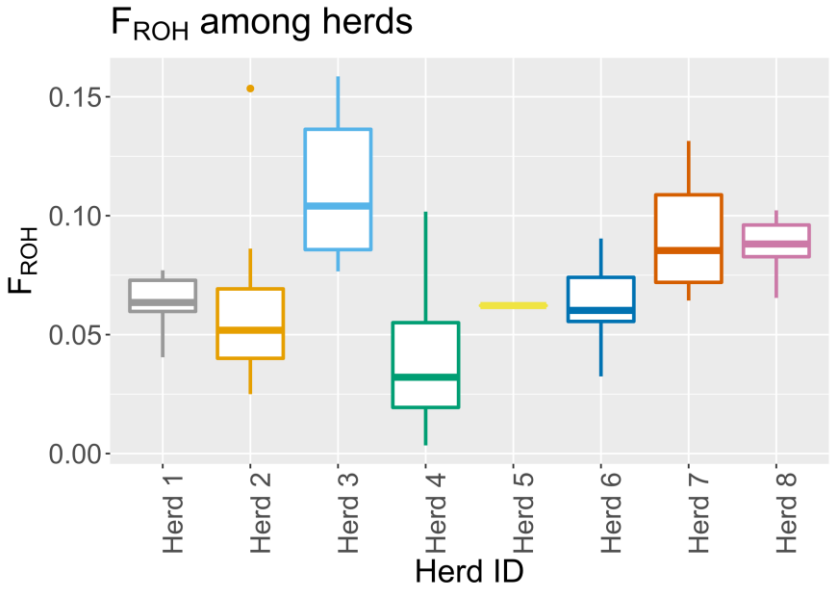


Figure 5 Inbreeding coefficient based on the ROH in each herd. The minimum length of a ROH was defined at 36 SNP, for further parameters see the methods section. The colours in the plot match the herd colours in previous plots.

The range of F_{ROH} calculated for the Swedish Landrace goat is similar to the values observed in the study of Berg et al. (2020) where the means for the two non-island populations were 0.074 and 0.115 for MG and Se respectively (for explanation about the breed abbreviations please refer to the legend of Figure 4). The F_{ROH} for the island population (Sk) was higher (0.347) than the values observed in in Figure 5. The study of Berg et al. (2020) used the same SNP chip as the one used in my thesis. In general, it is difficult to compare results based on identified ROH as all the input parameters are often not published.

3.3.3 Coancestry

As relatedness among individuals plays an important role in future inbreeding, this was also investigated. Figure 6 shows the mean approximated coancestry in each herd. The mean approximated coancestry for all the sampled individuals is meant by the heading "all" in Figure 6. In Figure 6 it is visible that herds 3 and 7 have the highest mean estimated coancestry and heard 4 has the lowest mean estimated coancestry. The mean of estimated coancestry in Herd 4 is lower than the average of all the sampled individuals. The pattern visible in Figure 6 is in line with the plots for observed heterozygosity and F_{ROH} (compare Figures 4 and 5). An interpretation of the low value for Herd 4 could be that it is close to the capital and therefore there is a better infrastructure to trade animals. Furthermore, the mean approximated coancestry for all farms is lower than the mean of the other coancestry values (see "all" in Figure 6). This also shows that there is a structure present in the population; animals in farms are generally more related to individuals from the same farm than to individuals from other farms.

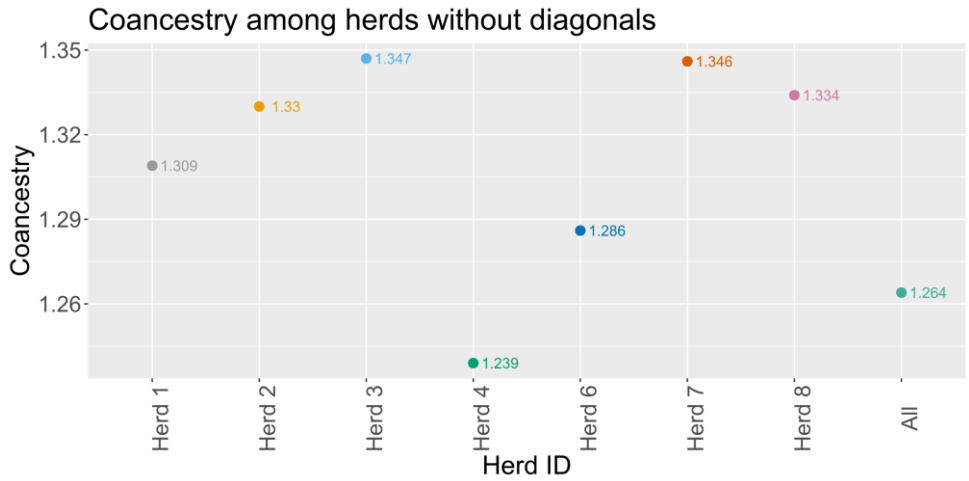


Figure 6 Estimated mean coancestries for 7 herds. Note that the means were calculated after removing the diagonals, therefore there is no value for Herd 5 where there was only one sample available. "All" stands for the mean estimated coancestry of all the sampled individuals, including the individual from Herd 5. The colours in the plot match the herd colours in previous plots.

3.4 ROH on Chromosome 6

For each chromosome a plot was created showing the ROH found in the individuals. All the 29 plots can be found in Appendix 3. Note that only the individuals that have a ROH on the given chromosome show up on the y axis of these plots. So, if an individual is not present in a plot that means that that individual did not have a ROH on that chromosome. Chromosome 6 shows something unique; 41 of the animals have a ROH in the region of 85-87 Mbps (see Figure 7, black rectangle).

Figure 8 zooms into this region of Chromosome 6 and shows the genes that can be found in this region and also all the individuals. The four casein genes that can be found in this region are *CSN1S1*, *CSN1S2*, *CSN2* and *CSN3*; the reference genome was ARS1 (Bickhart et al. 2017).

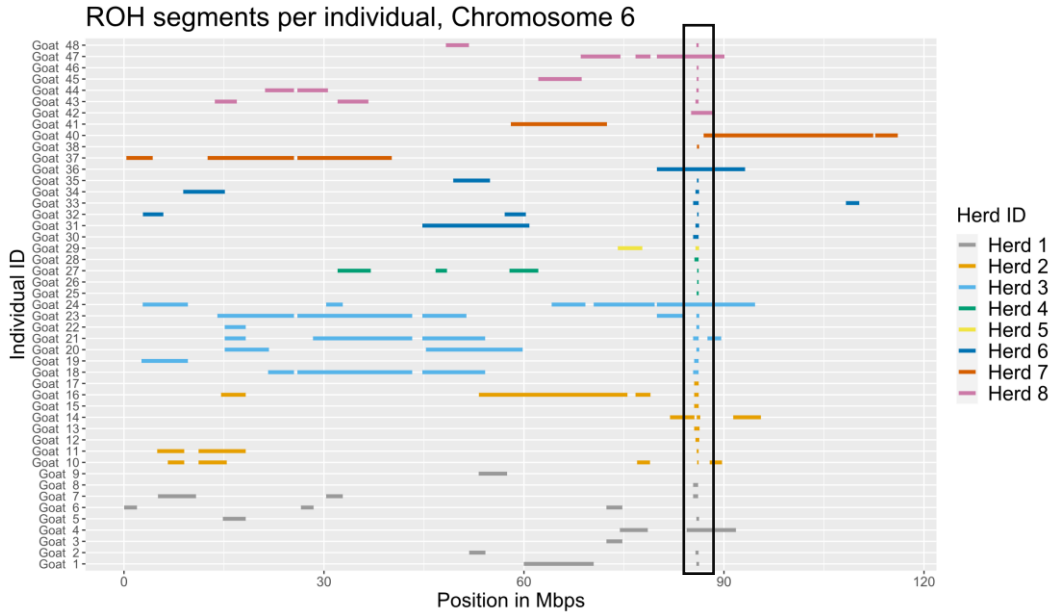


Figure 7 ROH on Chromosome 6. The black rectangle marks the region where 40 of the 48 genotyped individuals have a ROH. Figure 8 zooms in on this region. Individual 39 does not have a ROH on Chromosome 6.

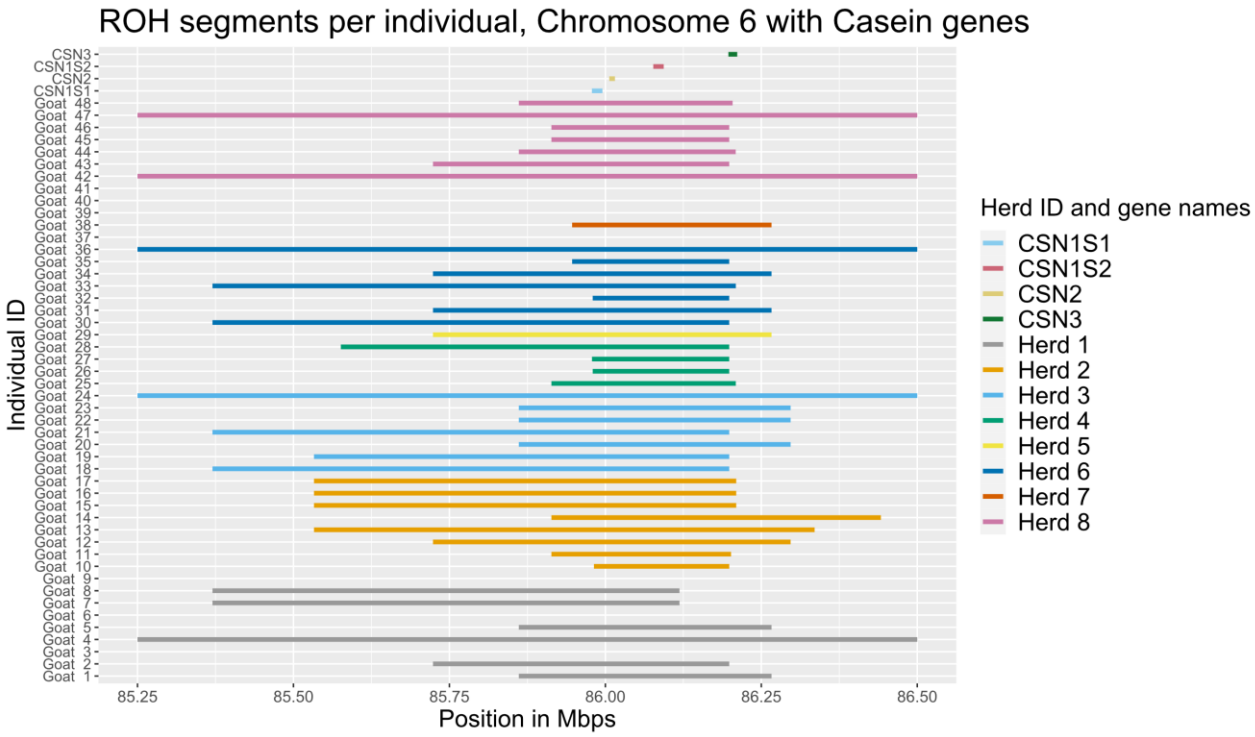


Figure 8 ROH at 85.25-86.5 Mbps on Chromosome 6. Note that for illustration in this figure all individuals are present on the y axis, even if they do not have a ROH in this region.

To investigate whether the ROH present in 41 of the individuals were the same haplotypes, a heatmap was created. Figure 9 shows this heatmap. The heatmap confirms that there is indeed little variation in this region and opposite haplotypes are rare. An exception from this is Goat 42 from Herd 8 that shows the opposite haplotype at several loci. Goat 42 is also the individual that has the genotype GG at the position of the one base pair deletion in exon 12 of the *CSN1S1* gene. The prevalence of this mutation was studied by Björk (2019), however her study did not include all the individuals mentioned here. Since Anna Johansson had the genotypes of all the individuals studied in this thesis with regard to the deletion, the deletion genotype of all the individuals was also compared to the ROH presented in Figure 7. The comparison shows that the homozygosity status is mostly similar between the ROH plot and the deletion genotypes. The only difference is that all the goats with the genotypes DA (so heterozygous for the deletion) showed up as having a ROH in the plot. The individuals in question are Goats 2, 10, 11, 13, 15, 16, 17, 38 and 43. The heterozygous individuals for the genotype DG did not show up as ROH.

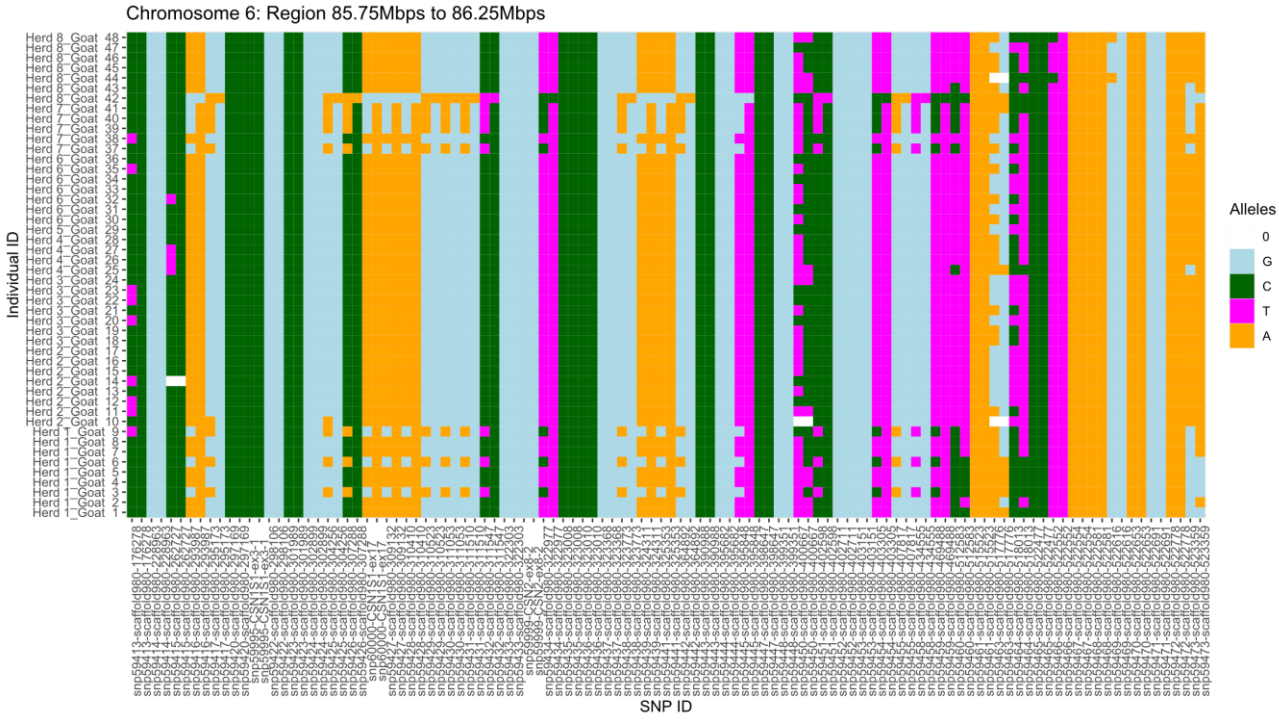


Figure 9 Heatmap of SNPs at the region of 85.75Mbps to 86.25Mbps on Chromosome 6. The colours indicate the different alleles, white stands for missing data. The SNP names are on the X axis. Note that there is not SNP for exon 12 of the CSN1S1 in this SNP chip.

4. Discussion

4.1 Discussion of results

The aim of this thesis was to study the genetic diversity in the Swedish Landrace Goat as there were no prior genetic studies conducted on any of the goat breeds in Sweden. The results shown here indicate that there is some structuring among the sampled herds, as shown by the PCoA and the Structure plots (see Figures 2 and 3). However, there are also herds that group together in the PCoA plot and show a lot of similarity in the Structure plots indicated by shared proportions of the same colour. The calculation of coancestry also provides some information on the structure of the populations as the mean coancestry is lower than the averages of all the other coancestries (see Figure 6). This indicates structuring between farms. Regarding the inbreeding some difference is visible between the herds, however, all the herds are in the range of the values that are observed in other European non-island goat populations (see Figure 4 regarding observed heterozygosities). Both the inbreeding measures, observed heterozygosity and F_{ROH}, gave the same picture when ranking the herds of the Swedish Landrace Goat with regard to level of inbreeding. The calculation of coancestry between and within herds aimed to assess the risk of future inbreeding. It shows that the herds, currently having a higher level of inbreeding are expected to follow the same trend unless new less related animals are introduced into these herds. Furthermore, the calculation of the coancestry showed that the average coancestry between all the sampled individuals is lower than of all but one individual herds. This result shows that future inbreeding could somewhat be avoided by trading more animals among farms.

The ROH found on Chromosome 6 was an interesting aspect of this thesis. The ROH found in the region of 85.25-86.50 Mbps on chromosome 6 could indicate the history of selection on the Casein genes in this region. As the Casein genes are responsible for more than 80% of the protein content in the milk (Ceballos et al. 2009) and the goats in question are dairy goats, this is a logical finding. Dagnachew et al. (2011) found that the deletion on exon 12 of the *CSN1S1,* studied in the Swedish goat population in the thesis of Björk (2019), leads to lower fat and protein percentage in the milk and also to a higher milk yield. This mutation was found to have a high frequency both in Norway and Sweden (Hayes et al. 2006; Björk

2019).This could mean that the selection of the goats in the past was based on the quantity of the produced milk. It was previously discovered that the haplotype of the one base pair deletion is only different from the A haplotype on the position of the deletion (A. M. Johansson personal communication 2022). This was also confirmed in the comparison of the deletion genotypes and the ROH detected with the medium density SNP chip; all of the individuals that are heterozygous for the deletion and the ancestral A allele had a ROH at this position. These findings indicate that before the deletion occurred there were two haplotypes, one with the A allele and one with the G allele, and that the mutation with the deletion occurred on the haplotype with the A allele. A recent study looking into the signatures of selection in Swiss goat breeds with whole genome sequencing data did not find any ROH on Chromosome 6, which was present in at least 80% of their samples (Signer-Hasler et al. 2022).

In order to see geographical patterns in the data, the farms were plotted on the map of Sweden and also the driving distances between the farms were calculated. Close attention was paid to the geographical location of the farms while analysing the results. The geographical locations were used as a measure of ease of exchange of animals between the farms. An example is the close grouping of an individual from Herd 3 with Herd 1 which can be explained by the proximity of these two herds (see Figure 2 of the PCoA). There are also examples in Figure 2 that cannot be explained simply by the geographic location of the farms (see Herds 8 and 3 that group close together but are more than 800km away from each other). However, there is another factor that might have an effect on the structuring and exchange of animals; the caprine arthritis encephalitis (CAE) status of the farms. Farms that are classified as CAE free can only keep this status if the animals they add to their herds are also from CAE free farms. On the other hand, if a farm is not classified as CAE free then it can purchase individuals from any farm regardless of the CAE status. This structuring of farms could potentially have an effect on the inbreeding level of the individuals in farms, if for example the number of farms that are CAE free is low. Unfortunately, the CAE status of the farms in this thesis was not available for all of them.

When discussing the results, it needs to be mentioned that the sampling of the farms for genotyping cannot be considered random. Only samples of farms could be genotyped that agreed to this procedure.

4.2 Future risks and threats to the Swedish Landrace Goat

There are some risks that should be considered when breeding the Swedish Landrace Goat in the future. The potential future risks will be discussed in this chapter.

Firstly, the Swedish goat breeding organization emphasizes the breeding against the deletion on exon 12 of *CSN1S1* in order to increase the protein yield of the milk produced (Svenska Getavelsförbundet 2021). Given the potential high prevalence of this mutation shown by Björk (2019), breeding against this mutation with a high intensity would result in a high level on inbreeding in the future.

Secondly, a potential risk for inbreeding is posed by the grouping of farms in CAE-free and not CAE-free. This grouping is epidemiologically relevant, however it should be closely monitored and studied how this grouping affects the within group coancestry of the CAE-free farms and the non-CAE free farms. These coancestries should be compared to the overall coancestry of the whole Swedish Landrace Goat population and monitored over time. Furthermore, this analysis would answer the question to what extent the CAE status of the farms affect the structure visible in this goat breed.

Thirdly, there is a lack of funding in the Swedish Goat Association which is a voluntary organization. This leads to problems when it comes to data organization and handling. It would benefit the Swedish goat breeders, if projects were created to upgrade the software, Elitlamm, used to store data for the Swedish goat breeds.

Lastly, I believe that it would be beneficial to define a proper breeding goal for the Swedish Landrace Goat. This however might be hard to structure as most goat keepers are small scale goat keepers. The previous point mentioned about data handling and organizing would be a prerequisite of a more organized breeding programme for the farmers. A breeding programme with the goal of improving the production traits in the Swedish Landrace Goat would help it to survive in the market for longer, and thereby secure the population.

4.3 Studies for the future

Something that I have encountered during my thesis is the lack of data with regard to goat breeds in Sweden. Therefore, there is room for several future studies. However, as a starting point, I think some data organization is necessary in order to be able to access all the data that currently exists for the Swedish goat breeds.

Furthermore, this thesis is not to be taken as the full description of the genomic diversity of the Swedish Landrace Goat. It is rather the beginning of several studies that are still necessary for the Swedish goat breeds. In this thesis the inbreeding level was studied, however this does not give any insight on the rate of inbreeding

from one generation to the other in the population. The rate of inbreeding could be calculated with a sample of the next generation in comparison to the results presented in this thesis or by calculating it from the pedigree data, given that this data can be accessed.

It would also be interesting to study the structure and the relationship of all the four Swedish goat breeds (Göinge Goat, the Jämt Goat and Lapp Goat in addition to the Swedish Landrace Goat). As the Swedish Landrace Goat is the goat breed in Sweden with the highest population number it is essential to assess the three other Swedish breeds as well. The Swedish goat breeds could also be compared to other European breeds with regard to their population structure and phylogeny.

Furthermore, even in absence of exact population data, which is the case for the Swedish goat breeds, it is possible to calculate the effective population size with the help of genomic data (Do et al. 2014).This could shed light on the future viability of these breeds.

Lastly, the finding of the ROH segment on Chromosome 6 opens up the question of looking for patterns of selection in the Swedish Landrace goats. As the Scandinavian goat breeds are said to be closely related (Svenska Getavelsförbundet 2021) it would be interesting if there are any region specific selection patterns. It also needs to be mentioned that short ROH are harder to detect (Purfield et al. 2012) and therefore there are probably more short ROH in the population studied here that were not found.

5. Conclusion

To study the population structure, a principal coordinate analysis and an admixture analysis were conducted. The level of inbreeding was investigated with three measures; observed heterozygosity, F_{ROH} and approximated coancestry. The results show that there is some structuring in the population and this structure is not solely due to the geographic location of these farms. The inbreeding level of the farms is comparable to other European non-island goat populations.

In case of a strong selection against the deletion in exon 12 of the *CSN1S1* the inbreeding could increase and therefore, should be closely monitored. The inbreeding level can also be affected by the restricted trading of the animals due to the CAE status of the farms.

Further research is needed in several areas to describe all the four Swedish goat breeds with regard to their inbreeding level and clustering in comparison with other populations. Furthermore, it is important to calculate the effective population size and the rate of inbreeding in these breeds to assess their future viability.

6. References

Andersson, E. (2019). *Böldsjuka och kaprin artrit encefalit hos svenska mjölkproducerande getter*. [Second cycle, A2E]. https://stud.epsilon.slu.se/14792/ [2022-01-12]

Barbato, M., Orozco-terWengel, P., Tapio, M. & Bruford, M.W. (2015). SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics*, 6. https://www.frontiersin.org/article/10.3389/fgene.2015.00109 [2022-01-24]

Berg, P., Groeneveld, L.F., Brekke, C., Våge, D.I., Sørheim, K.M. & Grøva, L. (2020). Genetic characterization of a small closed island population of Norwegian coastal goat. *Acta Agriculturae Scandinavica, Section A — Animal Science*, 69 (1–2), 47–52. https://doi.org/10.1080/09064702.2020.1729852

Bickhart, D.M., Rosen, B.D., Koren, S., Sayre, B.L., Hastie, A.R., Chan, S., Lee, J., Lam, E.T., Liachko, I., Sullivan, S.T., Burton, J.N., Huson, H.J., Nystrom, J.C., Kelley, C.M., Hutchison, J.L., Zhou, Y., Sun, J., Crisà, A., Ponce de León, F.A., Schwartz, J.C., Hammond, J.A., Waldbieser, G.C., Schroeder, S.G., Liu, G.E., Dunham, M.J., Shendure, J., Sonstegard, T.S., Phillippy, A.M., Van Tassell, C.P. & Smith, T.P.L. (2017). Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics*, 49 (4), 643–650. https://doi.org/10.1038/ng.3802

Biscarini, F., Cozzi, P., Gaspa, G. & Marras, G. (2019). *detectRUNS: Detect Runs of Homozygosity and Runs of Heterozygosity in Diploid Genomes*. Version: 0.9.6. https://CRAN.R-project.org/package=detectRUNS [2021-12-07]

Björk, A. (2019). *Detection of mutational sites in the CSN1S1 gene and analysis of αs1-casein composition of the milk in Swedish goats (Capra hircus)*. [Second cycle, A2E]. https://stud.epsilon.slu.se/14201/ [2021-12-15]

Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, 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, L.S., Morales, E.R., de la Torre Adarve, G., Castro, J.D., Martínez, L.P. & Sampelayo, M.R.S. (2009). Composition of goat and cow milk produced under similar conditions and analyzed by identical methodology. *Journal of Food Composition and Analysis*, 22 (4), 322–329. https://doi.org/10.1016/j.jfca.2008.10.020

Dagnachew, B.S., Thaller, G., Lien, S. & Ådnøy, T. (2011). Casein SNP in Norwegian goats: additive and dominance effects on milk composition and quality. *Genetics Selection Evolution*, 43 (1), 31. https://doi.org/10.1186/1297-9686-43-31

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J. & Ovenden, J.R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular*

Ecology Resources, 14 (1), 209–214. https://doi.org/10.1111/1755-0998.12157
Falconer, D.S. & Mackay, T.F.C. (1996). *Introduction to quantitative genetics*. 4th Edition. Essex, England: Longman.
Falush, D., Stephens, M. & Pritchard, J. (2007). Inference of population STRUCTURE using multilocus genotype data: dominant markers and null alleles. *Molecular ecology notes*, 7, 574–578. https://doi.org/10.1111/j.1471-8286.2007.01758.x
Falush, D., Stephens, M. & Pritchard, J.K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164 (4), 1567–1587. https://doi.org/10.1093/genetics/164.4.1567
FAO (2021). *Browse by country and species | Domestic Animal Diversity Information System (DAD-IS) | Food and Agriculture Organization of the United Nations*. *Domestic Animal Diversity Information System (DAD-IS)*. https://www.fao.org/dad-is/browse-by-country-and-species/en/ [2021-12-22]
Ganteil, A., Rodriguez-Ramilo, S.T., Ligonesche, B. & Larzul, C. (2021). Characterization of Autozygosity in Pigs in Three-Way Crossbreeding. *Frontiers in Genetics*, 11. https://www.frontiersin.org/article/10.3389/fgene.2020.584556 [2022-03-11]
Ghoreishifar, S.M., Rochus, C.M., Moghaddaszadeh-Ahrabi, S., Davoudi, P., Salek Ardestani, S., Zinovieva, N.A., Deniskova, T.E. & Johansson, A.M. (2021). Shared Ancestry and Signatures of Recent Selection in Gotland Sheep. *Genes*, 12 (3), 433. https://doi.org/10.3390/genes12030433
Hayes, B., Hagesæther, N., Ådnøy, T., Pellerud, G., Berg, P.R. & Lien, S. (2006). Effects on Production Traits of Haplotypes Among Casein Genes in Norwegian Goats and Evidence for a Site of Preferential Recombination. *Genetics*, 174 (1), 455–464. https://doi.org/10.1534/genetics.106.058966
Hayes, B.J., Bowman, P.J., Chamberlain, A.C., Verbyla, K. & Goddard, M.E. (2009). Accuracy of genomic breeding values in multi-breed dairy cattle populations. *Genetics Selection Evolution*, 41 (1), 1–9. https://doi.org/10.1186/1297-9686-41-51
Högberg, M. (2011). *Milk yield and composition in Swedish landrace goats (Capra hircus) kept together with their kids in two different systems*. https://stud.epsilon.slu.se/2551/ [2021-12-20]
Hubisz, M.J., Falush, D., Stephens, M. & Pritchard, J.K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, 9 (5), 1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x
Johansson, A.M. & Nelson, R.M. (2015). Characterization of genetic diversity and gene mapping in two Swedish local chicken breeds. *Frontiers in Genetics*, 6. https://www.frontiersin.org/article/10.3389/fgene.2015.00044 [2022-03-06]
Lencz, T., Lambert, C., DeRosse, P., Burdick, K.E., Morgan, T.V., Kane, J.M., Kucherlapati, R. & Malhotra, A.K. (2007). Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, 104 (50), 19942–19947. https://doi.org/10.1073/pnas.0710021104
Manunza, A., Noce, A., Serradilla, J.M., Goyache, F., Martínez, A., Capote, J., Delgado, J.V., Jordana, J., Muñoz, E., Molina, A., Landi, V., Pons, A., Balteanu, V., Traoré, A., Vidilla, M., Sánchez-Rodríguez, M., Sànchez, A., Cardoso, T.F. & Amills, M. (2016). A genome-wide perspective about the

diversity and demographic history of seven Spanish goat breeds. *Genetics Selection Evolution*, 48 (1), 52. https://doi.org/10.1186/s12711-016-0229-6
Meyermans, R., Gorssen, W., Buys, N. & Janssens, S. (2020). How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics*, 21, 94. https://doi.org/10.1186/s12864-020-6463-x
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
Nicoloso, L., Bomba, L., Colli, L., Negrini, R., Milanesi, M., Mazza, R., Sechi, T., Frattini, S., Talenti, A., Coizet, B., Chessa, S., Marletta, D., D'Andrea, M., Bordonaro, S., Ptak, G., Carta, A., Pagnacco, G., Valentini, A., Pilla, F., Ajmone-Marsan, P. & Crepaldi, P. (2015). Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. *Genetics Selection Evolution*, 47 (1), 1–10. https://doi.org/10.1186/s12711-015-0140-6
Patterson, N., Price, A.L. & Reich, D. (2006). Population Structure and Eigenanalysis. *PLOS Genetics*, 2 (12), e190. https://doi.org/10.1371/journal.pgen.0020190
Porras Hurtado, G.L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, A. & Lareu, M. (2013). An overview of STRUCTURE: Application, parameter settings, and supporting software. *Frontiers in genetics*, 4, 98. https://doi.org/10.3389/fgene.2013.00098
Pritchard, J., Wen, X. & Falush, D. (2010). Documentation for structure software: Version 2.3
Pritchard, J.K., Stephens, M. & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 155 (2), 945–959
Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., 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, 70. https://doi.org/10.1186/1471-2156-13-70
Rochus, C.M. & Johansson, A.M. (2017). Estimation of genetic diversity in Gute sheep: pedigree and microsatellite analyses of an ancient Swedish breed. *Hereditas*, 154 (1), 4. https://doi.org/10.1186/s41065-017-0026-4
Rochus, C.M., Jonas, E. & Johansson, 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
Signer-Hasler, H., Henkel, J., Bangerter, E., Bulut, Z., Drögemüller, C., Leeb, T. & Flury, C. (2022). Runs of homozygosity in Swiss goats reveal genetic changes associated with domestication and modern selection. *Genetics Selection Evolution*, 54 (1), 1–11. https://doi.org/10.1186/s12711-022-00695-w
Svenska Getavelsförbundet (2021). Avelsprogram för getter av svensk lantras från år 2020
Tave, D. (1999). Inbreeding and brood stock management. *Fisheries Technical Paper*. No. 392. Rome: FAO.
Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y., Donnadieu-Tonon, C., Eggen, A., Heuven, H.C.M., Jamli, S., Jiken, A.J., Klopp, C., Lawley, C.T., McEwan, J., Martin, P., Moreno, C.R., Mulsant, P., Nabihoudine, I., Pailhoux, E., Palhière, I., Rupp, R., Sarry, J., Sayre, B.L., Tircazes, A., Wang, J., Wang, W., Zhang, W. & Consortium, and the

I.G.G. (2014). Design and Characterization of a 52K SNP Chip for Goats. *PLOS ONE*, 9 (1), e86227. https://doi.org/10.1371/journal.pone.0086227

Upadhyay, M., Eriksson, S., Mikko, S., Strandberg, E., Stålhammar, H., Groenen, M.A.M., Crooijmans, R.P.M.A., Andersson, G. & Johansson, A.M. (2019). Genomic relatedness and diversity of Swedish native cattle breeds. *Genetics Selection Evolution*, 51 (1), 1–11. https://doi.org/10.1186/s12711-019-0496-0

VanRaden, P.M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, 91 (11), 4414–4423. https://doi.org/10.3168/jds.2007-0980

Wang, J. (2017). The computer program structure for assigning individuals to populations: easy to use but easier to misuse. *Molecular Ecology Resources*, 17 (5), 981–990. https://doi.org/10.1111/1755-0998.12650

Yurchenko, S., Sats, A., Tatar, V., Kaart, T., Mootse, H. & Jõudu, I. (2018). Fatty acid profile of milk from Saanen and Swedish Landrace goats. *Food Chemistry*, 254, 326–332. https://doi.org/10.1016/j.foodchem.2018.02.041

Zarruk, R.A.M.& & D.R.T.& D. (2018). *gmapsdistance: Distance and Travel Time Between Two Points from Google Maps*. https://CRAN.R-project.org/package=gmapsdistance

Zuur, A.F., Ieno, E.N. & Smith, G.M. (eds.) (2007). Principal coordinate analysis and non-metric multidimensional scaling. *Analysing Ecological Data*. New York, NY: Springer, 259–264. https://doi.org/10.1007/978-0-387-45972-1_15

Popular science summary

There are four goat breeds present in Sweden. However, no genetic studies were dedicated to investigating the current genetic diversity in these breeds. This thesis therefore aims to describe the population structure and the level of inbreeding in the goat breed with the largest population number, the Swedish Landrace goat. The Swedish Landrace goat is a dairy breed with a non-uniform appearance; fell colour and length can vary between individuals and there are both animals with and without horns.

In order to study the genetic diversity, 48 samples from eight farms with genetic marker information were used. To study the population structure, a principal coordinate analysis, and an admixture analysis were conducted. Both of these analyses can be used to describe the level of genetic exchange between the farms. The level of inbreeding was investigated with three measures; observed heterozygosity, F_{ROH} and approximated coancestry. These measures shed light on the current level of inbreeding and on the possible future inbreeding.

The results of this thesis show that there is some structuring in the population of the Swedish Landrace goat, meaning that the exchange of animals between some farms is less than among others. Furthermore, this structure is not solely due to the geographic location of these farms. The inbreeding level of the studied farms is comparable to other European non-island goat populations. Lastly, a potential selection pattern was identified on Chromosome 6 in the region of the Casein genes, that are responsible for most of the protein content in the milk.

There are also some areas of further research that were identified during this thesis. For example, further research is needed to calculate the effective population size and the rate of inbreeding in this breed. Furthermore, it is necessary to study the remaining three Swedish goat breeds as well regarding their genetic diversity.

Lastly, some risks were identified with regard to future inbreeding. There is an attempt by the Swedish Goat Association to reduce the prevalence of a mutation in the *CSN1S1* gene, which affects the protein content and the quantity of the produced milk. As the mutation is highly prevalent this reduces the number of individuals that can be selected and might lead to a high rate of inbreeding in the upcoming generations.

Appendix 1 - Simulation input for Structure

```
================  Parameter Set: locprior_10k_true  ==================

                    Running Length

 Length of Burnin Period: 10000
 Number of MCMC Reps after Burnin: 10000

                    Ancestry Model Info

 Use Admixture Model
  * Use Sampling Location Information
  * Use Population IDs as Sampling Location Information
  * Infer Alpha
  * Initial Value of ALPHA (Dirichlet Parameter for Degree of Admixture): 1.0
  * Use Same Alpha for all Populations
  * Use a Uniform Prior for Alpha
    ** Maximum Value for Alpha: 10.0
    ** SD of Proposal for Updating Alpha: 0.025

                    Frequency Model Info

 Allele Frequencies are Correlated among Pops
  * Assume Different Values of Fst for Different Subpopulations
  * Prior Mean of Fst for Pops: 0.01
  * Prior SD   of Fst for Pops: 0.05
  * Use Constant Lambda (Allele Frequencies Parameter)
  * Value of Lambda: 1.0

                    Advanced Options

 Estimate the Probability of the Data Under the Model
 Frequency of Metropolis update for Q: 10
```

Appendix 2 - Structure Plots

For each K the replicate with the highest likelihood is displayed below. K=4 can be found in the main text (see the Results section).

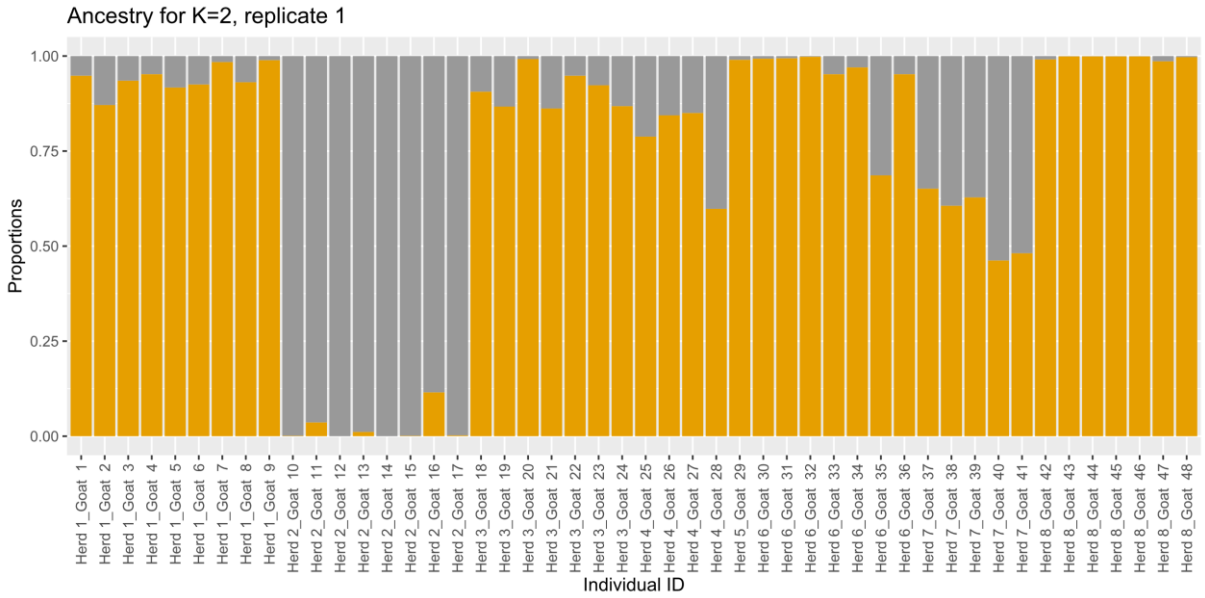


Figure 10 Results of structure analysis with K=2. The proportion of the 2 assumed ancestral populations in each individual is shown for all the 48 genotyped individuals.

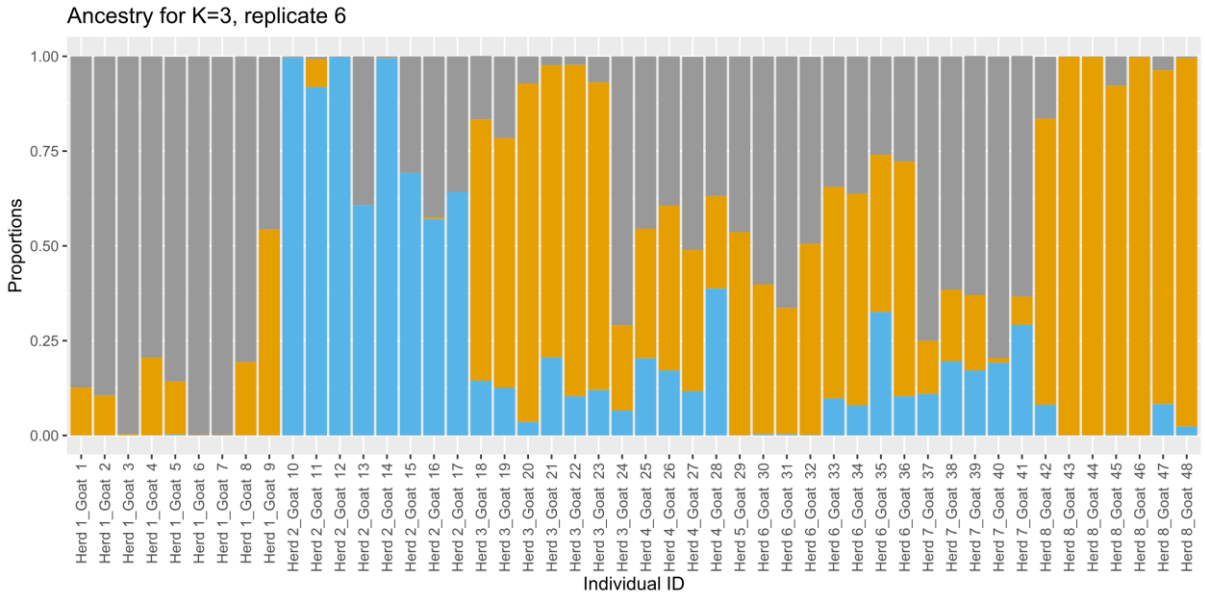


Figure 11 Results of structure analysis with K=3. The proportion of the 3 assumed ancestral populations in each individual is shown for all the 48 genotyped individuals.

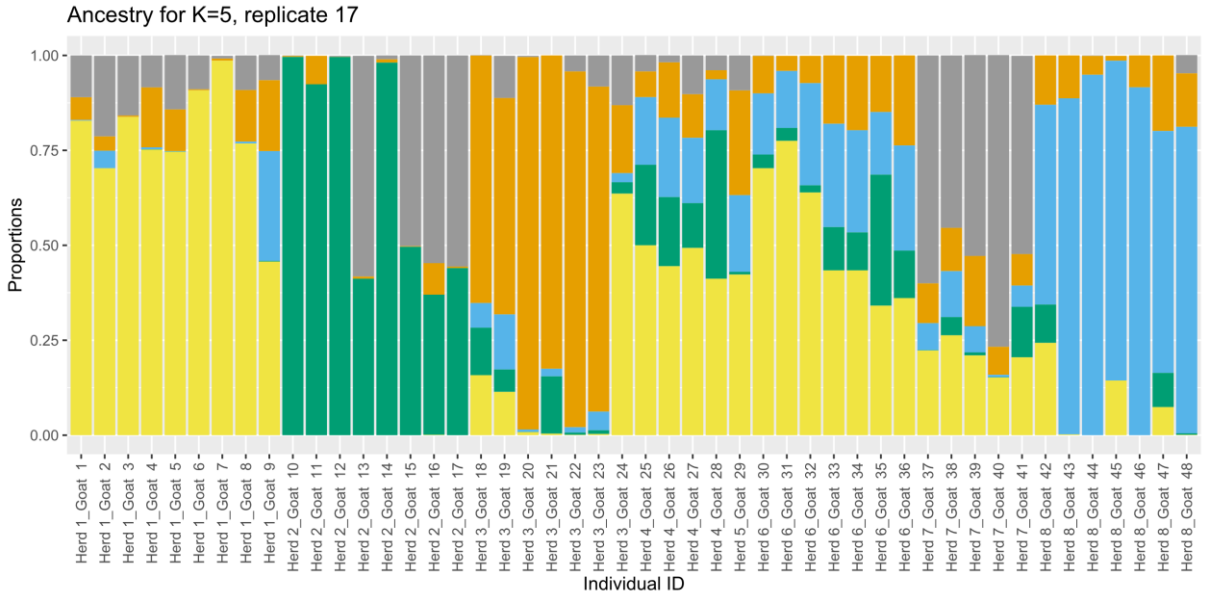


Figure 12 Results of structure analysis with K=5. The proportion of the 5 assumed ancestral populations in each individual is shown for all the 48 genotyped individuals.

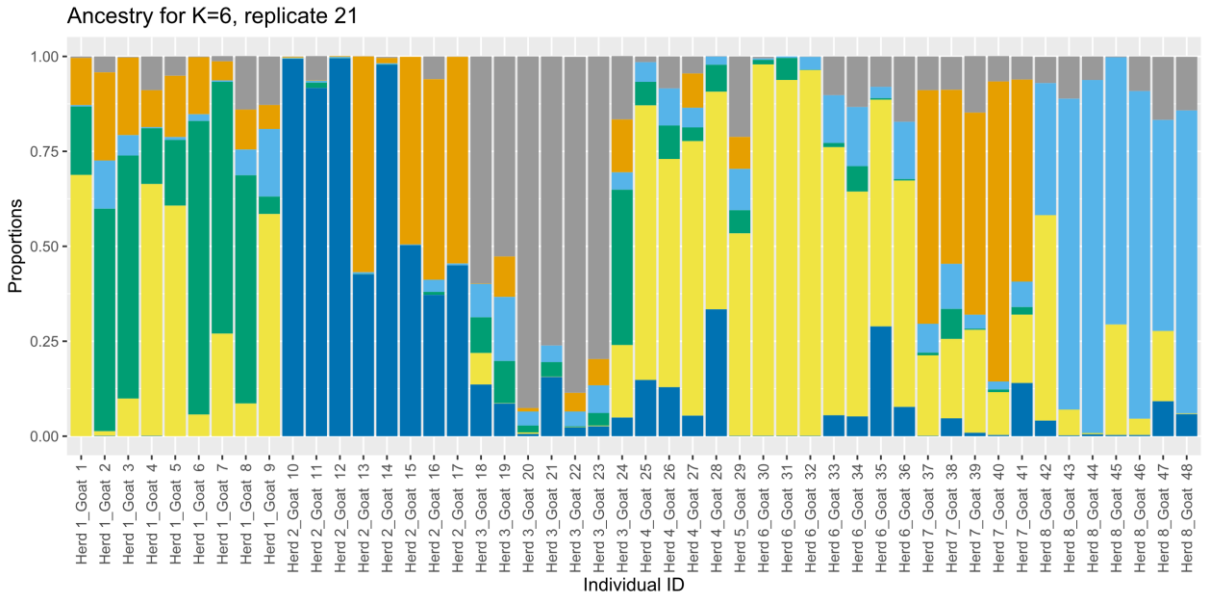


Figure 13 Results of structure analysis with K=6. The proportion of the 6 assumed ancestral populations in each individual is shown for all the 48 genotyped individuals.

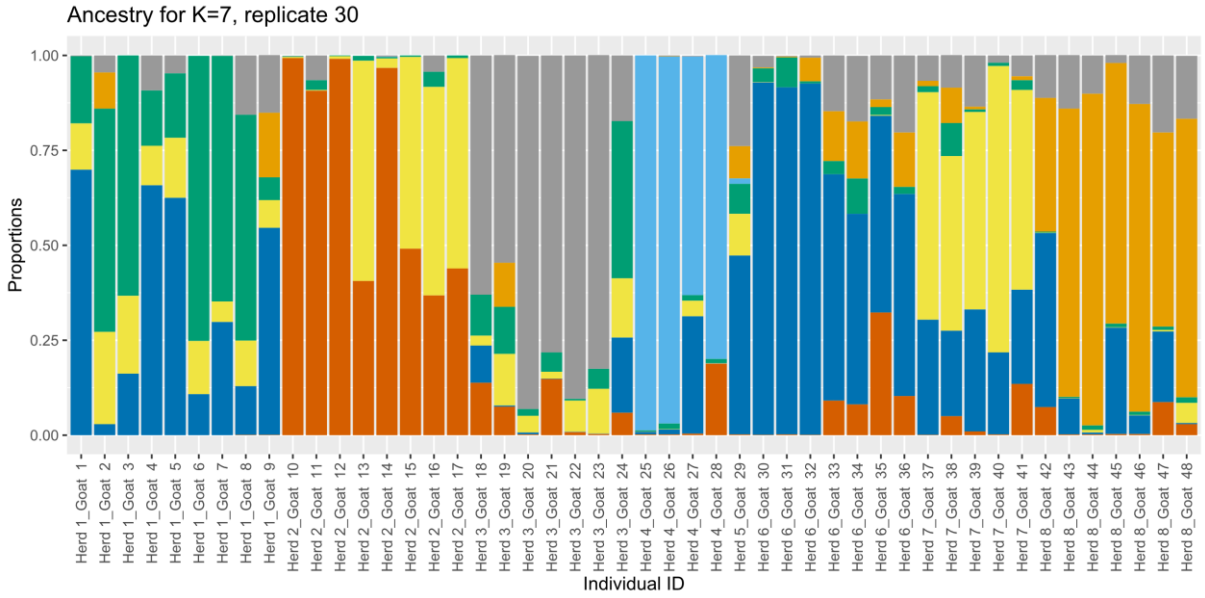


Figure 14 Results of structure analysis with K=6. The proportion of the 6 assumed ancestral populations in each individual is shown for all the 48 genotyped individuals.

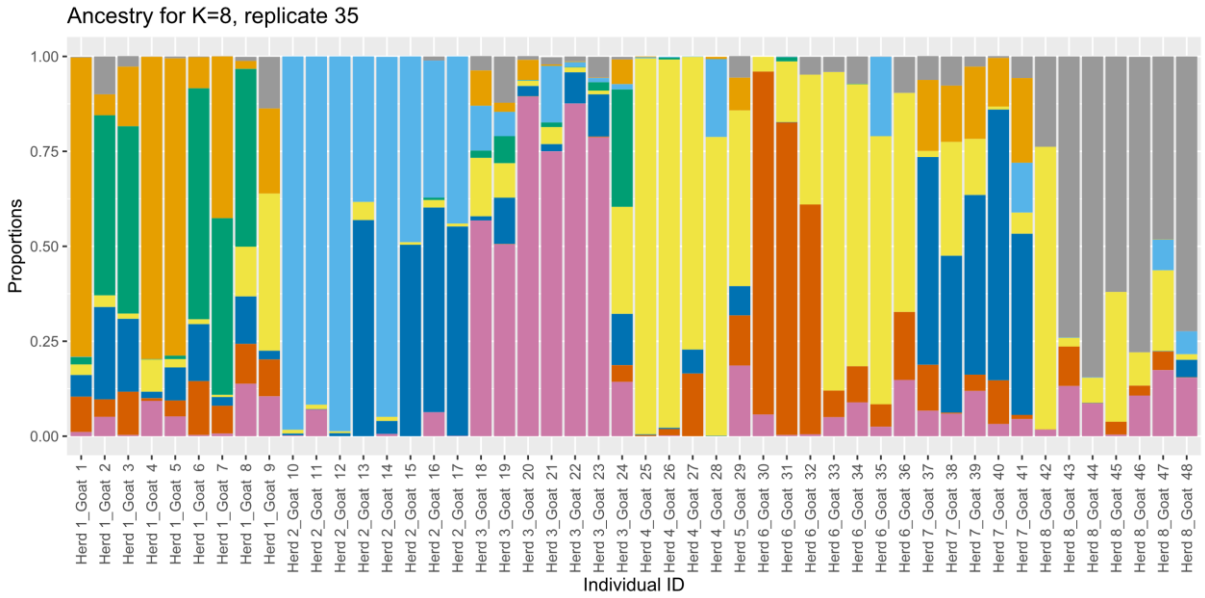


Figure 15 Results of structure analysis with K=6. The proportion of the 6 assumed ancestral populations in each individual is shown for all the 48 genotyped individuals.

Appendix 3 - ROH plots per Chromosome

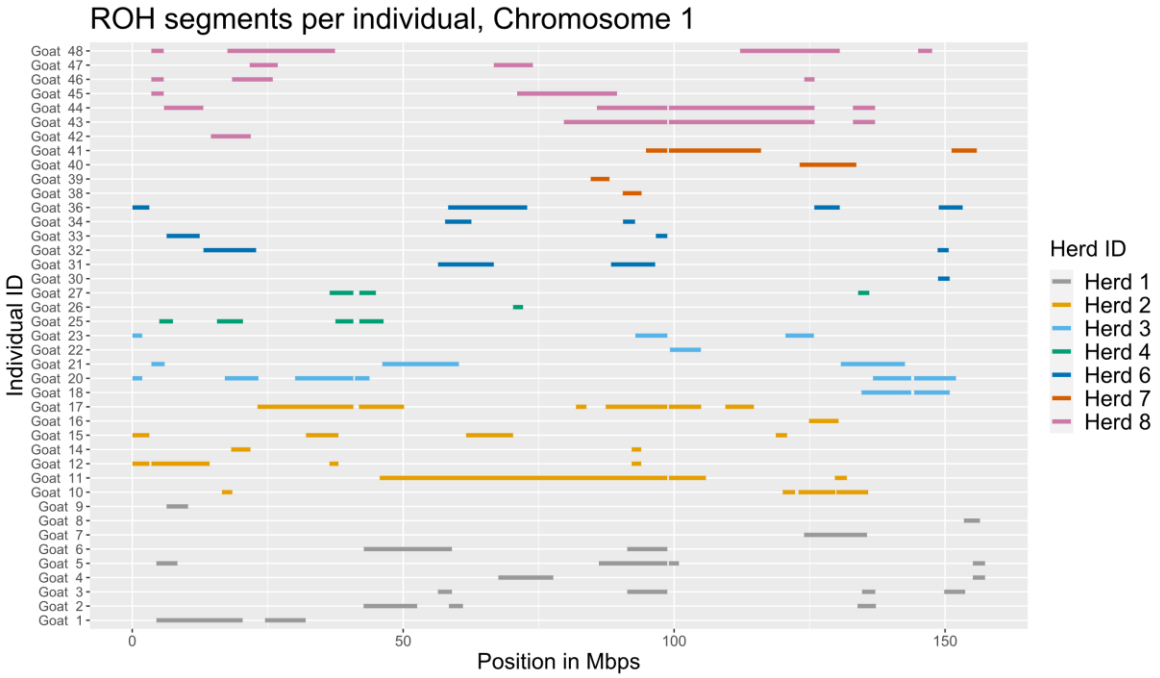


Figure 16 ROH on Chromosome 1. The individuals that do not show up in the plot did not have a ROH on this chromosome.

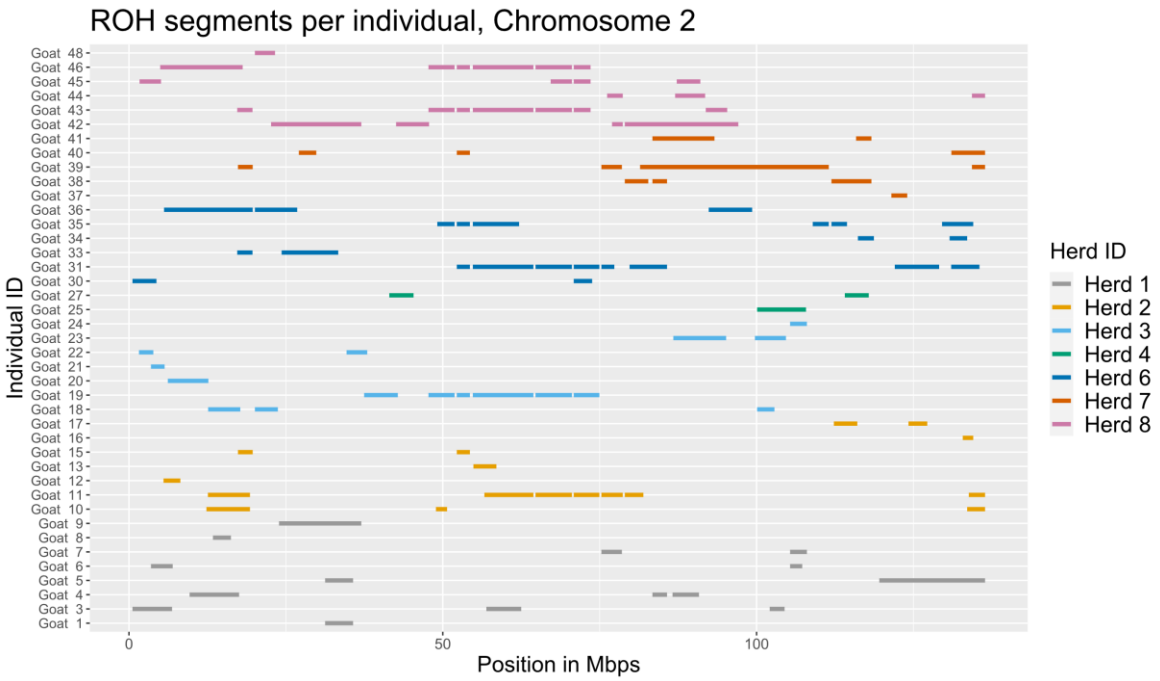


Figure 17 ROH on Chromosome 2. The individuals that do not show up in the plot did not have a ROH on this chromosome.

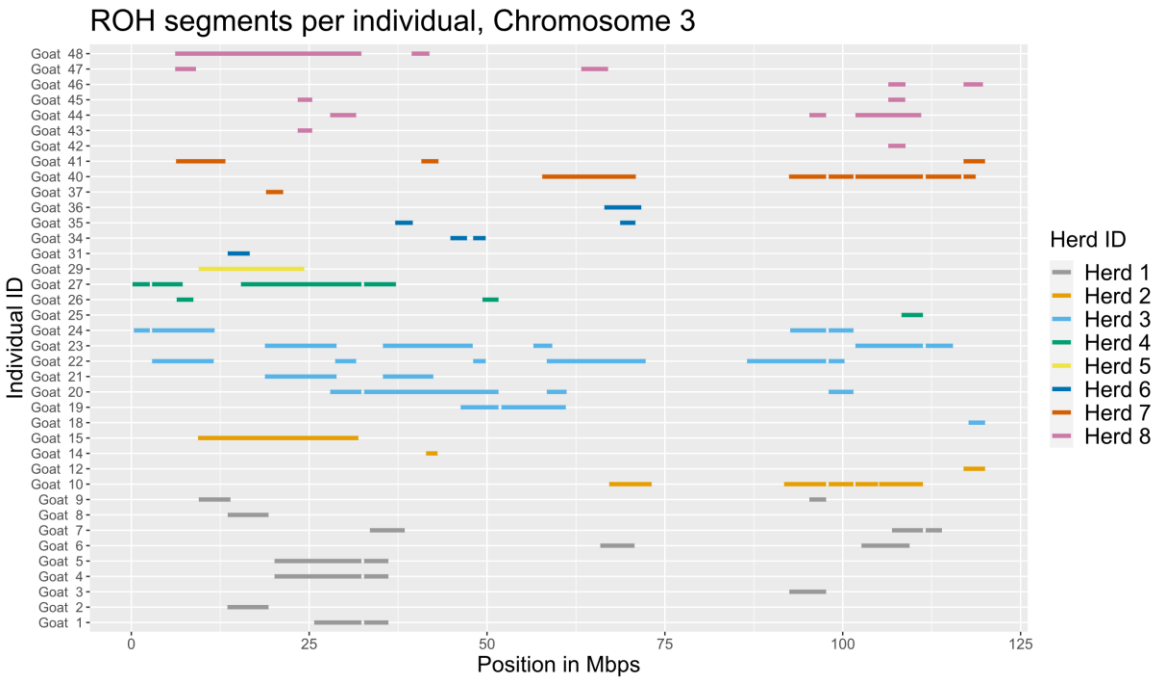


Figure 18 ROH on Chromosome 3. The individuals that do not show up in the plot did not have a ROH on this chromosome.

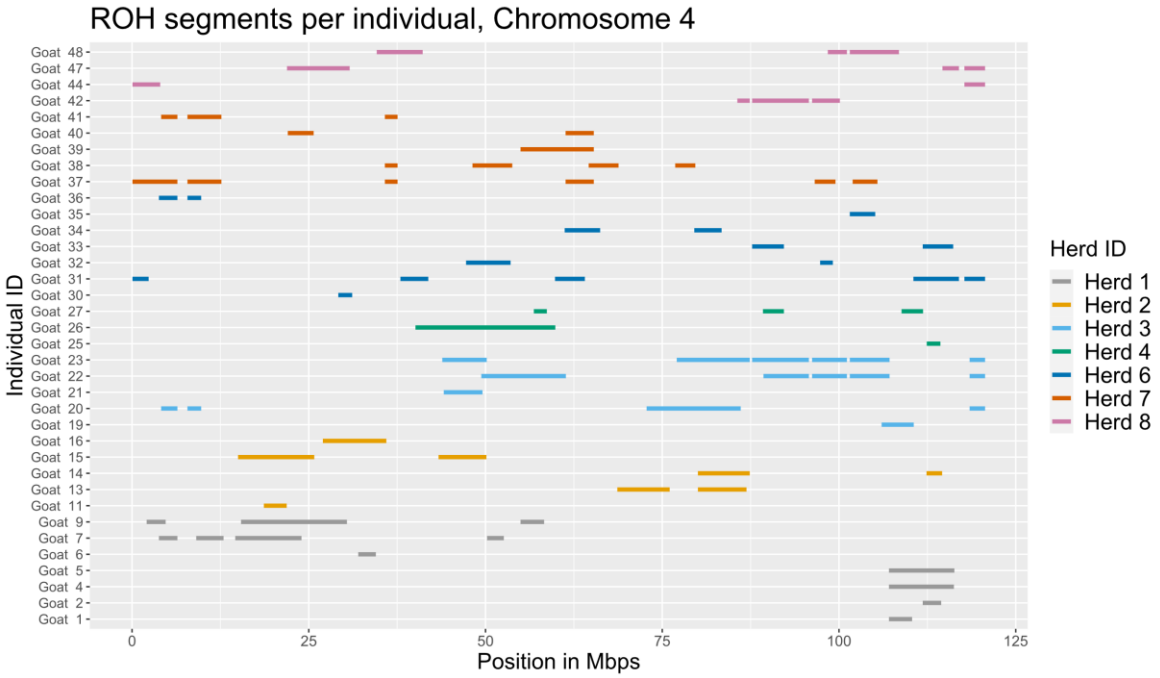


Figure 19 ROH on Chromosome 4. The individuals that do not show up in the plot did not have a ROH on this chromosome.

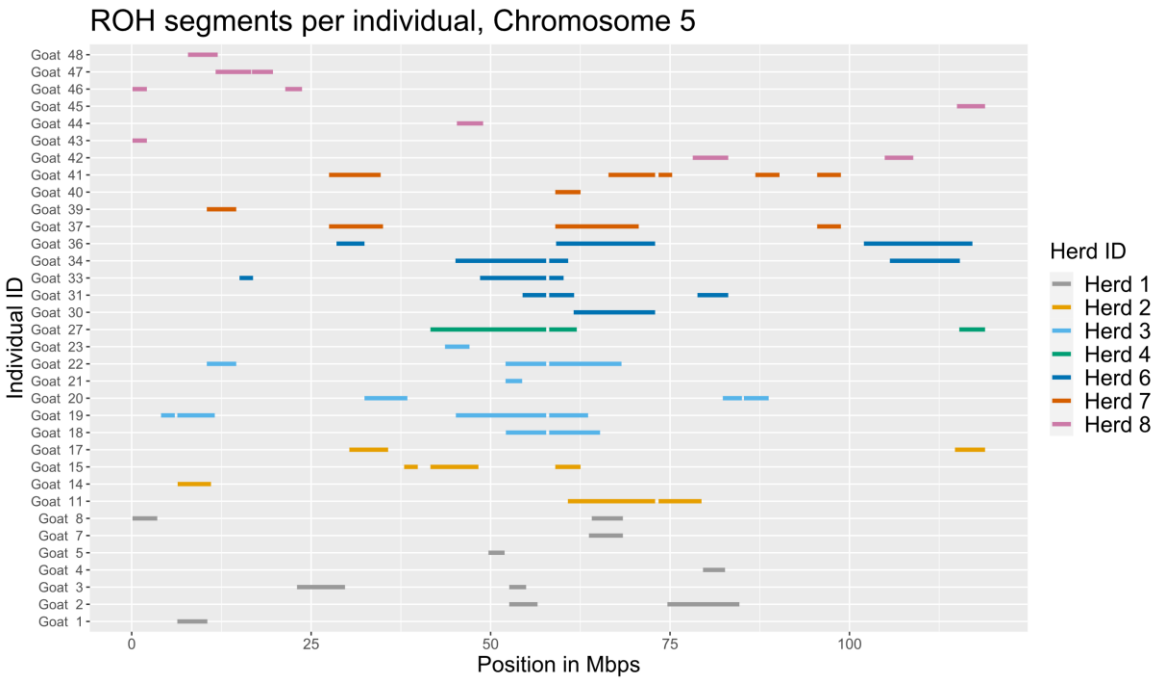


Figure 20 ROH on Chromosome 5. The individuals that do not show up in the plot did not have a ROH on this chromosome.

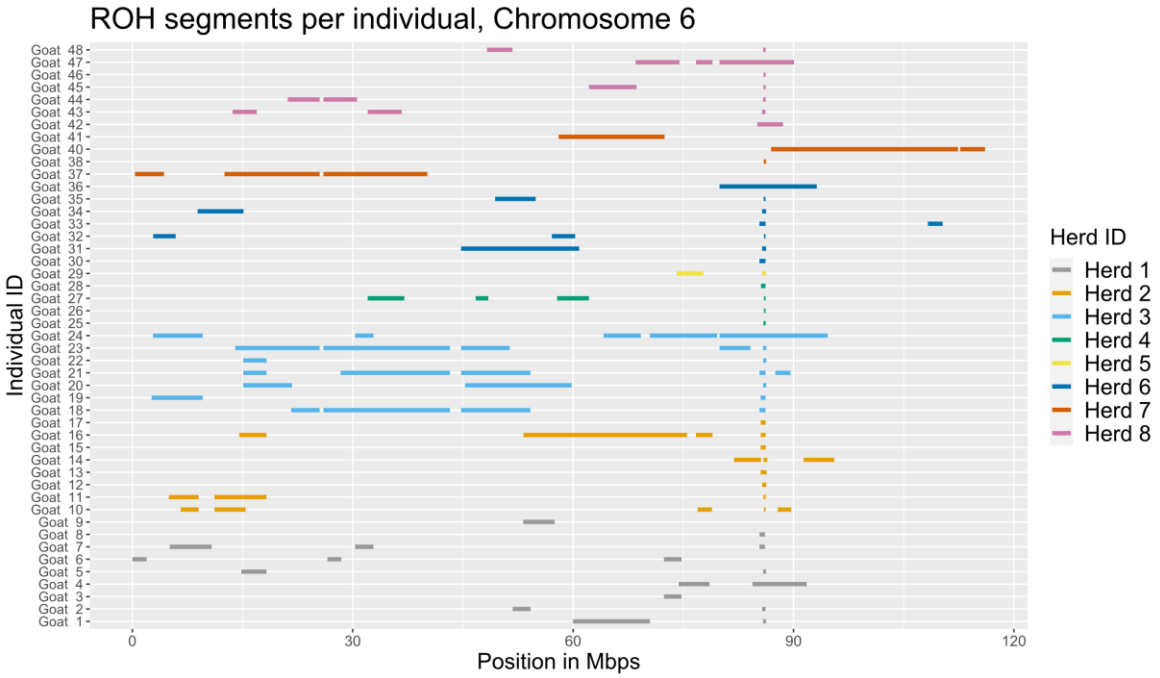


Figure 21 ROH on Chromosome 6. The individuals that do not show up in the plot did not have a ROH on this chromosome.

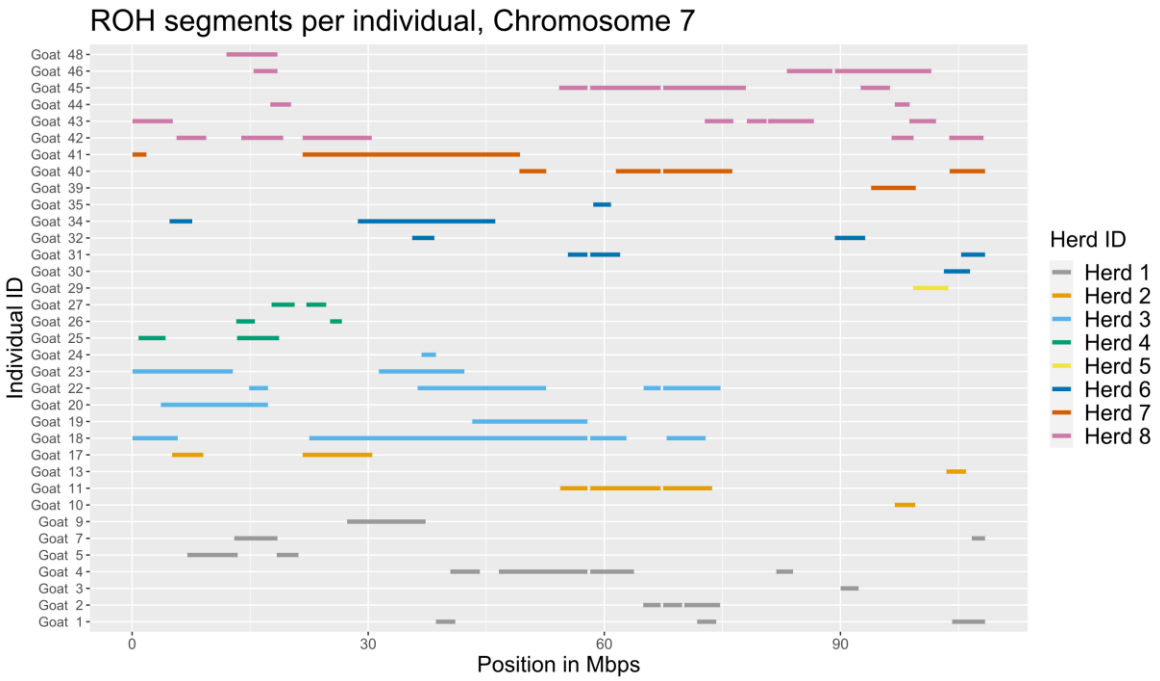


Figure 22 ROH on Chromosome 7. The individuals that do not show up in the plot did not have a ROH on this chromosome.

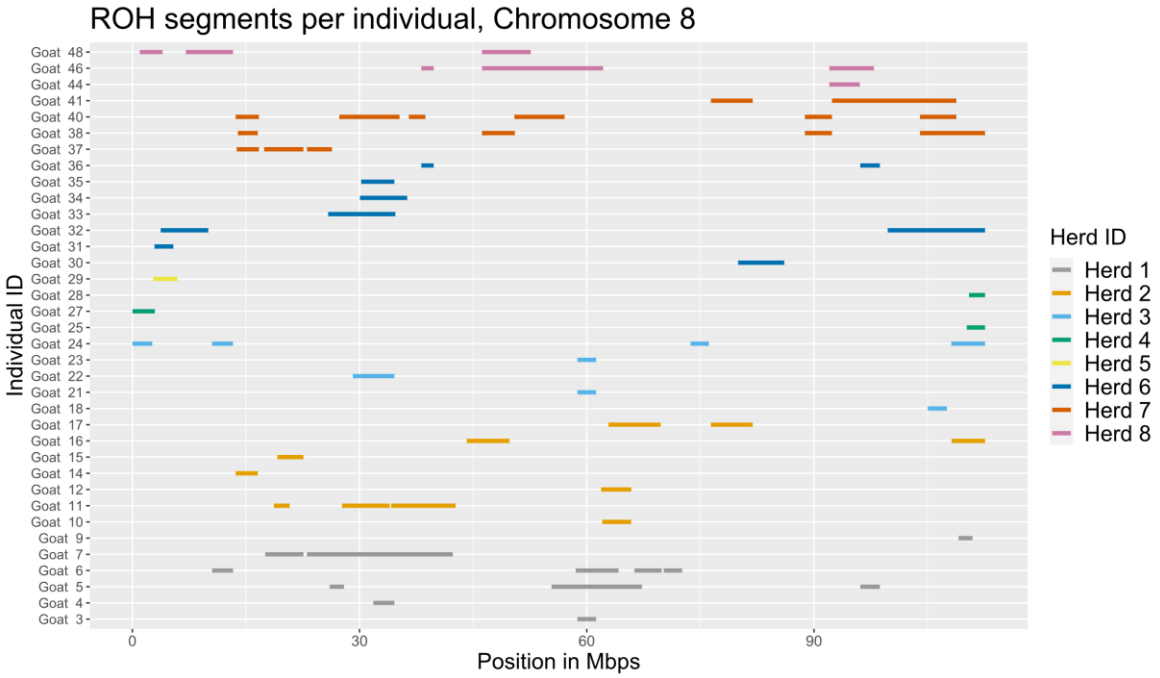


Figure 23 ROH on Chromosome 8. The individuals that do not show up in the plot did not have a ROH on this chromosome.

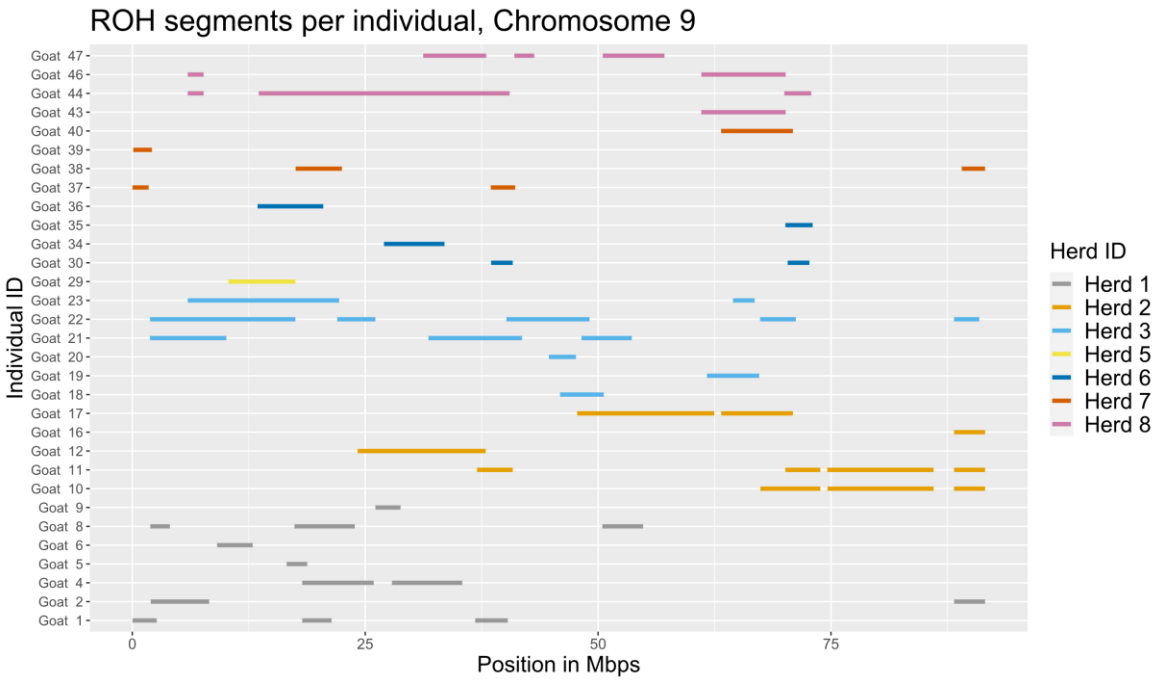


Figure 24 ROH on Chromosome 9. The individuals that do not show up in the plot did not have a ROH on this chromosome.

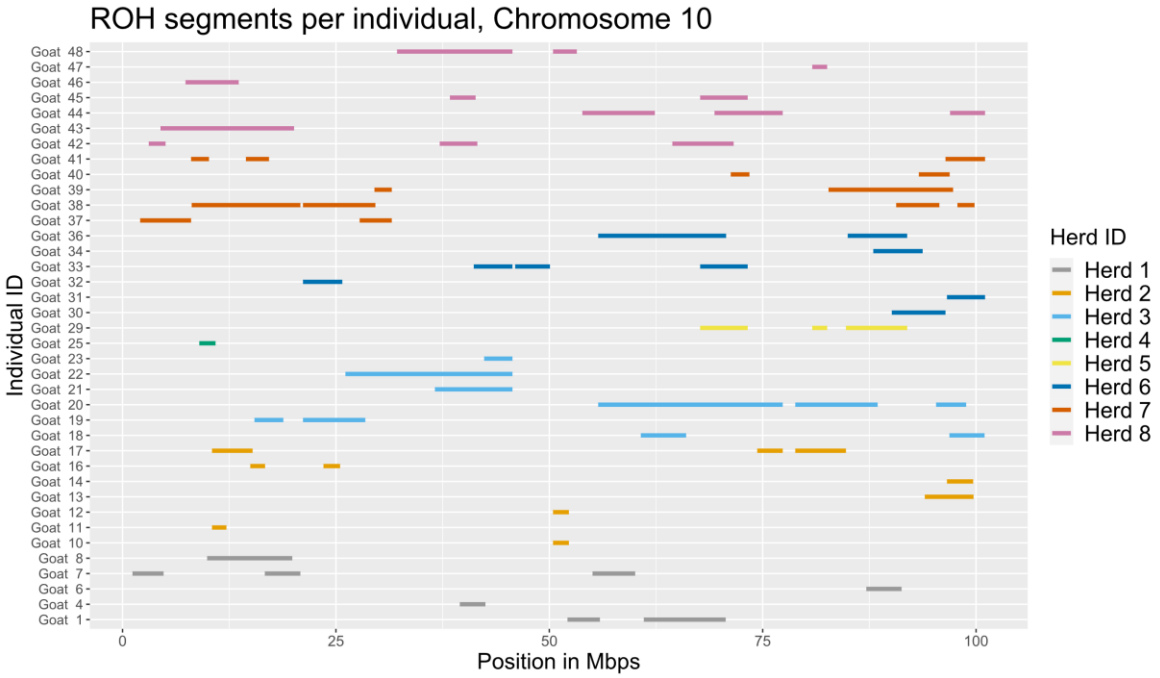


Figure 25 ROH on Chromosome 10. The individuals that do not show up in the plot did not have a ROH on this chromosome.

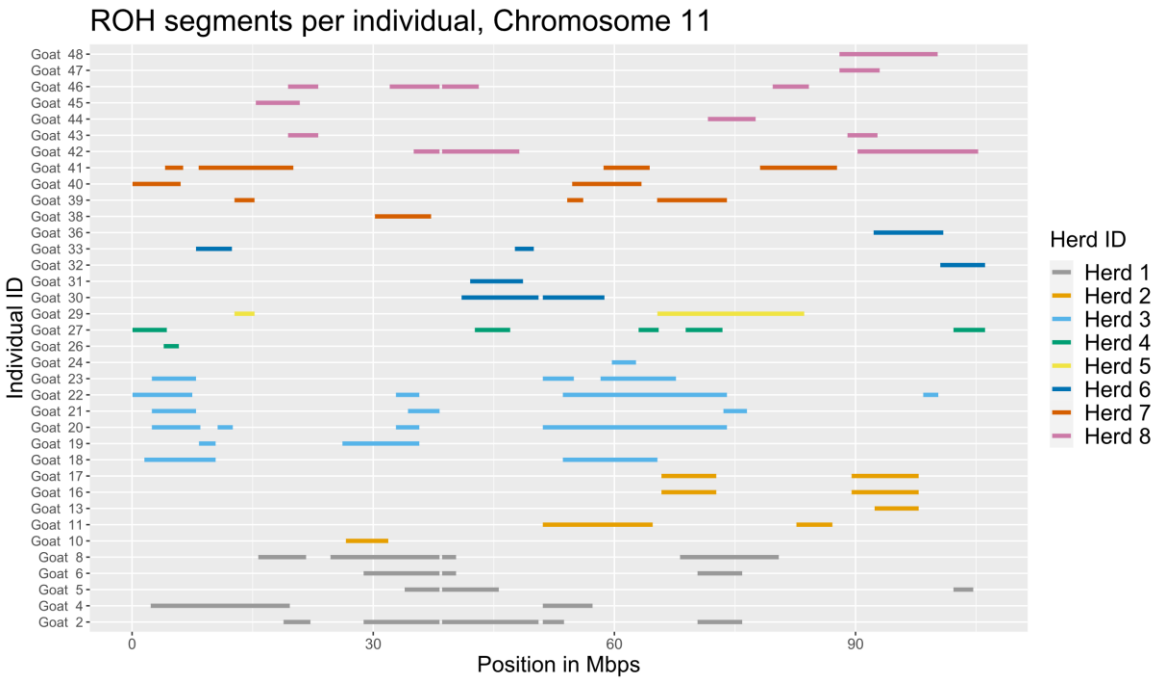


Figure 26 ROH on Chromosome 11. The individuals that do not show up in the plot did not have a ROH on this chromosome.

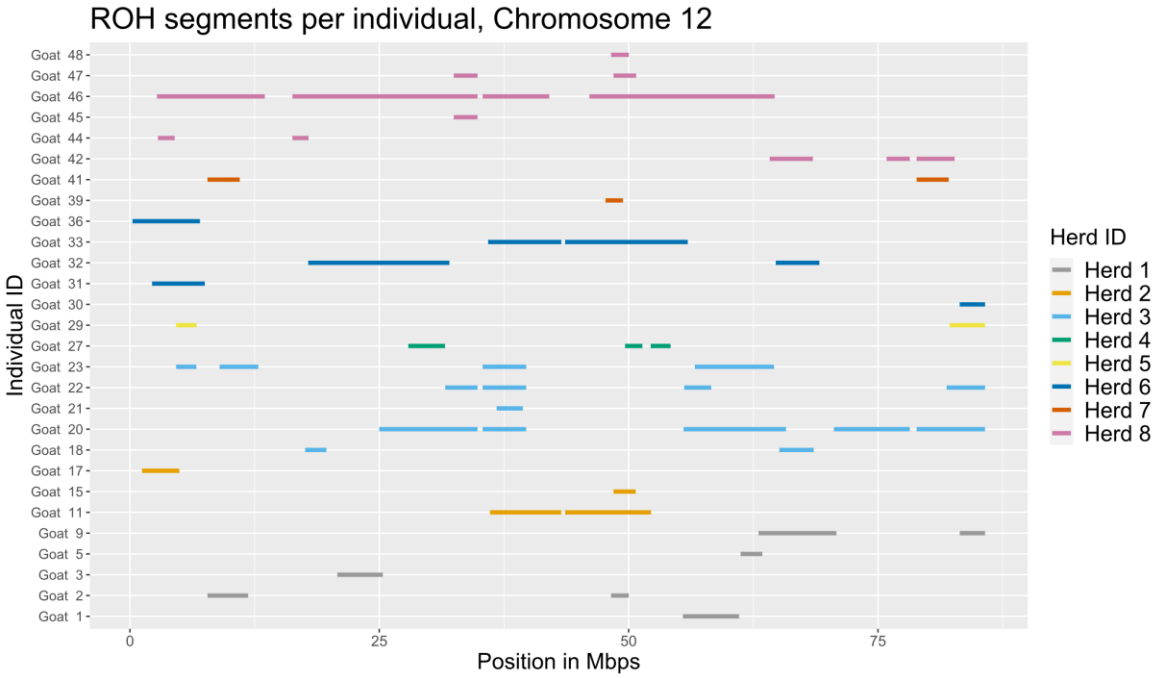


Figure 27 ROH on Chromosome 12. The individuals that do not show up in the plot did not have a ROH on this chromosome.

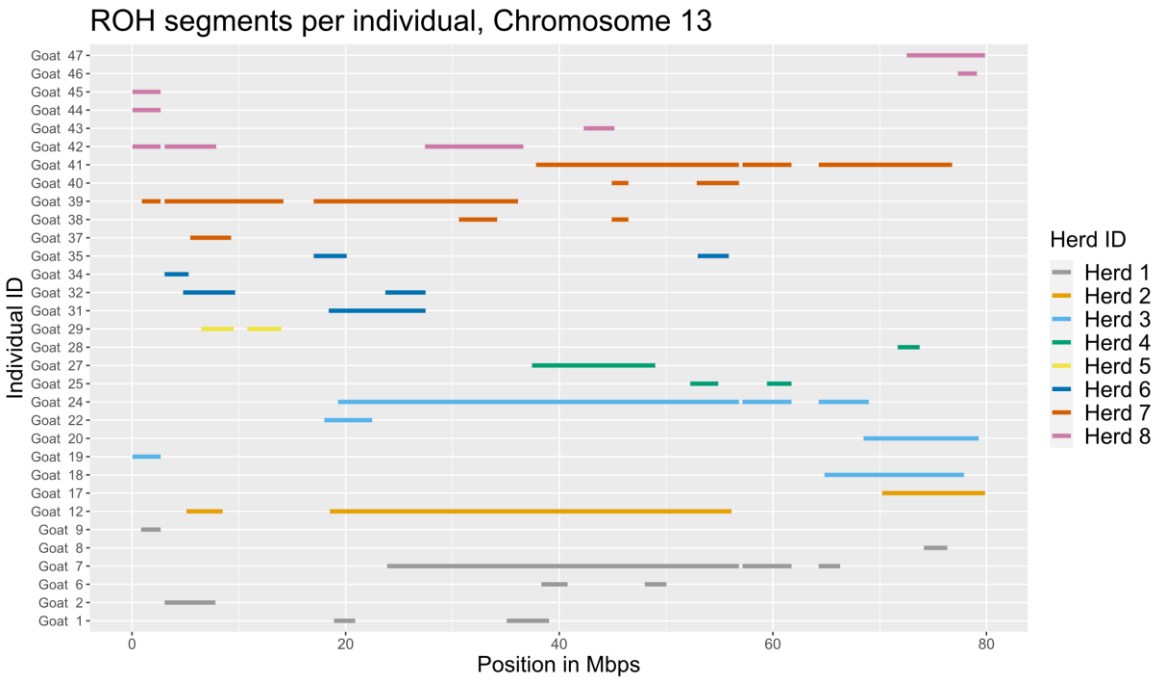


Figure 28 ROH on Chromosome 13. The individuals that do not show up in the plot did not have a ROH on this chromosome.

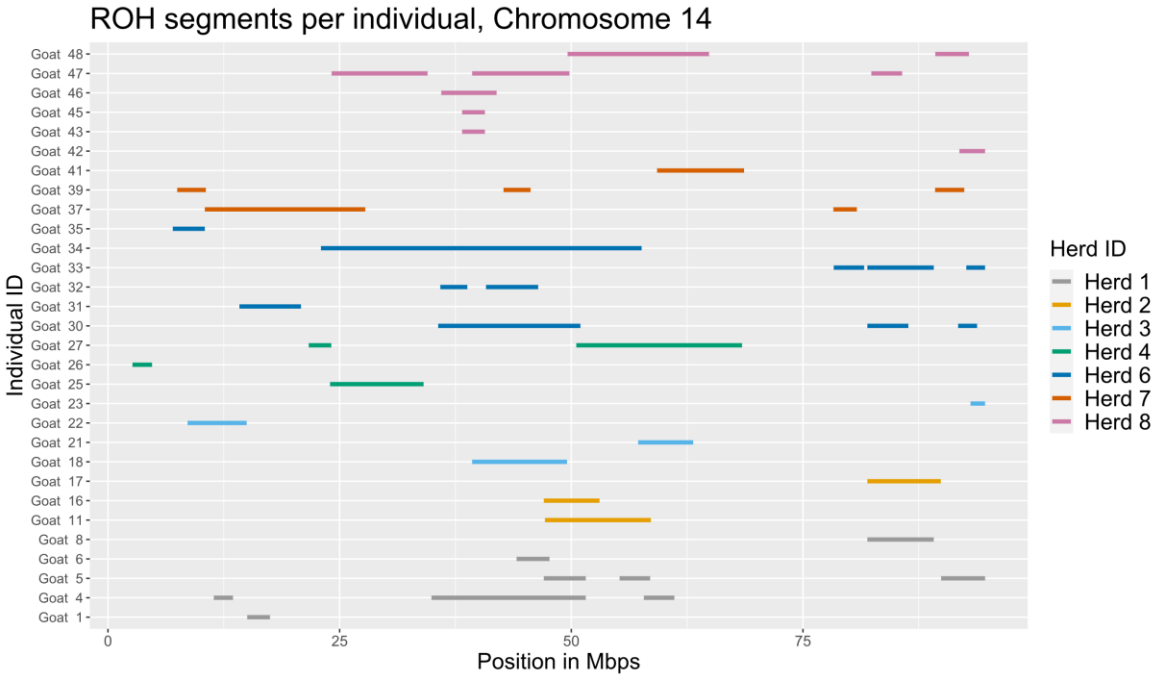


Figure 29 ROH on Chromosome 14. The individuals that do not show up in the plot did not have a ROH on this chromosome.

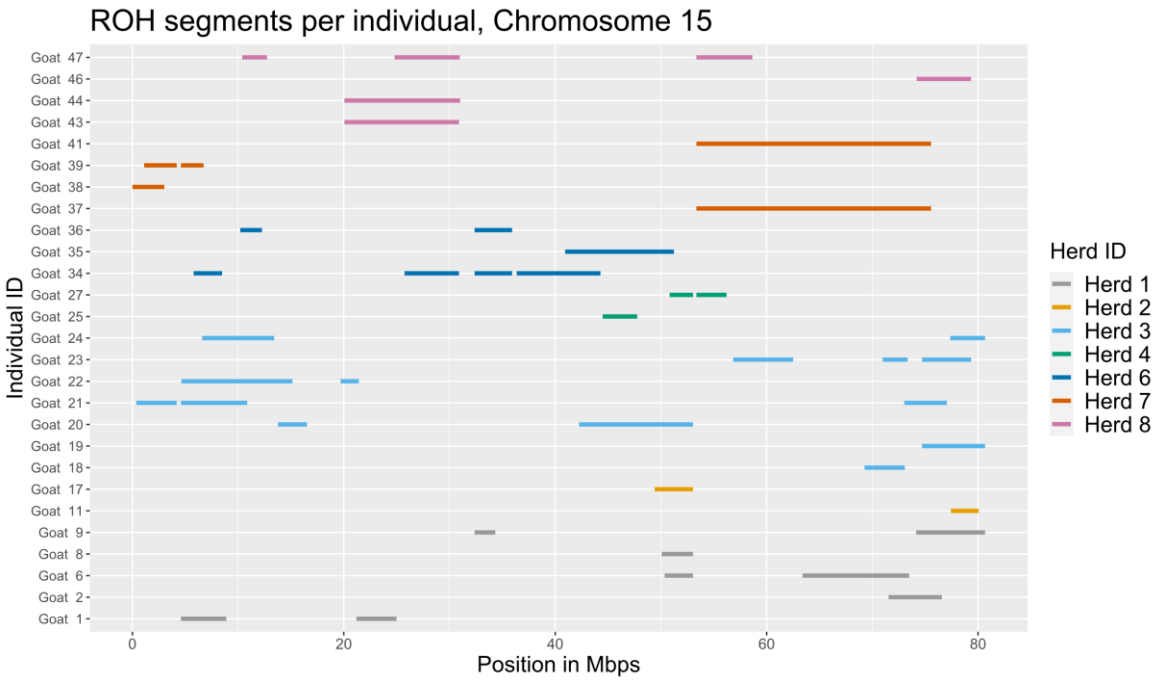


Figure 30 ROH on Chromosome 15. The individuals that do not show up in the plot did not have a ROH on this chromosome.

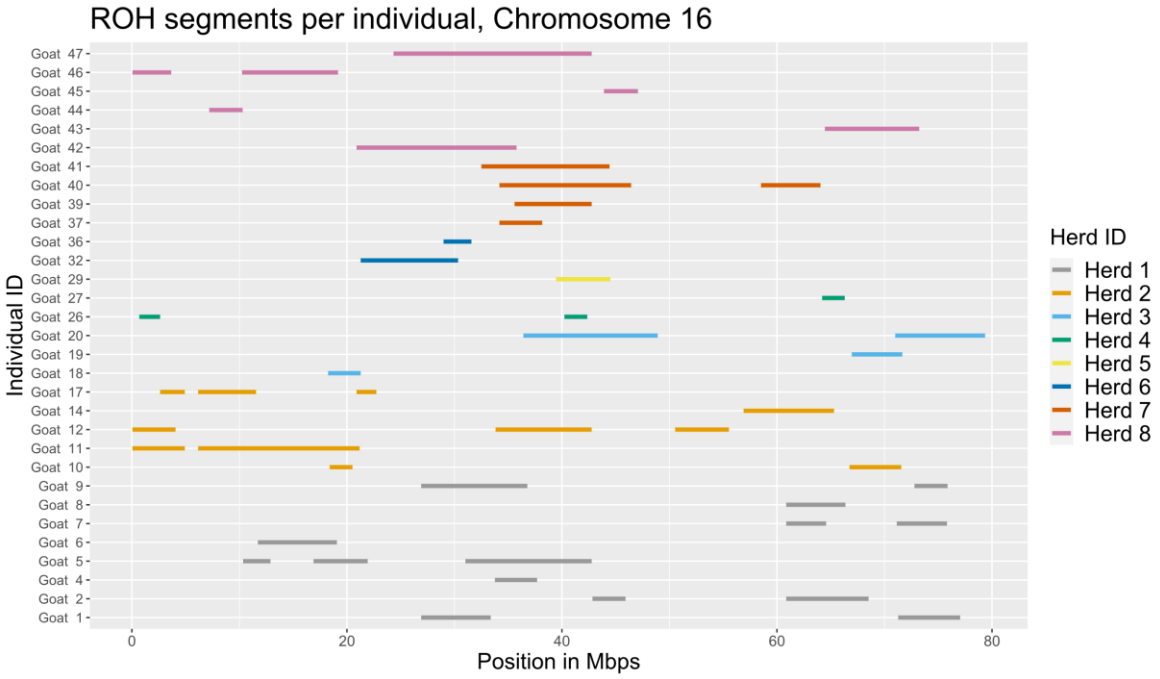


Figure 31 ROH on Chromosome 16. The individuals that do not show up in the plot did not have a ROH on this chromosome.

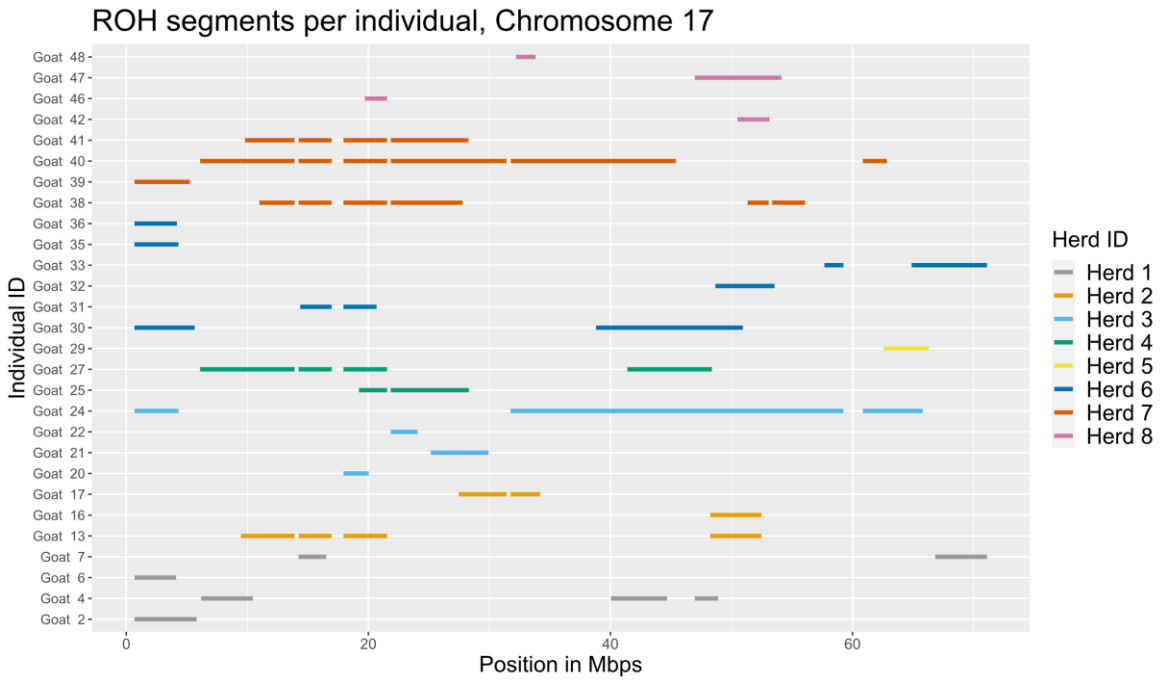


Figure 32 ROH on Chromosome 7. The individuals that do not show up in the plot did not have a ROH on this chromosome.

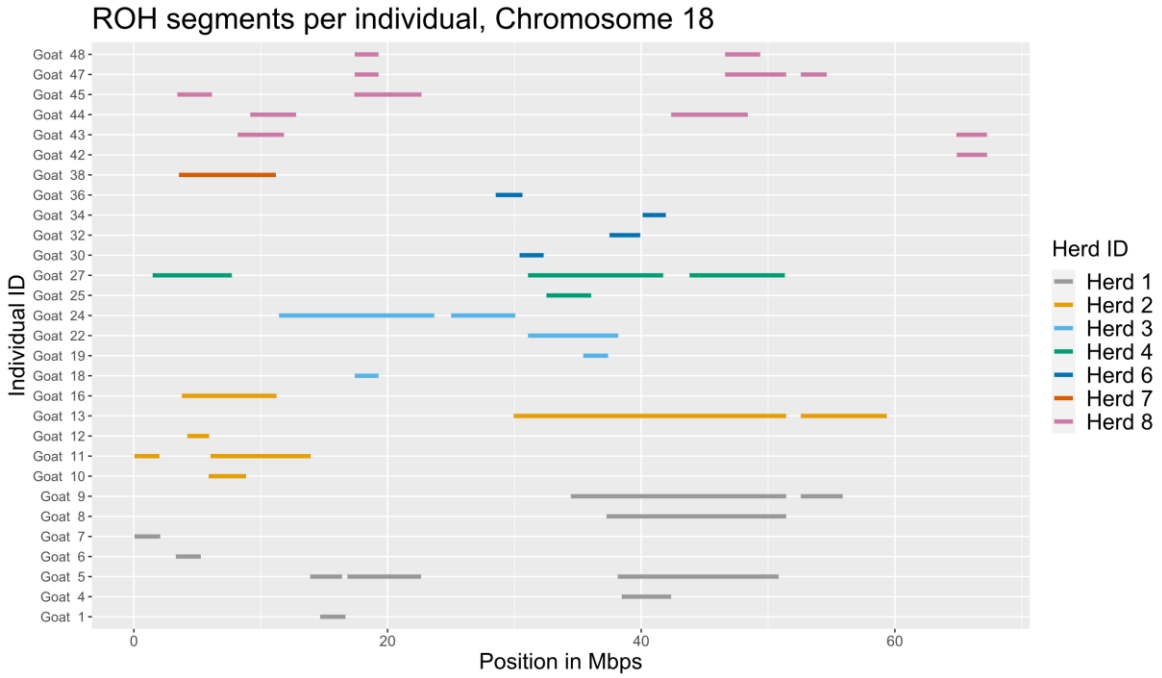


Figure 33 ROH on Chromosome 18. The individuals that do not show up in the plot did not have a ROH on this chromosome.

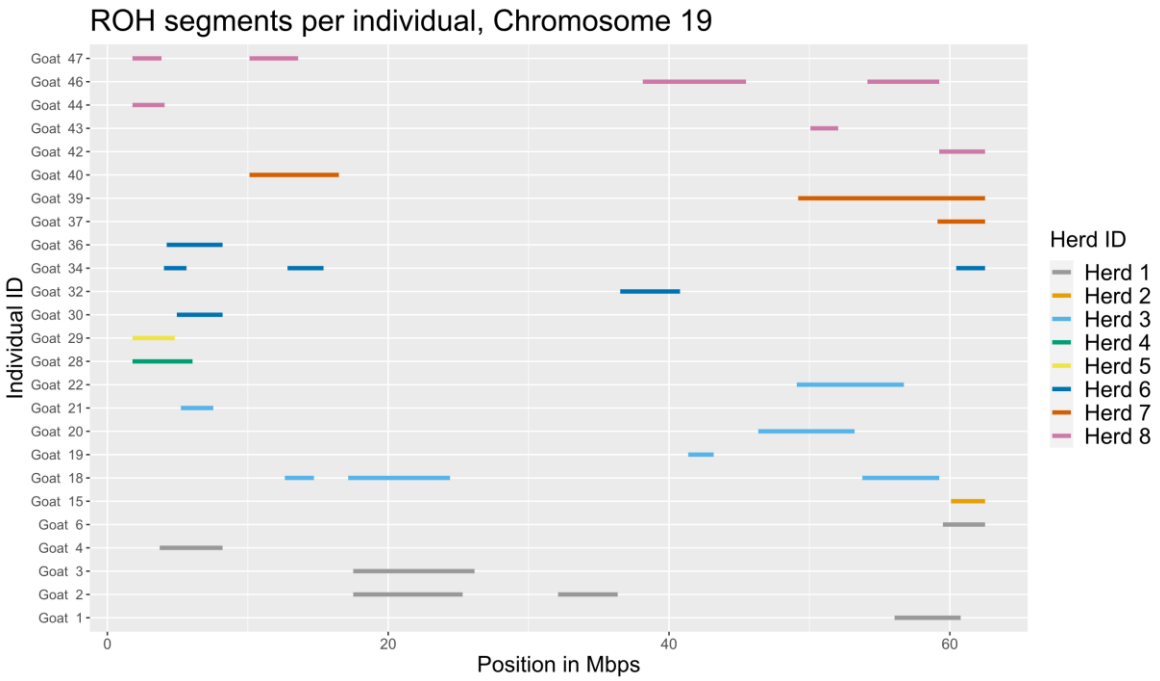


Figure 34 ROH on Chromosome 19. The individuals that do not show up in the plot did not have a ROH on this chromosome.

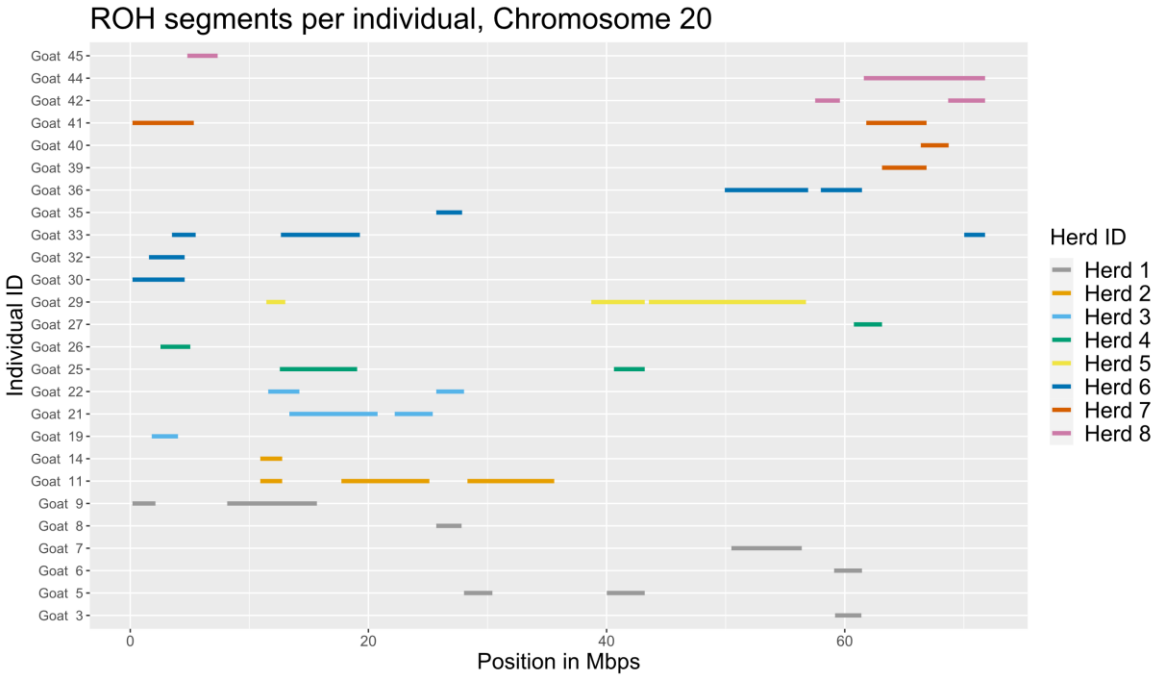


Figure 35 ROH on Chromosome 20. The individuals that do not show up in the plot did not have a ROH on this chromosome.

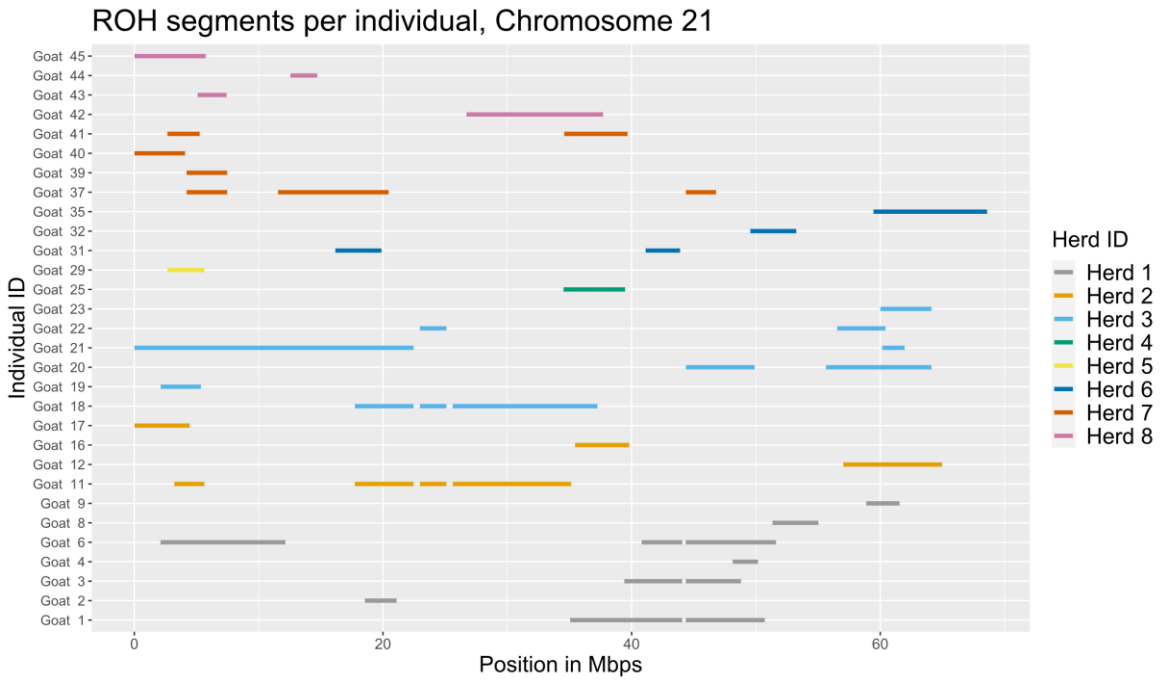


Figure 36 ROH on Chromosome 21. The individuals that do not show up in the plot did not have a ROH on this chromosome.

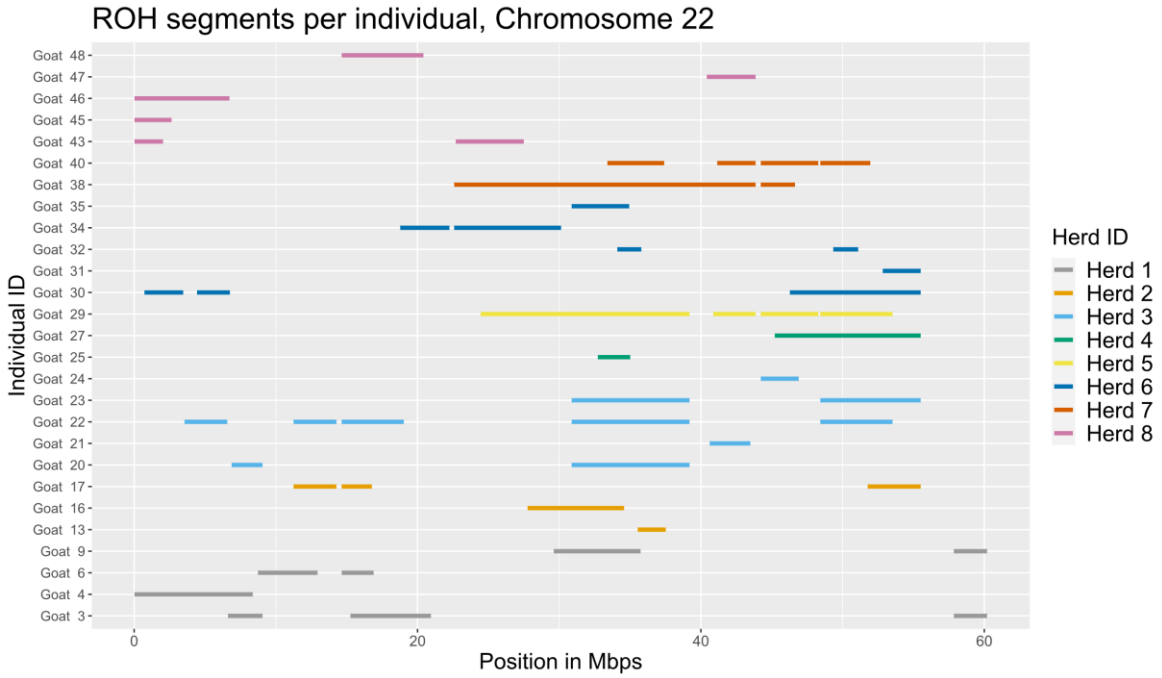


Figure 37 ROH on Chromosome 22. The individuals that do not show up in the plot did not have a ROH on this chromosome.

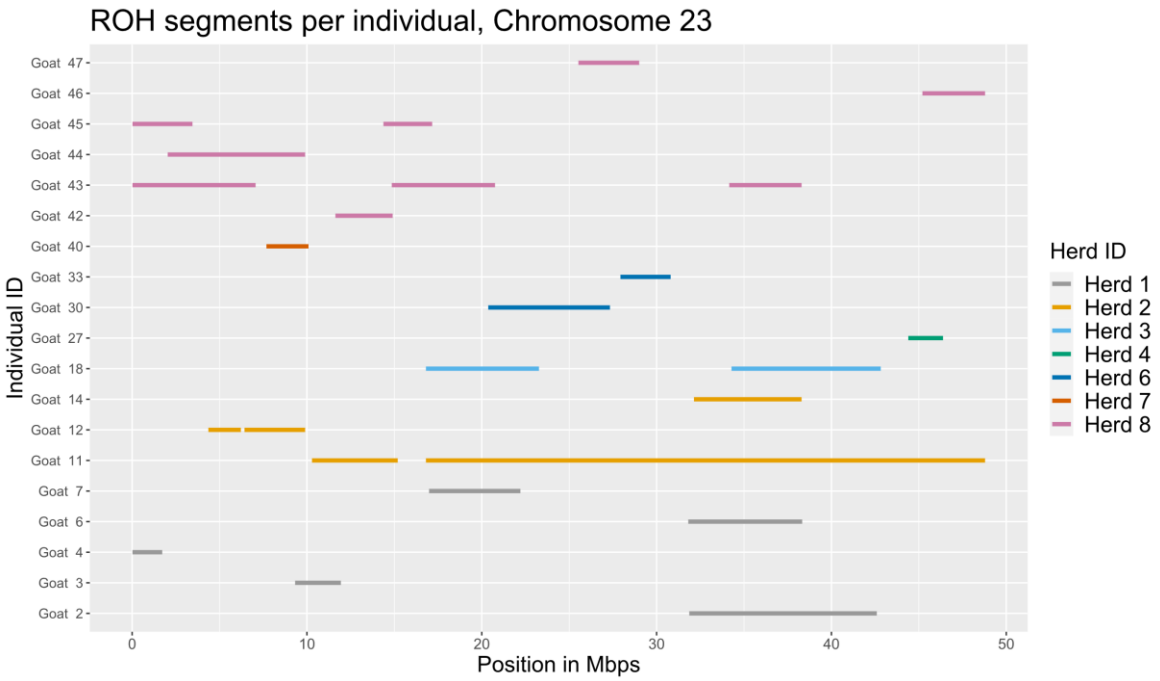


Figure 38 ROH on Chromosome 23. The individuals that do not show up in the plot did not have a ROH on this chromosome.

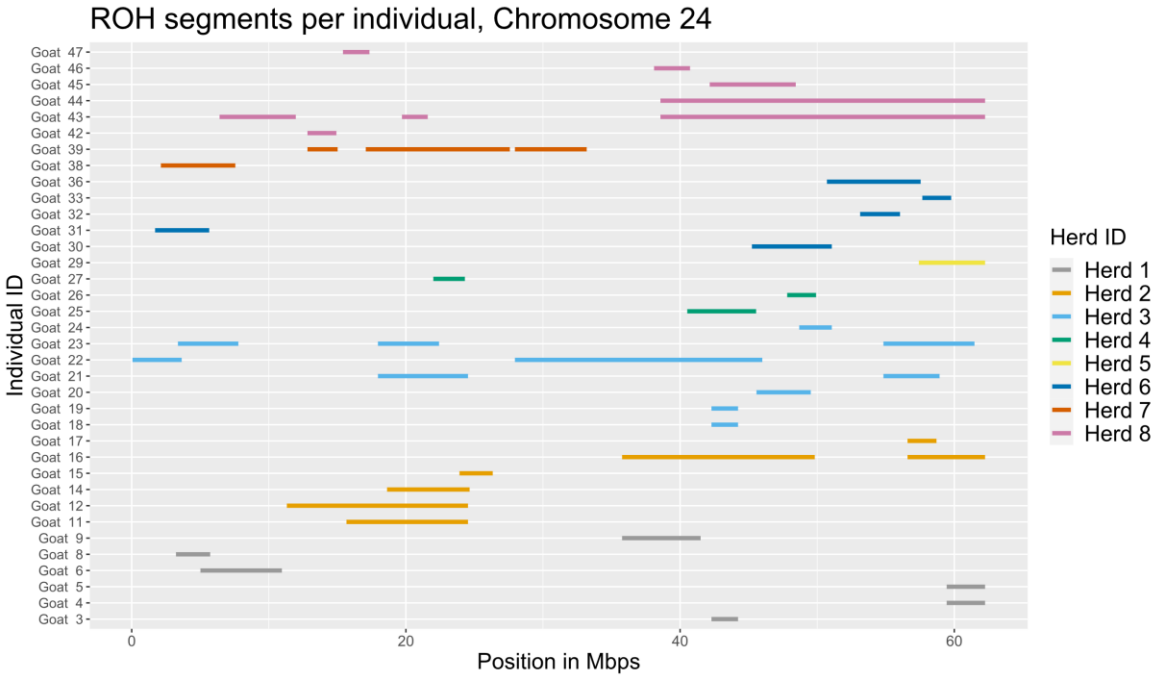


Figure 39 ROH on Chromosome 24. The individuals that do not show up in the plot did not have a ROH on this chromosome.

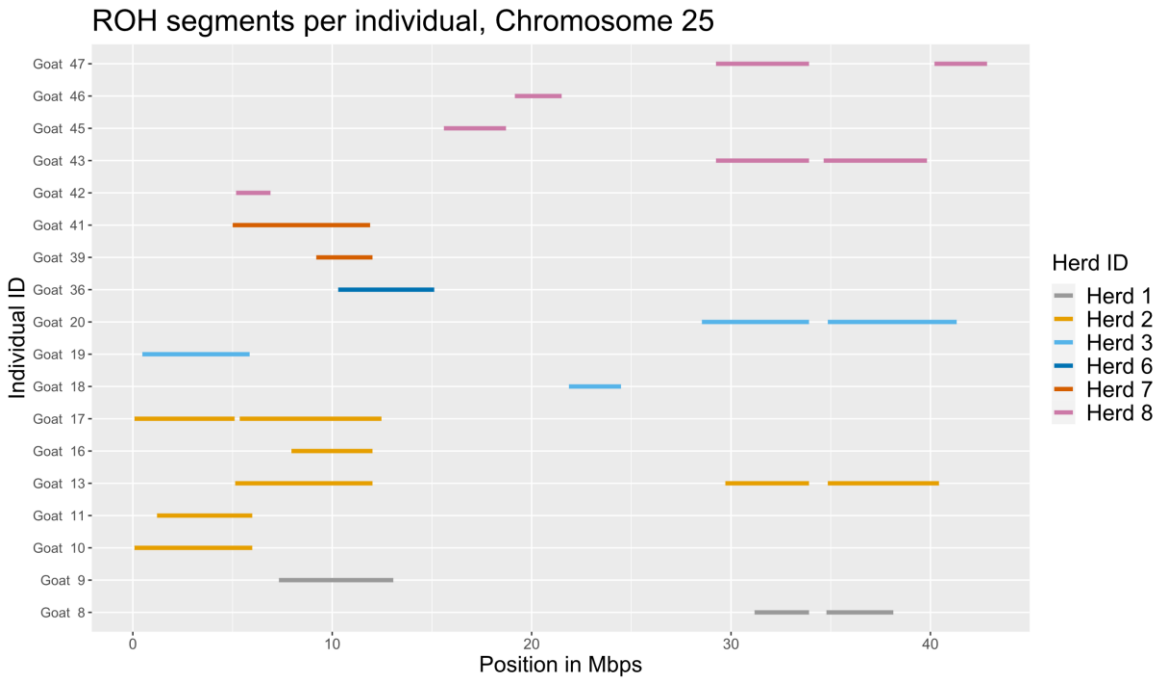


Figure 40 ROH on Chromosome 25. The individuals that do not show up in the plot did not have a ROH on this chromosome.

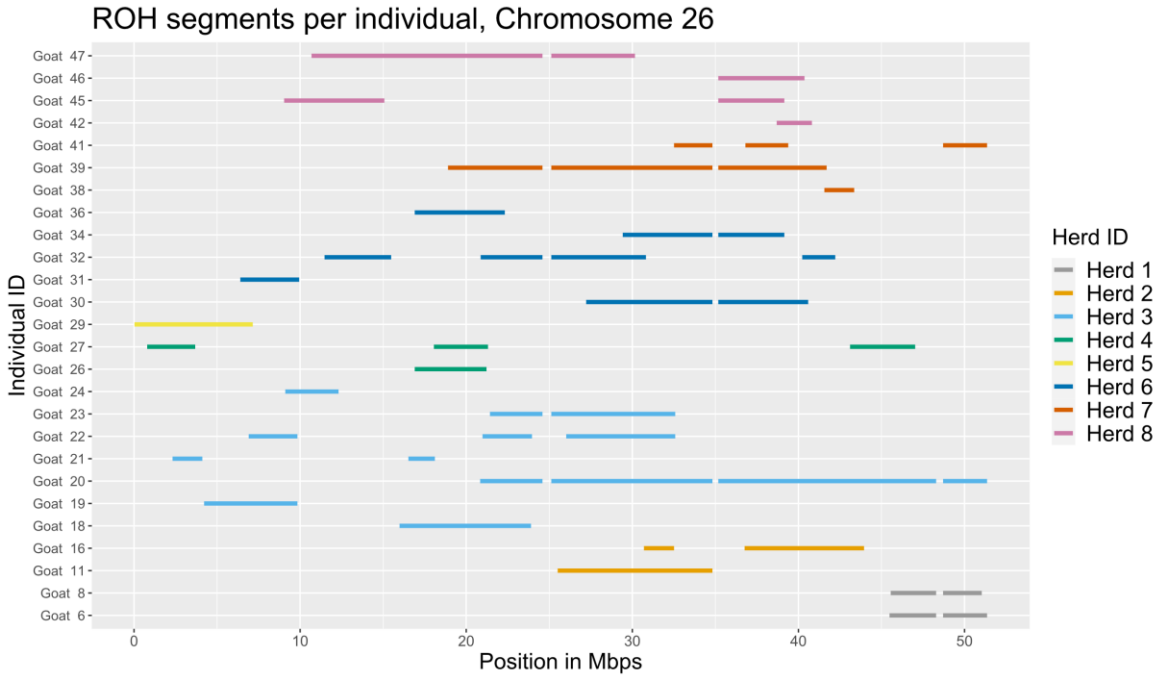


Figure 41 ROH on Chromosome 26. The individuals that do not show up in the plot did not have a ROH on this chromosome.

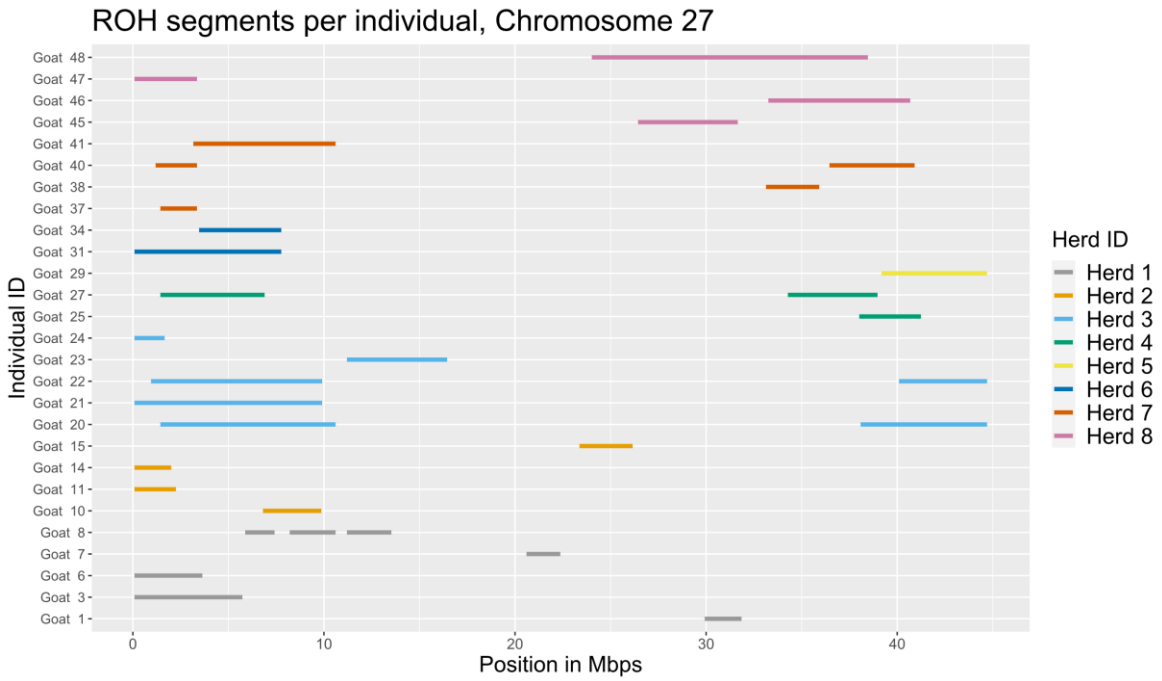


Figure 42 ROH on Chromosome 27. The individuals that do not show up in the plot did not have a ROH on this chromosome.

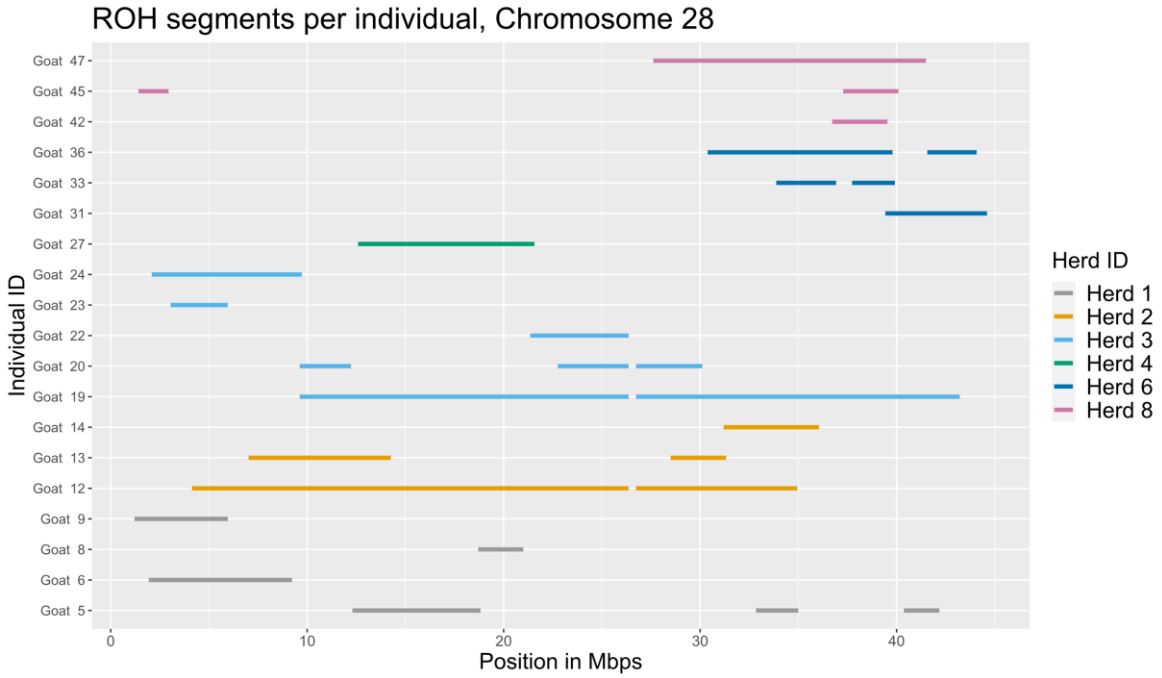


Figure 43 ROH on Chromosome 28. The individuals that do not show up in the plot did not have a ROH on this chromosome.

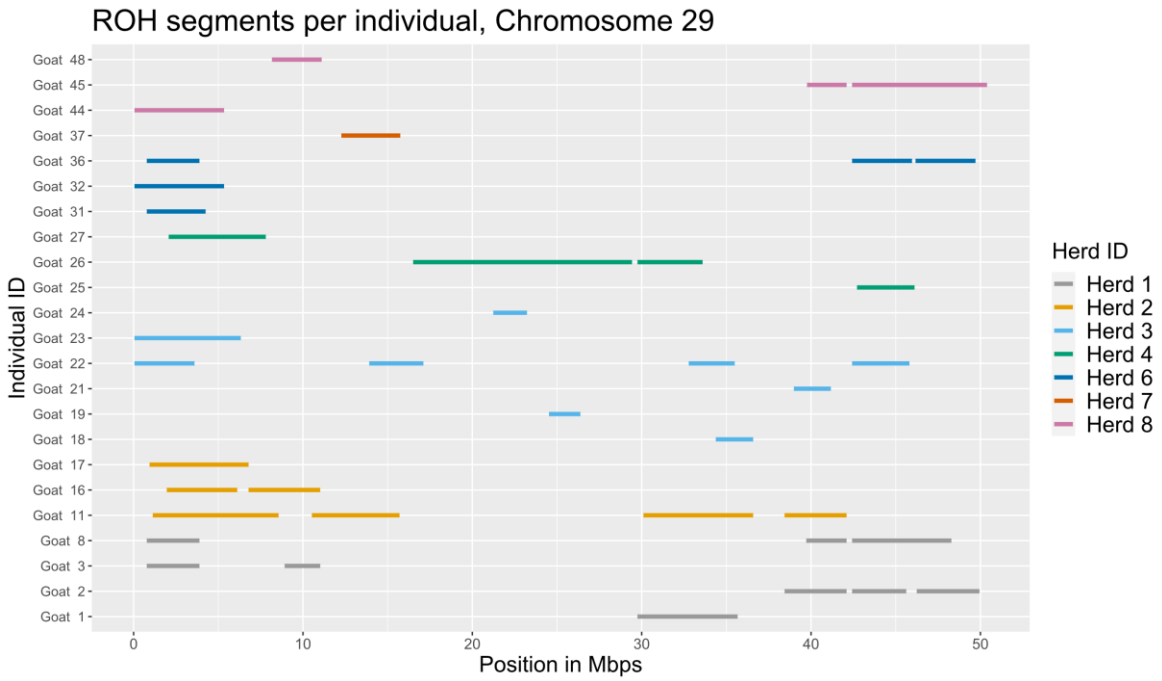


Figure 44 ROH on Chromosome 29. The individuals that do not show up in the plot did not have a ROH on this chromosome.

Appendix 4 – Data Management Plan

Data management plan belonging to the MSc thesis performed at the Animal Breeding and Genetics Department (SLU) by Bernadett Hegedüs, completed in March 2022.

Agreements

1. The data used in this thesis project have been described in this document and have been stored in a systematic manner (at least in separate folders for all sections as described below). Data includes all data as mentioned in the results section of your report.
2. The data management plan has been discussed with the MSc thesis supervisor and he/she has agreed on the location for data storage.
3. In case of confidentiality, contact details of the responsible person from the company/institution that has ownership of the data are mentioned in this document.
4. **The data and the scripts for data analysis can be found through Bernadett Hegedüs (bernadett.hegedus@wur.nl) and Anna Maria Johansson (anna.johansson@slu.se).**

Section A – Input files

File names	Created in	Remarks
Swedish_Univ_Eriksson_GOAT53KV1_20200722_FinalReport .txt	July, 2020	Genotype data from 48 goats

Comment: The raw genotype data uses the original farm names, so it is confidential.

Section B – Data analysis and plots for the results section

File names	Created in	Remarks
Convert_to_plink_goat_SNP.R	2021	Script that creates input files for Plink from raw data, mostly written by Anna Johansson.
Goat_analysis.R	2022	Script with headings according to the different analyses, also includes Plink commands and code for plots for PCoA, observed heterozygosity and ROH analysis. The calculation of driving distances between farms and the plotting of farms on the map of Sweden are also found in this script.
Goat_functions.R	2022	The script Goat_analysis sometimes calls functions; these can be found in this script.
Structure_plots.R	2022	This script creates the plots seen in the report from the output of the Software Structure.

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. Read about SLU's publishing agreement here:

- https://www.slu.se/en/subweb/library/publish-and-analyse/register-and-publish/agreement-for-publishing/.

☒ 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.