Genomic indicators of diversity in Austrian horse populations

Master's Thesis

for obtaining the academic degree Master of Science
in Livestock Science

Maria STRASSER, BSc
1145101

Supervisor:
Univ. Prof. Dipl.-Ing. Dr. Johann Sölkner

Co-Supervisor:
Ass.Prof. Dr. Gábor Mészáros

Department of Sustainable Agricultural Systems
Division of Livestock Science

Vienna, December 2017

University of Natural Resources and Life Sciences, Vienna

Declaration in lieu of oath

I herewith declare in lieu of oath that this thesis has been composed by myself without any inadmissible help and without the use of sources other than those given due reference in the text and listed in the list of references. I further declare that all persons and institutions that have directly or indirectly helped me with the preparation of the thesis have been acknowledged and that this thesis has not been submitted, wholly or substantially, as an examination document at any other institution.

Date	Signature

Acknowledgements

This master thesis was carried out in the year 2017 at the University of Natural Resources and Life Sciences, Vienna, at the department of Sustainable Agricultural Systems.

I would like to express my special thanks of gratitude to my co-supervisor Dr. Gábor Mészáros, who encouraged and guided me throughout the entire process of writing this thesis and without his technical support this work would not have been possible. I am also deeply grateful to my supervisor Dr. Johann Sölkner for his guidance and enthusiasm.

I am grateful to Keerthy for proofreading my work with tremendous motivation and passion. I would also like to thank Renan Hanada for always pushing me and being my source of motivation. Finally, I must express my very profound gratitude to my parents, Elfriede and Andreas Strasser for their unfailing support, patience and continuous encouragement throughout my years of study. This accomplishment would not have been possible without them.

Thank you.

Maria Strasser

List of abbreviations

Abbreviation	Explanation
ARGE	working group/consortium (Arbeitsgemeinschaft)
BMLFUW	The Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management (Bundesministerium für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft)
DNA	deoxyribonucleic acid
FAO	Food and Agriculture Organization
F_{ST}	fixation index
Gb	giga base pairs
HAF	Haflinger
kb	kilo base pairs
LD	linkage disequilibrium
MAF	minor allele frequency
Mb	mega base pairs
NOR	Noriker
PCA	principal component analysis
r^2	squared coefficient of correlation
SNP	single-nucleotide polymorphism
ZAP	Austrian Horse Breeding Association (Zentrale Arbeitsgemeinschaft Österreichischer Pferdezüchter)

Abstract

The promotion of health and welfare for horses is closely linked to the maintenance of genetic diversity. It is crucial to comprehend genetic diversity and relationships within and between populations in order to support constant development of the breed and to avoid a decline in animal genetic resources.

Genotypic information from a high density BeadChip is used to analyse and compare genetic variability of two Austrian horse breeds (Haflinger and Noriker). The genetic structure of Haflinger and Noriker populations is assessed through a principal components analysis based on distances between individuals, expressed as allele counts. The F_{ST} values are assessed. Nei's genetic distances between populations and between individuals are calculated and estimations of identity by state and identity by descent scores are determined. F are calculated based upon observed and expected autosomal homozygous genotype counts and are compared with F based on runs of homozygosity.

The principal components analysis shows a clear distinction between the two Austrian horse breeds, however, according to the pairwise F_{ST} between Haflinger and Noriker populations, there is only moderate genetic differentiation ($F_{ST} = 0.085$). Estimation of the mean identity by state scores shows similar values for both breeds, whereas identity by descent scores are higher in the Haflinger population (mean identity by descent score for Haflinger = 0.027, for Noriker = 0.015). F based upon observed and expected autosomal homozygous genotype counts express a higher heterozygosity than expected, although F based on runs of homozygosity indicate a recent decline in inbreeding for both breeds.

Indications showing possible connections between the Haflinger and Noriker breed have been detected which encourages fine mapping of their population structure in a follow up research.

Keywords: horse, genetic diversity, inbreeding, genetic relationships, runs of homozygosity

Zusammenfassung

Die Sicherstellung eines gewissen Maßes an genetischer Diversität trägt zur Verbesserung der Tiergerechtheit bei und gewährleistet auf lange Sicht den Fortbestand einer gesunden Pferdepopulation. Kenntnisse über das Ausmaß der genetischen Diversität und Verwandtschaft einer Population sind von entscheidender Bedeutung, um einem Verfall der genetischen Ressourcen nachhaltig entgegenwirken zu können.

In dieser Studie wurden molekulare Marker zur Bestimmung der genetischen Diversität zweier österreichischer Pferderassen (Haflinger und Noriker) herangezogen. Die genetische Strukturierung der Haflinger- und Norikerpopulationen wird durch eine Hauptkomponentenanalyse, basierend auf genetischen Distanzen, dargestellt. F_{ST}-Werte und Neis genetische Distanz zwischen Populationen und Individuen wurden berechnet und die Werte für „Identity-by-state“ und „Identity-by-descent“ (IBD) geschätzt. Des Weiteren wurden Inzuchtkoeffizienten berechnet und lange homozygote Strecken analysiert.

Die Hauptkomponentenanalyse trennt die beiden Pferderassen deutlich voneinander, allerdings weisen die F_{ST}-Werte lediglich auf eine moderate Differenzierung hin (F_{ST} = 0.085). Die Schätzung des Anteils der IBD in beiden Pferdepopulationen ergibt höhere Mittelwerte für die Haflingerpopulation (Haflinger = 0.027 und Noriker = 0.015). Die Schätzung der Inzuchtkoeffizienten deutet auf einen Rückgang innerhalb der letzten Jahre in beiden Populationen hin. Hinweise auf mögliche Verbindungen der Haflinger- und Norikerrasse wurden gefunden, wobei die Ergebnisse die Dringlichkeit für weitere Forschung zeigen.

Schlüsselwörter: Pferd, genetische Diversität, Inzucht, lange Regionen von Homozygotie

Table of contents

1 Introduction

Genetic diversity plays a major role in global food security. A high level of diversity is not only required to avoid inbreeding depression and to guaranty a certain adaptation potential, it is crucial for the survival of a breed. A decline in genetic diversity can result in a higher disease susceptibility and a reduction of production traits. A closer examination of the term "genetic diversity" is conducted in Chapter 2.2 of this thesis and further specifications regarding equine genetic diversity are included as well.

Genetic diversity in horse populations has already been investigated by various studies, which are mentioned in Chapter 2.2.1. However, there has not been an analysis of genomic indicators of diversity in Austrian horse populations prior to this thesis.

The present state of Austria's horse population is described in Chapter 2. In addition, Chapter 3 contains detailed information about the featured breeds (Haflinger and Noriker), and a description of the utilized data and methods. In Chapter 4, the results are presented and visualized. Furthermore, the outcome is reviewed and a discussion is included. Finally, in Chapter 5 the conclusion is reached.

1.1 Aim of the thesis

The aim of this thesis is to investigate the genetic diversity of two Austrian horse breeds (Haflinger and Noriker) based on SNP data from high density BeadChips.

2 Review of literature

2.1 Horses – facts and figures

Austria's horse population consists of 87,000 animals and contributes 0.15 % to the global horse population of 58,832,221 (Food and Agriculture Organization of the United Nations (FAO) 2014). According to the Institute for Industrial Research (2005), Austria's horse-associated industry is generating an annual production valued between 1.19 and 1.26 billion Euros and therefore constitutes an important economic factor for the country. Around 24,300 jobs are created by the equine sector, that is to say, three to four horses secure one employment contract. With regards to the number of employment contracts and overall production, the primary sector benefits the most from horses, as an economic factor (Institute for Industrial Research 2005).

About 75 % of Austrian horses are kept at agricultural holdings and approximately half of all horse farms are breeding farms (12,500) (BMLFUW 2015). The breeding population consists of around 17,000 broodmares and 1,800 breeding stallions (BMLFUW 2015).

According to Frickh (2012), 100,000 ha of forage area, 160,000 t/year feed grain and 180,000 t/year hay and straw are required for the maintenance of Austria's horse population.

Horse slaughter is of minor significance in Austria. The annual domestic consumption of 191 t horsemeat is equivalent to 930 slaughtered animals. However, a proportion of horsemeat is destined for the pet food market and to feed captive wild animals in zoos. Therefore, it is difficult to determine the amount of meat, that is actually consumed by humans (Humane Society International 2012). The producer price for horsemeat per kg live weight is currently 1€ (BMLFUW 2017).

A total of 30 horse breeding associations are officially recognized in Austria and are managing 51 horse breeds. Figure 1 (ZAP 2011, modifed) below shows, that the major breeds are Haflinger, Noriker and the Austrian Warmblood Horse.

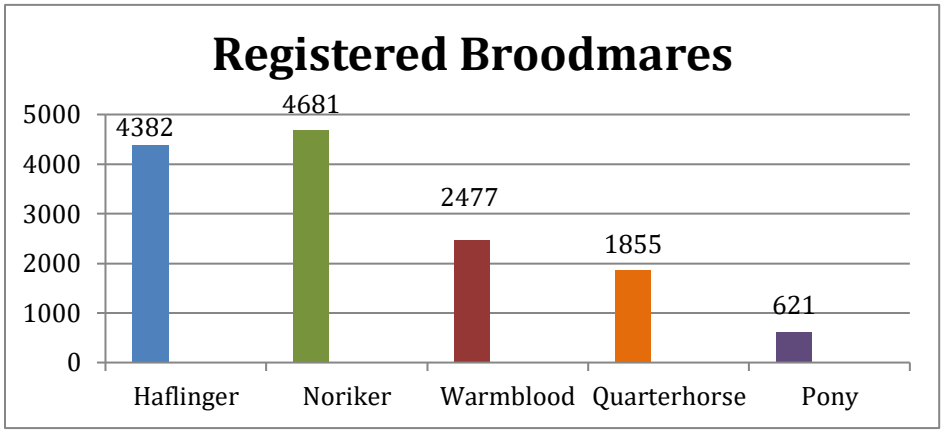


Figure 1: Registered Broodmares (ZAP 2011, modified)

2.1.1 Horse genome

The horse (*Equus caballus*) genome consists of 64 chromosomes (31 autosomes, two sex chromosomes)(National Center for Biotechnology Information 2017). The endangered subspecies of wild horse (*Equus ferus*), Przewalski, however, has 66 chromosomes due to a form of chromosomal rearrangement (Robertsonian translocation)(Goto et al. 2011). With a predicted genome size of 2.67 gb, the horse genome is smaller than the human and bovine one, yet larger than a dog's (McCue et al. 2012).

Whole genome shotgun sequencing has been completed in 2007 for Twilight - a thoroughbred mare - and allowed estimating the overall frequency of single nucleotide polymorphism (SNP) (one in 1500 bp on average) (McCue et al. 2012).

The domestic horse has been chosen for sequencing as a representative of the order *Perissodactyla* (horses, rhinoceroses and tapirs) due to its similarity with the human DNA (McCue et al. 2012). According to Wade et al. (2009), humans and horses have

16,617 orthologous genes out of 20,322 protein-coding genes, and 48 % of equine chromosomes show shared synteny to a single human chromosome suggesting, *inter alia*, shared regulatory mechanisms. Over 90 genetic conditions in horses are also present in human disorders such as muscular/respiratory diseases or infertility (Chowdhary et al. 2008). *Equus caballus* is therefore considered as a model organism and assists in the process of comprehending genetic aspects of diseases, both, in humans and horses (National Center for Biotechnology Information 2017).

2.2 Genetic diversity

Frankham et al. (2002) define genetic diversity as "the variety of alleles and genotypes present in a population". Common measurements for genetic diversity are allelic diversity; observed and expected heterozygosity; the proportion of polymorphic loci; and manifestly the frequencies of genotypes and alleles (Toro and Caballero 2005).

Genetic diversity can be described at various levels, however, in line with the current state of research, molecular markers are prominently used (Ellegren and Galtier 2016). The available SNP data has established the use of genomic indicators of diversity (FAO 2015). Genomic indicators of diversity may include, for example, measurements of genomic relatedness, inbreeding levels, fixation indices and genetic distances.

There is universal consensus regarding the importance of molecular data on within- and between-breed diversity for an adequate management of animal genetic resources (e.g. Weitzman 1993; Ruane 2000; Simianer 2005; Toro and Caballero 2005; Toro et al. 2009).

The importance of examination and determination of genetic polymorphism has increased even more due to the uncontrollably fast loss of animal genetic resources (FAO 2007). The threat is acute, considering the high rate at which breeds are disappearing (Frankham et al. 2002) with a total of 915 mammalian breeds at risk (Commission on genetic resources for food and agriculture 2013). Extinction is preceded by a loss of genetic diversity within the species, thus genetic polymorphism is indispensable for maintaining survival of a population (Oliehoek et al. 2009). Furthermore, genetic variability enables adaptation to climatic conditions, water/feed availability, emerging diseases and socio-economic conditions. Consequently it strengthens the global food security and is aiding the reduction of poverty and hunger by facilitating livelihood diversification and generating income (Commission on genetic resources for food and agriculture 2007).

The intense selection of commercial breeds is reducing genetic variability (Boettcher et al. 2010). Additionally, improved communication/transportation tools and state-of-the-art reproduction technologies - such as artificial insemination and embryo transfer - have led to a

global diffusion of genetic material (Groeneveld et al. 2009). These developments may be contributing factors to the growing concerns about the decline of genetic resources.

To ensure conservation, development and sustainable use of the world's livestock diversity the "Global Plan of Action for Animal Genetic Resources" is being implemented by 169 countries in cooperation with the FAO since 2007 (FAO 2007). While it is inevitable to lose a certain amount of breeds due to limited financial resources and constant alteration of the livestock production systems, it is the objective of the FAO to provide guidance and evaluate consciously which gene pools need to be preserved (FAO 2007). It is essential to understand and consider diversity at a species level and within/between breeds to make sustainable management decisions and avoid losing populations with unknown, unique characteristics (Commission on genetic resources for food and agriculture and FAO 2007).

2.2.1 Equine genetic diversity

The assessment of genomic indicators of diversity with molecular data is of importance to various sectors of the horse industry, including healthcare, selection and breeding (Bowling and Ruvinsky 2000). It may also help shed light on the domestication process of the horse, which has happened approximately 6,500 years ago in the Eurasian Steppe, however, the exact dating and place are still subject of discussion (FAO 2007). Petersen et al. (2013) suggest that the possibility of multiple domestication events, along with the continuation of gene flow between wild and domesticated horses, might prove to be the cornerstones of equine genetic diversity.

Although more than 400 different horse breeds have been formed primarily by selection (Shrestha 2017), the horse has never experienced such strong selective pressure regarding production traits, as is common in other livestock species (Petersen et al. 2013). The modern horse is mainly a result of breeding schemes, that focus on strength, endurance and speed (Lippold et al. 2011).

In general, equine genetic diversity varies, depending on breed regulations (definition, restrictions and amount of selection pressure); founding stock diversity; duration since establishment of the breed and geographic boundaries (Petersen et al. 2013).

As mentioned in the previous chapter, a decline of genetic diversity may be a harbinger of extinction. Horses have the highest proportion of "at-risk" breeds when compared to every other livestock species, with 23 % of all breeds being classified as either critical, critical-maintained, endangered, or endangered-maintained according to the risk status classification of the FAO (2007). 88 horse breeds have already become extinct and 91 breeds are endangered (Commission on genetic resources for food and agriculture 2013). Therefore, it is crucial to analyse the current level of diversity displayed in horse populations.

Studies, where molecular data has been used to assess genetic diversity within and among horse breeds have been conducted previously on Swiss breeds (Glowatzki-Mullis et al. 2006), Italian breeds (Felicetti et al. 2010), Canadian breeds (Plante et al. 2007), Tunisian breeds (Jemmali et al. 2017), Mexican breeds (Vázquez-Armijo et al. 2017) and Syrian horse populations (Almarzook 2017). However, the totality of studies evaluating genetic diversity in Austrian horse breeds is rather limited. Nevertheless, genetic diversity studies based on pedigree data have been conducted in Austria for the Haflinger (Druml et al. 2016) and the Noriker breed (Druml et al. 2009).

3 Animals, Data and Methods

3.1 Animals

In this thesis, genotypes of Haflinger and Noriker horses from Austria are analysed as described below.

3.1.1 Haflinger

Phenotype description and special characteristics:

The Haflinger breed invariably shows a chestnut coat colour with white mane and tail (ZAP 2015b). Markings are not desirable except on the head, and the height at withers should be between 140 and 150 cm (Tyrolean Haflinger Breeding Association 2016). Breeding goals aspire to create an elegant and harmonious type, which includes a noble head and a fairly divided croup in combination with a pronounced musculature (ZAP 2015b). Explicitly desired is a robust horse with a good nature and a strong character (Tyrolean Haflinger Breeding Association 2016).

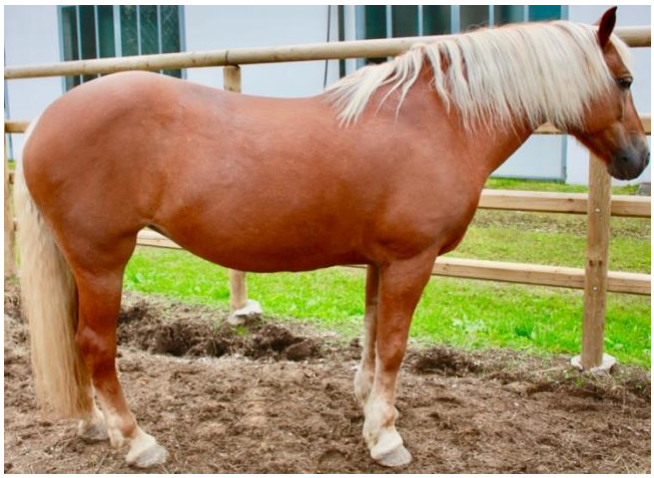


Figure 2: Haflinger broodmare in 2017. Photograph by Author, taken on September 22nd, 2017

Use:

Originally been used as a packhorse for farming in rough and steep terrain, the Haflinger has now developed into a general-purpose horse for sport and leisure riding (ZAP 2015b). The horse is still being used for military purposes in Austria and Germany. However, the major scope for utilization lies in riding and driving (Tyrolean Haflinger Breeding Association 2016). The current breeding population in Austria consists of more than 4300 registered mares and 150 stallions (ZAP 2015b). According to Mr. Schweisgut, former president of the World Haflinger Breeding & Sports Federation, 40 % of all Haflinger colts in Tyrol are slaughtered every year. He emphasizes, however, that they are not bred with the intention to slaughter them (Morawetz 2010).

History:

The founder of the breed, Folie 249, was a small, light, warmblood stallion in the South Tyrolean Alps in 1874 (Tyrolean Haflinger Breeding Association 2016). The founding stallion was the product of cross-breeding an Oriental stallion with a local mare of Galician origin (Messner 2014). Today, the breed is divided by seven sire lines, which are the progeny of Folie 249.

3.1.2 Noriker

Phenotype description and special characteristics:

The Noriker is a mid-weight coldblood horse, with an optimal height at withers between 156 and 162 cm (ZAP 2015a). The breed shows various coat colours such as black, bay, chestnut, blue-, brown- and red-roan, leopard- and tobiano-spotting (ARGE Noriker 2010). Formation of a Noriker is defined by a significantly split and heavily muscled croup (Association Noriker Austria 2015). In addition, great importance is attached to correct, pronounced joints and sturdy, resistant hooves (ARGE Noriker 2010). An essential breeding goal is the creation of a surefooted, enduring horse with a quiet temperament (ZAP 2015a).

Use:

This breed used to be exclusively utilized as a work horse to pull heavy wagons and transport goods (PferdAustria 2015). Today, however, it has transformed successfully into a modern leisure horse (Association Noriker Austria 2015), which is intensively used in the driving sport and a fundamental part of rural traditions (ZAP 2015a).

An important aspect, regarding the Noriker, is the endangerment of the breed. Similar to other European draft horses, the population size has decreased from 34,500 animals in the year 1968 to 2376 active breeding animals in 2004 (Druml et al. 2007). According to Arche Austria

(2015), approximately 4,600 mares and 180 stallions are registered in Austria. The major breeding areas are Salzburg, Carinthia and Upper Austria (Druml et al. 2007).

Since the Austrian Noriker is acknowledged as an endangered breed by conservation platforms, farmers, who precisely fulfil the requirements of the conservation breeding program are provided subsidies, 180€ for mares and 360€ for stallions per year (ÖNGENE 2015).

History:

The Noriker is one of the oldest indigenous draft horse breeds in Europe (Druml et al. 2007). Creation of the breed dates back to the Roman province Noricum, albeit systematic breeding started in the 16th century and was implemented by the archbishops of Salzburg (ARGE Noriker 2010). During this period, the breed has experienced influences of Neapolitan and Spanish blood lines. Furthermore, five bloodlines are present in the current population: Vulkan, Schaunitz, Elmar, Diamant and Nero (ARGE Noriker 2010). Today the stud book is closed and a pure breeding scheme is strictly followed (ZAP 2015a).

Figure 3: Noriker breeding stallion in 2015. Photograph by Author, taken on June 15th, 2015

3.2 Data

A total of 274 horses, 91 Haflinger (Haf) and 183 Noriker (Nor), have been used for the analysis. The original dataset consists of 101 mares (13 Haflinger, 90 Noriker) and 173 stallions (78 Haflinger, 93 Noriker). Genotyping was performed using the Axiom Equine Genotyping Array. The array features 670,796 markers and can be used for genotyping 20 different breeds, including Haflinger and the South German Coldblood, which shares common origins with the Noriker breed (Affymetrix, Inc. 2015).

3.2.1 Quality control process

The quality control process for the genotype dataset has been performed with PLINK v 1.9 (Purcell et al. 2015). Originally, the dataset contained 670,796 SNPs. Only SNPs located on autosomal chromosomes have been selected for the analysis, i.e., unplaced SNPs or SNPs

situated on sex chromosomes have been excluded (exclusion of 8.78 % of SNPs). Individuals with more than 5 % missing SNPs have been excluded (32 animals eliminated: 12 Haflinger and 20 Noriker) and SNPs with more than 10 % missing genotypes have also been discarded (47,894 SNPs eliminated). Minor allele frequency (MAF) was stipulated to be more than 1 %, which led to the withdrawal of 76,171 SNPs. After the application of these quality parameters, 487,850 SNPs and 242 individuals have been retained: 79 Haflinger and 163 Noriker.

For the analysis of runs of homozygosity, the MAF-filter has not been applied, which led to 564,021 remaining SNPs and 242 individuals.

For some calculations (fixation index (F_{ST}), identity by descent, inbreeding coefficient and principal component analysis), linkage disequilibrium (LD) pruning has been performed to exclude SNPs which show a squared coefficient of correlation (r^2) above the threshold of 0.7. This has reduced the number of SNPs from 487,850 to 328,005 (removing 159,845 SNPs).

3.3 Methods

The applied methods, such as estimation of F_{ST} pairwise values and Nei's genetic distances; principal component analysis; calculation of genomic relationships; and the genomic inbreeding coefficient are described in detail in this chapter.

3.3.1 Principal Component Analysis

Pearson (1901) developed this statistical technique to simplify a set of data by describing maximal variance with the fewest number of principal components. The first principal component describes the greatest variance and so forth. Its aim is to preserve as much of the relevant information as possible (Jolliffe 2002).

Principal components analysis (PCA) is extensively used to assess the genetic structure of a population from SNP data (Jianzhong and Amos 2012). Menozzi et al. (1978) conducted the first attempt at applying a PCA to population genetic data. It allows to summarize genetic variability without relying on the Hardy-Weinberg equilibrium (Jombart 2008).

Prior to conducting the PCA, the data has been prepared in the same way as described in detail in chapter 3.2.1, including LD pruning and the basic quality control steps.

The distances between individuals, expressed as allele counts, have been calculated using PLINK 1.9 (Purcell et al. 2015; Chang et al. 2015). The created file has further been processed with the software R (R Core Team 2017) and the PCA has been conducted separately and jointly for both populations (HAF, NOR, HAF & NOR). Afterwards the proportion of variation captured by each eigenvector (eigenvalues) has been computed and results of the PCA were visualized.

3.3.2 F_{ST} pairwise values and Nei's genetic distances

In population genetics the F_{ST} is widely used to analyse genetic variation (Holsinger and Weir 2009). According to Wright (1949), the F_{ST} quantifies the extent to which a polymorphic population can be subdivided into subpopulations. F_{ST} can be defined with following formula:

$$F_{ST} = \frac{(Ht - Hs)}{Ht}$$

Ht = expected heterozygosity of the overall population

Hs = mean expected heterozygosity across subpopulations

To calculate pairwise F_{ST} values for Haflinger and Noriker, all quality control steps, including LD pruning (see chapter 3.2.1) have been performed. The processed data file has further been prepared using PLINK and then imported into the R software (R Core Team 2017). For the calculation of both - F_{ST} pairwise values and Nei's genetic distances - the StAMPP (Statistical Analysis of Mixed Ploidy Populations) R package (Pembleton et al. 2013) has been used. The pairwise F_{ST} values (stamppFst) have been calculated based on Weir and Cockerham's (1984) updated version of Wright's method (1949). The outcome of this calculation is a matrix of pairwise F_{ST} values between populations.

To estimate genetic distances between populations and individuals, Nei's standard genetic distance is the most commonly used statistical method (Chakraborty et al. 2012). Nei (1972) formulated a measure of genetic distance based on the identity of genes. Accumulated allele differences per locus are determined and it is applicable for closely related populations within a species as well as distantly related species.

For this thesis, both, Nei's genetic distances between populations and between individuals have been estimated. As described above, the prepared data has been imported into the R software (R Core Team 2017) and Nei's genetic distance has been computed with the "Genetic Distance Calculation-function" of the StAMPP-package. Afterwards the neighbour-joining (NJ) method (Saitou and Nei 1987), as implemented in the R software-package APE 4.1 (Paradis et al. 2004), has been used to build the phylogenetic trees from the distance matrices and the results have been visualized using the "plot phylogenies-function".

3.3.3 Genomic relationships: Identity by descent and identity by state

The concepts of identity by state (IBS) and identity by descent (IBD) are of enormous importance for population genetics (Browning and Browning 2010).

The term IBS is used, if two individuals share one (IBS_1) or two (IBS_2) identical alleles at a given locus. IBS distances are calculated with following formula:

$$IBS_{distance} = \frac{(number\ of\ markers\ with\ IBS_2) + (0.5 \times number\ of\ markers\ with\ IBS_1)}{Number\ of\ non-missing\ markers}$$

Alleles, which are IBS could also be IBD, if they are inherited from a common ancestor (Gusev et al. 2009). IBD combines relatedness and inbreeding and is therefore an important quantification for genomic relationships within populations (Browning and Browning 2010).

For the purpose of this thesis a relationship matrix was constructed with the function "--distance-matrix" from PLINK 1.9 (Purcell et al. 2015; Chang et al. 2015). This produces a distance matrix from the quality controlled data sets for each breed separately, based on the formula:

$$1 - IBS_{distance}$$

Furthermore, another relationship matrix was formed with PLINK 1.9 (Purcell et al. 2015; Chang et al. 2015) based on calculations of Hamming's distances and IBS. The computation of Hamming's distances is frequently used for the comparison of DNA segments (Chang et al. 2015) by calculating the number of dissimilar components (Hamming 1950).
Modifications were applied to change the shape of the output matrix to a symmetric form (square) and to express distances with an IBS-matrix (ibs).

The third relationship matrix focused on IBD and was formed with the PLINK 1.9 (Purcell et al. 2015; Chang et al. 2015) "--genome" function. These calculations are not LD sensitive and therefore were applied on the LD pruned data set (see chapter 3.2.1) for both breeds.
All three relationship matrices were further processed with the software R (R Core Team 2017) to produce heatmaps and the mean and standard deviation has been calculated. Additionally, a histogram, to visualize the distribution of values, has been generated for the distance matrix, the IBS matrix and the IBD matrix.

3.3.4 Inbreeding coefficient and runs of homozygosity

In relation to population diversity, inbreeding plays an important role. While it refers to the mating of individuals, who have one or more common ancestors, it can also be used to represent the level of diversity of the genome (Wright 1922). Methods to quantify livestock inbreeding

have developed in the past years from being based on pedigree information to usage of molecular markers for inbreeding estimations (Curik et al. 2014). Monitoring inbreeding levels is crucial for animal breeder due to its various effects on the population. In addition to an increase of homozygosity, severe effects of inbreeding may include a reduction of population fitness (Charlesworth and Willis 2009) and increased occurrence of homozygous recessive defects (Alvarez et al. 2009). However, inbreeding is commonly used in commercially essential livestock populations to enhance uniformity (Curik et al. 2014).

Numerous molecular approaches to estimate levels of inbreeding exist, but for the purpose of this thesis the genomic inbreeding coefficient implemented in PLINK 1.9 (Purcell et al. 2015; Chang et al. 2015) (F_{PLINK}) is used. The estimation of F_{PLINK} for the quality controlled and LD pruned data set of each population is based upon observed and expected autosomal homozygous genotype counts for each sample (–het) and applied through following formula:

$$F = \frac{observed\ hom.count - expected\ hom.count}{total\ observations - expected\ hom.count}$$

hom. count = homozygous genotype counts

Another method to reliably estimate inbreeding levels, or autozygosity, is based on the fact that inbred individuals exhibit long homozygous sections of the genome, which are identical by descent. These so-called runs of homozygosity (ROH) can be produced by various mechanisms, however the main cause for emergence of ROH is inbreeding (Ferenčaković et al. 2013). The estimation of inbreeding coefficients through ROH (F_{ROH}) may be even more reliable than estimations based on pedigree information (Sölkner et al. 2010, Ferenčaković et al. 2013).

F_{ROH} can be defined with following formula:

$$F_{ROH} = \frac{\sum L_{ROH}}{L_{AUTOSOME}}$$

L_{ROH} = total length of all ROH in the genome of an individual

$L_{AUTOSOME}$ = length of the autosomal genome covered by SNPs

$L_{AUTOSOME}$ for the Axiom Equine Genotyping Array amounts to 2,242,739 kb. F_{ROH} is considerably higher for livestock populations than for humans due to smaller effective population sizes and artificial selection (Curik et al. 2014).

The data from both breeds has been prepared using PLINK 1.9 (Purcell et al. 2015; Chang et al. 2015) (no MAF filter nor LD pruning has been applied, see chapter 3.2.1) and ROH have been detected using the cgaTOH software 1.0.1. The software cgaTOH differentiates between TOH (tracts of homozygosity; identical with ROH) and cTOH. cTOH are identified, if a number

of individuals share the same ROH (for example, if more than 10 individuals display ROH at the same consecutive SNPs)(Zhang et al. 2013). An overview is given for these regions of the genome which are termed islands (Nothnagel et al. 2010) and for regions where ROH are rare in the population, known as deserts (Curik et al. 2014). ROH islands have been shown to be often the result of selection, whereas deserts might occur at loci associated with critical functions (Pemberton et al. 2012). In general, ROH islands and deserts show where on the genome abundance and lack of diversity are manifested.

On the basis of the length of ROH, five categories have been formed: ROH longer than 1 Mb (ROH > 1); ROH longer than 2 Mb (ROH > 2); ROH longer than 4 Mb (ROH > 4); ROH longer than 8 Mb (ROH > 8) and ROH longer than 16 Mb (ROH > 16). To exclude common and short ROH that occur due to LD, the minimum length of ROH has been set to 1 Mb.
The length of ROH reveals how many generations earlier inbreeding has occurred (Ferenčaković et al. 2013). Short ROH have most likely been affected by many recombination events and therefore suggest ancient inbreeding, whereas long ROH imply recent inbreeding incidents. According to Curik et al. (2014), 10 Mb long ROH imply that inbreeding happened five generations ago and 5 Mb long ROH originate from inbreeding events that have taken place ten generations back. These assumptions are based on studies on other livestock species, however, due to the scarcity of studies on ROH in horses, these propositions have been taken into account for the purpose of this thesis. The present categories for ROH lengths of 1 Mb, 2 Mb, 4 Mb, 8 Mb and 16 Mb may be linked to inbreeding events occurring 50, 25, 13, 6 and 3 generations ago, respectively.
The following parameter settings were taken into account:

- 1 heterozygous SNP and maximum of 4 missing SNPs were allowed for ROH>1
- 2 heterozygous SNPs and maximum of 8 missing SNPs were allowed for ROH>2
- 4 heterozygous SNPs and maximum of 16 missing SNPs were allowed for ROH>4
- 8 heterozygous SNPs and maximum of 32 missing SNPs were allowed for ROH>8
- 16 heterozygous SNPs and maximum of 64 missing SNPs were allowed for ROH>16
- ROH were called if 15 or more consecutive homozygous SNPs were present
- maximum physical gap between adjacent SNPs of 1 Mb
- minimum SNP overlap of 10 (for cTOH)

The R software (R Core Team 2017) has then been used to summarize the results from the cgaTOH software 1.0.1 (Zhang et al. 2013). Additionally, the function ggplot from the R package ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics version

2.2.1 (Wickham 2009) has been used to visualize ROH on each autosome for both breeds.

4 Results and Discussion

The first two principal components of the PCA for both breeds are illustrated in figure 4. About 12 % of the variation is described by the first principal component and 3.07 % by the second. The individuals belonging to the Noriker breed cluster loosely in the right half of the plot and the Haflinger population groups in the bottom left quadrant. Figure 4 shows a distinction between breeds, however, three individuals show anomalous clustering. Two Noriker horses and one Haflinger horse scatter in-between the two breed clusters. A possible explanation for these divergent individuals may be based on the assumption of Leroy et al. (2009) that in the past Haflinger were used as draught horses and therefore might have been crossbred with other European draught horse breeds. In compliance with this assumption, Druml et al. (2016) show with an analysis of Austria's current Haflinger population, that the Haflinger gene pool consists of 1.8 % Noriker genes.

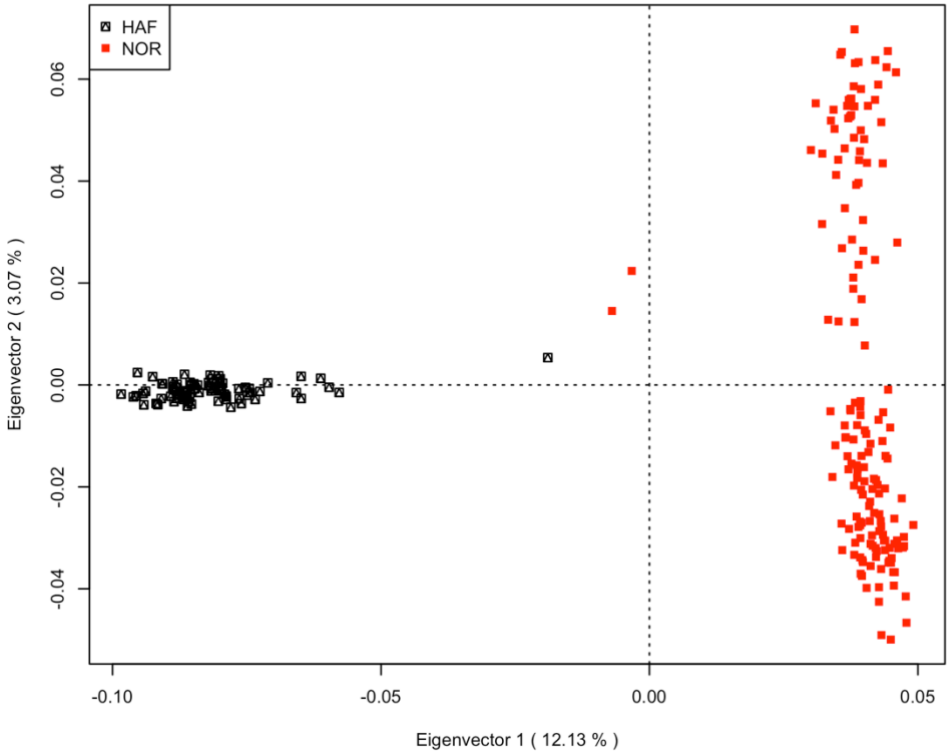


Figure 4: First and Second principal component of a PCA of allele frequencies from 242 individuals of two different breeds (Haflinger & Noriker)

Furthermore, Druml et al. (2016) express their concerns regarding a decline in genetic variability of the Haflinger horse, which can be reinforced with the present PCA results.

Whereas the Noriker population scatters across the plot with the second and third eigenvectors, the Haflinger breed remains in a relatively tight cluster until being separated by eigenvectors six and seven, as shown in figure 5. Despite, the fact of being considered as endangered, the PCA clustering proposes a higher level of genetic diversity for the Noriker breed in comparison with the Haflinger population.

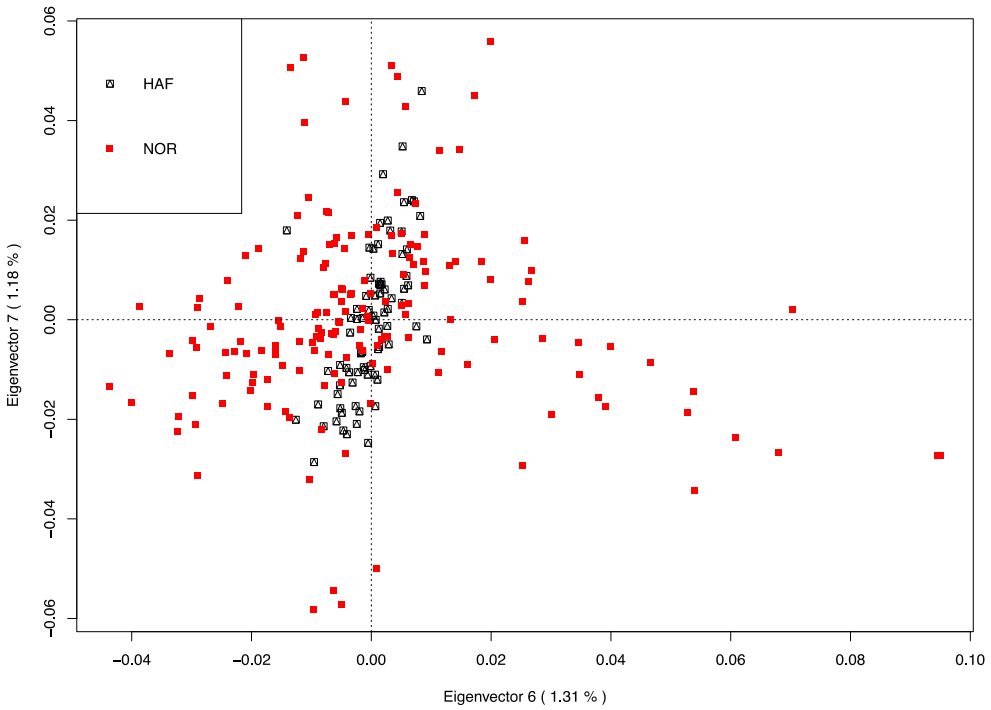


Figure 5: Sixth and Seventh principal component of a PCA of allele frequencies from 242 individuals of two different breeds (Haflinger & Noriker)

Figure 6 shows the first two principal components for the Haflinger breed. Almost 4 % of the variation is described by the first principal component and 3.44 % by the second. The PCA for the Noriker breed is shown in figure 7 with a value of 5.06 % of the variation explained by the first eigenvector and 2.67 % by the second.

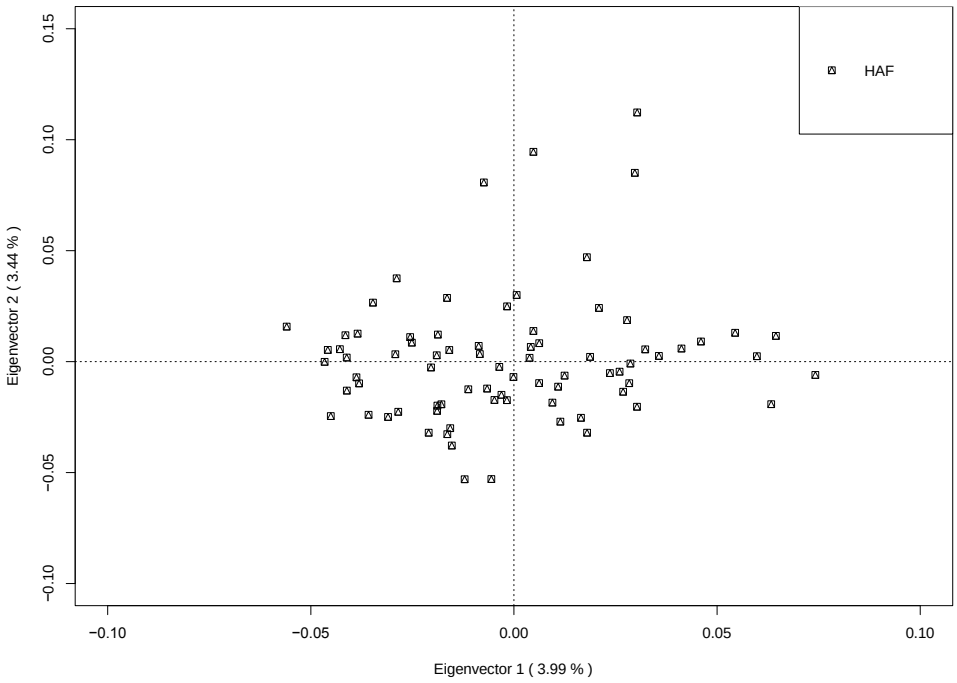


Figure 6: First and Second principal component of a PCA of allele frequencies from 79 individuals of the Haflinger breed

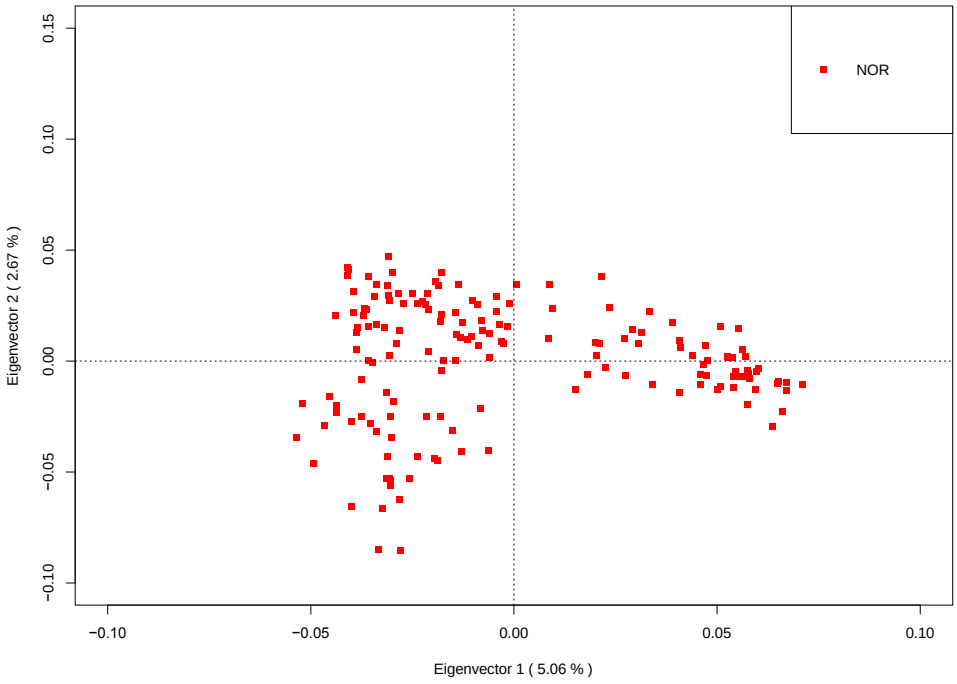


Figure 7: First and Second principal component of a PCA of allele frequencies from 163 individuals of the Noriker breed

The pairwise F_{ST} between Haflinger and Noriker populations was 0.0855. In other words, 8.5 % of the total genetic diversity is explained by between-breed differences. The remaining genetic variance is based on differences between individuals. According to the suggestions of

Wright (1984) and Hartl & Clark (2007), F_{ST} values between 0.05 and 0.15 indicate moderate genetic differentiation. Cañon et al. (2000) found similar results for Spanish horse breeds (0.08). Nei's genetic distance between populations amounts to 0.0367, suggesting a rather close connectedness between the two populations. While the Noriker breed was used in the establishment of the Haflinger population, the breeds are currently well separated, thus such a narrow estimated genetic distance was surprising.

The result of Nei's genetic distance calculations between individuals is visualized in form of a phylogenetic tree in figure 8 below. With the exception of one individual (marked with a red asterisk in figure 8) a clear distinction between the two breeds is visible. The assumption that interbreeding occurred between Haflinger and Noriker, which in some individual has been shown by our PCA results and studies of Druml et al. (2016) and Leroy et al. (2009) is not sufficiently supported by the structure of the phylogenetic tree. However, the structure of one branch of the Noriker breed (indicated in figure 8 with a bracket) seems to be more similar to the structuring of the Haflinger population than to the other branches of the Noriker breed.

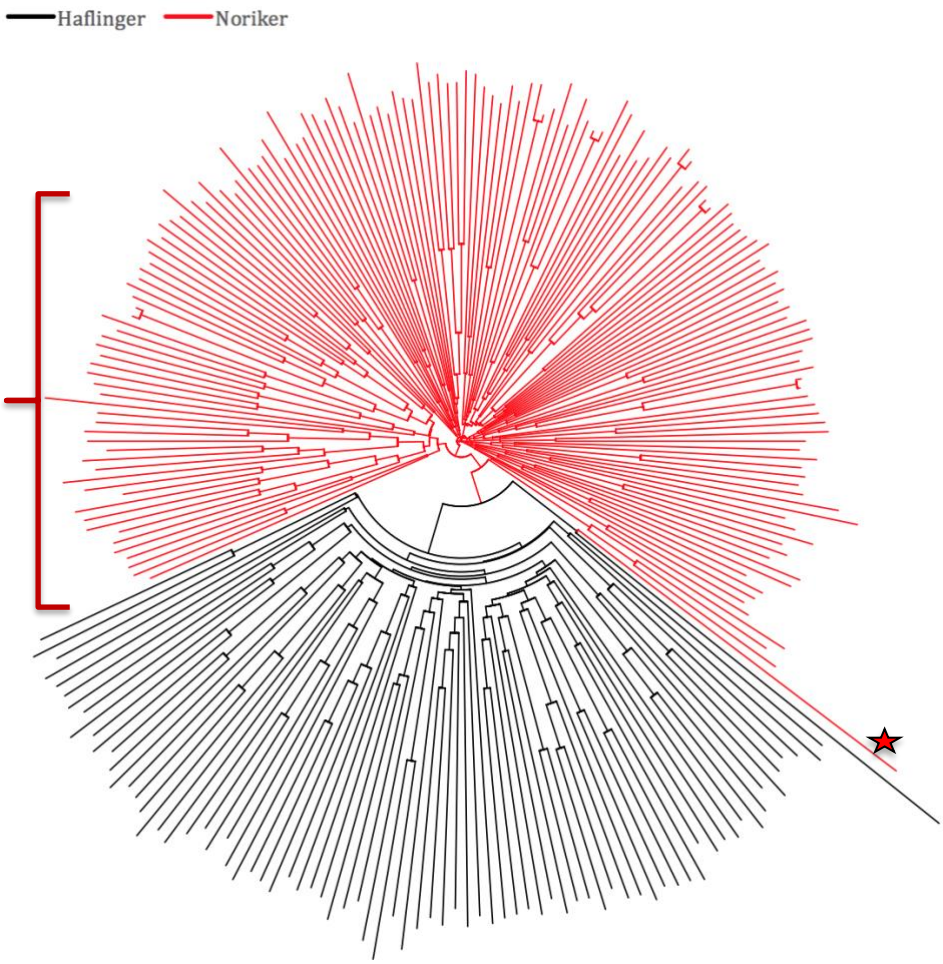


Figure 8: Phylogenetic tree based on Nei's genetic distances between individuals for Noriker (red) and Haflinger (black)

To further demonstrate genetic relationships within the Haflinger and Noriker population, both, IBS and IBD matrices are depicted below. The mean IBS distances are similar for both breeds (0.766 for the Haflinger breed and 0.767 for the Noriker breed), which is shown in Table 1 and the identity by state matrices in form of heatmaps for both breeds (see figure 9 for HAF and figure 11 for NOR). Histograms to visualize the distribution of IBS values show the Haflinger population in Figure 10 and the Noriker population in figure 12. Values for IBS can range between 0 and 1. A value of 0 would occur if all markers are IBS (e.g. duplicates or identical twins), whereas a value close to 1 indicates that very few individuals share identical alleles at a given locus. The values obtained in this analysis are rather high for both breeds, suggesting a higher allelic diversity. The heatmaps and histograms for the distance matrices for both breeds can be found in the Annex. They do not show any differentiation between the two breeds either.

Table 1: Mean and standard deviation (in brackets) for IBS and IBD in the Haflinger and Noriker population

	Identity by state	Identity by descent
Noriker	0.767 (0.012)	0.015 (0.050)
Haflinger	0.766 (0.015)	0.027 (0.056)

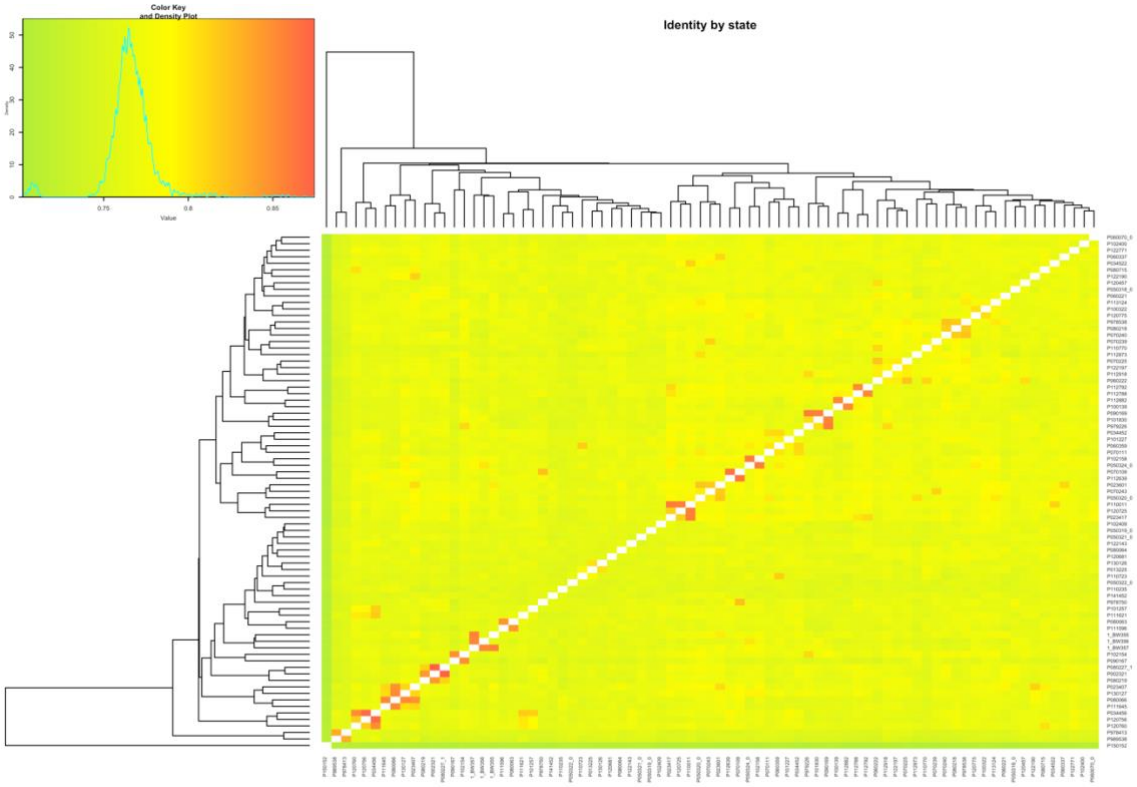


Figure 9: Identity by state matrix for the quality controlled data set of the Haflinger breed

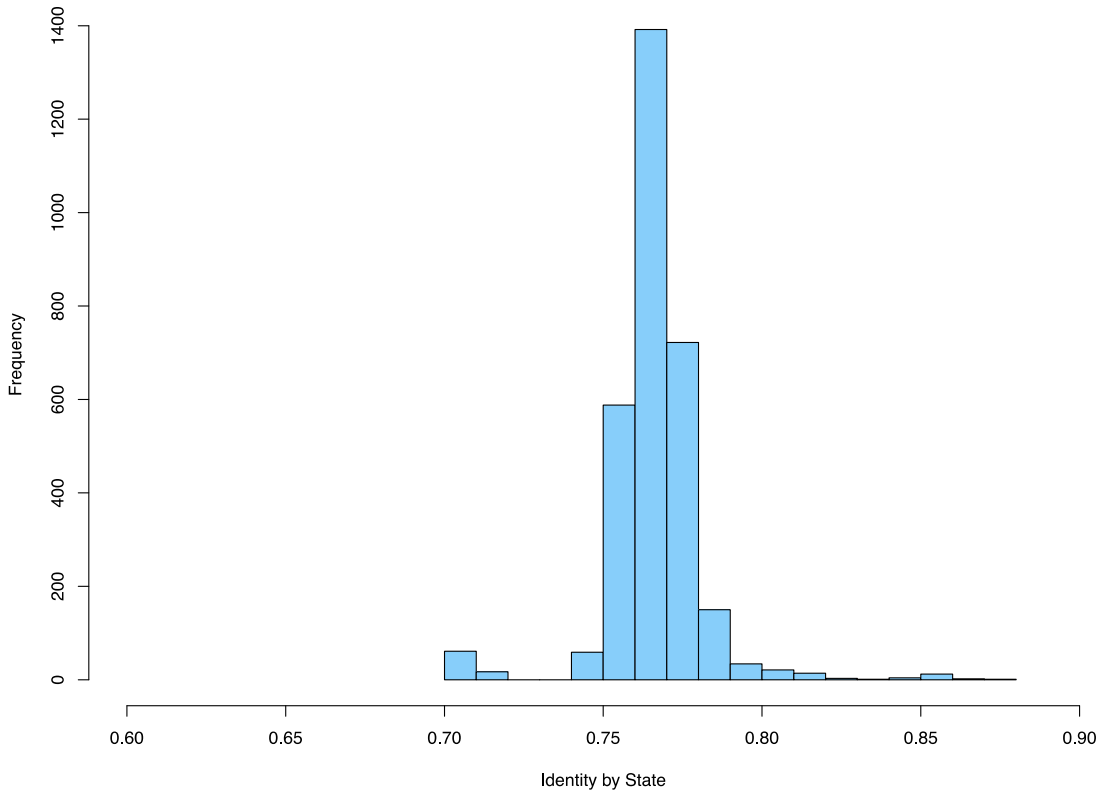


Figure 10: Histogram of identity by state matrix for the Haflinger breed

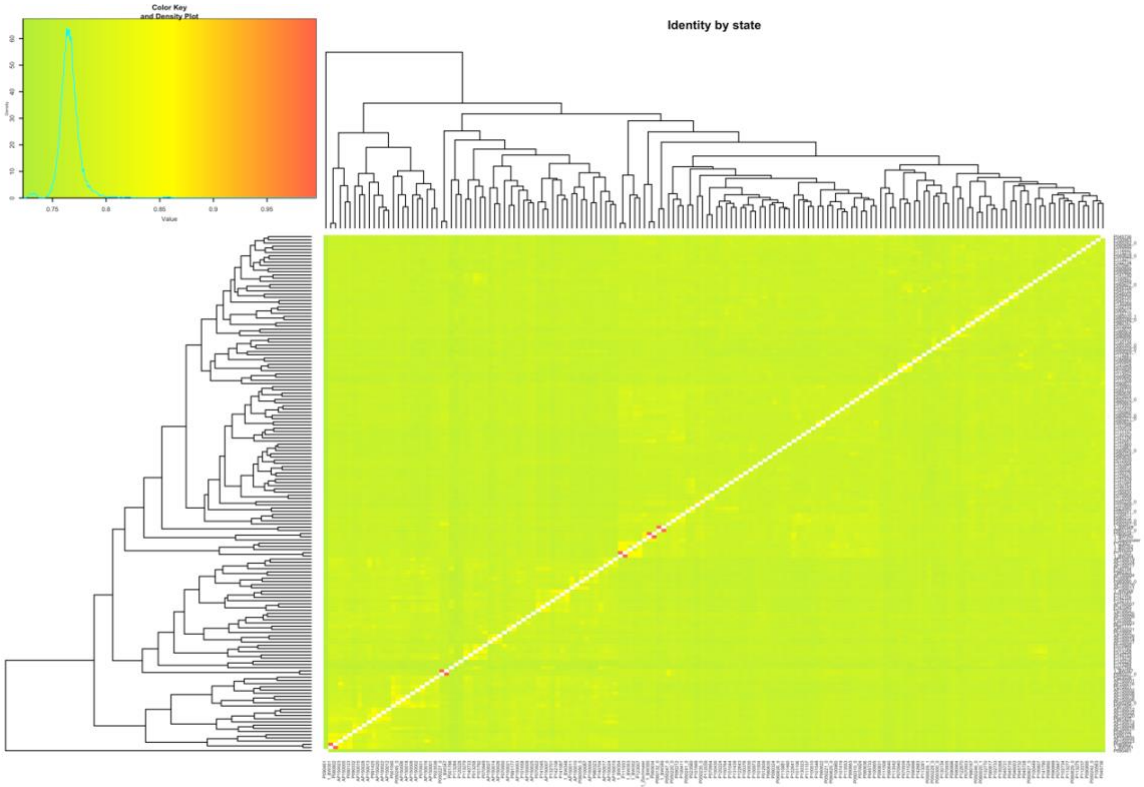


Figure 11: Identity by state matrix for the quality controlled data set of the Noriker breed

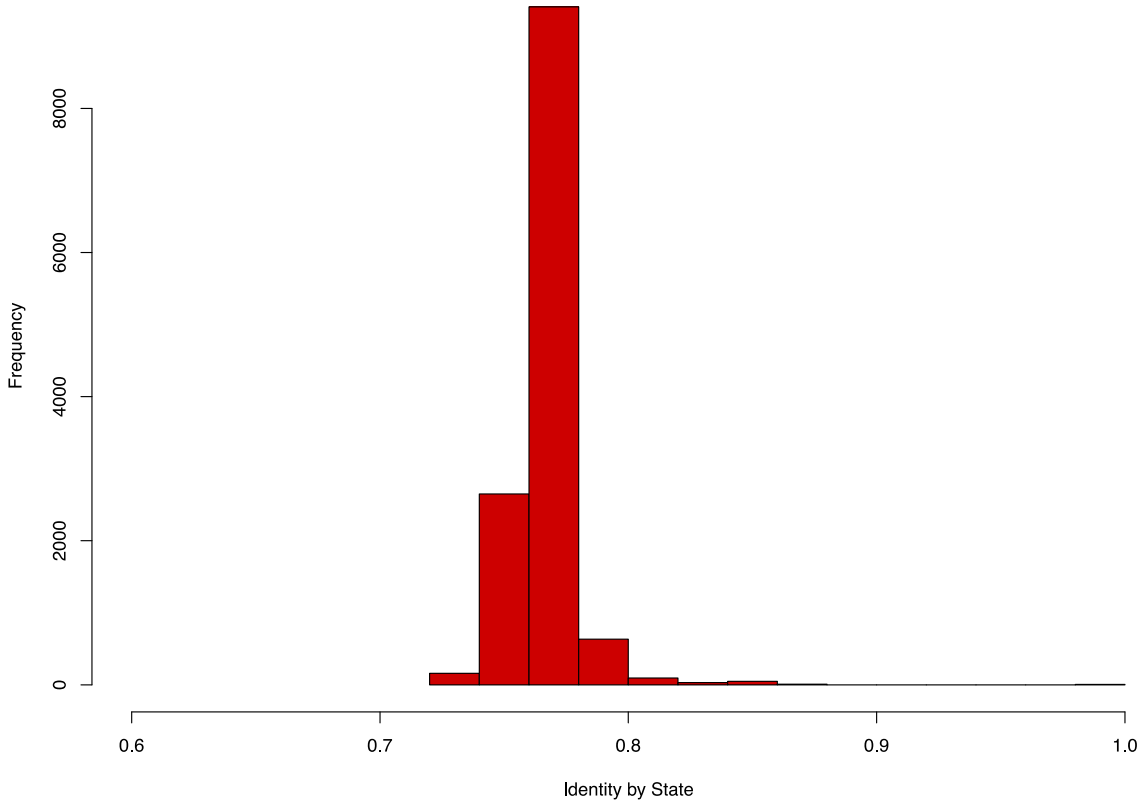


Figure 12: Histogram of identity by state matrix for the Noriker breed

Table 1 includes the mean estimates for IBD and the results are close to 0, which represents a population with mostly unrelated individuals. Nonetheless, the heatmaps based on the IBD matrices (see figure 13 for the Haflinger and figure 15 for the Noriker population) and the histograms (figure 14 HAF, Figure 16 NOR) show that both breeds include some individuals with high estimates. The mean IBD estimates are slightly higher for the Haflinger population (0.027 HAF, 0.015 NOR) as can be seen in Table 1. These higher values for some individuals could be due to first degree relationships. However, results above 0.5 proportion IBD are most likely due to sampling errors (e.g. inclusion of duplicates) (Browning and Browning 2010).

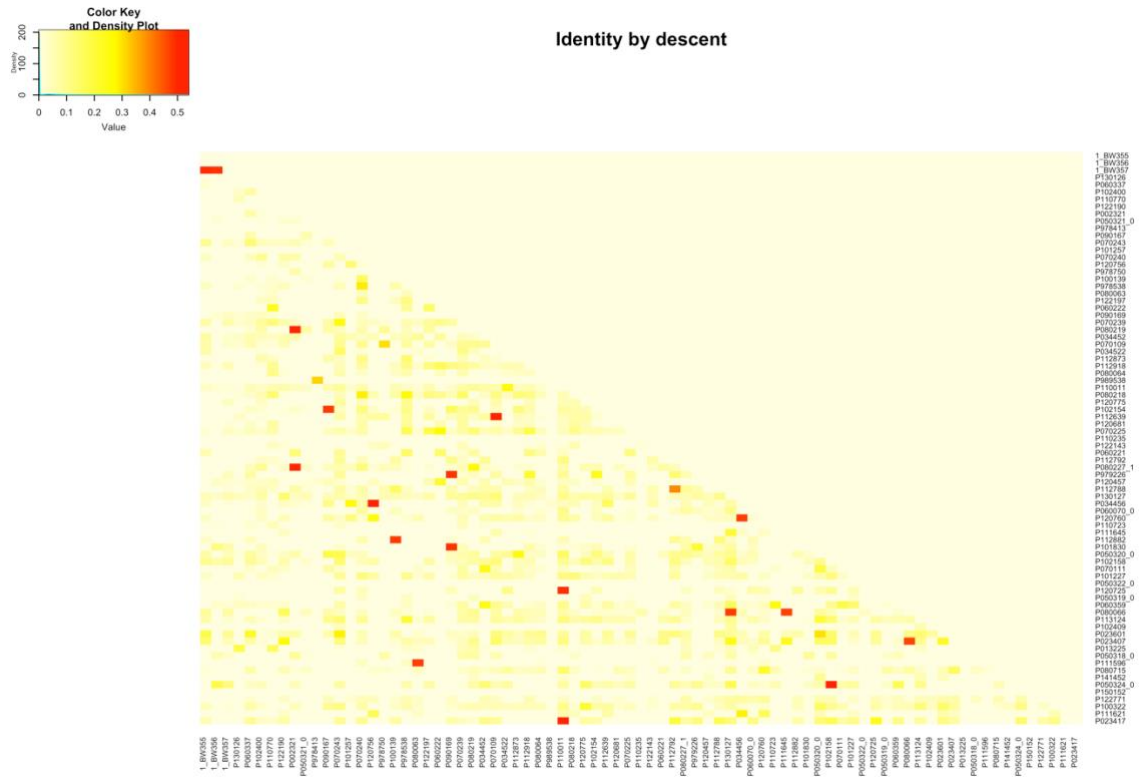


Figure 13: Identity by descent matrix for the quality controlled and LD pruned data set of the Haflinger breed

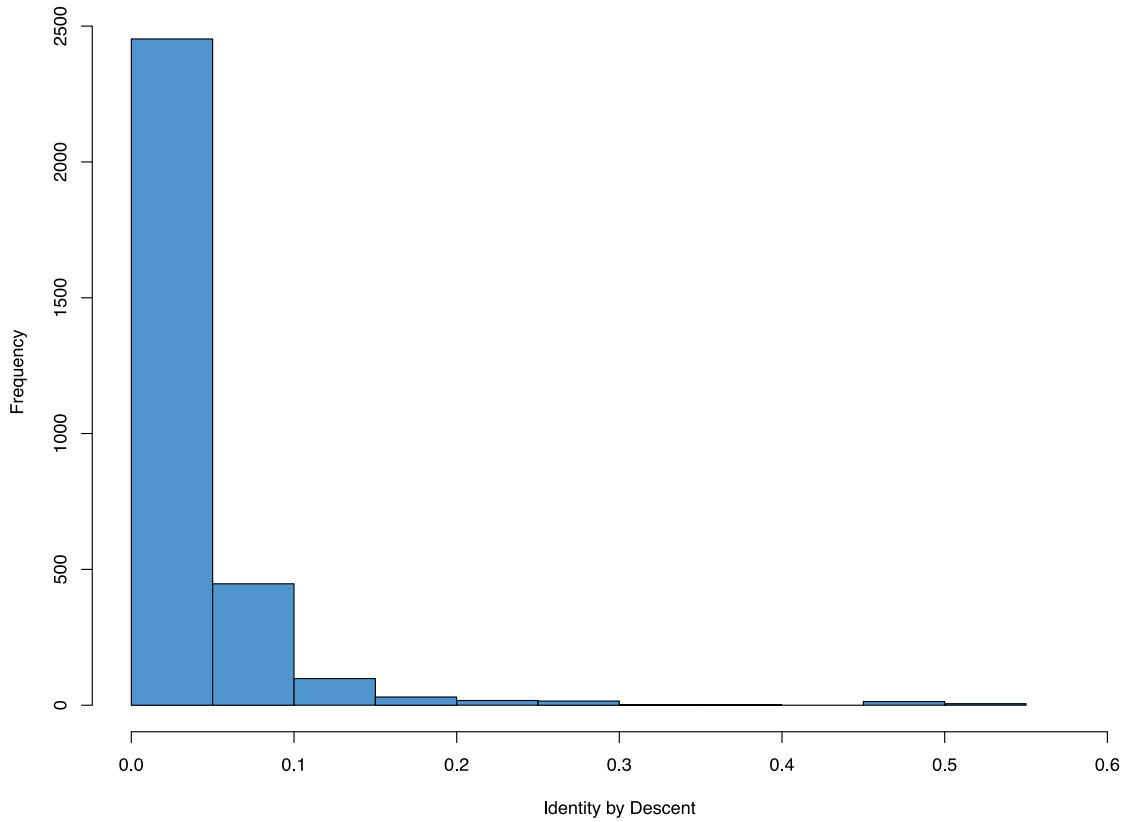


Figure 14: Histogram of identity by descent matrix for the Haflinger breed

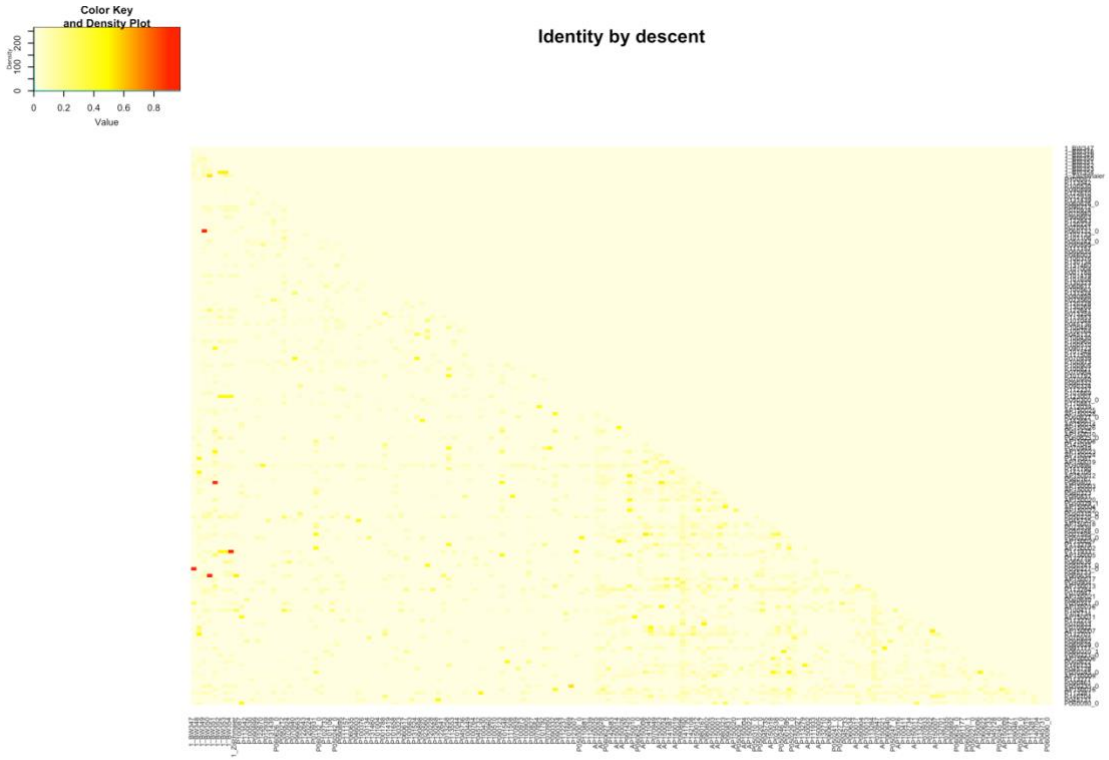


Figure 15: Identity by descent matrix for the quality controlled and LD pruned data set of the Noriker breed

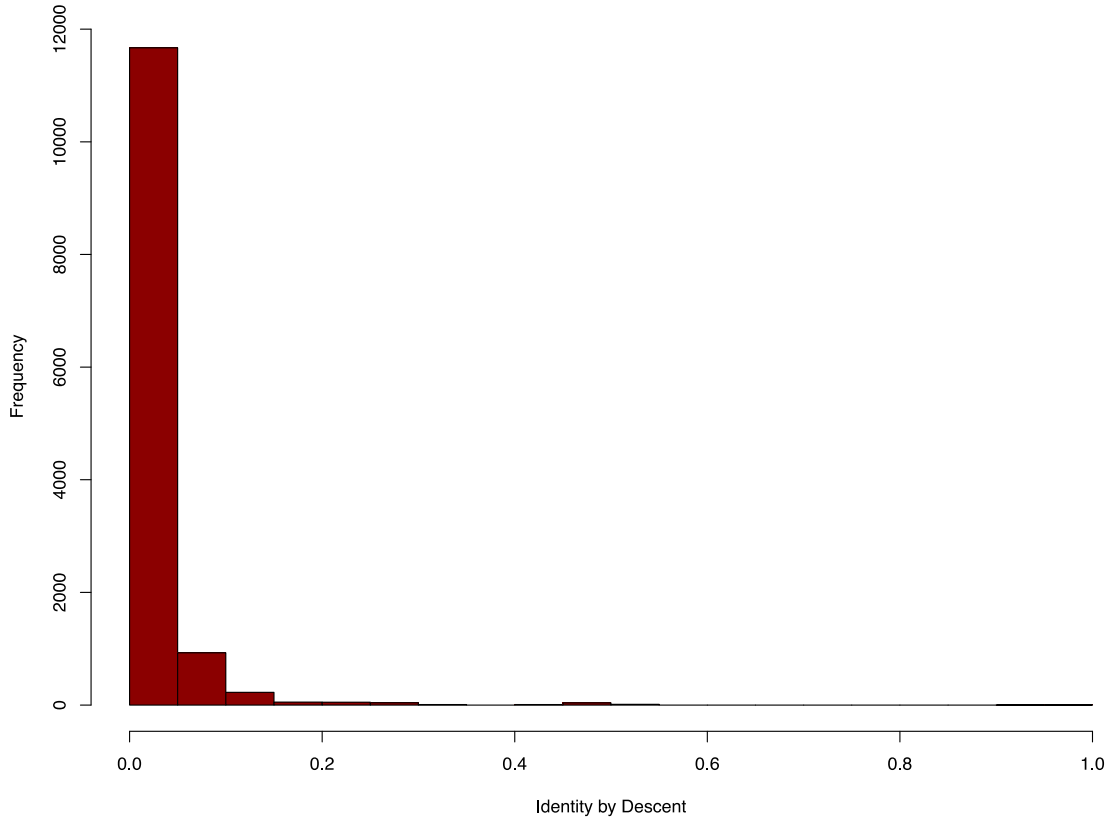


Figure 16: Histogram of identity by descent matrix for the Noriker breed

Table 2: Number of samples (N), individual inbreeding estimates (F_{PLINK}), mean and standard deviation (in brackets) and inbreeding coefficients based on runs of homozygosity (F_{ROH}) of different minimum lengths (> 1, > 2, > 4, > 8 or > 16 Mb) for quality controlled data (see chapter 3.2.1) of Noriker and Haflinger populations

		Individual inbreeding (F_{PLINK})			Inbreeding coefficients from ROH (F_{ROH})				
Breed	**N**	**Min**	**Max**	**Mean**	**$F_{ROH>1}$**	**$F_{ROH>2}$**	**$F_{ROH>4}$**	**$F_{ROH>8}$**	**$F_{ROH>16}$**
Noriker	163	-0.286	0.118	0.015 (0.05)	0.045	0.029	0.018	0.012	0.010
Haflinger	79	-0.144	0.127	-0.004 (0.05)	0.088	0.062	0.041	0.023	0.018

The calculation of the inbreeding coefficient (F_{PLINK}) for each individual based upon observed and expected homozygosity has shown several individuals with significant loss of heterozygosity (see Table 2). The highest individual value of F_{PLINK} (0.127) has been found in the Haflinger population, suggesting half-sib mating has occurred. This result conforms with the high IBD values estimated for some of the Haflinger individuals.

However, the average estimate of F_{PLINK} within breeds has been greater in the Noriker population. Values for F_{PLINK} can range from -1 to 1. Negative values result from a lower homozygote genotype count than expected by chance at the genome-wide level. For close to 60 % of the Haflinger population negative values have been estimated. Hence, the mean genomic inbreeding coefficient for the entire population is also negative and indicates a higher heterozygosity than expected. Purcell et al. (2007) and Li et al. (2011) state that negative values for inbreeding values are most likely produced by sampling errors and should therefore be set to zero for the purpose of calculating the mean. If this suggestion would have been taken into account, the results for F_{PLINK} would have been increased to a value of 0.018 for the Haflinger and 0.027 for the Noriker breed.

According to previous studies using molecular data, the genomic inbreeding coefficient for horse breeds can range between values of 0.02 (Mongolian horses) and 0.15 (Thoroughbred horses)(McCue et al. 2012). The observed F_{PLINK} values in this study are at the lower end of this range and the Haflinger population is even below the literature results. Druml et al. (2009) conducted a study on the Austrian Noriker population based on pedigree data and showed decreasing inbreeding levels in the horse breed (0.0121 based on five generations and 0.0324 for ten generation pedigrees). They suggest, that the use of modern breeding techniques, which allows breeders to expand the geographic area for stallion selection, might be a reason for this decline of inbreeding levels. For the Austrian Haflinger population, however, a recent study by Druml et al. (2016) suggests possible occurrence of an increase in inbreeding levels. The negative F_{PLINK} values estimated in this thesis cannot be considered to represent the entire Austrian Haflinger population. A higher number of genotyped individuals would be needed to give a better overview of the current inbreeding status of the Haflinger breed.

On the contrary to the results of F_{PLINK}, the results of F_{ROH} support this assumption of increasing inbreeding levels in the Austrian Haflinger population. Several studies proclaim that the ROH approach to calculate F is more reliable than pedigree approaches (Sölkner et al. 2010, Ferenčaković et al. 2013). F_{ROH} values were higher in the Haflinger population than in the Noriker population (see Table 2). The inbreeding coefficient for ROH >1 Mb ($F_{ROH>1}$) was highest for an individual of Haflinger origin (0.185). The lowest inbred horse belonged to the Noriker population with a value of 0.001.

The average total length of ROH > 1 Mb for Haflinger and Noriker were 197 Mb and 100 Mb, respectively. Considering the approach of Curik et al. (2014) to determine the number of generations back to the common ancestor, approximately three generations ago (ROH > 16 Mb) more inbreeding events occurred in the Haflinger population than in the endangered Noriker population. Figure 17 shows the average total length of ROH for each category and breed in Mb.

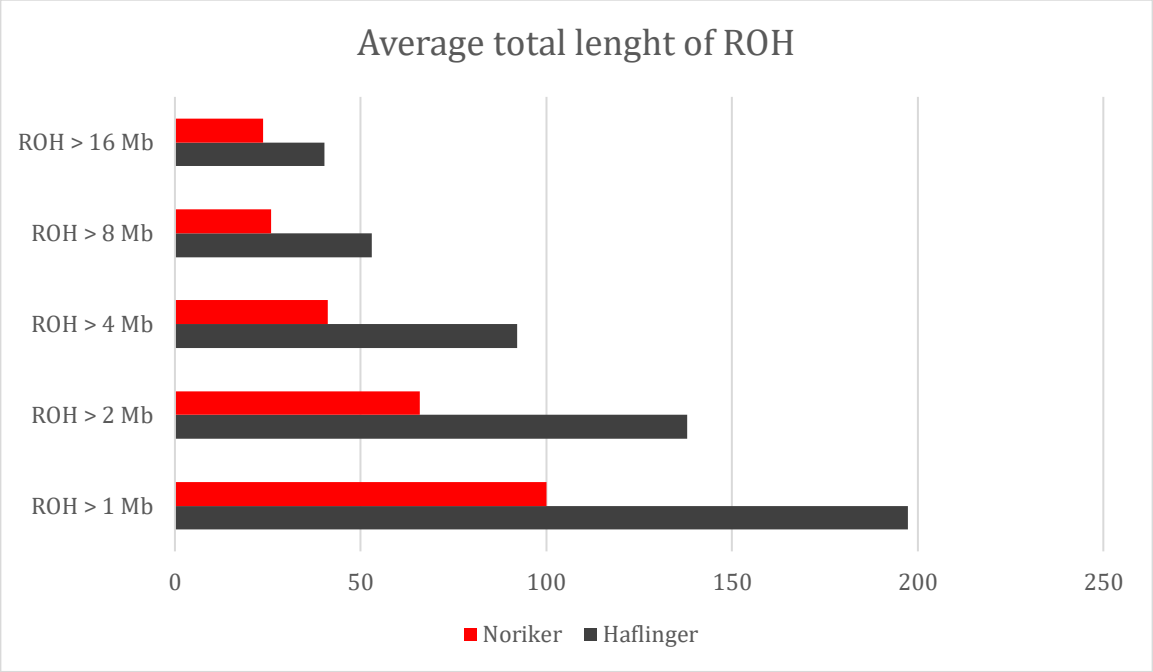


Figure 17: Average total length of ROH in Mb for the categories >1, >2, >4, >8, >16 Mb for a data set of 79 Haflinger and 163 Noriker

All horses had ROH longer than 1 Mb and 95 % of the individuals exhibited ROH longer than 2 Mb. Hitherto, analysis of ROH have scarcely been conducted for horse populations. Metzger et al. (2015) have analysed frequency and number of ROH for a very small group of horses, however, they were able to reveal signatures of positive selection for reproduction traits.

In the present study, 98 % of all Noriker horses display ROH on chromosome three (105 - 110 Mb). The Haflinger population exhibits ROH on the third chromosome as well between 30 and 40 Mb. ROH on chromosome three for both breeds are presented below in

figure 18. On chromosome three, which is 119.48 Mb long, a total of 1,016 genes have been located for the horse genome.

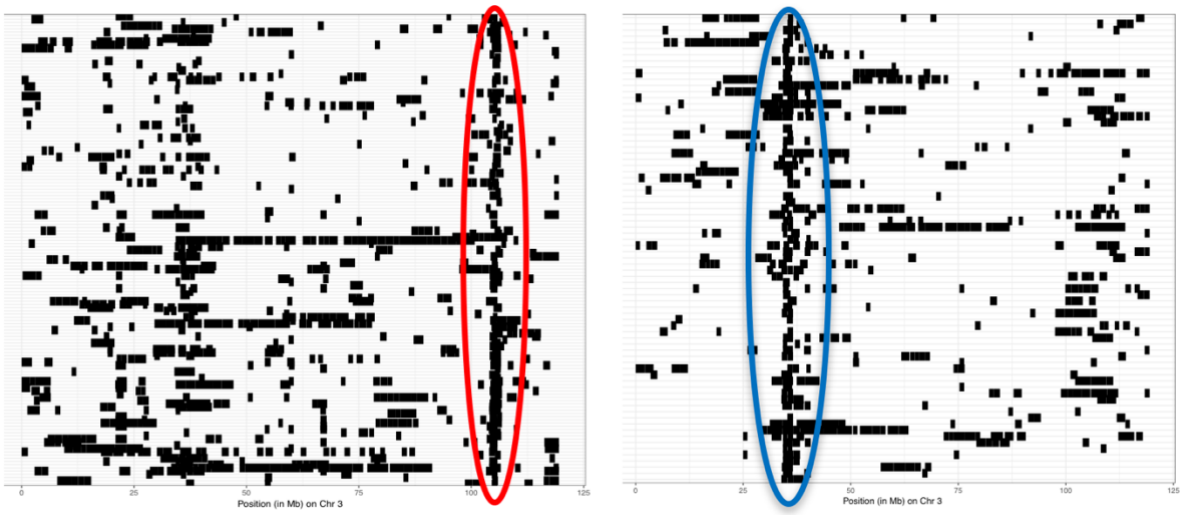


Figure 18: ROH on chromosome three for the Noriker (left) and Haflinger (right) breed

The Noriker population shows breed specific ROH at chromosome six and at chromosome eleven (both positioned around 30 Mb). Individuals of the Haflinger population display ROH at chromosome fifteen (62 – 68 Mb). These ROH within populations might imply signatures of selection and identification of genes, located in these areas, is strongly recommended (Metzger et al. 2015). Deserts are present on chromosome six (40 – 50 Mb) and chromosome 23 (at 40 Mb), further on chromosome 26 and 28 (both at 25 Mb), for Noriker and Haflinger, respectively. This overview of islands and deserts in Austria's horse population indicates the necessity of further research in this area.

5 Conclusion

To assess genetic diversity in horses is of particular significance due to the large number of already extinct breeds and the need for continuous adaptation to the current demands. Several studies have already investigated genetic diversity within and across horse breeds. However, heretofore, an evaluation of the Austrian horse population on the basis of genomic marker data has not been conducted.

The various methods, that have been utilized in this thesis, are capable of giving a comprehensive overview of the population structure and the present genetic diversity. The PCA has divided the Haflinger and Noriker populations clearly, with the exception of three individuals. These individuals may be showing anomalous clustering due to several reasons. As suggested in literature, there might have been an occurrence of crossbreeding between the two breeds. However, it must also be stated that these results could have been caused by sampling errors. Unfortunately, there has not been any information available on the sampling process of the utilized data. If the genotyped samples, for example, have been gathered from only one stud farm they would not sufficiently represent the entire population.

By looking at the phylogenetic tree based on Nei's genetic distances between individuals the clear distinction between Haflinger and Noriker horses presented in the PCA is reinforced. To better understand the results of the phylogenetic tree and to clarify the origin of the Haflinger and Noriker breed, it would be valuable to gather further information on other Austrian breeds. Additionally, a subpopulation of Noriker similar to the Haflinger population was identified. The subpopulation seems to have common features with the Haflinger population, but further investigations are needed to confirm the hypothesis.

Estimations of IBS and IBD have not been able to show as much disparity between populations as has been expected. Nevertheless, it can be concluded that the individuals of the Haflinger population are more interrelated than the Noriker population.

The inbreeding coefficients for both breeds are much lower than expected, particularly, because an endangered breed such as the Noriker, is expected to be strongly inbred due to a small population size. Moreover, there seems to have been a further decline in inbreeding, with regards to both populations, during these last preceding years. The inbreeding coefficient for the Austrian Haflinger population has also been identified as alarmingly increasing in literature, although the findings of this thesis suggest the opposite. According to F_{ROH} values, however, the Haflinger population has been more inbred than the Noriker population. The continuous monitoring of the inbreeding coefficient in both breeds is advisable.

In addition to calculating inbreeding levels, ROH have been used to detect the locations of abundance in diversity and lack of diversity on the Haflinger and Noriker genome. Various ROH islands and deserts have been found and they would need to be further investigated to identify the reasons behind their occurrence.

Moreover, while this thesis has been able to give an overview of the genetic diversity of the Austrian horse population, it demands for a more specific analysis of the population fine structure. In line with the objectives of the "Global Plan of Action for Animal Genetic Resources", it would be beneficial to merge all existing data on equine genetic diversity to facilitate further investigations. Only the greatest level of combined efforts can enable genetic diversity to prosper and to prevent its decline.

References

Affymetrix, Inc. 2015. “Axiom® Equine Genotyping Array - DataSheet.” http://www.affymetrix.com/catalog/prod1010002/AFFY/Axiom%26%23174%3B+Equine+Genotyping+Array#1_3. [23.07.2016]

Almarzook, S. 2017. “Genetic Diversity of Syrian Arabian Horses.” *Animal Genetics* 48 (March). https://doi.org/10.1111/age.12568.

Alvarez, G., Francisco C. Ceballos, and Celsa Quinteiro. 2009. “The Role of Inbreeding in the Extinction of a European Royal Dynasty.” *PLOS ONE* 4 (4):e5174. https://doi.org/10.1371/journal.pone.0005174.

Arche Austria. 2015. “Österreichischer Noriker.” Arche Austria. 2015. http://www.arche-austria.at/index.php?id=86. [12.09.2016]

ARGE Noriker. 2010. “Studbook and Breeding Goals.” Maishofen: Landespferdezuchtverband Salzburg. https://www.pferdezucht-austria.at/download/files/%7B2E2056B8-9100-46A2-9548-1168908295D2%7D/UZB_Noriker_Sbg_2011.pdf. [04.03.2017]

Association Noriker Austria. 2015. “Noriker - Breed Description.” Pferdezucht Austria. 2015. https://www.pferdezucht-austria.at/download/files/%7BD5E2021B-ED62-41F9-9A97-F05869137955%7D/ARGE%20folder%202009%20englisch%20net.pdf. [04.03.2017]

BMLFUW. 2015. “Horse Farming in Austria.” Federal Ministry of Agriculture, Forestry, Environment and Water Management. https://www.bmlfuw.gv.at/english/agriculture/Productionandmarkets/Animalproduction.html. [23.07.2016]

———. 2017. “Animal Production.” Sheep, Lamb, Horse. http://www.agraroekonomik.at/index.php?id=kontakt&L=1. [23.07.2016]

Boettcher, P.J., M. Tixier-Boichard, M.A Toro, H. Simianer, H. Eding, G. Gandini, S. Joost, et al. 2010. “Objectives, Criteria and Methods for Using Molecular Genetic Data in Priority Setting for Conservation of Animal Genetic Resources.” *International Society for Animal Genetics, Animal Genetics*, no. 41:64–77.

Bowling, Ann T., and Anatoly Ruvinsky. 2000. *The Genetics of the Horse*. CABI.

Browning, Sharon R., and Brian L. Browning. 2010. “High-Resolution Detection of Identity by Descent in Unrelated Individuals.” *The American Journal of Human Genetics* 86 (4):526–39. https://doi.org/10.1016/j.ajhg.2010.02.021.

Cañon, J., M. L. Checa, C. Carleos, J. L. Vega-Pla, M. Vallejo, and S. Dunner. 2000. “The Genetic Structure of Spanish Celtic Horse Breeds Inferred from Microsatellite Data.” *Animal Genetics* 31 (1):39–48.

Chakraborty, S., S.K. Bhattacharjee, and G.D. Sharma. 2012. “Comparative Study of Nei’s D with Other Genetic Distance Measures between Barak Valley Muslims and Other Nations for ABO Locus.” *Notulae Scientia Biologicae* 4 (1):12.

Chang, Christopher, Shaun Purcell, C Chow, LCAM Tellier, S Vattikuti, and JJ Lee. 2015. “Second-Generation PLINK: Rising to the Challenge of Larger and Richer Datasets.” *GigaScience* 4.

Charlesworth, D, and John H. Willis. 2009. “The Genetics of Inbreeding Depression.” *Nature Reviews Genetics* 10 (11):783–96. https://doi.org/10.1038/nrg2664.

Chowdhary, BP, N. Paria, and T Raudsepp. 2008. “Potential Applications of Equine Genomics in Dissecting Diseases and Fertility.” *Animal Reproductive Science* Volume 107 (Issues 3–4):208–18.

Commission on genetic resources for food and agriculture. 2013. “Status and Trends of Animal Genetic Resources.” Rome.

Commission on genetic resources for food and agriculture, and FAO. 2007. “GLOBAL PLAN OF ACTION FOR ANIMAL GENETIC RESOURCES and the INTERLAKEN DECLARATION.” Rome.

Curik, Ino, Maja Ferenčaković, and Johann Sölkner. 2014. “Inbreeding and Runs of Homozygosity: A Possible Solution to an Old Problem.” *Livestock Science* 166 (August):26–34. https://doi.org/10.1016/j.livsci.2014.05.034.

Druml, T., R. Baumung, and J. Sölkner. 2009. “Pedigree Analysis in the Austrian Noriker Draught Horse: Genetic Diversity and the Impact of Breeding for Coat Colour on Population Structure.” *Journal of Animal Breeding and Genetics* 126 (5):348–56. https://doi.org/10.1111/j.1439-0388.2008.00790.x.

Druml, T., Katharina Sauer, Julia Elsbacher, Gertrud Grilz-Seger, and Gottfried Brem. 2016. “Analysis of the Gene Pool, the Genetic Diversity and Inbreeding Ratios of Austrian Haf Linger Population.” *Züchtungskunde* 88 (September):379–394.

Druml, T., R. Baumung, J. Sölkner. 2007. “Morphological Analysis and Effect of Selection for Conformation in the Noriker Draught Horse Population.” Livestock Science 115.

Ellegren, H., and Nicolas Galtier. 2016. “Determinants of Genetic Diversity.” *Nature Reviews Genetics* 17 (7):422–33. https://doi.org/10.1038/nrg.2016.58.

FAO. 2007. “The State of the World’s Animal Genetic Resources for Food and Agriculture.” Rome.

———. 2014. “Live Animals - Food and Agriculture Organization of the United Nations.” http://www.fao.org/faostat/en/?#data/QA/visualize. [18.05.2017]

———. 2015. “The Second Report on The State of the World’s Animal Genetic Resources for Food and Agriculture.” Rome: commission on genetic resources for food and agriculture. [28.11.2017]

Felicetti, M., M. S. Lopes, A. Verini-Supplizi, A. Da Câmara Machado, M. Silvestrelli, D. Mendonça, and O. Distl. 2010. “Genetic Diversity in the Maremmano Horse and Its Relationship with Other European Horse Breeds.” *Animal Genetics* 41 Suppl 2 (December):53–55. https://doi.org/10.1111/j.1365-2052.2010.02102.x.

Ferenčaković, M., Johann Sölkner, and Ino Curik. 2013. “Estimating Autozygosity from High-Throughput Information: Effects of SNP Density and Genotyping Errors.” *Genetics Selection Evolution* 45:42. https://doi.org/10.1186/1297-9686-45-42.

Frankham, R., Ballou, J. D. & Briscoe, D. A. 2002. *Introduction to Conservation Genetics. Cambridge University Press. - Google-Suche*. Cambridge University Press:67-92.

Glowatzki-Mullis, M. L., J. Muntwyler, W. Pfister, E. Marti, S. Rieder, P. A. Poncet, and C. Gaillard. 2006. “Genetic Diversity among Horse Populations with a Special Focus on the Franches-Montagnes Breed.” *Animal Genetics* 37 (1):33–39. https://doi.org/10.1111/j.1365-2052.2005.01376.x.

Goto, H., Oliver A. Ryder, Allison R. Fisher, Bryant Schultz, Sergei L. Kosakovsky Pond, Anton Nekrutenko, and Kateryna D. Makova. 2011. “A Massively Parallel Sequencing Approach Uncovers Ancient Origins and High Genetic Variability of Endangered Przewalski’s Horses.” *Genome Biology and Evolution* 3:1096–1106. https://doi.org/10.1093/gbe/evr067.

Gusev, A., Jennifer K. Lowe, Markus Stoffel, Mark J. Daly, David Altshuler, Jan L. Breslow, Jeffrey M. Friedman, and Itsik Pe’er. 2009. “Whole Population, Genome-Wide Mapping of Hidden Relatedness.” *Genome Research* 19 (2):318–26. https://doi.org/10.1101/gr.081398.108.

Hamming, R. W. 1950. “Error Detecting and Error Correcting Codes.” *Bell System Technical Journal* 29 (2):147–60. https://doi.org/10.1002/j.1538-7305.1950.tb00463.x.

Hartl, D.L., and A.G. Clark. 2007. *Principles of Population Genetics*. 4th ed. Macmillan Education.

Holsinger, Kent E., and Bruce S. Weir. 2009. “Genetics in Geographically Structured Populations: Defining, Estimating and Interpreting FST.” *Nature Reviews. Genetics* 10 (9):639–50. https://doi.org/10.1038/nrg2611.

Humane Society International. 2012. “Horsemeat Production Worldwide.” http://www.hsi.org/assets/pdfs/Horses_meatproduction_worldwide.pdf.

Institute for Industrial Research. 2005. “The Economic Role of Horses in Austria.” Pferd Austria.

Jemmali, Bayrem, Mohamed Mezir Haddad, Nawel Barhoumi, Syrine Tounsi, Faten Lasfer, Amira Trabelsi, Belgacem Ben Aoun, et al. 2017. “Genetic Diversity in Tunisian Horse Breeds.” *Arch. Anim. Breed* 60 (June):153–160. https://doi.org/10.5194/aab-60-153-2017.

Jianzhong, Ma, and Christopher I. Amos. 2012. “Principal Components Analysis of Population Admixture.” *PLoS ONE* 7 (e40115).

Jolliffe, I.T. 2002. *Principal Component Analysis*. Second. Springer-Verlag, New York.

Jombart, Thibaut. 2008. “Adegenet: A R Package for the Multivariate Analysis of Genetic Markers.” *Bioinformatics* 24 (11):1403–5. https://doi.org/10.1093/bioinformatics/btn129.

Leroy, Grégoire, Lucille Callède, Etienne Verrier, Jean-Claude Mériaux, Anne Ricard, Coralie Danchin-Burge, and Xavier Rognon. 2009. “Genetic Diversity of a Large Set of Horse Breeds Raised in France Assessed by Microsatellite Polymorphism.” *Genetics, Selection, Evolution : GSE* 41 (1):5. https://doi.org/10.1186/1297-9686-41-5.

Li, Meng-Hua, Ismo Strandén, Timo Tiirikka, Marja-Liisa Sevón-Aimonen, and Juha Kantanen. 2011. “A Comparison of Approaches to Estimate the Inbreeding Coefficient and Pairwise Relatedness Using Genomic and Pedigree Data in a Sheep Population.” Edited by Petter Holme. *PLoS ONE* 6 (11):e26256. https://doi.org/10.1371/journal.pone.0026256.

Lippold, Sebastian, Nicholas J. Matzke, Monika Reissmann, and Michael Hofreiter. 2011. “Whole Mitochondrial Genome Sequencing of Domestic Horses Reveals Incorporation of Extensive Wild Horse Diversity during Domestication.” *BMC Evolutionary Biology* 11 (November):328. https://doi.org/10.1186/1471-2148-11-328.

McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, et al. 2012. “A High Density SNP Array for the Domestic Horse and Extant Perissodactyla: Utility for Association Mapping, Genetic Diversity, and Phylogeny Studies.” PLoS Genet 8(1): e1002451. doi:10.1371/journal.pgen.1002451.

Menozzi, P., A. Piazza, and L. Cavalli-Sforza. 1978. “Synthetic Maps of Human Gene Frequencies in Europeans.” *Science* 201 (4358):786–92.

Messner, E. 2014. “The Haflinger - History.” Haflinger.eu. 2014. http://www.haflinger.eu/en/haflinger/index.asp?MAID=343&LG=3&APP=8&NKey=haflinger. [18.05.2017]

Metzger, Julia, Matthias Karwath, Raul Tonda, Sergi Beltran, Lídia Águeda, Marta Gut, Ivo Glynne Gut, and Ottmar Distl. 2015. “Runs of Homozygosity Reveal Signatures of Positive Selection for Reproduction Traits in Breed and Non-Breed Horses.” *BMC Genomics* 16 (October):764. https://doi.org/10.1186/s12864-015-1977-3.

Morawetz, E., 2010. “Die Schattenseite Der Zucht | Pferderevue | Allgemein.” Pferderevue. 2010. http://www.pferderevue.at/?id=2500%2C4822097%2C%2C. [12.09.2016]

MR Dipl.-Ing. Dr. Johannes Frickh. 2012. “Die Bedeutung Der Pferdewirtschaft in Österreich.” Raumberg-Gumpenstein: lebensministerium.at.

National Center for Biotechnology Information. 2017. “Genome Equus Caballus.” NCBI. July 23, 2017. https://www.ncbi.nlm.nih.gov/genome?term=equus%20caballus. [18.05.2017]

Nei, Masatoshi. 1972. “Genetic Distance between Populations.” *The American Naturalist* 106 (949):283–92. https://doi.org/10.1086/282771.
Nothnagel, Michael, Timothy Tehua Lu, Manfred Kayser, and Michael Krawczak. 2010. “Genomic and Geographic Distribution of SNP-Defined Runs of Homozygosity in Europeans.” *Human Molecular Genetics* 19 (15):2927–35. https://doi.org/10.1093/hmg/ddq198.
Oliehoek, P.A., P. Bijma, and A. van der Meijden. 2009. “History and Structure of the Closed Pedigreed Population of Icelandic Sheepdogs.” *Genetics Selection Evolution* 41:39.
ÖNGENE. 2015. “Erhaltung Gefährdeter Nutztierrassen Im ÖPUL 2015.” Arche Austria. 2015. http://www.arche-austria.at/fileadmin/archeaustria/arche_austria/arche_austria/merkblatt_2015.pdf. [30.06.2017]
Paradis, E., J. Claude, and K. Strimmer. 2004. “APE: Analyses of Phylogenetics and Evolution in R Language.” *Bioinformatics* 20:289–90. https://doi.org/10.1093/bioinformatics/btg412.
Pearson, K. 1901. “On Lines and Planes of Closest Fit to Systems of Points in Space.” *Phil. Mag* 2:559–72. https://doi.org/10.1080/14786440109462720.
Pemberton, Trevor J., Devin Absher, Marcus W. Feldman, Richard M. Myers, Noah A. Rosenberg, and Jun Z. Li. 2012. “Genomic Patterns of Homozygosity in Worldwide Human Populations.” *The American Journal of Human Genetics* 91 (2):275–92. https://doi.org/10.1016/j.ajhg.2012.06.014.
Pembleton, LW, NOI Cogan, and JW Forster. 2013. “StAMPP: An R Package for Calculation of Genetic Differentiation and Structure of Mixed-Ploidy Level Populations.” *Molecular Ecology Resources* 13 (5):946–52.
Petersen JL, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, et al. 2013. “Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP Data.” *PLoS ONE* 8 (1). https://doi.org/10.1371/journal.pone.0054997.
PferdAustria. 2015. “Traditional Horse Breeds of Austria.”
Plante, Yves, Jose Luis Vega-Pla, Zoe Lucas, Dave Colling, Brigitte de March, and Fiona Buchanan. 2007. “Genetic Diversity in a Feral Horse Population from Sable Island, Canada.” *The Journal of Heredity* 98 (6):594–602. https://doi.org/10.1093/jhered/esm064.
Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A. R. Ferreira, David Bender, Julian Maller, et al. 2007. “PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses.” *The American Journal of Human Genetics* 81 (3):559–75. https://doi.org/10.1086/519795.
Purcell, S., C. Chang, CC. Chow, LCAM. Tellier, S. Vattikuti, and JJ. Lee. 2015. *PLINK* (version 1.9). GigaScience 4. www.cog-genomics.org/plink/1.9/.
R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
Ruane, J., 2000. “A Framework for Prioritizing Domestic Animal Breeds for Conservation Purposes at the National Level: A Nor- Wegian Case Study.” 14. Conservation Biology.
Saitou, N., and M. Nei. 1987. “The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees.” *Molecular Biology and Evolution* 4 (4):406–25.
Shrestha, Merina. 2017. “Genetics of Equine Insect Bite Hypersensitivity and Genetic Diversity in Horses.” Sweden: Swedish University of Agricultural Sciences and the Graduate School Wageningen Institute of Animal Sciences of Wageningen University. http://dx.doi.org/10.18174/396810.
Simianer, H., 2005. “Decision Making in Livestock Conservation.” 53. Ecological Economics.

Sölkner, Johann, Maja Ferencakovic, Birgit Gredler, and Ino Curik. 2010. “Genomic Metrics of Individual Autozygosity, Applied to a Cattle Population.” *Book of Abstracts of the 61st Annual Meeting of the European Association of Animal Production*, January.

Toro, M.A, and A Caballero. 2005. “Characterization and Conservation of Genetic Diversity in Subdivided Populations.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 360 (1459):1367–78. https://doi.org/10.1098/rstb.2005.1680.

Toro, M.A., A. Caballero, and J. Fernández. 2009. “Molecular Character- Ization of Breeds and Its Use in Conservation.” 120. Livestock Science.

Tyrolean Haflinger Breeding Association. 2016. “History of the Haflinger Breed.” 2016. http://www.haflinger-tirol.com/en/haflinger-breeding-association/history-of-the-haflinger-breed.html.

Vázquez-Armijo, José, Manuel Parra-Bracamonte, Miguel Velazquez, Ana María Sifuentes-Rincón, José Luis Tinoco-Jaramillo, Pascuala Ambriz-Morales, Williams Arellano-Vera, and Victor Ricardo Moreno-Medina. 2017. “Diversity and Effective Population Size of Four Horse Breeds from Microsatellite DNA Markers in South-Central Mexico.” *Archiv Fur Tierzucht - Archives Animal Breeding* 60 (June):137. https://doi.org/10.5194/aab-60-137-2017.

Wade, CM, E Giulotto, S Sigurdsson, M Zoli, S Gnerre, F Imsland, TL Lear, et al. 2009. “Genome Sequence, Comparative Analysis and Population Genetics of the Domestic Horse (Equus Caballus).” *Science (New York, N.Y.)* 326 (5954):865–67. https://doi.org/10.1126/science.1178158.

Weir, Bruce S., and C. Clark Cockerham. 1984. “Estimating F-Statistics for the Analysis of Population Structure.” *Evolution* 38 (6):1358–70.

Weitzman, M.L., 1993. “What to Preserve? An Application of Diversity Theory to Crane Conservation.” 108. Quarterly Journal of Economy.

Wickham, H. 2009. “ggplot2: Elegant Graphics for Data Analysis.” *Springer-Verlag New York*. http://ggplot2.org.

Wright, S. 1922. “Coefficients of Inbreeding and Relationship.” *The American Naturalist* 56 (645):330–38. https://doi.org/10.1086/279872.

———. 1949. “The Genetical Structure of Populations.” *Annals of Eugenics* 15:323–354.

Wright, S. 1984. *Evolution and the Genetics of Populations, Volume 4: Variability Within and Among Natural Populations*. University of Chicago Press.

ZAP. 2011. “Annual Report 2011.” Jahresbericht. Zentrale Arbeitsgemeinschaft für Pferdezüchter.

———. 2015a. “Noriker - Portrait of the Breed.” 2015. https://www.pferdezucht-austria.at/main.asp?kat1=13&kat2=347&kat3=308&vid=1&_=2. [18.05.2017]

———. 2015b. “Pferdezucht Austria - Breeds / Haflinger / Portrait of the Breeds.” 2015. https://www.pferdezucht-austria.at/main.asp?kat1=13&kat2=142&kat3=309&_=2. [20.05.2017]

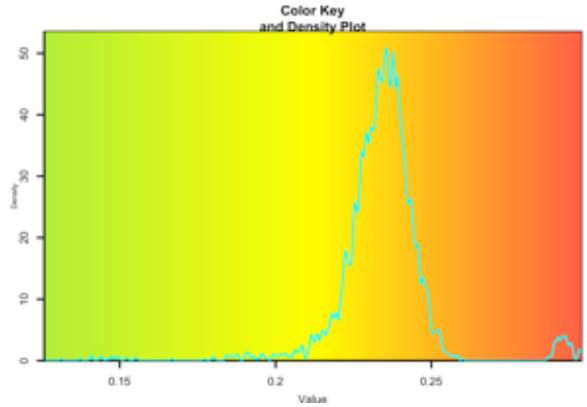
Zhang, Li, Mohammed S. Orloff, Sean Reber, Shengchao Li, Ye Zhao, and Charis Eng. 2013. “cgaTOH: Extended Approach for Identifying Tracts of Homozygosity.” *PLOS ONE* 8 (3):e57772. https://doi.org/10.1371/journal.pone.0057772.

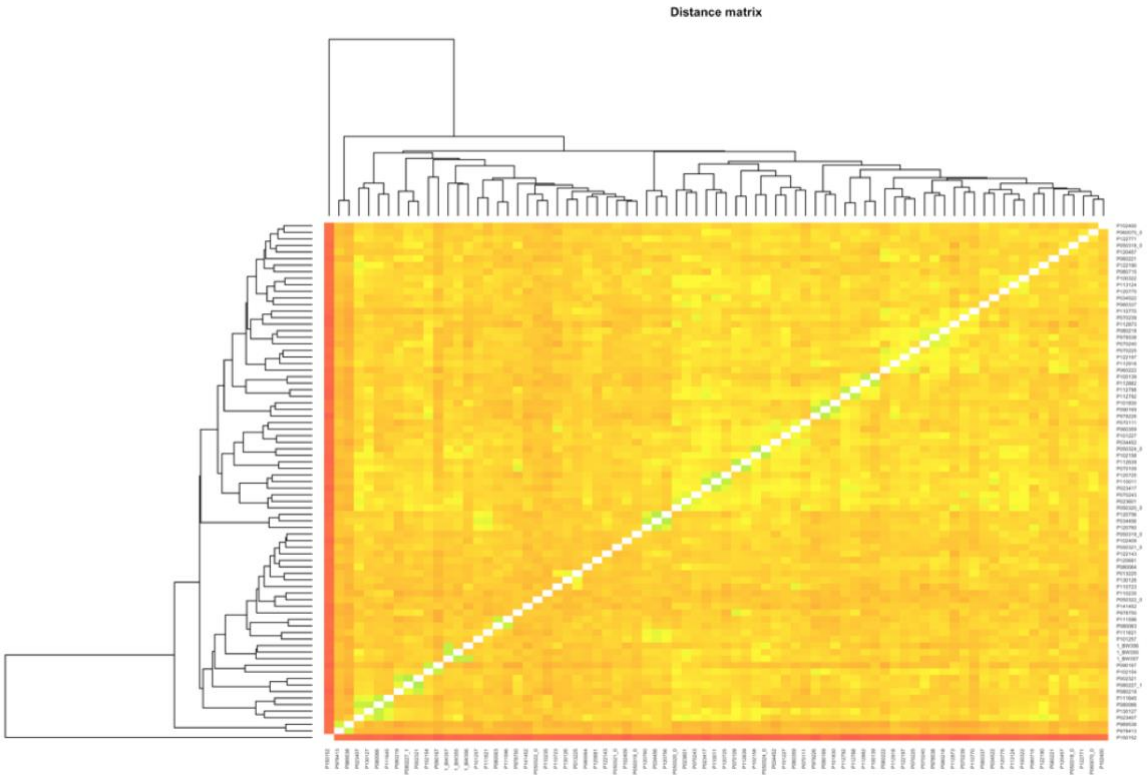
Table of Figures

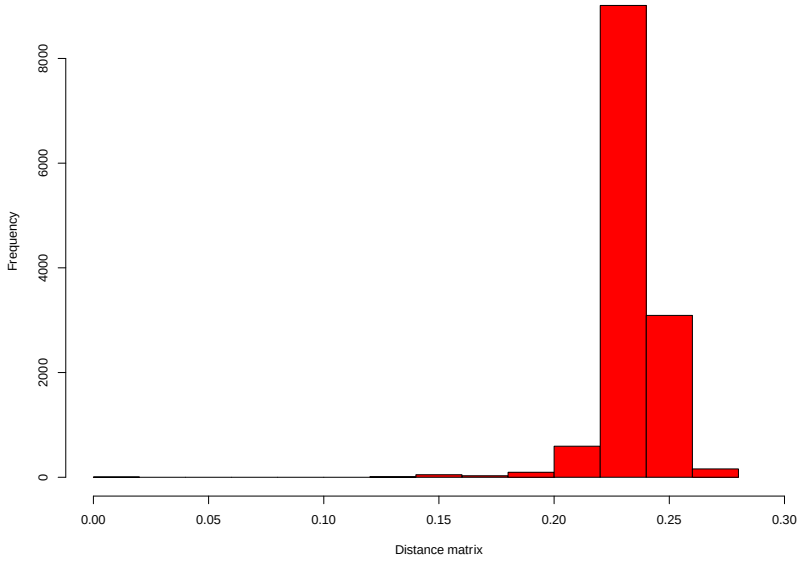
Appendix

I. Heatmaps/histograms for distance matrices

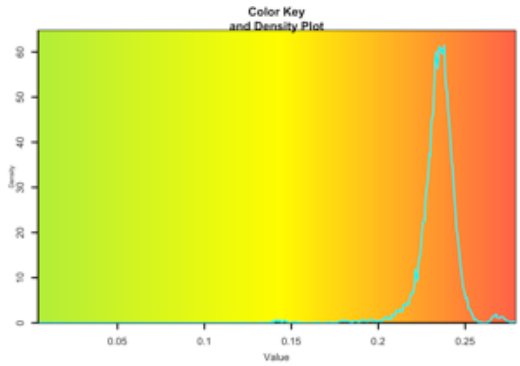
Noriker

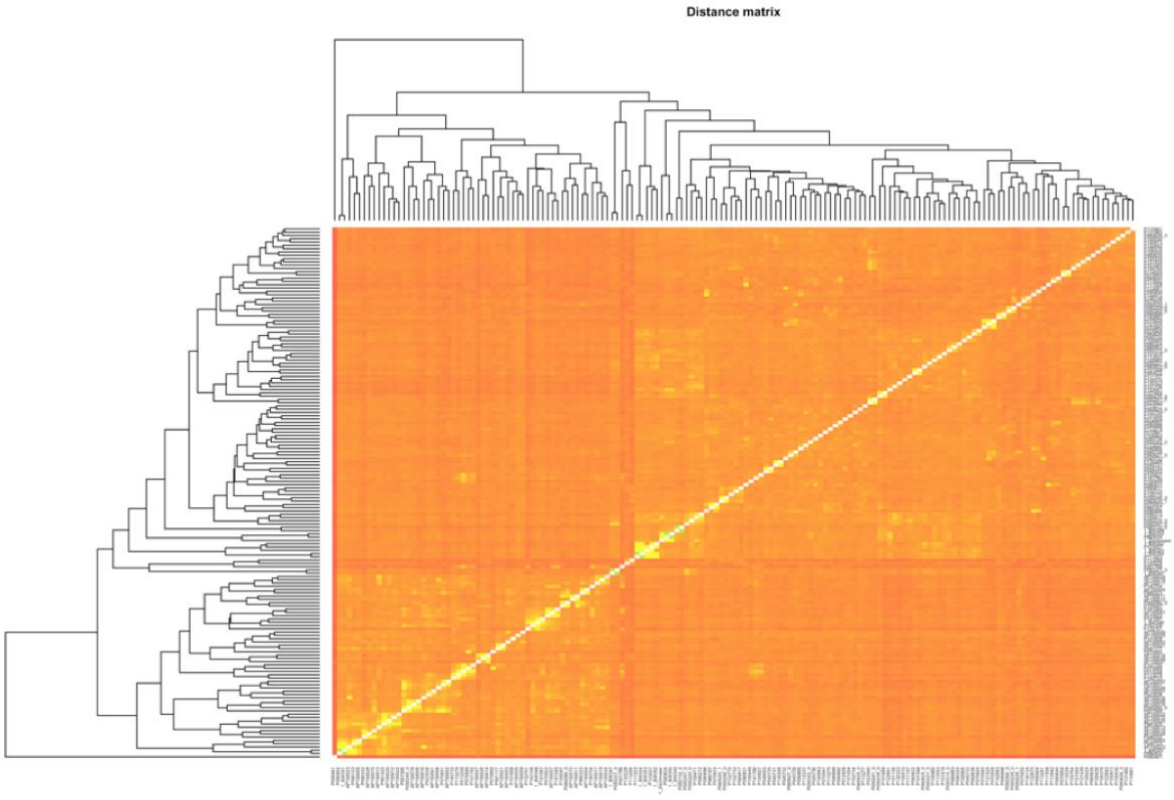


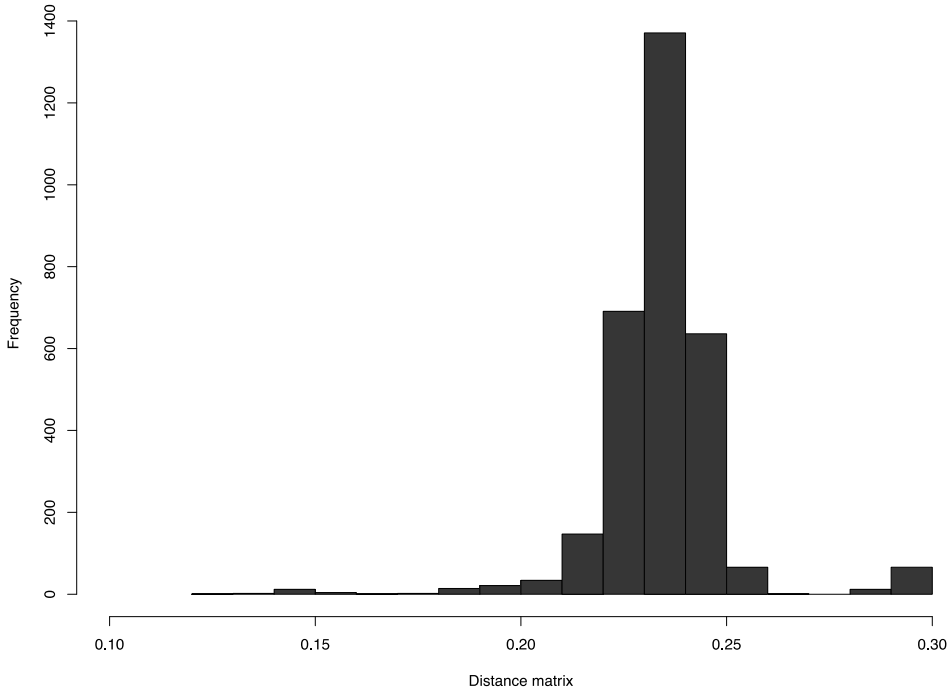



Frequency
0
2000
4000
6000
8000
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Distance matrix

Haflinger


Color Key
and Density Plot
Density
0
10
20
30
40
50
60
0.05
0.1
0.15
0.2
0.25
Value

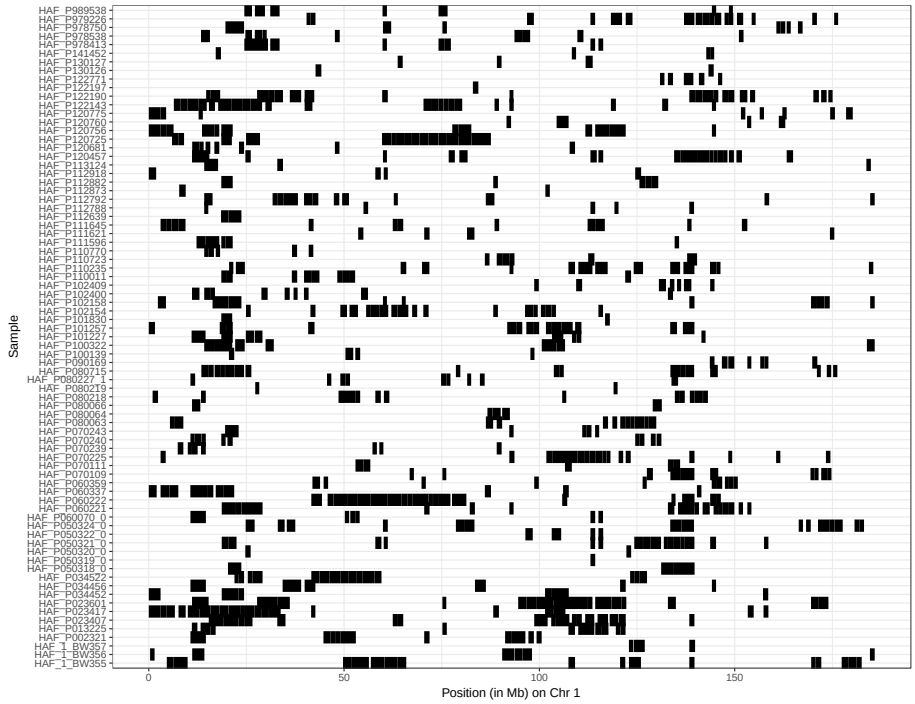

Distance matrix


Frequency
0
200
400
600
800
1000
1200
1400
0.10
0.15
0.20
0.25
0.30
Distance matrix

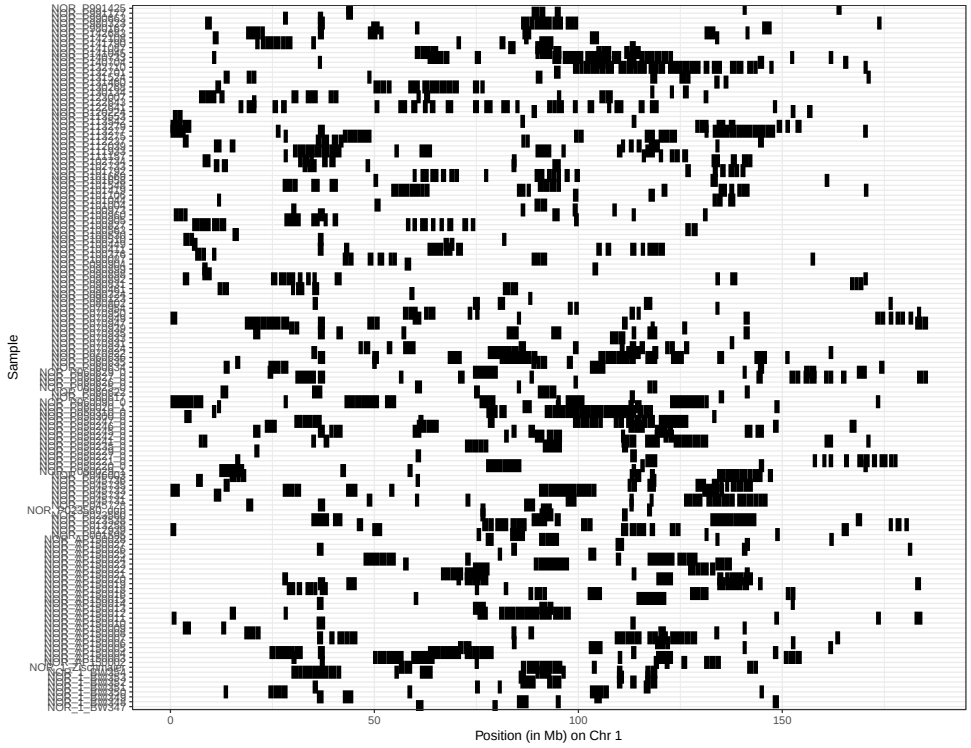
II. Runs of homozygosity >1 Mb

Chromosome 1

Haflinger

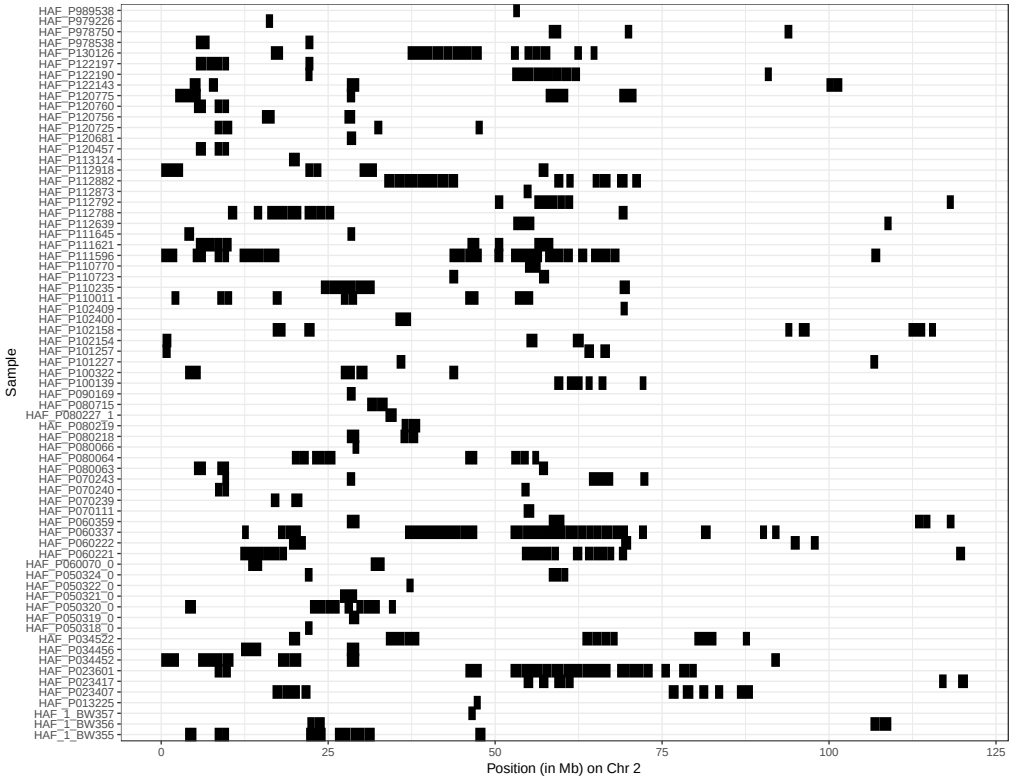


Noriker

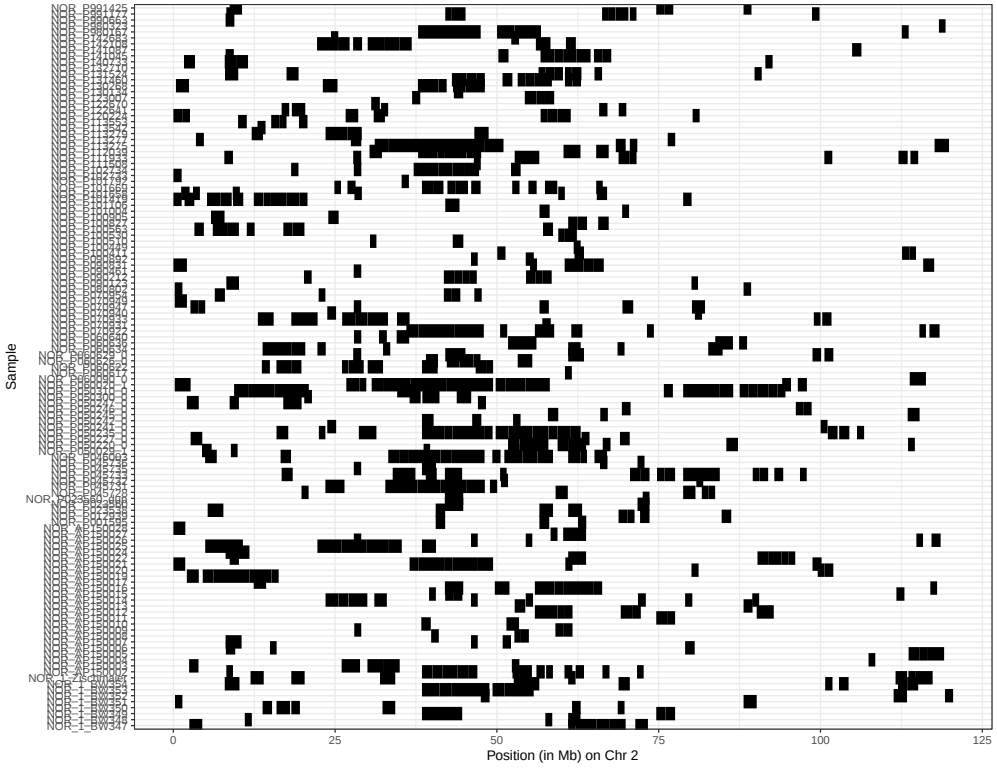


Chromosome 2

Haflinger

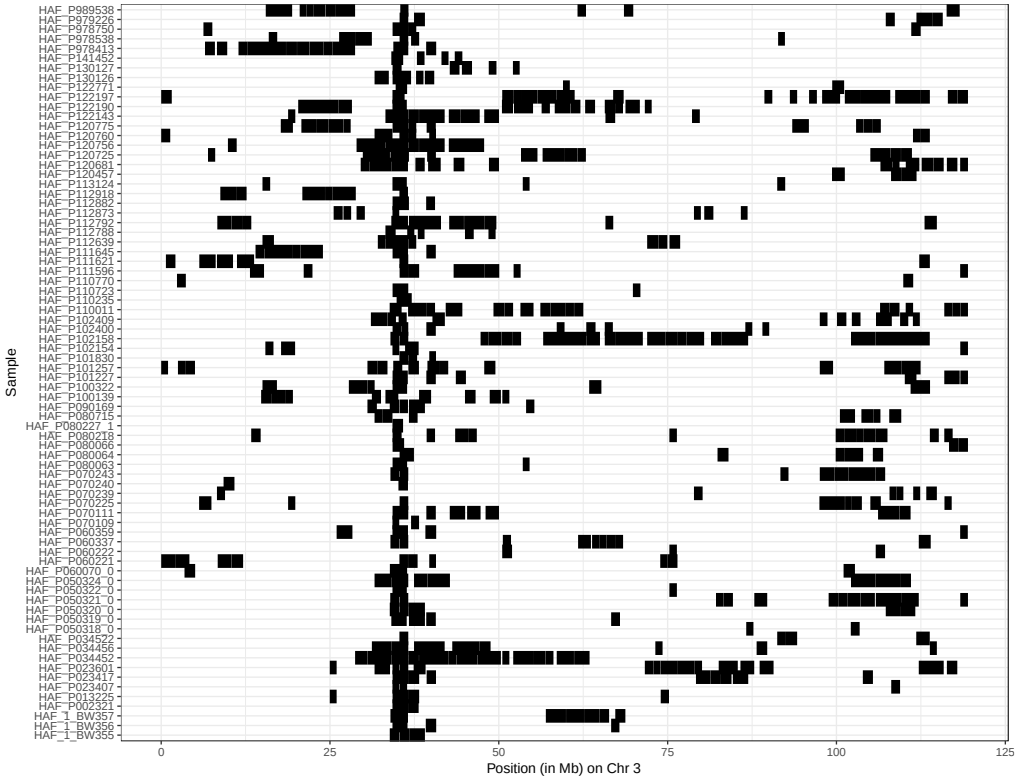


Noriker



Chromosome 3

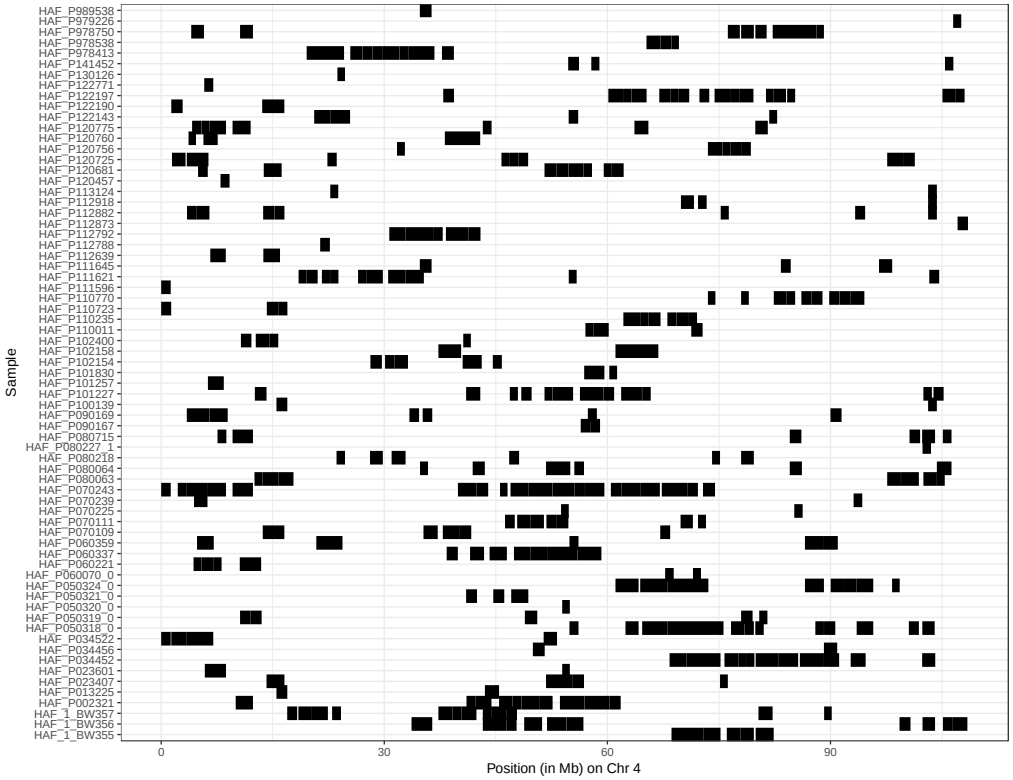
Haflinger



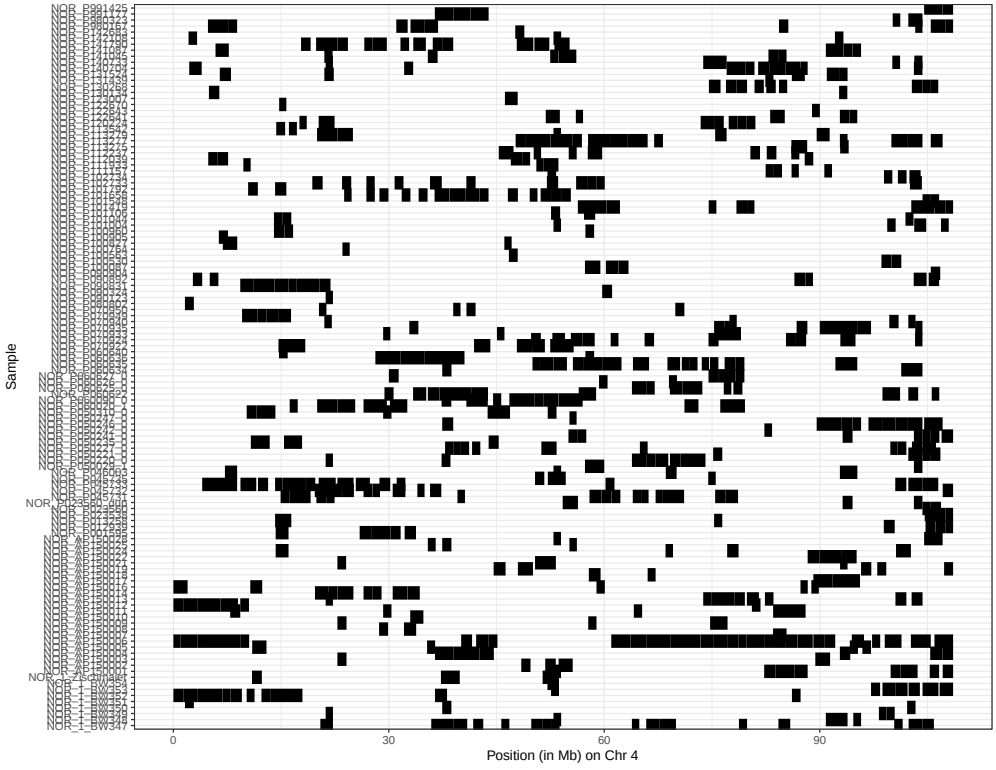
Noriker

Chromosome 4

Haflinger

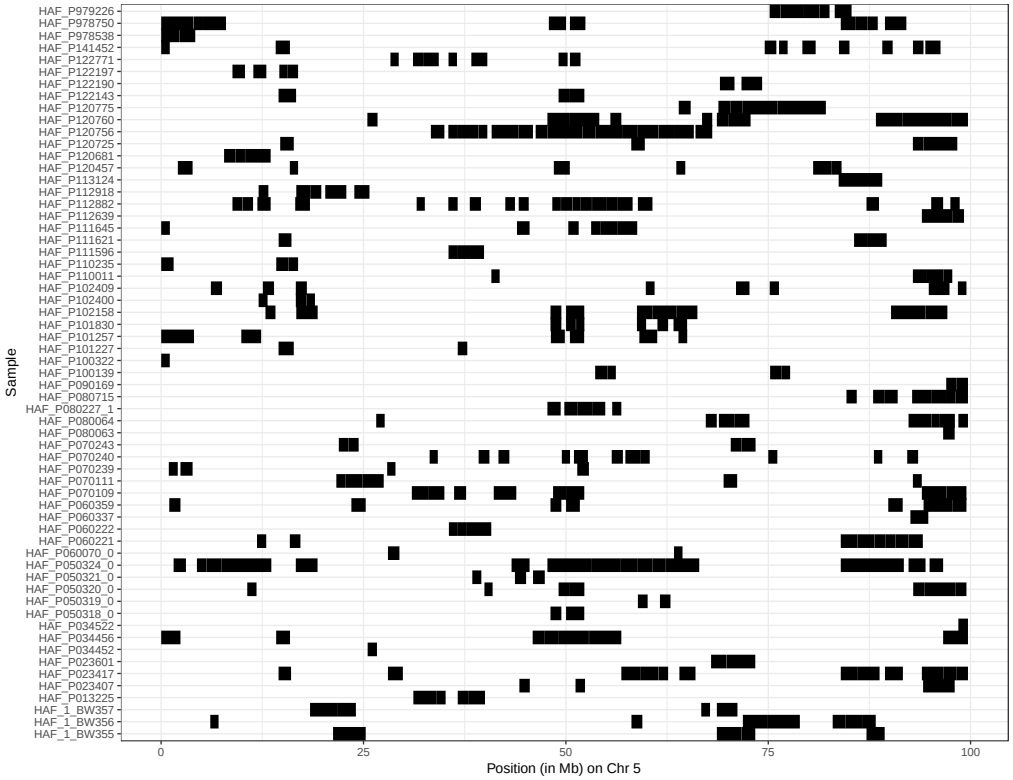


Noriker

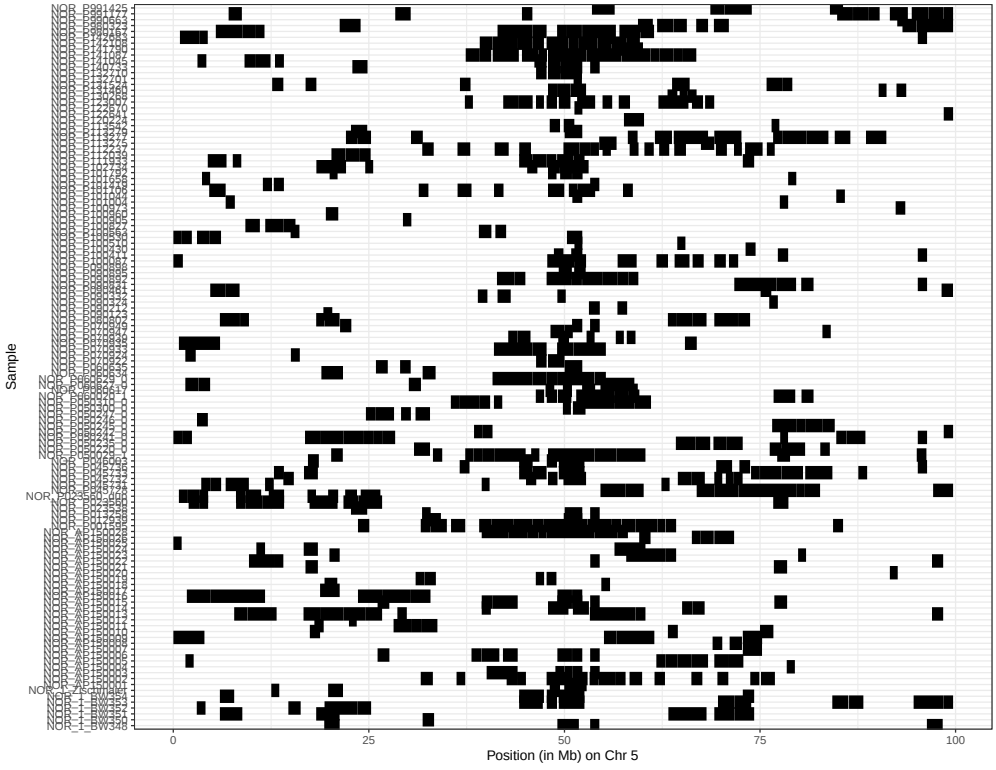


Chromosome 5

Haflinger

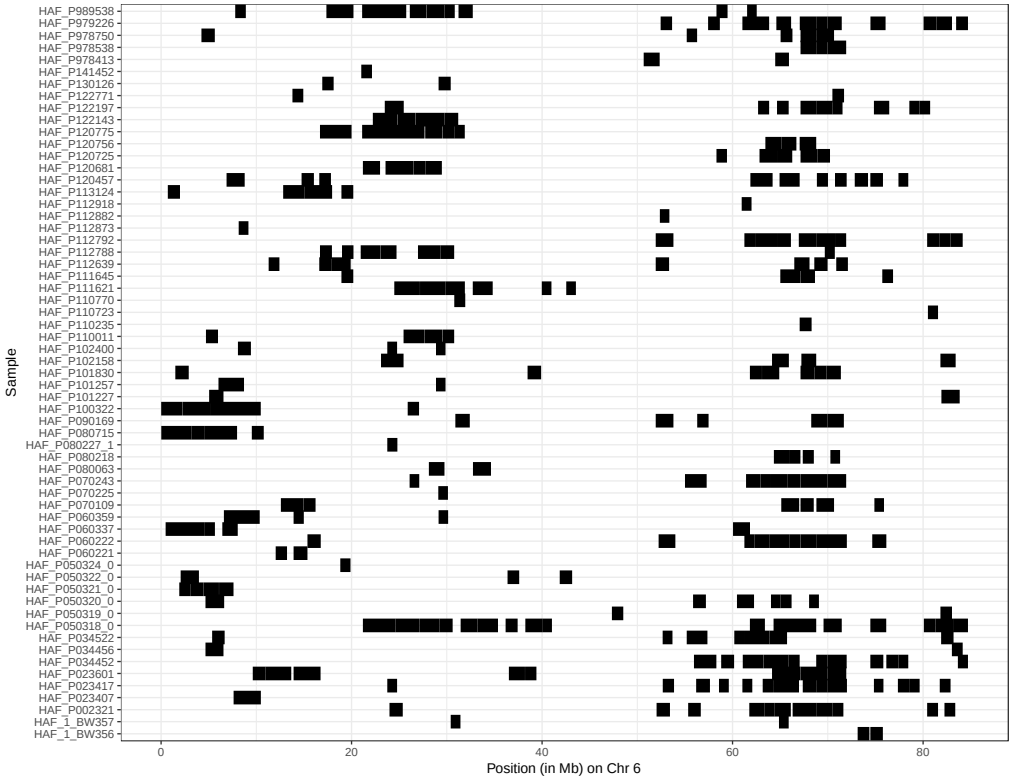


Noriker

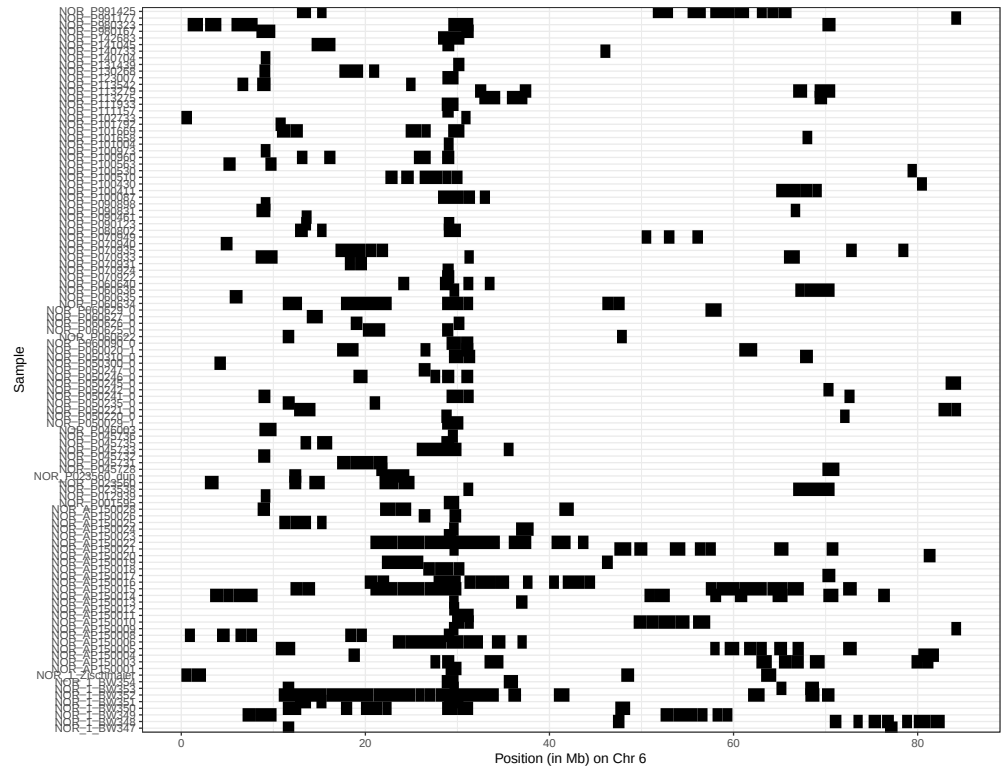


Chromosome 6

Haflinger

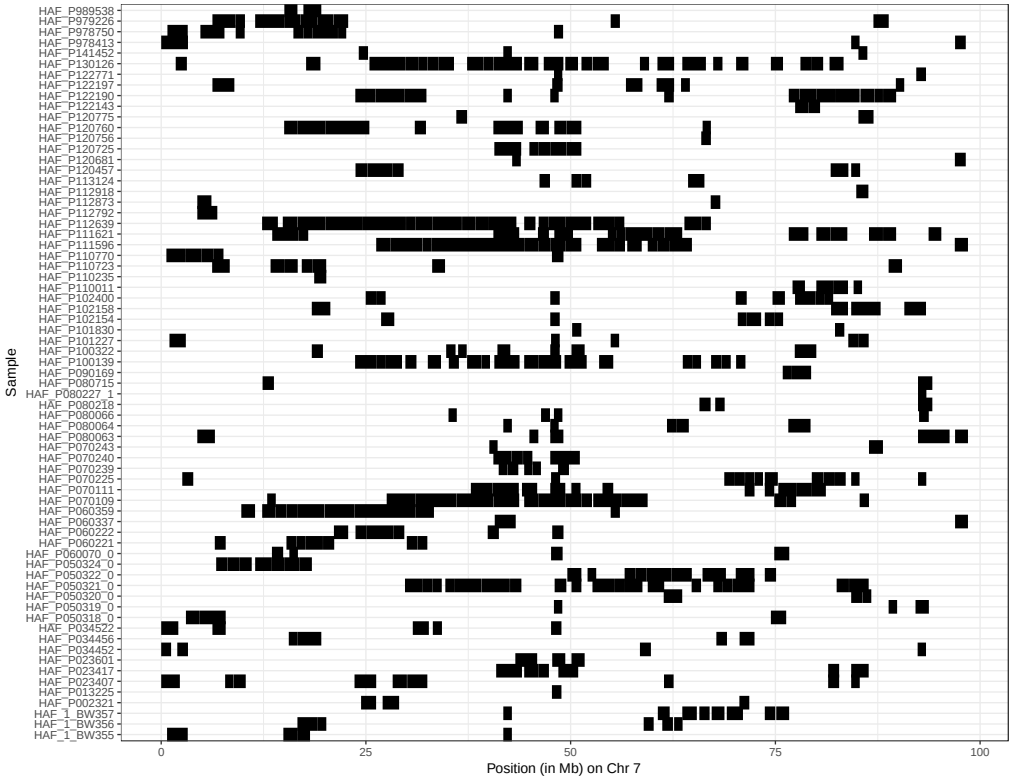


Noriker

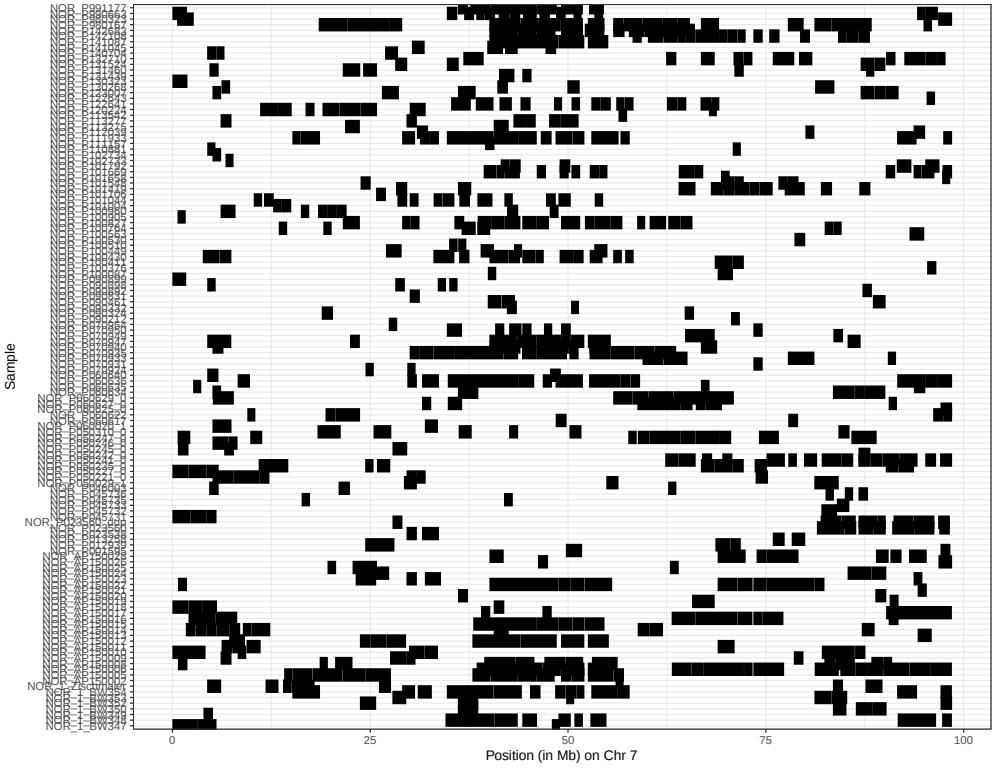


Chromosome 7

Haflinger



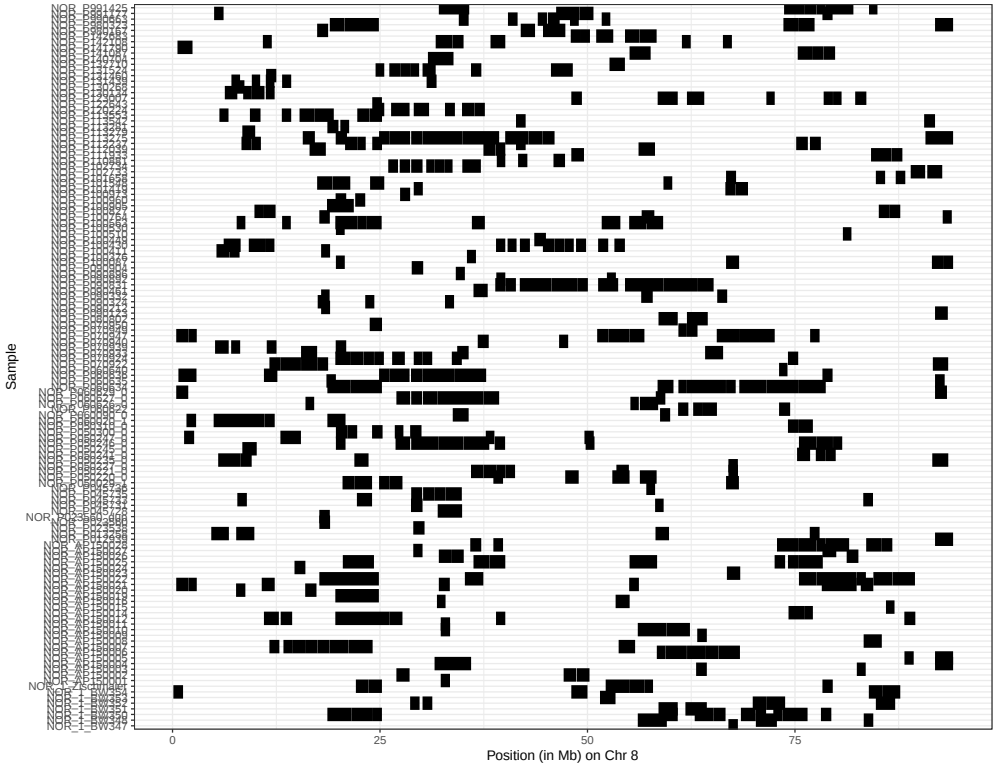
Noriker



Chromosome 8

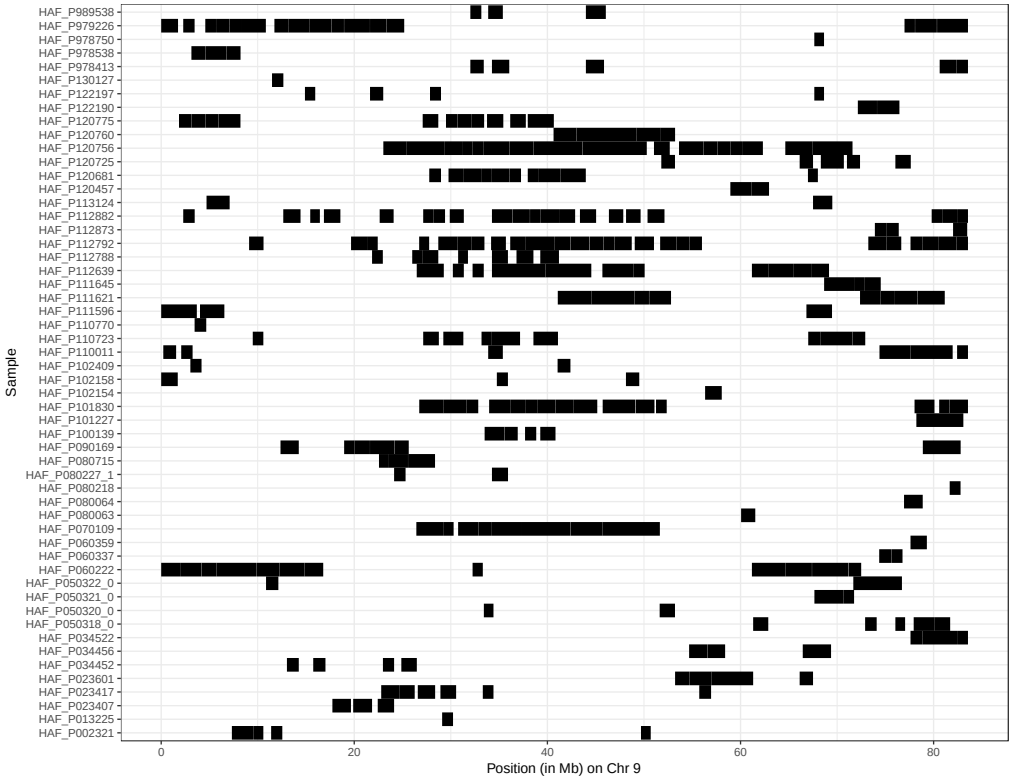
Haflinger

Noriker

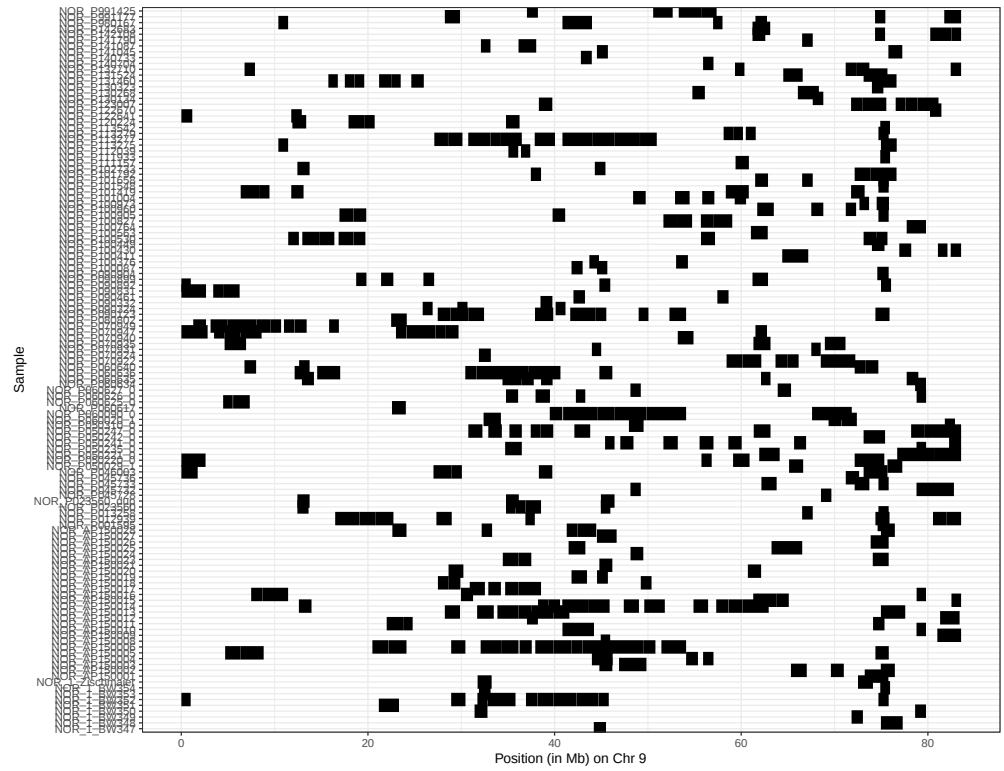


Chromosome 9

Haflinger

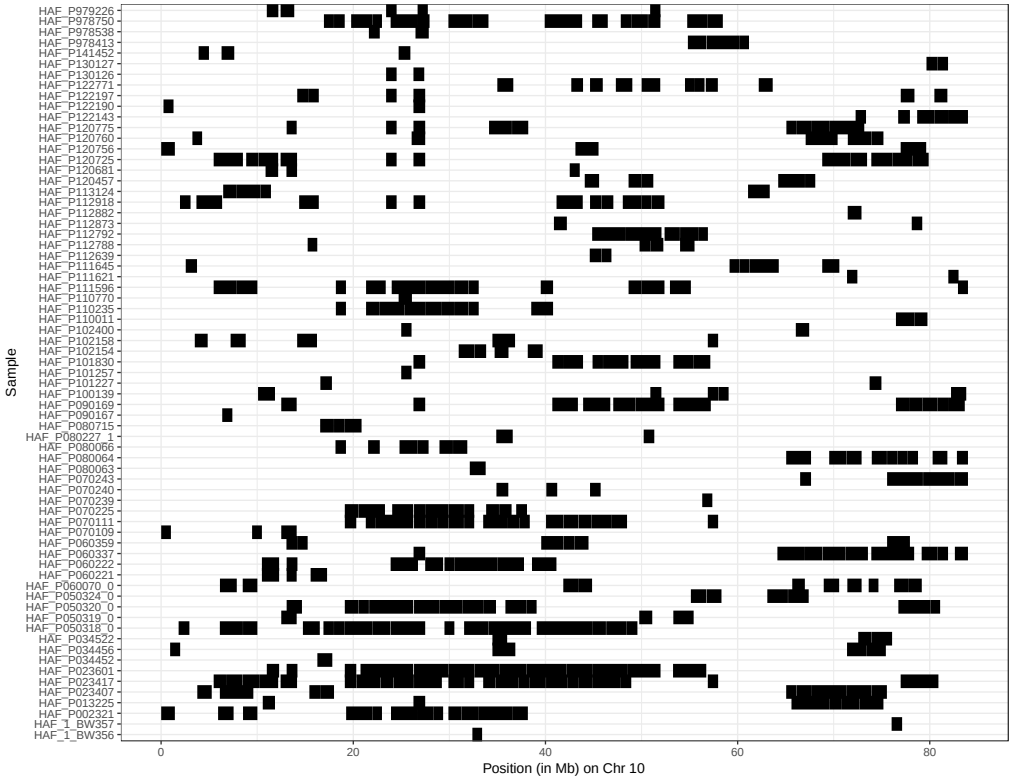


Noriker

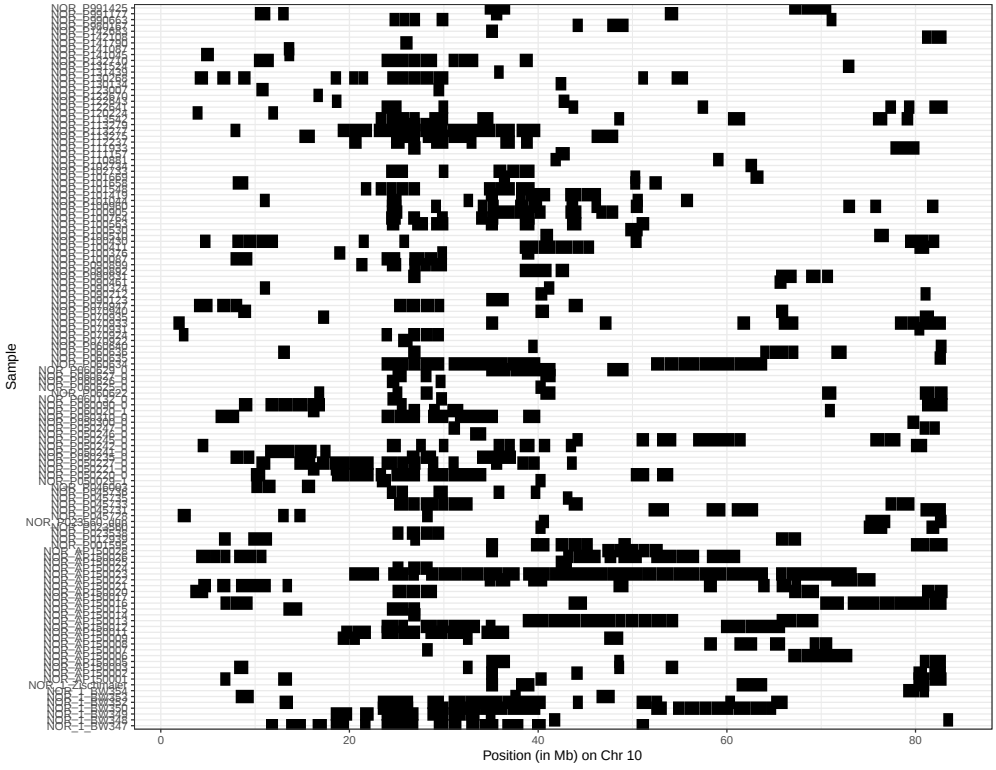


Chromosome 10

Haflinger

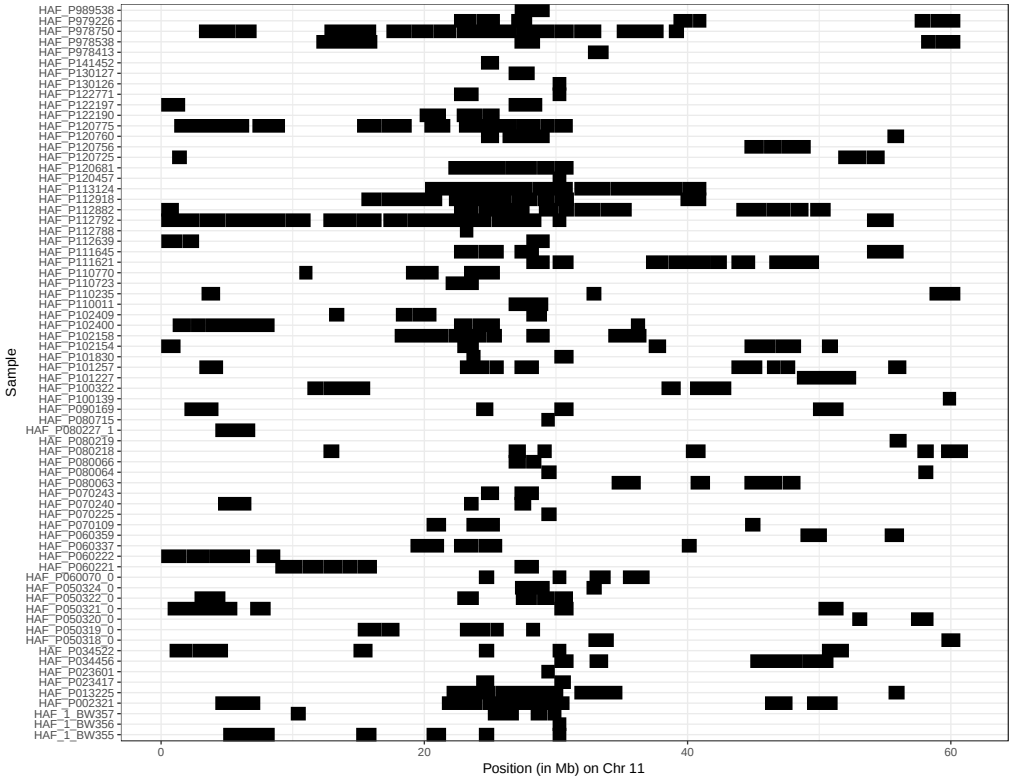


Noriker

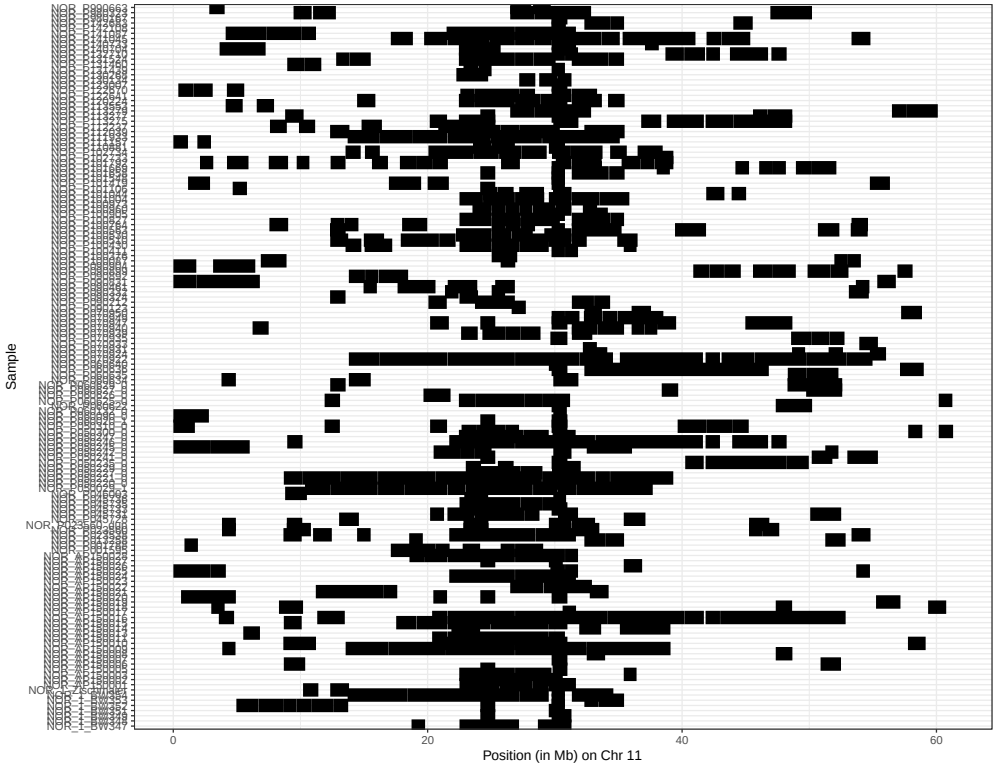


Chromosome 11

Haflinger

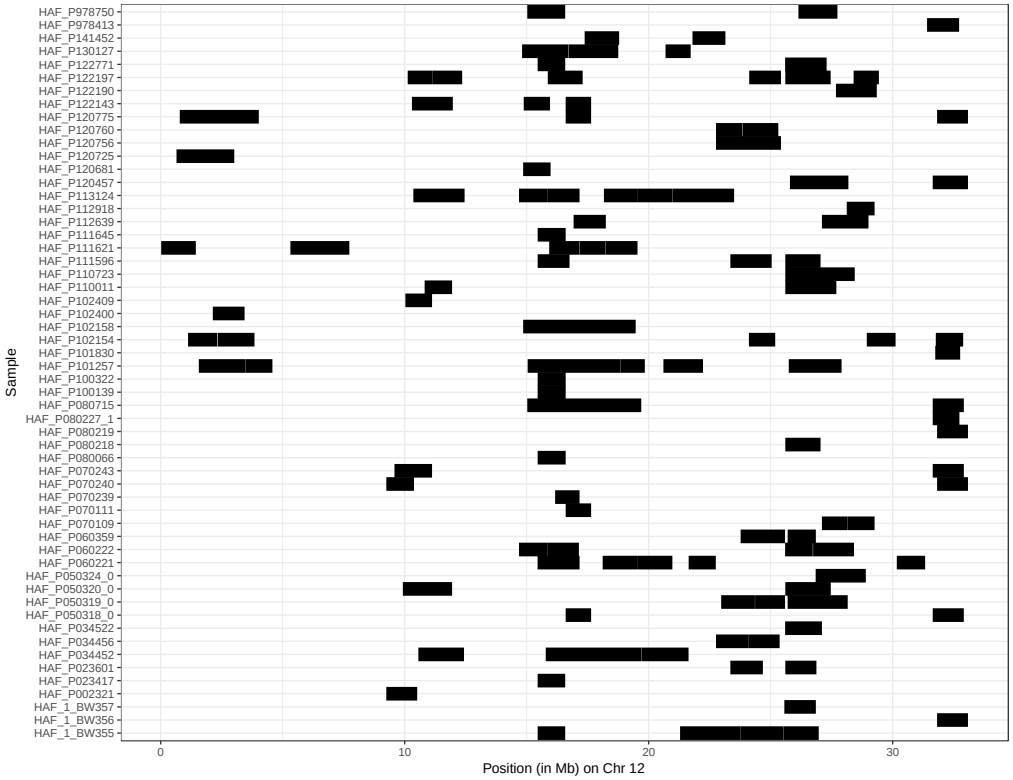


Noriker



Chromosome 12

Haflinger

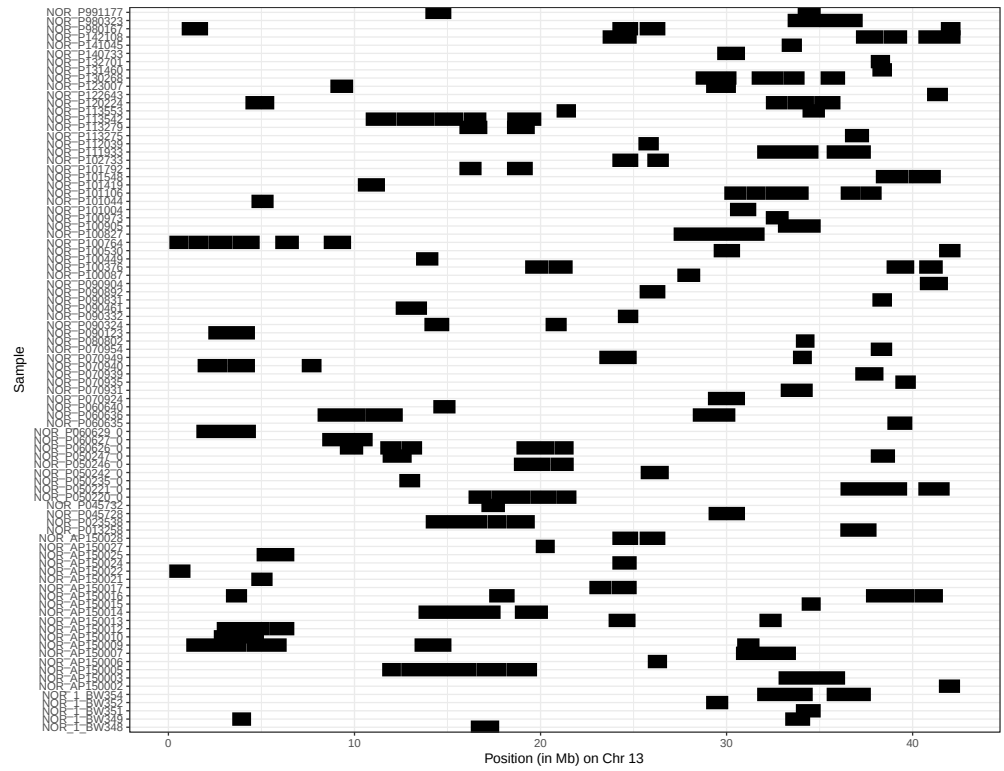


Noriker

Chromosome 13

Haflinger

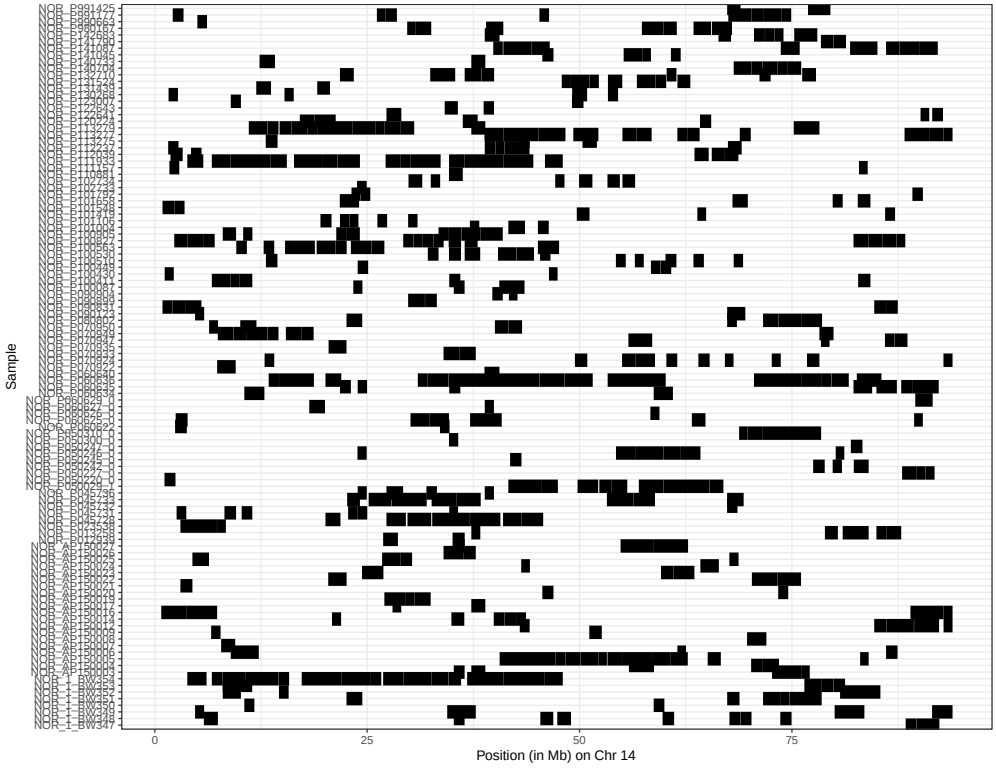
Noriker



Chromosome 14

Haflinger

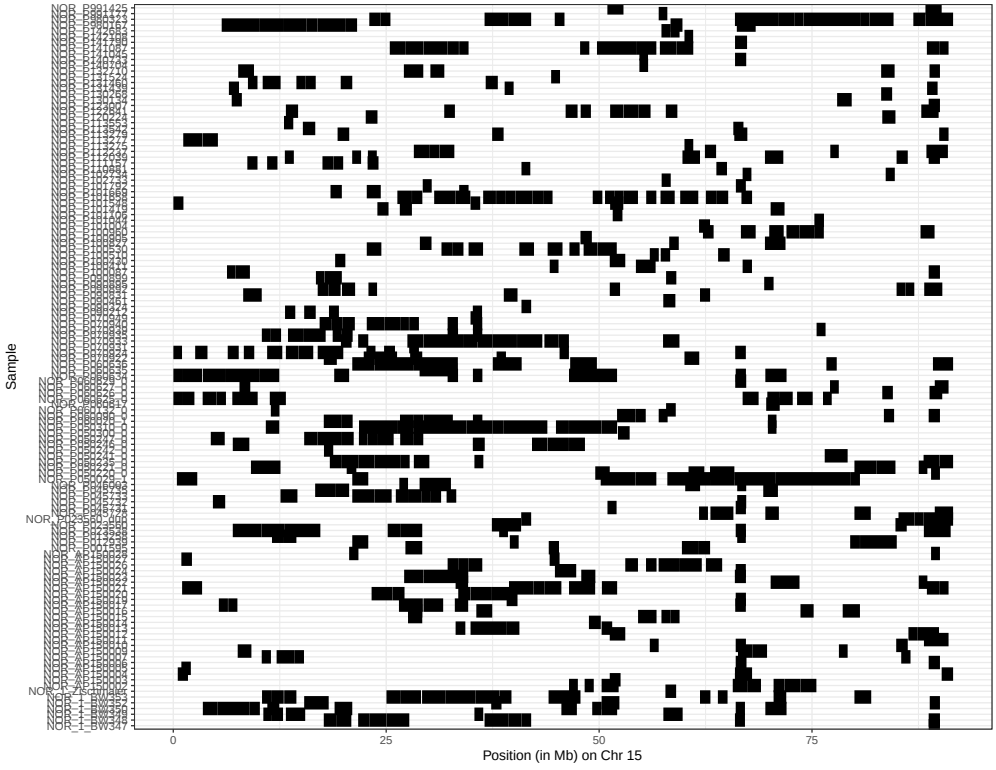
Noriker



Chromosome 15

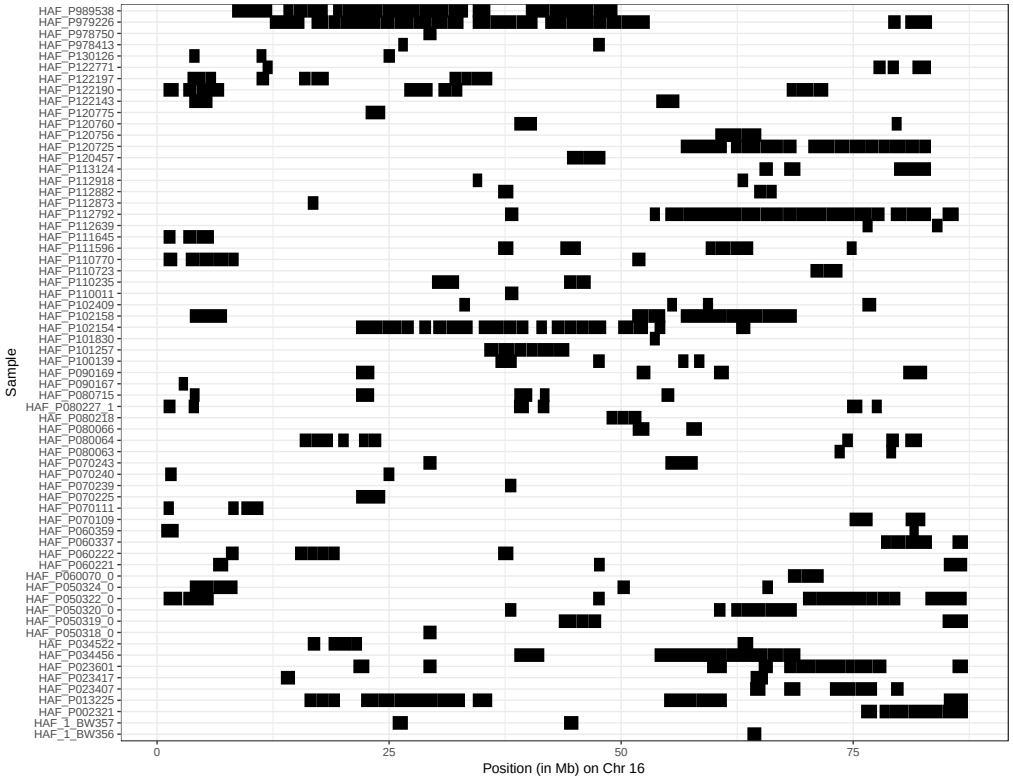
Haflinger

Noriker

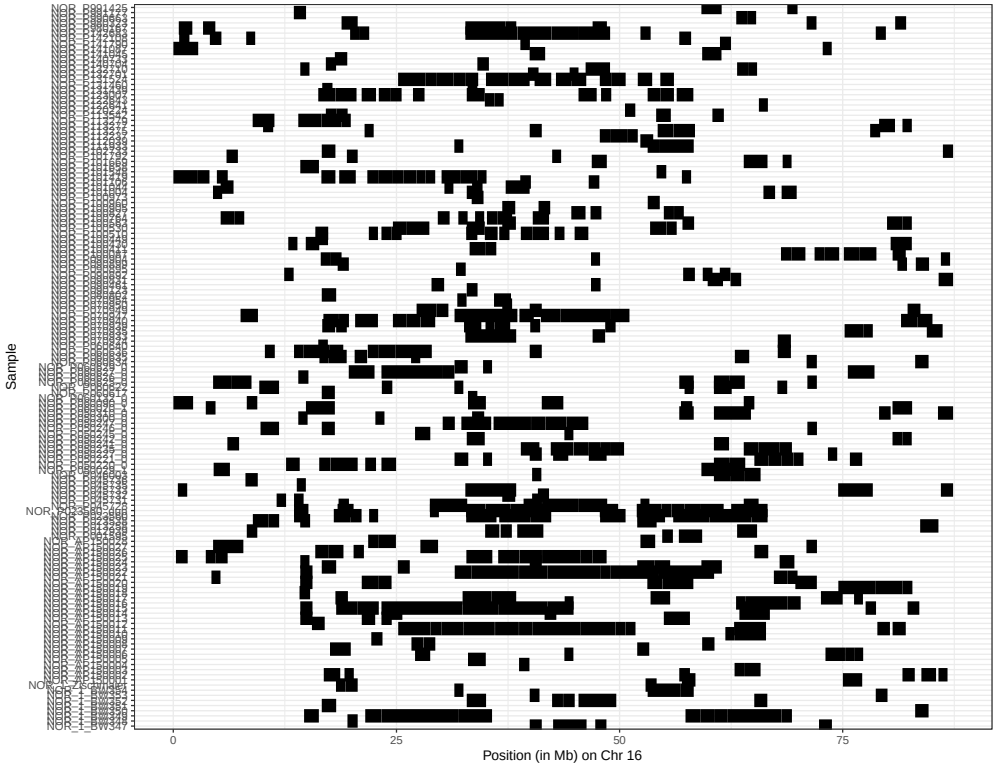


Chromosome 16

Haflinger

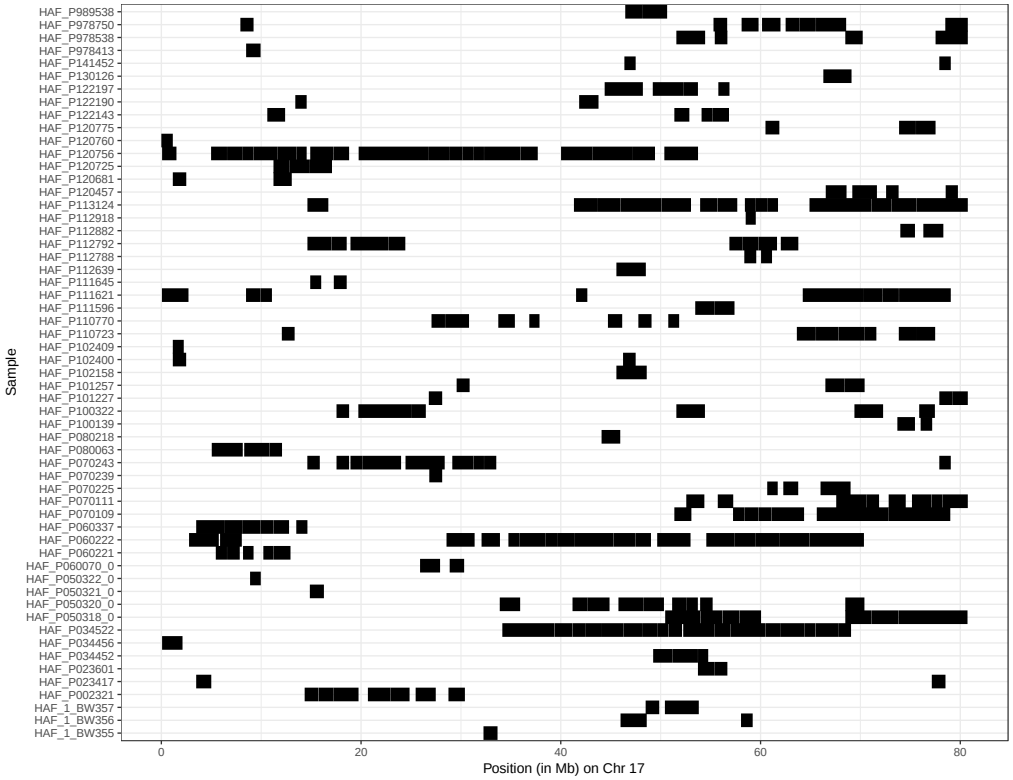


Noriker

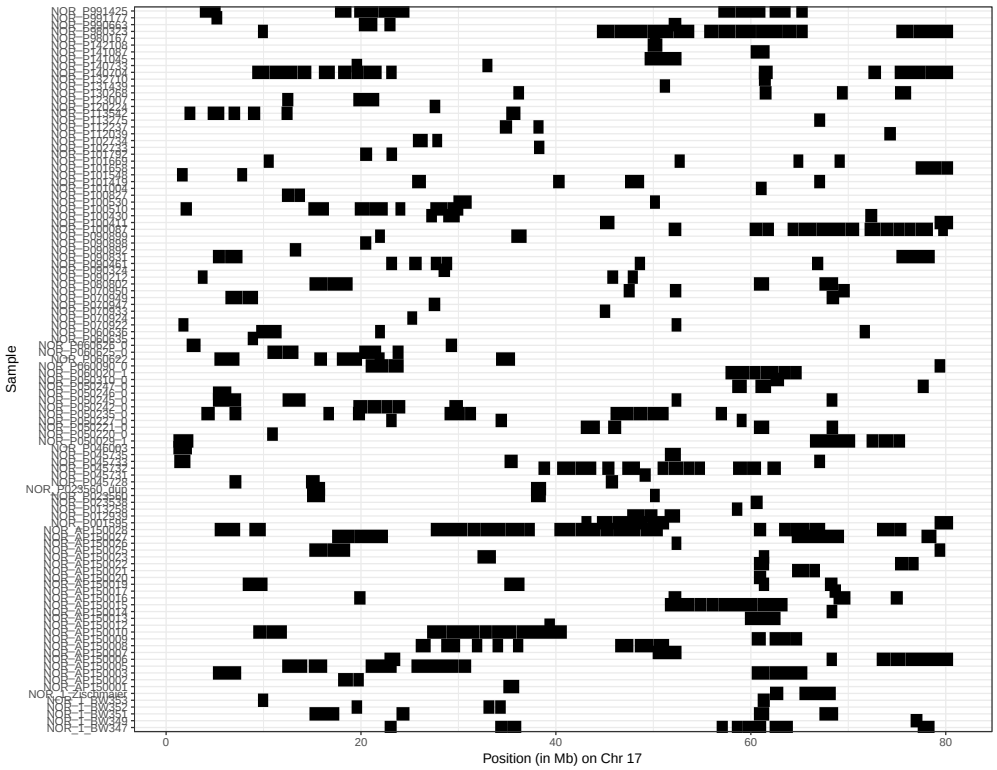


Chromosome 17

Haflinger

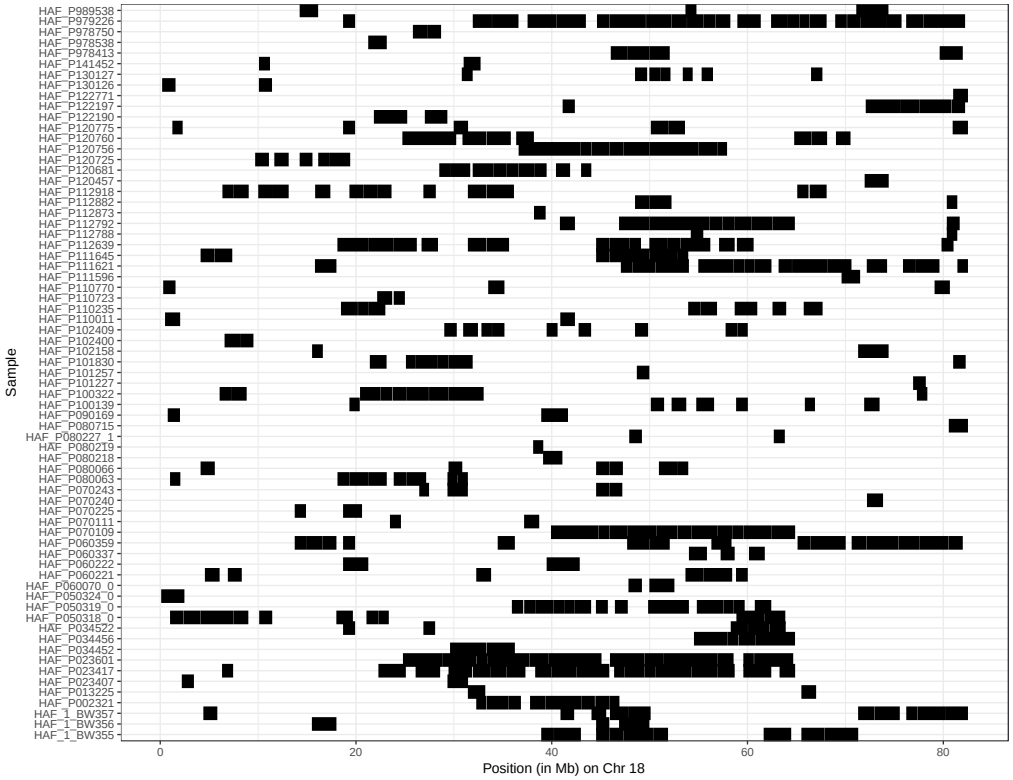


Noriker

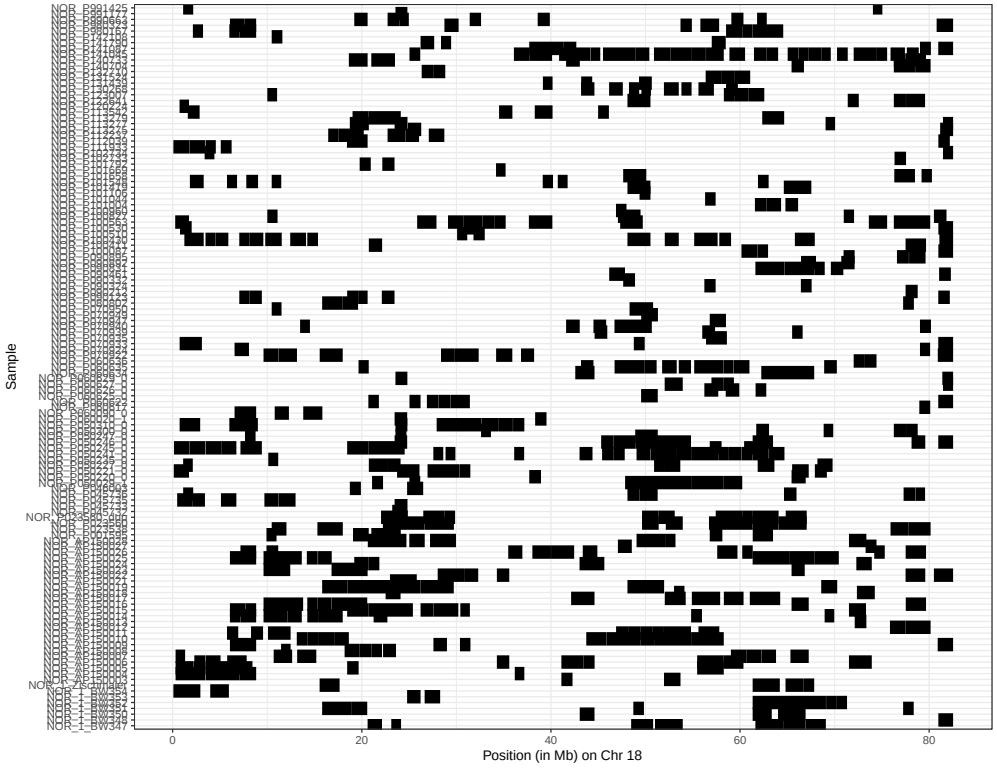


Chromosome 18

Haflinger

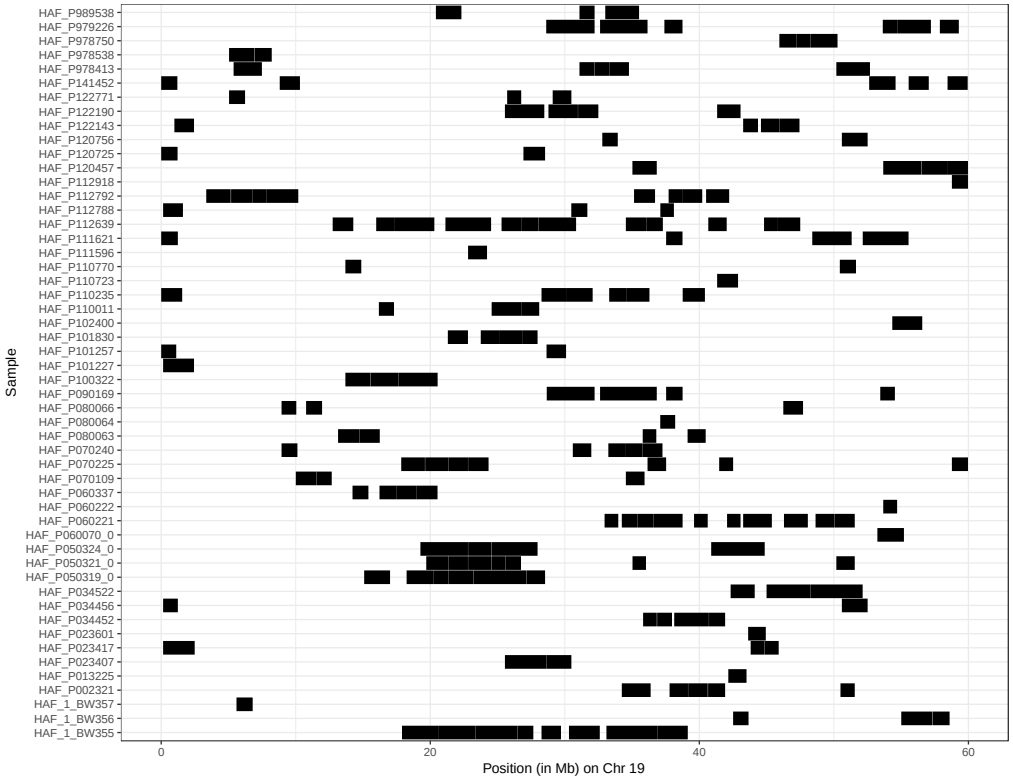


Noriker

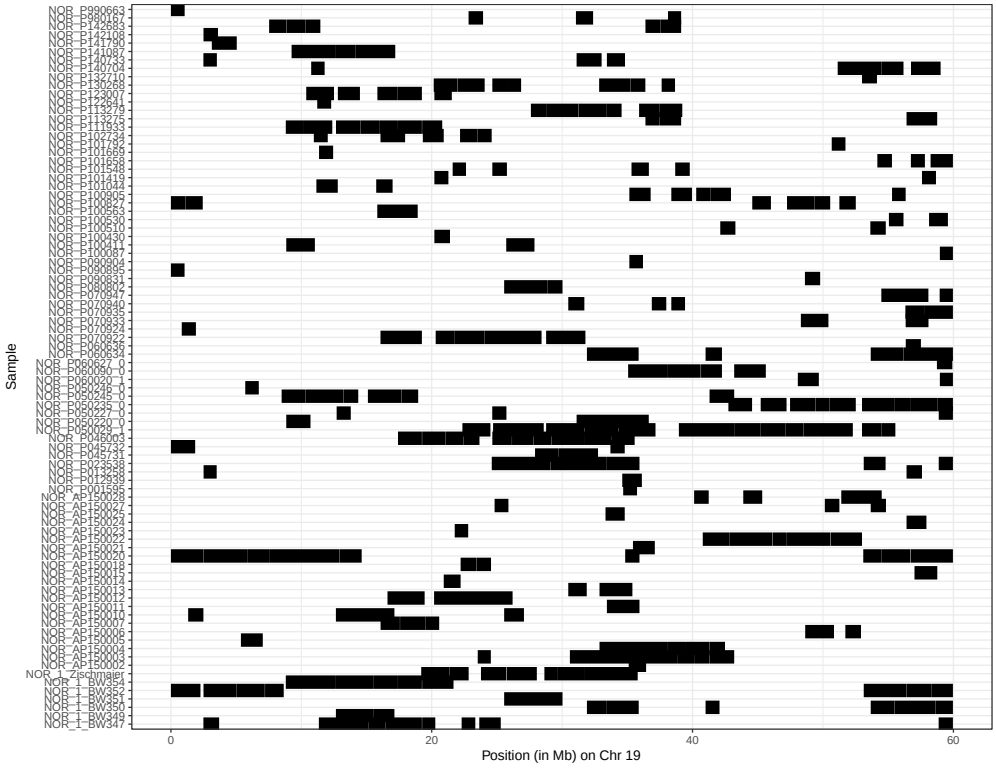


Chromosome 19

Haflinger



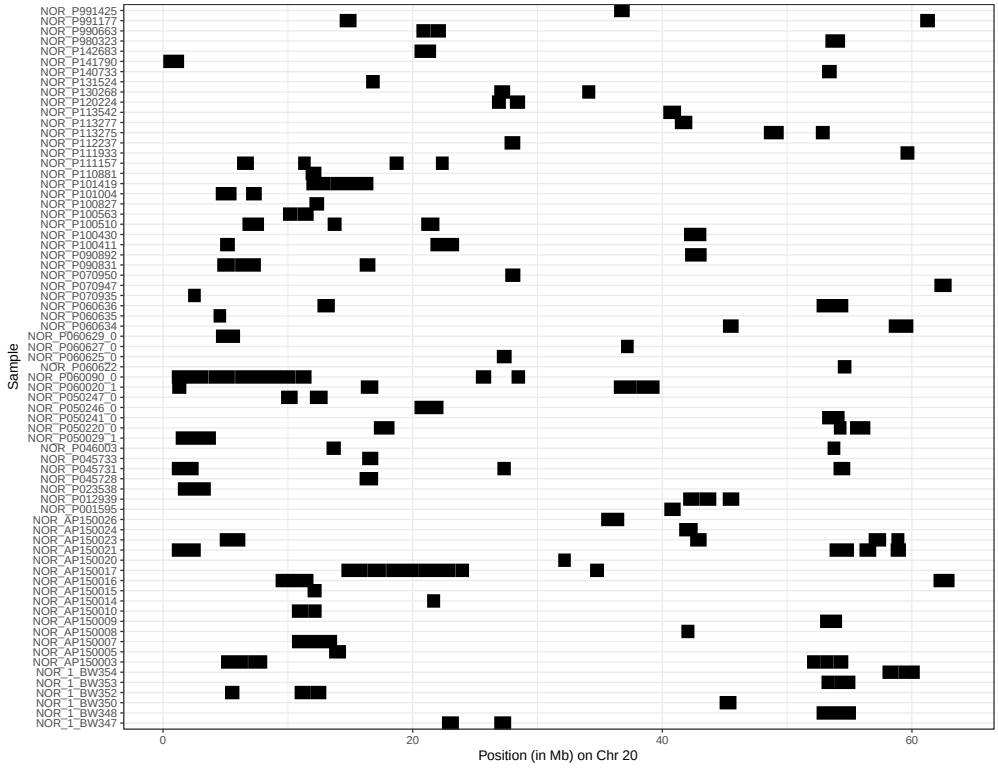
Noriker



Chromosome 20

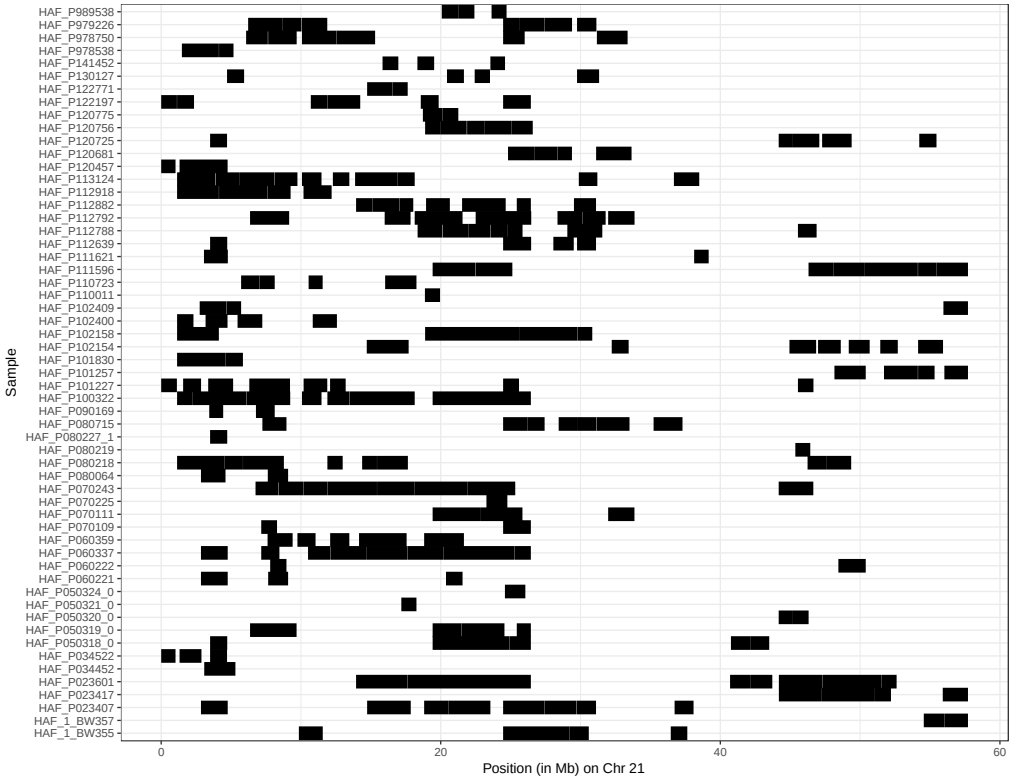
Haflinger

Noriker

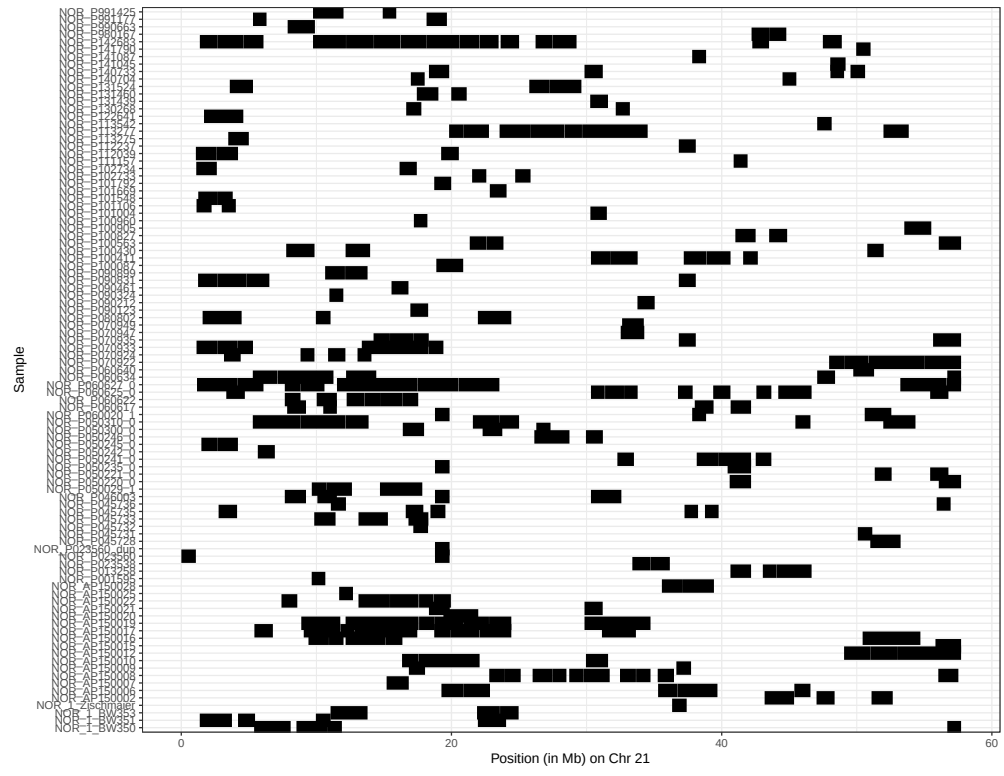


Chromosome 21

Haflinger

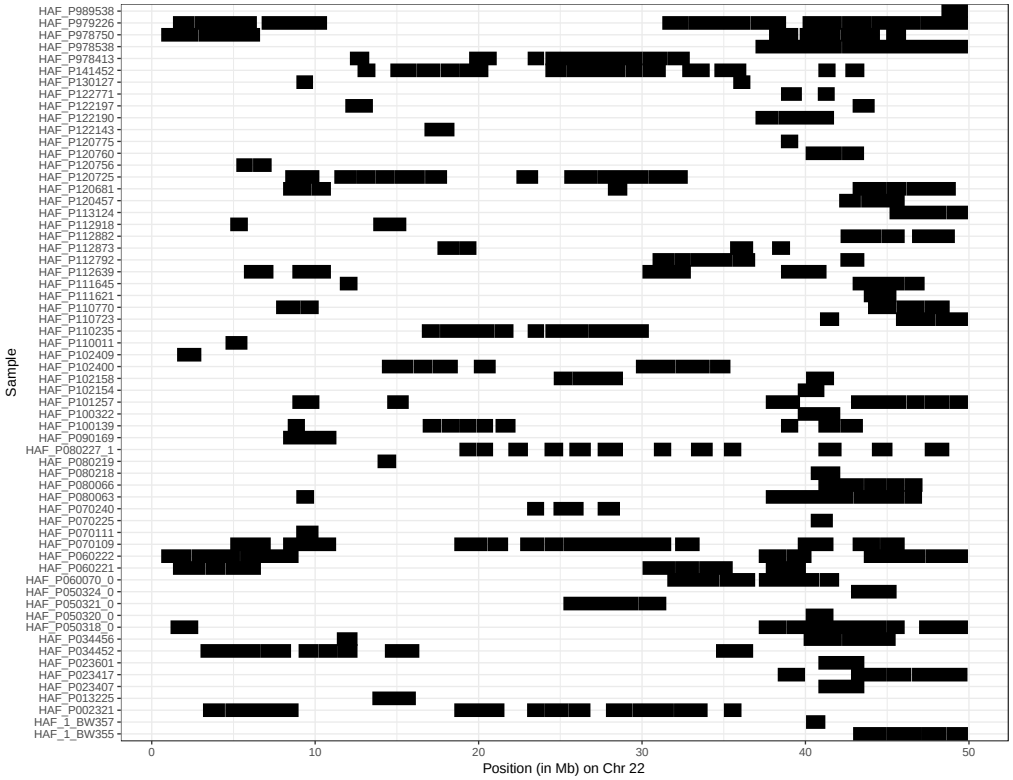


Noriker

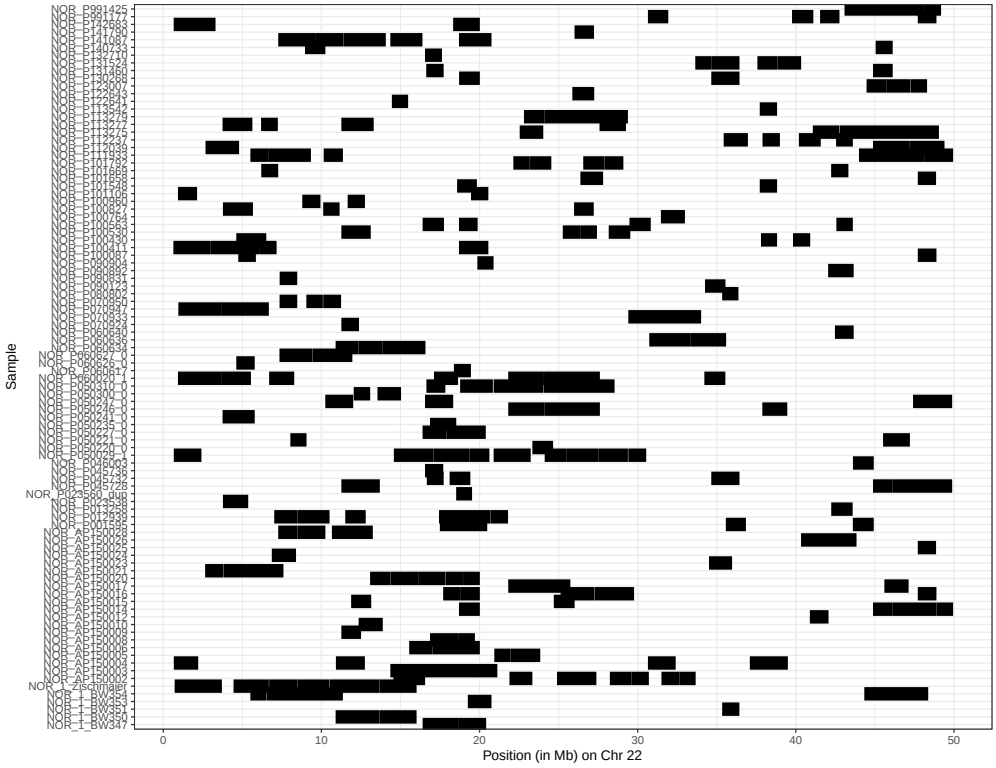


Chromosome 22

Haflinger



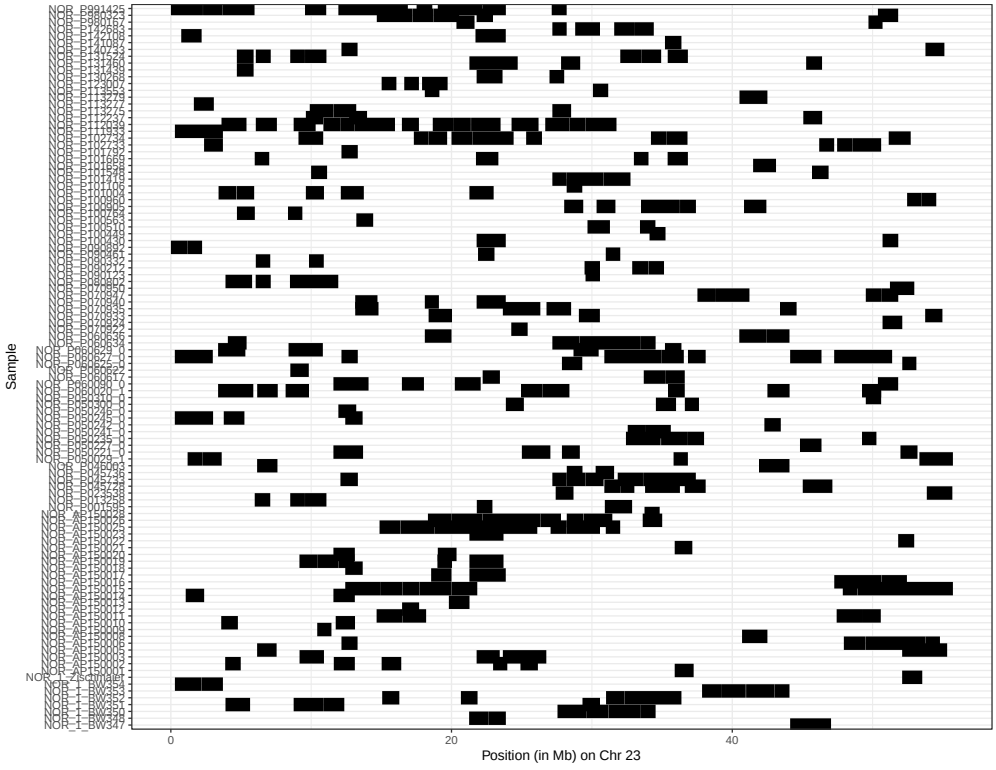
Noriker



Chromosome 23

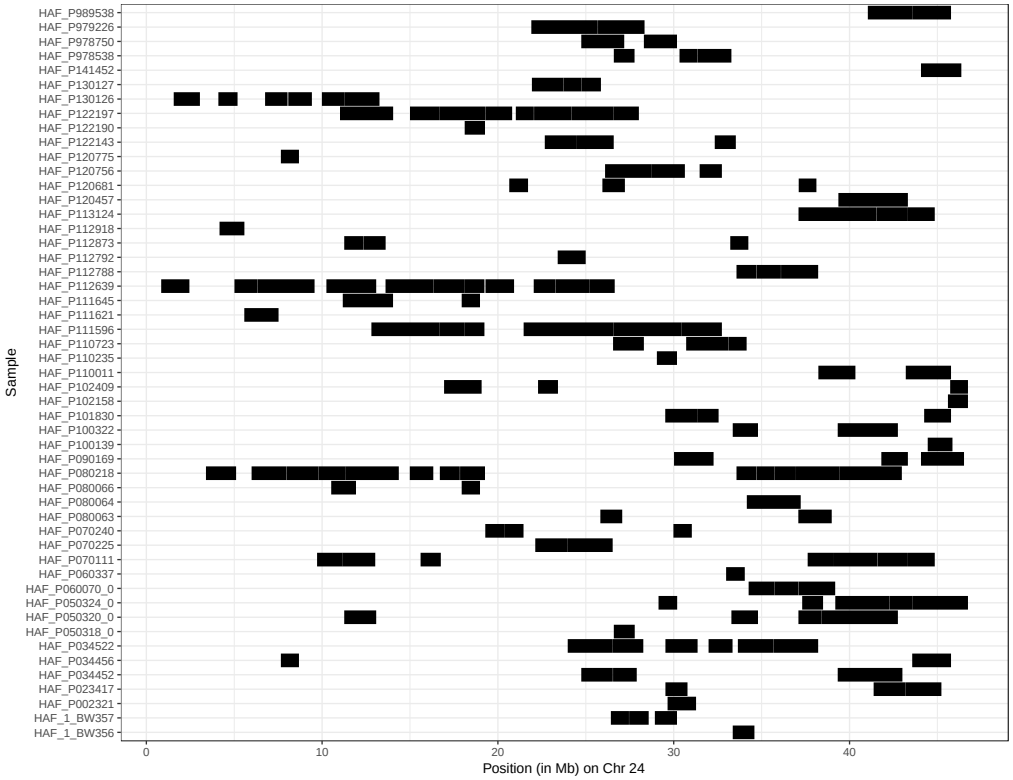
Haflinger

Noriker

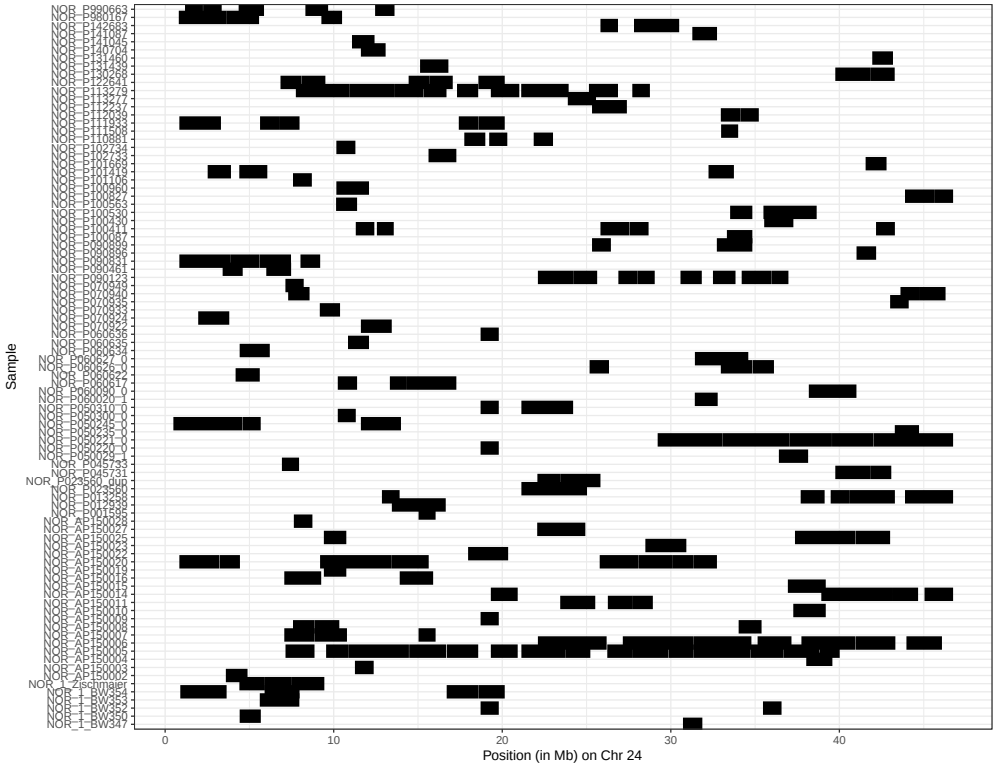


Chromosome 24

Haflinger

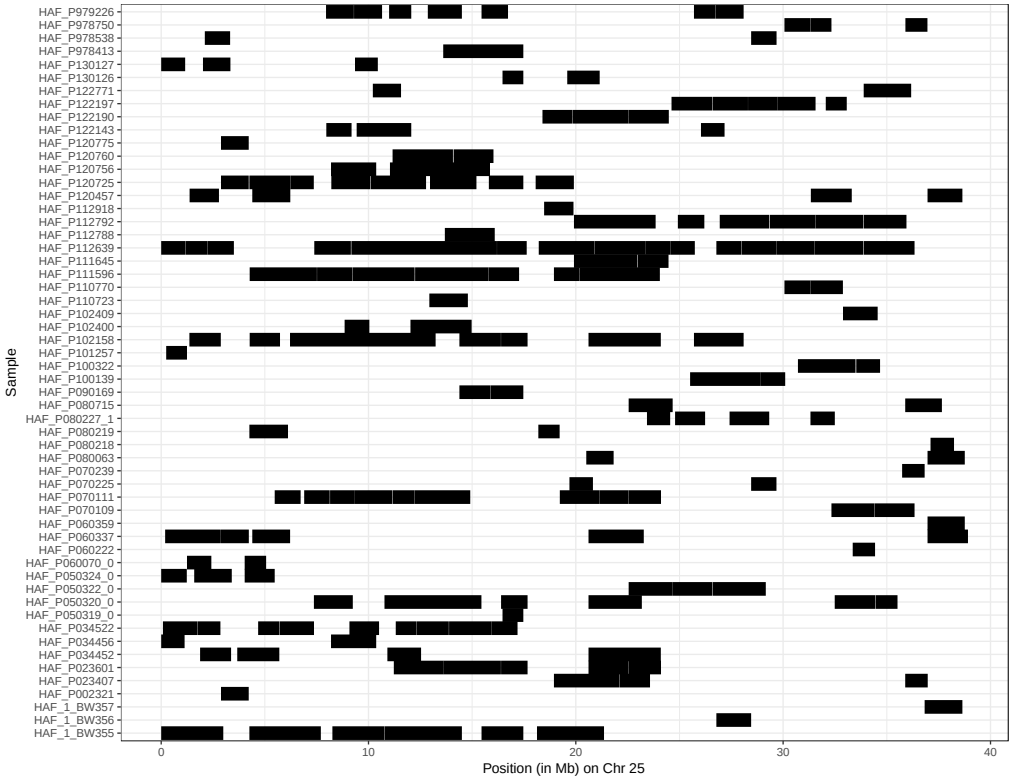


Noriker



Chromosome 25

Haflinger

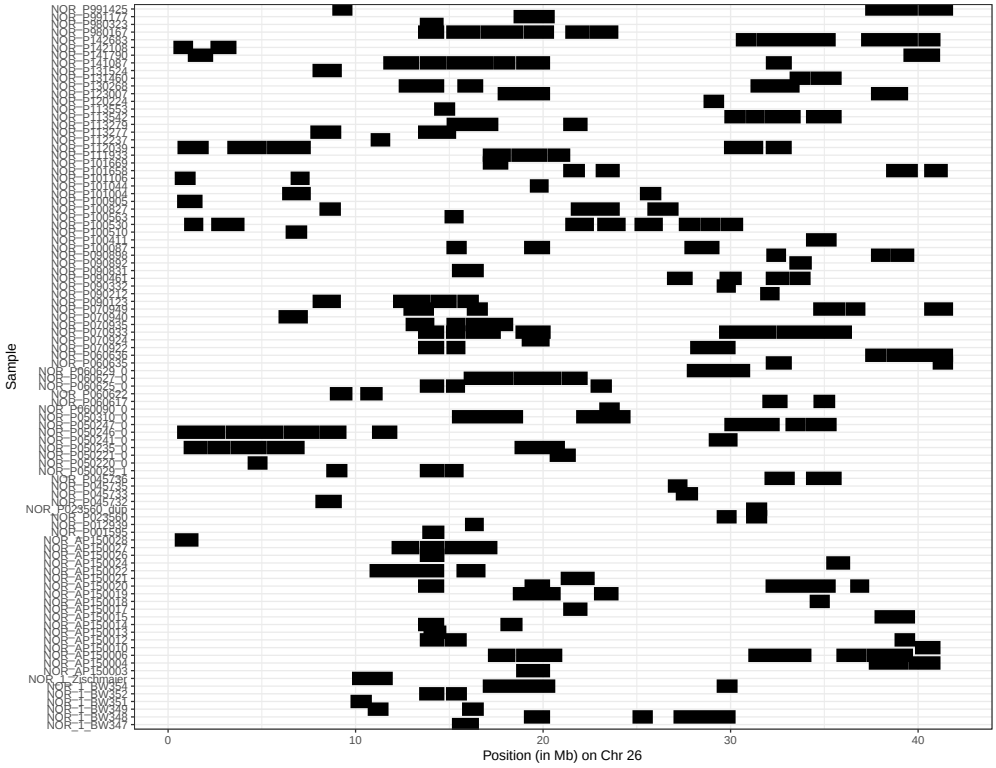


Noriker

Chromosome 26

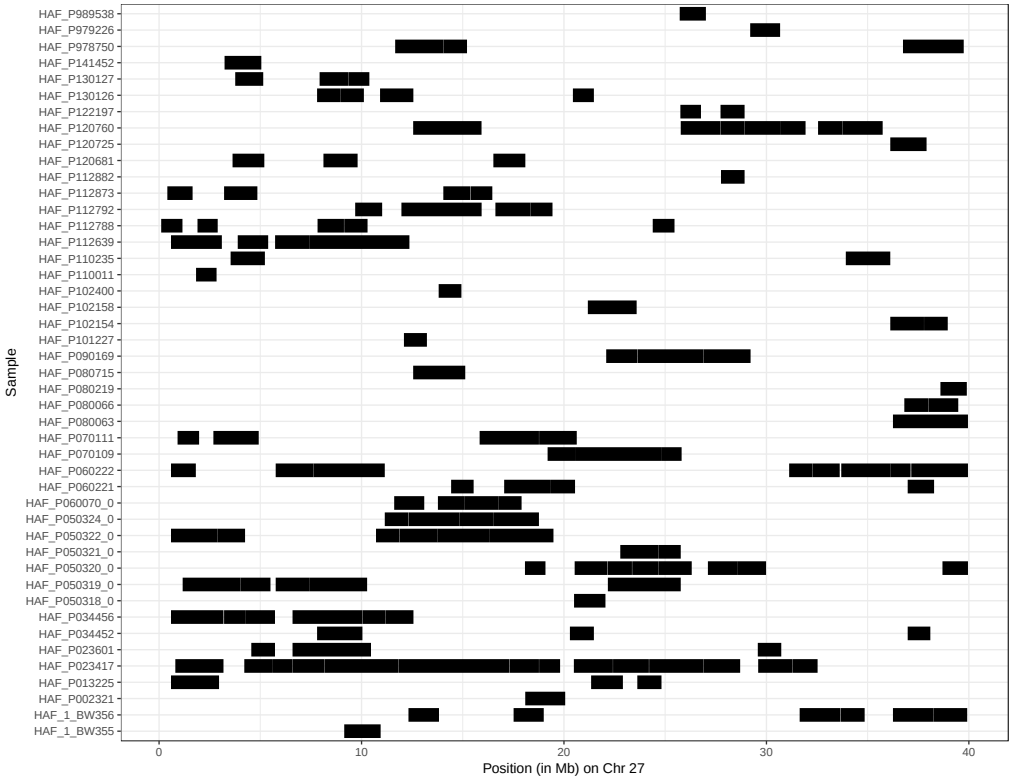
Haflinger

Noriker

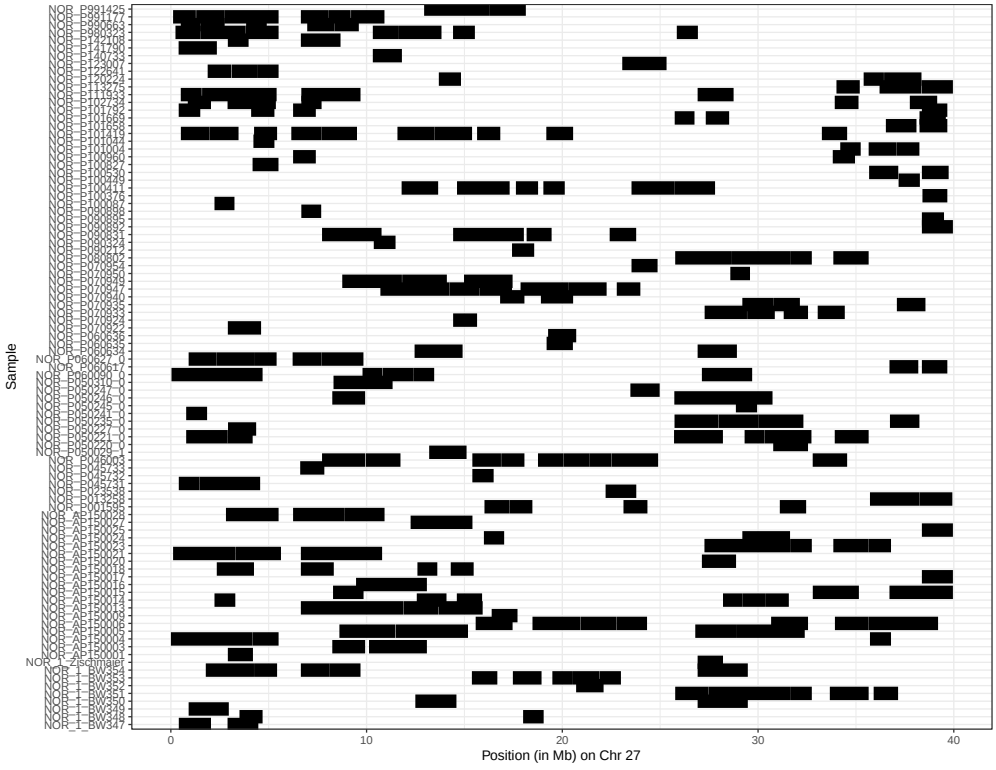


Chromosome 27

Haflinger

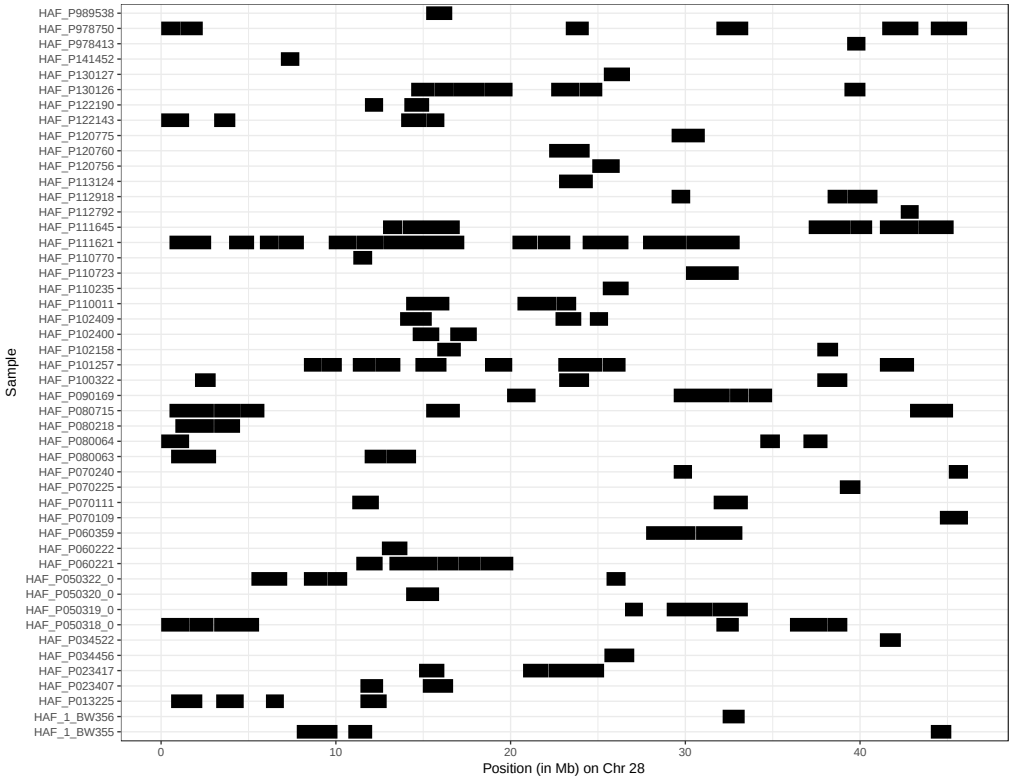


Noriker

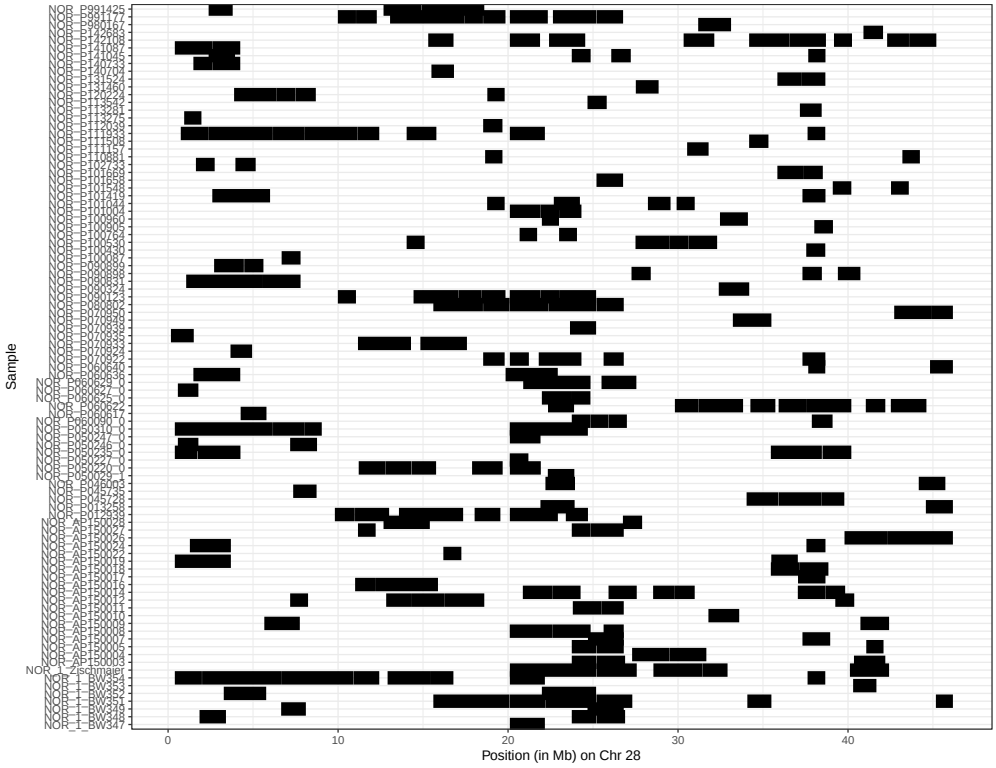


Chromosome 28

Haflinger

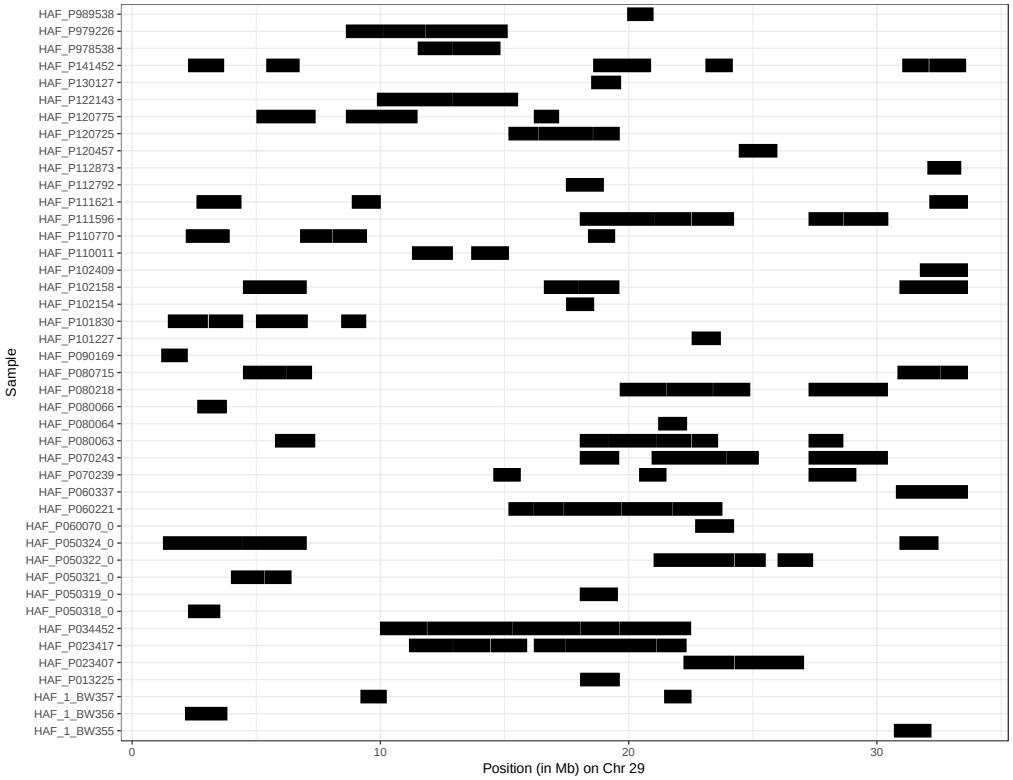


Noriker

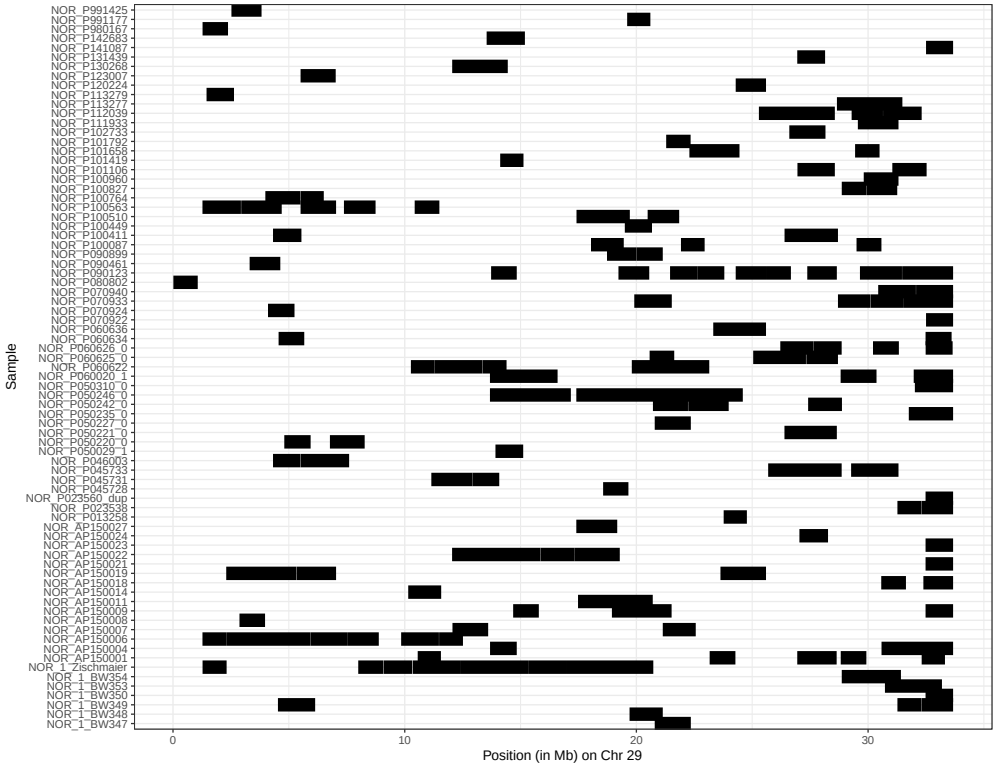


Chromosome 29

Haflinger

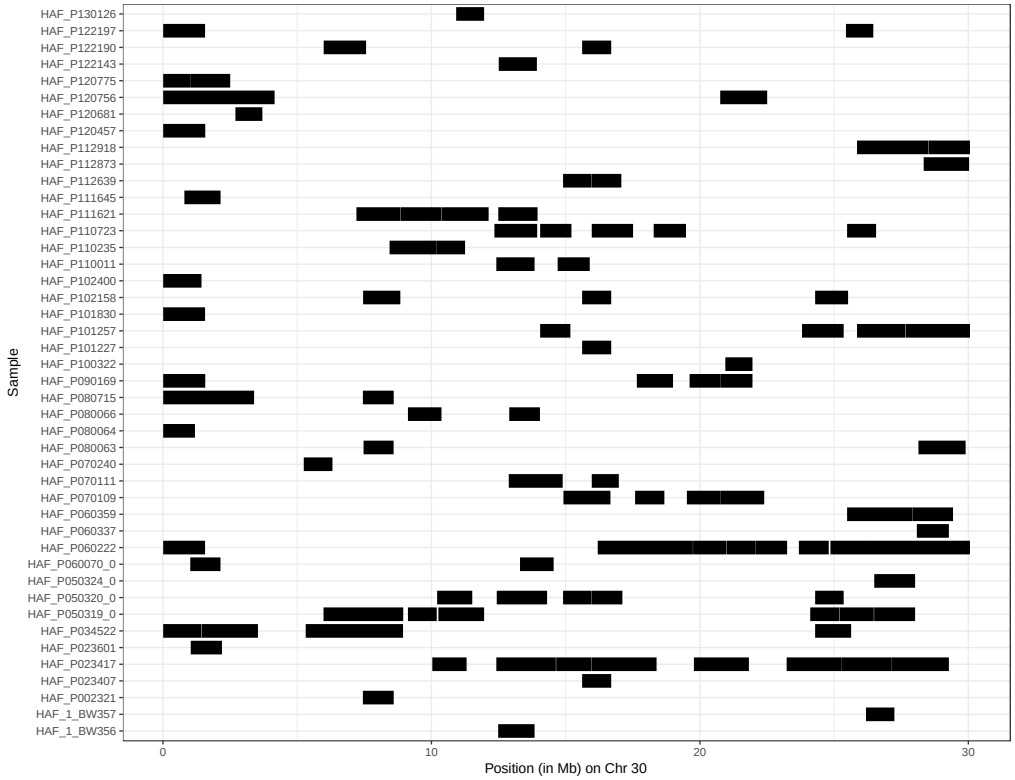


Noriker

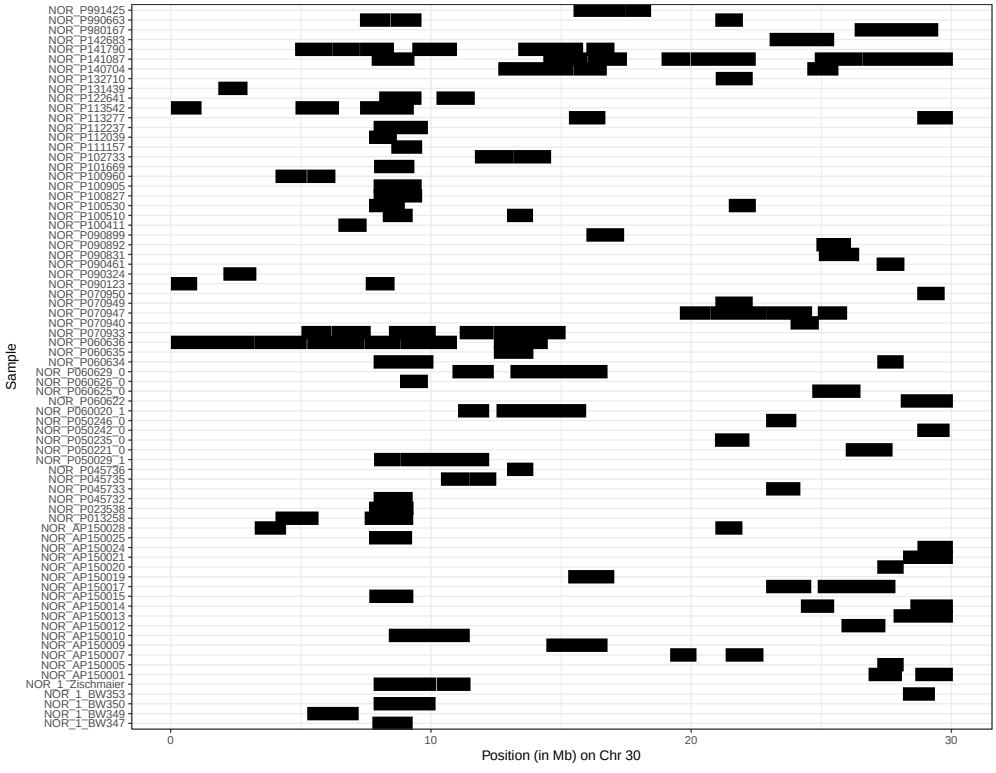


Chromosome 30

Haflinger

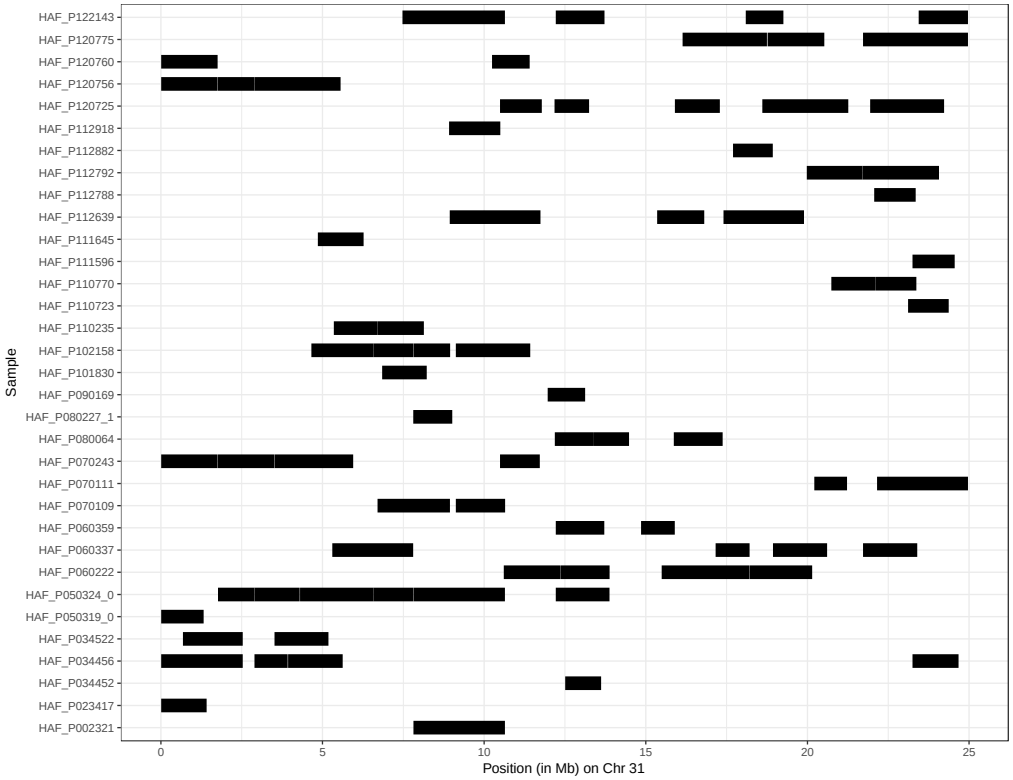


Noriker



Chromosome 31

Haflinger



Noriker