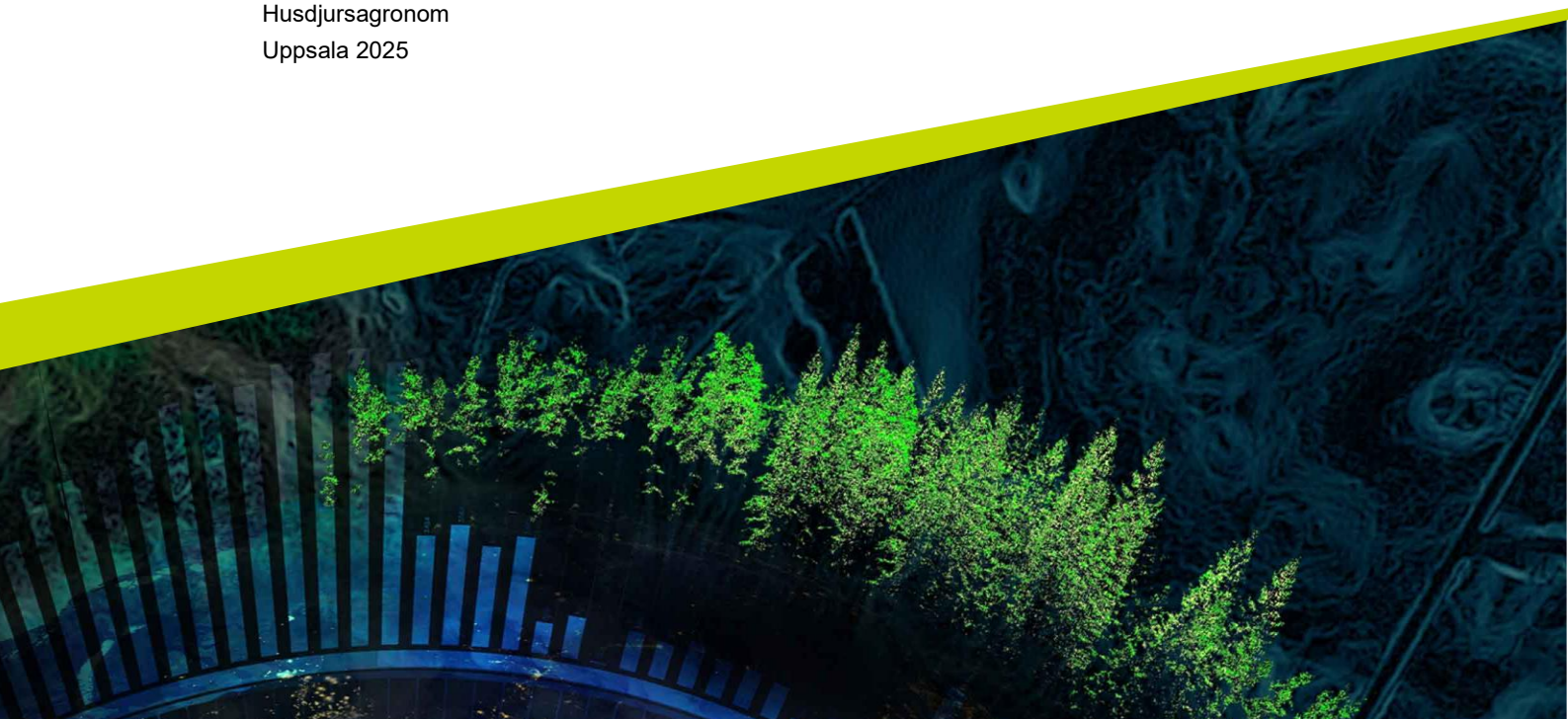
Simulating Inbreeding and Selection in Cattle using Runs of homozygosity

Detecting Hidden Islands in a Sea of Homozygosity

Kajsa Nyström

Independent project • 30 credits
Swedish University of Agricultural Sciences, SLU
Faculty of Veterinary Medicine and Animal Science/ Department of Animal Biosciences
Husdjursagronom
Uppsala 2025



Simulating Inbreeding and Selection in Cattle Using Runs of homozygosity. Detecting Hidden Islands in a Sea of Homozygosity

Simulering av inavel och urval hos kor med hjälp av homozygositetssträckor. Att upptäcka dolda öar i ett hav av homozygoti

Kajsa Nyström

Supervisor:	**Martin Johnsson, Swedish University of Agricultural Sciences, Department of Animal Biosciences**
Examiner:	Susanne Eriksson, University of Agricultural Sciences, Department of Animal Biosciences

Credits:	30 credits
Level:	A2E
Course title:	Självständigt arbete i husdjursvetenskap
Course code:	EX1023
Programme:	Animal Science
Course coordinating dept:	Department of Animal Biosciences
Year of publication:	2025
Keywords:	Runs of homozygosity, Selection, Inbreeding, Cattle, Runs of Homozygosity Islands, Effective population size.

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

Abstract

Inbreeding, genetic drift and selection are key forces that shape the genetic architecture of domestic cattle populations. This study investigates the genomic consequences of these forces by analysing runs of homozygosity (ROH) in both simulated and empirical data from Swedish native breeds (Fjäll, Rödkulla and SRB) and international breeds such as Holstein and Jersey. ROH defined as continuous stretches of homozygous genotypes, serve as genomic markers of inbreeding and selection and offer insights into both recent and ancient demographic processes.

Simulations were conducted using msprime to model historical population structures, and ROH were detected using PLINK in both simulated and empirical data. The results revealed distinct ROH patterns between native and commercial breeds. Swedish native breeds exhibited several short ROH segments, consistent with historical bottlenecks and small effective population sizes. In contrast, Holstein and Jersey cattle showed fewer but longer ROH segments, indicating more recent inbreeding due to intensive selection.

ROH islands, genomic regions consistently homozygous across individuals, were identified in both data types and further investigated for gene content. These regions might display selection for traits like disease resistance, environmental adaptability and fertility.

This demonstrates the value of ROH analysis in revealing the demographic and selective histories of cattle populations. It provides with insights for the preservation of rare breeds and the management of breeding programs.

Keywords: Runs of homozygosity, Selection, Inbreeding, Cattle, Runs of homozygosity islands, Effective population size.

Table of contents

List of tables

List of figures

Abbreviations

Abbreviation	Description
AI	Artificial Insemination
ANOVA	Analysis of variance
LD	Linkage disequilibrium
ROH	Runs of homozygosity
SNP	Single nucleotide polymorphism
MAF	Minor allele frequency
MB	Megabase
Ne	Effective population size
VCF	Variant call format

1. Introduction

The domestication of cattle, believed to have taken place approximately 10,000 years ago, initiated a complex evolutionary trajectory driven by both natural and artificial selection pressures (Purfield et al. 2012). Over time, this process has led to the emergence of diverse cattle breeds, each with distinct phenotypes shaped by local adaptation, genetic drift, and selection (Purfield et al. 2012; Liu et al. 2022). These mechanisms, acting at varying intensities across populations, have led to notable differences in genetic construction, including variation in levels of homozygosity (Hewett et al. 2023).

Swedish native cattle breeds, while genetically distinct and valuable for their adaptability and robustness, face increasing threats from the extensive use of high-productivity commercial breeds such as Holstein and Jersey (Upadhyay et al. 2019; Macharia et al. 2024). These native populations are typically small and isolated, making them particularly vulnerable to inbreeding and loss of genetic diversity (Ghoreishifar et al. 2020; Makanjuola et al. 2020). Several breeds are now considered endangered due to declining genetic diversity, with studies calling for monitoring and conservation efforts using genomic tools like ROH analysis (Harish et al., 2024; Adepoju et al., 2024; Hewett et al., 2023).

Some consequence of modern breeding practices such as the extensive use of selected sires, artificial insemination (AI), and embryo transfer offers a reduction in effective population size (Ne) and genetic diversity (Zhang et al. 2015). This is particularly noticed in highly selected commercial breeds (Zhang et al. 2015; Makanjuola et al. 2020). Selection practices, while accelerating genetic gain, have increased the prevalence of inbreeding and regions of the genome that are homozygous by descent (Purfield et al. 2012; MacLeod et al. 2013). As reviewed by Peripolli et al.(2017), runs of homozygosity (ROH) have become a valuable genomic recourse for assessing inbreeding levels and tracing historical demographic events.

1.1 Runs of homozygosity (ROH)

ROH are continuous stretches of homozygous genotypes that arise when an individual inherits identical haplotypes from both parents (Liu et al. 2022). The length and frequency of ROH segments offer valuable insights into population history where longer ROH segments typically signal recent inbreeding, whereas shorter ROH segments point to older events like genetic drift or bottlenecks (Purfield et al., 2012; Harish et al., 2024). Moreover, ROH analysis has been

shown to offer certain advantages over traditional pedigree-based approaches. Breeding organizations for more established breeds often possess well-documented pedigrees, which are often used to verify the lineage of the animals. However, breeding organisations for less established breeds are more constrained by incomplete or inaccurate data (Sumreddee et al. 2020).

ROH regions are more frequent and extended in populations that have experienced recent bottlenecks or are subject to high levels of inbreeding (Purfield et al. 2012; Harish et al. 2024). The distribution and size of ROH segments across genomes can reveal both recent and ancient demographic events (Gibson et al. 2006; Macharia et al. 2024). ROH is not only a marker of inbreeding but also a genomic signature of selection (Ceballos et al. 2018). When strong selection acts on a beneficial allele, the surrounding genomic region may become more homozygote, forming what is known as a run of homozygosity island (Ghoreishifar et al. 2020). Identifying ROH islands can therefore, reveal signatures of selection shaping different breeds, especially when contrasted against regions of the genome that are homozygous due to natural processes like genetic drift (Ghoreishifar et al. 2020; Makanjuola et al. 2020; Macharia et al. 2024).

1.1.2 ROH islands

Detecting ROH islands requires setting thresholds to define which genomic regions are considered shared ROH across individuals. However, there is currently no universally accepted standard for setting such thresholds. Whether based on a fixed proportion of individuals or a statistical percentile, all thresholding approaches introduce a degree of uncertainty (Ceballos et al. 2018; Nandolo et al. 2018; Meyermans et al. 2020). If the threshold is set too high, true signals of selection might be missed; if set too low, false positives could arise due to structural variation or SNP density gaps (Nandolo et al. 2018).

According to the findings in a study made by Meyermans et al. (2020) it is recommended to report genome-wide ROH coverage and conduct sensitivity analyses to assess how parameter settings affect ROH detection. Key threshold decisions include the minimum ROH length and the frequency threshold for defining a shared ROH region or "island". Each choice involves trade-offs: stringent thresholds reduce false positives but risk overlooking subtle signals, while relaxed thresholds increase detection rates at the cost of higher false discovery rates (Ceballos et al., 2018; Meyermans et al., 2020). Given this sensitivity, it is essential to transparently report all parameter choices and to interpret findings within the context of these methodological assumptions.

Ultimately, validating threshold strategies through simulations or comparisons between methods helps ensure that identified ROH islands reflect genuine biological signals of inbreeding or selection.

1.2 Aim

This study simulates genomic data from Swedish and international cattle breeds to identify runs of homozygosity (ROH) and homozygosity islands. By comparing simulated and empirical data, it aims to differentiate genomic patterns resulting from inbreeding and selection, and to examine how population size and historical structure influence ROH distribution over time.

2. Literature review

2.1 Native breeds and small populations

Native Swedish cattle breeds are generally well-adapted to harsh climatic conditions while still maintaining key productive traits (Upadhyay et al. 2019). These adaptations have been influenced by both by natural selection and deliberate artificial selection methods designed to enhance traits like fertility, growth and milk yield (Harish et al. 2024). Ghoreishifar et al. (2020) identified genomic regions under selection in Swedish cattle that are linked to economically important traits and environmental adaptation, emphasizing the influence of selection. Systematic breeding programs, while improving performance, have also contributed to a reduction in Ne. This genetic narrowing has been compounded by demographic declines, as reduced popularity of native breeds has led to population bottlenecks (Ghoreishifar et al. 2020). Together, both have accelerated the loss of genetic diversity. It is therefore important to recognize that both structured breeding and demographic pressures have played roles in shaping the current genetic landscape of these breeds (Adepoju et al. 2024). Even if population sizes are stabilized at current levels, the remaining genetic diversity may be insufficient to sustain healthy populations over time (Adepoju et al. 2024).

Small, native cattle populations often experience genetic drift and inbreeding, resulting in a lower effective population size (Ne) and a decline in genetic diversity. (Upadhyay et al. 2019; Ghoreishifar et al. 2020). The Swedish Red Polled (SRB) and Fjällnära breeds both show decreasing Ne in the last decade, correlating with extensive ROH (Upadhyay et al. 2019). Adepoju et al. (2024) investigated the population history of Swedish cattle breeds by employing linkage disequilibrium (LD)-based methods to estimate changes in Ne. The study by Adepoju et al. (2024) revealed a rapid and recent decline in Ne across various breeds, a trend that coincides with the implementation of systematic breeding practices. This decline in genetic diversity is primarily attributed to genetic drift and population bottlenecks, phenomena commonly observed in small populations (Adepoju et al., 2024). Genetic drift, defined as random changes in allele frequencies over generations, can result in the loss of genetic variants critical for maintaining genetic diversity (Peripolli et al. 2017; Upadhyay et al. 2019).

Maintaining both overall genetic diversity and key adaptive variants is essential to safeguard population health under climate change and future breeding schemes (Peripolli et al. 2017; Harish et al. 2024).

2.2 International breeds and populations with strong selection

Genomic selection has significantly influenced dairy cattle breeds, impacting highly heritable traits such as production traits, as well as traits with lower heritability, such as adaptability to the environment (Gutierrez-Reinoso et al., 2021). The sustainability of animal production and welfare largely depends on sustaining genetic diversity within livestock populations, which is essential for adaptation to environmental changes and long-term productivity (Makanjuola et al., 2020).

Commercial breeds like Holstein and Jersey have been subjected to modern breeding programs emphasizing economic traits such as production and health. This intense selection has led to shorter generation intervals and increased inbreeding, contributing to reduced genetic diversity (Makanjuola et al., 2020). Compared to less intensively selected breeds, such as the Rödkulla and Fjällnära cattle, Holsteins exhibit a notably higher frequency of long runs of homozygosity (ROH), suggesting recent occurrences of inbreeding (MacLeod et al. 2013; Upadhyay et al. 2019; Makanjuola et al. 2020; Harish et al. 2024)

Jersey cattle originated from a small island population, which consequently constrained their genetic diversity, thereby increasing their predisposition to inbreeding (Makanjuola et al., 2020). The implementation of genomic selection has further decreased their already limited population size, intensifying levels of inbreeding and homozygosity (Makanjuola et al., 2020). This highlights the importance of balancing genomic selection for productivity with strategies to preserve genetic diversity and minimize the negative consequences of inbreeding (Adepoju et al., 2024).

2.3 Detection and Genome-wide distribution ROH

ROH methods have been extensively utilized in both human- and animal-focused research and are therefore recognized as a robust approach for evaluating population history (Purfield et al. 2012; Peripolli et al. 2017; Hewett et al. 2023). Populations exhibiting high homozygosity are predisposed to recessive diseases and are notably susceptible to environmental changes (Purfield et al. 2012). The size, prevalence, and genomic distribution of ROH fluctuate significantly due to various correlated factors, including recombination, selection and ancestry (Hewett et al. 2023). Longer ROH segments are typically indicative of recent inbreeding, while shorter ROH indicates a more distant inbreeding (Purfield et al. 2012). Recombination, which gradually fragments long chromosome segments, is

shaped by the age of the most recent mutual ancestor shared by the parents (Purfield et al. 2012). Consequently, homozygous segments are typically longer in inbred populations compared to those with larger effective population sizes (Gibson et al. 2006).

ROH hotspots, defined as regions with a high concentration of ROH, are population dependent. Their occurrence is often attributed to recent inbreeding bottlenecks or selective breeding practices (Hewett et al. 2023). Hewett et al. (2023) preformed a combination of empirical analyses and evolutionary simulations to investigate the factors shaping ROH hotspots and ROH islands. The study included genomic analysis of red deer populations, using high density SNP to estimate ROH distribution. They also compared population studies to examine ROH patterns across populations and evaluate the effect of demographic history. Lastly, genetic simulations modelling selection strength, recombination rates and population histories were conducted to determined ROH formation (Hewett et al. 2023). The results identified selection strength as a key factor influencing the development of ROH hotspots; an elevated selection coefficient corresponded to increased ROH density (Hewett et al. 2023). Hewett et al. (2023) emphasize that both a large Ne and strong selection pressures are necessary for the formation of ROH hotspots. Consequently, in populations with a history of bottlenecks, genetic drift can have a more substantial impact on ROH patterns than selection (Ceballos et al., 2018).

ROH islands, which in contrast to hotspots, can be found across populations (Hewett et al. 2023). ROH islands are therefore, consistently present in multiple distinct populations rather than being restricted to a single group or breed. These islands contain genes that are often linked to fitness or production traits (Liu et al., 2022). ROH clustering can also be a result of technical artifacts or neutral processes such as low recombination or shared ancestry (Nandolo et al., 2018). Recombination gradually fragments homozygous regions over generations, unless constrained by recombination cold-spots or regions of strong linkage disequilibrium (LD), which may preserve ROH signatures (Hewett et al., 2023; Gibson et al., 2006).

In cattle ROH islands have been associated with traits like body size, fertility, and disease resistance (Zhang et al., 2015; Liu et al., 2022). ROH islands shared across breeds may also reflect similarities in recombination landscapes, as LD structure is often conserved across closely related populations (Gibson et al., 2006).

2.4 Effective population size

The population history has a more substantial influence on the distribution of ROH than recombination or selection. Although selection can produce genomic regions with high levels of ROH, such outcomes are dependent on specific conditions (Hewett et al. 2023). These conditions include a large Ne, strong selective pressure or population bottlenecks (Ceballos et al. 2018). In populations experiencing bottlenecks, the effects of genetic drift often surpass those of selection in which case the ROH patterns are more likely to reflect historical demographic events, rather than selection processes (Hewett et al. 2023). In populations with strong selective pressure some regions with increased homozygosity, usually with beneficial alleles, tend to become fixed which contributes to ROH formation (Mastrangelo et al. 2018; Hewett et al. 2023).

Breeds with small Ne, such as Swedish native breeds or commercial breeds subjected to modern breeding programs, are particularly vulnerable to reduced genetic diversity. This reduction is largely driven by inbreeding and selective pressures targeting specific productive traits (Upadhyay et al., 2019). Swedish native cattle have been selectively bred not only for productive traits but also to maintain production in challenging climates, making them resilient in harsh environmental conditions (Harish et al., 2024). However, their declining population size requires careful monitoring to lessen the risk of inbreeding and to preserve genetic diversity (Ghoreishifar et al. 2020; Harish et al. 2024).

The study by Hewett et al. (2023) identified selection as the primary driver of ROH hotspots. The same study also suggested that recombination rates may play a facilitative role in their emergence. Furthermore, the study highlighted that the primary mechanisms underpinning ROH formation are genetic drift and genome-wide recombination. Evidently demonstrating the influence of genetic drift on the formation of ROH hotpots (Hewett et al., 2023).

2.5 Detecting ROH: methodological consideration

Runs of homozygosity are detected using a range of computational methods that vary in their sensitivity and specificity. The choice of tool and its parameter settings can strongly influence the detection of ROH, making methodological transparency essential (Howrigan et al. 2011). Broadly, detection approaches can be grouped into two groups; sliding window methods, which scan the genome in fixed-size segments (e.g. PLINK) or likelihood-based methods (e.g. BCFtools), which statistically model genotype data (Howrigan et al. 2011; Narasimhan et al. 2016).

Each method carries assumptions and limitations. Window-based approaches may misidentify ROH if SNP density is low, while model-based approaches are computationally intensive and sensitive to parameter misspecification (Meyermans et al., 2020; Nandolo et al., 2018). Factors such as copy number variation (CNV), gaps in SNP coverage, and genotyping errors can all lead to false positives or negatives (Ceballos et al., 2018; Nandolo et al., 2018). These challenges necessitate careful quality control, including minor allele frequency (MAF) filtering, LD trimming, and setting thresholds for missing data and heterozygosity (Nandolo et al. 2018; Meyermans et al. 2020).

2.6 Genes

ROH islands and ROH hotspots often harbour genes linked to productions traits, disease resistance and environmental adaptation (Dixit et al. 2020). This highlights the role of ROH islands and the formation of breed specific characteristics (Peripolli et al. 2018). Homozygous regions often emerge in response to selective pressures, which further results in ROH islands or ROH hotspots (Zhang et al. 2015). ROH islands across populations suggests that they contain genes with an important role in selection or evolutionary processes (Peripolli et al. 2017, 2018; Dixit et al. 2020).

Previous studies have found several candidate genes where ROH islands are present. Dixit et al. (2020) conducted a genome-wide analysis of ROH in several native Indian cattle breeds, to investigate selection signatures and genetic diversity. The study found ROH islands in genomic regions with genes correlating to production traits. Some of the genes found were *PTGFR, CSN1S1* and *CSNS2* which are related to milk composition and milk yield (Dixit et al. 2020). Peripolli et al (2017) conducted a genome-wide analysis of ROH in dairy cattle to assess genomic inbreeding and identify ROH islands associated with selection. The study found 14 genomic regions with high ROH frequencies, which indicated strong selection. The same study found some candidate genes *TRAPPC9, IRS2* and *HSF1* within ROH islands which correlated with milk yield, milk composition, lactation and heat adaptation (Peripolli et al. 2017). The presence of ROH islands correlated with production traits suggests that these islands serve as markers of selection, which helps in the assessment of genes related to economic traits and adaptation (Peripolli et al. 2017). Analyzing ROH islands offers a framework for locating candidate genes that can refine breeding strategies targeting specific traits (Mastrangelo et al. 2018). This approach is especially relevant to populations with small Ne, where limited genetic diversity requires careful management to preserve both productivity and health traits (Mastrangelo et al. 2018; Dixit et al. 2020).

3. Method

The project integrates computational models of cattle genomes with population histories from Swedish breeds to predict the distribution of homozygosity and genomic inbreeding resulting from genetic drift. To investigate how population history, inbreeding, and selection influence ROH architecture in cattle, this study employed two complementary strategies. Firstly, analysis of empirical genomic data using PLINK for ROH detection, and secondly simulation of genomic data using msprime to model population-specific demographic scenarios. Lastly, all statistical analyses and graphs were performed in R version 4.3.1 (Rstudio Team 2023).

The methodological choices were made to ensure comparability between real and simulated data, while minimizing biases due to data quality or software-specific artifacts. The following section describes the data sources, filtering criteria, simulation design, and ROH detection pipeline in detail.

3.1 Data

The simulated population comprises Swedish breeds including Fjäll, Rödkulla, SRB, and Holstein. The genotyping of these samples was conducted using a 150K Illumina array, as previously described by Upadhyay et al. (2019). For the international breeds Holstein, Jersey and the published Holstein population history of Macleod et al., 2013 were used and the extraction was made by Adepoju et al. (2024) from the 1000 Bull genome project, run 9, downloaded from the European Nucleotide Archive (project accession PRJEB56689). The 1000 Bull genome project data is based on Illumina sequencing provided as genotypes in variant call format (VCF). The simulated breeds and empirical data were both derived from the works of Upadhyay et al. (2019) and Adepoju et al. (2024), respectively.

3.2 Simulation of population histories

Population genetic simulations were conducted using msprime v1.2, a coalescent-based simulator designed for efficient and scalable modelling of ancestral recombination graphs (Kelleher & Lohse 2020). Simulations aimed to replicate historical demographic events such as bottlenecks, population separations, and varying Ne, as described in Adepoju et al. (2024). The different cattle population consisted of Holstein, Jersey, Swedish red polled (SRB) and Swedish mountain breed (Fjäll). *Table 2* details the number of simulated individuals for each breed as well as the genotyping methods employed.

The simulation pipeline was based on publicly available scripts from https://github.com/mrtnj/cattle_population_history, which allow the modelling of cattle-like genome structures as previously presented by Adepoju et al. (2024). Each simulation generated a VCF-file containing 40 individuals per breed. The resulting VCFs were subjected to the same ROH detection pipeline as empirical data to allow direct comparison of ROH patterns across simulated and observed populations. This comparative framework enables the identification of selection signatures and demographic effects by contrasting expected and observed ROH distributions.

3.3 ROH detection using PLINK

ROH were detected using PLINK v1.9, a widely used software tool for population genomic analyses (Purfield et al. 2012; Meyermans et al. 2020). PLINK implements a sliding window approach (--homozyg) which identifies ROH based on user-defined criteria related to SNP density, window size, heterozygosity, and missingness. PLINK evaluates window is for homozygosity based on user-specified criteria, such as the maximum number of heterozygous and missing calls allowed (Chang et al. 2023)

The default procedure involves shifting the window one SNP at a time across the genome. If a sufficient proportion of the window meets the homozygosity criteria (i.e., the window “passes”), PLINK records the window as part of a potential ROH. Consecutive overlapping windows that pass the threshold are then merged to form continuous homozygous segments, which are subsequently defined as ROH (Chang et al. 2023).

Several settings were adjusted from PLINK v.1.9 defaults to optimize the detection of both long and short ROH. Which minimizes false positives and ensures compatibility between datasets of varying SNP density (Ceballos et al. 2018; Meyermans et al. 2020; Chang et al. 2023).

Settings of PLINK v.1.9 used in simulated datasets:

- Minimum number of SNPs per ROH: 100 (default PLINK v.1.9 setting)
- Minimum ROH length: 1000 kb (default PLINK v.1.9 setting)
- Maximum allowed gap between two SNPs: 1000 kb (default PLINK v.1.9 setting)
- Minimum SNP density: 1 SNP per 50 kb (default PLINK v.1.9 setting)

- Maximum heterozygous SNPs per window: 1 (default PLINK v.1.9 setting)
- Maximum missing SNPs per window: 5 (default PLINK v.1.9 setting)
- Window hit rate threshold: ≥ 5% (default PLINK v.1.9 setting)
- Sliding window size: 50 SNPs (default PLINK v.1.9 setting)

ROH segments were defined with a minimum length of 100 kb and a minimum of 100 SNPs. The maximum gap between SNPs was set to 1,000 kb and a minimum density of 1 SNP per 50 kb was used, as a default setting by PLINK v.1.9. A sliding window of 50 SNPs allowed at most one heterozygous and five missing calls per window. These settings were used to detect both long and short ROH and adjusted to reflect the SNP density of both empirical and simulated datasets (Meyermans et al. 2020). The minimum ROH length of 1000kb filters out very short ROH segments that might arise by chance. This filter allows for all detected ROH to be biologically meaningful and to still identify short ROH (Purfield et al. 2012).

Each ROH call was exported into .hom files, containing details such as chromosome, start and end positions, and ROH length per individual.

3.4 Filtering criterion and quality control

For the empirical SNP Chip data the breeds Fjäll, SRB, Rödkulla and Swedish Holstein breed were filtered (Adepoju et al. 2024). The high-density SNP chip was already converted into PLINK compatible .map –and -.ped file format.

The dataset was already filtered with a baseline filter consisting of following for the SNP chip from the native Swedish breeds:

- minor allele count (MAC): 1
- File format conversion: PLINK-compatible-.ped and .map files were used

This configuration ensured the removal of SNPs that were monomorphic within each breed, thereby retaining only polymorphic variants for analysis. Thereafter, a filter was implemented to fit this study.

The filter used was following for the empirical SNP chip data:

- Genotype missingness filter: 0.05
- Minimum number of SNPs per ROH: 100
- Minimum ROH length: 1000 kb (default PLINK v.1.9 setting)
- Maximum allowed SNP gap within ROH: 500 kb
- Minimum SNP density: 1 SNP per 50 kb
- Sliding window size: 50 SNPs (default PLINK v.1.9 setting)
- Maximum heterozygous SNPs per window: 1
- Maximum missing SNPs per window: 1
- Window hit rate threshold: ≥5%

Filtering for genotype quality (--geno 0.05) was enabled to remove low-quality variants, as missing or erroneous SNPs can artificially extend or shorten ROH segments (Ceballos et al. 2018). The settings for the simulated population and empirical data were set do differ as little as possible, without the risk of register bad quality data and still be able to compare the two results (Kim et al. 2022). *Table 1* shows the number of individuals that passed the criterion for the empirical data.

The maximum allowed SNP gap within ROH requires at least one SNP every 500kb within an ROH (Purfield et al. 2012). This limits the maximum distance allowed between consecutive SNPs within ROH. Larger gaps may create artificial joins, particularly in sparsely genotyped regions (Meyermans et al. 2020). Minimum number of SNPs per ROH filters out ROH formed by only a few adjacent SNPs, which are more likely to occur by chance, particularly in low-density datasets (Purfield et al. 2012). Minimum ROH length 1000kb specifies the minimum physical length (in kilobases) of an ROH. This helps to eliminate spurious short homozygous segments that may not be biologically meaningful (Purfield et al., 2012). Minimum SNP density requires at least one SNP every 50 kb within an ROH. This ensures consistent SNP coverage and avoids false ROH calls in low-density regions (Meyermans et al. 2020).

The dataset was already filtered with a baseline filter consisting of following for the empirical sequence data were:

- Biallelic SNP filter
- Minor allele frequency (MAF): 0.000001
- Chromosomes analysed 1-29

Added filtering to empirical sequence data:

- Minimum number of SNPs per ROH: 100
- Minimum ROH length: 1000 kb (default PLINK v.1.9 setting)
- Maximum allowed SNP gap within ROH: 500 kb
- Minimum SNP density: 1 SNP per 50 kb
- Sliding window size: 50 SNPs (default PLINK v.1.9 setting)
- Maximum heterozygous SNPs per window: 1
- Maximum missing SNPs per window: 3
- Window hit rate threshold: ≥5%
- Genotype missingness filer: 0.05

The maximum missing SNPs per window (--homozyg-window-missing 3) allows for some missingness while avoiding fragmentation of true ROH, which in the sequence data is set to 3 and in the SNP chip data is set to 1. This accounts for the difference in data quality between SNP Chip and sequence data. The settings of the sequence data were slightly more relaxed to accommodate for higher variant density and complexity of sequence data (Ceballos et al. 2018).

Table 1 Specification of breed and number of individuals that passed the qualification criteria from empirical data.

Breed	Number of individuals	Method
Fjäll	23	SNP Chip
Rödkulla	18	SNP Chip
Holstein	24	SNP Chip

SRB	25	SNP Chip
Holstein (International)	146	Sequencing

Table 2 Specification and breed and number of individuals from simulated data.

Breed	Number of individuals	Method
Fjällko	40	SNP Chip
Rödkulla	40	SNP Chip
Holstein	40	SNP Chip
SRB	40	SNP Chip
Holstein (International)	40	Sequencing
Holstein (Macleod et al., 2013)	40	Sequencing
Jersey	40	Sequencing

3.5 Genome tiling and window-based ROH overlap analysis

To facilitate ROH analysis across the genome, the reference genome was partitioned into non-overlapping windows of 1 megabase (Mb) using the GenomicRanges package in R. The window size of 1 Mb was selected to match the expected scale of ROH segments typically found in cattle populations. While smaller windows are computationally feasible, this resolution provides a practical balance between detecting meaningful ROH signals and avoiding fragmentation of longer ROH due to window boundaries. Partial windows at chromosome ends were excluded to avoid biases. Each individual's ROH was mapped against the tiled genome using overlap counting:

- The number of ROH segments overlapping each 1 Mb window was counted per individual.

- A final window matrix was constructed, where each row represents an individual, each column corresponds to a genetic window, and each cell contains the count of overlapping Runs of Homozygosity (ROH) within that specific window for each individual.

3.5.1 Identification of ROH hotspots and ROH islands

To identify ROH hotspots, the summed overlap counts per window were calculated across all individuals in a population. The threshold for defining ROH

islands was set to the mean + 2 standard deviations (SD) of the window overlap distribution for the simulated populations (Purfield et al. 2012).

The threshold was justified based on properties of the normal distribution, where approximately 95% of observations fall within ±2SD. Under a normal distribution, approximately 2.5% of the windows are expected to exceed the +2SD threshold by chance. This approach has been adopted to ensure that only genomic regions with an exceptionally high degree of homozygosity are flagged for further analysis (Purfield et al. 2012; Gorssen et al. 2021).

Formula mean:

$$\bar{x} = \frac{\Sigma_{i=1}^{n} x_i}{n}$$

Formula median:

$$median = \frac{n+1}{2}$$

Formula standard deviation:

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

Windows with counts above this threshold were classified as ROH islands. The threshold was set to justify the balancing of false positives (random variation) against sensitivity to true selective signals.

3.5.2 Categorizing ROH by length

To estimate the distribution of ROH segments both within and between the populations, the ROH segments were classified by length. The different lengths were conducted based on previous studies with similar methods (Gibson et al. 2006; Zhang et al. 2015; Kim et al. 2022):

- Short: 1–3 Mb
- Medium: 3–6 Mb
- Long: >6 Mb

These lengths follow established criteria for ROH and align with the previously decided 1 Mb tiling resolution (Gibson et al. 2006; Kim et al. 2017).

3.5.3 Statistical analysis and visualization

All analyses were performed in R version 4.3.1 (Rstudio Team 2023) using the tidyverse suite, including dplyr for data handling and ggplot2 for visualization. For each simulated and empirical population, descriptive statistics were calculated: Mean, median, minimum, maximum, interquartile range of window overlap counts and ROH frequency distributions were plotted across the genome. Comparisons between simulated and empirical ROH patterns were visualized through density plots and genome-wide ROH maps.

An analysis of variance (ANOVA) and a Kruskal-Walli's test was performed to compare genomic inbreeding coefficients (ROH %) across the different populations. The aim was to assess whether there were statistically significant differences between the simulated means and those derived from empirical data. Since both tests yielded consistent results, only the ANOVA output is presented in *Figure 15* and *Figure 16* for clarity.

3.6 Locating genes

Genes located within genomic regions identified as ROH hotspots and ROH islands were examined to assess potential associations between homozygous regions and beneficial traits. Threshold values were applied to define ROH regions, after which the corresponding genomic coordinates were used to identify overlapping genes. Genes located within ROH islands were annotated using Ensembl release 109. Genes without known annotations were reported by their Ensembl gene ID only (Ensembl 2024).

4. Results

This section presents the results from both simulated and empirical analyses of ROH in Swedish and international cattle breeds. The simulations were designed to reflect historical population dynamics and were analysed in parallel with real genomic data to facilitate direct comparisons.

Overall, the results highlight clear differences in ROH distribution patterns between native Swedish breeds and highly selected international breeds. Swedish breeds generally exhibited a larger number of shorter ROH segments, suggestive of ancient inbreeding and historical bottlenecks. In contrast, international breeds such as Holstein and Jersey displayed fewer but longer ROH segments, consistent with recent inbreeding and intensive selection pressure.

ROH islands and ROH hotspots were identified in both simulated and empirical datasets. These islands varied in location and size across breeds and were further investigated for gene content to assess possible functional relevance.

4.1 ROH distribution across the genome

To visualise the distribution of ROH over the genome when the genome is divided into 1 Mb windows. The plots illustrate differences in ROH accumulation patterns between breeds. Generally, the Swedish native breeds showed several regions with high or moderate ROH counts. The Fjäll and Rödkulla breeds displays several regions with moderate ROH clustering (*Figure 1* and *Figure 3)*, while Holstein shows a more distinct presence of high frequency of ROH (*Figure 2)*. The sequence data of Holstein displayed the highest level of genomic inbreeding among all the breeds (*Figure 5).*

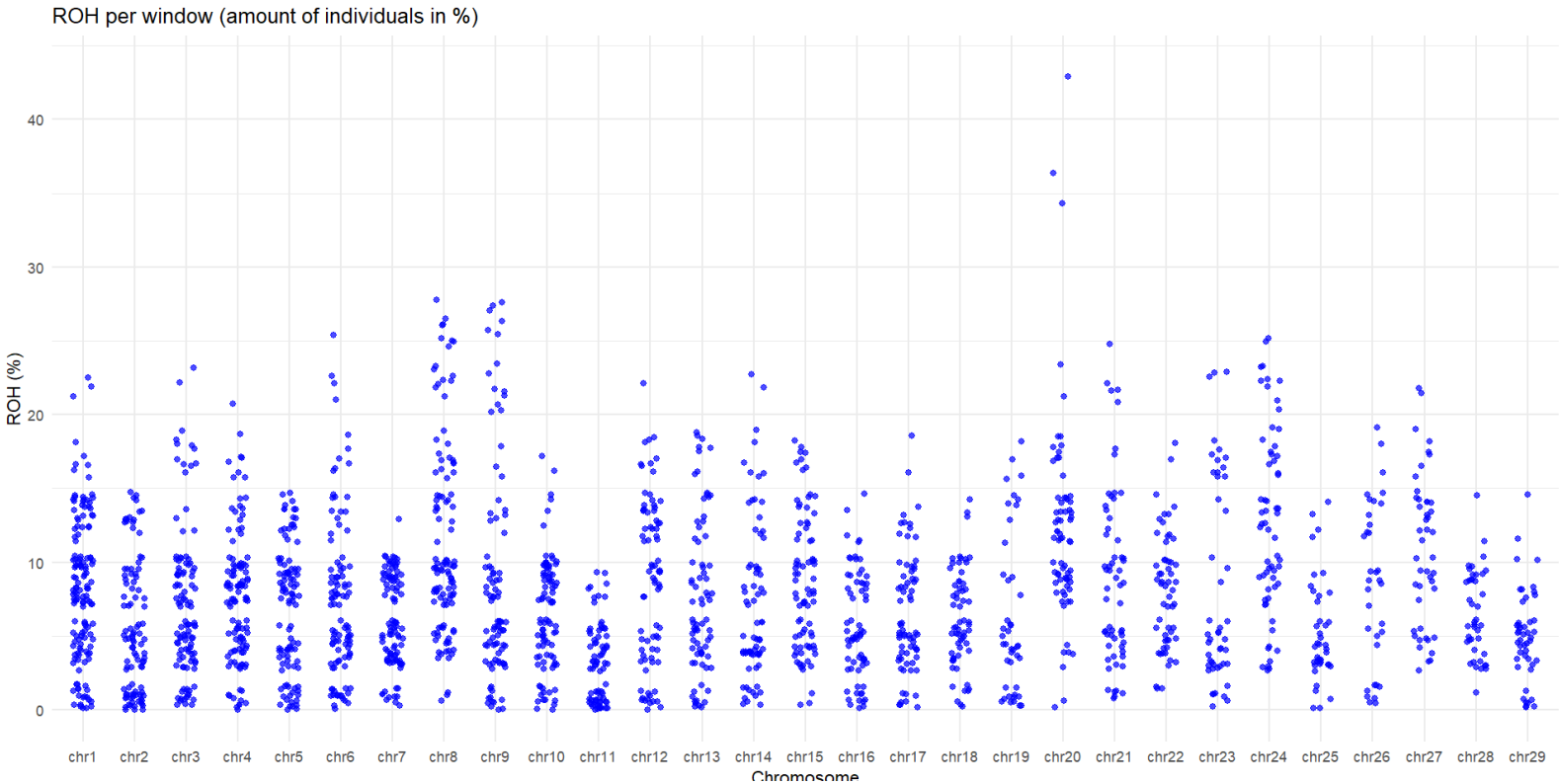


Figure 1 scatterplot depicts the distribution of "ROH-window-counts" across genome of empirical Fjäll SNP Chip data.

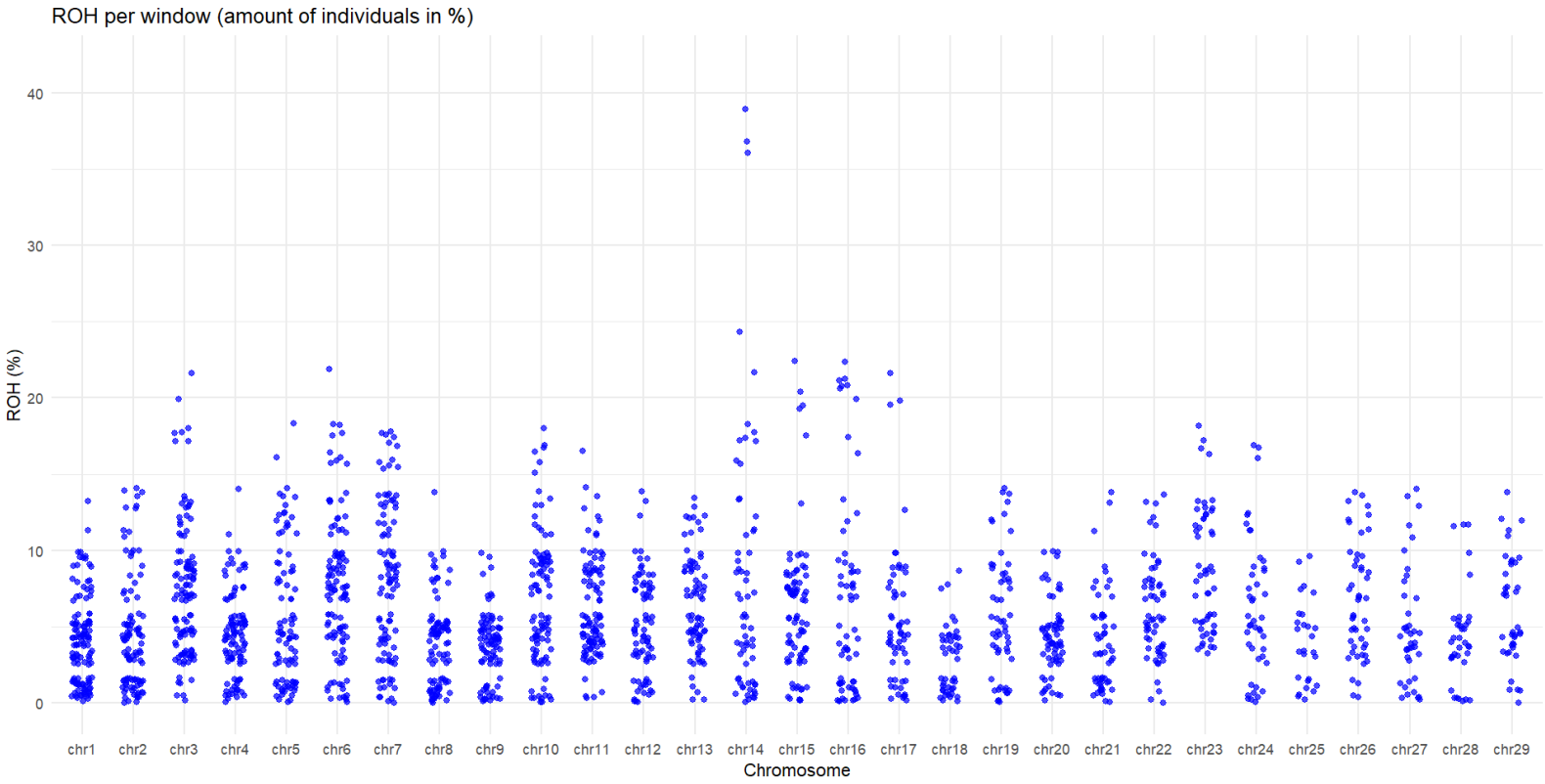


Figure 2 scatterplot depicts the distribution of "ROH-window-counts" across genome of empirical Holstein SNP Chip data.

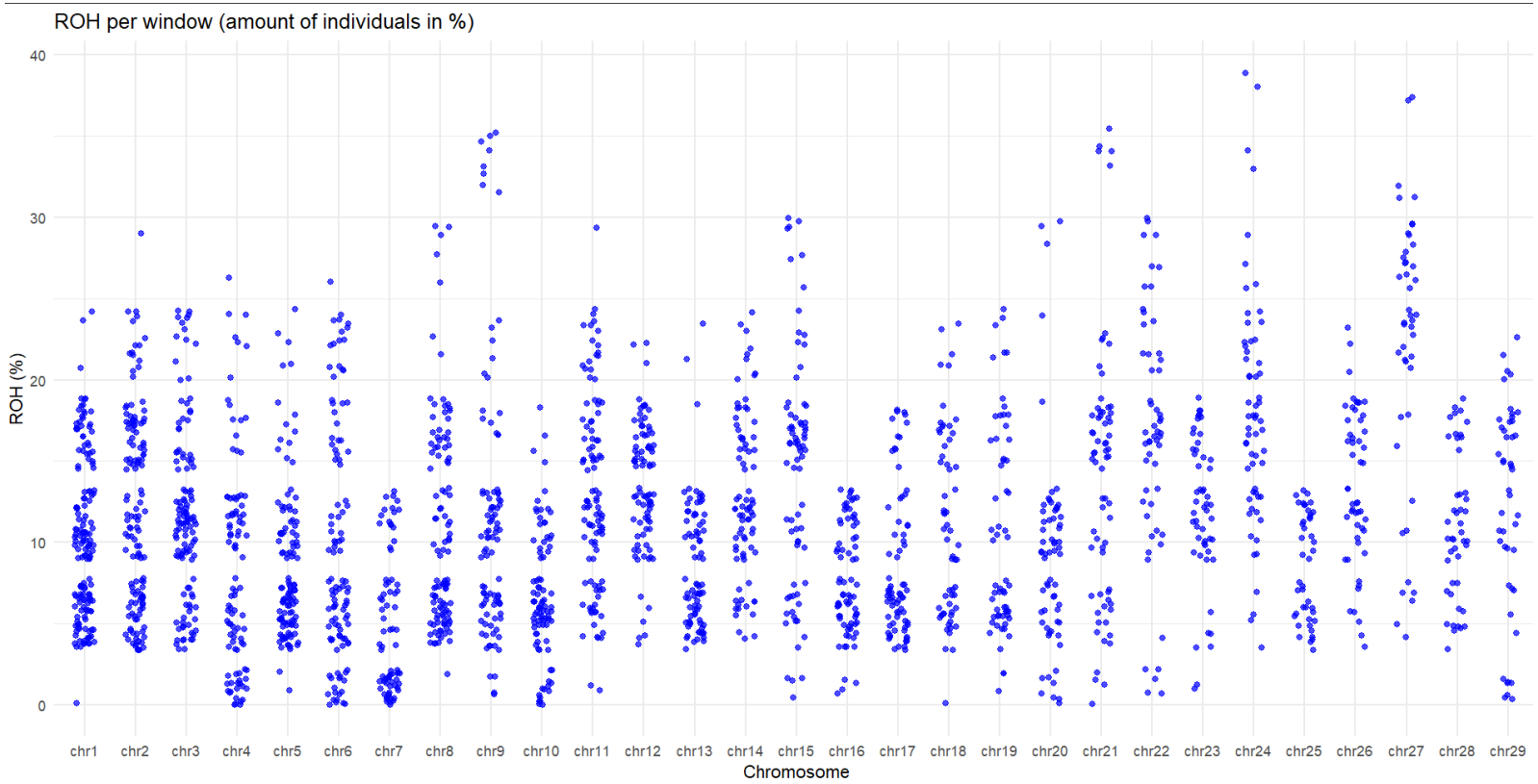


Figure 3 This scatterplot depicts the distribution of "ROH-window-counts" across genome of empirical Rödkulla SNP Chip data.

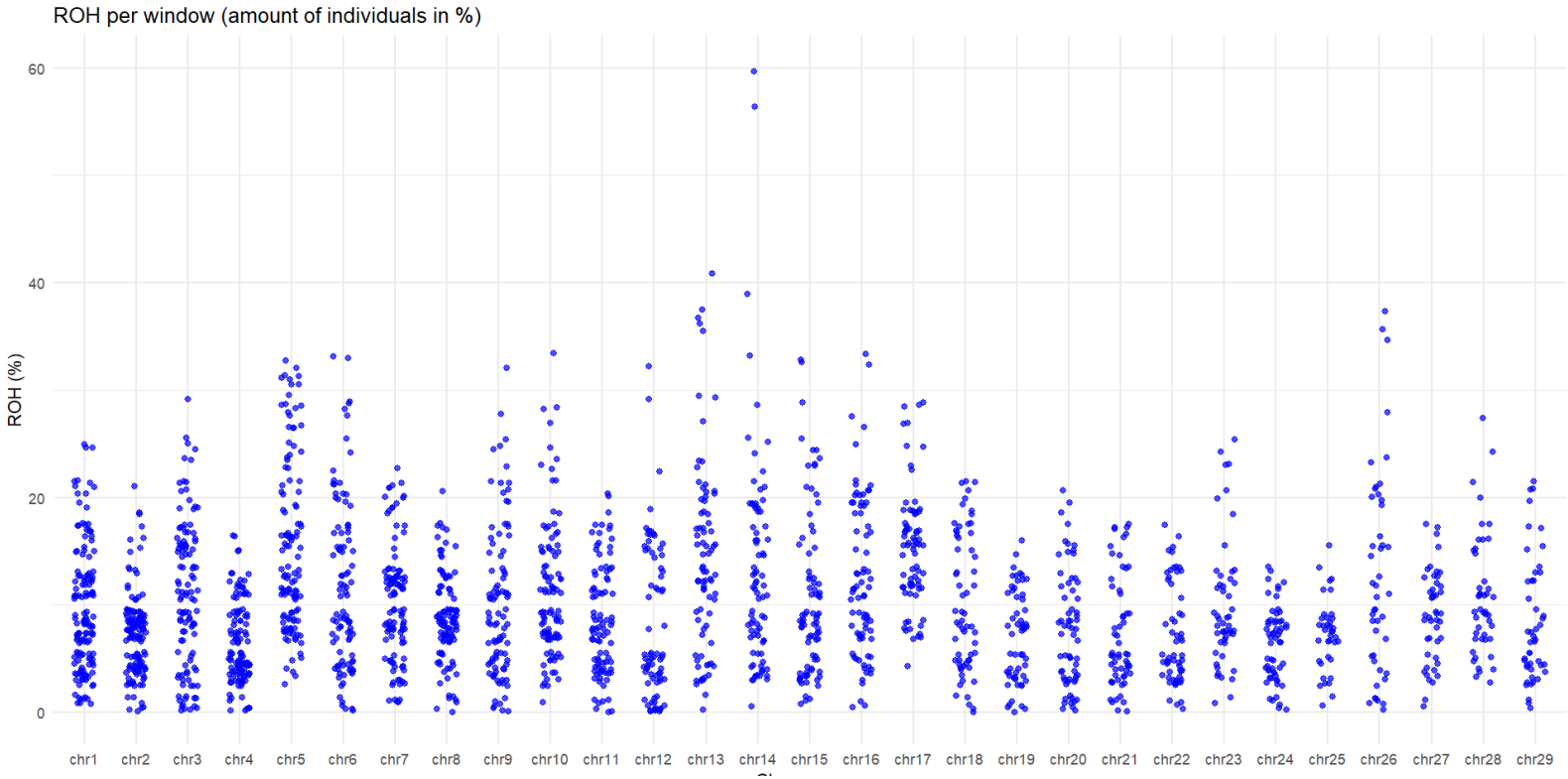


Figure 4 This scatterplot depicts the distribution of "ROH-window-counts" across the genome of empirical SRB SNP Chip data.

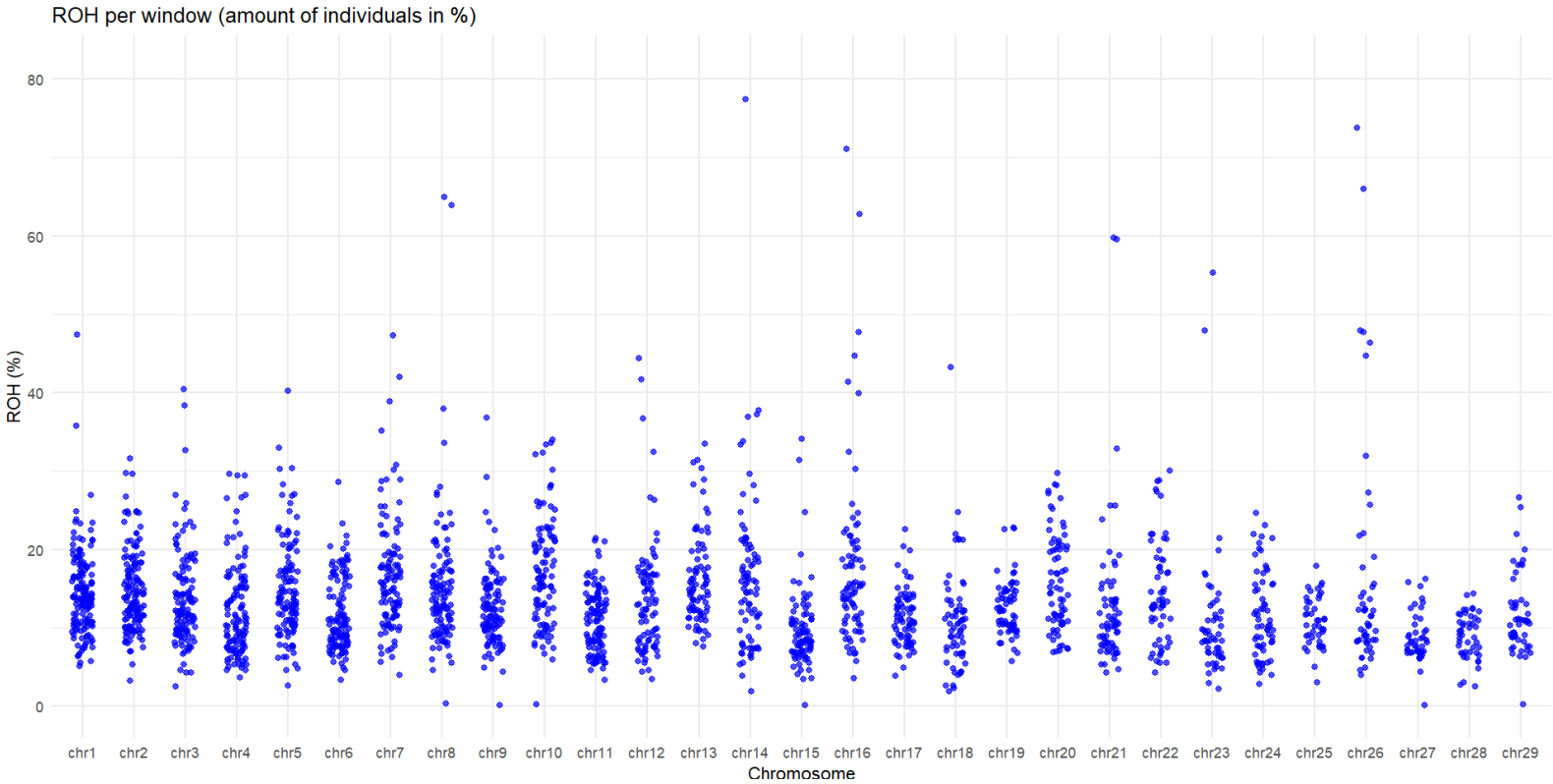


Figure 5 This scatterplot depicts the distribution of "ROH-window-counts" across genome of empirical sequence data of Holstein.

4.2 Detection of ROH hotspots and ROH islands

ROH were analysed across simulated cattle genomes by scanning non-overlapping 1 Mb windows and defining ROH islands as regions exceeding the mean window homozygosity by two standard deviations.

The Jersey simulation *Figure 11* exhibit the highest ROH coverage and longest ROH segments. In contrast, the two Holstein populations showed relatively fewer ROH segments above the threshold value, although the segments tended to be noticeably long. The Swedish native breeds (Fjäll, Rödkulla and Swedish Red Polled) displayed many homozygous regions. However, they were generally shorter ROH segments compared to the Holstein breed.

The scatterplots show the distribution of ROH per window over the genome and visualizes the threshold and mean in the different populations. The x-axis displays chromosomes numbered from 1 to 29 (chr1 to chr29), while the y-axis measures the number of ROH per window. Each blue dot represents a data point showing ROH frequency for a specific chromosome. The green line is representing the mean, and the red line represents the threshold value.

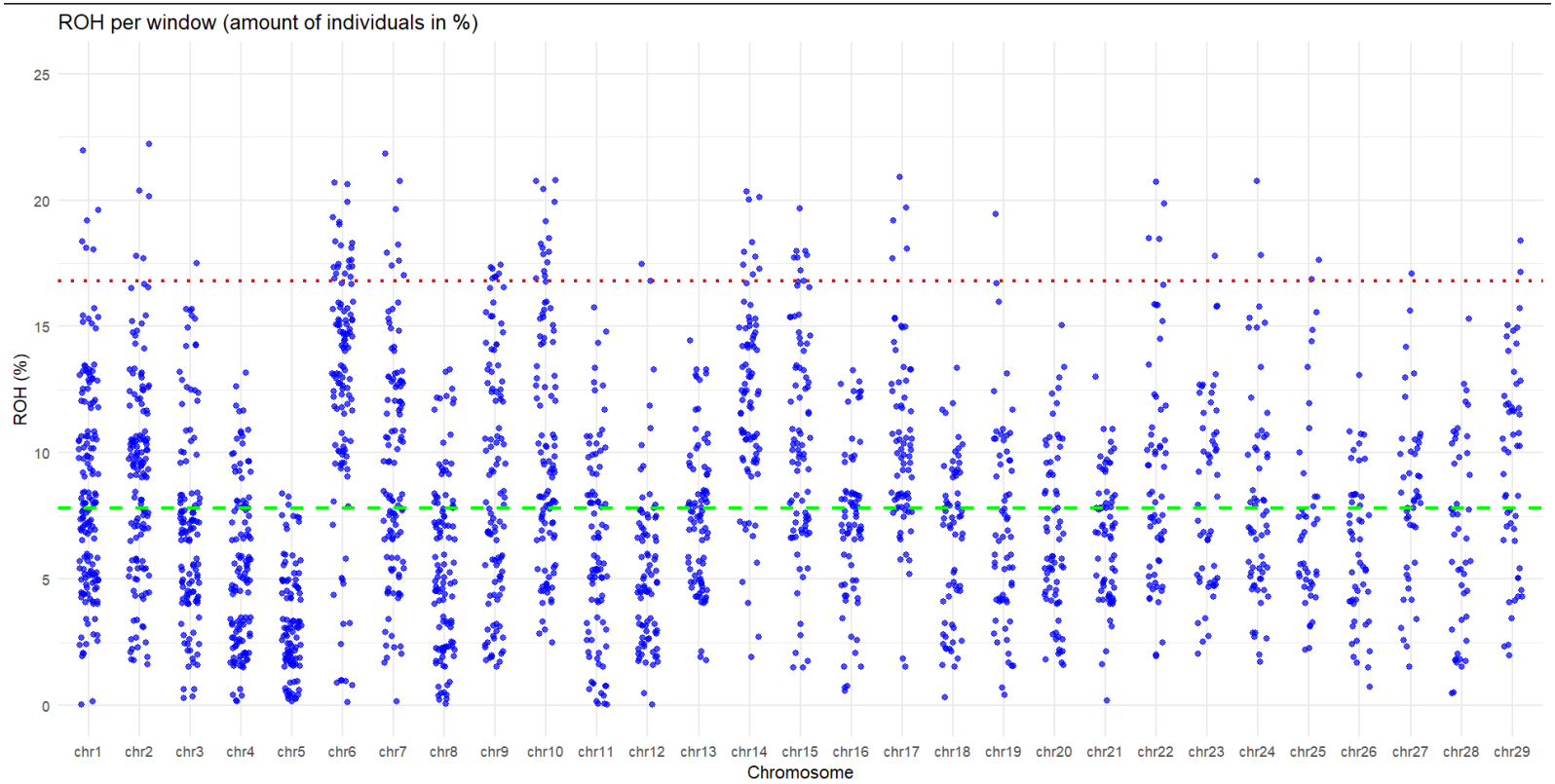


Figure 6 The scatterplot illustrates the distribution of ROH per window across different chromosomes in simulated Fjäll cattle.

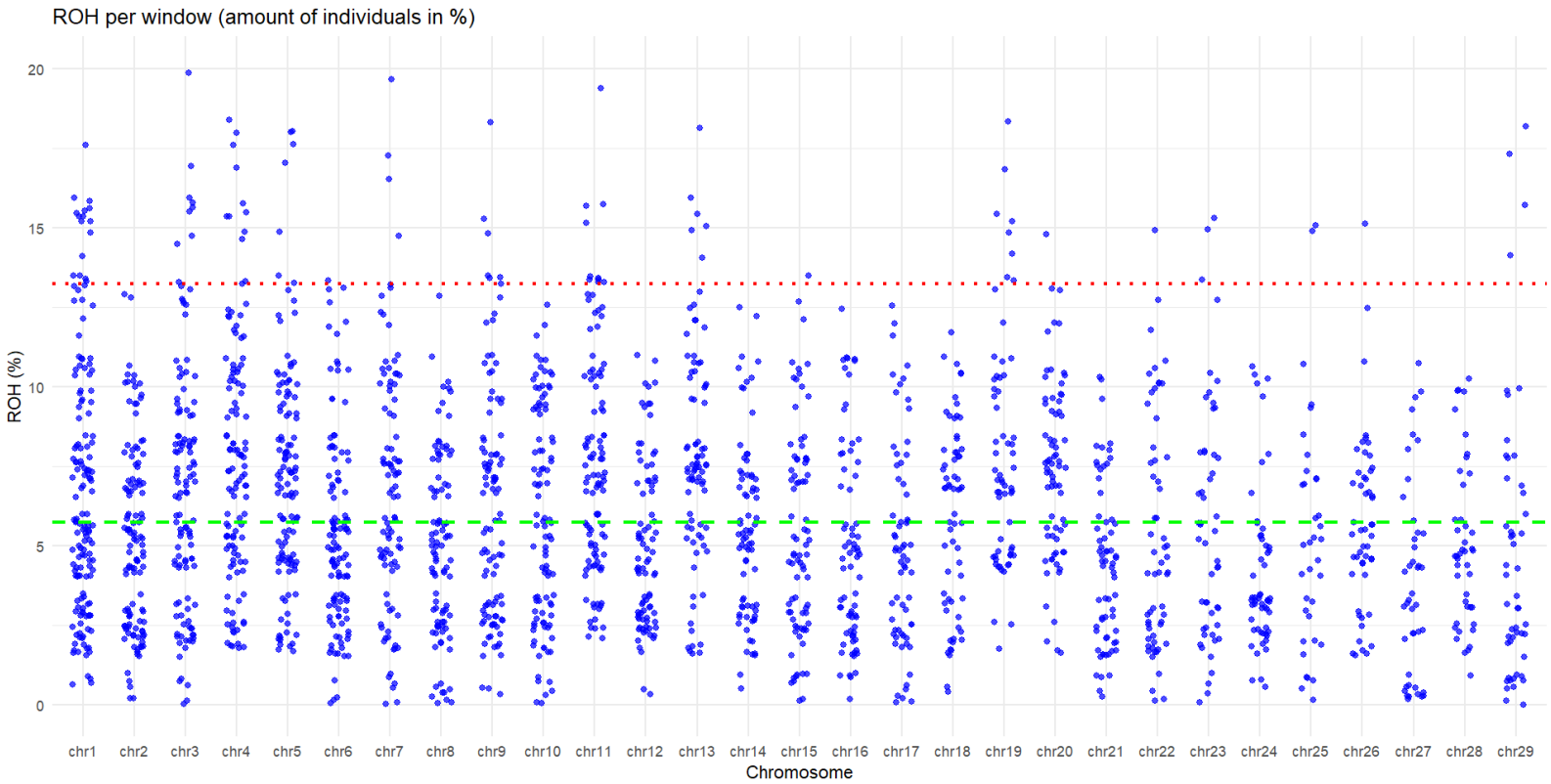


Figure 7 The scatterplot illustrates the distribution of ROH per window across different chromosomes in simulated Holstein SNP chip cattle.

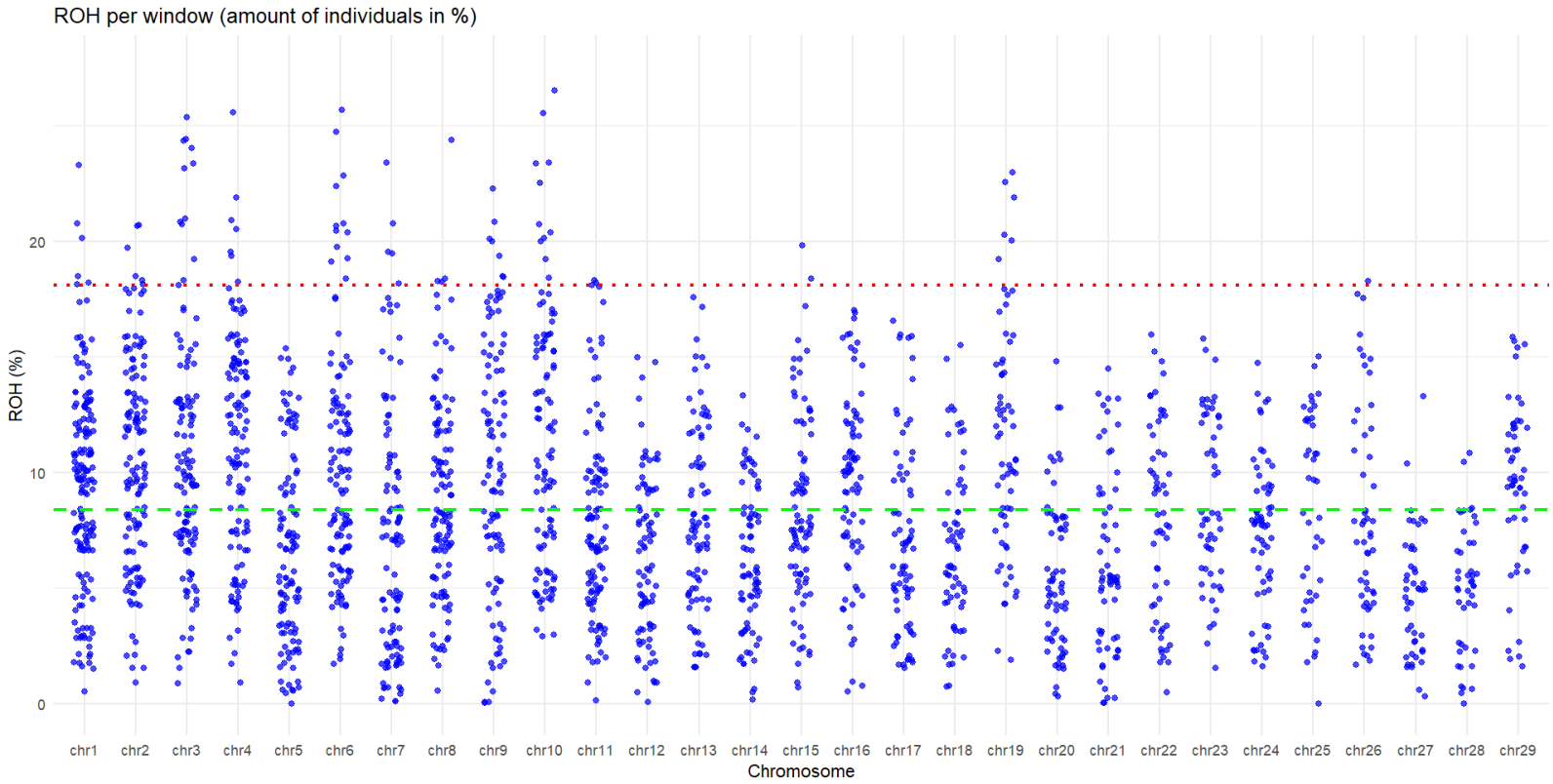


Figure 8 The scatterplot illustrates the distribution of ROH per window across different chromosomes in simulated Holstein sequence data.

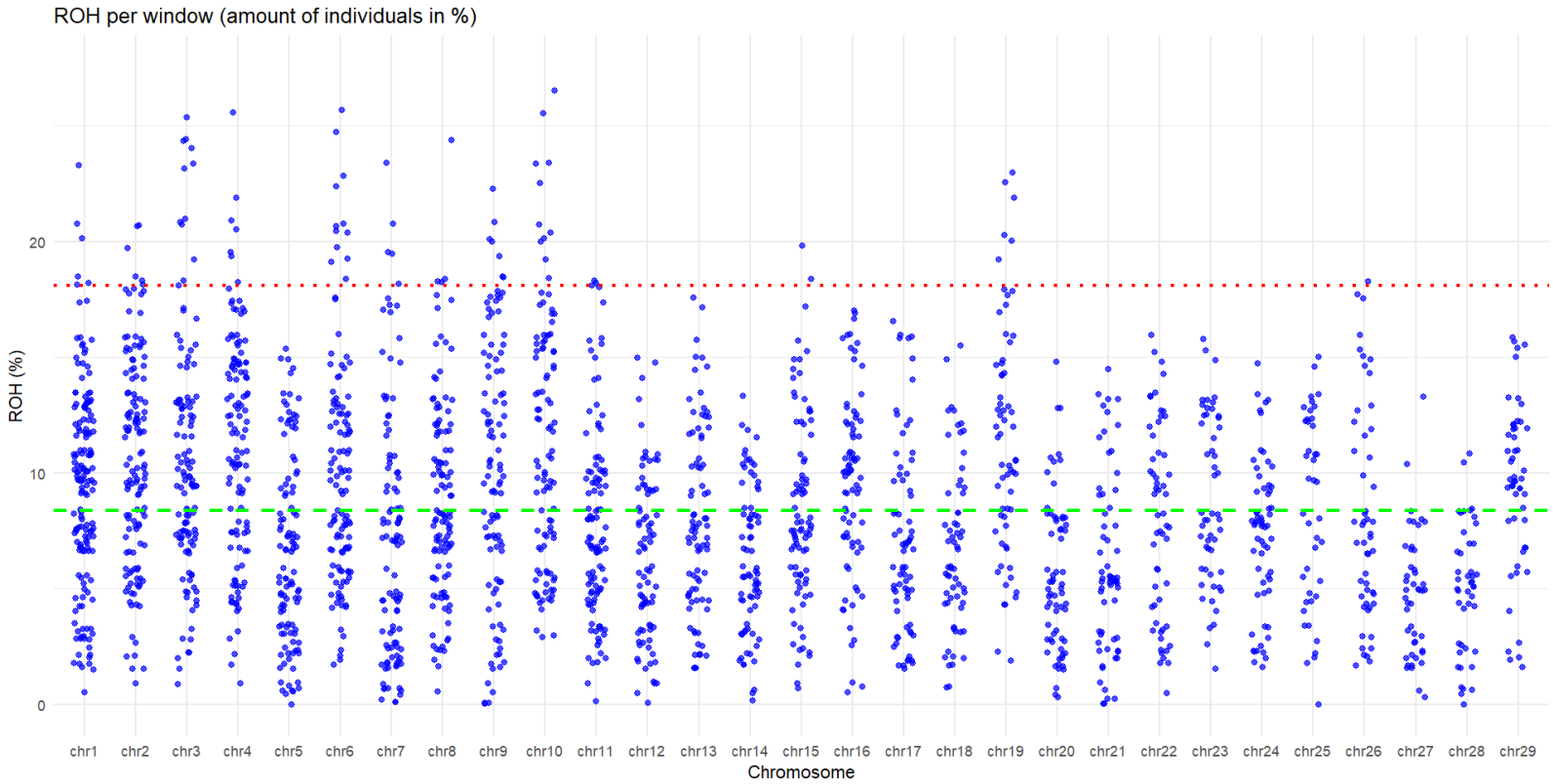


Figure 9 The scatterplot illustrates the distribution of ROH per window across different chromosomes in simulated Rödkulla cattle.

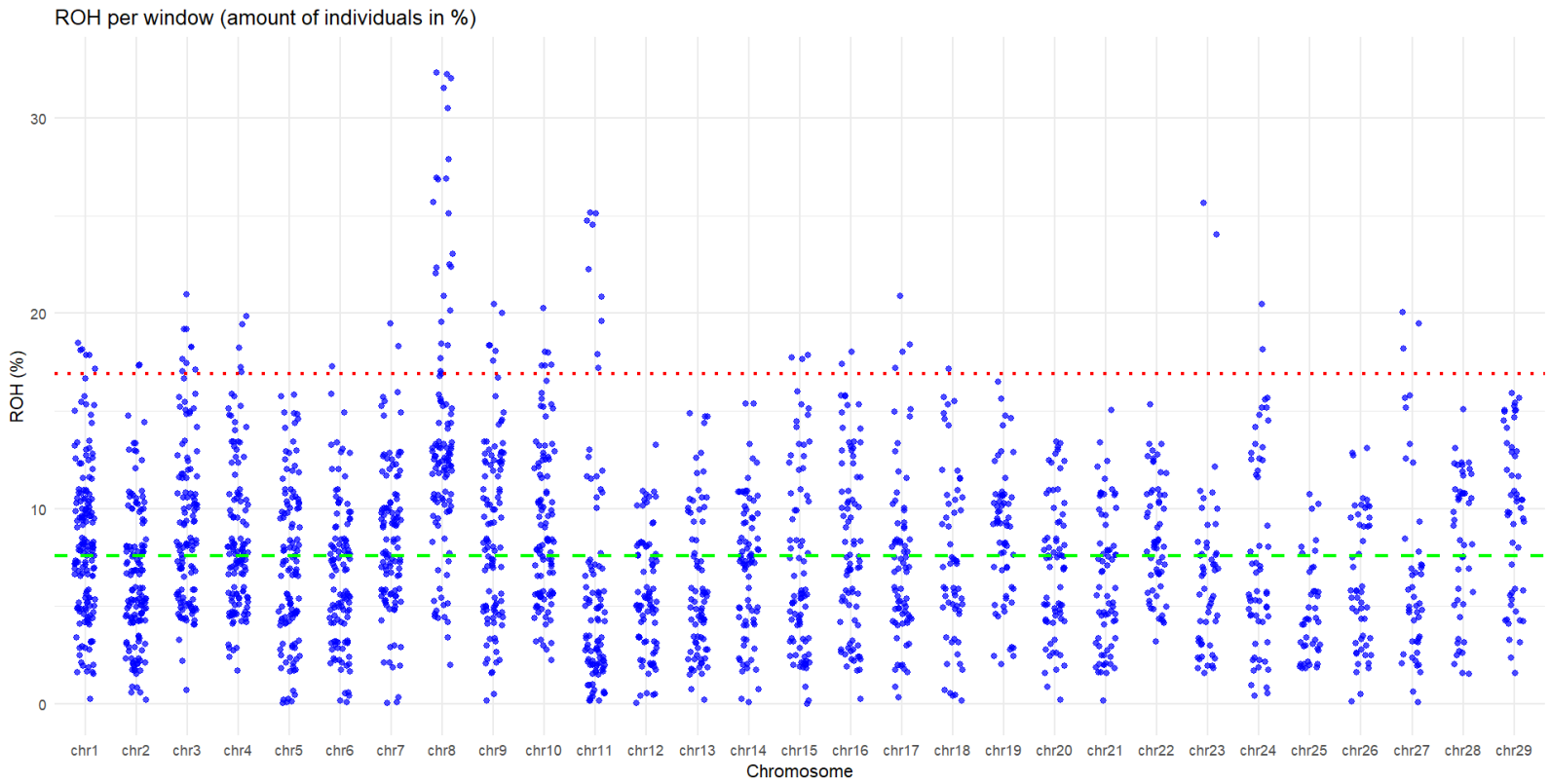


Figure 10 The scatterplot illustrates the distribution of ROH per window across different chromosomes in simulated SRB cattle. In this scatterplot, each point represents a genomic window.

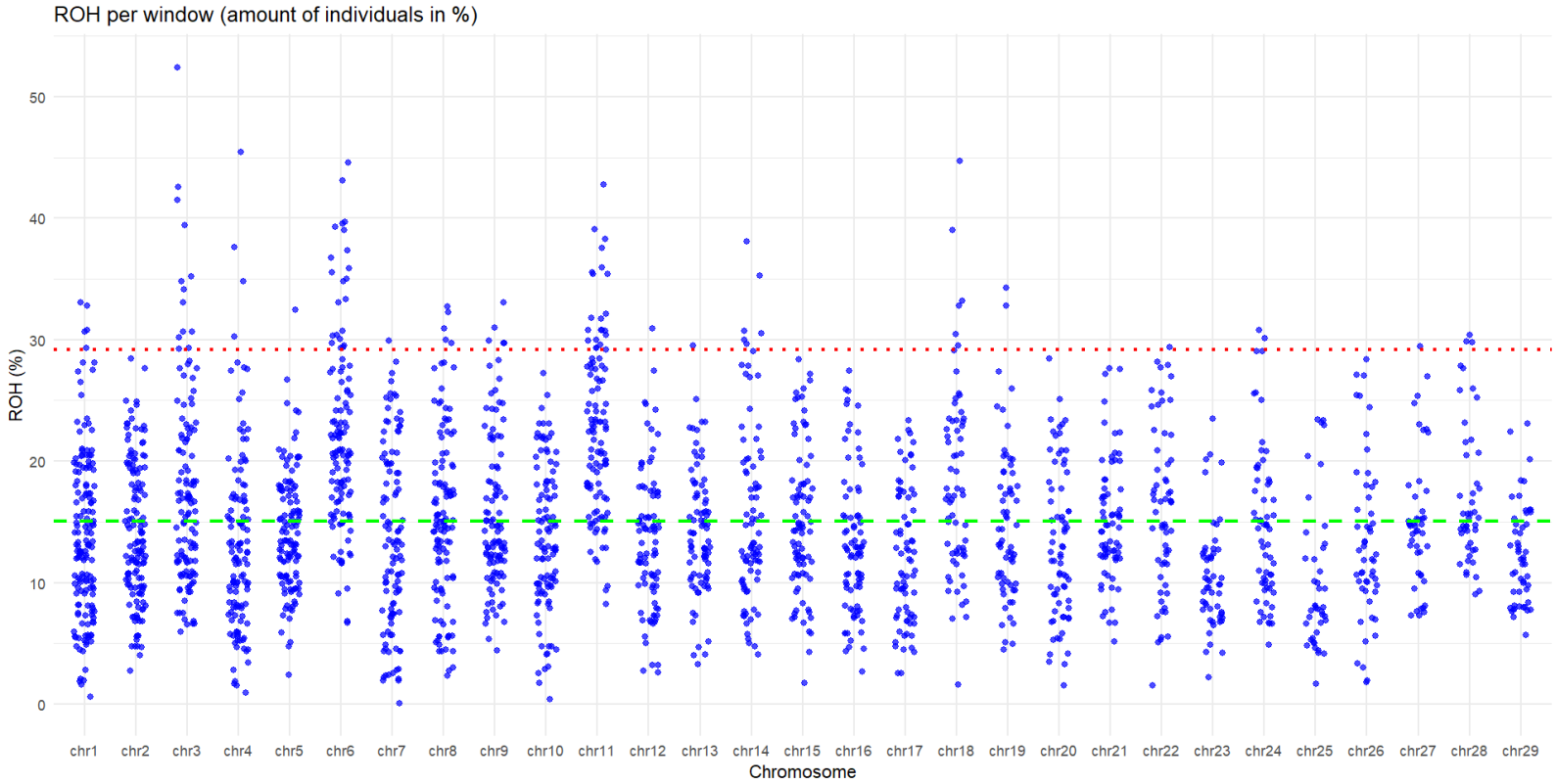


Figure 11 The scatterplot illustrates the distribution of ROH per window across different chromosomes in simulated Jersey cattle.

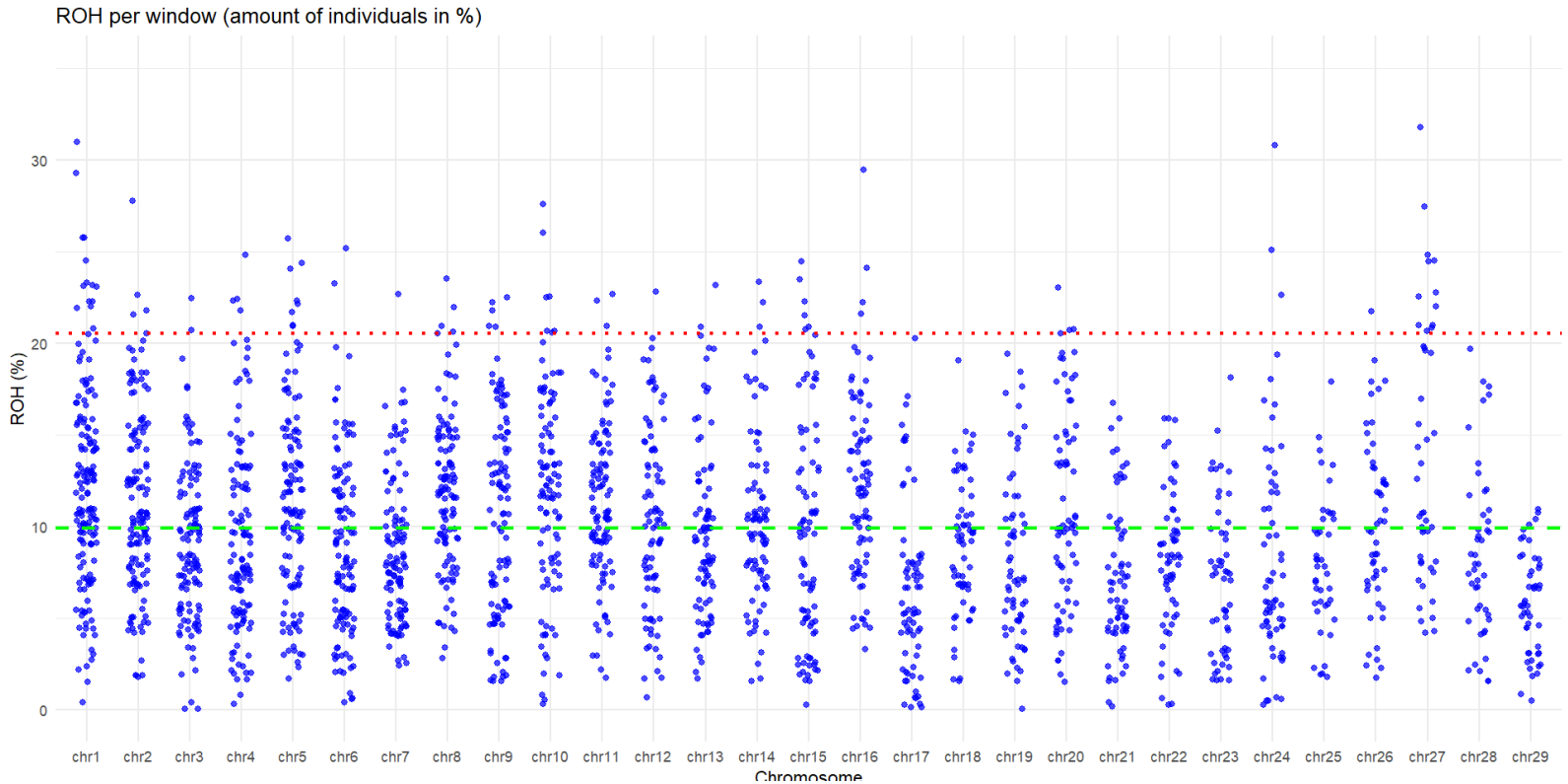


Figure 12 The scatterplot illustrates the distribution of ROH per window across different chromosomes in empirical Holstein Macleod 2013 data.

4.3 ROH coverage and distribution

The differences in ROH coverage by length class indicate that each breed has a distinct genomic history. As displayed in Figure 13, the Swedish native breeds showed mostly medium (3-6Mbp) and small (1-3 Mbp) ROH coverage.

In contrast, as shown in *Figure 14*, Holstein cattle display a high number of ROHs that predominantly fall into the medium (3–6 Mbp) and large (>6Mbp) categories. Although their ROH frequency is high, these segments cover only a small portion of the genome. Meanwhile, Jersey cattle exhibit the broadest coverage of large (>6 Mbp) and medium (3–6 Mbp) ROH.

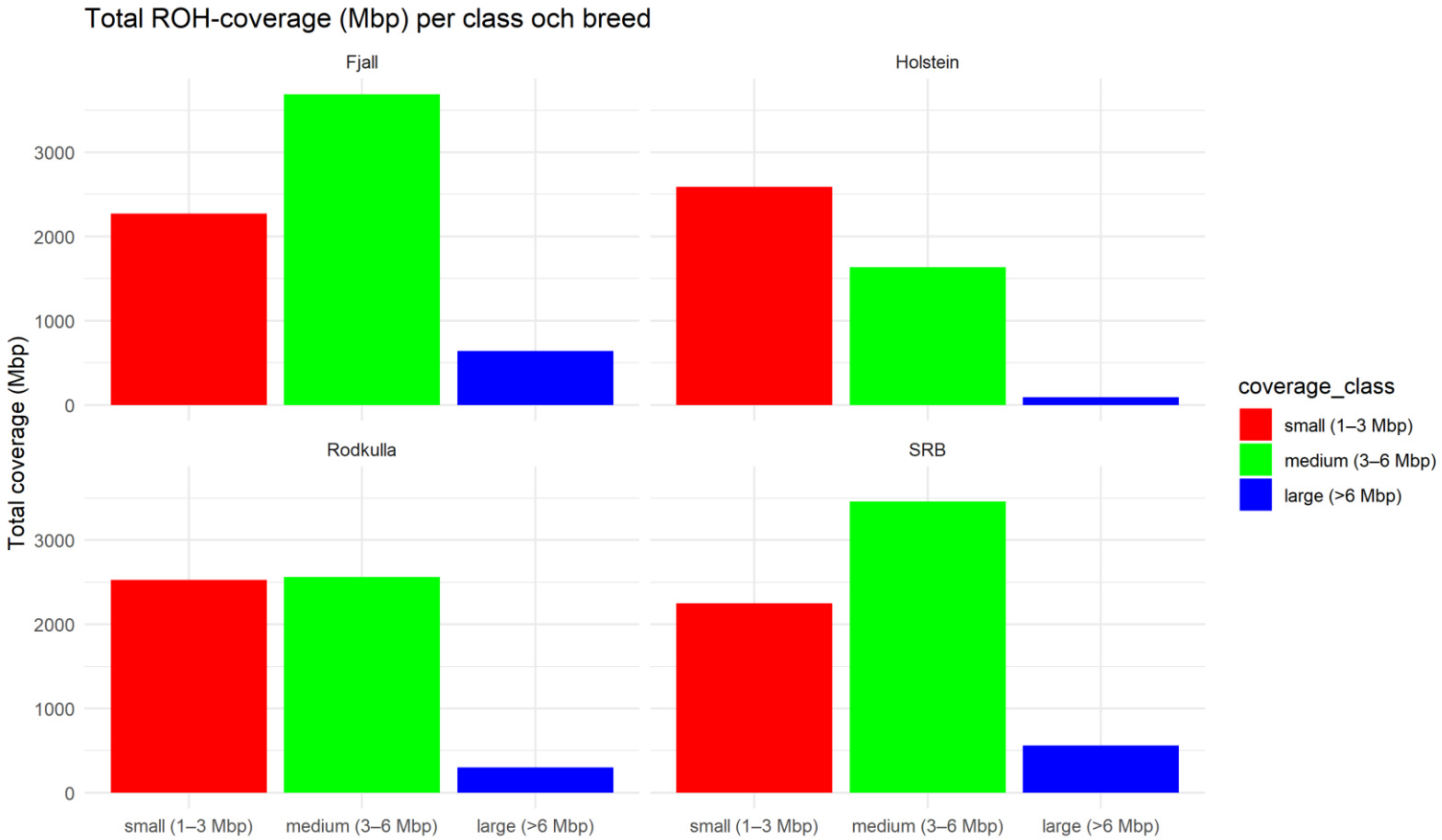


Figure 13 Total ROH coverage (Mbp) per class and breed," compares Runs of Homozygosity (ROH) coverage across four cattle breeds: Fjäll, Holstein, Rödkulla, and SRB. The ROH segments are classified into three categories based on size: small (1-3 Mbp), medium (3-6 Mbp)

Figure 14 Total ROH coverage (Mbp) per class and breed. Compares Runs of Homozygosity (ROH) coverage across three cattle breeds: Holstein Macleod2013, Holstein_seq (international), and Jersey.

4.4 Implications of mean and comparison between breeds

The ANOVA was preformed to show if the ROH inbreeding coefficients differentiated between the empirical data and simulated data. The results gave an indication on how well the simulation model represented reality considering selection, bottlenecks and genetic drift. The results in *Figure 15* and *Figure 16* shows that the simulated sequence data might underestimate the ROH distribution in native breeds with small Ne. Whereas, the big sequence dataset was overestimated in ROH count. A post-hoc Tukey Honest Significant Difference (HSD) test was performed to pinpoint which breeds differ significantly in their ROH levels. The results demonstrated disparities displayed in *Figure 15* and *Figure 16*.

Holstein (international), which is the empirical sequence dataset for Holstein exhibit the highest genomic inbreeding coefficient (F_ROH), which is significantly greater than all other populations. This suggests higher levels of genomic homozygosity. The resulted difference between the empirical Holstein sequence data and the simulation indicates that a factor not accounted for in the simulated is affecting the empirical ROH inbreeding coefficient in the sequence data. The native Fjäll populations did correlate quite well between the empirical dataset and the simulation. However, there is a slight overestimation of ROH count in the simulation.

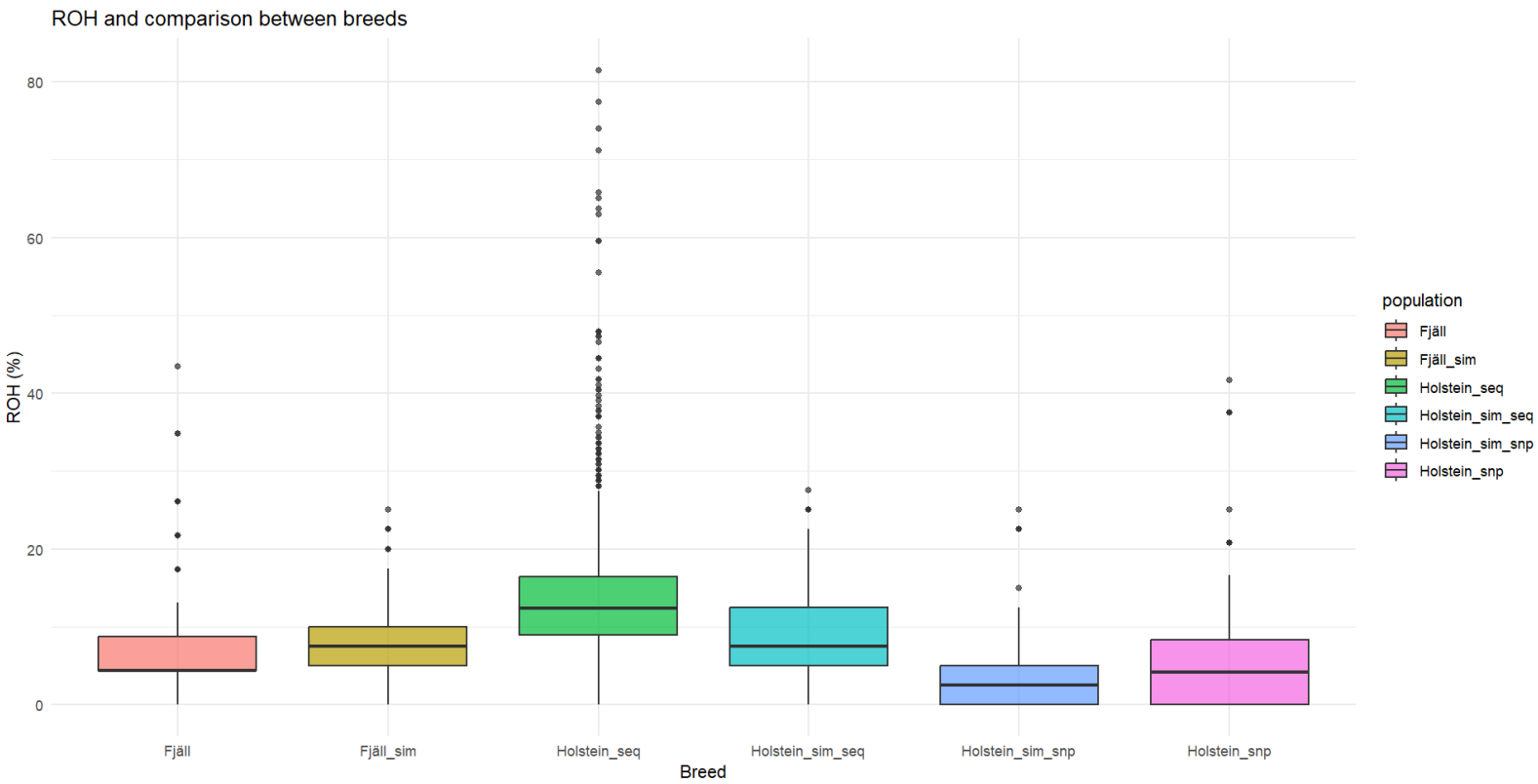


Figure 15 Visual represents the distribution of genomic inbreeding coefficient (F_ROH) across six different cattle populations based on ANOVA. In this figure ROH (%) visualises the inbreeding coefficient on the y-axis and breeds on the x-axis.

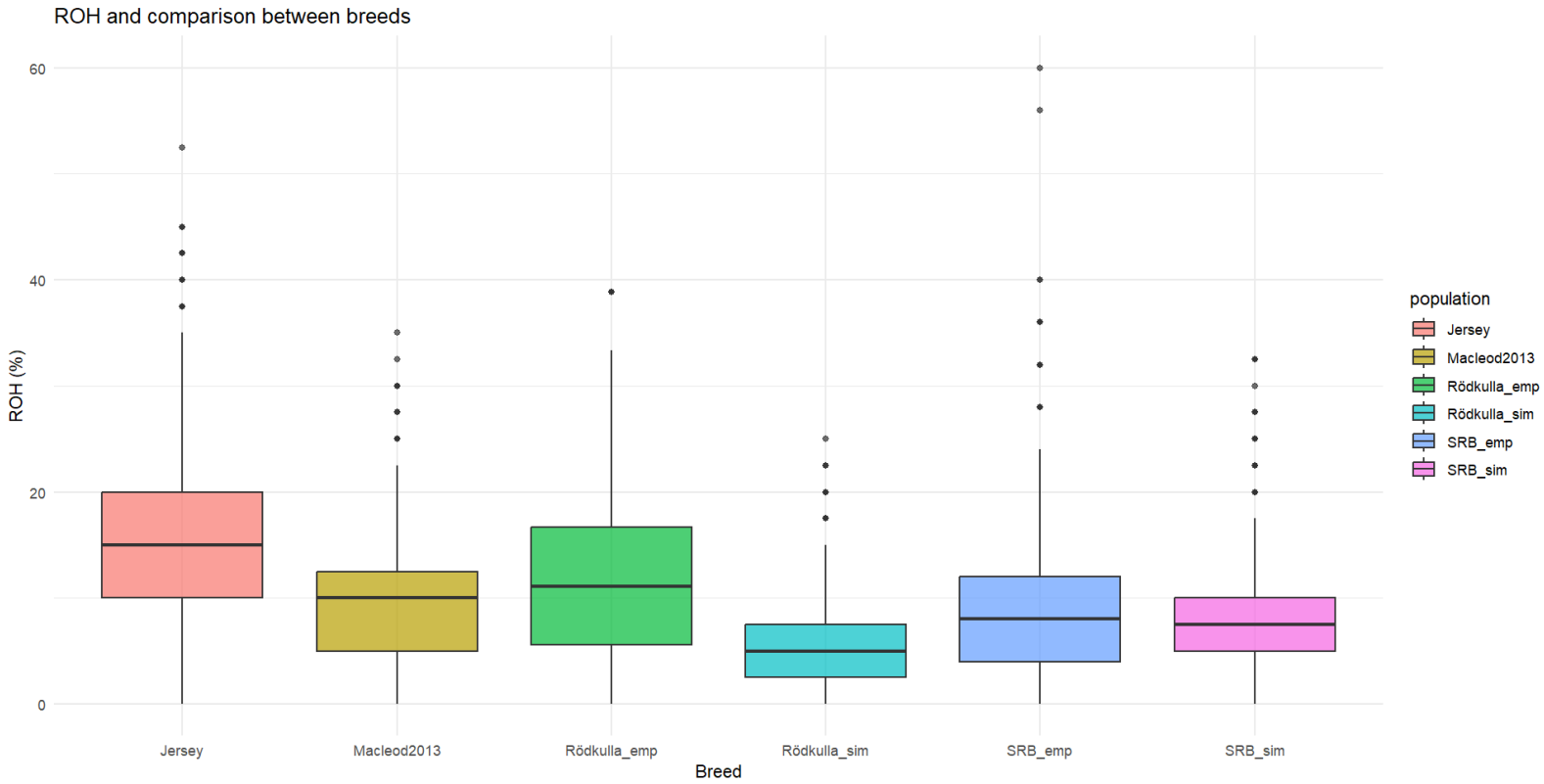


Figure 16 Visual represents the distribution of genomic inbreeding coefficient (F_ROH) across six different cattle populations based on ANOVA. In this figure ROH (%) visualises the inbreeding coefficient on the y-axis and breeds on the x-axis.

4.5 Genes found in ROH islands and ROH hotspots

The ROH hotspots and islands exceeding the defined threshold were mapped and subsequently annotated using the Ensembl genome browser (Ensembl, 2024). The table below presents the genomic coordinates of each ROH region (in Mb), along with the corresponding genes identified within those regions. For genes lacking functional annotation, only the Ensembl transcript ID is listed. For annotated genes, both the ID and known gene function are provided. ROH regions found in the same genomic location across multiple breeds are classified as ROH islands, while those occurring in only one breed are referred to as ROH hotspots.

Table 3 Genomic regions with detected ROH hotspots or islands.

Ensembl transcript ID	Gene	Breed	Region (Mb)	Trait	Reference
ENSBTAG00000009188 ENSBTAG00000009187 ENSBTAG00000001258 ENSBTAG00000067223 ENSBTAG00000043993 ENSBTAG00000003064 NSBTAG000000003063 ENSBTAG00000003059	*GART* *DNAJC28* *TMEM50B* *EPCIP* *PAXBP* *SYNJ* *CFAP*	Fjäll, Holstein	1: 2.0-3.0	Fertility Thermoregulation and polled phenotype - - - - -	(Fritz et al. 2013) (Fernandes Júnior et al. 2020; Ghoreishifar et al. 2020)
ENSBTAG00000017310 ENSBTAG00000023718 ENSBTAG00000020762	*EVA1C* *MRAP* *HUNK* -	Fjäll, Holstein	1:3.0-4.0	Polled phenotype Visceral fat accumulation - -	(Ghoreishifar et al. 2020) (Novo et al. 2022)

ENSBTAG00000018854					
ENSBTAG00000002796 ENSBTAG00000020041 ENSBTAG00000011415 ENSBTAG00000008759 ENSBTAG00000008753 ENSBTAG00000005288 ENSBTAG00000011419	*KLHL3* *FAM13* *ETF1* *CDC23* *BRD8* *GRFA3* *HSPA9*	Fjäll Holstein	7:4.9–5.0	Temperament and stress response Energy Metabolism - Environmental adaptation Growth regulation Nervous system function Stress response and cell function	(Ruiz-De-La-Cruz et al. 2024) (Liang et al. 2021) (Liu et al. 2022) (Johnston et al. 2021) (Wang et al. 2023) (Tripathy et al. 2021)
ENSBTAG00000007782 ENSBTAG00000007785 ENSBTAG00000008758 ENSBTAG00000030529 ENSBTAG00000010063	*MYOT* *PKD2L2* *KIF20A* *FAM53C* *KDM3B*	Fjäll Holstein	7: 4.9-5.0	Meat quality Fertility - - Fertility	(Silva et al. 2020) (Nani & Peñagaricano 2020) (Ebenezer Samuel King et al. 2022)
ENSBTAG00000004415 ENSBTAG00000033747 ENSBTAG00000010798 ENSBTAG00000030520 ENSBTAG00000017957 ENSBTAG00000002286	- *SIL1* *SLC23A1* *PROB1* *SPATA24* *DNAJC18* *ECSCR* *STING1*	Fjäll	7:5.0-5.1	- - Environmental adaptation and distribution of vitamin C - - Heat stress/ environmental adaptation - Thermogenesis and energy metabolism	 (Ayalew et al. 2023) (Liu et al. 2022) (Huang et al. 2023)

ENSBTAG00000030518 ENSBTAG00000002296					
ENSBTAG00000001174 ENSBTAG00000005748 ENSBTAG00000011203 ENSBTAG00000044050 ENSBTAG00000004243 ENSBTAG00000004243 ENSBTAG00000004243	*MRPL15* *SOX17* *RP1* *XKR4* - - -	Fjäll Holstien	14: 22.0-23.0	High feed conversion ratio Embryo development Microtubule-associated transport Feed intake and Fedd conversion ratio - - -	(de las Heras-Saldana et al. 2019) (Abdullah et al. 2023) (Michot et al. 2016) (Lindholm-Perry et al. 2012)
ENSBTAG00000005898 ENSBTAG00000020034 ENSBTAG00000049910 ENSBTAG00000019147 ENSBTAG00000018570	*TGS1* *LYN* *CHCHD7* *RSPS20* *SDR16C5*	Fjäll SRB Holstein	14:23.0-24.0	Pleiotropic Pleiotropic Pleiotropic Pleiotropic Milk and energy metabolism	(Ghoreishifar et al. 2020) (Laodim et al. 2024)
ENSBTAG00000007111 ENSBTAG00000009517 ENSBTAG00000026825 ENSBTAG00000054356	*STEAP3* *DBI* *TMEM37* *CFAP221* *PTPN4* *TMEM177* *EPB4L5*	Holstein	2: 7.1-7.2	Involved in iron metabolism Insulin and energy metabolism Disease resistance Milk content and ion transport - -	(Xing et al. 2025) (Huang et al 2019) (Ghoreishifar et al. 2020) (Singh et al. 2022)

ENSBTAG00000020855 ENSBTAG00000009058 ENSBTAG00000018236				Phenotypic traits and environmental adaptation	(Chen et al. 2023)
ENSBTAG00000032021 ENSBTAG00000049058 ENSBTAG00000011682	*RALB* *INHBB* *GLI2*	Holstein	2:7.2-7.3	- Milk content Growth, reproduction and milk production	(Singh et al. 2022) (Yurchenko et al. 2018)
ENSBTAG00000063036	-	Holstein	3: 7.1-7.2	-	
ENSBTAG00000058741 ENSBTAG00000055278	*U6* *NEGR1*	Holstein	3: 7.2-7.3	- Milk production and energy metabolism	(Wang et al. 2022)
ENSBTAG00000016159 ENSBTAG00000003454 ENSBTAG00000005560 ENSBTAG00000017492	*NPBWR1* *RGS20* *ST18* *PCMTD1*	Holstein	14:21.-22.0	Feed intake and Temperament Environmental adaptation and production traits - Fertility	(Mota et al. 2017) (Adhikari 2024) (Mota et al. 2017)
ENSBTAG00000018272 ENSBTAG00000006609 ENSBTAG00000001401 ENSBTAG00000020518 ENSBTAG00000058678	*RERE* *ERRFI1* *SLC45A1* *PARK7*	Holstein	16:45.0-46.0	- - Reproduction Temperament	(Sun et al. 2023) (Paredes-Sánchez et al. 2023)

ENSBTAG00000051176					
ENSBTAG00000005685 ENSBTAG00000008860	*HSD11B2* *RIPOR1*	Holstein	18: 35.0-36.0		
ENSBTAG00000022890 ENSBTAG00000004650 ENSBTAG00000021003	*MBP* *ZNF516* *GALR1*	Rödkulla	24: 20.0-30.0	- Environmental adaptation and production traits Fertility and reproduction	(Shashank et al. 2021) (Melo et al. 2018)
ENSBTAG00000001724 ENSBTAG00000061558 ENSBTAG00000022887	*PTGR3*	Rödkulla	24: 30.0–40.0	- - Milk production	 (Pegolo et al. 2018)
ENSBTAG00000012191 ENSBTAG00000000957 ENSBTAG00000047706 ENSBTAG00000007422 ENSBTAG00000008886 ENSBTAG00000003081	*WWC2* *CDKN2AIP* - *TRAPPC11* *STOX2* *RWDD4*	Rödkulla	27: 14.0–15.0	Mastitis resistance - - Carcass fat Foetal growth -	(Zheng et al. 2025) (Martins et al. 2021) (Prakash et al. 2020)
ENSBTAG00000018863 ENSBTAG00000063178 ENSBTAG00000014704	*PRIMPOL* - *HELT* *SNX25* *ANKRD37*	Rödkulla	27: 15.0–16.0	- - - Immunoregulation -	 (Ruvinskiy et al. 2022)

ENSBTAG00000010351 ENSBTAG00000018453 ENSBTAG00000034430 ENSBTAG00000008682	*CFAP96* *TLR3*			- Pleiotropic	(Chen et al. 2020)
ENSBTAG00000034393 ENSBTAG00000021263 ENSBTAG00000009501 ENSBTAG00000003572 ENSBTAG00000066762	*FAM149A* *CYP4V2* *KLKB1* *F11*	Rödkulla	27: 16.0–17.0	- Meat quality - Blood coagulation	(Shi et al. 2024) (Krappmann et al. 2011)
ENSBTAG00000015188 ENSBTAG00000049772	*KLF6*	SRB	13: 44.0–45.0	Milk production	(Liu et al. 2023)
ENSBTAG00000004954	*TOX*	SRB	14: 25.0 26.0	Reproduction	(De Camargo et al. 2015)

5. Discussion

The aim of the study was to simulate genomic data from Swedish and international cattle breeds to identify runs of homozygosity (ROH) and homozygosity islands. A comparison of the simulated and empirical data, aimed to differentiate genomic patterns resulting from inbreeding and provide insight to how population size and historical structure influence ROH distribution over time.

The coverage and distribution of ROH across the different species demonstrated the accumulation of ROH in cattle. Overall, native Swedish breeds (Rödkulla, SRB and Fjällko) carry shorter segments of ROH throughout the genome, indicating earlier population bottlenecks. Holstein and Jersey, exhibit fewer but longer ROH segments, indicative of more recent inbreeding. In contrast, a comparison with native Swedish breeds revealed ROH islands both within and between populations, likely reflecting regions with varying levels of homozygosity that may harbour functionally important genes.

5.1 ROH distribution across the genome and ROH islands

The genome-wide ROH distribution reflects demographic differences between breeds. ROH coverage, defined as the number of individuals with ROH in a given genomic window, indicates how frequently regions are homozygous across a population. Swedish native breeds showed moderate ROH coverage, consistent with shared ancestry, while high-coverage regions were rare. The differences in ROH coverage by length class indicate that each breed has a distinct genomic history. The Swedish native breeds, such as Fjäll, generally exhibit shorter ROH, reflecting their higher genetic diversity. This finding aligns with the observations reported by Harish et al. (2024). In contrast, Holstein cattle showed evident peaks of high ROH coverage, likely indicating more recent selection. This is especially clear in the sequence data for the Holstein breed. The genomic patterns mirror effective population size (Ne). A large Ne that remains over time with only some modest inbreeding or selective pressure yields many short ROH, whereas a small Ne or recent inbreeding or bottlenecks produces long ROH (Purfield et al., 2012). Strong selection can also reduce Ne, as selection strategies can increase homozygosity around certain alleles and reduce genetic diversity in populations. The two Holstein populations showed relatively fewer ROH segments above the threshold value, although the ROH segments tended to be noticeably long.

The Swedish native breeds (Fjäll, Rödkulla and Swedish Red Polled) displayed many homozygous regions. However, they were generally shorter ROH segments

compared to the Holstein breed. This pattern indicates that the native breeds have a diverse ancestral background and possibly distant inbreeding as previously described by Harish et al. (2024). The ROH pattern displayed in Jersy suggests more recent inbreeding and genetic isolation considering the distribution and frequency of ROH regions (Zhang et al. 2015; Huson et al. 2020).

The results demonstrate a pattern that reflects the genomic impact of both selection and inbreeding. The Rödkulla and Fjäll breeds historically descended from relatively small, distinct Ne with limited crossbreeding (Upadhyay et al. 2019). The Swedish breeds as Rödkulla and Fjäll are known for their limited gene pools with few contribution from commercial breeds (Upadhyay et al. 2019). As a result, some ROH islands can be observed in the native breeds. The Jersey breed, originating from an isolated island population, has undergone intensive selective breeding in commercial programs, a history that is reflected in the distinctive distribution of ROH segments throughout its genome (Huson et al. 2020). The results indicate overall that long-term selection and small Ne have driven the accumulation of ROH in these simulated breeds consistent with the patterns seen in real cattle populations.

5.2 Implications of ROH

The long ROH segments observed in Holstein and Jersey breeds reflect recent inbreeding and reduced Ne, consistent with other studies presented by Purfield et al. (2012). In contrast, the majority of shorter ROH in native Swedish breeds suggests older shared ancestry and weaker recent inbreeding. Nevertheless, the identification of breed-specific ROH islands suggests that some regions may have become more homozygous due to drift or selection.

From a conservation perspective, these insights emphasize the value of native breeds (Harish et al. 2024). Conservation programs could aim to maintain this variability by incorporating genomic tools (such as ROH mapping) into breeding strategies. ROH mapping can provide insights into breeding histories and can highlight areas of the genome that have been affected by selection or genetic drift. This can be used in monitoring inbreeding levels, detecting harmful alleles, and identifying genomic regions associated with important traits. This would allow for more sustainable and healthy breeding practices (Peripolli et al. 2017; Ceballos et al. 2018). Efforts should also focus on minimizing further inbreeding, particularly in small populations by carefully managing selection and promoting genetic exchange.

ROH has been widely used as indicators inbreeding levels based on whole-genome sequencing (Zhang et al. 2015). In livestock, ROH have proven useful in identifying genomic regions under frequent selection, often appearing as ROH islands or hotspots. These regions typically exhibit elevated homozygosity around selected loci compared to the rest of the genome (Mastrangelo et al., 2018). Additionally, Purfield et al. (2012) found a strong association between extensive linkage disequilibrium (LD) and a high incidence of ROH. High LD often results from reduced recombination, drift, or selection, allowing long haplotypes to persist and form ROH. This suggests that regions with strong LD are more prone to ROH accumulation, particularly in populations affected by inbreeding or selection. In this study, such correlations could indicate genomic regions shaped by both demographic history and selection for traits like production or disease resistance.

5.3 Genes in ROH islands

The genes found within ROH islands or ROH hotspots further illustrate the effects of historical population bottlenecks, inbreeding, and selective breeding practices. Some of the ROH islands contain only one of a few annotated genes, or uncharacterised genes. Mastrangelo et al. (2018) suggest that this might reflect selection acting on uncharacterized regulatory regions or due to genetic drift, the fixation of non-coding DNA. Most hotspots and islands found for the Holstein breed were related to economic traits as milk production and feed conversion ratio. However, there were a significant amount of adaptation related genes as well. The genes found that were related to immune response might according to Mastrangelo et al. (2018) be caused by local adaptation which reflects natural or artificial selection for disease resistance. Some traits found in ROH regions for Rödkulla and relates to economically important traits, including milk content and meat quality. Which could be considered as a consequence of selection.

Inbreeding increases autozyosity and may trigger homozygous recessive alleles that express unfavourable phenotypes. High autozygosity resulting from inbreeding might also affect fertility and contribute to increased susceptibility to mastitis (Peripolli et al. 2017). Therefore, further investigation could be done with unfavourable phenotypes and their correlation to ROH islands.

Genes associated with production traits such as milk yield, feed efficiency, and carcass quality, were more frequently found in homozygous regions of the Holstein population. This reflects the influence of intensive selection. Swedish native breeds exhibited a higher number of genes related to environmental adaptation, although both native and international breeds showed ROH patterns associated with traits relevant to both production and environmental adaptation.

Overall, ROH regions tend to hold functional variants, including both advantageous and potentially deleterious alleles. Therefore, the genes identified within these ROH segments offer promising candidates for further investigation into traits such as fertility, robustness, health and environmental adaptability in Swedish cattle (Zhang et al. 2015; Ghoreishifar et al. 2020).

5.3.1 Candidate genes in ROH islands and ROH hotspots

Swedish native ROH islands contained *GART,* a gene implicated in fertility and embryonic viability. Fritz et al. (2013) specifically associated deleterious *GART* to prenatal death in Holstein, underscoring its relevance to reproduction and fertility. Another island found in native Swedish breeds contained the gene *DNAJC2* which has been liked to both the polled trait and heat stress response (Fernandes Júnior et al. 2020; Ghoreishifar et al. 2020). These findings align with previous work suggesting that ROH islands may serve as genomic signatures of selection, as they often harbour genes related to adaptive and economically valuable traits (Zhang et al. 2015). In the Swedish native breeds ROH islands and hotspots include genes involved in thermoregulation, metabolism and fertility consistent with adaptation to harsh climates. However, some genes related to milk production as *RSPS20* and *SDR16C5* were found in ROH islands in both Swedish native breeds and commercial breeds. This highlights their relevance in traits of economic importance (Ghoreishifar et al. 2020; Laodim et al. 2024). Rödkulla interestingly displayed hotspots in genes related to milk production *PTGR3* but also carcass quality *TRAPPC11* and mastitis resistance *WWC2* (Pegolo et al. 2018; Martin et al. 2023; Zheng et al. 2025). This highlights Rödkulla's genetic robustness and underscores the breed's relevance not only for conservation efforts but also as a valuable genetic resource for sustainable breeding programs.

5.4 Methodological discussion

The simulations captured key contrasts between cattle breeds. However, some methodological limitations should be noted. The filters applied on both SNP chip and sequence datasets could be adjusted to a more forgiving or strict setting to avoid false positives or false negatives for the datasets.

The settings used in PLINK for ROH detection influence the outcome and interpretation of the results. In this study, non-default parameters such as minimum ROH length, SNP density, SNP gap, and window thresholds were adjusted to balance sensitivity to both short and long ROH while minimizing false positives in the empirical data. However, these parameter selections are known to impact the number, length and genomic distribution of detected ROH

(Meyermans et al., 2020). For example, the minimum ROH length of 500 kb may have excluded biologically relevant short ROH, particularly in sequence data where resolution is higher (Ceballos et al. 2018). Allowing only one heterozygous SNP per window may prevent overestimation of ROH but could fragment ROH in regions with slight genotyping error. The choice of SNP density threshold (1 SNP per 50 kb) is also critical, as too strict a requirement may bias against detecting ROH in low-density genomic regions. Future work could explore the effect of varying these parameters systematically or use other methods as likelihood-based methods (e.g. BCFtools), which statistically model genotype data (Howrigan et al. 2011; Narasimhan et al. 2016).

The detection of ROH islands and ROH hotspots is sensitive to the choice of threshold parameters. The ROH islands were identified using a threshold of mean + 2 SD in window overlap counts. While this approach is common, it remains somewhat arbitrary and there is a risk of losing subtle signals of selection or drift (Meyermans et al. 2020). Stringent thresholds reduce false positives might underestimate real homozygosity signals, whereas more tolerant settings risk including technical noise or regions affected by structural variation (Ceballos et al. 2018; Nandolo et al. 2018). Furthermore, sequencing and SNP-chip data have different densities and distribution of genetic markers across the genome, which can impact ROH detection. Sequence data can identify shorter ROH that might be missed in SNP chip-based genotyping because sequencing data captures more variants, allowing for more precise identification of long and short ROH. SNP chip data typically include fewer and more widely spread markers, which can fail to detect short ROH, especially in regions with low SNP coverage (Purfield et al. 2012; Kim et al. 2017).

5.4.1 Filter criterion

The study applied several filters on both empirical sequence data and SNP chip data. This was to exclude SNPs with high missingness and poor quality. At the same time the different datasets were filtered to be comparable with each other and with the simulations.

The filters consequently removed rare alleles which could disproportionately affect the detection of ROH in small or inbred populations. Small populations often have informative variations in low-frequency alleles (Purfield et al. 2012). Similarly, applying a uniform MAF threshold across both high-and-lower density datasets could bias comparisons especially when SNP-chip data and high sequence data are analysed in parallel (Ceballos et al. 2018).

For future studies a methodological improvement could include testing a range of filtering thresholds, especially for heterozygous and missingness, to assess influence of ROH detection. To minimize platform effects or harmonizing the datasets the use of either high density SNP chip datasets or whole genome sequencing datasets. According to Meyermans et al (2020) testing multiple ROH detection parameter sets evaluate the robustness across different demographic scenarios. The same article suggests validating empirical data with either simulated or pedigree data to assess sensitivity and ROH detection under different filters.

5.4.2 Difference between simulation and empirical data

A notable observation displayed in the results is the mismatch in genomic inbreeding coefficient, or the proportion if genome that lies within ROH for each individual in the different populations, between empirical and simulated data in both Swedish and Holstein breeds. In the case of native Swedish breeds, the simulation-based demographic reconstructions, which estimate population history from observed linkage disequilibrium (LD), appeared to overestimate the impact of historical bottlenecks. This was evident in the higher genomic inbreeding coefficients in the simulated sequence data compared to the results of the empirical sequence data. As explained by Adepoju et al. (2024) such bottlenecks have occurred historically. However, the extent simulated demographic may not fully reflect the long-term resilience or occasional gene flow that preserved more genetic diversity in these populations than expected. One possible explanation is that the simulation underestimated historical Ne due to noisy LD estimates or short marker distances. This has been discussed in earlier evaluations of LD-based demographic inference, where it was shown that small sample sizes or biased marker density can affect Ne estimation at certain time depth (MacLeod et al. 2013). To improve realism, future simulations could increase historical Ne in Swedish breeds during key generations.

In the Holstein breed, particularly in the high-coverage sequence data, the simulations underestimated both the number and size of ROH. This discrepancy likely reflects the genetic diversity and admixture within international Holstein populations, highlighting a limitation of the simulation model. These factors capture recent intense inbreeding due to artificial insemination, short generation intervals, and heavy selection for production traits. These breeding strategies lead to rapid coalescence of haplotypes and extensive ROH, especially long tracts, which were more prominent in the empirical data than the simulated Holstein population. Earlier work by Zhang et al. (2015) and Peripolli et al. (2017) emphasizes how intense selection for milk yield in Holstein cattle has markedly decreased Ne and increased autozygosity. A potential solution would be to lower

Ne in the last few generations of the Holstein simulation to reflect this intensified selection and founder effects.

These mismatches demonstrate the dual importance of validating simulation parameters and interpreting ROH patterns considering both demographic and selective forces. The comparison between simulated and empirical data, as demonstrated here, offers a practical approach to calibrate demographic events.

Future research could include phenotypic association studies within ROH islands to identify candidate genes under selection. Combining ROH mapping with expression and genomic data could help determine whether homozygous regions harbour adaptive or deleterious variants. Furthermore, integrating simulations with genome-wide association data and environmental variables could clarify the functional impact of homozygosity and correlated candidate genes in different breeding contexts.

6. Conclusion

This study applied ROH analysis to both simulated and empirical genomic data from Swedish and international cattle breeds to examine the impacts of inbreeding and selection. The findings confirm ROH as a powerful tool for quantifying genomic inbreeding and identifying both recent and historical demographic events.

Swedish breeds, such as Fjäll and Rödkulla, exhibited numerous shorter ROH segments, indicating historical bottlenecks and isolation, whereas international breeds like Holstein and Jersey showed fewer but longer ROH, reflecting recent inbreeding driven by artificial selection. These patterns highlight how genetic drift and selection shape ROH islands, with certain adaptive genes potentially becoming more homozygous over time.

By comparing simulated and real genomic data, this study demonstrated that ROH analysis can effectively detect selection signatures and pinpoint key genomic regions. These insights are crucial for preserving endangered native breeds and enhancing genomic selection strategies for commercial cattle populations.

References

Abdullah, M., Rehman, M.S., Rehman, M.S.N., AlKahtane, A.A., Al-Hazani, T.M., Hassan, F. & Rehman, S. ur (2023). Genome-Wide Identification, Evolutionary and Mutational Analysis of the Buffalo Sox Gene Family. *Animals*, 13 (14), 2246. https://doi.org/10.3390/ani13142246

Adepoju, D., Ohlsson, J.I., Klingström, T., Rius-Vilarrasa, E., Johansson, A.M. & Johnsson, M. (2024). Population history of Swedish cattle breeds: estimates and model checking. bioRxiv. https://doi.org/10.1101/2024.10.03.616479

Adhikari, M. (2024). From Genes to Pastures: Unraveling Ancestry, Admixture, and Environmental Adaptations of Beef Cattle in Hawai‘i. https://hdl.handle.net/10125/108317 [2025-05-14]

Ayalew, W., Wu, X., Tarekegn, G.M., Sisay Tessema, T., Naboulsi, R., Van Damme, R., Bongcam-Rudloff, E., Edea, Z., Enquahone, S. & Yan, P. (2023). Whole-Genome Resequencing Reveals Selection Signatures of Abigar Cattle for Local Adaptation. *Animals*, 13 (20), 3269. https://doi.org/10.3390/ani13203269

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

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2023). Whole-genome analysis toolset. *Gigascience*, 4 (1), s13742-015-0047–8. https://doi.org/10.1186/s13742-015-0047-8

Chen, J., Wang, Y., Qi, X., Cheng, H., Chen, N., Ahmed, Z., Chen, Q., Lei, C. & Yang, X. (2023). Genome-wide analysis emancipates genomic diversity and signature of selection in Altay white-headed cattle of Xinjiang, China. *Frontiers in Genetics*, 14. https://doi.org/10.3389/fgene.2023.1144249

Chen, Y., Yang, Y., Li, C., Li, R., Xiao, H. & Chen, S. (2020). Genetic diversity of *TLR3* and *TLR8* genes among five Chinese native cattle breeds from southwest China. *Livestock Science*, 232, 103895. https://doi.org/10.1016/j.livsci.2019.103895

Dixit, S.P., Singh, S., Ganguly, I., Bhatia, A.K., Sharma, A., Kumar, N.A., Dang, A.K. & Jayakumar, S. (2020). Genome-Wide Runs of Homozygosity Revealed Selection Signatures in Bos indicus. *Frontiers in Genetics*, 11. https://doi.org/10.3389/fgene.2020.00092

Ebenezer Samuel King, J.P., Kumaresan, A., Talluri, T.R., Sinha, M.K., Raval, K., Nag, P., Karuthadurai, T. & Aranganathan, V. (2022). Genom-wide analysis identifies single nucleotide polymorphism variations and altered pathways associated with poor semen quality in breeding bulls. *Reproduction in Domestic Animals*, 57 (10), 1143–1155. https://doi.org/10.1111/rda.14185

Fernandes Júnior, G.A., de Oliveira, H.N., Carvalheiro, R., Cardoso, D.F., Fonseca, L.F.S., Ventura, R.V. & de Albuquerque, L.G. (2020). Whole-genome sequencing provides new insights into genetic mechanisms of tropical adaptation in Nellore (Bos primigenius indicus). *Scientific Reports*, 10 (1), 9412. https://doi.org/10.1038/s41598-020-66272-7

Fritz, S., Capitan, A., Djari, A., Rodriguez, S.C., Barbat, A., Baur, A., Grohs, C., Weiss, B., Boussaha, M., Esquerré, D., Klopp, C., Rocha, D. & Boichard, D. (2013). Detection of Haplotypes Associated with Prenatal Death in Dairy Cattle and Identification of Deleterious Mutations in GART, SHBG

and SLC37A2. *PLOS ONE*, 8 (6), e65550. https://doi.org/10.1371/journal.pone.0065550
Ghoreishifar, S.M., Eriksson, S., Johansson, A.M., Khansefid, M., Moghaddaszadeh-Ahrabi, S., Parna, N., Davoudi, P. & Javanmard, A. (2020). Signatures of selection reveal candidate genes involved in economic traits and cold acclimation in five Swedish cattle breeds. *Genetics Selection Evolution*, 52 (1), 52. https://doi.org/10.1186/s12711-020-00571-5
Gibson, J., Morton, N.E. & Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics*, 15 (5), 789–795. https://doi.org/10.1093/hmg/ddi493
Gorssen, W., Meyermans, R., Janssens, S. & Buys, N. (2021). A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species. *Genetics Selection Evolution*, 53 (1), 2. https://doi.org/10.1186/s12711-020-00599-7
Harish, A., Lopes Pinto, F.A., Eriksson, S. & Johansson, A.M. (2024). Genetic diversity and recent ancestry based on whole-genome sequencing of endangered Swedish cattle breeds. *BMC Genomics*, 25 (1), 89. https://doi.org/10.1186/s12864-024-09959-9
de las Heras-Saldana, S., Clark, S.A., Duijvesteijn, N., Gondro, C., van der Werf, J.H.J. & Chen, Y. (2019). Combining information from genome-wide association and multi-tissue gene expression studies to elucidate factors underlying genetic variation for residual feed intake in Australian Angus cattle. *BMC Genomics*, 20 (1), 939. https://doi.org/10.1186/s12864-019-6270-4
Hewett, A.M., Stoffel, M.A., Peters, L., Johnston, S.E. & Pemberton, J.M. (2023). Selection, recombination and population history effects on runs of homozygosity (ROH) in wild red deer (Cervus elaphus). *Heredity*, 130 (4), 242–250. https://doi.org/10.1038/s41437-023-00602-z
Howrigan, D.P., Simonson, M.A. & Keller, M.C. (2011). Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *BMC Genomics*, 12 (1), 460. https://doi.org/10.1186/1471-2164-12-460
Huson, H.J., Sonstegard, T.S., Godfrey, J., Hambrook, D., Wolfe, C., Wiggans, G., Blackburn, H. & VanTassell, C.P. (2020). A Genetic Investigation of Island Jersey Cattle, the Foundation of the Jersey Breed: Comparing Population Structure and Selection to Guernsey, Holstein, and United States Jersey Cattle. *Frontiers in Genetics*, 11. https://doi.org/10.3389/fgene.2020.00366
Johnston, D., Earley, B., McCabe, M.S., Kim, J., Taylor, J.F., Lemon, K., Duffy, C., McMenamy, M., Cosby, S.L. & Waters, S.M. (2021). Messenger RNA biomarkers of Bovine Respiratory Syncytial Virus infection in the whole blood of dairy calves. *Scientific Reports*, 11 (1), 9392. https://doi.org/10.1038/s41598-021-88878-1
Kelleher, J. & Lohse, K. (2020). Coalescent Simulation with msprime. I: Dutheil, J.Y. (red.) *Statistical Population Genomics*. Springer US. 191–230. https://doi.org/10.1007/978-1-0716-0199-0_9
Kim, S.-J., Ka, S., Ha, J.-W., Kim, J., Yoo, D., Kim, K., Lee, H.-K., Lim, D., Cho, S., Hanotte, O., Mwai, O.A., Dessie, T., Kemp, S., Oh, S.J. & Kim, H. (2017). Cattle genome-wide analysis reveals genetic signatures in trypanotolerant N'Dama. *BMC Genomics*, 18 (1), 371. https://doi.org/10.1186/s12864-017-3742-2
Kim, Y.-M., Seong, H.-S., Kim, Y.-S., Hong, J.-K., Sa, S.-J., Lee, J., Lee, J.-H., Cho, K.-H., Chung, W.-H., Choi, J.-W. & Cho, E.-S. (2022). Genome-

Wide Assessment of a Korean Composite Pig Breed, Woori-Heukdon. *Frontiers in Genetics*, 13. https://doi.org/10.3389/fgene.2022.779152
Krappmann, K., Weikard, R., Gerst, S., Wolf, C. & Kühn, Ch. (2011). A genetic predisposition for bovine neonatal pancytopenia is not due to mutations in coagulation factor XI. *The Veterinary Journal*, 190 (2), 225–229. https://doi.org/10.1016/j.tvjl.2010.10.007
Laodim, T., Koonawootrittriron, S., Elzo, M.A., Suwanasopee, T., Jattawa, D. & Sarakul, M. (2024). Genetic factors influencing milk and fat yields in tropically adapted dairy cattle: insights from quantitative trait loci analysis and gene associations. *Animal Bioscience*, 37 (4), 576–590. https://doi.org/10.5713/ab.23.0246
Liang, C., Wang ,Guohua, Raza ,Sayed Haidar Abbas, Wang ,Xiaoyu, Li ,Bingzhi, Zhang ,Wenzhen & and Zan, L. (2021). FAM13A promotes proliferation of bovine preadipocytes by targeting Hypoxia-Inducible factor-1 signaling pathway. *Adipocyte*, 10 (1), 546–557. https://doi.org/10.1080/21623945.2021.1986327
Lindholm-Perry, A.K., Kuehn, L.A., Smith, T.P.L., Ferrell, C.L., Jenkins, T.G., Freetly, H.C. & Snelling, W.M. (2012). A region on BTA14 that includes the positional candidate genes LYPLA1, XKR4 and TMEM68 is associated with feed intake and growth phenotypes in cattle. *Animal Genetics*, 43 (2), 216–219. https://doi.org/10.1111/j.1365-2052.2011.02232.x
Liu, Y., Han, B., Zheng, W., Peng, P., Yang, C., Jiang, G., Ma, Y., Li, J., Ni, J. & Sun, D. (2023). Identification of genetic associations and functional SNPs of bovine KLF6 gene on milk production traits in Chinese holstein. *BMC Genomic Data*, 24 (1), 72. https://doi.org/10.1186/s12863-023-01175-w
Liu, Y., Zhao, G., Lin, X., Zhang, J., Hou, G., Zhang, L., Liu, D., Li, Y., Li, J. & Xu, L. (2022). Genomic inbreeding and runs of homozygosity analysis of indigenous cattle populations in southern China. *PLoS ONE*, 17 (8), e0271718. https://doi.org/10.1371/journal.pone.0271718
Macharia, J.K., Kim, J., Kim, M., Cho, E., Munyaneza, J.P. & Lee, J.H. (2024). Characterisation of runs of homozygosity and inbreeding coefficients in the red-brown Korean native chickens. *Animal Bioscience*, 37 (8), 1355–1366. https://doi.org/10.5713/ab.23.0514
MacLeod, I.M., Larkin, D.M., Lewin, H.A., Hayes, B.J. & Goddard, M.E. (2013). Inferring Demography from Runs of Homozygosity in Whole-Genome Sequence, with Correction for Sequence Errors. *Molecular Biology and Evolution*, 30 (9), 2209–2223. https://doi.org/10.1093/molbev/mst125
Makanjuola, B.O., Miglior, F., Abdalla, E.A., Maltecca, C., Schenkel, F.S. & Baes, C.F. (2020). Effect of genomic selection on rate of inbreeding and coancestry and effective population size of Holstein and Jersey cattle populations. *Journal of Dairy Science*, 103 (6), 5183–5199. https://doi.org/10.3168/jds.2019-18013
Martin, C.A., Sheppard, E.C., Illera, J.C., Suh, A., Nadachowska-Brzyska, K., Spurgin, L.G. & Richardson, D.S. (2023). Runs of homozygosity reveal past bottlenecks and contemporary inbreeding across diverging populations of an island-colonizing bird. *Molecular Ecology*, 32 (8), 1972–1989. https://doi.org/10.1111/mec.16865
Martins, R., Brito, L.F., Machado, P.C., Pinto, L.F.B., Silva, M.R., Schenkel, F.S. & Pedrosa, V.B. (2021). Genome-wide association study and pathway analysis for carcass fatness in Nellore cattle measured by ultrasound. *Animal Genetics*, 52 (5), 730–733. https://doi.org/10.1111/age.13129
Mastrangelo, S., Sardina, M.T., Tolone, M., Di Gerlando, R., Sutera, A.M., Fontanesi, L. & Portolano, B. (2018). Genome-wide identification of runs

of homozygosity islands and associated genes in local dairy cattle breeds. *Animal*, 12 (12), 2480–2488. https://doi.org/10.1017/S1751731118000629
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 (1), 94. https://doi.org/10.1186/s12864-020-6463-x
Michot, P., Chahory, S., Marete, A., Grohs, C., Dagios, D., Donzel, E., Aboukadiri, A., Deloche, M.-C., Allais-Bonnet, A., Chambrial, M., Barbey, S., Genestout, L., Boussaha, M., Danchin-Burge, C., Fritz, S., Boichard, D. & Capitan, A. (2016). A reverse genetic approach identifies an ancestral frameshift mutation in RP1 causing recessive progressive retinal degeneration in European cattle breeds. *Genetics Selection Evolution*, 48 (1), 56. https://doi.org/10.1186/s12711-016-0232-y
Mota, R. r., Guimarães, S. e. f., Fortes, M. r. s., Hayes, B., Silva, F. f., Verardo, L. l., Kelly, M. j., de Campos, C. f., Guimarães, J. d., Wenceslau, R. r., Penitente-Filho, J. m., Garcia, J. f. & Moore, S. (2017). Genome-wide association study and annotating candidate gene networks affecting age at first calving in Nellore cattle. *Journal of Animal Breeding and Genetics*, 134 (6), 484–492. https://doi.org/10.1111/jbg.12299
Nandolo, W., Utsunomiya, Y.T., Mészáros, G., Wurzinger, M., Khayadzadeh, N., Torrecilha, R.B.P., Mulindwa, H.A., Gondwe, T.N., Waldmann, P., Ferenčaković, M., Garcia, J.F., Rosen, B.D., Bickhart, D., van Tassell, C.P., Curik, I. & Sölkner, J. (2018). Misidentification of runs of homozygosity islands in cattle caused by interference with copy number variation or large intermarker distances. *Genetics Selection Evolution*, 50 (1), 43. https://doi.org/10.1186/s12711-018-0414-x
Nani, J.P. & Peñagaricano, F. (2020). Whole-genome homozygosity mapping reveals candidate regions affecting bull fertility in US Holstein cattle. *BMC Genomics*, 21 (1), 338. https://doi.org/10.1186/s12864-020-6758-y
Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C. & Durbin, R. (2016). BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*, 32 (11), 1749–1751. https://doi.org/10.1093/bioinformatics/btw044
Novo, L.C., Cavani, L., Pinedo, P., Melendez, P. & Peñagaricano, F. (2022). Genomic Analysis of Visceral Fat Accumulation in Holstein Cows. *Frontiers in Genetics*, 12. https://doi.org/10.3389/fgene.2021.803216
Paredes-Sánchez, F.A., Sifuentes-Rincón, A.M., Lara-Ramírez, E.E., Casas, E., Rodríguez-Almeida, F.A., Herrera-Mayorga, E.V., Randel, R.D., Paredes-Sánchez, F.A., Sifuentes-Rincón, A.M., Lara-Ramírez, E.E., Casas, E., Rodríguez-Almeida, F.A., Herrera-Mayorga, E.V. & Randel, R.D. (2023). Identification of candidate genes and SNPs related to cattle temperament using a GWAS analysis coupled with an interacting network analysis. *Revista mexicana de ciencias pecuarias*, 14 (1), 1–22. https://doi.org/10.22319/rmcp.v14i1.6077
Pegolo, S., Mach, N., Ramayo-Caldas, Y., Schiavon, S., Bittante, G. & Cecchinato, A. (2018). Integration of GWAS, pathway and network analyses reveals novel mechanistic insights into the synthesis of milk proteins in dairy cows. *Scientific Reports*, 8 (1), 566. https://doi.org/10.1038/s41598-017-18916-4
Peripolli, E., Munari, D.P., Silva, M.V.G.B., Lima, A.L.F., Irgang, R. & Baldi, F. (2017). Runs of homozygosity: current knowledge and applications in livestock. *Animal Genetics*, 48 (3), 255–271. https://doi.org/10.1111/age.12526
Prakash, C., Gaur, G.K., D R, P. & Sahoo, N.R. (2020). Distribution analysis of single nucleotide polymorphisms related to feet and legs and their

association with lameness in Vrindavani cattle. *Tropical Animal Health and Production*, 52 (2), 851–858. https://doi.org/10.1007/s11250-019-02076-w
Purfield, D.C., Berry, D.P., McParland, S. & Bradley, D.G. (2012). Runs of homozygosity and population history in cattle. *BMC Genetics*, 13 (1), 70. https://doi.org/10.1186/1471-2156-13-70
Rstudio Team (2023). *RStudio: Integrated Development Environment for R (Version 2023.06.1*. https://www.posit.co/ [2025-04-27]
Ruiz-De-La-Cruz, G., Welsh, T.H., Randel, R.D. & Sifuentes-Rincón, A.M. (2024). A Comprehensive Systematic Review Coupled with an Interacting Network Analysis Identified Candidate Genes and Biological Pathways Related to Bovine Temperament. *Genes*, 15 (8), 981. https://doi.org/10.3390/genes15080981
Ruvinskiy, D., Igoshin, A., Yurchenko, A., Ilina, A.V. & Larkin, D.M. (2022). Resequencing the Yaroslavl cattle genomes reveals signatures of selection and a rare haplotype on BTA28 likely to be related to breed phenotypes. *Animal Genetics*, 53 (5), 680–684. https://doi.org/10.1111/age.13230
Shashank, C.G., Silpa, M.V., Ezhil Vadhana, P., Devaraj, C., Tiwari, M. & Sejian, V. (2021). Climate Change and Livestock Production: Significance of Studying Multiple Stressors Impact in Cattle in the Changing Climate Scenario. I: Sejian, V., Chauhan, S.S., Devaraj, C., Malik, P.K., & Bhatta, R. (red.) *Climate Change and Livestock Production: Recent Advances and Future Perspectives*. Springer. 59–72. https://doi.org/10.1007/978-981-16-9836-1_6
Shi, M., Huang, L., Meng, S., Wang, H., Zhang, J., Miao, Z. & Li, Z. (2024). Identification of several lncRNA-mRNA pairs associated with marbling trait between Nanyang and Angus cattle. *BMC Genomics*, 25 (1), 696. https://doi.org/10.1186/s12864-024-10590-x
Silva, D.B.S., Fonseca, L.F.S., Pinheiro, D.G., Magalhães, A.F.B., Muniz, M.M.M., Ferro, J.A., Baldi, F., Chardulo, L.A.L., Schnabel, R.D., Taylor, J.F. & Albuquerque, L.G. (2020). Spliced genes in muscle from Nelore Cattle and their association with carcass and meat quality. *Scientific Reports*, 10 (1), 14701. https://doi.org/10.1038/s41598-020-71783-4
Singh, A., Kumar, A., Gondro, C., Pandey, A.K., Dutt, T. & Mishra, B.P. (2022). Genome Wide Scan to Identify Potential Genomic Regions Associated With Milk Protein and Minerals in Vrindavani Cattle. *Frontiers in Veterinary Science*, 9. https://doi.org/10.3389/fvets.2022.760364
Sumreddee, P., Toghiani, S., Hay, E.H., Roberts, A., Aggrey, S.E. & Rekaya, R. (2020). Runs of homozygosity and analysis of inbreeding depression. *Journal of Animal Science*, 98 (12), skaa361. https://doi.org/10.1093/jas/skaa361
Sun, T., Pei, S., Liu, Y., Hanif, Q., Xu, H., Chen, N., Lei, C. & Yue, X. (2023). Whole genome sequencing of simmental cattle for SNP and CNV discovery. *BMC Genomics*, 24 (1), 179. https://doi.org/10.1186/s12864-023-09248-x
Tripathy, K., Sodhi, M., Kataria, R.S., Chopra, M. & Mukesh, M. (2021). In Silico Analysis of HSP70 Gene Family in Bovine Genome. *Biochemical Genetics*, 59 (1), 134–158. https://doi.org/10.1007/s10528-020-09994-7
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), 56. https://doi.org/10.1186/s12711-019-0496-0
Wang, L., Gao, P., Li, C., Liu, Q., Yao, Z., Li, Y., Zhang, X., Sun, J., Simintiras, C., Welborn, M., McMillin, K., Oprescu, S., Kuang, S. & Fu, X. (2023). A

single-cell atlas of bovine skeletal muscle reveals mechanisms regulating intramuscular adipogenesis and fibrogenesis. *Journal of Cachexia, Sarcopenia and Muscle*, 14 (5), 2152–2167. https://doi.org/10.1002/jcsm.13292

Wang, P., Li, X., Zhu, Y., Wei, J., Zhang, C., Kong, Q., Nie, X., Zhang, Q. & Wang, Z. (2022). Genome-wide association analysis of milk production, somatic cell score, and body conformation traits in Holstein cows. *Frontiers in Veterinary Science*, 9. https://doi.org/10.3389/fvets.2022.932034

Xing, Y., Mi, S., Dari, G., Zhang, Z., Chen, S. & Yu, Y. (2025). Ferroptosis-Related Genes as Molecular Markers in Bovine Mammary Epithelial Cells Challenged with Staphylococcus aureus. *International Journal of Molecular Sciences*, 26 (6), 2506. https://doi.org/10.3390/ijms26062506

Yurchenko, A.A., Daetwyler, H.D., Yudin, N., Schnabel, R.D., Vander Jagt, C.J., Soloshenko, V., Lhasaranov, B., Popov, R., Taylor, J.F. & Larkin, D.M. (2018). Scans for signatures of selection in Russian cattle breed genomes reveal new candidate genes for environmental adaptation and acclimation. *Scientific Reports*, 8 (1), 12984. https://doi.org/10.1038/s41598-018-31304-w

Zhang, Q., Guldbrandtsen, B., Bosse, M., Lund, M.S. & Sahana, G. (2015). Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC Genomics*, 16 (1), 542. https://doi.org/10.1186/s12864-015-1715-x

Zheng, W., Wu, P., Zhu, M., Ullah, Y., Zhao, Z., Cao, S., Li, G., Ou, S., He, K. & Xu, Y. (2025). Establishment and Validation of a Method for the Identification of Recessive Mastitis Resistance Genes in Dairy Cows. *Genes*, 16 (5), 485. https://doi.org/10.3390/genes16050485

Popular science summary

Cattle are typically bred for improved production traits, but such selection practices also impact genetic diversity. This study explores the manifestation of these effects across various cattle breeds, emphasizing Swedish native breeds, Rödkulla, Fjäll, and SRB, while comparing them to internationally selected breeds such as Holstein and Jersey.

Runs of homozygosity (ROH) are extensive, continuous segments of homozygous genotypes that arise when individuals inherit identical haplotypes from both parents. ROH patterns serve as valuable tools for reconstructing population history and identifying genomic markers associated with genetic drift, inbreeding and selection. The quantity, size, and genomic positioning of ROH offer insights into both recent and historical demographic events.

This study combined empirical genotyping data with simulated demographic models to explore breed-specific histories and selection shape ROH patterns. The results revealed that Swedish native breeds exhibited many short ROH, consistent with historical bottlenecks and moderate inbreeding, but also some genetic diversity. In contrast, Holstein and Jersey showed fewer but much longer ROH segments, an indication of modern inbreeding and intensive selection.

Regions of the genome where ROH occurred frequently across populations, known as ROH islands, were also identified. In native breeds, genes associated with traits like heat tolerance, metabolism, and disease resistance were commonly observed in ROH islands, suggesting adaptation to local environments. International breeds showed ROH islands containing genes related to milk production, growth, and reproductive traits.

Overall, this study highlights the value of ROH analysis as a genomic tool for assessing the impact of breeding strategies. The findings emphasize the importance of conserving genetic diversity in native breeds and demonstrate how genomic insights can support more sustainable and resilient cattle breeding programs.

Tack

Först och främst vill jag uttrycka min djupa tacksamhet till min handledare, Martin Johnsson, vars outtröttliga tålamod och klara förklaringar hjälpt mig att färdigställa arbetet. Jag är innerligt tacksam för varje gång han tog sig tid att svara på mina frågor oavsett hur små, upprepande eller förvirrade de än kan ha varit.

Jag vill även rikta ett stort tack till alla som tagit sig tid att läsa min text och bidragit med värdefull feedback och konstruktiva tips. Stort tack till min familj som stöttat mig under hela arbetet. Ett särskilt tack går till min vän Anna, vars ständiga stöd och engagemang burit mig genom tvivel, sena nätter och långa telefonsamtal. Din uppmuntran har betytt mer än du anar.

Tack!

Publishing and archiving

Approved students' theses at SLU can be published online. As a student you own the copyright to your work and in such cases, you need to approve the publication. In connection with your approval of publication, SLU will process your personal data (name) to make the work searchable on the internet. You can revoke your consent at any time by contacting the library.

Even if you choose not to publish the work or if you revoke your approval, the thesis will be archived digitally according to archive legislation.

You will find links to SLU's publication agreement and SLU's processing of personal data and your rights on this page:

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

☒ YES, I, Kajsa Nyström, have read and agree to the agreement for publication and the personal data processing that takes place in connection with this

☐ NO, I/we do not give my/our permission to publish the full text of this work. However, the work will be uploaded for archiving and the metadata and summary will be visible and searchable.