Inbreeding and selection on the X chromosome in domestic animal populations

Shihabi, Mario

University of Zagreb

FACULTY OF AGRICULTURE

Mario Shihabi

INBREEDING AND SELECTION ON THE X CHROMOSOME IN DOMESTIC ANIMAL POPULATIONS

DOCTORAL DISSERTATION

Zagreb, 2024

Sveučilište u Zagrebu

AGRONOMSKI FAKULTET

Mario Shihabi

INBRIDING I SELEKCIJA NA X KROMOSOMU U POPULACIJAMA DOMAĆIH ŽIVOTINJA

DOKTORSKI RAD

Zagreb, 2024.

University of Zagreb

FACULTY OF AGRICULTURE

Mario Shihabi

INBREEDING AND SELECTION ON THE X CHROMOSOME IN DOMESTIC ANIMAL POPULATIONS

DOCTORAL DISSERTATION

Supervisor: Prof. Ino Čurik, PhD.H

Zagreb, 2024

Sveučilište u Zagrebu

AGRONOMSKI FAKULTET

Mario Shihabi

INBRIDING I SELEKCIJA NA X KROMOSOMU U POPULACIJAMA DOMAĆIH ŽIVOTINJA

DOKTORSKI RAD

Mentor: Prof. dr. sc. dr. h. c. Ino Čurik

Zagreb, 2024.

Bibliography data:

- **Scientific area:** Biotechnical sciences
- **Scientific field:** Agriculture
- **Scientific branch:** Genetics and breeding of plants, animals and microorganisms
- **Institution:** University of Zagreb Faculty of Agriculture, Department of Animal Science
- **Supervisor of doctoral thesis:** Prof. Ino Čurik, PhD.H
- **Number of pages:** 116
- **Number of figures:** 27
- **Number of tables:** 7
- **Number of appendices:** 13
- **Number of references:** 176
- **Doctoral thesis defence date:** 17.07.2024.
- **Committee for doctoral thesis defence at the University of Zagreb Faculty of Agriculture:**
  1. Prof. Vlatka Čubrić Čurik, PhD (University of Zagreb Faculty of Agriculture)
  2. Assoc. Prof. Maja Ferenčaković, PhD (University of Zagreb Faculty of Agriculture)
  3. Prof. Luboš Vostry, PhD (Czech University of Life Sciences Prague)

This thesis has been stored in:

National and University Library in Zagreb, Ulica Hrvatske bratske zajednice 4 p.p 550, 10 000 Zagreb,
Library of the University of Zagreb Faculty of Agriculture, Svetošimunska cesta 25, 10 000 Zagreb.

The topic of the dissertation was approved at the meeting of the Faculty Council of the University of Zagreb Faculty of Agriculture on 07.03.2023 and approved at the meeting of the Senate of the University of Zagreb on 20.06.2023.

UNIVERSITY OF ZAGREB

FACULTY OF AGRICULTURE

DECLARATION OF ORIGINALITY

I, **Mario Shihabi**, declare that I have composed solely by myself the thesis titled:

INBREEDING AND SELECTION ON THE X CHROMOSOME IN DOMESTIC ANIMAL POPULATIONS

With my signature I confirm that:

- I am the sole author of this thesis;
- This thesis is an original report of my research, and references have been provided for all supporting literature and resources;
- I am familiar with the provisions of the Code of Ethics of the University of Zagreb (Article 19).

Zagreb, 17.07.2024.

Mario Shihabi

Evaluation of doctoral thesis

Doctoral thesis was evaluated by PhD defense committee comprised of:

1. Prof. Vlatka Čubrić Čurik, PhD
   University of Zagreb Faculty of Agriculture
2. Assoc. Prof. Maja Ferenčaković, PhD
   University of Zagreb Faculty of Agriculture
3. Prof. Luboš Vostry, PhD
   Czech University of Life Sciences Prague

Doctoral thesis was defended at the University of Zagreb Faculty of Agriculture on 17.07.2024. in front of the PhD defense committee comprised of:

1. Prof. Vlatka Čubrić Čurik, PhD _____
   University of Zagreb Faculty of Agriculture
2. Assoc. Prof. Maja Ferenčaković, PhD _____
   University of Zagreb Faculty of Agriculture
3. Prof. Luboš Vostry, PhD _____
   Czech University of Life Sciences Prague

Supervisor information

Prof. Ino Čurik, PhD.H

Ino Čurik was born on March 25, 1966, in Zagreb. He is married and the father of Tin Josip and Laura Ina. He graduated in 1991 with a degree in "Animal Science" from the University of Zagreb Faculty of Agriculture. During his studies, he received the Rector's Award and the May Award. He earned his Master's degree in 1995 in "Biomathematics" from the University of Zagreb Faculty of Science. He obtained his PhD in 2000 from the University of Zagreb Faculty of Agriculture, defending his dissertation titled "Inbreeding in Empirical and Theoretical Populations." He has successfully supervised ten defended doctoral theses. Since 2016, he has been a full professor with tenure at the Faculty of Agriculture, where he teaches (as the lead instructor) "Principles of Domestic Animal Breeding," "Quantitative Genetics," and "Population Genetics." He is the author or co-author of over 90 scientific papers (all indexed in the WOS Core Collection, cited over 2200 times with an h-index of 25), many published in prestigious journals (Nature Genetics, PLoS Genetics, Heredity, Animal Genetics, BMC Genomics, Conservation Genetics, GSE, Frontiers in Genetics, Journal of Dairy Science, Journal of Animal Science, Current Biology, Evolutionary Applications, MBE, Journal of Animal Science and Biotechnology, etc.). He is an associate editor of Frontiers in Genetics and a member of the editorial boards of Livestock Science, Genes, World Rabbit Science, etc. Since 2007, he has been a member of the coordinating committee of the International Symposium "Animal Science Days." From 2016 to 2019, he led the project "MendTheGap," the highest-ranked project out of 546 accepted in the H2020-TWINN-2015 call. He received the 2013 "Annual State Award for Science" (Republic of Croatia) in biotechnical sciences and the "Charter" for contributions to scientific, teaching, and professional work, awarded by the University of Zagreb Faculty of Agriculture, in 2013. The Biotechnical Faculty of the University of Ljubljana awarded him in 2020 for his outstanding research and teaching contributions. He was awarded an honorary doctorate by the University of Natural Resources and Life Sciences, Vienna, Austria, in 2022. The University of Zagreb awarded him the Praemium Biotechnicum in 2023 for scientific and educational achievements, promotion of biotechnical disciplines, knowledge transfer, and education of young scientists and experts in the biotechnical field. In his youth, he was a successful chess player, being the junior vice-champion of Croatia in 1984, vice-champion at the 1983 International Chess Festival for Juniors (Biel-Bienne, Switzerland), and junior champion of Croatia in 1981.

Acknowledgments

I would like to express my sincere gratitude to the following individuals:

To my mentor, Professor Ino Čurik, for guiding me towards genetics during my studies and life. Thank you for your time, professional guidance, trust, patience, advice and suggestions through various protocols and development of this thesis, as well as throughout the entire PhD studies, and even well before.

To Professor Vlatka Čubrić Čurik, for providing me with opportunities, patience, knowledge, help and constructive advice during my studies.

To Associate Professor Maja Ferenčaković for steering me through numerous situations during my studies (whenever it was necessary, which was often the case).

To Assistant Professor Vladimir Brajković and Ivana Držaić, PhD, for all the coffee breaks and their always available help and friendly advice at all times.

To Professor Luboš Vostry for your perseverance and human simplicity, which served as an example to me. You are a gentleman in every sense of the word.

To Yuri Utsonomiya, PhD, for providing the genotypes of the Nellore cattle breed and for your constructive advice, and to Professor Donagh Berry, Professor Gregor Gorjanc and Professor Tom Druet for their constructive advice during the preparation of this doctoral dissertation.

To Professor Ante Ivanković, Professor Darija Lemić and Professor Ivica Kos for numerous life advices in both professional and personal matters.

To Mrva, the good spirit of the Faculty of Agriculture, for countless times of saving me when I did not know what, where or how.

To all my friends who made my student days fulfilling. To all the members of KSA and UPSAF for the gatherings, parties and all the agricultural discussions during my studies, especially to Sven, Dino and Luka for involving me in it. Because of you, I grew to love this institution. Thanks to Antun and Skejo for numerous constructive agronomic and biological discussions.

To my brothers Amir and Fuad, for their support during my studies. To my cat, Cico, for enduring my high school and university education with me.

Finally, to my mother, Anka, and my father, Kamel, for their support, patience and trust during my education, especially when it was most needed. This dissertation is as much yours as it is mine, because without you, it would not have been possible.

I thank my class teacher for this message from my youth, which I recall more frequently as I get older and which somehow becomes the final thought of this dissertation:

“Let your smile change the world, don't let the world change your smile.”

The research presented in this dissertation was conducted within the framework of the Croatian Science Foundation project IP-2018-01-8708; "Application of NGS methods in the assessment of genomic variability of ruminants – ANAGRAMS".

Abstract

In domestic animals, genomic studies of inbreeding and selection have mainly focused on autosomes, neglecting the X chromosome. This neglect is significant as the X chromosome influences many important traits and has unique characteristics that may lead to more pronounced effects of inbreeding and positive selection. In addition, in its small part, PAR, inbreeding avoidance might occur. Furthermore, hemizygous haplotypes on nonPAR in males clearly reveal haplotype structure, enabling detection of positive selection signals and investigation of phylogenetic relationships. Therefore, the main objectives of this dissertation were to evaluate and compare *F* on the X chromosome and autosomes in domestic animal populations with special focus on PAR and to develop a new method for identifying positive selection signals based on the difference in haplotype richness of nonPAR in males. Each population used was represented by high density Illumina genotypes with an adequate number of both males and females. Five different inbreeding coefficients were used on two distinct populations for cattle (Croatian cattle breeds and Nellore), dogs (Labrador Retriever and Patagonian Sheepdog) and sheep (Croatian sheep breeds and Soay). Conversely, a new method called Haplotype Richness Drop (HRiD) was established and tested alongside classical methods (eROHi, iHS, and nSL) in metapopulation of native Croatian sheep breeds. Each identified signal underwent functional characterization, gene annotation and MJN. Higher inbreeding was found on the X chromosome compared to autosomes in all populations using F_{ROH_SVS} and $F_{ROH\_RZooROH}$ (most reliable), while no differences were found using F_{LH1}, F_{VR1} and F_{YA2}, with greater variability observed using all five coefficients. No difference in *F* between sexes at PAR or compared to autosomes was found. Using HRiD, four signals were identified and consistently validated, with the same most significant signal across all four methods (from 13.04 to 13.62 Mb). Overall, 14 positive selection signals (12 regions) were identified with 34 genes, with high concordance (86%) with other studies of sheep. The results demonstrate the high accuracy and reliability of HRiD and show that HRiD can be used comprehensively or in scenarios where only male genotypes are available, which is common in livestock where genomic breeding values are predominantly performed for males. Moreover, MJN is shown to provide useful additional information when analysing haplotypes identified as selection signals (derived versus ancestral haplotype or control for population structure-induced disorders). In general, the results emphasize the importance of including the X chromosome in inbreeding estimation and selection identification in domestic animal populations, while the new HRiD method opens up new possibilities in identifying signals using heterogametic sex haplotypes.

Key words: X chromosome, inbreeding, selection, domestic animal populations

Prošireni sažetak

Inbriding i selekcija temeljni su genetski procesi koji oblikuju genetsku osnovu populacija, duboko utječući na njihovu genetsku raznolikost i definirajuća svojstva. Kod domaćih životinja ti su procesi posebno izraženi zbog težnje za većom produktivnošću. Razvoj molekularnih markera, posebice SNP-ova, doveo je do stvaranja različitih metoda za procjenu inbridinga i identifikaciju pozitivnih selekcijskih signala, koji se obično primjenjuju i na populacije divljih i domaćih životinja. Međutim, istraživanja genomskog inbridinga i selekcije primarno su usredotočena na autosome, uvelike zanemarujući X kromosom unatoč njegovoj značajnoj ulozi u utjecaju na ekonomski i evolucijski važna svojstva. Važnost X kromosoma dodatno je istaknuta njegovim posebnim karakteristikama u usporedbi s autosomima, koji mogu dovesti do izraženijih učinaka inbridinga i pozitivne selekcije. Nasuprot tome, PAR, mali dio X kromosoma gdje se još uvijek događa rekombinacija između mužjaka i ženke, ako je relativno velik (kod domaćih životinja), sugerira prisutnost sila koje pogoduju rekombinaciji, kao što je izbjegavanje inbridinga. Nadalje, hemizigotni haplotipovi na nonPAR području kod muških jedinki daju jasnu sliku strukture haplotipa, omogućujući razvoj novih metoda za otkrivanje pozitivnih selekcijskih signala i proučavanje filogenetskih odnosa. Stoga su glavni ciljevi ove disertacije bili procijeniti i usporediti razine genomskog inbridinga na X kromosomu i autosomima u populacijama domaćih životinja s posebnim osvrtom na PAR, te uspostaviti novu metodu za identifikaciju pozitivnih selekcijskih signala na temelju razlike u bogatstvu haplotipova nonPAR područja kod muških jedinki. Svaka korištena populacija bila je predstavljena visoko kvalitetnim (dokazano nakon QC-a) Illumina SNP čip genotipovima visoke gustoće s odgovarajućim brojem muških i ženskih jedinki. Korišteno je pet različitih inbriding koeficijenata (F_{ROH_SVS}, $F_{ROH\_RZooROH}$, F_{LH1}, F_{VR1} i F_{YA2}) na dvije različite populacije goveda (metapopulacija autohtonih hrvatskih pasmina goveda i nellore), pasa (labrador retriver i patagonijski ovčar) i ovaca (metapopulacija autohtonih hrvatskih pasmina ovaca i soay). Nadalje, uspostavljena je nova metoda nazvana „Haplotype Richness Drop” (HRiD) i testirana u kombinaciji s klasičnim metodama (eROHi, iHS i nSL) u metapopulaciji autohtonih hrvatskih pasmina ovaca. U svakom identificiranom signalu provedena je detaljna funkcionalna karakterizacija, anotacija gena i filogenetska analiza (MJN). Kao osnova za obje analize, PAR je lokaliziran u svakoj vrsti, u rasponu od 133,20 do 139,00 Mb u goveda, od 0,00 do 6,59 Mb u pasa i od 0,00 do 7,04 Mb u ovaca, što je u skladu s njihovim referentnim genomima. Na X kromosomu utvrđen je veći inbriding u odnosu na autosome u svim populacijama upotrebom koeficijenata F_{ROH_SVS} i $F_{ROH\_RZooROH}$ (najpouzdaniji koeficijenti), dok nisu pronađene razlike upotrebom koeficijenata F_{LH1}, F_{VR1} i F_{YA2}. Pouzdanost prve skupine koeficijenata (F_{ROH_SVS} i $F_{ROH\_RZooROH}$) pripisuje se njihovoj

osnovi u IBD statusu, za razliku od druge skupine koja se temelji na IBS statusu alela. Veća varijabilnost u inbridingu pronađena je na X kromosomu u usporedbi s autosomima upotrebom svih pet koeficijenata u svim populacijama, sa specifičnim obrascima uočenim na X kromosomu. Ova povećana varijabilnost i veći inbriding na X kromosomu mogu biti posljedica njegovih jedinstvenih obrazaca nasljeđivanja i smanjene efektivne veličine populacije. Nikakva razlika u inbridingu između spolova u PAR-u ili prilikom njihove usporedbe s autosomnim vrijednostima nije pronađena ni u jednoj populaciji korištenjem bilo kojeg koeficijenta, što ukazuje na njihovu sličnu genetsku dinamiku. U metapopulaciji izvornih hrvatskih pasmina ovaca identificirano je ukupno 14 pozitivnih selekcijskih signala (u 12 regija) koji obuhvaćaju ukupno 34 kandidat gena, uz visoku podudarnost (86%) s drugim studijama na ovcama. Novom HRiD metodom identificirana su četiri signala: od 13,04 do 13,62 Mb, od 56,64 do 58,09 Mb, od 73,57 do 74,54 Mb (podijeljeno u dva signala), te od 115,30 do 115,73 Mb. Najznačajniji signal je također bio najznačajniji prema drugim metodama, a sva četiri identificirana signala dosljedno su validirana. Filogenetski odnosi unutar ovih signala, analizirani korištenjem MJN-ova, pokazuju da su najčešći haplotipovi unutar signala HRiD_w1 i HRiD_w3,4 haplotipovi predaka koji su podložni pozitivnoj selekciji, dok su derivirani haplotipovi favorizirani u signalima HRiD_w2 i HRiD_w5. Rezultati pokazuju visoku točnost i pouzdanost HRiD metode te da se HRiD može učinkovito koristiti uz eROHi, iHS i nSL metode ili u scenarijima u kojima su dostupni samo muški genotipovi, što je uobičajeno u stočarstvu gdje se pretežno provode procjene genomskih uzgojnih vrijednosti kod muških jedinki. Štoviše, pokazalo se da filogenetske analize mogu pružiti korisne dodatne informacije u analizi haplotipova identificiranih kao selekcijski signali, bilo u smislu povijesnih ili deriviranih statusa favoriziranih haplotipova ili kontroliranjem potencijalnih problema uzrokovanih strukturom populacije koji se mogu pojaviti pri analizi metapopulacija. Općenito, rezultati naglašavaju važnost uključivanja X kromosoma u procjenu inbridinga i identifikaciju selekcije u populacijama domaćih životinja, dok nova HRiD metoda otvara nove mogućnosti u identificiranju pozitivnih signala selekcije korištenjem heterogametskih spolnih haplotipova.

Ključne riječi: X kromosom, inbriding, selekcija, populacije domaćih životinja

TABLE OF CONTENTS

List of abbreviations

Abbreviation	Meaning
ABCB7	gene ATP Binding Cassette Subfamily B Member 7
AMOT	gene Angiomotin
AP1S2	gene Adaptor Related Protein Complex 1 Subunit Sigma 2
AR	gene Androgen Receptor
ARX	gene Aristaless Related Homeobox
ATRX	gene ATRX chromatin remodeler
BLAD	Bovine Leukocyte Adhesion Deficiency
CA5B	gene Carbonic Anhydrase 5B
CCNB3	gene Cyclin B3
CDC42	gene Cell Division Cycle 42
CHM	gene CHM Rab Escort Protein
cM	centiMorgan
DACH2	gene Dachshund Family Transcription Factor 2
DGKK	gene Diacylglycerol Kinase Kappa
DOCK11	gene Dedicator Of Cytokinesis 11
EFHC2	gene EF-Hand Domain Containing 2
EHH	extended haplotype homozygosity
eROHi	extreme Runs Of Homozygosity islands
F	inbreeding coefficient
F_A	ancestral inbreeding coefficient
FGF16	gene Fibroblast Growth Factor 16
F_{LH1}	Li and Horvitz inbreeding coefficient
F_{ROH}	inbreeding coefficient based on Runs Of Homozygosity
F_{VR1}	VanRaden inbreeding coefficient
F_{YA2}	Yang inbreeding coefficient
GBV	genomic breeding value
GRM	genomic relationship matrix
GRPR	gene Gastrin Releasing Peptide Receptor
GWAS	Genome-Wide Association Studies
HMM	Hidden Markov Model
H_O	observed heterozygosity
HRiD	Haplotype Richness Drop
HRRs	Heterozygosity Rich Regions
HW	Hardy–Weinberg
IBD	identical by descent
IBS	identical by state
iHS	integrated Haplotype Score
Kb	kilobase
KIAA2022	gene Neurite Extension And Migration Factor
KLHL13	gene Kelch Like Family Member 13

Abbreviation	Meaning
LD	linkage disequilibrium
LHFPL1	gene LHFPL tetraspan subfamily member 1
MAGEE2	gene MAGE Family Member E2
MAGT1	gene Magnesium Transporter 1
Mb	megabase
MIR221	gene MicroRNA 221
MJN	Median-Joining Network
MSY	male-specific region
mtDNA	mitochondrial DNA
NDP	gene Norrin Cystine Knot Growth Factor NDP
NGS	Next Generation Sequencing
n_h	haplotype richness
nonPAR	non-Pseudo Autosomal Region
nSL	number of Segregating Sites by Length
OPHN1	gene Oligophrenin 1
PAB	Pseudo Autosomal Boundary
PAR	Pseudo Autosomal Region
PLS3	gene Plastin 3
POLA1	gene DNA Polymerase Alpha 1, Catalytic Subunit
QC	Quality Control
QTL	Quantitative Trait Loci
r	correlation
RLIM	gene Ring Finger Protein, LIM domain interacting
ROH	Runs Of Homozygosity
SE	standard error
SHROOM2	gene Shroom Family Member 2
SHROOM4	gene Shroom Family Member 4
SNP	Single Nucleotide Polymorphism
TMEM27	gene Collectrin
UPRT	gene Uracil Phosphoribosyltransferase homolog
WDR44	gene WD Repeat Domain 44
YIPF6	gene Yip1 Domain Family Member 6
ZDHHC15	gene Zinc Finger DHHC-Type Palmitoyltransferase 15
ZRSR2	gene Zinc Finger CCCH-Type, RNA binding motif and Serine/Arginine Rich 2

List of tables

List of figures

1. INTRODUCTION

Inbreeding and selection are fundamental genetic processes that shape the genetic landscape of populations and species, impacting domestic animals and exerting profound effects on their genetic diversity and defining traits. In domestic animal populations, these processes are ongoing and driven by the relentless pursuit of higher productivity. Inbreeding, characterized by the mating of closely related individuals (Wright, 1933), can often be inevitable in genetically small populations, but it is also practiced intentionally to fix advantageous alleles for desirable traits within the population. Simultaneously, selection, driven by both environmental pressures and human intervention, favors the spread of advantageous genetic variants (Faure and Mills, 2014). Through selective breeding, in which only the most desirable individuals are chosen for reproduction, inbreeding is often increased, leading to reduced genetic variability within the population. While this approach may yield short-term uniformity and higher profits, it carries the long-term risk of accumulating deleterious and lethal mutations that have not yet manifested or been detected (Bosse *et al*., 2019). In domestic animals, such genetic mutations tend to accumulate more frequently due to artificial selection and the lower natural selection pressure in a human-dominated environment. Hence, the study of inbreeding and selection behavior holds paramount importance, especially for the conservation and genetic improvement of domestic animal populations (Brook *et al*., 2002), as well as to improve our understanding of the genetic basis of economically and evolutionarily important traits.

Historically, monitoring inbreeding relied on estimating inbreeding coefficients (*F*) through pedigree data, while selection was tracked through phenotypic observations. However, recent advances in bioinformatic tools and molecular-genetic technology (Next Generation Sequencing, NGS) have revolutionized our ability to estimate inbreeding or to track selection, manifesting as signals of selection in the population genome. Genomic data, particularly Single Nucleotide Polymorphisms (SNPs) from sequences or previously generated SNP arrays, have transformed our understanding of genetic dynamics in domestic animal populations. These high density genotyping tools have allowed researchers to assess genetic diversity, determine the degree of inbreeding, and identify regions of the genome subject to selection with unprecedented accuracy (Kristensen *et al*., 2010; Curik *et al*., 2014; Gouveia *et al*., 2014; Lukic *et al*., 2023). Genomic assessment of inbreeding not only corrects for pedigree related deficiencies, such as Mendelian sampling, but also provides valuable insight into the specific chromosomes and genomic regions where inbreeding is most pronounced. The leading method for estimating genomic inbreeding is based on Runs Of Homozygosity (ROH; Lencz *et al*.,

2007), where ROHs represent autozygous segments, and the sum of all ROHs in a genomic dataset can be considered the proportion of autozygosity (F_{ROH}; McQuillan *et al*., 2008). Furthermore, the power of selection in driving changes in allele frequencies can be tracked through genomic data. Positive selection, for instance, can lead to a reduction in genomic diversity not only at the selected locus but also in nearby regions of the population (Harris *et al*., 2018). This reduction results in a characteristic signal of increased haplotype homozygosity, also referred to as a signal of positive selection. These selection sweeps can take the form of hard or soft sweeps, with hard sweeps involving the rapid increase of one haplotype to high frequency and soft sweeps encompassing multiple haplotypes or previously neutral haplotypes. A range of methods have been developed to detect signals of positive selection. Notable methods for identifying soft sweeps include the integrated Haplotype Score (iHS; Voight *et al*., 2006) and the number of Segregating Sites by Length (nSL; Ferrer-Admetlla *et al*., 2014), while extreme Runs Of Homozygosity islands (eROHi; Boyko *et al*., 2010) excels in identifying hard sweeps. Balancing selection, on the other hand, which favors the maintenance of genetic variation within a population, remains challenging to identify (Fijarczyk and Babik, 2015). Some identification methods exist, such as Heterozygosity Rich Regions (HRRs; Marras *et al*. 2018), but they suffer from low power and a high frequency of false positives.

Despite the enormous potential of genomic data to elucidate inbreeding and selection, one critical aspect of the genome has often been overlooked — the X chromosome. While there have been numerous studies dedicated to estimating inbreeding or detecting selection signals on autosomes, the sex chromosome (X or Z) remains extremely poorly represented in domestic animal populations, with the exception of a few comprehensive (VanRaden *et al*., 2011; Kardos *et al*., 2015; Zavarez *et al*., 2015; Chen *et al*., 2018; Nadachowska-Brzyska *et al*., 2019; Manzari *et al*., 2019; Cesarani *et al*., 2022) and individual studies (Ma *et al*., 2014; Wu *et al*., 2016; Liu *et al*., 2018; Zhu *et al*., 2020; Wang *et al*., 2021a; Curik *et al*., 2022; Rajawat *et al*., 2024), leaving much room for a better understanding of inbreeding and selection behavior as well as good potential for methodological improvements. This is particularly surprising considering the X chromosome's particular characteristics compared to autosomes, such as the genome size (~5%), the low mutation rate (0.015 mutations/megabases/generation), the lower recombination rates (2/3), and consequently the lower effective population size (3/4) and the higher linkage disequilibrium (Schaffner, 2004). In domestic animal populations, females are homogametic and males are heterogametic with one Y chromosome and one X chromosome between which recombination is maintained only in the Pseudo Autosomal Region (PAR). PAR is a unique segment of sequence homology between differentiated

sex chromosomes where recombination occurs during meiosis. Because of monoploidy, the rest of the Y chromosome cannot be used to calculate inbreeding and identify signals. However, the possibility of using the X chromosome and PAR is clear.

The occurrence of ploidy differences between hemizygous males, excluding the PAR, and diploid females is a crucial aspect from a population genomics perspective. For instance, estimating genomic inbreeding and employing certain methods to identify positive selection, such as eROHi, are not feasible for a substantial part of the male X chromosome. Conversely, their hemizygous status provides precise haplotype information that can improve the accuracy of required phasing for other methods such as iHS and nSL. The difference in haplotype richness within this region of the X chromosome also provides an opportunity to develop a new method for detecting signals of positive selection.

Moreover, when considering the impact of sex-linked inheritance on numerous economically important traits, as well as overall health and survival, it becomes evident that the X chromosome deserves increased attention. As highlighted by Robinson *et al*. (2014), inbreeding on the X chromosome is associated with elevated female mortality, resulting in a skewed sex ratio favoring males and a reduction in effective population size, which increases the risk of extinction of the breed. Another indication of the X chromosome's importance lies in its high gene density and reduced recombination rate relative to autosomes, which may lead to higher inbreeding and more pronounced positive selection. In addition, special attention should be given to the PAR singularly. As emphasized by Otto *et al*. (2011), if PAR is relatively large compared to the rest of the sex chromosomes, as in most domestic animals (~5% of the total X chromosome), this phenomenon could be attributed to classical forces that promote recombination. These forces include the avoidance of autozygosity (inbreeding) and the potential presence of balancing selection in males, females, or both.

1.1. Hypotheses and objectives of research

Hypotheses:

1. The level of genomic inbreeding on the X chromosome is higher than the level of inbreeding on autosomes.
2. The variability of genomic inbreeding on the X chromosome is greater than the variability of inbreeding on autosomes.
3. The level of genomic inbreeding in the Pseudo Autosomal Region (PAR) is different between the sexes.
4. Based on the difference in the haplotype richness of the X chromosome of the heterogametic sex, it is possible to identify the signals of positive selection.

Objectives:

1. To evaluate and compare the levels of genomic inbreeding on the X chromosome and autosomes in domestic animal populations.
2. To evaluate and compare the levels of genomic inbreeding in the Pseudo Autosomal Region (PAR) in male and female individuals in domestic animal populations.
3. Establish a new method for identifying signals of positive selection based on the difference in the haplotype richness of the X chromosome of the heterogametic sex.
4. To identify signals of positive selection on the X chromosome in the metapopulation of the native Croatian sheep breeds.

2. REVIEW OF RELEVANT LITERATURE

2.1. The genomics era in domestic animals

Genomics is a multidisciplinary field dedicated to the comprehensive analysis and understanding of an individual or population at the genome level, encompassing both genes and non-coding sequences (Lesk, 2017). It integrates principles and techniques from genetics, molecular biology, bioinformatics, and computational biology to elucidate the structure, function, evolution, and regulation of genomes. This exploration carries profound implications across diverse scientific domains, spanning from medicine to agriculture, evolutionary biology and environmental science. In the context of domestic animals, genomics offers unparalleled opportunities to improve breeding strategies (Meuwissen *et al*., 2016), enhance disease resistance (Pal and Chakravarty, 2020) and maintain genetic diversity (Romanov *et al*., 2023).

2.1.1. Evolution of genomics: molecular-genetic marker advancements

The field of genomics has undergone remarkable evolution, primarily driven by advances in molecular-genetic technology and the development of bioinformatics tools and resources for data analysis and interpretation. Molecular-genetic markers, specific sequences of DNA used to identify unique genetic characteristics within an individual's genome, have played a pivotal role (Vignal *et al*., 2002). These markers enable the detection of genetic variations at the DNA sequence level, surpassing the limitations of previously used phenotypic markers in morphology and protein analysis (Montaldo and Meza-Herrera, 1998). They serve as invaluable tools for analyzing genetic diversity, mapping genes, detecting population polymorphisms, and studying evolutionary relationships between species. Among the most notable molecular-genetic markers are microsatellites, mitochondrial DNA (mtDNA), the Y chromosome, and Single Nucleotide Polymorphisms (SNPs) (Dekkers, 2012). Initially, microsatellites, also known as short tandem repeats (e.g., ATATATATATAT), emerged as versatile molecular markers in the late 20th century, finding extensive applications in population genetics studies of domestic animals (Freeman *et al*., 2016). Despite their continued utility, microsatellites have progressively been supplanted by SNP arrays and whole-genome sequence data in many analyses due to their higher resolution and efficiency.

2.1.2. Single Nucleotide Polymorphisms (SNPs)

Single Nucleotide Polymorphisms (SNPs) are single base pair variations occurring at specific genomic positions (Figure 1). Their significance extends across various scientific fields, due to their widespread occurrence and informative nature (Visscher *et*

al., 2017). SNPs primarily arise from mutations that have proliferated via genetic drift or selective pressures and are ubiquitous throughout the genome, encompassing both coding and non-coding regions. They are categorized based on their genomic location and functional implications. Within coding regions, SNPs, also referred to as coding SNPs or cSNPs, can induce amino acid substitutions, potentially altering protein structure and function. Conversely, SNPs residing in non-coding regions exert regulatory effects on gene expression by modulating elements such as promoters, enhancers, and microRNA binding sites. These non-coding SNPs can influence gene expression levels and splicing patterns, contributing to phenotypic variability and disease susceptibility (Maurano *et al*., 2012). Given these considerations, the development of a set of SNP markers specific to a particular species of domestic animals became imperative.

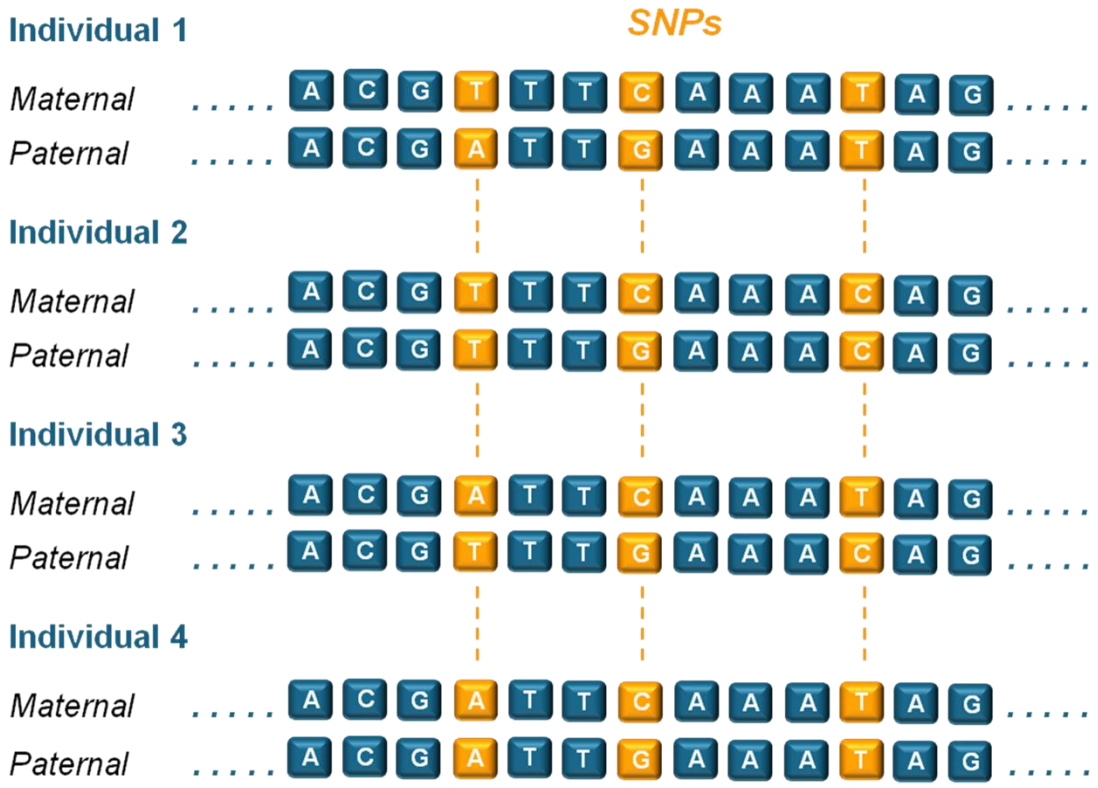


Figure 1. Schematic illustration of SNPs on a short sequence of the genome across several individuals.

2.1.2.1. SNP arrays

The advent of high-throughput sequencing technologies has significantly advanced the detection and genotyping of SNPs on a genome-wide scale. SNP arrays, also referred to as SNP chips or SNP genotyping arrays, have emerged as a transformative innovation, facilitating the simultaneous evaluation of thousands to several millions of SNPs across multiple samples. Since their introduction in the early 2000s, SNP arrays have undergone

continual enhancements, improving their performance, resolution, and marker density (Fan *et al*., 2010). These arrays are constructed using oligonucleotide probes designed to target specific SNP within the genome, ensuring robust and reliable genotyping. A notable feature of SNP arrays is their ability to accommodate a wide range of SNP variability. High-information SNPs, prevalent at a frequency of at least 1% within high number of diverse populations, are strategically selected for inclusion on the array.

Various manufacturers offer SNP array platforms for specific species of domestic animals, each boasting unique capabilities and coverage. For example, Illumina's Infinium and iScan platforms provide customizable SNP arrays with varying marker densities and sample throughput. Similarly, Affymetrix offers the Axiom and GeneChip platforms, employing photolithography and microarray technologies for SNP genotyping. These arrays have their roots in large-scale SNP discovery projects, such as the BovineSNP50 Consortium and the Porcine SNP Discovery Consortium, which identified millions of SNPs across diverse livestock breeds. In general, SNP arrays can be categorized into two main groups: those with low density, such as the Illumina BovineSNP50 BeadChip (53,218 SNPs), the Illumina EquineSNP50 BeadChip (54,602 SNPs) and the Illumina OvineSNP50 BeadChip (54,241 SNPs); and those with high density, such as the Illumina BovineHD BeadChip (777,000 SNPs), the Illumina CanineHD BeadChip (172,000 SNPs) and the Illumina OvineHD BeadChip (606,006 SNPs). While low density arrays are sufficient for certain population structure analyses, high density SNP arrays are more accurate for evaluating inbreeding and selection signals due to their higher resolution.

In the realm of domestic animal populations, SNP arrays have heralded a paradigm shift in breeding programs and genetic analyses. Their utilization has revolutionized various facets of animal husbandry, offering unprecedented insights and opportunities. SNP arrays serve as indispensable tools for genomic analyses such as genomic inbreeding estimation and the identification of genomic regions associated with economically significant traits, e.g. milk production in dairy cows (Raven *et al*., 2014), wool quality in sheep (Arzik *et al*., 2023) and meat quality in pigs (Sanchez *et al*., 2014). In addition, SNP arrays enable the estimation of genomic breeding values (GBVs), empowering breeders to make informed decisions to enhance desired phenotypes and overall productivity (Calus, 2010).

2.2. Inbreeding

Inbreeding, a common practice in animal breeding, refers to the mating of closely related individuals within a population (Wright, 1933). While it often occurs in genetically small populations, high inbreeding is also the result of intensive selection for high productivity. The mating of closely related individuals or the frequent use of founders can lead to the fixation of alleles for desirable traits, but it also elevates the overall degree of homozygosity. Therefore, inbreeding has been used strategically for centuries to fix desirable traits or to establish new breeds (in the last 200 years) or lines of domestic animals. However, the increase in overall homozygosity increases the likelihood of fixing alleles with negative effects, which may be closely linked to positive alleles due to linkage disequilibrium (LD). This phenomenon can lead to negative consequences of inbreeding, which have long been recognized in domestic animal populations and generally fall into two categories:

a) an increased prevalence of lethal or deleterious recessive mutations/disorders: e.g. Bovine Leukocyte Adhesion Deficiency, BLAD (Czarnik *et al*., 2007),

b) inbreeding depression, a decrease in the population mean for a quantitative trait: e.g. fertility (Ferenčaković *et al.*, 2017), milk yield (Bjelland *et al.*, 2013) and survivability (Mc Parland *et al.*, 2007).

Inbreeding depression is explained by two theories. The dominance theory, which defines inbreeding depression by the expression of fixed negative recessive alleles (which are not expressed in the heterozygous form) and assumes the superiority of heterozygotes over the average of homozygotes (Jones, 1917). And the overdominance theory, which states that inbreeding depression occurs when the heterozygote is superior to both homozygotes, and the frequency of heterozygotes is reduced (due to inbreeding), leading to decreased expression of overdominance (Charlesworth and Charlesworth, 1999). Inbreeding depression is estimated by regressing a trait on the inbreeding coefficient (*F*), which emphasizes the need for precise phenotype measurements and a reliable estimate of inbreeding.

2.2.1. Inbreeding estimation methods

Inbreeding is quantified by the inbreeding coefficient (*F*), serving as a numerical measure of the degree of inbreeding in an individual or population (Ballou, 1983). It reflects the probability that two alleles within a locus are identical by descent (IBD). Throughout history, inbreeding coefficients were estimated solely based on the pedigree. However, with advances in bioinformatics and molecular genetics, more and more

genomic estimation methods are being developed based on markers representing the variable sites of the genome. Both traditional (pedigree based) and modern (genomic) methods for estimating *F* are briefly described below.

2.2.1.1. Pedigree based inbreeding estimation

Pedigree based inbreeding estimation has been a cornerstone in quantifying the level of inbreeding within animal populations. It relies on the construction of pedigree records (Figure 2), which trace the ancestry of individuals over multiple generations. There are several methods for estimating *F* from pedigrees, including Path Analysis, Gene Dropping, and Tabular Method, and multiple softwares for their application, such as ENDOG (Gutiérrez and Goyache, 2005) and GRAIN (Baumung *et al*., 2015).

1. **Path Analysis** is the most commonly used method, based on tracing paths within the pedigree. It involves tracking the paths in the pedigree from the individual for which *F* is being calculated back to each common ancestor of the parents. The probability that an individual has inherited two alleles identical by descent (IBD) is computed. The path is closed to avoid revisiting individuals already accounted for in that path. *F* of an individual is equal to the sum of contributions from each path, including *F* of each common ancestor (F_A). The following formula (1) is typically used in the calculation (Ballou, 1983):

$$F_x = \sum_{\square}^{k} \left(\frac{1}{2}\right)^{m} (1 + F_A) \quad (1)$$

   where:

   - *k* = number of paths through all common ancestors,
   - *m* = number of individuals in the path,
   - F_A = inbreeding coefficient of the common ancestor.

2. **Gene Dropping** involves a simulation procedure where hypothetical alleles are assigned to each founder in the population, followed by observing the hypothetical segregation of parental alleles according to Mendel's laws (MacCluer *et al*., 1986).

3. **Tabular Method** relies on the additive relationship matrix between all individuals in the pedigree, representing shared alleles between two individuals expressed as the average relationship between one individual and the parents of the other (Chang *et al*., 1991).

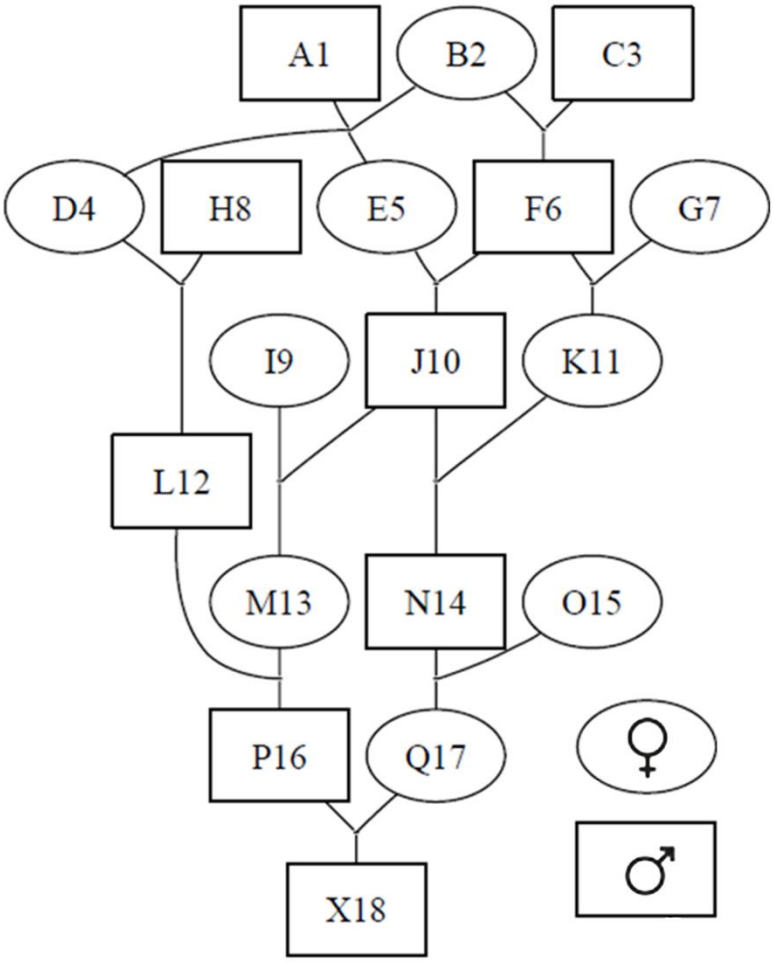


Figure 2. Schematic illustration of pedigree.

Overall, the pedigree based inbreeding estimation has several notable limitations. It does not account for inbreeding in the base population because it is unkown. It also does not take into account the stochastic nature of recombination, as pointed out by Ferenčaković *et al*. (2013). In addition, frequent errors in pedigrees due to inaccurate records, misinterpretation and misidentification are common (Curik *et al*., 2002). Therefore, caution should be exercised when interpreting results obtained using pedigree based methods.

2.2.1.2. Genomic based inbreeding estimation

The emergence of genomic technologies has revolutionized the estimation of *F*. Genomic based methods leverage the vast information contained within an individual's genome to provide more precise and comprehensive estimates of inbreeding levels (Alemu *et al*., 2021). Not only do these methods correct pedigree deficiencies (Ferenčaković *et al*., 2013; Curik *et al*., 2002), but they also offer insights into the precise chromosomes and locations where inbreeding is more pronounced. The estimates are

based on markers that represent variable sites of the genome, with SNPs being the most utilized markers in recent times.

In recent years, numerous methods for genomic based inbreeding estimation have been proposed, yet a consensus on the most suitable method remains elusive. These methods encompass various approaches, including maximum likelihood methods (Milligan, 2003; Wang, 2007), moment-based methods (Purcell *et al*., 2007), correlations between genetic effects (Yang *et al*., 2010; VanRaden, 2008), simple measures of homozygosity (Li and Horvitz, 1953), and analysis of Runs Of Homozygosity (ROH; Lencz *et al*., 2007). Basically, they are all based on the identity-by-state (IBS) status of the genotyped markers, except the method based on ROHs, which uses linkage map information (IBD status). The selection of the optimal method depends on several parameters, such as the number of markers, alleles, individuals, relatedness within the population, mating structure, and intended use (Wang, 2011; Goudet *et al*., 2018). A notable effort in this direction was undertaken by Alemu *et al*. (2021), who conducted a comprehensive comparison of the above methods using whole-genome sequences of 245 individuals of the Holstein breed. Their findings revealed that methods focusing on homozygosity, such as the F_{ROH} coefficient (Lencz *et al*., 2007; McQuillan *et al*., 2008), exhibited remarkable efficiency in estimating inbreeding levels, regardless of allele frequency or age. Further insights into the performance of popular methods of genomic inbreeding were provided by Caballero *et al*. (2022). They compared coefficients including F_{LH1} and F_{LH2} (Li and Horvitz, 1953), F_{VR1} and F_{VR2} (VanRaden, 2008), F_{YA1} and F_{YA2} (Yang *et al*., 2010) and F_{ROH} using simulations in populations under different mating scenarios. While most coefficients showed a fairly high correlation with true inbreeding (F_{IBD}), there were exceptions for F_{VR2} and F_{LH2}, which performed poorly across various conditions. Notably, F_{YA2} emerged as the most accurate estimator in terms of inbreeding depression, and in particular, F_{ROH} consistently proved to be a highly accurate estimator of inbreeding across most simulated scenarios.

Therefore, owing to its proven accuracy and widespread use as a reliable method for estimating inbreeding, the F_{ROH} coefficient is employed in this thesis, alongside the coefficients F_{LH1}, F_{VR1}, and F_{YA2}.

2.2.1.2.1. Inbreeding based on Runs Of Homozygosity (F_{ROH})

The analysis of Runs Of Homozygosity (ROH) has emerged as a powerful tool in population genomics of domestic animals for various analyses. ROHs are continuous, uninterrupted segments of homozygous genotypes (Lencz *et al*., 2007), representing parts of the genome where an individual has inherited identical haplotypes from both parents

(Figure 3). These segments can be classified into those arising due to inbreeding (IBD or autozygous segments) and those stemming from recombination events carrying alleles present from long time ago (IBS segments). According to the principle of recombination, the longer the ROH, the higher the probability that represents inbreeding. Consequently, segments shorter than a species-specific threshold, dependent on the level of LD, are typically not considered in inbreeding estimation, and on average larger segments represent recent inbreeding and vice versa. The expected length of autozygous segments follows an exponential distribution with an average of ½ g Morgans, where g represents the number of generations from the common ancestor (Howrigan *et al*., 2011).

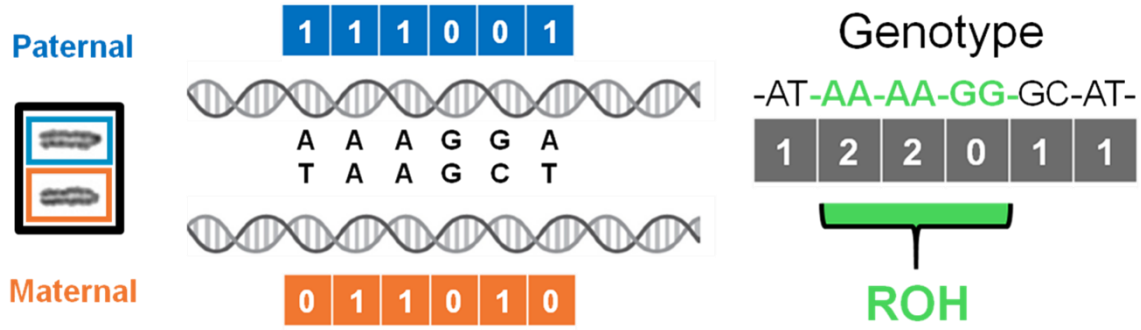


Figure 3. Representation of an ROH segment within an individual.

In light of these principles, McQuillan *et al*. (2008) introduced the F_{ROH} coefficient as a genomic measure of inbreeding. F_{ROH} is defined as the proportion of autozygosity (spanning from 0 to 1) derived from the sum of all ROHs representing autozygous segments within a given genomic information. The general formula (2) is as follows (McQuillan *et al*., 2008):

$$F_{ROH} = \frac{L_{ROH}}{L_{InfoUsed}} \quad (2)$$

where:

- L_{ROH} = total length of all ROHs representing autozygous segments within the genomic information of interest in individual,
- $L_{InfoUsed}$ = length of genomic information of interest covered by molecular markers.

In general, there are two main approaches for determining ROHs: the empirical approach; algorithms for counting genotypes by observers, and the statistical approach; model-based algorithms (Hidden Markov Model; HMM).

1. **Empirical approach** scans genomic information of interest using sliding window (Purcell *et al*., 2007) or consecutive SNPs (Marras *et al*., 2015) of a specified size along the genome to detect segments with homozygous genotypes. The user defines input parameters such as the minimum length of an ROH in megabases (Mb), the minimum number of SNPs within an ROH, and the maximum distance between two SNPs within an ROH. In addition, a certain number of heterozygotes and missing SNPs can be allowed to tolerate genotyping errors and omissions (Curik *et al*., 2014). The allowance of heterozygotes and missings depends on various factors including the genotypic error rate of the SNP array used (or sequencing error), the length of the analyzed ROH, and the marker density. Several softwares implement this approach, including Plink (Purcell *et al*., 2007; Chang *et al*., 2015), detectRUNS (Biscarini *et al*., 2018), and SNP & Variation Suite (SVS) v8.7.0 software package (Golden Helix, Inc., Bozeman, MT, www.goldenhelix.com). However, if a higher number of heterozygotes and missing values are to be allowed, SVS Golden Helix is considered the most reliable software due to its option of not allowing consecutive missing values and heterozygotes. This feature is crucial given the low probability of two consecutive genotyping errors and the higher occurrence of single errors.

2. **Statistical approach** relies on models using HMM to determine ROHs. The first such approach was proposed by Leutenegger *et al.* (2003), but it relied on a two-states HMM, assuming that each marker belongs to either a non-IBD or an IBD segment. The model proposed by Druet and Gautier (2017) expanded previous model by considering multiple classes of IBD segments based on age. Their model, akin to an exponential mixture model, provides a better fit to individual genotypic data and refines the genomic partitioning of inbreeding into stretches of IBD segments with potentially different ancestry. A crucial aspect of this statistical approach is its reliance on a one-order Markov process to define the transition probabilities between successive hidden states. Moreover, the assumption of a known genetic map in the HMM specification underscores the necessity for accurate genetic information. Studies have demonstrated that in scenarios without interference between recombination sites, this model serves as a robust approximation of the IBD process along the genome (Leutenegger *et al*., 2003; Thompson, 2008; Druet and Gautier, 2017; Solé *et al*., 2017). This approach is implemented in RZooROH software (Druet and Gautier, 2017).

A comparative analysis of the empirical and statistical approaches can be found in the study conducted by Solé *et al*. (2017), which examined a large number of cattle

individuals genotyped with the Illumina BovineHD BeadChip. The study revealed remarkably high correlation (r = 0.95) between F_{ROH} estimated from ROHs determined by Plink (empirical approach) and F_{ROH} estimated from ROHs determined by RZooROH (statistical approach). However, it is important to note that the correlation is calculated using F_{ROH} values of individuals. A comparison based on average F_{ROH} values at local genomic sites may provide different insights into their similarities and differences at differently characterized areas (areas with different recombination rates, different marker density, etc.).

In addition, it is essential to highlight that the marker density used to identify ROH segments has a significant impact on the effectiveness of the estimate. Purfield *et al*. (2012) investigated differences in ROHs detected using the empirical approach between two widely employed SNP arrays in cattle: Illumina BovineSNP50 BeadChip (50K), and Illumina BovineHD BeadChip (HD). The study suggests that the 50K SNP array is suitable only for identifying ROHs longer than 5 Mb, as its lower density leads to a higher error rate in detecting heterozygosity within ROHs of observed lengths. Hence, analyses based on such arrays may overestimate the number of segments shorter than 4 Mb. This underscores the superiority of high density SNP arrays for accurate estimation of overall inbreeding, as SNP arrays with lower density are suitable only for detecting recent inbreeding.

2.2.1.2.2. Inbreeding based on homozygosity derivations (F_{LH1})

The F_{LH1} inbreeding coefficient, initially proposed by Li and Horvitz (1953), is based on the deviation of the observed frequency of homozygotes from the expected values under Hardy–Weinberg (HW) equilibrium when current allele frequencies are considered. The formula (3) is expressed as follows (Li and Horvitz, 1953):

$$F_{LH1} = 1 - \frac{\Sigma_{k=1}^{S} x_k(2 - x_k)}{\Sigma_{k=1}^{S} 2p_k(1 - p_k)} \quad (3)$$

where:

- S = total number of markers (SNPs),
- x_k = number of minor alleles of marker k (i.e., 0, 1 or 2 copies),
- p_k = frequency (initial, current or 0.5) of the minor allele.

In this formula, the variance of allele frequencies is factored into the calculation by summing up variances across all SNPs. Given its reflection of deviations from HW equilibrium, this coefficient spans from -1 to 1, where positive values indicate a surplus of

observed homozygosity compared to expectations, and vice versa. Notably, F_{LH1} exhibits a correlation of 1 with the average count of homozygous SNPs, as emphasized by Caballero *et al*. (2022), and can be obtained using software tools such as Plink and GCTA (Yang *et al*., 2011).

2.2.1.2.3. Inbreeding based on genetic drift derivations (F_{VR1}, F_{YA2})

The inbreeding coefficients derived from genetic drift reflect changes in allele frequencies due to random effects within populations. Two notable coefficients in this category are the F_{VR1} and F_{YA2} coefficients. Both coefficients are based on the estimation of the realized genomic relationship matrix (GRM) among individuals within a population and rely on correlations between genetic effects. Thus, they span from -1 to 1, as F_{LH1}, but they do not provide useful information on whether heterozygosity has declined or increased (unlike F_{LH1}). However, the main difference between them is that F_{VR1} is based on the variance of the additive genetic values, while F_{YA2} is based on the correlation between uniting gametes (Caballero *et al*., 2022).

The F_{VR1} coefficient, introduced by VanRaden (2008) and also referred to as F_{GRM} (Sole *et al*., 2017; Forutan *et al*., 2018; Alemu *et al*., 2021; Villanueva *et al*., 2021), captures the extent of inbreeding by assessing the correlation between the actual and expected genetic values of individuals, considering the overall genetic covariance within the population. It can be obtained from the diagonal of the GRM obtained by GCTA with the option --make-grm-alg 1. The general formula (4) is:

$$F_{VR1} = \frac{\Sigma_{k=1}^{S} x_k(2 - x_k)}{\Sigma_{k=1}^{S} 2p_k(1 - p_k)} - 1 \quad (4)$$

where the parameters align with those in formula (3), and all SNP contributions are equally weighted.

The F_{YA2} coefficient, introduced by Yang *et al*. (2010) and also known as F_{UNI} (Sole *et al*., 2017; Alemu *et al*., 2021), is grounded in the correlation between uniting gametes, with homozygous genotypes weighted by the inverse of their allele frequencies. Notably, F_{YA2} is expected to exhibit lower sampling variance compared to F_{VR1}, as highlighted by Yang *et al*. (2011). F_{YA2} estimation is implemented in GCTA. The general formula (5) is:

$$F_{YA2} = \frac{1}{S} \sum_{k=1}^{S} \frac{x_k^2 - (1 + 2p_k)\, x_k + 2p_k^2}{2p_k(1 - p_k)} \quad (5)$$

using the parameters as in formula (3), but the weighting by the variance of allele

frequencies is done for each SNP separately, thereby assigning a greater weight to rare alleles.

Detailed descriptions of the relationships between inbreeding coefficients based on genetic drift derivations can be found in the Appendix of Caballero *et al*. (2021), while a detailed derivation of the F_{YA2} coefficient is provided in the Appendix of Caballero *et al*. (2022).

2.3. Selection

Selection, a cornerstone of evolutionary biology, exerts a profound influence on the genetic composition of populations by favoring individuals with advantageous traits (Saccheri and Hanski, 2006), leading to an increase in the frequency of beneficial genotypic states over generations. This dynamic process shapes the evolutionary trajectory of species, including domestic animals, and plays a pivotal role in the enhancement of desirable characteristics. The application of selection within the context of domestic animals is fundamentally divided into natural and artificial selection (Faure and Mills, 2014). Natural selection results from environmental pressures that favor the survival and reproduction of individuals best adapted to their environment. In contrast, artificial selection is driven by human decisions, where breeders select individuals with desirable traits for reproduction. This deliberate choice accelerates the process of genetic change in domestic animal populations compared to what would occur naturally.

2.3.1. Outcomes of selection in domestic animals

The pursuit of selective breeding in domestic animals has been oriented towards improving economically and culturally valuable traits, leading to remarkable advancements in traits like milk yield, growth rate and fertility. Despite these gains, selection has also precipitated unintended consequences (Rauw *et al*., 1998), notably the reduction of genetic diversity within populations and the increase in genetic divergence between populations. The diminution of genetic variance within populations poses a threat to their adaptability and overall health, highlighting the imperative for balanced breeding programs that judiciously consider both productivity and genetic health. Conversely, increasing genetic divergence between populations has led to the creation of distinct breeds (Mason, 1973), each adapted to specific environments and purposes. This breed diversity is a valuable genetic resource, providing a pool of traits that can be used to meet changing environmental conditions and market demands.

2.3.2. Manifestation of selection

The study of selection, particularly its manifestation in both phenotypic and genotypic forms, traces back to the foundational observations of Charles Darwin (Darwin, 1859). Darwin's work laid the groundwork for understanding the nuanced interplay between selection pressures and the resultant genetic and phenotypic diversity within and between populations. This historical perspective underscores the evolutionary underpinnings of selective breeding practices, providing insight into the genetic diversity observed in contemporary domestic animal breeds. By examining the legacy of selection

through the lens of modern scientific methods, the complex genetic tapestry of domesticated species is continuously unraveled, shedding light on the strategies that have shaped their development through millennia.

Historically, the indicators of selection were only observable through phenotypic manifestations, allowing for the classification of selection into directional, stabilizing, and disruptive categories (Pelabon *et al*., 2010). Directional selection targets the improvement of a specific trait, engendering a shift in the mean value of the trait within the population. Stabilizing selection, in contrast, favors phenotypic averages, curtailing variability, whereas disruptive selection, though less common in livestock breeding, promotes phenotypic extremes, potentially leading to the emergence of distinct subpopulations.

The advent of genomic technologies has transformed the ability to discern and analyze selection patterns within the genomes of domestic animals. Since selection operates by altering allele frequencies, genomic data unveil the footprints of selection, which can be identified through various statistical methods. These methods are broadly categorized into two groups: those that integrate both phenotypic and genotypic data, such as Genome-Wide Association Studies (GWAS; Uffelmann *et al*., 2021) and Quantitative Trait Loci (QTL) mapping (Georges, 2007), and those that rely exclusively on genotypic data to uncover regions of the genome under selection, such as various methods for identification of selection signals (Gouveia *et al*., 2014). The primary distinction between them lies in their focus and application. GWAS and QTL mapping are directed towards identifying SNPs that have a significant impact on specific traits. This is achieved by correlating variations in genotypes with aditive part of phenotypic variations, thereby pinpointing genomic regions that contribute to observable traits. On the other hand, methods dedicated solely to analyzing genotypic data provide insights into the broader landscape of selection within a population. These methods are invaluable, especially when phenotypic data are limited or unavailable, offering a panoramic view of the genomic regions subjected to selection pressures.

Furthermore, based on the preferred genotypic state of the SNPs, selection can be divided into positive selection, which favors homozygosity, and balancing selection, which favors heterozygosity.

2.3.2.1. Positive and balancing selection

Positive selection occurs when specific alleles confer a significant advantage to individuals in a population, leading to an increase in the frequency of these alleles over generations. This process can result in the rapid fixation of advantageous alleles, but it

also has the side effect of reducing genetic diversity around the selected locus (Figure 4), a phenomenon known as a "selective sweep" (Harris *et al*., 2018). Selective sweeps can be categorized into "hard" and "soft" sweeps. A hard sweep involves a novel advantageous allele rapidly becoming dominant in the population, reducing genetic diversity in the surrounding genomic region due to LD. In contrast, soft sweeps may occur from standing genetic variation that becomes beneficial in a changing environment, or from multiple new mutations arising simultaneously. Although soft sweeps also reduce genetic diversity, a higher level of genetic variation is maintained compared to hard sweeps. Both phenomena manifest through increased haplotype homozygosity, which indicates positive selection and is referred to as a signal of positive selection. The quest to uncover signals of positive selection within genomes has spurred the development of various statistical methods. Prominent methods for detecting soft sweeps include the integrated Haplotype Score (iHS; Voight *et al*., 2006) and the number of Segregating Sites by Length (nSL; Ferrer-Admetlla *et al*., 2014), whereas extreme Runs Of Homozygosity islands (eROHi; Boyko et al., 2010) excels in identifying hard sweeps.

Balancing selection, on the other hand, maintains genetic variation at specific locus within a population by favoring heterozygous genotypes. In this context, individuals possessing two distinct alleles at a particular locus (heterozygotes) exhibit a selective advantage over those with two identical alleles (homozygotes), leading to the maintenance of both alleles in the population. Unlike the distinct genomic consequences observable in regions affected by positive selection, the signals of balancing selection do not consistently extend to the surrounding genomic areas. This lack of a uniform pattern complicates the task of discerning loci genuinely under balancing selection from those that may appear as such due to genotyping inaccuracies or anomalies (Figure 4). Consequently, the identification of balancing selection remains challenging (Fijarczyk and Babik, 2015). There are some identification methods, such as Heterozygosity Rich Regions (HRRs; Marras *et al*. 2018), but they suffer from low power and a high frequency of false positives.

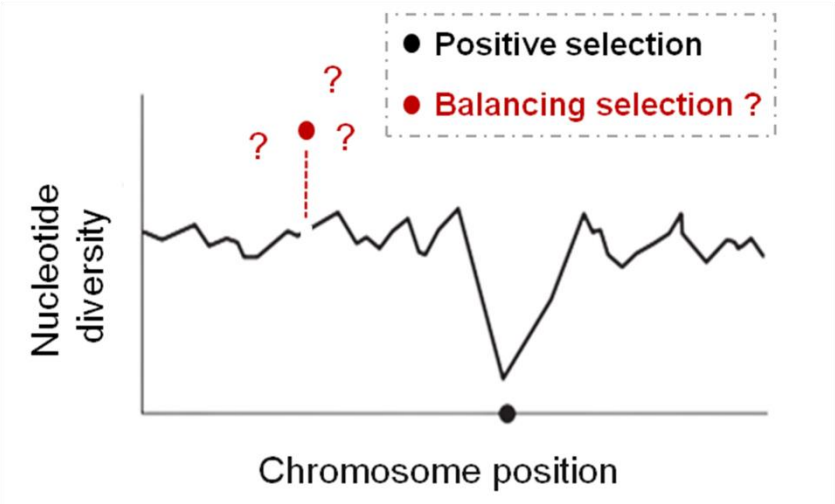


Figure 4. Positive and Balancing selection.

2.3.3. Methods for identification of positive selection signals

The pursuit of identifying positive selection signals has become a focal point for animal geneticists, as these signals are pivotal in identifying specific genes and mutations that confer a selective advantage in particular domestic animal population (Zhao *et al*., 2015). Various methods have been developed, encompassing techniques that evaluate LD, the spectrum of allele frequencies, and patterns of localized genetic variability (Qanbari and Simianer, 2014). While these techniques collectively aim to spotlight regions demonstrating decreased haplotype homozygosity, they each offer unique insights based on the origin, nature, and prevalence of advantageous mutations. eROHi, iHS and nSL are three complementary and commonly used methods for identifying selection signals in domestic animal populations (Utsunomiya *et al*., 2015; Saravanan *et al*., 2020). These within-population methods are based on genomic information from a single representative sample of a target population.

2.3.3.1. Extreme Runs Of Homozygosity islands (eROHi)

The extreme Runs Of Homozygosity islands (eROHi), conceptualized by Boyko *et al*. (2010) and also referred to as ROH islands (Nothnagel *et al*., 2010), are recognized as genomic indicators of strong positive selection. These regions are characterized by an unusually high degree of homozygosity that exceeds the thresholds derived from genome wide analysis. The process for detecting eROHi involves the computation of ROHs through an empirical approach (described in subchapter 2.2.1.2.1.), followed by the identification of regions with an extremely high frequency of SNPs in ROH. Considering that signals may also arise from shorter ROHs, indicating IBS segments and thus long-term adaptive responses, it is permissible to adjust the minimum ROH length criteria downward for eROHi analysis compared to the criteria for inbreeding assessment (IBD

segments), if supported by the density of the genomic data. Hence, high density SNP arrays or sequencing data are preferable, as well as to accurately delineate the boundaries of ROHs. To improve the accuracy and reliability of eROHi as an indicator of positive selection, the establishment of robust statistical thresholds is crucial. These thresholds can be derived from empirical distributions of SNP frequencies in ROHs within the population or from simulations modeling the expected distribution under neutral evolutionary processes (Kardos *et al*., 2017).

2.3.3.2. Integrated Haplotype Score (iHS)

The Integrated Haplotype Score (iHS), developed by Voight *et al*. (2006), is a sophisticated tool for identifying signals of positive selection in populations. This method is based on the observation that alleles under strong selection pressure exhibit a distinct pattern of LD decay characterized by increased haplotype homozygosity beyond what is typically observed under a neutral evolutionary model (Qanbari and Simianer, 2014). The iHS calculation involves comparing the extended haplotype homozygosity (EHH; Sabeti *et al*., 2002) around both derived and ancestral alleles at each core SNP, with the results normalized across frequency bins throughout the genome. This normalization allows for effective comparisons across SNPs, irrespective of their allele frequencies.

iHS quantifies the evidence of selection at or near each SNP, assigning both significantly high positive and high negative values to indicate the presence of selection, whether favoring derived or ancestral alleles. Positive iHS values point towards recent selection for derived alleles, manifesting as unusually long haplotypes when compared to the ancestral state. Conversely, negative values suggest a selection preference for ancestral alleles. However, as Voight *et al*. (2006) note, it is more accurate to look for windows of consecutive SNPs with extreme iHS values rather than treating each outlier individually. This approach is adapted as selection signals tend to create clusters of extreme iHS values across the sweep region, whereas in a neutral model these values are more evenly dispersed.

Prior to the calculation of iHS values, phasing of genotypic data into haplotypes is necessary, a process that can be facilitated by software like Shapeit2 (Delaneau and Marchini, 2014). In addition, accurate ancestral status of each SNP in phased data is critical for accurate iHS analysis and can be achieved through various approaches, such as identifying the most common allele across different breeds, referencing the most common allele in the known ancestor of a species (e.g. the gray wolf for dogs), or consulting publicly available genetic databases. Subsequently, tools like the rehh R

package (Gautier and Vitalis, 2012) and Selscan software (Szpiech and Hernandez, 2014) can be utilized for computing iHS values.

However, it should be noted that iHS has its limitations, particularly in detecting selection signals at low allele frequencies or those approaching fixation, which emphasizes the specificity of the method for certain types of selection scenarios (Voight *et al*. 2006).

2.3.3.3. Number of Segregating Sites by Length (nSL)

The number of Segregating Sites by Length (nSL), introduced by Ferrer-Admetlla *et al*. (2014), is another haplotype-based statistic designed to detect signals of positive selection in a population. While it shares conceptual similarities with iHS, nSL distinguishes itself by its robustness and independence from genetic maps. Central to nSL, like iHS, is the principle of haplotype homozygosity. However, nSL assesses haplotype length through the count of segregating sites (polymorphic sites) rather than the decay of LD, rendering it less influenced by local recombination rates. Thus, iHS can be viewed as nSL with some additional randomness due to the spacing between segregating sites. Ferrer-Admetlla *et al*. (2014) noted that in conditions of low recombination, the variance ratio of iHS to nSL would increase, underscoring the distinct yet complementary nature of these statistics. Consequently, nSL proves especially valuable in identifying sweeps from standing genetic variation and incomplete sweeps and offers an alternative perspective on the haplotype structure surrounding beneficial alleles.

As with iHS, both extreme positive and negative nSL values are indicative of selection, an outlier window analysis for identifying significant signals is suggested, and prior to calculating nSL values, phasing of genotypic data is required, incorporating ancestral allele information. Subsequently, nSL values can be computed on haplotypes using software tools such as selscan.

2.4. Genomic studies of inbreeding and selection in domestic animals

Genomic studies have greatly improved our understanding of inbreeding and selection within domestic animal populations. Regarding inbreeding, genomic analyses have been performed across a broad spectrum of breeds across all domestic species (Howard *et al*., 2017), using different inbreeding coefficients individually or in comparison. In addition, genomic inbreeding within the context of inbreeding depression, has been extensively examined in numerous breeds for a wide range of important traits (Leroy, 2014). These traits include, but are not limited to, production efficiency (Doekes *et al*., 2019), growth rates (Silió *et al*., 2013), and reproductive success (Saura *et al*., 2015). Such advances have improved the management of animal populations by facilitating the accurate identification and mitigation of inbreeding and its effects, thus enabling more effective breeding strategies for genetic health and diversity.

Simultaneously, extensive genomic regions and genes associated with selection have been identified under diverse selection pressures among different breeds of all domestic animal species. These include regions and genes that significantly influence production traits such as milk yield (Yuan *et al*., 2019), growth (Kijas *et al*., 2012), and fertility (Schnabel *et al*., 2005), as well as fitness traits like disease resistance (Li *et al*., 2020). Moreover, genomic research in the field of selection has not only highlighted genes and regions pivotal for particular traits but has also revealed the genetic basis underlying the adaptation of various breeds to specific environmental conditions or market demands, thereby supporting better future decision-making.

Nonetheless, it is important to note that almost all genomic research on inbreeding and selection has focused exclusively on autosomes, neglecting the sex chromosome (X or Z) except for a few comprehensive (VanRaden *et al*., 2011; Kardos *et al*., 2015; Zavarez *et al*., 2015; Chen *et al*., 2018; Nadachowska-Brzyska *et al*., 2019; Manzari *et al*., 2019; Cesarani *et al*., 2022) and individual studies (Ma *et al*., 2014; Wu *et al*., 2016; Liu *et al*., 2018; Zhu *et al*., 2020; Wang *et al*., 2021a; Curik *et al*., 2022; Rajawat *et al*., 2024). Given the unique characteristics of the X chromosome and its importance for numerous traits critical to both economic and evolutionary success, it is imperative to place greater emphasis to the X chromosome in the studies of inbreeding and selection in domestic animals.

2.5. X chromosome

The X chromosome plays an important role in the genomic architecture of mammals, including domestic animals. It is distinguished not only by its unique inheritance patterns but also by its critical role in sex determination and in harboring genes vital for numerous biological functions beyond sex-specific traits. The biology of the X chromosome has been thoroughly studied, both generally and specifically within the context of domestic animals, offering valuable insights into its complex nature and functions (Vaiman, 2002; Schafner, 2004; Raudsepp *et al*., 2012). As highlighted by Schafner (2004), the X chromosome is characterized by particular features compared to autosomes, such as the size of the genome (about 5% of the total genome), the low mutation rate (0.015 mutations/Mb/generation), the lower recombination rates (2/3), and consequently the lower effective population size (3/4) and the higher LD. In addition, the X chromosome is sex-linked (Figure 5), with females possessing two copies (XX) and males possessing one X and one Y chromosome (XY), resulting in different selection pressures and inheritance patterns between the sexes.

In general, the X chromosome can be divided into two regions: the non-Pseudo Autosomal Region (nonPAR) and the Pseudo Autosomal Region (PAR). The nonPAR, which comprises most of the X chromosome, does not undergo recombination with the Y chromosome in males, rendering hemizygous in them and subject to unique evolutionary forces. In females, it is inherited similarly to autosomes, while in males, it follows the inheritance pattern of the Y chromosome. This region is rich in genes, many of which are critical for vital functions and sex-specific traits. Conversely, the PAR, though a smaller segment located at one (domestic animals) or both ends (humans) of the X and Y chromosomes, still undergoes recombination between these chromosomes during male meiosis. Such recombination is essential for the proper pairing and segregation of sex chromosomes during male meiosis, making the PAR's characteristics more similar to autosomes than to the distinct features of the nonPAR. The physical domain of the PAR extends from the terminal ends of the sex chromosomes to the Pseudo Autosomal Boundary (PAB). The existence and variability of the PAR in size and gene content across almost all domestic species have been well-documented (Quilter *et al*., 2002; Young *et al*., 2008; Raudsepp and Chowdhary, 2008; Wilkerson *et al*., 2008; Das *et al*., 2009; Johnson *et al*., 2019; Liu *et al*., 2019), with bovids being a notable exception due to X chromosome rearrangements that have relocated the PAR to the other terminal end of the X chromosome (Raudsepp *et al*., 2012).

2.5.1. Evolution of sex chromosomes

In domestic animals, there are two types of sex-determining chromosome systems: the X and Y chromosomes, which are prevalent among most species (mammals), and the Z and W chromosomes, specific to avian species (birds). These systems have independently evolved from autosomal ancestors on multiple occasions (Ohno, 2013). The primary distinction between them is that in the ZW system, females are heterogametic (carrying one Z and one W chromosome), while males are homogametic (carrying two Z chromosomes). Despite this difference, the X and Z chromosomes exhibit considerable evolutionary and biological similarities, resembling the autosomes from which they originated (Namekawa and Lee, 2009).

In the main sex system, the X and Y chromosomes have derived from a single autosomal lineage, remain homologous and recombine at their ends, within one or two PARs (Figure 5). Outside these zones, their evolutionary paths diverge significantly (Charlesworth, 1991; Rice, 1996). This evolutionary transition is driven by the suppression of recombination through multiple inversions in specific segments of these chromosomes, leading to the differentiation of the X and Y chromosomes in terms of size, gene content and function.

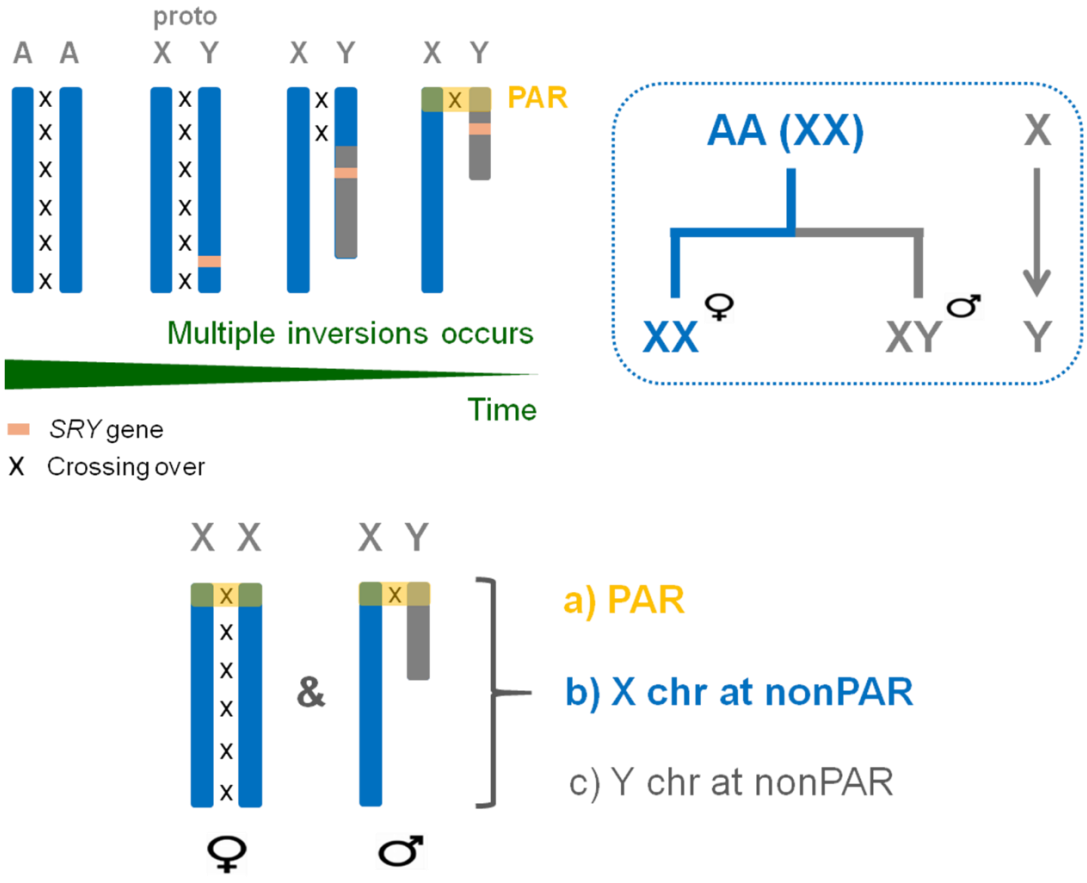


Figure 5. Evolutionary genesis of the XY sex system in domestic animals.

The Y chromosome, although not applicable to all population genetics analyses such as estimating inbreeding and identifying selection signals, has provided insights into paternal lineages, population history, and male-mediated evolutionary processes of various domestic animal species (Ganguly *et al*., 2020) due to its unique characteristics (Figure 6). Two critical features contributing to this are its lack of recombination in the male-specific region (MSY) during meiosis and its inheritance only through the male lineage. Within the MSY, crucial genes govern male-specific traits and functions, including spermatogenesis, sex determination (the SRY gene) and fertility.

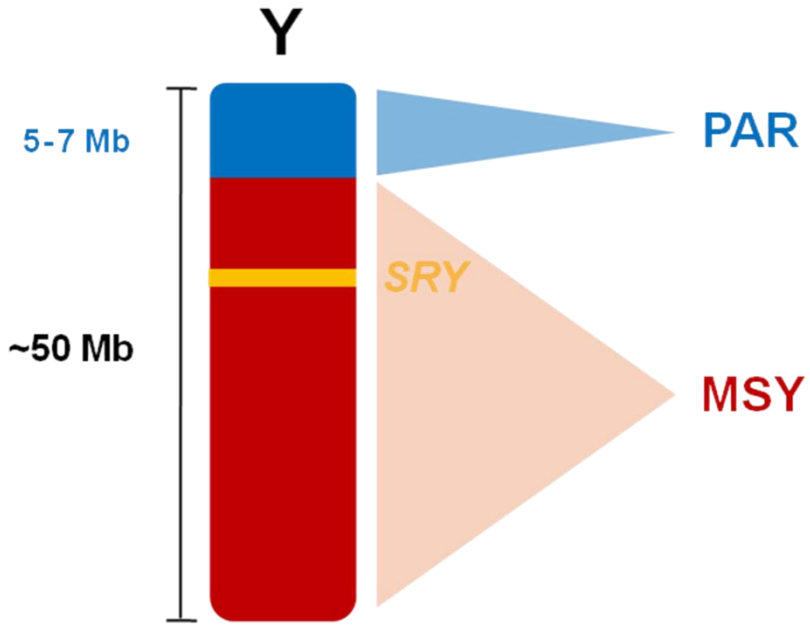


Figure 6. Y chromosome in domestic animals.

In contrast, the X chromosome has maintained a higher gene density and a more stable structure compared to the Y chromosome (Waterson *et al*., 2002), which has lost most of its sequence and genes, evolving into a unique pattern of repetitive sequences, with variability outside PAR caused solely by mutations (Skaletsky *et al*., 2003; Jobling and Tyler-Smith, 2003). The nonPAR of the X chromosome, has evolved to accumulate genes that are vital for general cellular functions, as well as those specifically influencing sex and reproduction. This region undergoes X inactivation in females to equalise gene expression between the sexes, a process that has significant implications for the evolution of gene expression regulation on the X chromosome (Carrel and Willard, 2005). The evolutionary maintenance of PAR, on the other hand, suggests a balancing act between the need for sex chromosome segregation and the preservation of essential autosomal genes (Otto *et al*., 2011). The evolutionary dynamics of the PAR are influenced by mechanisms that counteract the degradation of the Y chromosome, ensuring the

continuation of recombination and the maintenance of gene flow between the X and Y chromosomes.

2.5.2. Inbreeding and the X chromosome

Inbreeding exerts a distinctive impact on the X chromosome, differentiated by its unique inheritance patterns and biological functions. This effect can be broadly categorized into two areas: the impact on the nonPAR, which constitutes the majority of the X chromosome, and the impact on the PAR.

The impact on the nonPAR is profound due to its high gene density, including genes critical for survival, fertility, and other essential physiological functions. Furthermore, the nonPAR's unique features, such as hemizygosity in males and the dosage compensation mechanisms like X inactivation in females, render it particularly susceptible to inbreeding's consequences. For males, being hemizygous means that any detrimental alleles are expressed without the chance of being masked by an alternative allele. On the other hand, females, despite having two copies, are not completely protected due to X inactivation and the possibility of increased expression of detrimental alleles should they become homozygous. Studies, such as those by Robinson *et al*. (2014), have shown that inbreeding on the X chromosome (specifically the nonPAR) correlates with elevated female mortality, which can shift the sex ratio toward males, decrease effective population size, and potentially heighten the risk of extinction for certain breeds. This increase in mortality of the homogametic sex is attributed to a rise in the frequency of harmful, sex-limited alleles, with increased X chromosomal inbreeding serving as an indicator. Moreover, an increase in the nonPAR inbreeding coefficient can induce inbreeding depression for various significant traits in females (Curik *et al*., 2022). In addition, the nonPAR's reduced rate of recombination compared to autosomes could contribute to higher levels of inbreeding (Cotter *et al*., 2024).

In contrast, the effects on the PAR may be profound, but for other reasons, such as the maintenance of essential autosomal genes, which may be important for both sexes. Otto *et al*. (2011) pointed out that if the PAR is relatively large compared to the rest of the sex chromosomes, as is the case in most domestic animals (about 5% of the total X chromosome), this phenomenon could be attributed to classical forces that promote recombination, such as the avoidance of autozygosity (inbreeding) in males, females, or both. The PAR thus warrants focused attention, particularly as it remains underexplored (alongside the total X chromosome) within the context of inbreeding research in domestic animal populations.

2.5.2.1. Inbreeding estimation on the X chromosome

The hemizygous nature of the X chromosome in males requires specific methodological considerations for inbreeding estimation, diverging from the autosomes. When analysing the PAR solely, standard pedigree and genomic methods applicable to autosomes are directly translatable and equally effective for estimating *F*, across both sexes, attributed to its homologous nature.

Conversely, if the analysis extends to or focuses exclusively on the nonPAR, the estimation is viable only for females, given their X chromosome diploidy. In addition, this scenario demands some adjustments for female assessments compared to those based on autosomes. Path analysis through pedigree data, adapted for the X chromosome, excludes male pathways in the calculation (Kudo and Sakaguchi, 1963; Johnston *et al*., 2019). Specifically, any path involving two or more male individuals is disregarded, recognizing the non-transmission of the X chromosome from father to son. Softwares such as KinInbcoefX (Zheng and Bourgain, 2009) facilitate these specialized calculations by incorporating sex information within pedigree data. Furthermore, Fernando and Grossman (1990) have outlined rules for constructing affinity matrices (tabular method) specific to the X chromosome, notably differentiating by halving diagonal values for males and excluding paternal links.

Simultaneously, the modification of genomic methods for inbreeding estimation across the entire X chromosome has not yet been sufficiently researched (Calderón *et al*., 2009). Existing studies (VanRaden *et al*., 2011; Zavarez *et al*., 2015; Yengo *et al*., 2019; Cotter *et al*., 2024) adopt same values of parameters used for autosomes directly to the X chromosome due to a lack of specific guidelines for parameter adjustments. Although this strategy is practical, it might not adequately reflect the complexities of X-linked inheritance, underscoring a necessity for tailored parameter values that consider the unique properties of the X chromosome, such as the differential recombination rates.

2.5.3. Positive selection signals and the X chromosome

The study of positive selection on the X chromosome, characterized by its unique properties and high density of important genes for various traits, is essential for expanding our understanding of animal genetics, both in general and within specific populations. Its analysis, either independently or in conjunction with autosomes, enhances the ability to uncover the genetic background of distinct breeds subjected to various selection pressures due to its size and critical importance (Chen *et al*., 2018). As a result, a clearer

picture of the genetic background of diverse essential traits is obtained, an important aspect in the field of breeding and conservation strategies.

Moreover, it has been proposed that selection on genes located on the X chromosome could be more efficient, as all allele effects are fully exposed to selection when expressed in hemizygous males (Vicoso and Charlesworth, 2006). Nevertheless, determining whether a more pronounced selection signal can be expected on the X chromosome than on autosomes remains a challenge, as the intensity of selection is also influenced by factors such as genetic drift, mutation and recombination rates, which are quite different on the X chromosome. Thus, further empirical analyses are essential to better understand the interplay between selection and these genetic factors on the X chromosome.

In terms of identifying signals of positive selection, genomic methods used for autosomes, such as eROHi, nSL, and iHS, are adaptable and demonstrate similar effectiveness when applied to the X chromosome in females. However, their applicability in males is compromised due to the hemizygous nature of their X chromosome. This limitation emphasizes the need to develop specific methods for identifying positive selection signals on the X chromosome in males.

2.5.3.1. Hemizygous status: central point for development of a new positive selection method

The differentiation in ploidy between hemizygous males (excluding the PAR) and diploid females represents a crucial aspect from a positive selection perspective. For instance, employing certain methods to identify positive selection, such as eROHi, proves impractical for a considerable portion of the male X chromosome (nonPAR). Yet, the hemizygous state provides precise haplotype information that not only enhances the accuracy of phasing required (Choi *et al*., 2018) for other methods, such as iHS and nSL, but crucially, lays the basis for pioneering methods aimed at identifying signals of positive selection on the X chromosome in males.

This accurate haplotype information, by facilitating the extraction of exact haplotypes over various lengths, could prove vital in detecting positive selection signals through the identification of regions showing a marked reduction in allele (haplotype) diversity. To investigate regions of diminished haplotype diversity, the approach proposed by Kimura and Crow (1964), based on the calculation of the effective number of alleles, can be adapted. The adaptation involves modifying the approach to compute and compare the effective number of alleles across windows of hemizygous genomic data.

Furthermore, using these haplotypes, phylogenetic network analysis can be employed within each identified signal to clarify their phylogenetic relationships, distinguishing between derived and ancestral haplotypes if ancestral information is available.

3. MATERIALS AND METHODS

3.1. Genotypic data

To explore inbreeding and selection on the X chromosome with greater precision, the utilization of high density genotypic data, alongside a sufficient number of males and females, is essential. Therefore, the analysis encompassed two distinct populations for each of the following domestic animal species: cattle, dogs, and sheep, all genotyped with high density SNP arrays (Table 1).

- For cattle (genotyped with Illumina BovineHD BeadChip; 777,000 SNPs), a metapopulation of three native Croatian breeds (Croatian Busha, Istrian Cattle - Boškarin and Slavonian Syrmian Podolian Cattle) represented by 33 males and 79 females evenly distributed across the breeds, and a population of the Nellore breed represented by 963 males and 1267 females, were used.
- For dogs (genotyped with Illumina CanineHD BeadChip; 172,115 SNPs), a population of Labrador Retriever breed represented by 145 males and 129 females, and a population of Patagonian Sheepdog breed represented by 105 males and 41 females were utilized.
- For sheep (genotyped with Illumina OvineHD BeadChip; 606,006 SNPs), a metapopulation of eight native Croatian breeds (Istria Sheep, Pag Island Sheep, Dubrovnik Ruda, Cres Island Sheep, Rab Island Sheep, Krk Island Sheep, Dalmatian Pramenka and Lika Pramenka) represented by 101 males and 101 females evenly distributed among the breeds, and a population of Soay breed represented by 87 males and 102 females were included.

Genotypic data for the Labrador Retriever were obtained from Binversie *et al*. (2020), for the Patagonian Sheepdog from Barrios *et al*. (2022), and for the Soay from Johnston *et al*. (2016), through the GEO NCBI and Dryad online repositories. The genotypes for the Croatian native cattle and sheep breeds (sampled and genotyped in Croatia as part of the ANAGRAMs project) and for the Nellore (sampled and genotyped in Brazil) were ceded for this study. All samples were collected from animals bred by registered breeders, who provided details on the animals' origins and farm locations. Sampling was done in accordance with National Gene Bank protocols, from which DNA was isolated using a commercial kit.

In these six domestic animal populations, comprehensive analysis of inbreeding on the X chromosome and PAR were conducted, while a new proposed method for

identifying signals of positive selection was introduced and evaluated on a metapopulation of the native Croatian sheep breeds. This evaluation involved a comparative analysis with other complementary intra-population methods: eROHi, iHS and nSL. To accurately determine the ancestral allele status of each SNP in phased data, which is a critical step for precise iHS and nSL analysis, and for constructing the phylogenetic relationship (derived versus ancestral haplotype) for each identified signal, the genotypes of 10 mouflons (the closest ancestors of sheep) were also ceded from the ANAGRAMs project.

3.2. Quallity Control and PAR localization

For the Quality Control (QC) of genotypes, SAS 9.4. (SAS Institute, Cary, NC) and Plink software were employed. The initial step was to maintain only X chromosomal and autosomal SNPs from the Illumina final report of each of the six domestic animal populations, that were located on one of the specified reference genomes: ARS-UCD 1.2 for cattle, CanFam3.1 for dogs and Oar v4.0 for sheep. Subsequently, questionable genotyped SNPs were excluded using the following parameters: GenTrain <0.4, GenCall ≤0.8, call rate <0.9 and SNPs deviating from Hardy-Weinberg equilibrium with $p < 10^{-7}$, and then questionable individuals (call rate <0.95).

To localize PAR, the observed heterozygosity (H_O) was calculated separately for males and females in each population with the R programming language (R Core Team, 2021), assuming that the nonPAR region in males has an H_O value of 0 due to their hemizygosity. All heterozygous SNPs located in the nonPAR region in males were considered as SNPs with questionable assembly (highlighted in red in Figure 8, 9 and 10) and excluded from further analyses. A detailed overview of the genotypic data for each dataset after QC is presented in Table 1.

Table 1. Number of samples and SNPs per each domestic animal population.

Population	Before QC		After QC				
	Males	Females	Males	Females	autoSNPs	chrXSNPs	parSNPs
Croatian cattle breeds	33	79	33	79	610 457	30 871	1994
Nellore	963	1267	953	1255	709 591	34 011	3181
Labrador Retriever	145	129	145	126	111 198	2981	319
Patagonian Sheepdog	105	41	105	40	82 650	2802	283
Croatian sheep breeds	101	101	100	101	391 532	18 983	1232
Soay	87	102	86	102	393 273	15 309	1562

autoSNPs – number of autosomal SNPs; chrXSNPs – number of X chromosomal SNPs; parSNPs – number of SNPs located within PAR.

3.3. Estimation and evaluation of X chromosomal and PAR inbreeding

A comprehensive approach was taken to thoroughly estimate and evaluate genomic inbreeding on the X chromosome and PAR in domestic animal populations. This involved the calculation of several inbreeding coefficients: F_{ROH}, F_{LH1}, F_{VR1} and F_{YA2}, across different genomic regions. For the X chromosome, calculations were conducted exclusively for females due to its diploid nature. In contrast, the PAR was analyzed separately for males and females, as were the autosomal coefficients that served as the basis for comparison.

3.3.1. Inbreeding based on Runs Of Homozygosity (F_{ROH})

The F_{ROH} coefficient, recognized for its accuracy and reliability, was employed as the primary method for estimating genomic inbreeding. Two distinct approaches were used to identify ROHs representing autozygous (IBD) segments: the empirical consecutive approach and the statistical approach.

In the empirical consecutive approach, ROHs were determined for each individual using the SVS Golden Helix software. Five categories of ROH lengths were defined: 1-2 Mb, 2-4 Mb, 4-8 Mb, 8-16 Mb and >16 Mb. The following criteria were used: minimum ROH length was set to 1 Mb and 15 SNPs, maximum gap and maximum density parameters were disregarded, and heterozygotes and missingness were allowed based on the ROH length category (Table 2) according to Ferenčaković *et al*. (2013). No consecutive missing values and heterozygotes were allowed as this is more robust when genotyping errors and/or misassembling occur. The value of 1 Mb for the minimum ROH length was set under the assumption that the autozygosity of the individual originates from common ancestors up to 50 generations in the past (1 cM = 1 Mb).

Subsequently, F_{ROH} for ROH longer than 1 Mb and representing total inbreeding (F_{ROH_SVS}) was calculated using formula (2). In addition, F_{ROH} were calculated for lengths >2 Mb ($F_{ROH\_SVS>2Mb}$) indicating inbreeding within the last 25 generations, >4 Mb ($F_{ROH\_SVS>4Mb}$) within 12 generations, and >8 Mb ($F_{ROH\_SVS>8Mb}$) indicating recent inbreeding (up to 6 generations).

Table 2. Number of heterozygotes and missingness allowed per each ROH length category in the empirical consecutive approach.

	ROH length category				
	1-2 Mb	2-4 Mb	4-8 Mb	8-16 Mb	>16 Mb
Heterozygotes	1	2	4	8	16
Missingness	4	8	16	32	64

In the statistical approach, ROHs were determined for each individual using the RZooROH software. In the model, 1 cM corresponded to 1 Mb and ROHs were defined in seven HBD categories, ranging from HBD1 to HBD7. Each category reflects a specific depth of ancestry: HBD1 corresponds to inbreeding from one generation ago, HBD2 from two generations ago, HBD3 from four generations ago and HBD7 from 64 generations ago. Then, including all identified ROHs, the total inbreeding ($F_{ROH\_RZooROH}$) was calculated using formula (2) and divided into F_{ROH} for each individual HBD category (from $F_{ROH\_RZooROH\_HBD1}$ to $F_{ROH\_RZooROH\_HBD7}$).

In addition, genomic inbreeding (both F_{ROH_SVS} and $F_{ROH\_RZooROH}$) was assessed and compared regionally across the X chromosome and autosomes using a sliding window analysis equivalent in size to PAR and a sliding of 0.2 Mb. This was done to get a better insight into the behavior of inbreeding in PAR and the X chromosome, i.e. to see exactly where inbreeding is more pronounced and to compare inbreeding in PAR with windows of the same length.

3.3.2. Inbreeding based on IBS status (F_{LH1}, F_{VR1} and F_{YA2})

As additional reliable genomic inbreeding coefficients, F_{LH1}, F_{VR1} and F_{YA2} were determined. The F_{LH1} coefficient was estimated using the Plink software employing the --het option, while the F_{VR1} and F_{YA2} coefficients were both determined using the GCTA software, with the F_{VR1} estimated as the diagonal of the GRM employing the --make-grm-alg 1 option and the F_{YA2} estimated employing the --ibc option.

3.3.3. Comparison and visualization of estimated inbreeding coefficients

Comparison and visualization of estimated inbreeding coefficients across genomic landscapes (X chromosome, PAR and autosomes) were conducted using the R programming language. For each domestic animal population, the strength of the linear relationships among various estimated inbreeding coefficients within and between the X chromosome, PAR and autosomes in each sex was assessed employing Pearson correlation analysis (*r*).

3.4. Development and validation of a new method for identifing positive selection signals on the X chromosome

With the idea of maximising the use of all available genotyping information, a new method called Haplotype Richness Drop (HRiD) was designed to exploit the distinct genetic architecture of males for pinpointing positive selection signals on the X chromosome. To evaluate the effectiveness and reliability of HRiD, positive selection signals on the X chromosome were identified in a metapopulation of native Croatian sheep breeds using HRiD alongside three established methods: eROHi, iHS and nSL. While the iHS and nSL analyses used the complete dataset comprising 201 genotypes (both males and females), eROHi was applied exclusively to the 101 female genotypes and HRiD exclusively to the 100 male haplotypes.

3.4.1. Haplotype Richness Drop (HRiD): explanation and derivation of the concept

On the nonPAR part of the X chromosome, male genotypes are hemizygous, making it easy to derive exact haplotypes of different lengths. Leveraging this feature, a new method, HRiD, has been developed that uses data from male haplotypes to detect positive selection signals on the X chromosome. HRiD is based on the calculation of the effective number of alleles, a concept described by Kimura and Crow (1964) as "the expected value of the sum of squares of the allele frequencies, or, in simpler terms, the inverse of the effective number of alleles a population maintains." This concept, recognized in conservation genetics as an indicator of allelic richness, symbolized as A_e or n_a (Allendorf *et al*., 2013; Greenbaum *et al*., 2014), is adapted in HRiD to measure haplotype richness (n_h) and referred to as the effective number of haplotypes.

The main assumption is that the presence of positive selection leads to a sudden decrease in the effective number of haplotypes, which was measured by calculating Haplotype Richness Drop values (HRiD) defined by the formula (6):

$$HRiD_{\mathrm{w_{i+1}}} = \frac{n_{h_{\mathrm{w_i}}} + n_{h_{\mathrm{w_{i+2}}}}}{2n_{h_{\mathrm{w_{i+1}}}}} \quad (7)$$

where $n_{h_{\mathrm{w_i}}}$ represents the effective number of haplotypes of the i^{th} sliding window (haplotype) under study (i = 1, ..., w_i, where w_i = 503 and w_{i+2} = 505).

For the first (w_1) and last window (w_{i+2}), the formula of the numerator is slightly different to allow identification of selection signals in these two windows. Specifically, the numerator for the first window is $2n_{h_{\mathrm{w_2}}}$, while the numerator for the last window is $2n_{h_{\mathrm{w_{i+1}}}}$. It is

important to note that the effective number of haplotypes within each window is determined by the inverse of the sum of the squares of their haplotype frequencies, a calculation performed using the R programming language. In the absence of selection, HRiD values should fluctuate around the value of one, whereas positive selective sweeps would lead to higher positive values because n_h is much lower compared to surrounding regions (Figure 7). With this approach, the selection signals detected by HRiD do not depend on the heterogeneity of recombination rates. The size of the window was set to 70 SNPs with a slider of 35 SNPs (average ≈ 500 Kb and 250 Kb) to allow direct comparison of signals with those obtained by other methods. HRiD is expected to efficiently detect signals of positive selection that resemble hard sweeps. HRiD values were normalized and converted to −log(P) values. Windows (haplotypes) with −log(P) ≥3.3 corresponded to a minimum HRiD value of 2.8 and were considered significant.

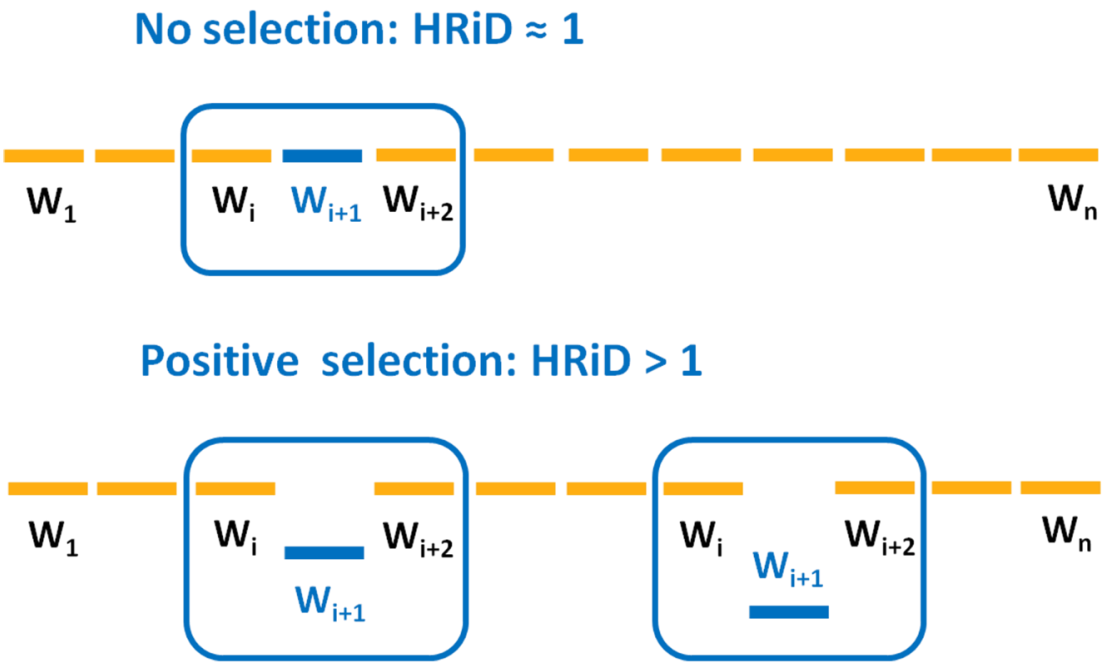


Figure 7. Concept of the HRiD analysis.

3.4.2. Extreme Runs Of Homozygosity islands (eROHi)

Through a consecutive empirical approach, in addition to the ROHs identified for the F_{ROH} estimation (see Table 2), a sixth category of ROH length was defined: 0.25-1 Mb for each female, using the SVS Golden Helix software with no heterozygotes or missing SNPs allowed. While the eROHi method typically focuses on ROHs longer than 1 Mb, indicating an origin within the last 50 generations, the inclusion of much shorter ROHs (>0.25 Mb) was intended to more accurately track selection patterns that may have arisen in the last 200 generations. This analysis is only possible if the density of the SNP array

allows it, which is the case here (an average of 35 SNPs covered 0.25 Mb), as a high density SNP array was used.

Subsequently, consecutive SNPs that occurred with a high frequency within the ROHs were utilized as indicators of regions potentially subject to positive selection. The frequency of each SNP within the ROHs was calculated and then normalized by the mean frequency, with the transformed value expressed as −log(P). SNPs exhibiting a −log(P) value of ≥3.3 were considered outliers, whereas regions with consecutive outliers were considered significant. The significance threshold [−log(P) ≥3.3] corresponded to a frequency of 0.396 (40 individuals) and was calculated using the simpleM method (Gao *et al*., 2008). At the end of the analysis, the signals are ranked according to the highest −log(P) value within a signal (“peak of signal”).

3.4.3. Integrated Haplotype Score (iHS) and number of Segregating Sites by Length (nSL)

Initially, the haplotype phasing required for iHS and nSL analysis was performed with Shapeit2 software, utilizing the --chrX option to incorporate both male and female genotypes. The VCF file was recoded to designate the ancestral allele as the reference and the derived allele as the alternative, using the most frequent allele in mouflons as the basis for the ancestral information. Subsequently, the iHS values for each SNP were calculated using the R package rehh and the nSL values were calculated using the software Selscan without restrictions (the allowed values for gap-scale and max-gap were higher than the maximum distance in data set used).

All iHS and nSL values were normalised within the frequency bin size of 0.025, and −log(P) values were calculated assuming two-sided tests to account for both significantly positive and negative iHS or nSL values. A sliding window approach was adopted (500 kb size; 100 kb slide), and SNPs exhibiting −log(P) >2 were flagged as outliers. Non-overlapping windows containing more than 10% outliers were identified as significant signals, ranked by their proportion of outliers. Furthermore, because both methods are ratio-based, they are limited in their ability to detect sweeps that are very close to fixation, so for any SNP with a minor allele frequency <5%, a calculation was not possible. Therefore, they were assigned the value NA, but they were retained for the construction of haplotypes in adjacent SNPs and are also informative for the eROHi and HRiD methods.

3.4.4. Analysis of Median-Joining Networks (MJNs) within identified signals

To illustrate the phylogenetic relationship between ancestral and derived haplotypes in the selection signals obtained by HRiD and classical methods, several Median-Joining Networks (MJNs) were constructed. The phylogenetic analysis was based on 100 male sheep and five male mouflons, which were assumed to represent the ancestral haplotypes. First, the SNPs of the X chromosome were converted to fasta format using the R package seqRLFP (Ding and Zhang, 2012) and visualised using MEGA7 (Kumar *et al*., 2016). In addition, DnaSP (Rozas *et al*., 2017) was used to derive unique haplotypes for each selection signal identified, while MJNs (Bandelt *et al*., 1999) were generated using both Arlequin v. 3.5.2.2 (Excoffier and Lischer, 2010) and PopART software (Leigh and Bryant, 2015).

3.4.5. Gene annotation and functional characterization of identified signals

Gene annotation within the identified signals of positive selection was conducted using SVS Golden Helix software based on positions in the Oar v4.0 reference sheep genome. Functional analysis of these candidate genes was carried out utilizing platforms such as UniProt (https://www.uniprot.org/) and GeneCards (https://www.genecards.org). In addition to genomic information from sheep (UniProt), genomic information from other species, including humans (GeneCards) and cattle (UniProt), was also used.

4. RESULTS

4.1. PAR localization

The H_O of males and females of SNPs located on the X chromosome are shown in Figure 8 for cattle (8A – metapopulation of native Croatian cattle breeds, 8B – Nellore), Figure 9 for dogs (9A – Labrador Retriever, 9B – Patagonian Sheepdog) and Figure 10 for sheep (10A – metapopulation of native Croatian sheep breeds, 10B – Soay).

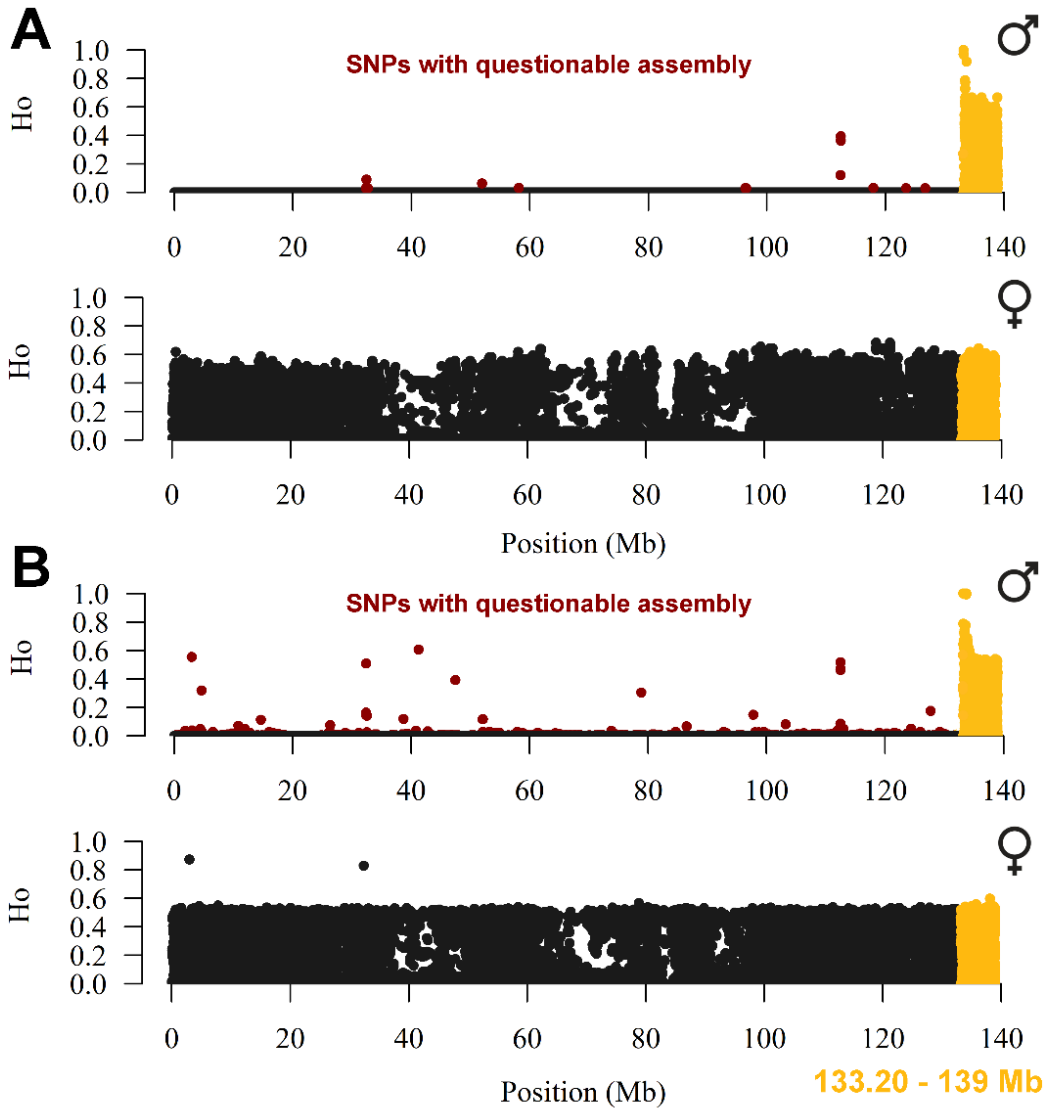


Figure 8. H_O of males and females of X chromosomal SNPs (SNPs within PAR are coloured in yellow) in metapopulation of native Croatian cattle breeds **(A)** and Nellore **(B)** representing cattle.

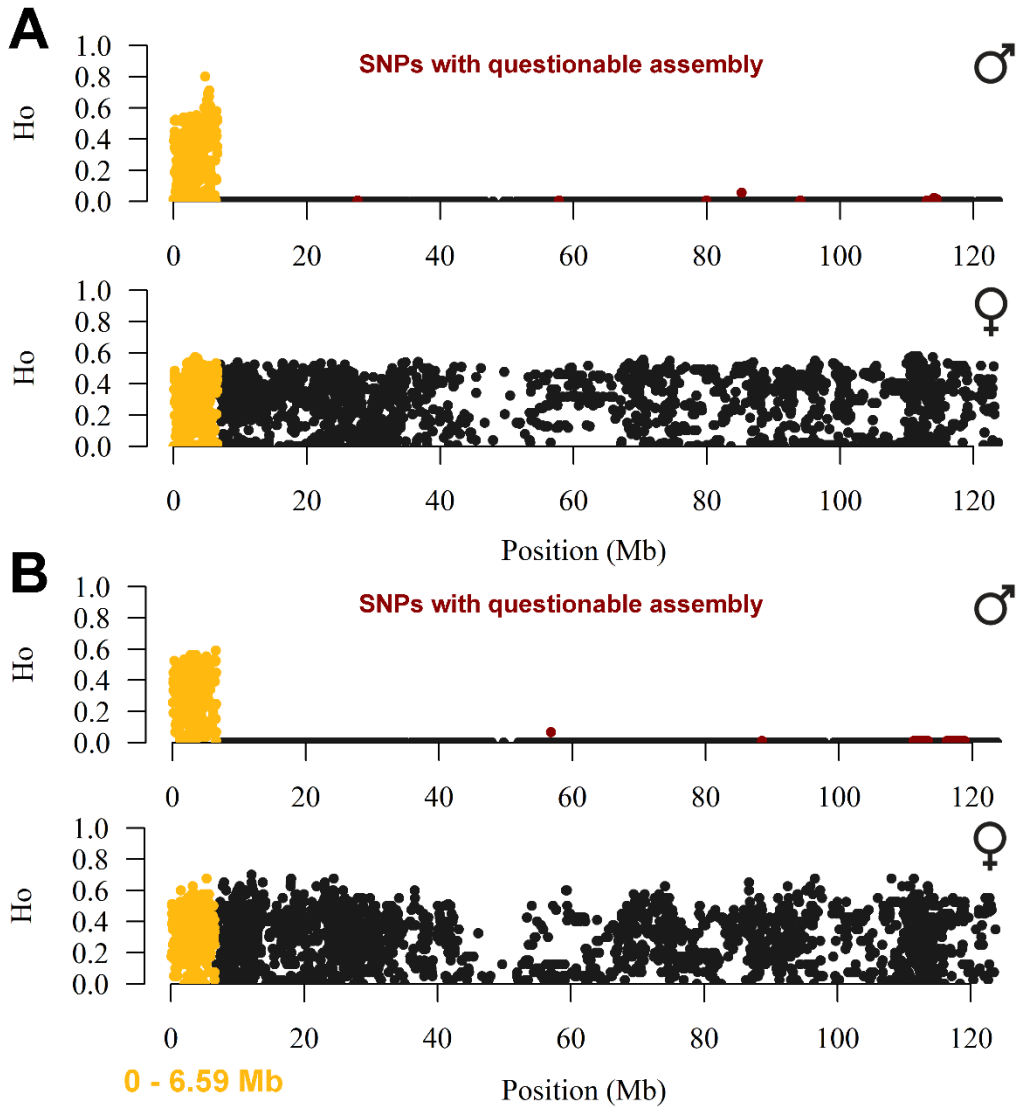


Figure 9. *H*$_O$ of males and females of X chromosomal SNPs (SNPs within PAR are coloured in yellow) in Labrador Retriever **(A)** and Patagonian Sheepdog **(B)** representing dogs.

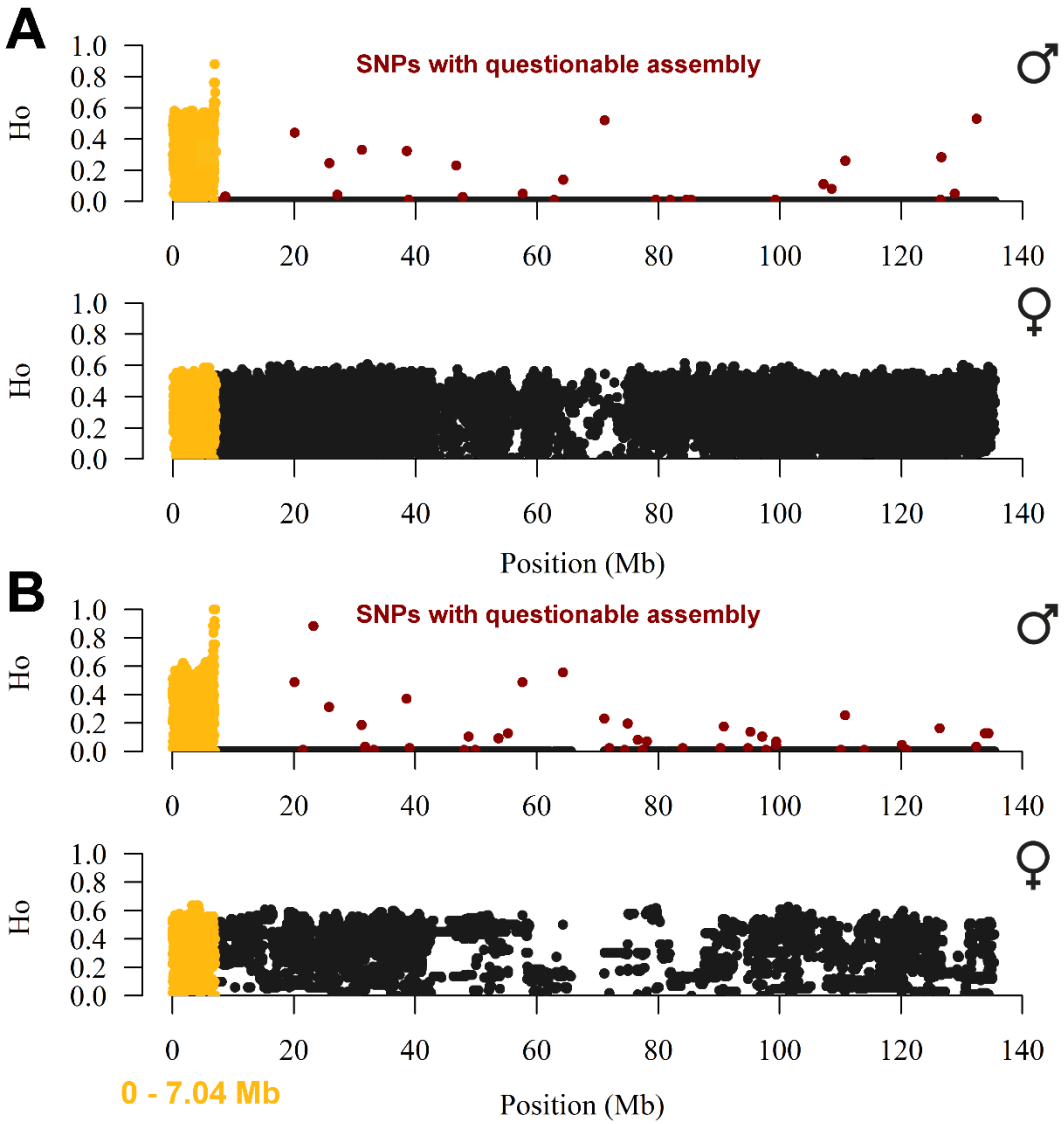


Figure 10. *H_O* of males and females of X chromosomal SNPs (SNPs within PAR are coloured in yellow) in metapopulation of native Croatian sheep breeds **(A)** and Soay **(B)** representing sheep.

In all populations, the average *H_O* value in females was approximately the same across the entire X chromosome, whereas in males an *H_O* value of 0, indicative of hemizygosity (nonPAR region), were observed ranging from 0.00 – 133.20 Mb in cattle, 6.59 – 123.72 Mb in dogs and 7.04 – 135.40 Mb in sheep. Thus, the PAR likely extends from 0.00 to 7.04 Mb in sheep (size of 7.04 Mb), from 0.00 to 6.59 Mb in dogs (size of 6.59 Mb), and conversely, at the other end of the X chromosome, from 133.20 to 139.00 Mb in cattle (size of 5.80 Mb). These findings are consistent with the mapped PAR in their reference genomes.

4.2. Genomic inbreeding coefficients estimated across X chromosome, PAR and autosomes

The mean values with associated standard errors (SE) of genomic inbreeding coefficients estimated on autosomes and PAR separately for females and males, and on the X chromosome for females only, of each domestic animal population are presented in Tables 3 and 4. For the coefficients F_{VR1}, F_{LH1} and F_{YA2}, mean values across autosomes, X chromosome, and PAR generally hovered around zero, indicating minimal discrepancies between these genomic landscapes within all populations.

In contrast, coefficients F_{ROH_SVS} and $F_{ROH\_RZooROH}$ demonstrated distinct patterns. In general, the X chromosome showed higher mean F_{ROH} values compared to autosomes and PAR. Specifically, in females, mean value on the X chromosome ranged from 0.124 in Nellore to 0.422 in Soay (F_{ROH_SVS}), and from 0.176 in Patagonian Sheepdog to 0.453 in Soay ($F_{ROH\_RZooROH}$). Corresponding autosomal values were notably lower, ranging from 0.046 in Nellore to 0.255 in Soay (F_{ROH_SVS}), and from 0.101 in Patagonian Sheepdog to 0.321 in Soay ($F_{ROH\_RZooROH}$), as well as the PAR values, which ranged from 0.013 in Nellore to 0.158 in Labrador Retriever (F_{ROH_SVS}), and from 0.017 in Nellore to 0.239 in Soay ($F_{ROH\_RZooROH}$). In males, similar patterns are observed, with autosomal inbreeding ranging from 0.044 in Croatian sheep breeds to 0.253 in Soay (F_{ROH_SVS}), and from 0.099 in Patagonian Sheepdog to 0.319 in Soay ($F_{ROH\_ RZooROH}$), and PAR inbreeding showing variations from 0.023 in Croatian sheep breeds to 0.142 in Croatian cattle breeds (F_{ROH_SVS}), and from 0.047 in Patagonian Sheepdog to 0.259 in Soay ($F_{ROH\_ RZooROH}$). Table 5 further elucidates these differences by expressing the ratio of mean F_{ROH} values between the X chromosome, autosomes and PAR in each population. For example, in Nellore females, the F_{ROH_SVS} ratio of the X chromosome to autosomes indicates that inbreeding at the X chromosome was approximately 2.7 times higher. Similarly, the ratio of the X chromosome to PAR was 9.531:1, highlighting an even more pronounced difference. This demonstrates a varied pattern of F_{ROH} across different genomic regions and between sexes, highlighting the unique genetic dynamics in each population.

The mean values and SE of additional F_{ROH_SVS} and $F_{ROH\_RZooROH}$ coefficients, for different minimum ROH lengths and HBD categories respectively, estimated on autosomes, X chromosome and PAR of each domestic animal population are presented separately for females in Appendices 1 and 3, and for males in Appendices 2 and 4.

Table 3. Mean values with associated standard errors (SE) of genomic inbreeding coefficients estimated on autosomes, X chromosome and PAR in females of each domestic animal population.

Population	Genomic landscape	F_{ROH_SVS}	$F_{ROH\_RZooROH}$	F_{VR1}	F_{LH1}	F_{YA2}
Croatian cattle breeds	Autosomes	0.087±0.009	0.211±0.008	0.057±0.007	0.058±0.007	0.050±0.010
	X chromosome	0.255±0.013	0.264±0.012	0.053±0.015	0.053±0.006	0.020±0.018
	PAR	0.065±0.023	0.110±0.022	0.051±0.021	0.051±0.019	0.046±0.024
Nellore	Autosomes	0.046±0.001	0.148±0.000	-0.007±0.001	-0.006±0.000	-0.003±0.001
	X chromosome	0.124±0.002	0.236±0.002	-0.009±0.003	0.003±0.002	-0.003±0.002
	PAR	0.013±0.002	0.017±0.001	0.002±0.004	0.002±0.002	0.003±0.003
Labrador Retriever	Autosomes	0.213±0.006	0.220±0.005	0.048±0.006	0.048±0.006	0.044±0.007
	X chromosome	0.396±0.012	0.272±0.013	0.064±0.019	0.064±0.014	0.066±0.018
	PAR	0.158±0.020	0.172±0.019	0.048±0.021	0.049±0.016	0.055±0.024
Patagonian Sheepdog	Autosomes	0.109±0.009	0.101±0.009	0.026±0.009	0.026±0.008	0.024±0.010
	X chromosome	0.367±0.017	0.176±0.021	0.045±0.025	0.045±0.019	0.034±0.023
	PAR	0.041±0.020	0.044±0.018	0.007±0.023	0.007±0.021	-0.001±0.026
Croatian sheep breeds	Autosomes	0.062±0.007	0.137±0.006	0.060±0.007	0.060±0.006	0.058±0.007
	X chromosome	0.144±0.014	0.226±0.013	0.080±0.015	0.080±0.015	0.078±0.015
	PAR	0.057±0.018	0.097±0.018	0.074±0.019	0.073±0.018	0.071±0.019
Soay	Autosomes	0.255±0.002	0.321±0.002	-0.010±0.003	-0.010±0.002	-0.009±0.003
	X chromosome	0.422±0.009	0.453±0.009	-0.021±0.015	-0.021±0.009	-0.015±0.015
	PAR	0.112±0.014	0.239±0.014	-0.042±0.022	-0.042±0.015	-0.034±0.021

Table 4. Mean values with associated standard errors (SE) of genomic inbreeding coefficients estimated on autosomes and PAR in males of each domestic animal population.

Population	Genomic landscape	F_{ROH_SVS}	$F_{ROH\_RZooROH}$	F_{VR1}	F_{LH1}	F_{YA2}
Croatian cattle breeds	Autosomes	0.136±0.017	0.242±0.014	0.070±0.017	0.070±0.012	0.056±0.018
	PAR	0.142±0.038	0.173±0.036	0.081±0.042	0.081±0.031	0.062±0.042
Nellore	Autosomes	0.072±0.001	0.166±0.001	0.003±0.002	0.003±0.001	0.004±0.001
	PAR	0.029±0.004	0.100±0.004	-0.019±0.006	-0.019±0.004	-0.014±0.005
Labrador Retriever	Autosomes	0.199±0.004	0.206±0.004	0.030±0.005	0.036±0.005	0.029±0.005
	PAR	0.091±0.012	0.118±0.012	-0.013±0.015	-0.013±0.011	-0.014±0.015
Patagonian Sheepdog	Autosomes	0.107±0.006	0.099±0.006	0.029±0.007	0.029±0.007	0.028±0.007
	PAR	0.039±0.013	0.047±0.013	0.024±0.016	0.024±0.014	0.024±0.015
Croatian sheep breeds	Autosomes	0.044±0.005	0.119±0.005	0.043±0.005	0.044±0.005	0.042±0.005
	PAR	0.023±0.011	0.051±0.012	0.028±0.012	0.028±0.011	0.028±0.012
Soay	Autosomes	0.253±0.002	0.319±0.002	-0.014±0.003	-0.014±0.002	-0.013±0.002
	PAR	0.124±0.018	0.259±0.016	-0.030±0.025	-0.030±0.017	-0.018±0.024

Table 5. Ratio between mean F_{ROH} values (lower value scaled to 1) estimated on autosomes, X chromosome and PAR in each domestic animal population, stratified by method and sex.

F_{ROH}	Ratio	Croatian cattle breeds	Nellore	Labrador Retriever	Patagonian Sheepdog	Croatian sheep breeds	Soay
F_{ROH_SVS}	chrX_F:Auto_F	2.931:1.000	2.696:1.000	1.859:1.000	3.367:1.000	2.323:1.000	1.655:1.000
	chrX_F:PAR_F	3.923:1.000	9.531:1.000	2.506:1.000	8.951:1.000	2.526:1.000	9.768:1.000
	PAR_F:Auto_F	1.000:1.338	1.000:3.538	1.000:1.348	1.000:2.659	1.000:1.088	1.000:2.277
	PAR_M:Auto_M	1.044:1.000	1.000:2.483	1.000:2.187	1.000:2.744	1.000:1.913	1.000:2.040
	PAR_F:PAR_M	1.000:2.185	1.000:2.231	1.736:1.000	1.051:1.000	2.428:1.000	1.000:1.107
	Auto_F:Auto_M	1.000:1.563	1.000:1.565	1.070:1.000	1.019:1.000	1.409:1.000	1.008:1.000
$F_{ROH\_RZooROH}$	chrX_F:Auto_F	1.251:1.000	1.595:1.000	1.236:1.000	1.743:1.000	1.650:1.000	1.411:1.000
	chrX_F:PAR_F	2.400:1.000	13.882:1.000	1.581:1.000	4.000:1.000	2.330:1.000	1.895:1.000
	PAR_F:Auto_F	1.000:1.918	1.000:8.706	1.000:1.279	1.000:2.295	1.000:1.412	1.000:1.343
	PAR_M:Auto_M	1.000:1.399	1.000:1.660	1.000:1.746	1.000:2.106	1.000:2.333	1.000:1.232
	PAR_F:PAR_M	1.000:1.573	1.000:5.882	1.458:1.000	1.000:1.068	1.902:1.000	1.000:1.084
	Auto_F:Auto_M	1.000:1.147	1.000:1.122	1.068:1.000	1.020:1.000	1.151:1.000	1.066:1.000

chrX_F = X chromosome_Females, Auto_F = Autosomes_Females, Auto_M = Autosomes_Males, PAR_F = PAR_Females, PAR_M = PAR_Males.

To show the variability of each estimated genomic inbreeding coefficient, boxplots showing the distribution of estimates on autosomes, X chromosome and PAR separated by sex in each species are presented in Figures 11, 12 and 13. Almost all X chromosome and PAR inbreeding coefficients showed greater variation compared to the autosome estimates. At the same time, the median values on the X chromosome were higher than on the autosomes and PAR in the case of F_{ROH_SVS} and $F_{ROH\_RZooROH}$, while they were approximately the same in the case of F_{VR1}, F_{LH1} and F_{YA2}.

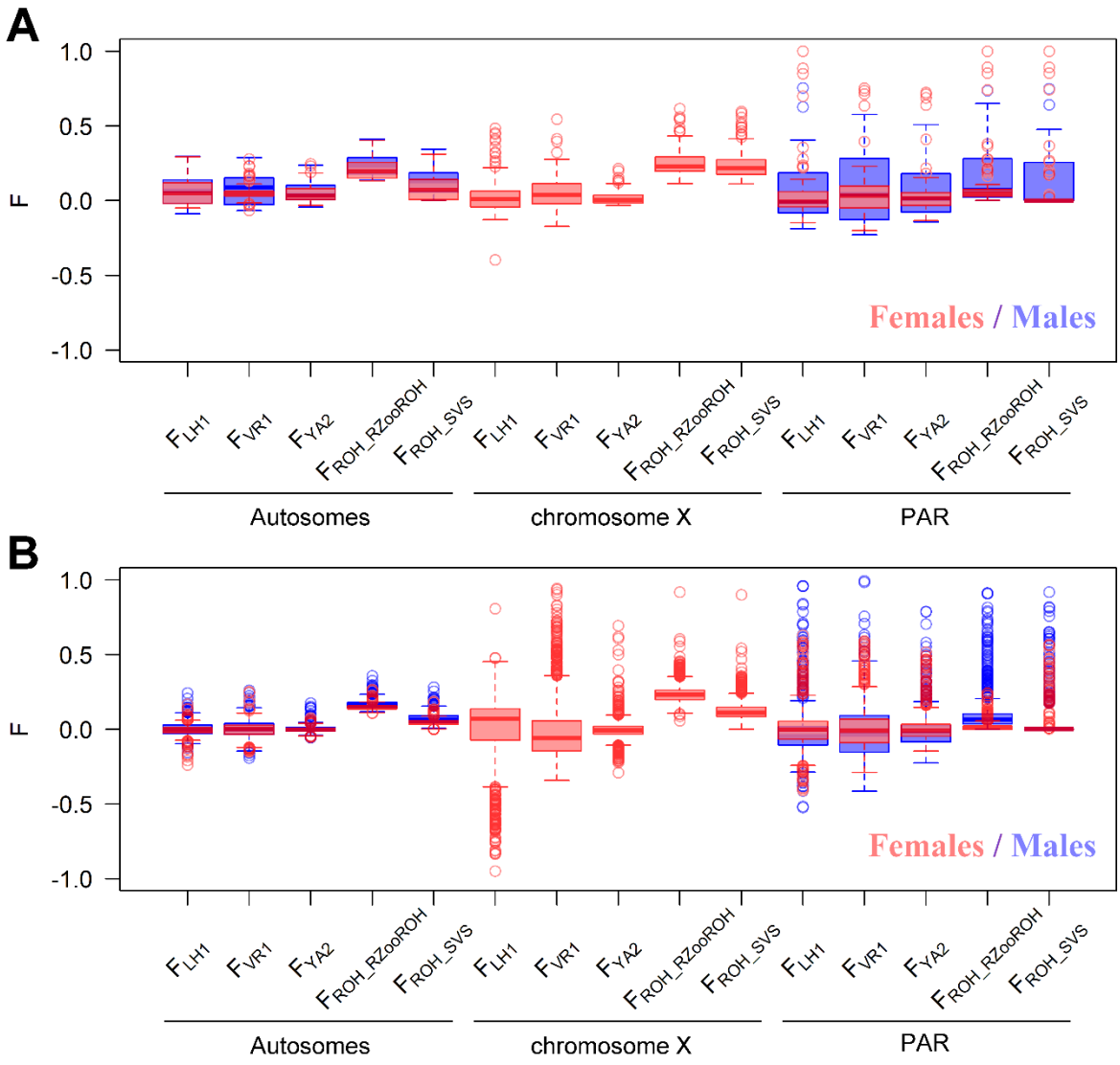


Figure 11. Distribution of genomic inbreeding coefficients estimated on autosomes, X chromosome and PAR in metapopulation of native Croatian cattle breeds **(A)** and Nellore **(B)** representing cattle, separated by sex.

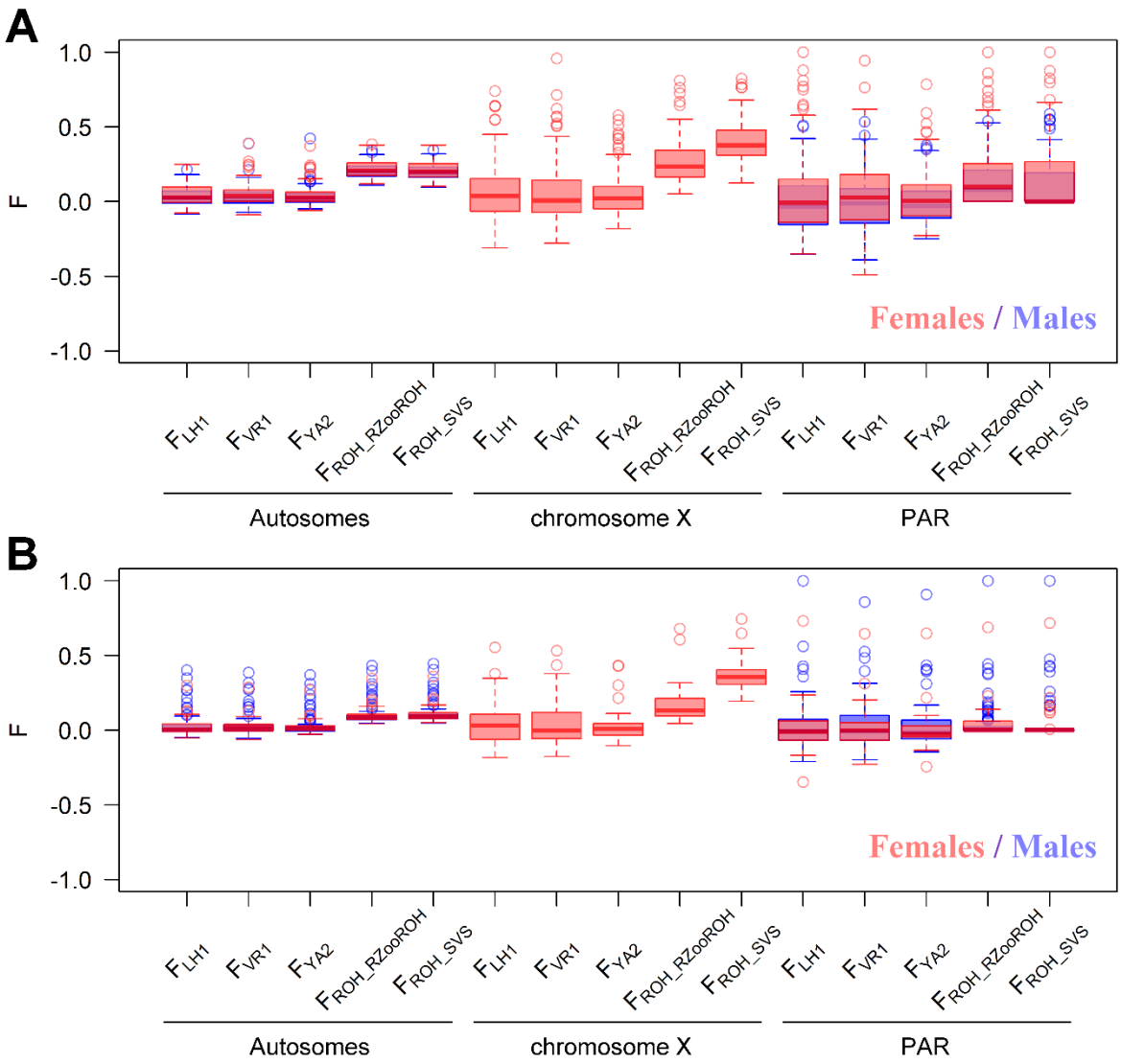


Figure 12. Distribution of genomic inbreeding coefficients estimated on autosomes, X chromosome and PAR in Labrador Retriever **(A)** and Patagonian Sheepdog **(B)** representing dogs, separated by sex.

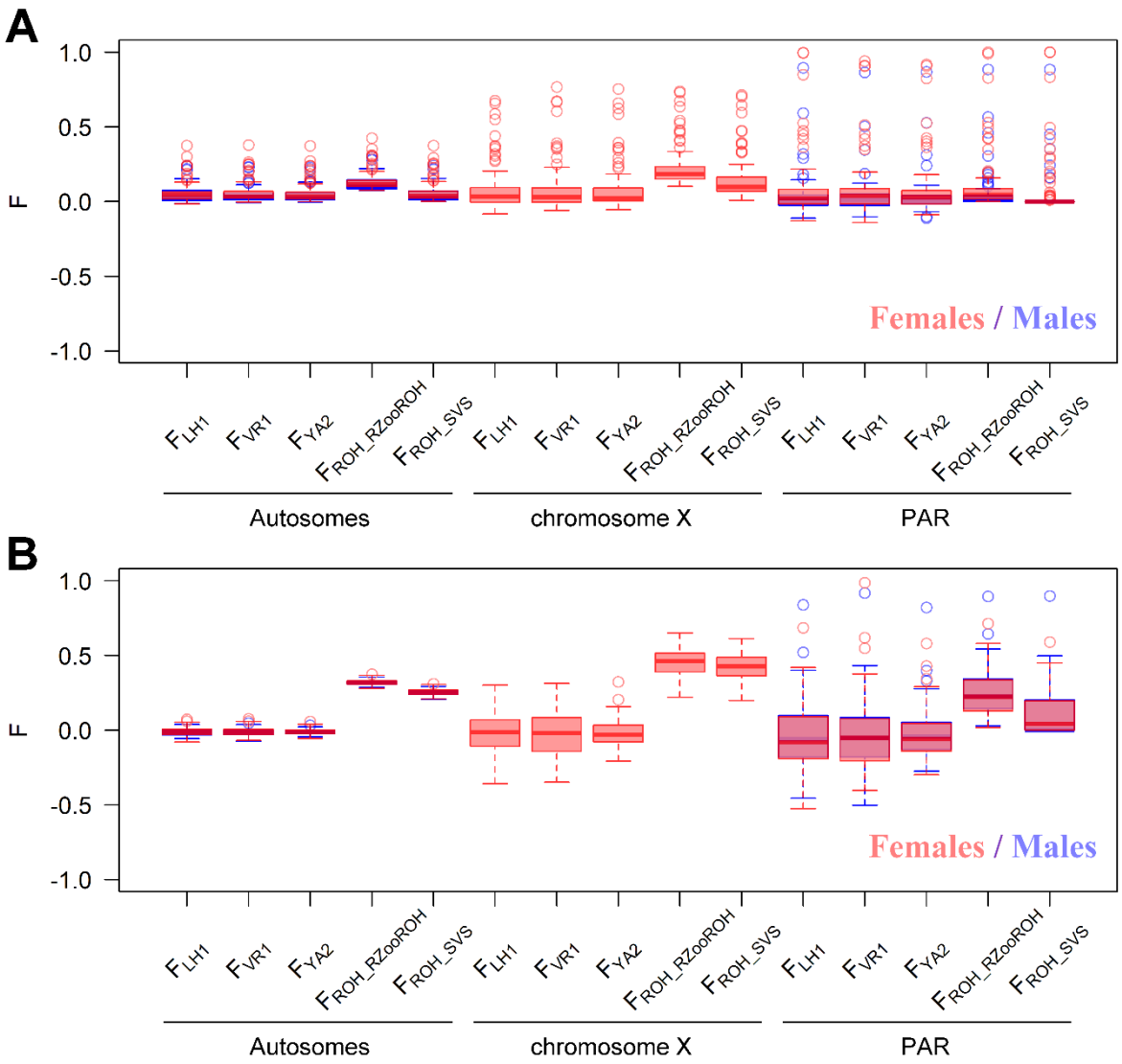


Figure 13. Distribution of genomic inbreeding coefficients estimated on autosomes, X chromosome and PAR in metapopulation of native Croatian sheep breeds **(A)** and Soay **(B)** representing sheep, separated by sex.

In addition, boxplots showing the distribution of F_{ROH_SVS} values estimated using different minimum ROH lengths and $F_{ROH\_RZooROH}$ values divided by each HBD category across different genomic landscapes in males and females of each domestic animal population are presented in Appendices 5-10.

4.2.1. Assesment and comparison of regional F_{ROH}

Regional inbreeding variation (F_{ROH_SVS} and $F_{ROH\_RZooROH}$) on the X chromosome with a window size equal to PAR size and boxplot visualisation of regional estimates for each chromosome separated by sex, are shown for each domestic animal population in Figures 14-19.

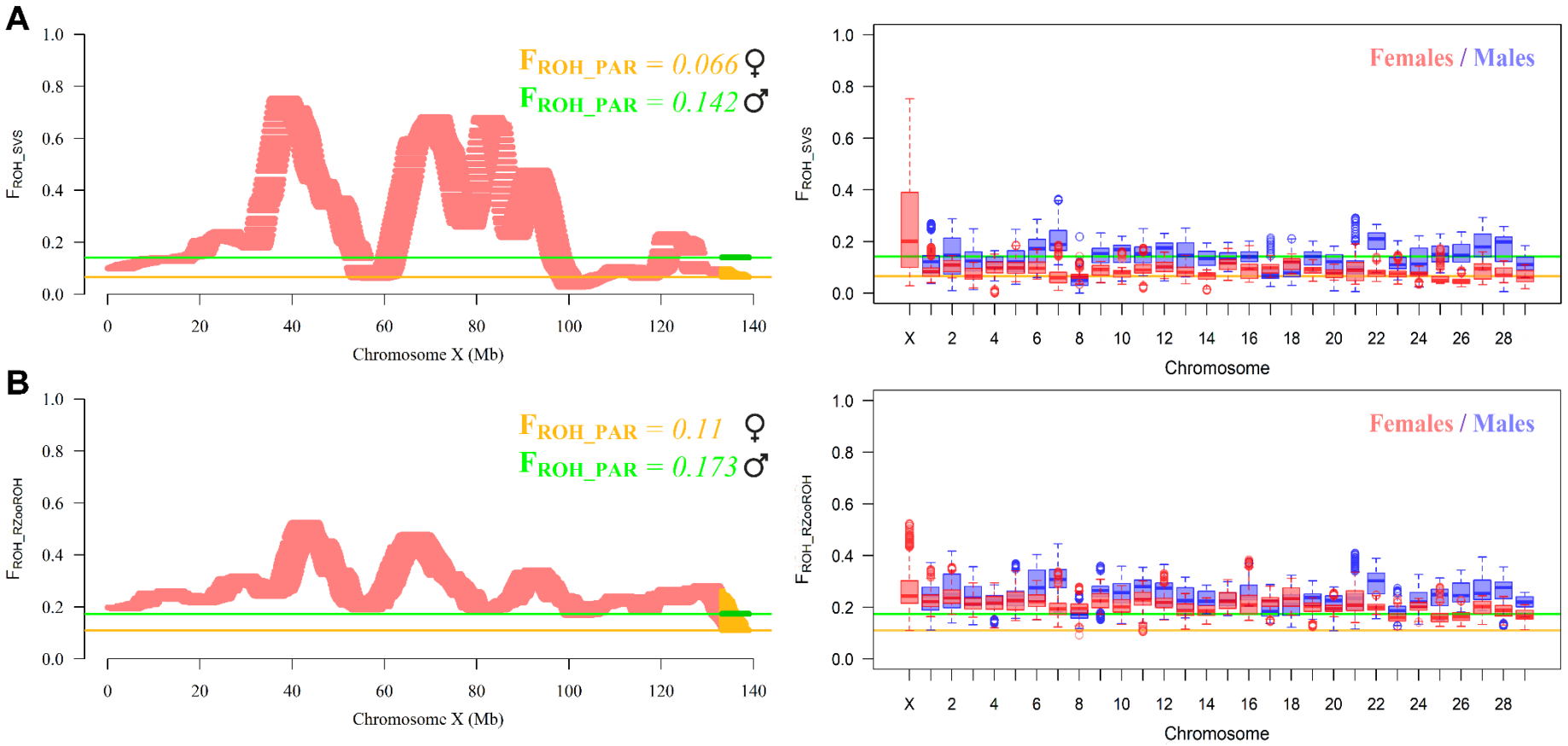


Figure 14. Regional inbreeding variation (F_{ROH}) on the X chromosome and distribution of regional estimates for each chromosome separated by sex, estimated with SVS **(A)** and RZooROH **(B)** in metapopulation of native Croatian cattle breeds.

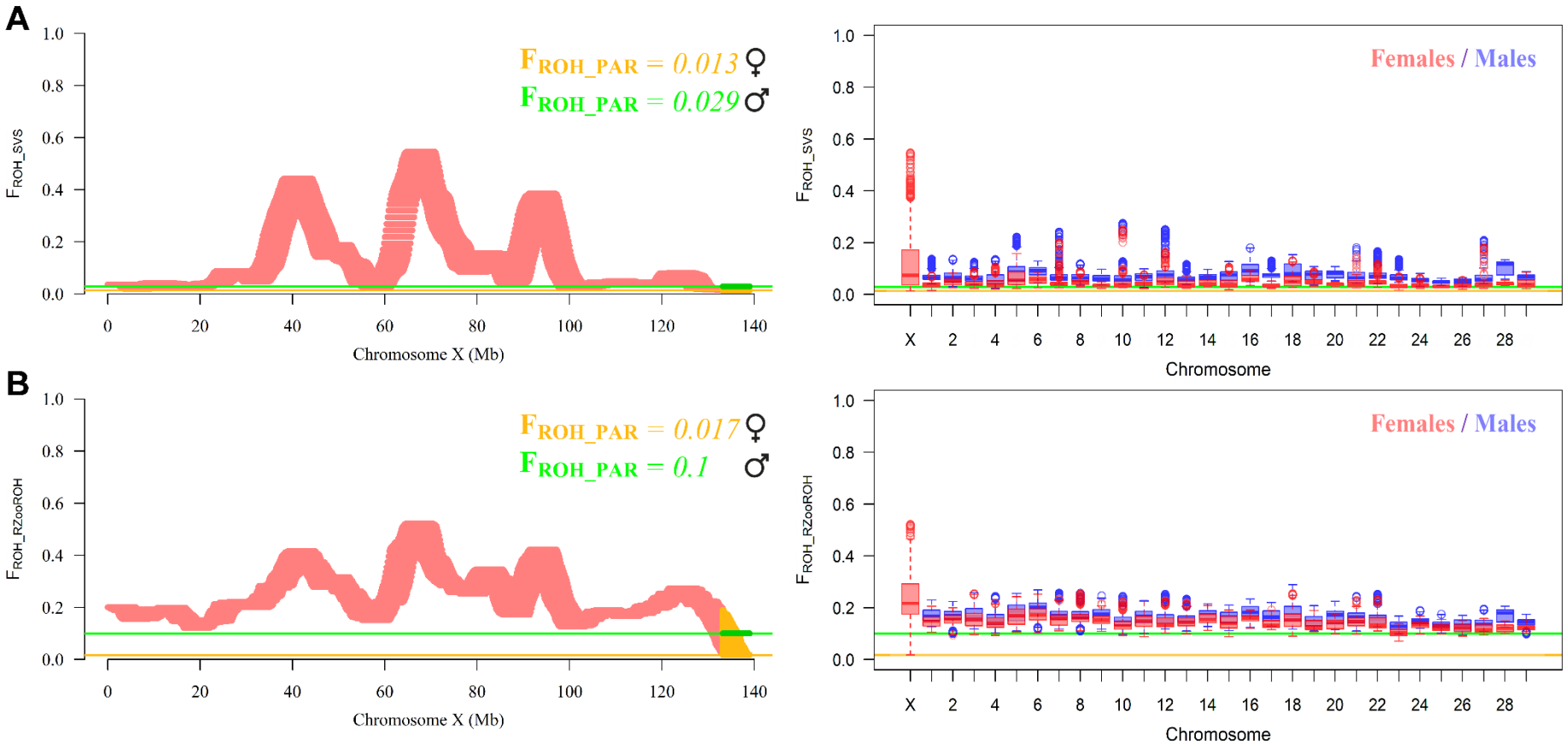


Figure 15. Regional inbreeding variation (F_{ROH}) on the X chromosome and distribution of regional estimates for each chromosome separated by sex, estimated with SVS **(A)** and RZooROH **(B)** in Nellore.

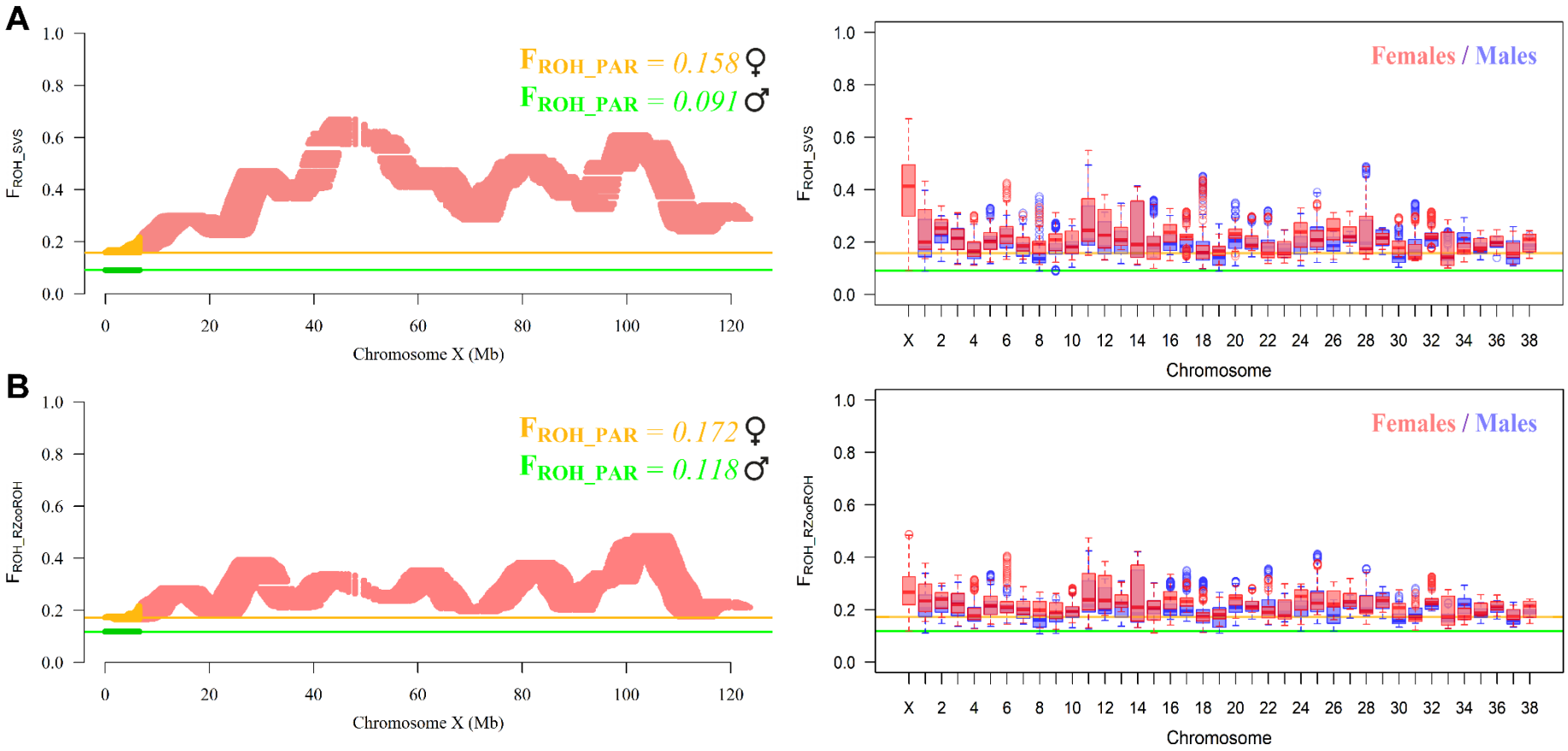


Figure 16. Regional inbreeding variation (F_{ROH}) on the X chromosome and distribution of regional estimates for each chromosome separated by sex, estimated with SVS **(A)** and RZooROH **(B)** in Labrador Retriever.

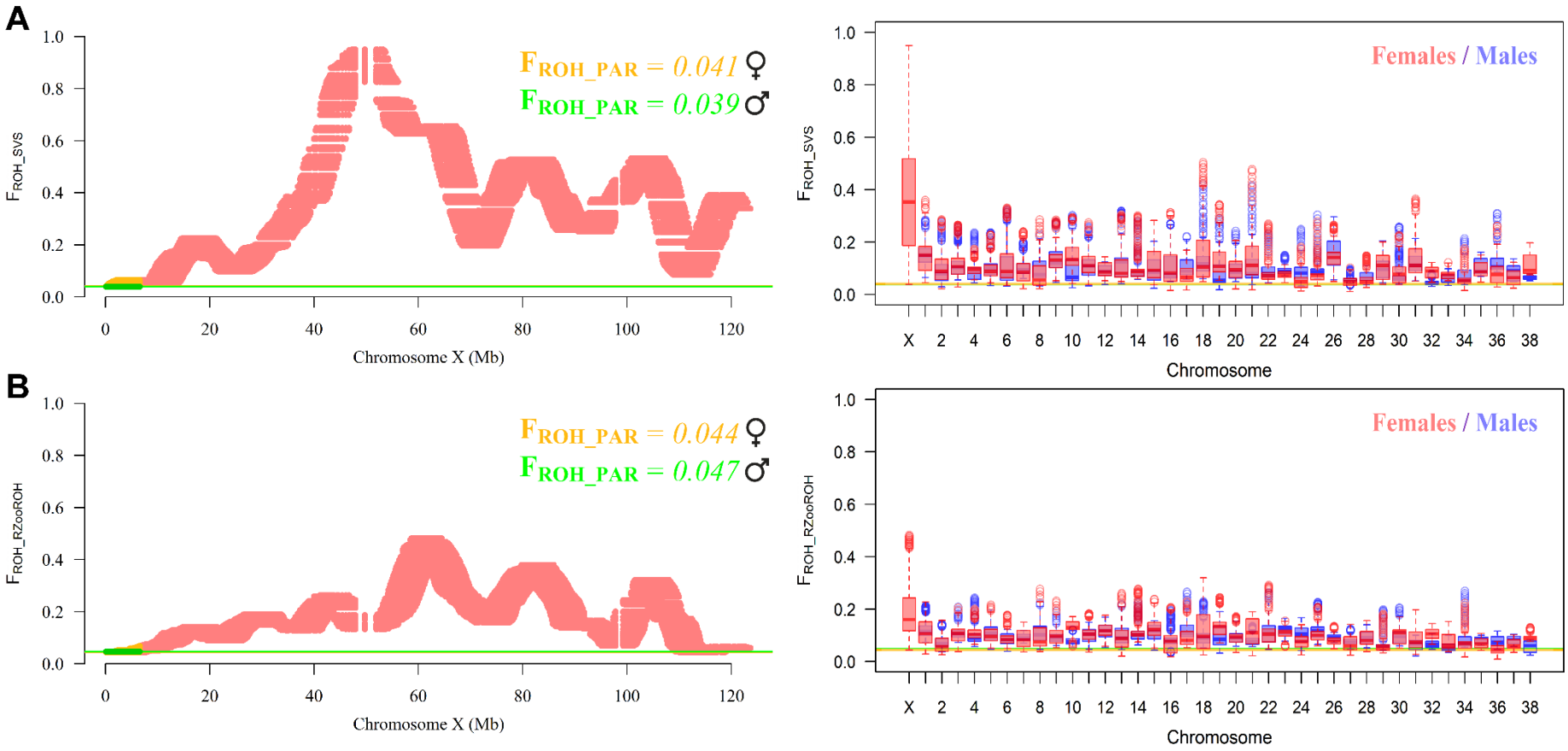


Figure 17. Regional inbreeding variation (F_{ROH}) on the X chromosome and distribution of regional estimates for each chromosome separated by sex, estimated with SVS **(A)** and RZooROH **(B)** in Patagonian Sheepdog.

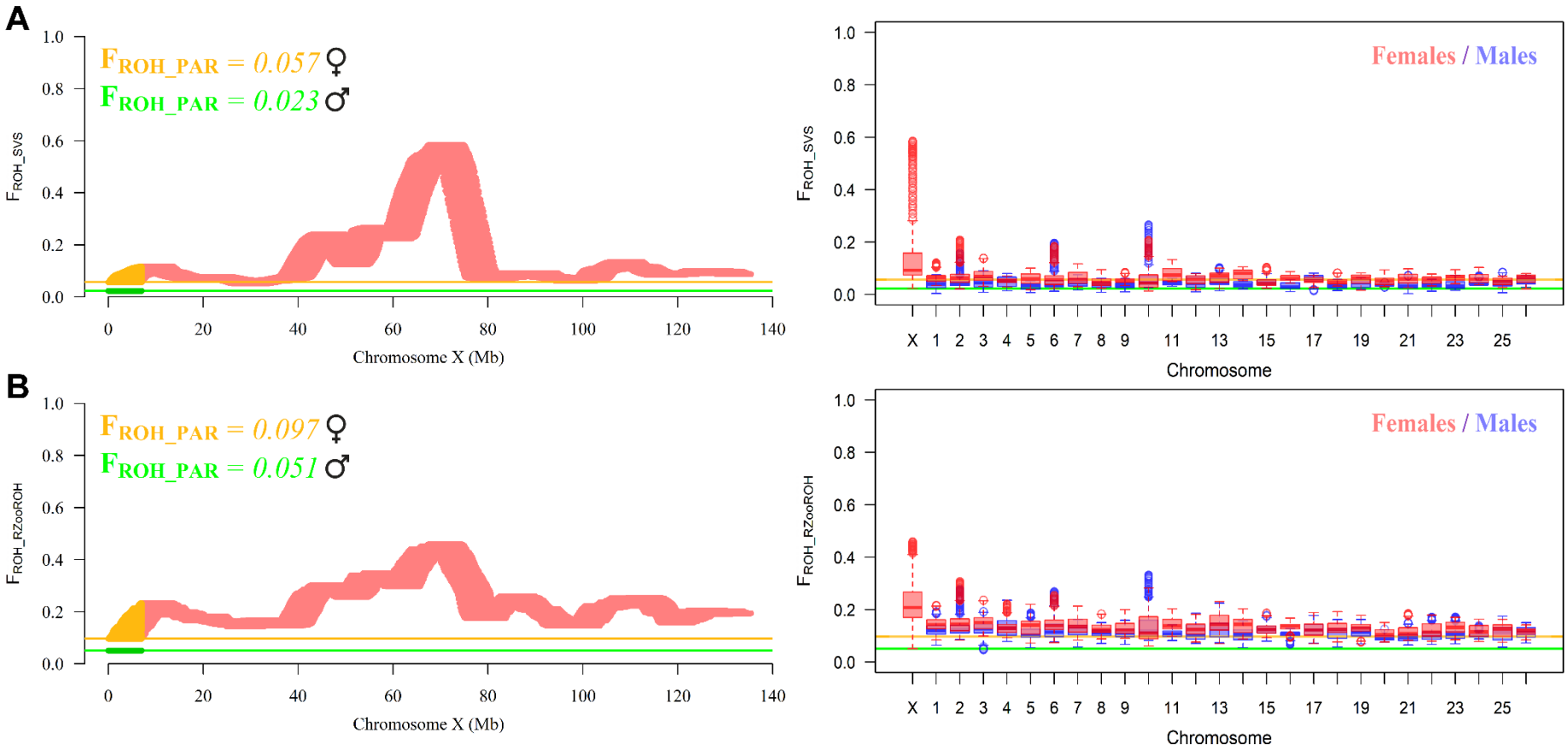


Figure 18. Regional inbreeding variation (F_{ROH}) on the X chromosome and distribution of regional estimates for each chromosome separated by sex, estimated with SVS **(A)** and RZooROH **(B)** in metapopulation of native Croatian sheep breeds.

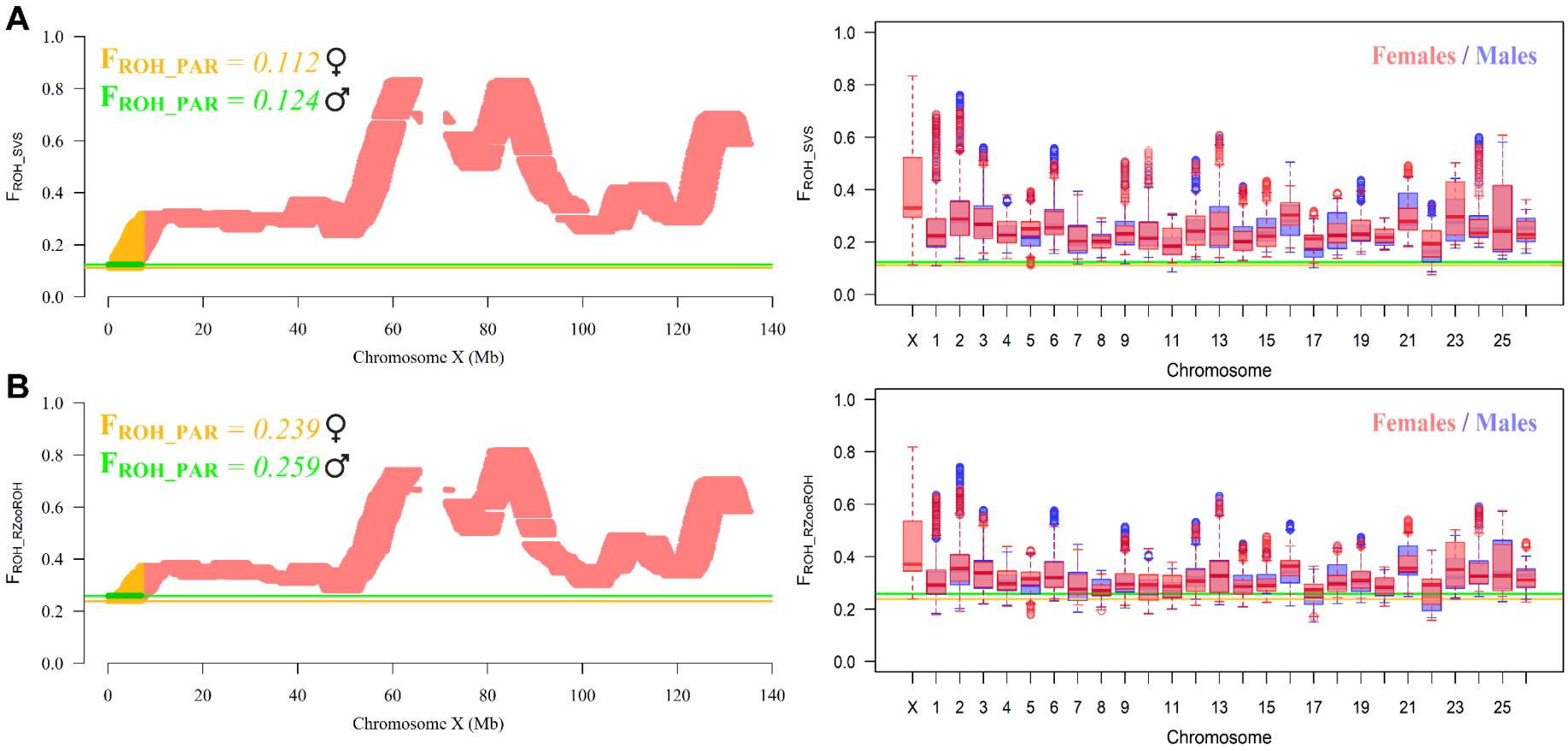


Figure 19. Regional inbreeding variation (F_{ROH}) on the X chromosome and distribution of regional estimates for each chromosome separated by sex, estimated with SVS **(A)** and RZooROH **(B)** in Soay.

Overall, specific patterns were observed on the X chromosome in both F_{ROH_SVS} and $F_{ROH\_RZooROH}$, which exhibited greater regional variability than the autosomes. The F_{ROH_SVS} estimates on the X chromosome were generally more variable than those of $F_{ROH\_RZooROH}$, except in Nellore and Croatian sheep breeds, where they were roughly equivalent but exhibited more outliers. Simultaneously, the median F_{ROH_SVS} and $F_{ROH\_RZooROH}$ values on the X chromosome were higher than those on all other chromosomes, except in Soay, where they matched the median values of autosomes 16, 21 and 23. When comparing median F_{ROH_SVS} and $F_{ROH\_RZooROH}$ values on the X chromosome, F_{ROH_SVS} values were much higher in dogs, while $F_{ROH\_RZooROH}$ values were much higher in Nellore and Croatian sheep breeds.

F_{ROH_SVS} and $F_{ROH\_RZooROH}$ within PAR showed different patterns across populations when compared between sexes and with other X chromosomal and autosomal regional estimates. In cattle, values within PAR were lower in females than in males. Specifically, $F_{ROH\_RZooROH}$ was lowest in females within PAR in Nellore compared to all genomic windows, lowest in Croatian cattle breeds compared to the rest of the X chromosome, and among the lowest in Croatian cattle breeds compared to autosomal estimates. However, there were regions on the rest of the X chromosome as well as on the autosomes where F_{ROH_SVS} values were lower. In dogs, PAR values were higher in females than males in Labrador Retriever, while similar values were observed between the sexes in Patagonian Sheepdog. When PAR was compared with the rest of the X chromosome and autosomes, F_{ROH_SVS} exibited lower values than $F_{ROH\_RZooROH}$, although some autosomal regions still had lower values for both coefficients. In sheep, the PAR typically showed the lowest inbreeding levels compared to the rest of the X chromosome. In Croatian sheep breeds, inbreeding was higher in females than in males, whereas in Soay, it was the reverse. Notably, some autosomal regions exhibited lower inbreeding levels than those found within PAR.

4.2.2. Pearson correlation (*r*) between estimated inbreeding coefficients

Pearson correlations (*r*) among various estimated inbreeding coefficients across autosomes, the X chromosome, and PAR are presented, separately by sex for each population, in Figures 20-25.

Generally, the lowest *r* values were observed between different genomic landscapes for both sexes. In females, *r* values between the same coefficients estimated on the X chromosome and autosomes ranged from 0.35 to 0.53 for Croatian cattle breeds, 0.01 to 0.33 for Nellore, 0.26 to 0.48 for Labrador Retriever, 0.35 to 0.43 for Patagonian Sheepdog, 0.66 to 0.68 for Croatian sheep breeds, and -0.07 to 0.06 for Soay. Between

the X chromosome and PAR, *r* values varied from 0.26 to 0.41 in Croatian cattle breeds, 0.09 to 0.28 in Nellore, 0.27 to 0.35 in Labrador Retriever, 0.17 to 0.41 in Patagonian Sheepdog, 0.47 to 0.50 in Croatian sheep breeds, and from -0.07 to 0.15 in Soay. Between PAR and autosomes, *r* values varied from 0.09 to 0.22 in Croatian cattle breeds, -0.04 to 0.32 in Nellore, 0.31 to 0.35 in Labrador Retriever, 0.27 to 0.45 in Patagonian Sheepdog, 0.38 to 0.41 in Croatian sheep breeds, and -0.06 to -0.02 in Soay. At the same time, in males, *r* values among the same coefficients estimated between PAR and autosomes ranged from 0.39 to 0.60 for Croatian cattle breeds, 0.05 to 0.24 for Nellore, 0.08 to 0.19 for Labrador Retriever, 0.38 to 0.49 for Patagonian Sheepdog, 0.05 to 0.41 for Croatian sheep breeds, and -0.11 to -0.04 for Soay.

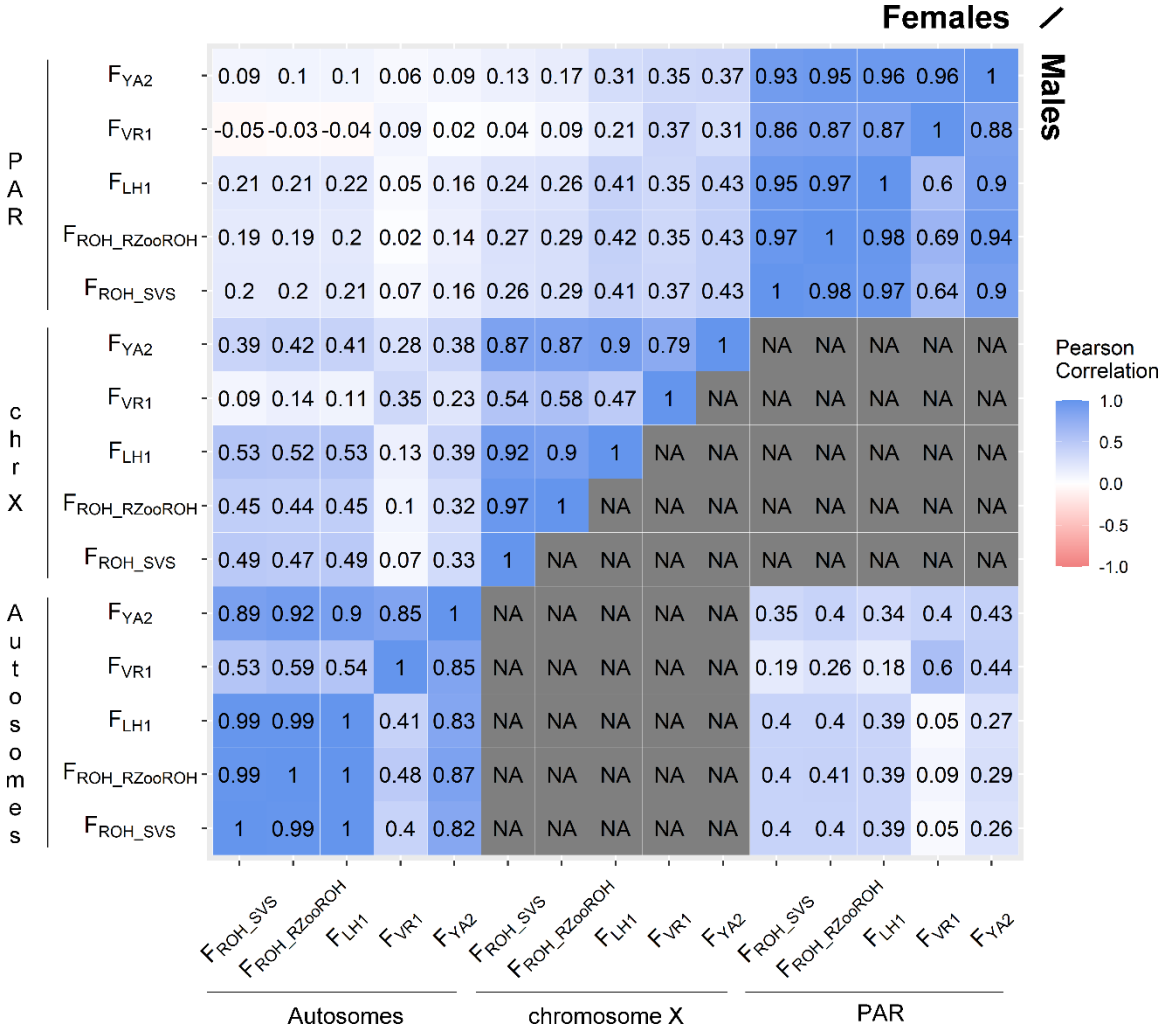


Figure 20. Pearson correlations (*r*) between estimated genomic inbreeding coefficients for males (below diagonal) and females (above diagonal) in metapopulation of native Croatian cattle breeds.

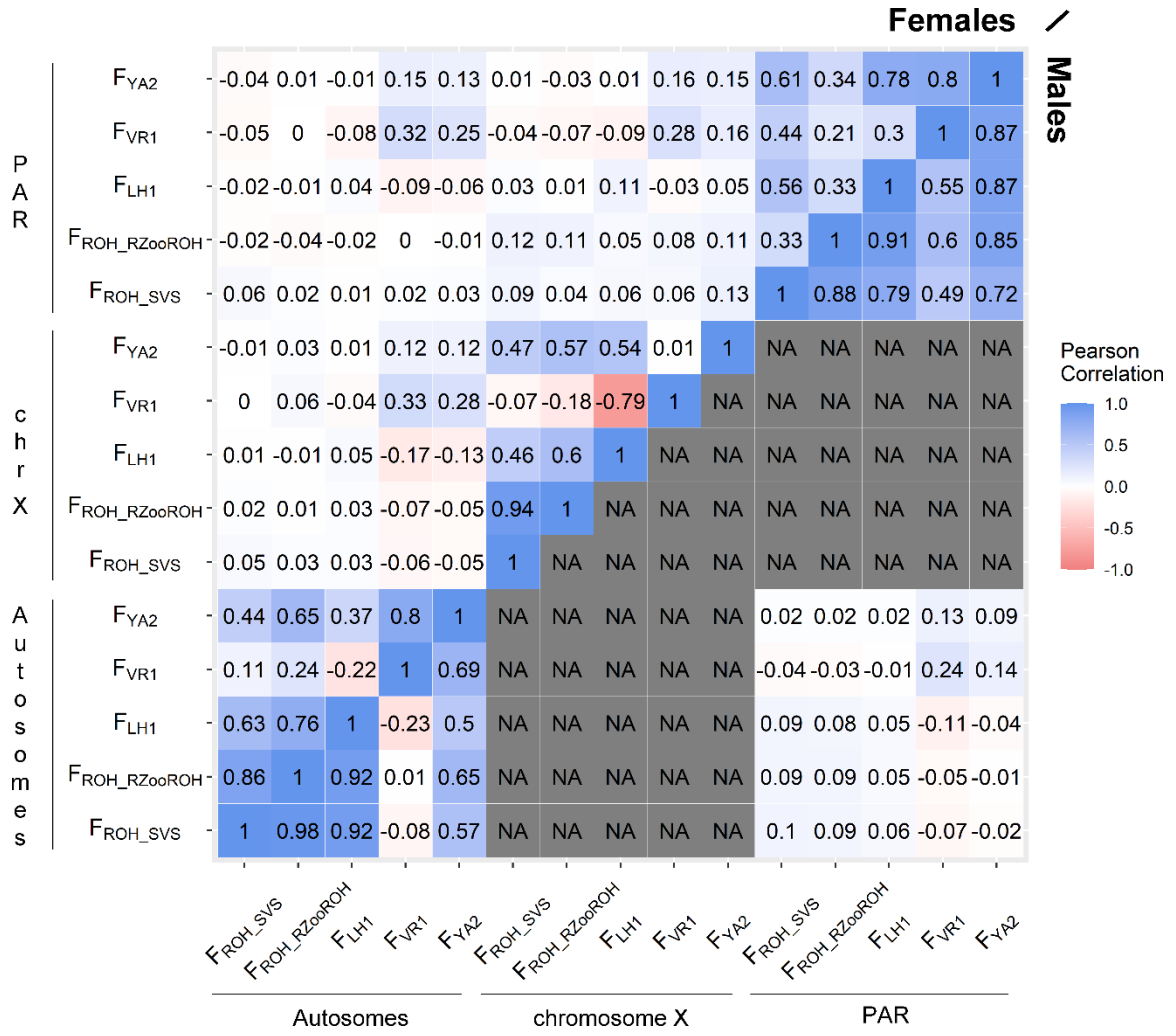


Figure 21. Pearson correlations (*r*) between estimated genomic inbreeding coefficients for males (below diagonal) and females (above diagonal) in Nellore.

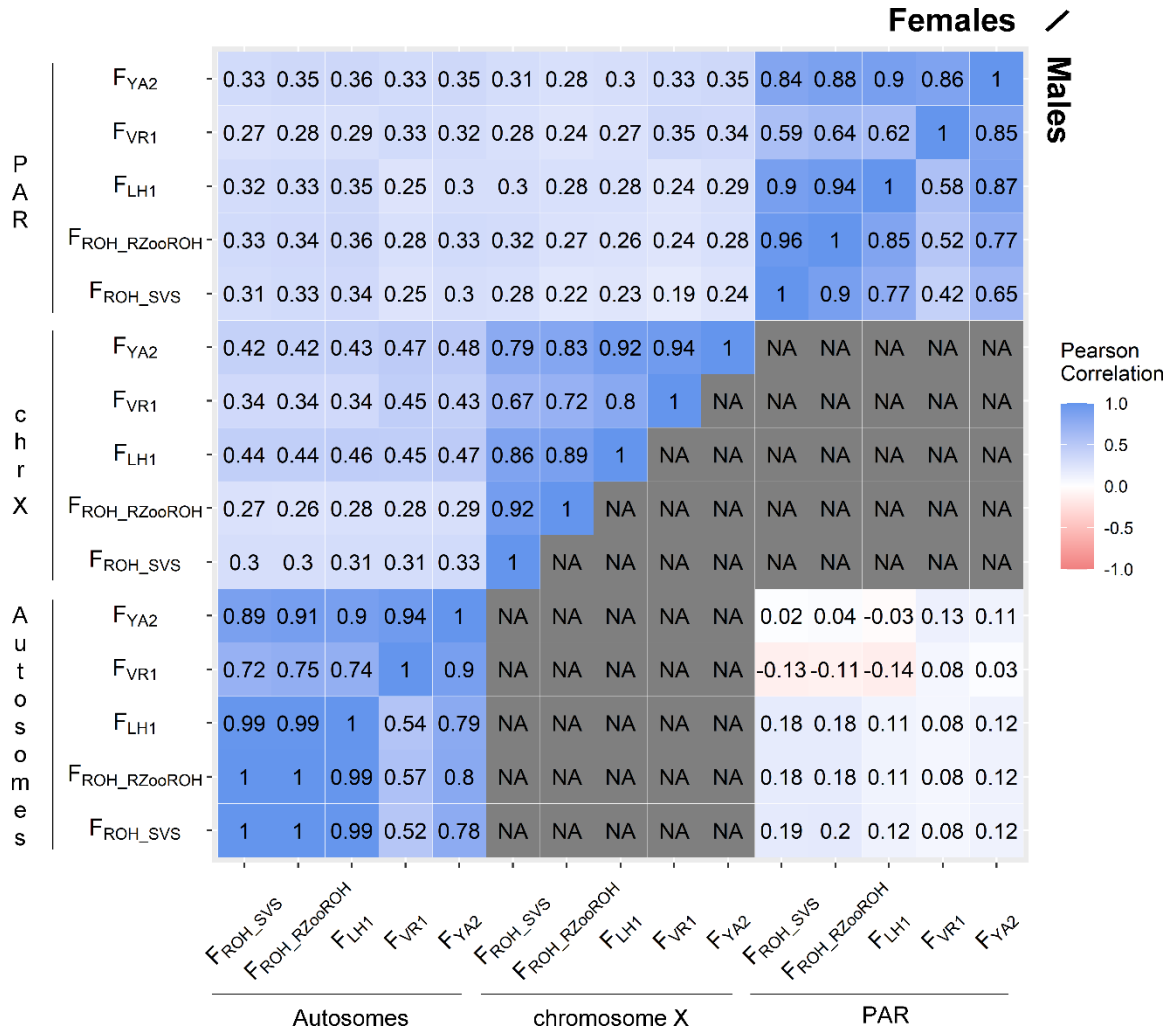


Figure 22. Pearson correlations (*r*) between estimated genomic inbreeding coefficients for males (below diagonal) and females (above diagonal) in Labrador Retriever.

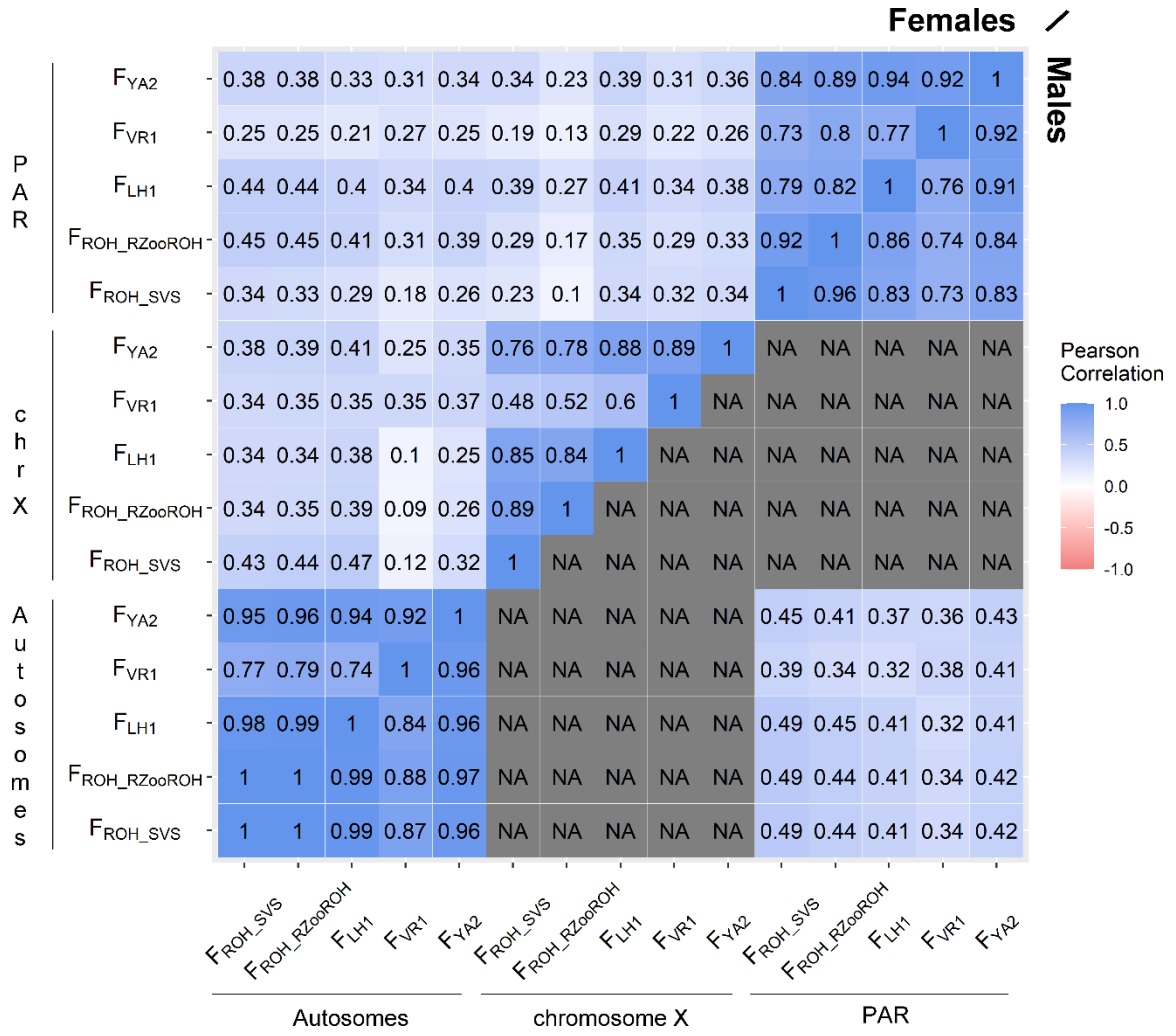


Figure 23. Pearson correlations (*r*) between estimated genomic inbreeding coefficients for males (below diagonal) and females (above diagonal) in Patagonian Sheepdog.

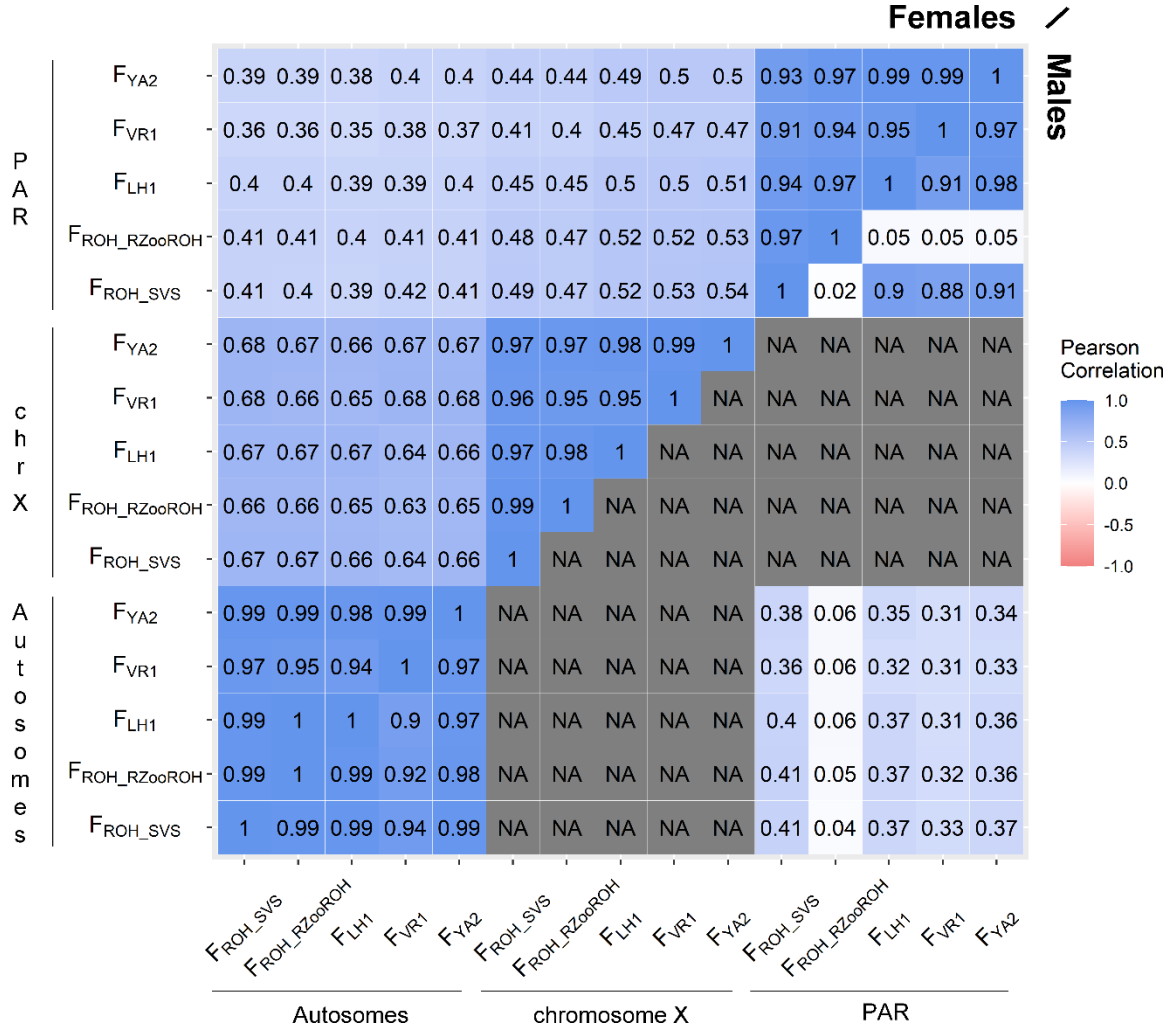


Figure 24. Pearson correlations (*r*) between estimated genomic inbreeding coefficients for males (below diagonal) and females (above diagonal) in metapopulation of native Croatian sheep breeds.

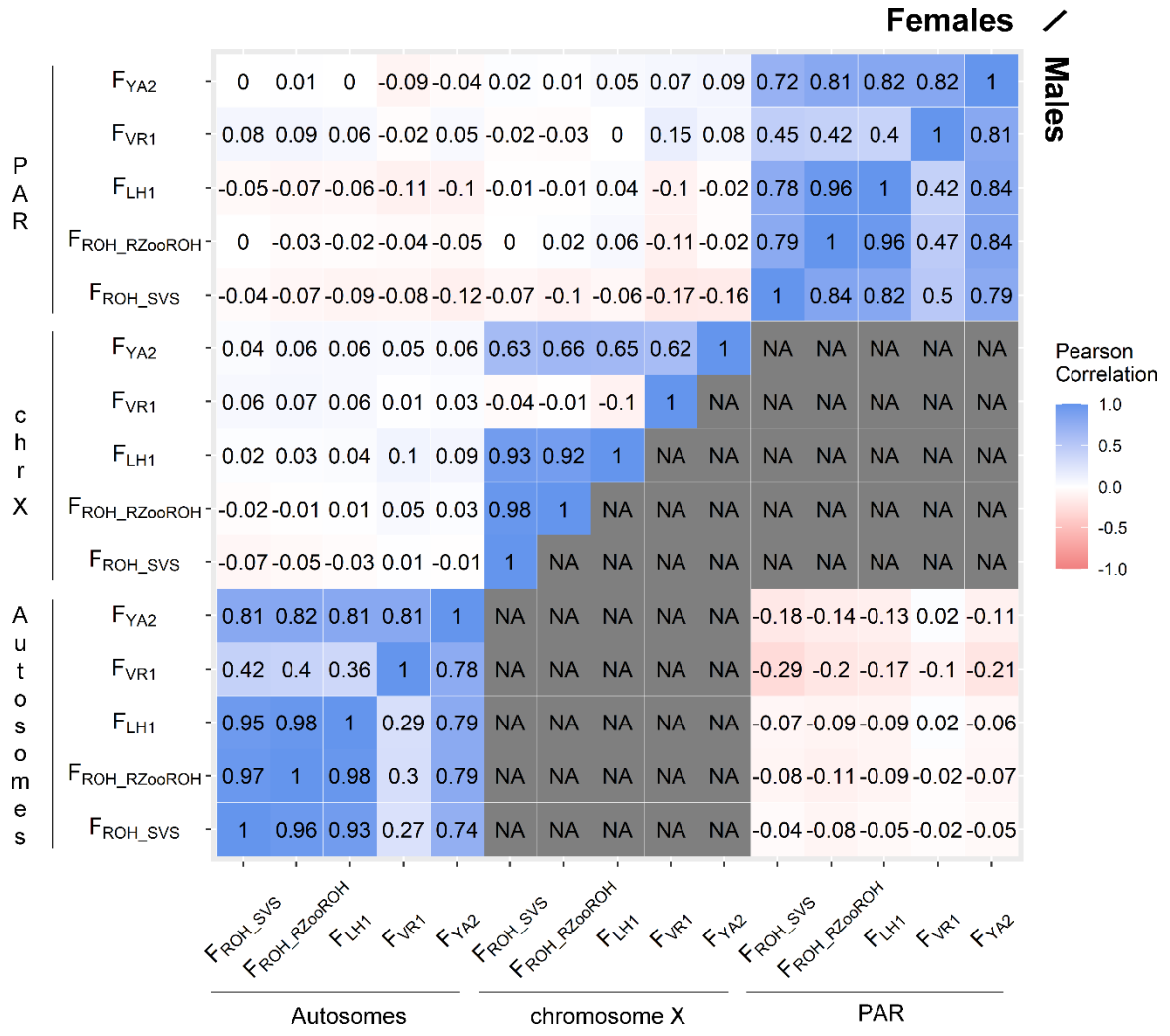


Figure 25. Pearson correlations (*r*) between estimated genomic inbreeding coefficients for males (below diagonal) and females (above diagonal) in Soay.

When comparing inbreeding coefficients estimated within same genomic landscapes in each population, the highest *r* values were noted between F_{ROH_SVS} and $F_{ROH\_RZooROH}$ on autosomes, ranging from 0.96 to 1.00 (with an exception of 0.86 in Nellore females) and on the X chromosome ranging from 0.89 to 0.99. Furthermore, F_{VR1} exhibited the lowest linear relationship with the other coefficients, recording a minimum value of -0.79 with F_{LH1} on the X chromosome in Nellore females.

4.3. Positive selection signals identified on the X chromosome

4.3.1. Signals identified by classical eROHi, iHS and nSL methods

The visualization of positive selection signals on a Manhattan plot, analyzed with three classical methods: eROHi, iHS and nSL, are shown in Figure 26 (A, B and C). Using these methods, ten genomic regions with 12 positive selection signals were identified. The genomic region extending from 13.10 to 13.69 Mb showed the strongest evidence of positive selection, as it was identified by all three methods. This region had the highest −log(P) value of 16.5 in eROHi analysis and the highest proportion of outliers in iHS (17/50) and nSL (35/50) analyses. The results of iHS and nSL were very similar, as five selection signals (from 13.10 to 13.60 Mb, from 32.20 to 32.80 Mb, from 41.00 to 41.50 Mb, from 63.20 to 63.80 Mb, and from 110.10 to 110.80 Mb) were identified by both methods (Figure 26B,C). However, some other signals were identified only by iHS (from 42.50 to 43.00 Mb) or by nSL (from 51.40 to 51.90; from 63.80 to 64.30 Mb; and from 64.60 to 65.10 Mb). One selection signal identified by eROHi (eROHi_4), ranging from 51.63 to 51.94 Mb (Figure 26A), was also identified by nSL method (nSL_w6) (Figure 26C), while three selection signals (from 21.96 to 22.26 Mb, from 83.78 to 84.28 Mb, and from 112.53 to 112.72 Mb) were identified by eROHi only (Figure 26A).

4.3.2. Signals identified by new HRiD method

The visualization of positive selection signals on a Manhattan plot, analyzed with new HRiD method, is shown in Figure 26 (D). HRiD was able to identify four (five if signal located from 73.57 to 74.54 is split into two signals) genomic regions displaying positive selection patterns. The most significant selection signal identified by HRiD with a −log(P) value of 56.5 was located from 13.04 to 13.62 Mb (HRiD_w1) and largely overlapped with those identified by the other three classical methods (eROHi, iHS, and nSL).

This result confirms both the reliability of HRiD in identifying selection signals and the validity of this particular positive selection signal. All other selection signals identified by HRiD (from 56.64 to 58.09 Mb, from 73.57 to 74.20 Mb, from 73.90 to 74.54 Mb, and from 115.30 to 115.73 Mb) were not confirmed by other methods. Due to the methodological approach in defining the window size used by HRiD, the signals from 73.57 to 74.20 Mb and from 73.90 to 74.54 Mb were reported as two separate signals, although it would be more appropriate to consider them as a single extended signal from 73.57 to 74.54 Mb. Part of this signal (HRiD_w4) exhibited the lowest n_h value (1.9), which could be influenced by a low recombination rate. Further detailed information about the significance level of the identified selection signals can be found in Tables 6 and 7.

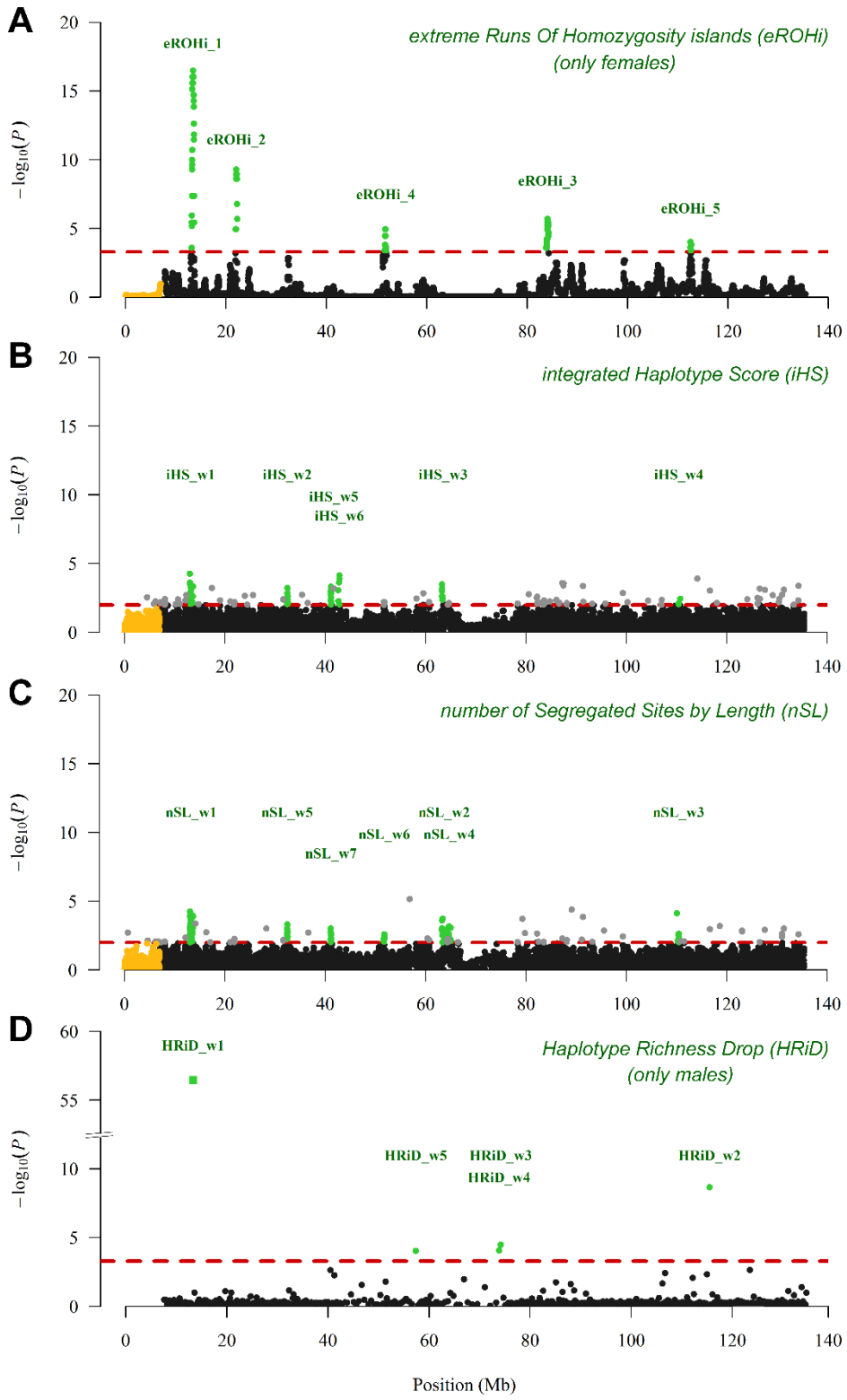


Figure 26. Visualisation of positive selection signals in the Manhattan plot analyzed on the X chromosome (SNPs within PAR are coloured in yellow) using three classical (eROHi, iHS and nSL) and one new (HRiD) method in metapopulation of native Croatian sheep breeds. SNPs or windows above the dashed threshold line (in red) that were considered significant are coloured green, except for single SNP outliers (grey) observed in iHS and nSL analyses.

4.3.3. Phylogenetic relationship within identified signals

The phylogenetic relationships analyzed using MJNs between all male haplotypes (including mouflons representing the ancestors) of each signal identified by HRiD method are highlighted in Figure 27, while MJNs for all mapped signals with respect to their breed origin are shown in Appendices 11 and 12. For HRiD, MJNs were only performed for the four selection signals, as two signals (HRiD_w3 and HRiD_w4) were considered as one signal. Consequently, the haplotypes located within this combined signal were longer (105 SNPs) than the haplotypes (70 SNPs) located within the other three selection signals. The most common haplotype is likely to be the most favorable haplotype selected along with the neighboring haplotypes (assumed here to be no more than three mutations away). Following this concept, the most common mouflon haplotype for the HRiD_w1 selection signal was in the group of favorable haplotypes, indicating that the ancestral haplotype was subject to positive selection (Figure 27A). The same pattern, indicating that the ancestral haplotype was under positive selection, was observed for the merged selection signal HRiD_w3,4 (HRiD_w3 and HRiD_w4) (Figure 27C). In contrast, favorable haplotypes under positive selection in the HRiD_w2 and HRiD_w5 signals were considered derived because they were distant from the ancestral haplotypes present in the mouflons (Figures 27B, D).

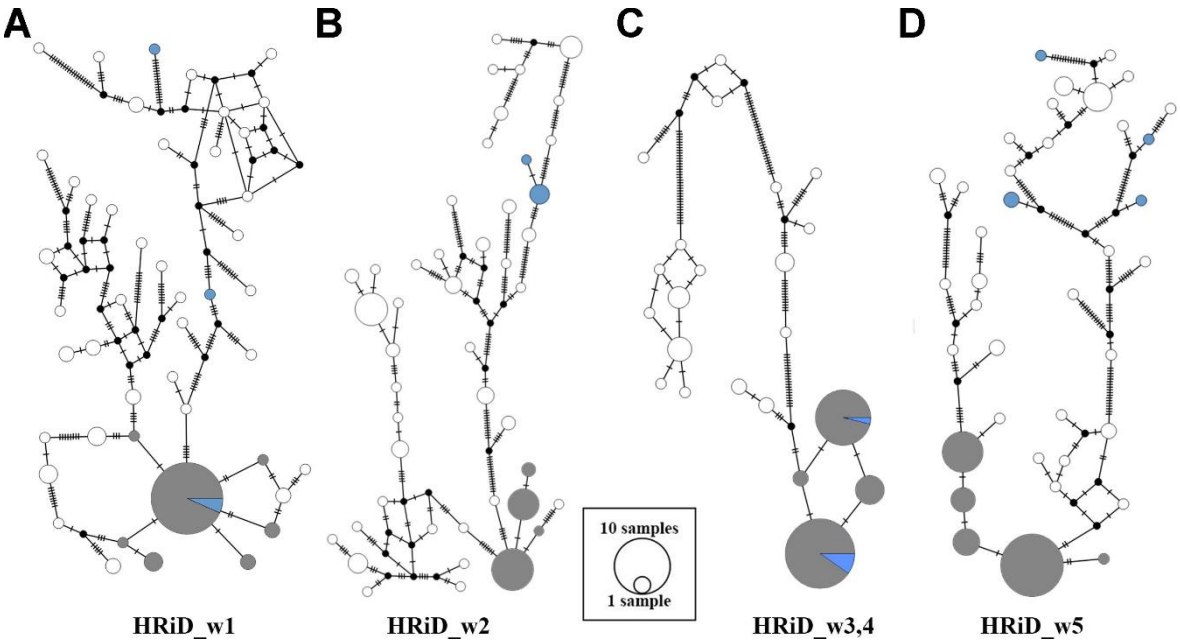


Figure 27. MJNs between ancestral and derived haplotypes within the selection signals identified by HRiD; (**A**) HRiD_w1 (70 SNPs from 13.4 to 13.62 Mb), (**B**) HRiD_w2 (70 SNPs from 115.30 to 115.73), (**C**) HRiD_w3 and HRiD_w4 (105 SNPs from 73.57 and 74.54) and (**D**) HRiD_w5 (70 SNPs from 56.64 and 58.09). The most common haplotypes with adjacent haplotypes no more than three mutations apart are coloured grey, whereas the mouflon haplotypes (representing ancestral haplotypes) are coloured light blue.

4.3.4. Gene annotation and functional characterization of identified signals

The mapping statistics and annotation of genes for identified selection signals on the X chromosome in metapopulation of native Croatian sheep breeds, analyzed by three classical methods and one new method, are detailed in Table 6 and Table 7, respectively. A total of 34 genes in 12 identified regions exhibited patterns of positive selection. Notably, no annotated genes were present in two of these regions: in the region from 32.20 to 32.80 Mb (identified by iHS and nSL), and in the region from 83.78 to 84.28 Mb (identified exclusively by eROHi).

Table 6. Description of mapping statistics and annotation of genes inside positive selection signals identified by three classical methods (eROHi, iHS and nSL) on the X chromosome in metapopulation of native Croatian sheep breeds.

Signal name	Position (Mb)	SNPs*	−log(P)$	Candidate genes #
eROHi_1	13.17−13.69	59/59	16.5	***CA5B***, ***ZRSR2***, ***AP1S2***, ***GRPR***
eROHi_2	21.96−22.26	35/35	9.3	*POLA1*, *ARX*
eROHi_3	83.78−84.28	73/73	5.7	No annotated genes found
eROHi_4	51.63−51.94	33/33	4.9	***DGKK***, ***CCNB3***
eROHi_5	112.53−112.72	11/11	4.0	*PLS3*
iHS_w1	13.10−13.60	17/50	4.3	***TMEM27***, ***CDC42***, ***CA5B***, ***ZRSR2***, ***AP1S2***, ***GRPR***
iHS_w2	32.20−32.70	13/55	3.2	No annotated genes found
iHS_w3	63.20−63.70	6/35	3.5	*RLIM*, ***KIAA2022***, ***ABCB7***
iHS_w4	110.30−110.80	5/36	2.4	***DOCK11***, *WDR44*, *KLHL13*
iHS_w5	41.00−41.50	9/65	3.3	***NDP***, ***EFHC2***
iHS_w6	42.50−43.00	6/60	4.1	*MIR221*
nSL_w1	13.10−13.60	35/50	4.3	***TMEM27***, ***CDC42***, ***CA5B***, ***ZRSR2***, ***AP1S2***, ***GRPR***
nSL_w2	63.30−64.30	19/44	3.7	***KIAA2022***, ***ABCB7***, *UPRT*, *ZDHHC15*, *MAGEE2*
nSL_w3	110.10−110.60	12/39	4.1	***DOCK11***
nSL_w4	64.60−65.10	10/39	3.2	*MAGT1*, *ATRX*, *FGF16*
nSL_w5	32.30−32.80	14/61	3.3	No annotated genes found
nSL_w6	51.40−51.90	8/56	2.6	*SHROOM4*, ***DGKK***, ***CCNB3***
nSL_w7	41.00−41.50	9/65	3.0	***NDP***, ***EFHC2***

*Number of significant/all SNPs within the signal (window): −log(P) ≥3.3 for eROHi, −log(P) ≥2 for iHS and nSL method. $The highest −log(P) value for the individual SNP within the signal (window). #Genes identified with at least two methods (eROHI, iHS or nSL) as positive selection candidates are bolded.

Table 7. Description of mapping statistics and annotation of genes inside positive selection signals identified by new HRiD method on the X chromosome in metapopulation of native Croatian sheep breeds.

Signal name	Position (Mb)	n_a *	n_h $	HRiD	−log(P)#	Candidate genes≠
HRiD_w1	13.04−13.62	42	5.4	9.6	56.5	***TMEM27, CDC42, CA5B, ZRSR2, AP1S2, GRPR***
HRiD_w2	115.30−115.73	36	13.3	4.2	8.7	*AMOT, LHFPL1*
HRiD_w3	73.90−74.54	13	4.3	3.2	4.5	*DACH2*
HRiD_w4	73.57−74.20	10	1.9	3.1	4.1	*CHM, DACH2*
HRiD_w5	56.64−58.09	33	6.9	3.1	4.0	*AR, OPHN1, YIPF6*

*Total number of unique alleles (haplotypes). $Effective number of alleles (haplotypes). #−log(P) value refers to the significance of the signal (window). ≠Genes additionally identified as positive selection candidates by other methods (eROHI, iHS or nSL) are bolded.

4.3.4.1. Candidate genes assigned to the selection signal between 13.04 and 13.69 Mb

Four genes (*CA5B*, *ZRSR2, AP1S2* and *GRPR*) were found within the main signal of all four methods and can be considered as major candidates, ahead of *TMEM27* and *CDC42*, which were not identified by eROHi only. *CA5B* (Carbonic Anhydrase 5B) expression is localized in mitochondria and involved in biological functions such as reversible hydration of carbon dioxide and response to bacteria. In addition, *CA5B* may play an important role in growth, development, energy storage and utilization of porcine skeletal muscle (Guo *et al*., 2021). *ZRSR2* (Zinc Finger CCCH-Type, RNA binding motif and Serine/Arginine Rich 2) may play a role in network interactions during spliceosome formation and has been linked to sex determination in cattle (Peterson, 2020). *AP1S2* (Adaptor Related Protein Complex 1 Subunit Sigma 2) has been linked to abnormal responses to novelty, while *GRPR* (Gastrin Releasing Peptide Receptor) regulates multiple functions of the gastrointestinal tract and central nervous system and has been linked to regulation of the reproductive system in boars (Ma *et al*., 2018). *TMEM27* (Collectrin) is important for amino acid transport, while *CDC42* (Cell Division Cycle 42) regulates signaling pathways that control various cellular functions such as cell morphology, migration, endocytosis and cell cycle progression.

4.3.4.2. Candidate genes assigned to the selection signal between 21.96 and 22.26 Mb

POLA1 (DNA Polymerase Alpha 1, Catalytic Subunit) and *ARX* (Aristaless Related Homeobox) genes were mapped to the selection signal identified by eROHi. *POLA1* gene was associated with DNA replication and RNA primer synthesis. In addition, Starokadomskyy *et al*. (2021) linked it to growth, intellectual abilities and immune disorders, while *ARX* is thought to be involved in CNS development.

4.3.4.3. Candidate genes assigned to the two selection signals between 41.00 and 43.00 Mb

NDP and *EFHC2* genes were found in the signal identified by iHS and nSL located from 41.00 to 41.50 Mb. *NDP* (Norrin Cystine Knot Growth Factor NDP) encodes a secreted protein with a cysteine knot motif that activates the Wnt/beta-catenin signalling pathway and has been associated with dysplasia in dogs (Joyce *et al*., 2021), whereas *EFHC2* (EF-Hand Domain Containing 2) is associated with fear recognition and harm avoidance in humans (Blaya *et al*., 2009).

The iHS signal identified from 42.50 to 43.00 Mb contains only the gene *MIR221* (MicroRNA 221), which has been associated with the regulation of milk fat, protein synthesis and mammary gland development in sheep (Duman *et al*., 2021).

4.3.4.4. Candidate genes assigned to the selection signal between 51.40 and 51.94 Mb

DGKK and *CCNB3* genes were mapped with eROHi and nSL, whereas S*HROOM4* was mapped with nSL only. *DGKK* (Diacylglycerol Kinase Kappa) is involved in oxidative stress response, while *CCNB3* (Cyclin B3) plays an essential role in cell cycle control and was dispensable for spermatogenesis in mice (Karasu and Keeney, 2019). In addition, *SHROOM4* (Shroom Family Member 4) plays an important role in regulating cytoskeletal architecture, brain development and cognition.

4.3.4.5. Candidate genes assigned to the selection signal between 56.64 and 58.09 Mb

Three genes (*AR, OPHN1* and *YIPF6*) found in HRiD signal are associated with tail fatness in sheep (Moradi *et al*., 2012). The gene *AR* (Androgen Receptor) is also important for prostate development, urogenital system, and reproduction and has been linked to carcass traits in cattle by Choi *et al*. (2010). *OPHN1* (Oligophrenin 1) has been linked to abnormal response to novelty, while *YIPF6* (Yip1 Domain Family Member 6) has been linked to intestinal epithelial cell development.

4.3.4.6. Candidate genes assigned to the two selection signals between 63.20 and 65.10 Mb

In the first signal, from 63.20 to 64.30 Mb, *KIAA2022* and *ABCB7* genes were mapped with both iHS and nSL, *RLIM* only with iHS, while *UPRT*, *ZDHHC15* and *MAGEE2* genes were mapped only with nSL. *KIAA2022* (Neurite Extension And Migration Factor) has been linked to nervous system development, while *ABCB7* (ATP Binding Cassette Subfamily B Member 7) is involved in the transport of heme from mitochondria to the cytosol and has therefore been linked to mitochondrial iron accumulation. *RLIM* (Ring Finger Protein, LIM domain interacting) was associated with ligase activity and transcriptional corepressor activity. It has also been linked to mouse lung development (Kammoun *et al*., 2018) and spermiogenesis (Wang *et al*., 2021b). *UPRT* (Uracil Phosphoribosyltransferase homolog) was associated with nucleoside metabolic process, lactation and female pregnancy, whereas *MAGEE2* (MAGE Family Member E2) may play a role as a tumor antigen.

The second signal identified from 64.60 to 65.10 Mb by nSL contained *MAGT1*, *ATRX* and *FGF16* genes. *MAGT1* (Magnesium Transporter 1) is associated with the immune system and glycosylation (Blommaert *et al*., 2019), while *ATRX* (ATRX chromatin remodeler) has numerous functions in development. *FGF16* (Fibroblast Growth Factor 16) is associated with embryonic development, cell growth, morphogenesis, tissue repair, tumor growth and proper heart development.

4.3.4.7. Candidate genes assigned to the selection signal between 73.57 and 74.54 Mb

Two genes (*CHM* and *DACH2*) were mapped by HRiD. Annotations associated with *CHM* (CHM Rab Escort Protein) include GTPase activator activity and Rab geranylgeranyltransferase activity and have been linked to milk production in cattle (Stella *et al*., 2010). *DACH2* (Dachshund Family Transcription Factor 2) may be involved in the regulation of organogenesis and myogenesis and may play a role in premature ovarian failure.

4.3.4.8. Candidate genes assigned to the selection signal between 110.10 and 110.80 Mb

Three genes (*DOCK11*, *WDR44* and *KLHL13*) were mapped by iHS, whereas the *DOCK11* gene was mapped by nSL only. *DOCK11* (Dedicator Of Cytokinesis 11) is involved in the polarisation processes of epithelial cells. Annotations of this gene include guanyl nucleotide exchange factor activity and binding of small GTPases. *WDR44* (WD Repeat Domain 44) may be involved in vesicle recycling, while *KLHL13* (Kelch Like Family Member 13) is required for proper chromosome segregation and completion of

cytokinesis and underlies the female pluripotency phenotype in mammals (Genolet *et al*., 2021).

4.3.4.9. Candidate genes assigned to the selection signal between 112.53 and 112.72 Mb

The signal identified by eROHi contains only the *PLS3* (Plastin 3) gene, which is related to the binding of calcium ions and actin and may play a role in the regulation of bone development.

4.3.4.10. Candidate genes assigned to the selection signal between 115.30 and 115.73 Mb

Two genes, *AMOT* (Angiomotin) and *LHFPL1* (LHFPL tetraspan subfamily member 1), were mapped by HRiD. *AMOT* has been linked to convergent evolution and domesticated adaptation to high-altitude environments in humans (Witt and Huerta-Sánchez, 2019) and, together with *LHFPL1*, to hypoxia adaptation in dogs (Wu *et al*., 2016).

5. DISCUSSION

In the realm of inbreeding and selection within domestic animal populations, the X chromosome has often been overlooked despite its biological importance (Vaiman, 2002; Schafner, 2004; Otto *et al*., 2011; Raudsepp *et al*., 2012), as most genomic studies focus only on autosomes. This dissertation aimed to fill that gap by specifically examining the X chromosome to understand its role in inbreeding and its response to selective pressures in domestic animal populations. Accordingly, the thesis is structured into two main sections: the first addressing inbreeding and the second focusing on positive selection, both of which are pivotal and interrelated phenomena in the context of domestic animal genetics.

In the first section, inbreeding coefficients were estimated specifically on the X chromosome (females only) and PAR solely and compared with estimates on autosomes in six populations of three domestic animal species: cattle, dogs and sheep. Each species was represented by two distinct populations, all genotyped using high density SNP arrays. Five different inbreeding methods were utilized: F_{ROH_SVS}, $F_{ROH\_RZooROH}$, F_{LH1}, F_{VR1} and F_{YA2}, all proven for their reliability by Cabalero *et al*. (2022). This multifaceted approach enabled a thorough comparison of inbreeding levels across different genomic regions, offering insights into the unique dynamics of the X chromosome versus autosomes in each domestic animal population with special attention given to PAR in each sex separately.

The second part of the dissertation introduced a new method, HRiD, designed to identify positive selection signals specifically on the X chromosome by utilizing genomic information from males. HRiD was empirically tested on metapopulation of native Croatian sheep breeds (one of the six populations where inbreeding was also assessed), where its findings were compared with those from other complementary intra-population methods (eROHi, iHS and nSL). The development of this method was essential, considering that classical methods like eROHi cannot identify positive selection on the X chromosome in males due to their hemizygous nature and considering that breeding farms predominantly genotype males because of their shorter generation interval (Jonas and Koning, 2015).

For this dissertation, all populations were carefully selected, ensuring that only those populations genotyped with high density SNP arrays with high success rates were considered, and that each population was only considered if it was adequately represented by both male and female individuals. High density genotypes are essential to obtain more reliable and accurate results in both inbreeding estimation and identification of positive selection signals, as they provide greater coverage and higher resolution (Fan

et al., 2010). The validation of these genotypes, as evidenced by high retention rates of individuals and SNPs (Table 1) even under stringent QC parameters (GenTrain <0.4, GenCall ≤0.8, call rate <0.9 for SNPs and <0.95 for individuals, SNPs deviating from HW equilibrium with $p < 10^{-7}$), supports the high success rates of genotyping in each population used. Conversely, ensuring a sufficient number of females is crucial for an adequate estimation of inbreeding on the X chromosome, which can only be assessed in females. Similarly, a sufficient number of males and females is crucial for valid comparisons of inbreeding within PAR between the sexes and for a reliable comparison of HRiD (based on male haplotypes) with other methods for identifying signals of positive selection. This is confirmed by the fact that, after QC, populations of Croatian cattle breeds, Nellore, Labrador Retriever, Patagonian Sheepdog, Croatian sheep breeds and Soay included 33, 953, 145, 105, 100 and 86 males and 79, 1255, 126, 40, 101 and 102 females.

5.1. Localization of PAR in domestic animal species

Following QC, the precise localization of PAR in each domestic animal species was essential, serving as a foundation for investigating inbreeding and selection on the X chromosome. Such step was crucial as comparing inbreeding between sexes at PAR required an initial accurate identification of this region in each species. Moreover, establishing the PAR's boundaries was imperative to facilitate the implementation of HRiD method, which relies on haplotypes from nonPAR segment of the male X chromosome.

To accurately assess PAR, heterozygosity analysis was predicated on the assumption that nonPAR in males would exhibit a heterozygosity value of zero, indicating hemizygosity. The findings confirmed that this methodology was effective, as it aligned with the genetic structure anticipated from reference genomic data. The exploration of PAR in domestic animals has unveiled substantial variation in the extension and localization across species, such as cattle, dogs and sheep. In cattle, PAR was located at the distal end of the X chromosome, extending from 133.20 Mb to 139.00 Mb. This span, approximately 5.80 Mb in length, coincided with findings in other bovine genomics studies, such as those by Johnson *et al*. (2019), which similarly mapped the bovine PAR to the terminal end due to specific X chromosome rearrangements. Conversely, Liu *et al*. (2019) reported a larger PAR, measuring 6.84 Mb, but based on an assembly of the X chromosome from *Bos taurus indicus*, that is approximately 146 Mb in length, rather than 139 Mb. For dogs, PAR extended from 0.00 Mb to 6.59 Mb, aligning with the findings from studies like those conducted by Young *et al*. (2008), which have characterized the canine PAR to extend approximately 6.60 Mb, with PAB contained within a 2 Kb region. Sheep displayed a PAR spanning from 0.00 Mb to 7.04 Mb, a finding that corroborates the genomic configurations found by Chen *et al*. (2018) and reflects similarities with other small ruminants, suggesting conserved evolutionary mechanisms that regulate sex chromosome recombination (Das *et al*., 2009). The slightly larger PAR in sheep, compared to that in cattle and dogs, might suggest a greater area for genetic recombination, potentially changing genetic variability and influencing trait selection in breeding programs.

Notably, PAB of all three species is contained within the gene *SHROOM2* (Shroom Family Member 2; Young *et al*., 2008; Id-Lahoucine *et al*., 2022; Li *et al*., 2023). As highlighted by Li *et al*. (2023), *SHROOM2* was not only a candidate gene for body weight in sheep, but also exhibited strong LD (r^2 = 0.99) within a 0.3 Mb region at the X chromosome, significantly associated with weaning and yearling weights, suggesting its broader influence in genetic traits linked to sex differentiation. There was a significant

increase in heterozygosity in this small region among males compared to females in both cattle and sheep populations (Figures 8 and 10), whereas in dogs (Figure 9), the increase was less pronounced in Labrador Retriever population and absent in Patagonian Sheepdog population. Such variation between males and females further underscores the importance of researching PAR in assessing inbreeding distinctly by sex, highlighting its critical role in understanding genetic management and conservation strategies.

5.2. Inbreeding at the X chromosome

Upon comparing mean values of each inbreeding coefficient used between the X chromosome and autosomes, a distinct grouping of results emerged, irrespective of the population analyzed. Two main groups can be observed: one consisting of F_{ROH_SVS} and $F_{ROH\_RZooROH}$ coefficients, and the other comprising F_{LH1}, F_{VR1} and F_{YA2} coefficients. The foundational principles of each coefficient explain the observed grouping: the first group, based on the IBD status of alleles ranging from 0 to 1 and estimated using ROHs representing autozygous segments, and second group, based on the IBS status of alleles ranging from -1 to 1 and estimated from deviations from expected values. Given the fundamental differences in the underlying principles, it was crucial to evaluate the results for each group separately. More attention was paid to the first group, as F_{ROH} has been shown to be a highly accurate estimator of inbreeding in most scenarios (Alemu *et al*., 2021; Caballero *et al*., 2022).

Within first group, the difference between the two coefficients lies in their methodological approach. F_{ROH_SVS} is based on an empirical approach for determining ROH, whereas $F_{ROH\_RZooROH}$ employs a statistical approach. Despite these methodological disparities, using both coefficients, mean values on the X chromosome in each domestic animal population were significantly higher than the corresponding autosomal values (Table 3), particularly noted in Soay (F_{ROH_SVS} = 0.422; $F_{ROH\_RZooROH}$ = 0.453), Labrador Retriever (F_{ROH_SVS} = 0.396) and Patagonian Sheepdog (F_{ROH_SVS} = 0.367). The findings align with those observed on human populations by Cotter *et al*. (2024), where each 1% increase in autosomal ROH were associated with a 2.1% increase in X chromosomal ROH, and each 1% increase in autosomal IBD sharing corresponded to a 1.6% increase in X chromosomal IBD sharing. These higher X-linked inbreeding levels could have a stronger effect compared to autosomal inbreeding levels, with the potential consequences of increased female mortality and skewed sex ratios in favor of males, as highlighted by Robinson *et al*. (2014), as well as induced inbreeding depression for various significant traits in females, as pointed out by Curik *et al*. (2022). The pronounced manifestation could be particularly consequential in Soay, which exhibited the highest estimates of both X chromosomal and autosomal inbreeding. As Bérénos *et al*. (2016) outlined, their extended maternal care could result in maternal inbreeding depression, impacting fitness-related traits such as weight and survival rates.

From the perspective of first group, the careful parameter setting provided a high reliability in the estimation and comparison of inbreeding between different genomic landscapes. In the empirical approach, the use of high density SNP arrays allowed the

establishment of a minimum ROH length of 1 Mb, a standard that reflects an equivalent of 50 generations ago and is considered a benchmark for F_{ROH} estimation (Purfield *et al*., 2012; Ferenčaković *et al*., 2013). Additionally, the use of SVS Golden Helix software, rather than alternatives like Plink or detectRUNS for ROH determination, provided the flexibility to allow a specific number of heterozygotes and missing values that are not consecutive per each category (Table 2), enhancing the robustness of inbreeding estimation. However, there were still certain limitations common to all studies of inbreeding that included the X chromosome. For example, the LD values used for the coefficients at both the X chromosome and autosomes (1 cM = 1 Mb) may not accurately reflect the lower recombination expected on nonPAR part of the X chromosome (2/3 of autosomes; Schaffner, 2004), where a different ratio might be more appropriate. The values were set as such because the standard had not yet been determined and were consistent with those in other studies (Zavarez *et al*., 2015; Chen *et al*., 2018; Curik *et al*., 2022). This discrepancy underscores the need for further research to establish appropriate values specific to the X chromosome for different species and populations, potentially through simulations using software like AlphaSimR (Gaynor *et al*., 2021), although current tools lack modifications tailored for the X chromosome specific characteristics. Nonetheless, regardless of the overall set ratio between cM and Mb values, specific patterns on the X chromosome were observed in each population with much larger fluctuations in inbreeding compared to autosomes, indicating a significantly large difference in the rate of recombination between individual parts of nonPAR.

Within second group, all three coefficients exhibited similar results, where mean values of inbreeding at the X chromosome were not significantly different from mean values of inbreeding at autosomes in each population (Table 3). Some populations even showed slightly lower values on the X chromosome than on autosomes, such as Croatian cattle breeds and Soay using all three coefficients, and Nellore using F_{VR1} coefficient. However, it is important to note that there was generally a very narrow range of mean values for all coefficients across all populations, ranging from -0.021 to 0.080 on the X chromosome and from -0.010 to 0.060 on autosomes, further emphasizing the similarity between these coefficients due to the nature of their estimation. So far, no studies had used these coefficients to estimate inbreeding using the X chromosome, which is not surprising given the generally very low number of studies that involved the X chromosome. Although, despite proving the high reliability of both groups in estimating inbreeding, Caballero *et al*. (2022) also recorded much lower mean values for the second group compared to the first group using various simulations, with these three coefficients also showing similar results.

In analyzing the variability of inbreeding at the X chromosome in comparison to autosomes, major differences were observed. Using all five coefficients, greater variation is observed on the X chromosome than on autosomes in all populations studied (with the exception of Croatian cattle breeds in the case of F_{YA2}), as illustrated in Figures 11 to 13. This increased variability of inbreeding on the X chromosome suggest distinct evolutionary dynamics at play, potentially due to its unique inheritance patterns and reduced effective population size (Schaffner, 2004). Figures 14 to 19 further highlight regional variations in inbreeding on the X chromosome, estimated using first group coefficients. The results of each population reveal significant fluctuations in inbreeding levels across various regions of the X chromosome. These findings underscore the importance of considering the X chromosome in genomic studies to capture the full extent of genetic diversity and inbreeding levels in domestic animal populations.

When the mean inbreeding values at PAR were compared between the sexes in each population, no significant differences were found for any of the five coefficients used (Tables 3 and 4). The same applied to their median values (Figures 11-13). Regionally, the inbreeding levels at PAR showed lower values compared to equally sized windows across the X chromosome and similar values compared to equally sized windows across autosomes, as shown in Figures 14 to 19. These results indicate that the assertion made by Otto *et al*. (2011), which suggested that the presence of a large PAR in domestic animals could be attributed to forces promoting recombination, such as autozygosity avoidance, was not confirmed. This suggests that the genetic dynamics of PAR are more similar to those on autosomes and between the sexes than previously hypothesized. However, other forces promoting recombination, such as balancing selection, may be at play, particularly in PAB (*SHROOM2* gene), where high levels of heterozygosity were observed, although identifying these forces remains challenging, as pointed out by Fijarczyk and Babik (2015).

To get a better insight into the behavior of inbreeding across genomic landscapes, Pearson correlations were evaluated between all five used coefficients at the X chromosome, autosomes and PAR, separately for males and females. As shown in Figures 20 to 25, the general trend revealed lower *r* values between different genomic landscapes compared to within the same landscape, indicating distinct behavior of inbreeding coefficients across these regions. Specifically, within the same genomic landscapes, the highest *r* values were generally observed between the first group coefficients, with exceptions at PAR in Nellore females (r = 0.33) and Croatian sheep breed males (r = 0.02). The high *r* values between F_{ROH_SVS} and $F_{ROH\_RZooROH}$ coefficients were also observed by Solé *et al*. (2017), but they used only high density genomic

information from autosomes (r = 0.95). However, a closer examination of Figures 16 and 17 shows that, while they exhibit the same trend, a pronounced difference in inbreeding levels is apparent in regions with lower SNP coverage between these two coefficients. This suggests that when using low density SNP arrays, these coefficients might show greater differences in inbreeding estimation with lower correlations. In contrast, r values between the same coefficients estimated between the X chromosome, autosomes and PAR were generally close to zero in Nellore and Soay, with some negative values observed. Slightly higher correlations were achieved in other populations, with the highest r values observed between the X chromosome and autosomes in the Croatian sheep breeds females, ranging from 0.66 to 0.68. Very low r value (0.08) was also reported by Curik *et al*. (2022) between the X chromosome and autosomes in the Czech Holstein cattle population, but using only F_{ROH_SVS} coefficient on low density SNP data.

To summarize, mean inbreeding values at the X chromosome in comparison to autosomes showed higher values in each domestic animal population using first group coefficients, which was not true for second group coefficients. The variability at the X chromosome was also greater than the variability at autosomes using all five coefficients. Specific patterns of much greater variability in inbreeding were observed across the X chromosome compared to autosomes, emphasizing the importance of including the X chromosome in genomic studies of inbreeding in domestic animal populations. In PAR, no significant differences in mean inbreeding values were found between sexes or compared to autosomal values, suggesting similar genetic dynamics. In contrast, correlations between these regions indicated that all three (X chromosome, autosomes and PAR) exhibited distinct inbreeding behaviors in each population.

5.3. Positive selection at the X chromosome

The genotypes of metapopulation of native Croatian sheep breeds were ideal for testing HRiD for several reasons. The genotyping included 202 sheep individuals (101 females and 101 males, with one male individual excluded after QC) from 105 farms and 10 mouflons (five females and five males), providing a representative sample and minimizing the influence of specific families on the results. Additionally, both mouflons and sheep breeds were from Croatia, ensuring that mouflons provided relatively accurate ancestral information for phasing and constructing MJNs. The equal number of male and female individuals allowed for an adequate comparison between different methods. Specifically, eROHi was performed only on female genotypes, iHS and nSL were performed on all genotypes (utilizing the "--chrX" option in Shapeit2 software), while HRiD was performed only on male haplotypes (SNPs outside PAR). Moreover, the use of a high density SNP array and the very high quality of genotyping, with 18,983 SNPs after QC and a mean distance between adjacent SNPs of 7.13 Kb, increased the accuracy of the results. For example, in estimating eROHi, it was possible to detect ROHs as short as 0.25 Mb and analyse selection signals over an extended period (approximately 200 generations). Álvarez *et al*. (2020) also tracked selection signals over a long-time scale, but their analyses were based on the estimation of HBD segments using statistical approach (described in subchapter 2.2.1.2.1.). An empirical approach was chosen here to evaluate autozygosity caused by both inbreeding and selection (Curik *et al*., 2002), whereas statistical approach focuses more on deviations from HW equilibrium caused by inbreeding rather than selection (Druet and Gautier, 2017). However, the assumption that eROHi method better captures selection-induced autozygosity remains to be verified by computer simulations.

There is no other study that has mapped positive selection on the sex chromosome using only male (XY) or female (XZ) genomic information. From this perspective, HRiD offers an interesting possibility for use in conjunction with other methods or when only male genotypic information is available, which is often the case in genomic breeding value estimations.

A total of 12 regions on a 135.4 Mb long sheep X chromosome were identified in Croatian sheep breeds that exhibited genomic patterns of positive selection (14 signals). While 11 selection signals were identified using three classical methods (eROHi, iHS and nSL), three additional signals were identified using HRiD. The most reliable candidate region under positive selection were between 13.04 and 13.69 Mb, as it was identified with the highest significance using all four methods. In the 12 identified regions, 34 genes

were annotated, as two regions (from 32.20 to 32.80 Mb and from 83.78 to 84.28 Mb) had no genes annotated (Tables 6 and 7).

Almost all (11 of 14) selection signals identified in this dissertation were also observed in other studies on domestic or wild sheep populations, although the signal intervals did not completely overlap (Appendix 13). Exceptions included selection signals mapped from 21.96 to 22.26 Mb and from 83.78 to 84.28 Mb by eROHi, and a signal mapped from 115.30 to 115.73 Mb by HRiD, none of which were reported in other sheep studies. However, the functional characterization of genes within the 115.30-115.73 Mb region (*AMOT* and *LHFPL1*) suggests potential roles in adaptation to hypoxia (details in the Results section), despite their absence in other studies. The most reliable signal overlapped with the region mapped from 13.20 to 13.60 Mb (annotated for the *CA5B*, *ZRSR2*, *AP1S2* and *GRPR* genes) by Chen *et al*. (2018) in a large study of 68 sheep breeds worldwide (using iHS) and in a comparison between sheep and mouflon (using XP-EHH). Both methods classified part of this region (13.2-13.4 Mb; *CA5B*, *ZRSR2*, *AP1S2)* as the primary signal, highlighting its importance and the biological functions of the annotated genes. The region mapped from 32.20 to 32.80 Mb (iHS and nSL), which lacks annotated genes, was also identified by Zhu *et al*. (2015), Liu *et al*. (2016) and Chen *et al*. (2018). Similarly, the region mapped from 41.00 to 43.00 Mb (iHS and nSL), containing *NDP* and *EHC2* genes, was confirmed by Zhu *et al*. (2015), Liu *et al*. (2016) and Cesarani *et al*. (2022). Given that *EFHC2* is associated with fear recognition and harm avoidance, this signal may be linked to extensive husbandry practices (e.g., fear of guard dogs and wolves), characteristic of Croatian native breeds. The region mapped from 51.40 to 51.94 Mb (eROHi and nSL) was identified in two studies by Zhu *et al*. (2015, 2020). The large signal mapped from 56.64 to 58.09 Mb by HRiD, which includes the *AR*, *OPHN1* and *YIPF6* genes, was found in four other studies (Liu *et al*., 2016; Chen *et al*., 2018; Manzari *et al*., 2019; Cesarani *et al*., 2022). This concordance is notable because this region was not detected by eROHi, iHS, and nSL but coincided with the most significant individual nSL outlier [-log(P) = 5.17] at position 56.81 Mb, demonstrating HRiD's complementary potential in identifying positive selection signals. The region from 64.60 to 65.10 Mb (annotated for *MAGT1*, *ATRX* and *FGF16*) was mapped only by nSL and identified by Chen *et al*. (2018) and Zhu *et al*. (2020) in domestic sheep, as well as by Kardos *et al*. (2015) for the candidate genes *ATRX* and *FGF16* in bighorn sheep. Another selection signal mapped using HRiD, but not classical methods, corresponds to the region from 73.57 to 74.54 Mb (annotated for genes *CHM* and *DACH2*) and was also identified by Zhu *et al*. (2015, 2020), further highlighting the usefulness of HRiD method. The signal mapped from 110.10 to 110.80 Mb (iHS and nSL), containing the *DOCK11*, *WDR44* and

KLHL13 genes, was also identified as a positive selection signal by Chen *et al*. (2018), while the signal assigned to (eROHi) the region from 112.53 to 112.72 Mb (*PLS3*) was likewise identified by Zhu *et al*. (2015).

By analysing the phylogenetic relationship (MJN) between all male haplotypes within each signal identified by HRiD, it was possible to determine whether the ancestral or derived haplotype was subject to selection (Figure 27). Given that mouflons, representing the ancestral haplotype, are well adapted to their natural environment and have not been subjected to artificial selection, it is hypothesized that signals associated with the ancestral haplotype are candidate regions related to natural adaptation (HRiD_w1 and HRiD_w3,4).

In this dissertation, eight closely related but distinct breeds were treated as a single unit (metapopulation), which means possible genomic differences between breeds could influence the results. For instance, a high frequency of specific haplotypes in only a few breeds could lead to the misidentification of selection signals. To address this, an MJN analysis was conducted for each selection signal to verify whether favorable haplotypes were uniformly represented across all breeds. The results, detailed in Appendices 11 and 12, indicate a relatively even distribution of breeds within the selected haplotypes. This supports the hypothesis that the identified selection signals likely reflect a long-term adaptive response to the local (Mediterranean) environment and the production system in use. In contrast, short-term selection, either natural or artificial, would result in "private" haplotypes occurring in only one or a few breeds. Particular attention was also paid to the observed haplotype frequency distribution between "continental" and "island" breeds (Appendices 11 and 12). Although major haplotypes under selection were identified in most signals, three signals (mapped from 32.20 to 32.80 Mb, 42.50 to 43.00 Mb, and 110.10 to 110.80 Mb) did not clearly indicate which haplotype was selected, raising doubts about their reliability.

From a methodological perspective, the results of the new HRiD method may be sensitive to the definition of window size, similar to other methods, as the age and strength of selection are functionally related to haplotype size. However, HRiD has several advantages over other methods, including lower sensitivity to variation in recombination rates and the ability to perform phylogenetic analyses on male haplotypes in genomic regions exhibiting selection signals. Overall, the basic concept of HRiD method is sound, though there is room for further improvement, which is anticipated in the near future. One critical and challenging parameter in HRiD is determining the appropriate window size. Due to the varying SNP array density across regions, it is recommended to set the

window size based on the number of SNPs rather than base pair units to ensure that the calculated n_h values are more representative (i.e., determined using the same number of SNPs). Moreover, as with other methods (iHS and nSL), the window size should be tailored to the population under study. For this research, the number of SNPs was set to 70 (35), corresponding to an average of 500 kb (250 kb), consistent with other methods. Given that the genomic composition of Croatian sheep breeds has significantly changed due to environmental adaptations and sustainable production, the focus was on signals that have persisted in the population for a long time (approximately 100 to 200 generations). HRiD is less efficient at detecting selection signals longer than the defined window size because adjacent windows may also be subject to selection. This may have been the case, for example, with HRiD_w3,4 (Figures 26D and 27C), where the analysis power was reduced, although positive selection was still detected.

In summary, a new method for identifying positive selection on the X chromosome, HRiD, was introduced. This method utilizes male genotypic information to identify signals by detecting genomic regions with a sudden decrease in haplotype richness. The results demonstrate that HRiD can be effectively used alongside the eROHi, iHS and nSL methods or in scenarios where only male genotypes are available, which is common in livestock where genomic breeding value estimates are predominantly performed for males. In metapopulation of native Croatian sheep breeds, 14 positive selection signals (across 12 regions) were identified at the X chromosome using all four methods, encompassing a total of 34 annotated genes. The high repeatability (86%) of these findings is notable, as 12 of the identified selection signals were also confirmed in other studies on sheep. Moreover, it is shown that phylogenetic analyses, such as MJN, can provide useful additional information in the analysis of haplotypes identified as selection signals, either in terms of the ancestral or derived status of the advantageous selected haplotypes or by controlling for the potential confounding caused by population structure that may occur when analysing metapopulations. Overall, the results highlight the importance of the X chromosome in the adaptive architecture of domestic ruminants as well as in selection in general, while the new HRiD method opens new avenues for research.

6. CONCLUSIONS

Based on the conducted research of inbreeding and selection on the X chromosome in domestic animal populations using high density genomic data, the following conclusions can be made:

1. The exact location of the PAR on the X chromosome was determined in each of the two distinct populations used for cattle (metapopulation of native Croatian cattle breeds and Nellore), dogs (Labrador Retriever and Patagonian Sheepdog) and sheep (metapopulation of native Croatian sheep breeds and Soay). The PAR was located from 133.20 to 139.00 Mb in cattle, from 0.00 to 6.59 Mb in dogs and from 0.00 to 7.04 Mb in sheep, which is consistent with their reference genomes.

2. Genomic inbreeding on the X chromosome and autosomes was estimated and compared using five different coefficients in each of the six domestic animal populations, fulfilling **objective O1**. Higher inbreeding was found on the X chromosome compared to autosomes in all populations using the F_{ROH_SVS} and $F_{ROH\_RZooROH}$ coefficients, while no differences were found using the F_{LH1}, F_{VR1} and F_{YA2} coefficients. Therefore, **hypothesis H1** is accepted for the coefficients F_{ROH_SVS} and $F_{ROH\_RZooROH}$, but rejected for the coefficients F_{LH1}, F_{VR1} and F_{YA2}. This indicates that whether inbreeding is higher on the X chromosome than on autosomes depends on the method used, with the most reliable coefficients suggesting higher inbreeding on the X chromosome. In addition, greater variability in inbreeding was found on the X chromosome compared to autosomes using all five coefficients in all populations, with specific patterns observed across the X chromosome. Therefore, **hypothesis H2** is accepted.

3. Genomic inbreeding in the PAR was estimated and compared separately for each sex with five different coefficients in each of the six domestic animal populations, achieving **objective O2**. Furthermore, PAR values were compared to those obtained from autosomes. No difference in inbreeding between sexes at the PAR or compared to autosomal values (in almost all cases) was found. Therefore, **hypothesis H3** is rejected, suggesting similar genetic dynamics in the context of inbreeding at the PAR between sexes and in comparison to autosomes.

4. A new method, named Haplotype Richness Drop (HRiD), for identifying signals of positive selection based on the difference in haplotype richness of the X chromosome in males has been developed, fulfilling **objective O3**. Using this method, in conjunction with classical methods (eROHi, iHS and nSL), positive

selection signals on the X chromosome were identified in the metapopulation of native Croatian sheep breeds, achieving **objective O4**. A total of 14 positive selection signals (across 12 regions) were identified, encompassing a total of 34 annotated genes, with high concordance (86%) with other studies on sheep. The high accuracy and reliability of the HRiD method were demonstrated, as the most significant signal identified by HRiD, between 13.04 and 13.69 Mb, was also the most significant one using all three classical methods and all four identified signals were consistently validated. Therefore, **hypothesis H4** is accepted.

5. In general, the results emphasize the importance of including the X chromosome in inbreeding estimation and selection identification in domestic animal populations, while the new HRiD method opens up new possibilities in identifying positive selection signals using heterogametic sex haplotypes.

7. REFERENCES

Alemu S.W., Kadri N.K., Harland C., Faux P., Charlier C., Caballero A., Druet T. (2021). An evaluation of inbreeding measures using a whole-genome sequenced cattle pedigree. Heredity 126: 410-423.

Allendorf F.W., Luikart G., Aitken S.N. (2013). Conservation and the Genetics of Populations. Hoboken: John Wiley & Sons.

Álvarez I., Fernández I., Traoré A., Pérez-Pardal L., Menéndez-Arias N.A., Goyache F. (2020). Genomic scan of selective sweeps in Djallonké (West African Dwarf) sheep shed light on adaptation to harsh environments. Sci Rep 10: 1-13.

Arzik Y., Kizilaslan M., Behrem S., White S.N., Piel L.M., Cinar M.U. (2023). Genome-Wide Scan of Wool Production Traits in Akkaraman Sheep. Genes 14: 713.

Ballou J. (1983). Calculating inbreeding coefficients from pedigrees. In Genetics and conservation: a reference for managing wild animal and plant populations (pp. 509-520). Benjamin/Cummings Publishing Company. Menlo Park, California, USA.

Bandelt H.J., Forster P., Röhl A. (1999). Median-joining Networks for Inferring Intraspecific Phylogenies. Mol Biol Evol 16: 37-48.

Barrios N., González-Lagos C., Dreger D.L., Parker H.G., Nourdin-Galindo G., Hogan A.N., Gómez M.A., Ostrander E.A. (2022). Patagonian Sheepdog: Genomic analyses trace the footprints of extinct UK herding dogs to South America. PLoS Genet 18: e1010160.

Baumung R., Farkas J., Boichard D., Mészáros G., Sölkner J., Curik I. (2015). GRAIN: a computer program to calculate ancestral and partial inbreeding coefficients using a gene dropping approach. J Anim Breed Genet 132: 100-108.

Bérénos C., Ellis P.A., Pilkington J.G., Pemberton J.M. (2016). Genomic analysis reveals depression due to both individual and maternal inbreeding in a free-living mammal population. Mol Ecol 25: 3152-3168.

Binversie E.E., Baker L.A., Engelman C.D., Hao Z., Moran J.J., Piazza A.M., Sample, S.J., Muir P. (2020). Analysis of copy number variation in dogs implicates genomic structural variation in the development of anterior cruciate ligament rupture. Plos One 15: e0244075.

Biscarini F., Cozzi P., Gaspa G., Marras G. (2018). detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes. CRAN (The Comprehensive R Archive Network).

Bjelland D.W., Weigel K.A., Vukasinovic N., Nkrumah J.D. (2013) Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. J Dairy Sci 96: 4697-4706.

Blaya C., Moorjani P., Salum G.A., Gonçalves L., Weiss L.A., Leistner-Segal S., Manfro G.G., Smoller J.W. (2009). Preliminary evidence of association between EFHC2, a gene implicated in fear recognition, and harm avoidance. Neurosci Lett 452: 84-86.

Blommaert E., Péanne R., Cherepanova N.A., Rymen D., Staels F., Jaeken J., Race V., Keldermans L., Souche E., Corveleyn A., Sparkes R. (2019). Mutations in MAGT1 lead to a glycosylation disorder with a variable phenotype. Proc Natl Acad Sci 116: 9865-9870.

Bosse M., Megens H.J., Derks M.F., de Cara Á.M., Groenen M.A. (2019). Deleterious alleles in the context of domestication, inbreeding, and selection. Evol Appl 12: 6-17.

Boyko A.R., Quignon P., Li L., Schoenebeck J.J., Degenhardt J.D., Lohmueller K.E., Zhao K., Brisbin A., Parker H.G., vonHoldt B.M., Cargill M., Auton A., Reynolds A., Elkahloun A. G., Castelhano M., Mosher D.S., Sutter N.B., Johnson G.S., Novembre J., Hubisz M.J., Siepel A., Wayne R.K., Bustamante C.D., Ostrander E.A. (2010). A Simple Genetic Architecture Underlies Morphological Variation in Dogs. Plos Biol 8: e1000451.

Brook B.W., Tonkyn D.W., O'Grady J.J., Frankham R. (2002). Contribution of inbreeding to extinction risk in threatened species. Conserv Ecol 6.

Caballero A., Villanueva B., Druet T. (2021). On the estimation of inbreeding depression using different measures of inbreeding from molecular markers. Evol Appl 14: 416-428.

Caballero A., Fernández A., Villanueva B., Toro M.A. (2022). A comparison of marker-based estimators of inbreeding and inbreeding depression. Genet Sel Evol 54: 82.

Calderón R., Aresti U., Ambrosio B., González-Martín A. (2009). Inbreeding coefficients for X-linked and autosomal genes in consanguineous marriages in Spanish populations: the case of Guipúzcoa (Basque Country). Ann Hum Genet 73: 184-195.

Calus M.P. (2010). Genomic breeding value prediction: methods and procedures. Animal 4: 157-164.

Carrel L., Willard H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434: 400-404.

Cesarani A., Gaspa G., Correddu F., Dimauro C., Macciotta N.P. (2022). Unravelling the effect of environment on the genome of Sarda breed ewes using Runs of Homozygosity. J Anim Breed Genet 139: 292-306.

Chang H.L., Fernando R.L., Grossman M. (1991). On the principle underlying the tabular method to compute coancestry. Theor Appl Genet 81: 233-238.

Chang C.C., Chow C.C., Tellier L.C., Vattikuti S., Purcell S.M., Lee J.J. (2015). Second-generation Plink: Rising to the challenge of Larger and Richer Datasets. GigaSci 4: s13742.

Charlesworth B. (1991). The evolution of sex chromosomes. Sci 251: 1030-1033.

Charlesworth B., Charlesworth D. (1999). The genetic basis of inbreeding depression. Genet Res 74: 329–340.

Chen Z.-H., Zhang M., Lv F.-H., Ren X., Li W.-R., Liu M.-J., Nam K., Bruford M.W., Li M.-H. (2018). Contrasting Patterns of Genomic Diversity Reveal Accelerated Genetic Drift but Reduced Directional Selection on X-Chromosome in Wild and Domestic Sheep Species. Gen Biol Evol 10: 1282-1297.

Choi B., Ryu K., Bong J., Lee J., Choy Y., Son S., Han O. Baik M. (2010). Comparison of steroid hormone concentrations and mRNA levels of steroid receptor genes in longissimus dorsi muscle and subcutaneous fat between bulls and steers and association with carcass traits in Korean cattle. Livest Sci 131: 218-226.

Choi Y., Chan A.P., Kirkness E., Telenti A., Schork N.J. (2018). Comparison of Phasing Strategies for Whole Human Genomes. Plos Genet 14: e1007308.

Cotter D.J., Severson A.L., Kang J.T., Godrej H.N., Carmi S., Rosenberg N.A. (2024). Modeling the effects of consanguinity on autosomal and X-chromosomal runs of homozygosity and identity-by-descent sharing. G3 14: jkad264.

Curik I., Sölkner J., Stipic N. (2002). Effects of models with finite loci, selection, dominance, epistasis and linkage on inbreeding coefficients based on pedigree and genotypic information. J Anim Breed Genet 119: 101-115.

Curik I., Ferenčaković M., Sölkner J. (2014). Inbreeding and Runs of Homozygosity: a Possible Solution to an Old Problem. Livestock Sci 166: 26-34.

Curik I., Vostra-Vydrova H., Shihabi M., Sölkner J., Vostry L. (2022). Estimation of sex chromosome inbreeding depression on milk production in cattle. In Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP) Technical and species orientated innovations in animal breeding, and contribution of genetics to solving societal challenges (pp. 874-877). Wageningen Academic Publishers.

Czarnik U, Grzybowski G, Kamiński S, Prusak B, Zabolewicz T. (2007). Effectiveness of a program aimed at the elimination of BLAD-carrier bulls from Polish Holstein-Friesian cattle. J Appl Genet 48: 375-377.

Darwin, C. (1859). On the origin of species: facsimile of the first edition.

Das P.J., Chowdhary B.P., Raudsepp T. (2009). Characterization of the bovine pseudoautosomal region and comparison with sheep, goat, and other mammalian pseudoautosomal regions. Cytogenet Genome Res 126: 139-147.

Delaneau O., Marchini J. (2014). Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nat Commun 5: 3934.

Dekkers J. (2012). Application of genomics tools to animal breeding. Curr Genomics 13: 207-212.

Ding Q., Zhang J. (2012). seqRFLP: Simulation and Visualization of Restriction Enzyme Cutting Pattern from DNA Sequences. R Package Version 1.0.1. Available at: https://github.com/helixcn/seqRFLP (Accessed January 25, 2022).

Doekes H.P., Veerkamp R.F., Bijma P., de Jong G., Hiemstra S.J., Windig J.J. (2019). Inbreeding depression due to recent and ancient inbreeding in Dutch Holstein–Friesian dairy cattle. Genet Sel Evol 51: 1-16.

Druet T., Gautier M. (2017). A model-based approach to characterize individual inbreeding at both global and local genomic scales. Mol Ecol 26: 5820-5841.

Duman E., Özmen Ö., Kul S. (2021). Oar-miR-16b and oar-miR-27a: Negatively correlated with milk yield and milk protein in sheep. Anim Biotechnol 1-14.

Excoffier L., Lischer H.E.L. (2010). Arlequin Suite Ver 3.5: a New Series of Programs to Perform Population Genetics Analyses under Linux and Windows. Mol Ecol Resour 10: 564-567.

Fan B., Du Z.Q., Gorbach D.M., Rothschild M.F. (2010). Development and application of high-density SNP arrays in genomic studies of domestic animals. Asian-Aust J Anim Sci 23: 833-847

Faure J.M., Mills A.D. (2014). Improving the adaptability of animals by selection. In Genetics and the behavior of domestic animals (pp. 291-316). Academic Press.

Ferenčaković M., Sölkner J., Curik I. (2013). Estimating Autozygosity from High-Throughput Information: Effects of SNP Density and Genotyping Errors. Genet Sel Evol 45: 42.

Ferenčaković M., Sölkner J., Kapš M., Curik I. (2017). Genome-wide mapping and estimation of inbreeding depression of semen quality traits in a cattle population. J Dairy Sci 100: 4721-4730.

Fernando R.L., Grossman M. (1990). Genetic evaluation with autosomal and X-chromosomal inheritance. Theor Appl Genet 80: 75-80.

Ferrer-Admetlla A., Liang M., Korneliussen T., Nielsen R. (2014). On Detecting Incomplete Soft or Hard Selective Sweeps Using Haplotype Structure. Mol Biol Evol 31: 1275-1291.

Fijarczyk A., Babik W. (2015). Detecting balancing selection in genomes: limits and prospects. Mol Ecol 24: 3529-3545.

Forutan M., Ansari Mahyari S., Baes C., Melzer N., Schenkel F.S., Sargolzaei M. (2018). Inbreeding and runs of homozygosity before and after genomic selection in North American Holstein cattle. BMC Genomics 19: 1-12.

Freeman A.R., Bradley D.G., Nagda, S., Gibson, J.P., Hanotte O. (2006). Combination of multiple microsatellite data sets to investigate genetic diversity and admixture of domestic cattle. Anim Genet 37: 1-9.

Ganguly I., Jeevan C., Singh S., Dixit S.P., Sodhi M., Ranjan A., Kumar S., Sharma A. (2020). Y-chromosome genetic diversity of Bos indicus cattle in close proximity to the centre of domestication. Sci Rep 10: 9992.

Gao X., Starmer J., Martin E.R. (2008). A Multiple Testing Correction Method for Genetic Association Studies Using Correlated Single Nucleotide Polymorphisms. Genet Epidemiol 32: 361-369.

Gautier M., Vitalis R. (2012). Rehh: an R Package to Detect Footprints of Selection in Genome-wide SNP Data from Haplotype Structure. Bioinformatics 28: 1176-1177.

Gaynor R.C., Gorjanc G., Hickey J.M. (2021). AlphaSimR: an R package for breeding program simulations. G3 11: jkaa017.

Genolet O., Monaco A.A., Dunkel I., Boettcher M., Schulz E.G. (2021). Identification of X-chromosomal genes that drive sex differences in embryonic stem cells through a hierarchical CRISPR screening approach. Genome Biol 22: 1-41.

Georges M. (2007). Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. Annu. Rev. Genomics Hum Genet 8: 131-162.

Goudet J., Kay T., Weir B.S. (2018) How to estimate kinship. Mol Ecol 27: 4121-4135.

Gouveia J.J.D.S., Silva M.V.G.B.D., Paiva S.R., Oliveira S.M.P.D. (2014). Identification of selection signatures in livestock species. Genet Mol Biol 37: 330-342.

Greenbaum G., Templeton A.R., Zarmi Y., Bar-David S. (2014). Allelic Richness Following Population Founding Events - A Stochastic Modeling Framework Incorporating Gene Flow and Genetic Drift. PLoS One 9: e115203.

Guo J., Fan X., Yang Y., Liang G., Tang Z. (2021). Sequence Characteristics and Expression Analysis of CA5B Gene in Pigs. Act Vet Zoot Sin 52: 322-330.

Gutiérrez J.P., Goyache F. (2005). A note on ENDOG: a computer program for analysing pedigree information. J Anim Breed Genet 122: 172-176.

Harris, A.M., Garud, N.R., DeGiorgio, M. (2018). Detection and classification of hard and soft sweeps from unphased genotypes by multilocus genotype identity. Genetics 210: 1429-1452.

Howard J.T., Pryce J.E., Baes C., Maltecca C. (2017). Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. J Dairy Sci 100: 6009-6024.

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: 460.

Id-Lahoucine S., Casellas J., Fonseca P. A., Suárez-Vega A., Schenkel FS., Cánovas A. (2022). Deviations from mendelian inheritance on bovine X-chromosome revealing recombination, sex-of-offspring effects and fertility-related candidate genes. Genes 13: 2322.

Jobling M.A., Tyler-Smith C. (2003). The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4: 598-612.

Johnson T., Keehan M., Harland C., Lopdell T., Spelman R.J., Davis S.R., Rosen B.D., Smith T.P.L., Couldrey C. (2019). Identification of the pseudoautosomal region in the Hereford bovine reference genome assembly ARS-UCD1. 2. J Dairy Sci 102: 3254-3258.

Johnston S.E., Bérénos C., Slate J., Pemberton, J.M. (2016). Conserved genetic architecture underlying individual recombination rate variation in a wild population of Soay sheep (Ovis aries). Genet 203: 583-598.

Johnston H.R., Keats B.J., Sherman S.L. (2019). Population genetics. In Emery and Rimoin's Principles and Practice of Medical Genetics and Genomics (pp. 359-373). Academic Press.

Jonas E., Koning D.J.D. (2015). Genomic selection needs to be carefully assessed to meet specific requirements in livestock breeding programs. Front Genet 6: 49.

Jones D.F. (1917). Dominance of linked factors as a means of accounting for heterosis. Genet 2: 466.

Joyce H., Burmeister L.M., Wright H., Fleming L., Oliver J.A., Mellersh C. (2021). Identification of a variant in NDP associated with X-linked retinal dysplasia in the English cocker spaniel dog. PloS One 16: e0251071.

Kammoun M., Maas E., Criem N., Gribnau J., Zwijsen A., Vermeesch J.R. (2018). RLIM enhances BMP signalling mediated fetal lung development in mice. bioRxiv: 507921.

Karasu M.E., Keeney S. (2019). Cyclin B3 is dispensable for mouse spermatogenesis. Chromosoma 128: 473-487.

Kardos M., Luikart G., Bunch R., Dewey S., Edwards W., McWilliam S., Stephenson J., Allendorf F.W., Hogg J.T., Kijas J. (2015). Whole-genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. Mol Ecol 24: 5616-5632.

Kardos M., Qvarnström A., Ellegren H. (2017). Inferring Individual Inbreeding and Demographic History from Segments of Identity by Descent in Ficedula Flycatcher Genome Sequences. Genetics 205: 1319-1334.

Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R., San Cristobal M., Servin B., McCulloch R., Whan V., Gietzen K., Paiva S., Barendse W., Ciani E., Raadsma H., McEwan J., Dalrymple B. (2012). Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol 10: e1001258.

Kimura M., Crow J.F. (1964). The Number of Alleles that Can Be Maintained in a Finite Population. Genetics 49: 725-738.

Kristensen T.N., Pedersen K.S., Vermeulen C.J., Loeschcke V. (2010). Research on inbreeding in the 'omic'era. Trends Ecol Evol 25: 44-52.

Kudo A., Sakaguchi K. (1963). A Method for Calculating the Inbreeding Coefficient II. Sex-Linked Genes. Am J Hum Genet 15: 476-480.

Kumar S., Stecher G., Tamura K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33: 1870-1874.

Leigh J.W., Bryant D. (2015). Popart: Full-feature Software for Haplotype Network Construction. Methods Ecol Evol 6: 1110-1116.

Lencz T., Lambert C., DeRosse P., Burdick K.E., Morgan T.V., Kane J.M., Kucherlapati R., Malhotra A.K. (2007). Runs of Homozygosity Reveal Highly Penetrant Recessive Loci in Schizophrenia. Proc Natl Acad Sci USA 104: 19942-19947.

Leroy G. (2014). Inbreeding depression in livestock species: review and meta-analysis. Anim Genet 45: 618-628.

Lesk A.M. (2017). Introduction to genomics. Oxford University Press.

Leutenegger A.L., Prum B., Ge´nin E., Verny C., Lemainque A., Clerget-Darpoux F., Thompson E.A. (2003). Estimation of the inbreeding coefficient through use of genomic data. Am J Hum Genet 73: 516-523.

Li C.C., Horvitz D.G. (1953) Some methods of estimating the inbreeding coefficient. Am J Hum Genet 5: 107.

Li R., Li C., Chen H., Li R., Chong Q., Xiao H., Chen S. (2020). Genome-wide scan of selection signatures in Dehong humped cattle for heat tolerance and disease resistance. Anim Genet 51: 292-299.

Li Y., Yang H., Guo J., Yang Y., Yu Q., Guo Y., Zhang C., Wang Z., Zuo P. (2023). Uncovering the candidate genes related to sheep body weight using multi-trait genome-wide association analysis. Front Vet Sci 10: 1206383.

Liu Z., Ji Z., Wang G., Chao T., Hou L., Wang J. (2016). Genome-wide analysis reveals signatures of selection for important traits in domestic sheep from different ecoregions. BMC Genomics 17: 1-14.

Liu X.X., Pan J.F., Zhao Q.J., He X.H., Pu Y.B., Han J.L., Ma Y.H., Jiang L. (2018). Detecting selection signatures on the X chromosome of the Chinese Debao pony. J Anim Breed Genet 135: 84-92.

Liu R., Low W.Y., Tearle R., Koren S., Ghurye J., Rhie A., Phillippy A.M., Rosen B.D., Bickhart D.M., Smith T.P., Hiendleder S., Williams J.L. (2019). New insights into mammalian sex chromosome structure and evolution using high-quality sequences from bovine X and Y chromosomes. BMC Genomics 20: 1-11.

Lukic B., Curik I., Drzaic I., Galić V., Shihabi M., Vostry L., Cubric-Curik V. (2023). Genomic signatures of selection, local adaptation and production type characterisation of East Adriatic sheep breeds. J Anim Sci Biotechol 14: 142.

Ma Y., Zhang H., Zhang Q., Ding X. (2014). Identification of selection footprints on the X chromosome in pig. Plos One 9: e94911.

Ma Z., Zhang Y., Su J., Li X., Yang S., Qiao W., Suo C., Lei Z. (2018). Distribution of the pig gastrin-releasing peptide receptor and the effect of GRP on porcine Leydig cells. Peptides 99: 142-152.

MacCluer J.W., VandeBerg J.L., Read B., Ryder O.A. (1986). Pedigree analysis by computer simulation. Zoo Biol 5: 147-160.

Manzari Z., Mehrabani-Yeganeh H., Nejati-Javaremi A., Moradi M.H., Gholizadeh M. (2019). Detecting Selection Signatures in Three Iranian Sheep Breeds. Anim Genet 50: 298-302.

Marras G., Gaspa G., Sorbolini S., Dimauro C., Ajmone-Marsan P., Valentini A., Williams J.L., Macciotta N.P.P. (2015). Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim Genet 46: 110-121.

Marras G., Wood B.J., Makanjuola B., Malchiodi F., Peeters K., Van As P., Baes C.F., Biscarini F. (2018). Characterization of runs of homozygosity and heterozygosity-rich regions in a commercial turkey (Meleagris gallopavo) population. In Proceedings of 11th World Congress on Genetics Applied to Livestock Production (pp. 11.16), Auckland, New Zealand.

Mason I.L. (1973). The role of natural and artificial selection in the origin of breeds of farm animals: A critique of Darwin's ‘The variation of animals and plants under domestication’. Z. Tierzücht. Züchtungsbiol 90: 229-244.

Maurano M.T., Humbert R., Rynes E., Thurman R.E., Haugen E., Wang H., Reynolds A.P., Sandstrom R., Qu H., Brody J., Shafer A., Neri F., Lee K., Kutyavin T., Stehling-

Sun S., Johnson A.K., Canfield T.K., Giste E., Diegel M., Bates D., Hansen R.S., Neph S., Sabo P.J., Heimfeld S., Raubitschek A., Ziegler S., Cotsapas C., Sotoodehnia N., Glass I., Sunyaev S.R., Kaul R., Stamatoyannopoulos J.A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Sci 337: 1190-1195.

Mc Parland S., Kearney J. F., Rath M., Berry D.P. (2007). Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. J Dairy Sci 90: 4411-4419.

McQuillan R., Leutenegger A.L., Abdel-Rahman R., Franklin C.S., Pericic M., Barac-Lauc L., Smolej-Narancic N., Janicijevic B., Polasek O., Tenesa A., MacLeod A.K., Farrington S.M., Rudan P., Hayward C., Vitart V., Rudan I., Wild S.H., Dunlop M.G., Wright A.F., Campbell H., Wilson J.F. (2008) Runs of homozygosity in European populations. Am J Hum Genet 83: 359-372.

Meuwissen T., Hayes B., Goddard M. (2016). Genomic selection: A paradigm shift in animal breeding. Anim Front 6: 6-14.

Milligan B.G. (2003) Maximum-likelihood estimation of relatedness. Genetics 163:1153-1167.

Montaldo H.H., Meza-Herrera C.A. (1998). Use of molecular markers and major genes in the genetic improvement of Livestock. Electron J Biotechnol 1: 1-7.

Moradi M.H., Nejati-Javaremi A., Moradi-Shahrbabak M., Dodds K.G., McEwan J.C. (2012). Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. BMC Genet 13: 1-15.

Nadachowska-Brzyska K., Burri R., Ellegren H. (2019). Footprints of Adaptive Evolution Revealed by Whole Z Chromosomes Haplotypes in Flycatchers. Mol Ecol 28: 2290–2304.

Namekawa S.H., Lee J.T. (2009). XY and ZW: is meiotic sex chromosome inactivation the rule in evolution?. PLoS Genet 5: e1000493.

Nothnagel M., Lu T.T., Kayser M., Krawczak M. (2010). Genomic and Geographic Distribution of SNP-Defined Runs of Homozygosity in Europeans. Hum Mol Genet 19: 2927-2935.

Ohno S. (2013). Sex chromosomes and sex-linked genes (Vol. 1). Springer Science & Business Media.

Otto S.P., Pannell J.R., Peichel C.L., Ashman T.L., Charlesworth D., Chippindale A.K., Delph L.F., Guerrero R.F., Scarpino S.V., McAllister B.F. (2011). About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. Trends Genet 27: 358-367.

Pal A., Chakravarty A.K. (2020). Disease resistance for different livestock species. In Genetics and breeding for disease resistance of livestock (pp. 271-296). Academic Press: Cambridge, MA, USA.

Pelabon C., Hansen T.F., Carter A.J., Houle D. (2010). Evolution of variation and variability under fluctuating, stabilizing, and disruptive selection. Evol 64: 1912-1925.

Peterson E.K. (2020). Novel Polymorphisms of ZRSR2 and GPM6B Gene Homologs and Their Use in Sex Identification of Bovine and Porcine Species. [Doctoral dissertation]: Utah State University.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J., Sham P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-575.

Purfield D., Berry D., McParland S., Bradley D. (2012). Runs of homozygosity and population history in cattle. BMC Genetics 13: 70.

Qanbari S., Simianer H. (2014). Mapping Signatures of Positive Selection in the Genome of Livestock. Livestock Sci 166: 133-143.

Quilter C.R., Blott S.C., Mileham A.J., Affara N.A., Sargent C.A., Griffin D.K. (2002). A mapping and evolutionary study of porcine sex chromosome gene. Mamm. Genome 13: 588-594.

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Rajawat D., Panigrahi M., Nayak S.S., Bhushan B., Mishra B.P., Dutt T. (2024). Dissecting the genomic regions of selection on the X chromosome in different cattle breeds. 3 Biotech 14: 1-16.

Raudsepp T., Chowdhary B.P. (2008). The horse pseudoautosomal region (PAR): characterization and comparison with the human, chimp and mouse PARs. Cytogenet. Genome Res 121: 102-109.

Raudsepp T., Das P.J., Avila F., Chowdhary B.P. (2012). The pseudoautosomal region and sex chromosome aneuploidies in domestic species. Sex Dev 6: 72-83.

Rauw W.M., Kanis E., Noordhuizen-Stassen E.N., Grommers F.J. (1998). Undesirable side effects of selection for high production efficiency in farm animals: a review. Livestock Product Sci 56: 15-33.

Raven L.A., Cocks B.G., Hayes B.J. (2014). Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics 15: 1-14.

Rice W.R. (1996). Evolution of the Y sex chromosome in animals. Biosci 46: 331-343.

Robinson S.P., Simmons L.W., Kennington W.J. (2014). Sex-biased mortality associated with inbreeding in Drosophila melanogaster. BMC Evol Biol 14: 51.

Romanov, M.N., Sölkner, J., Zinovieva, N. A., Wimmers, K., & Weigend, S. (2023). Traditional and up-to-date genomic insights into domestic animal diversity. Front Genet 13: 1117708.

Rozas J., Ferrer-Mata A., Sánchez-DelBarrio J.C., Guirao-Rico S., Librado P., Ramos-Onsins S.E., Sánchez-Gracia A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol Evol 34: 3299-3302.

Sabeti P.C., Reich D.E., Higgins J.M., Levine H.Z.P., Richter D.J., Schaffner S.F., Gabriel S.B., Platko J.V., Patterson N.J., McDonald G.J., Ackerman H.C., Campbell S.J., Altshuler D., Cooper R., Kwiatkowski D., Ward R., Lander E.S. (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419: 832-837.

Saccheri I., Hanski I. (2006). Natural selection and population dynamics. Trends Ecol Evol 21: 341-347.

Sanchez M.P., Tribout T., Iannuccelli N., Bouffaud M., Servin B., Tenghe A., Dehais P., Muller N., Del Schneider M.P., Mercat M.J., Bidanel J.P., Gilbert H. (2014). A genome-wide association study of production traits in a commercial population of Large White pigs: evidence of haplotypes affecting meat quality. Genet Sel Evol 46: 1-12.

Saravanan K.A., Panigrahi M., Kumar H., Bhushan B., Dutt T., Mishra B.P. (2020). Selection Signatures in Livestock Genome: A Review of Concepts, Approaches and Applications. Livestock Sci 241: 104257.

SAS Institute Inc. (2013). SAS® 9.4 Guide to Software Updates and Product Changes. Cary, NC: SAS Institute Inc.

Saura M., Fernández A., Varona L., Fernández A.I., de Cara M.Á.R., Barragán C., Villanueva B. (2015). Detecting inbreeding depression for reproductive traits in Iberian pigs using genome-wide data. Genet Sel Evol 47: 1-9.

Schaffner S.F. (2004). The X chromosome in population genetics. Nat Rev Genet 5: 43-51.

Schnabel R.D., Sonstegard T.S., Taylor J.F., Ashwell M.S. (2005). Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. Anim Genet 36: 408-416.

Silió L., Rodríguez M.C., Fernández A., Barragán C., Benítez R., Óvilo C., Fernández A.I. (2013). Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics J Anim Breed Genet 130: 349-360.

Skaletsky H., Kuroda-Kawaguchi T., Minx P.J., Cordum H.S., Hillier L., Brown L.G., Repping S., Pyntikova T., Ali J., Bieri T., Chinwalla A., Delehaunty A., Delehaunty K., Du H., Fewell G., Fulton L., Fulton R., Graves T., Hou S.F., Latrielle P., Leonard S., Mardis E., Maupin R., McPherson J., Miner T., Nash W., Nguyen C., Ozersky P., Pepin K., Rock S., Rohlfing T., Scott K., Schultz B., Strong C., Tin-Wollam A., Yang S.P., Waterston R.H., Wilson R.K., Rozen S., Page D.C. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423: 825-837.

Solé M., Gori A. S., Faux P., Bertrand A., Farnir F., Gautier M., Druet T. (2017). Age-based partitioning of individual genomic inbreeding levels in Belgian Blue cattle. Genet Sel Evol 49: 1-18.

Starokadomskyy P., Escala Perez-Reyes A., Burstein E. (2021). Immune dysfunction in mendelian disorders of POLA1 deficiency. J Clin Immunol 41: 285-293.

Stella A., Ajmone-Marsan P., Lazzari B., Boettcher P. (2010). Identification of selection signatures in cattle breeds selected for dairy production. Genetics 185: 1451-1461.

Szpiech Z.A., Hernandez R.D. (2014). Selscan: an Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. Mol Biol Evol 31: 2824-2827.

Thompson E.A. (2008). The IBD process along four chromosomes. Theor Popul Biol 73: 369-373.

Uffelmann E., Huang Q.Q., Munung N.S., De Vries J., Okada Y., Martin A.R., Lappalainen T., Posthuma D. (2021). Genome-wide association studies. Nat Rev Methods Primers 1: 59.

Utsunomiya Y.T., Pérez O'Brien A.M., Sonstegard T.S., Sölkner J., Garcia J.F. (2015). Genomic data as the “htchhiker's guide” to cattle adaptation: tracking the milestones of past selection in the bovine genome. Front Genet 6: 125042.

Vaiman D. (2002). Fertility, sex determination, and the X chromosome. Cytogenet. Genome Res 99: 224-228.

VanRaden P.M. (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91: 4414-4423.

VanRaden P.M., Olson K.M., Wiggans G.R., Cole J.B., Tooker M. E. (2011). Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. J Dairy Sci 94: 5673-5682.

Vicoso B., Charlesworth B. (2006). Evolution on the X Chromosome: Unusual Patterns and Processes. Nat Rev Genet 7: 645-653.

Vignal A., Milan D., SanCristobal M., Eggen A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. Genet Sel Evol 34: 275-305.

Villanueva B., Fernández A., Saura M., Caballero A., Fernández J., Morales-González E., Toro M.A., Pong-Wong R. (2021). The value of genomic relationship matrices to estimate levels of inbreeding. Genet Sel Evol 53: 1-17.

Visscher P.M., Wray N.R., Zhang Q., Sklar P., McCarthy M.I., Brown M.A., Yang J. (2017). 10 years of GWAS discovery: biology, function, and translation. Am J Hum Genet 101: 5-22.

Voight B.F., Kudaravalli S., Wen X., Pritchard J.K. (2006). A Map of Recent Positive Selection in the Human Genome. PLoS Biol 4: e72.

Wang J. (2007) Triadic IBD coefficients and applications to estimating pairwise relatedness. Genet Res 89: 135-153.

Wang J. (2011) COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. Mol Ecol Resour 11: 141-145.

Wang T., Zhou M., Guo J., Guo Y.Y., Ding K., Wang P., Wang Z.P. (2021a). Analysis of selection signatures on the Z chromosome of bidirectional selection broiler lines for the assessment of abdominal fat content. BMC Genomic Data 22: 18.

Wang F., Gervasi M.G., Bošković A., Sun F., Rinaldi V.D., Yu J., Wallingford M.C., Tourzani D.A., Mager J., Zhu L.J., Rando O.J. (2021b). Deficient spermiogenesis in mice lacking Rlim. Elife 10: e63556.

Waterston R.H., Lindblad-Toh K., Birney E., Rogers J., Abril J.F., Agarwal P., Agarwala R., Ainscough R., Alexandersson M., An P., Antonarakis S.E. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature 420: 520-562.

Wilkerson A.J.P., Raudsepp T., Graves T., Albracht D., Warren W., Chowdhary B.P., Skow L.C., Murphy W.J. (2008). Gene discovery and comparative analysis of X-degenerate genes from the domestic cat Y chromosome. Genomics 92: 329-338.

Witt K.E., Huerta-Sánchez E. (2019). Convergent evolution in human and domesticate adaptation to high-altitude environments. Phill Trans R Soc B 374: 20180235.

Wright S. (1933). Inbreeding and homozygosis. Proc Natl Acad Sci 19: 411-420.

Wu H., Liu Y.H., Wang G.D., Yang C.T., Otecko N.O., Liu F., Wu S.F., Wang L., Yu L., Zhang Y.P. (2016). Identifying molecular signatures of hypoxia adaptation from sex chromosomes: A case for Tibetan Mastiff based on analyses of X chromosome. Sci Rep 6: 35004.

Yang J., Benyamin B., McEvoy B.P., Gordon S., Henders A.K., Nyholt D.R., Madden P.A., Heath A.C., Martin N.G., Montgomery G.W., Goddard M.E., Visscher P.M. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42: 565-569.

Yang J., Lee S.H., Goddard M.E., Visscher P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88: 76-82.

Yengo L., Wray N.R., Visscher P.M. (2019). Extreme inbreeding in a European ancestry sample from the contemporary UK population. Nat Commun 10: 3719.

Young A.C., Kirkness E.F., Breen M. (2008). Tackling the characterization of canine chromosomal breakpoints with an integrated in-situ/in-silico approach: the canine PAR and PAB. Chromosome Res 16: 1193-1202.

Yuan Z., Li W., Li F., Yue X. (2019). Selection signature analysis reveals genes underlying sheep milking performance. Archiv Anim Breed 62: 501-508.

Zavarez L.B., Utsunomiya Y.T., Carmo A.S., Neves H.H., Carvalheiro R., Ferenčaković M., Pérez O'Brien A.M., Curik I., Cole J.B., Van Tassell C.P., da Silva M.V.G.B., Sonstegard T.S., Sölkner J., Garcia J.F. (2015). Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. Front Genet 6: 125037.

Zhao F., McParland S., Kearney F., Du L., Berry D.P. (2015). Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genet Sel Evol 47: 1-12.

Zheng Q., Bourgain C. (2009). Kininbcoef: Calculation of kinship and inbreeding coefficients. Available at: https://www.stat.uchicago.edu/ (Accessed April 12, 2022).

Zhu C., Fan H., Yuan Z., Hu S., Zhang L., Wei C., Zhang Q., Zhao F., Du L. (2015). Detection of selection signatures on the X chromosome in three sheep breeds. Int J Mol Sci 16: 20360-20374.

Zhu C., Li M., Qin S., Zhao F., Fang S. (2020). Detection of copy number variation and selection signatures on the X chromosome in Chinese indigenous sheep with different types of tail. Asian-Aust J Anim Sci 33: 1378.

8. CURRICULUM VITAE

Mario Shihabi, MSc, was born on March 6, 1997, in Zagreb, Croatia. He completed his elementary and high school education as a pharmaceutical technician in Šibenik. In 2015, he enrolled in the Agroecology undergraduate studies at the University of Zagreb Faculty of Agriculture, earning his bachelor's degree in 2019. That same year, he began his graduate studies of Animal Genetics and Breeding, graduating in 2021. During his graduate studies, Mario received the Rector's Award and the Award for the best student of graduate studies at the University of Zagreb Faculty of Agriculture, achieving a perfect GPA of 5.0 and earning 12 ECTS surplus credits. Even as a student, Mario gained experience presenting his research at an international conference and was honored with the Best Poster Award at the "Animal Science Days 2021-ASD" held in Hungary.

Immediately following his graduation, he secured a position as an assistant at the Department of Animal Science within the same faculty, where he is currently working on the Croatian Science Foundation project "Application of NGS Methods in Assessing Genomic Variability of Ruminants-ANAGRAMS". During his PhD, Mario has contributed to nine published papers, two of which he was the first author. He has successfully presented his research at five conferences, received the Best Poster Award at the "International Symposium of Agriculture 2023-SA" held in Dubrovnik, and delivered an invited lecture at the "Animal Science Days 2023-ASD" held in Slovenia. Additionally, he participated in the organization of the "Animal Science Days 2022-ASD" held in Zadar, served as a co-trainer at the AlphaSimR Course, and received several grants and scholarships, including WCGALP2022 Scholarship, Faculty of Agriculture Foundation, Fondazione Edmund Mach Scholarship, City of Zagreb Support, and EAAP2024 Scholarship. As an assistant, Mario is involved in teaching two courses, with his primary areas of expertise encompassing genomics, bioinformatics, and population genetics of domestic animals.

In addition to his academic commitments, he is a member of several extracurricular activities. Through his membership in the Faculty of Agriculture student club (KSA), he has learned the importance of interpersonal relationships in business associations. Working first as a member and later as a project planner and vice president has greatly improved his teamwork and communication skills. Moreover, as a member of the Futsal team of the Faculty of Agriculture at the University of Zagreb since 2015 and the captain since 2018, he has become accustomed to taking responsibility and developed a competitive mentality that drives him forward. In addition, he was elected by the students as a member of the Student Council of the Faculty of Agriculture, the

Extracurricular Activities Commission, and the Faculty Council of the Faculty of Agriculture for the period from 2019 to 2023. Having started his PhD in 2021, with a desire to foster social connections and contribute to his community, he joined the Association of PhD students of the Faculty of Agriculture (UPSAF) and was elected as vice president for the year 2023.

List of publications:

a) Scientific publications from group a1:

1. **Shihabi M.**, Lukic B., Cubric-Curik V., Brajkovic V., Oršanić M., Ugarković D., Vostry L., Curik I. (2022) Identification of Selection Signals on the X-Chromosome in East Adriatic Sheep: A New Complementary Approach. Front Genet 13: 887582. DOI: 10.3389/fgene.2022.887582
2. Drzaic I., Curik I., Lukic B., **Shihabi M.**, Li M.H., Kantanen J., Mastrangelo S., Ciani E., Lenstra J.A., Cubric-Curik V. (2022) High-Density Genomic Characterization of Native Croatian Sheep Breeds. Front Genet 13: 940736. DOI: 10.3389/fgene.2022.940736
3. Djokic M., Drzaic I., **Shihabi M.**, Markovic B., Cubric-Curik V. (2023) Genomic Diversity Analyses of Some Indigenous Montenegrin Sheep Populations. Diversity 15: 640. DOI: 10.3390/d15050640
4. Vostry L., Vostra-Vydrova H., Moravcikova N., Kasarda R., Cubric-Curik V., Brzakova M., Solkner J., **Shihabi M.**, Moreno J.A.H., Spehar M., Curik I. (2023) Genomic diversity and population structure of the Czech Holstein cattle. Livest Sci 273: 105261. DOI: 10.1016/j.livsci.2023.105261
5. Drzaic I., Orehovački V., Moravčikova N., **Shihabi M.**, Curik I., Vostry L., Kasarda R., Soelkner J., Cubric-Curik V. (2023) Genomic characterization and diversity of indigenous goat breeds from Croatia. Livest Sci 105388. DOI: 10.1016/j.livsci.2023.105388
6. Lukic B., Curik I., Drzaic I., Galić V., **Shihabi M.**, Vostry L., Cubric-Curik V. (2023) Genomic signatures of selection, local adaptation and production type characterisation of East Adriatic sheep breeds. J Anim Sci Biotechnol 14: 1-17. DOI: 10.1186/s40104-023-00936-y
7. Vostry L., Vostra-Vydrova H., Moravcikova N., Kasarda R., Margetin M., Rychtarova J., Drzaic I., **Shihabi M.**, Cubric-Curik V., Sölkner J., Curik I. (2024) Genomic analysis of conservation status, population structure and admixture in local Czech and Slovak dairy goat breeds. J Dairy Sci. DOI: 10.3168/jds.2023-24607

b) Scientific publications from group a3:

1. **Shihabi M.**, Glavaš V., Cubric-Curik V., Zorc M., Dovč P., Curik I. (2022). Genome-wide signals of positive selection identified in Livestock Guardian Dogs. DOI: 10.3920/978-90-8686-940-4_753
2. Curik I., Vostra-Vydrova H., **Shihabi M.**, Sölkner J., Vostry L. (2022). Estimation of sex chromosome inbreeding depression on milk production in cattle. DOI: 10.3920/978-90-8686-940-4_205

List of publications from the Croatian Scientific Bibliography (CROSBI):
https://www.croris.hr/osobe/profil/37895

9. APPENDICES

Appendix 1. F_{ROH} mean values and SE estimated with the empirical consecutive approach (SVS) using different minimum ROH lengths on autosomes, X chromosome and PAR in females of each domestic animal population.

Population	Genomic landscape	F_{ROH}	$F_{ROH>2Mb}$	$F_{ROH>4Mb}$	$F_{ROH>8Mb}$
Croatian cattle breeds	Autosomes	0.087±0.009	0.079±0.009	0.071±0.009	0.058±0.007
	X chromosome	0.255±0.013	0.198±0.014	0.152±0.015	0.102±0.016
	PAR	0.065±0.023	0.056±0.022	0.046±0.021	/
Nellore	Autosomes	0.046±0.001	0.034±0.001	0.025±0.005	0.014±0.000
	X chromosome	0.124±0.002	0.084±0.002	0.054±0.002	0.024±0.002
	PAR	0.013±0.002	0.007±0.001	0.002±0.001	/
Labrador Retriever	Autosomes	0.213±0.006	0.178±0.006	0.139±0.006	0.091±0.005
	X chromosome	0.396±0.012	0.337±0.013	0.266±0.014	0.195±0.015
	PAR	0.158±0.020	0.110±0.019	0.046±0.015	/
Patagonian Sheepdog	Autosomes	0.109±0.009	0.071±0.009	0.046±0.008	0.029±0.007
	X chromosome	0.367±0.017	0.319±0.018	0.245±0.020	0.176±0.021
	PAR	0.041±0.020	0.018±0.018	0.018±0.018	/
Croatian sheep breeds	Autosomes	0.062±0.007	0.054±0.007	0.047±0.007	0.038±0.006
	X chromosome	0.144±0.014	0.118±0.014	0.091±0.015	0.067±0.015
	PAR	0.057±0.018	0.050±0.180	0.038±0.017	/
Soay	Autosomes	0.255±0.002	0.191±0.002	0.109±0.002	0.045±0.002
	X chromosome	0.422±0.009	0.373±0.009	0.295±0.010	0.209±0.010
	PAR	0.112±0.014	0.053±0.011	0.016±0.006	/

Appendix 2. F_{ROH} mean values and SE estimated with the empirical consecutive approach (SVS) using different minimum ROH lengths on autosomes and PAR in males of each domestic animal population.

Population	Genomic landscape	F_{ROH}	$F_{ROH>2Mb}$	$F_{ROH>4Mb}$	$F_{ROH>8Mb}$
Croatian cattle breeds	Autosomes	0.136±0.017	0.127±0.016	0.115±0.016	0.092±0.014
	PAR	0.142±0.038	0.106±0.038	0.023±0.023	/
Nellore	Autosomes	0.072±0.001	0.059±0.001	0.047±0.001	0.032±0.001
	PAR	0.029±0.004	0.016±0.003	0.004±0.002	/
Labrador Retriever	Autosomes	0.199±0.004	0.164±0.004	0.125±0.004	0.080±0.004
	PAR	0.091±0.012	0.035±0.010	0	/
Patagonian Sheepdog	Autosomes	0.107±0.006	0.068±0.007	0.044±0.006	0.029±0.006
	PAR	0.039±0.013	0.026±0.012	0.010±0.000	/
Croatian sheep breeds	Autosomes	0.044±0.005	0.038±0.005	0.032±0.005	0.025±0.004
	PAR	0.023±0.011	0.017±0.010	0.009±0.009	/
Soay	Autosomes	0.253±0.002	0.187±0.002	0.106±0.002	0.043±0.002
	PAR	0.124±0.018	0.044±0.015	0.008±0.008	/

Appendix 3. F_{ROH} mean values and SE estimated with the statistical approach (RZooROH) for each HBD category on autosomes, X chromosome and PAR in females of each domestic animal population.

Pop	Gen land	F_{ROH}	F_{ROH_HBD1}	F_{ROH_HBD2}	F_{ROH_HBD3}	F_{ROH_HBD4}	F_{ROH_HBD5}	F_{ROH_HBD6}	F_{ROH_HBD7}
CCB	Auto	0.211±0.008	0.006±0.003	0.041±0.006	0.003±0.001	0	0	0	0.161±0.002
	chrX	0.264±0.012	0.027±0.012	0.031±0.011	0.010±0.005	0.001±0.001	0	0	0.195±0.006
	PAR	0.110±0.022	0	0.016±0.013	0.011±0.011	0	0	0	0.083±0.016
N	Auto	0.148±0.000	0	0.004±0.003	0.005±0.000	0	0	0	0.138±0.000
	chrX	0.236±0.002	0.002±0.001	0.008±0.001	0.005±0.001	0.001±0.000	0	0	0.220±0.001
	PAR	0.017±0.001	0	0	0	0	0	0	0.017±0.001
LR	Auto	0.220±0.005	0.002±0.002	0.008±0.003	0.034±0.005	0.072±0.004	0	0	0.104±0.001
	chrX	0.272±0.013	0.022±0.009	0.041±0.012	0.068±0.011	0.029±0.007	0.032±0.006	0.012±0.003	0.068±0.004
	PAR	0.172±0.019	0.001±0.000	0.003±0.003	0.024±0.011	0.016±0.009	0.007±0.004	0.012±0.006	0.110±0.014
PS	Auto	0.101±0.009	0	0.009±0.004	0.017±0.005	0.005±0.002	0	0	0.069±0.002
	chrX	0.176±0.021	0.028±0.020	0.025±0.012	0.011±0.007	0.014±0.007	0.023±0.009	0.006±0.003	0.069±0.006
	PAR	0.044±0.018	0	0.017±0.017	0	0	0	0	0.027±0.008
CSB	Auto	0.137±0.006	0.001±0.000	0.026±0.006	0.008±0.002	0	0	0	0.102±0.001
	chrX	0.226±0.013	0.037±0.013	0.008±0.006	0.011±0.004	0.002±0.002	0	0	0.168±0.005
	PAR	0.097±0.018	0.013±0.010	0.003±0.003	0.010±0.008	0.001±0.001	0	0	0.069±0.012
S	Auto	0.321±0.002	0	0	0.011±0.002	0.020±0.002	0	0	0.290±0.001
	chrX	0.453±0.009	0.008±0.005	0.023±0.008	0.091±0.013	0.085±0.012	0.014±0.006	0.003±0.003	0.229±0.008
	PAR	0.239±0.014	0	0	0.006±0.004	0	0.004±0.004	0.004±0.004	0.228±0.014

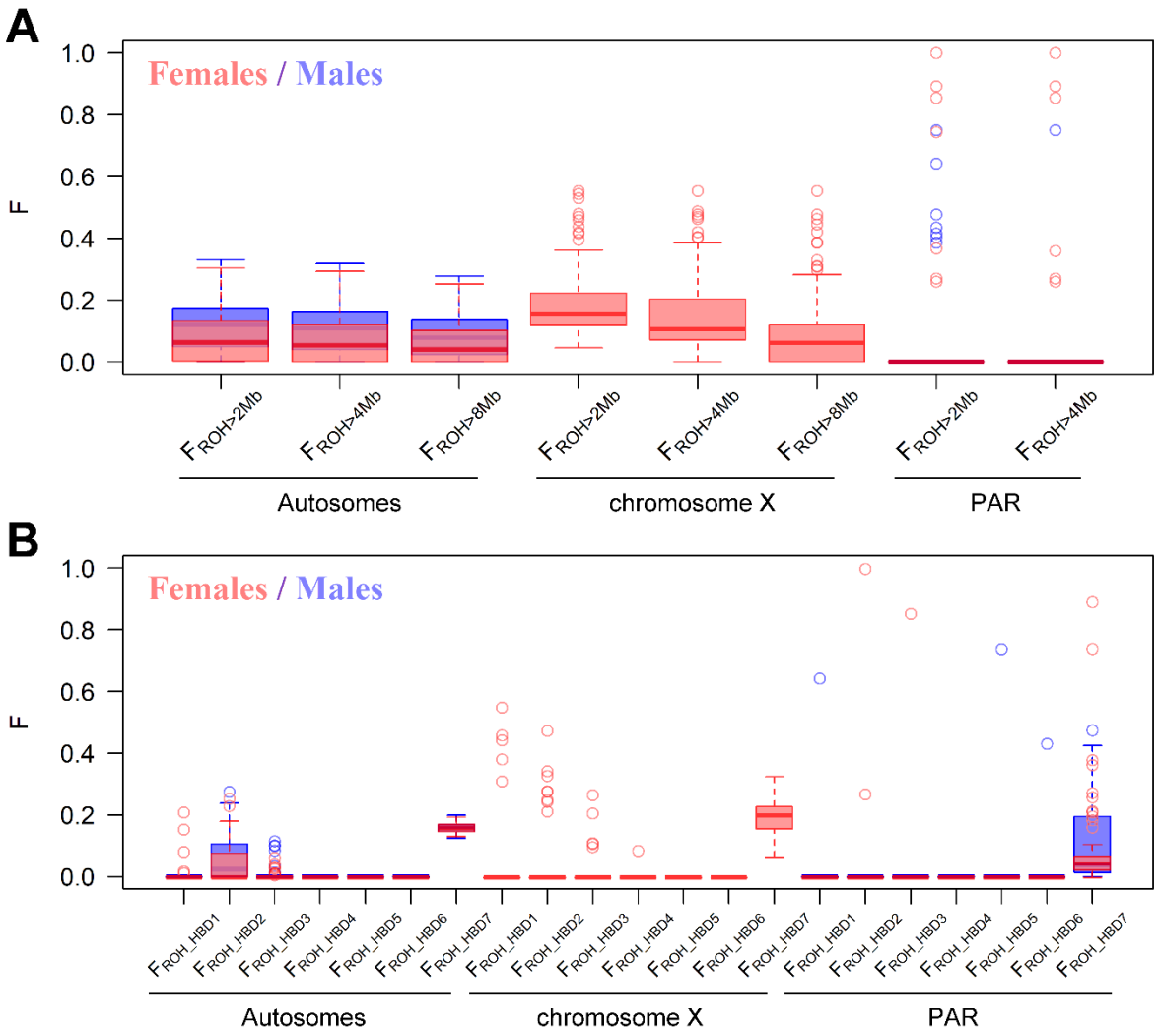
Pop = Population: CCB = Croatian cattle breeds, N = Nellore, LR = Labrador Retriever, PS = Patagonian Sheepdog, CSB = Croatian sheep breeds, S = Soay. Gen land = Genomic landscape: Auto = Autosomes, chrX = X chromosome.

Appendix 4. F_{ROH} mean values and SE estimated with the statistical approach (RZooROH) for each HBD category on autosomes and PAR in males of each domestic animal population.

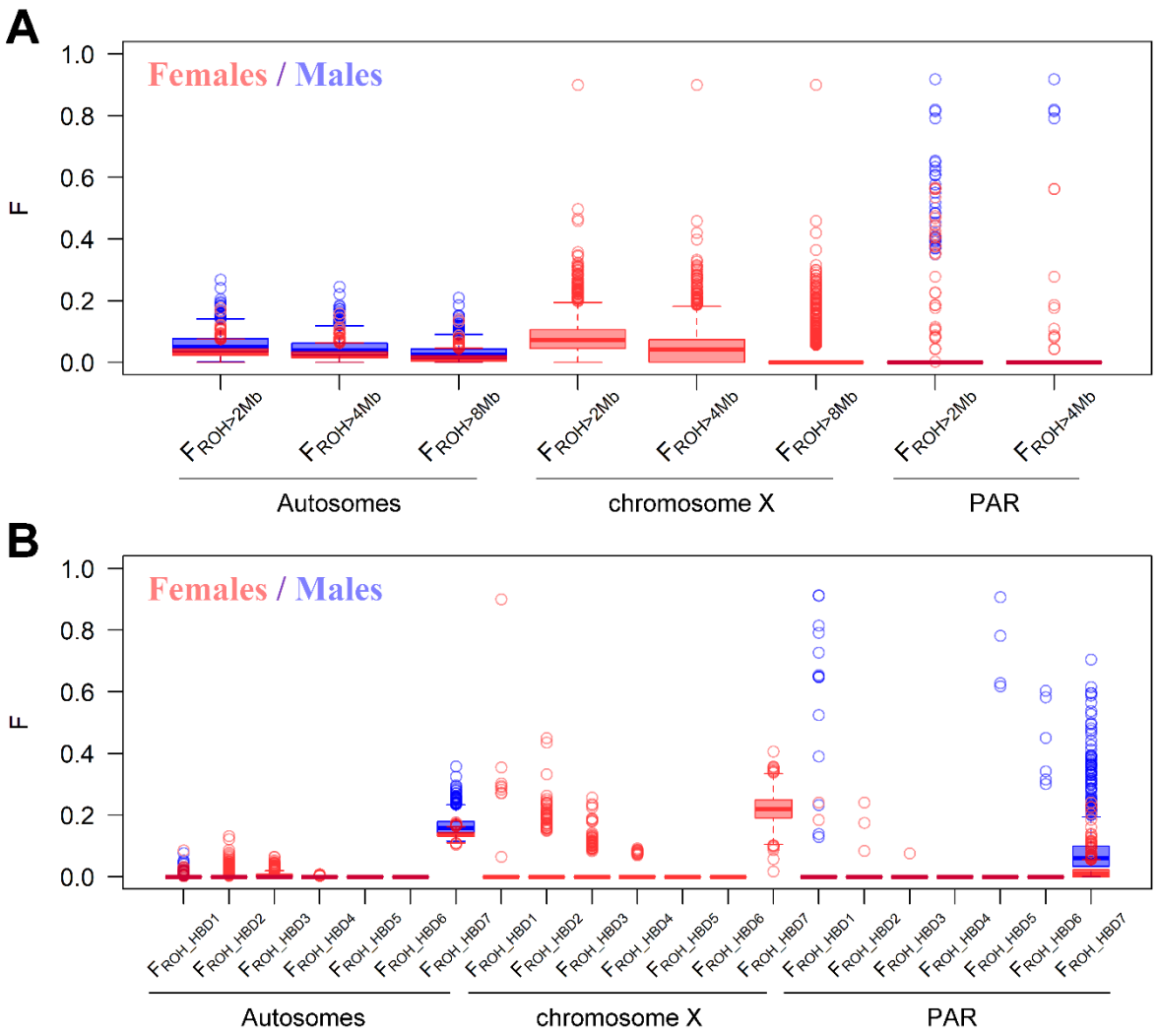
Pop	Gen land	F_{ROH}	F_{ROH_HBD1}	F_{ROH_HBD2}	F_{ROH_HBD3}	F_{ROH_HBD4}	F_{ROH_HBD5}	F_{ROH_HBD6}	F_{ROH_HBD7}
CCB	Auto	0.242±0.014	0	0.069±0.015	0.014±0.006	0	0	0	0.160±0.003
	PAR	0.173±0.036	0.019±0.019	0	0	0	0.022±0.022	0.013±0.013	0.118±0.026
N	Auto	0.166±0.001	0.001±0.000	0	0	0	0	0	0.165±0.001
	PAR	0.100±0.004	0.009±0.003	0	0	0	0.003±0.002	0.003±0.001	0.086±0.003
LR	Auto	0.206±0.004	0.001±0.000	0.003±0.002	0.028±0.004	0.072±0.004	0	0	0.102±0.001
	PAR	0.118±0.012	0	0	0	0	0.040±0.004	0.018±0.007	0.096±0.010
PS	Auto	0.099±0.006	0.003±0.002	0.008±0.004	0.014±0.004	0.006±0.001	0	0	0.069±0.001
	PAR	0.047±0.013	0.010±0.010	0	0	0	0	0.015±0.008	0.022±0.005
CSB	Auto	0.119±0.005	0.001±0.000	0.015±0.004	0.007±0.002	0	0	0	0.097±0.002
	PAR	0.051±0.012	0	0	0	0	0	0.009±0.009	0.042±0.008
S	Auto	0.319±0.002	0	0	0.011±0.002	0.019±0.002	0	0	0.290±0.002
	PAR	0.259±0.016	0	0	0	0	0.005±0.005	0.005±0.005	0.249±0.016

Pop = Population: CCB = Croatian cattle breeds, N = Nellore, LR = Labrador Retriever, PS = Patagonian Sheepdog, CSB = Croatian sheep breeds, S = Soay. Gen land = Genomic landscape: Auto = Autosomes, chrX = X chromosome.

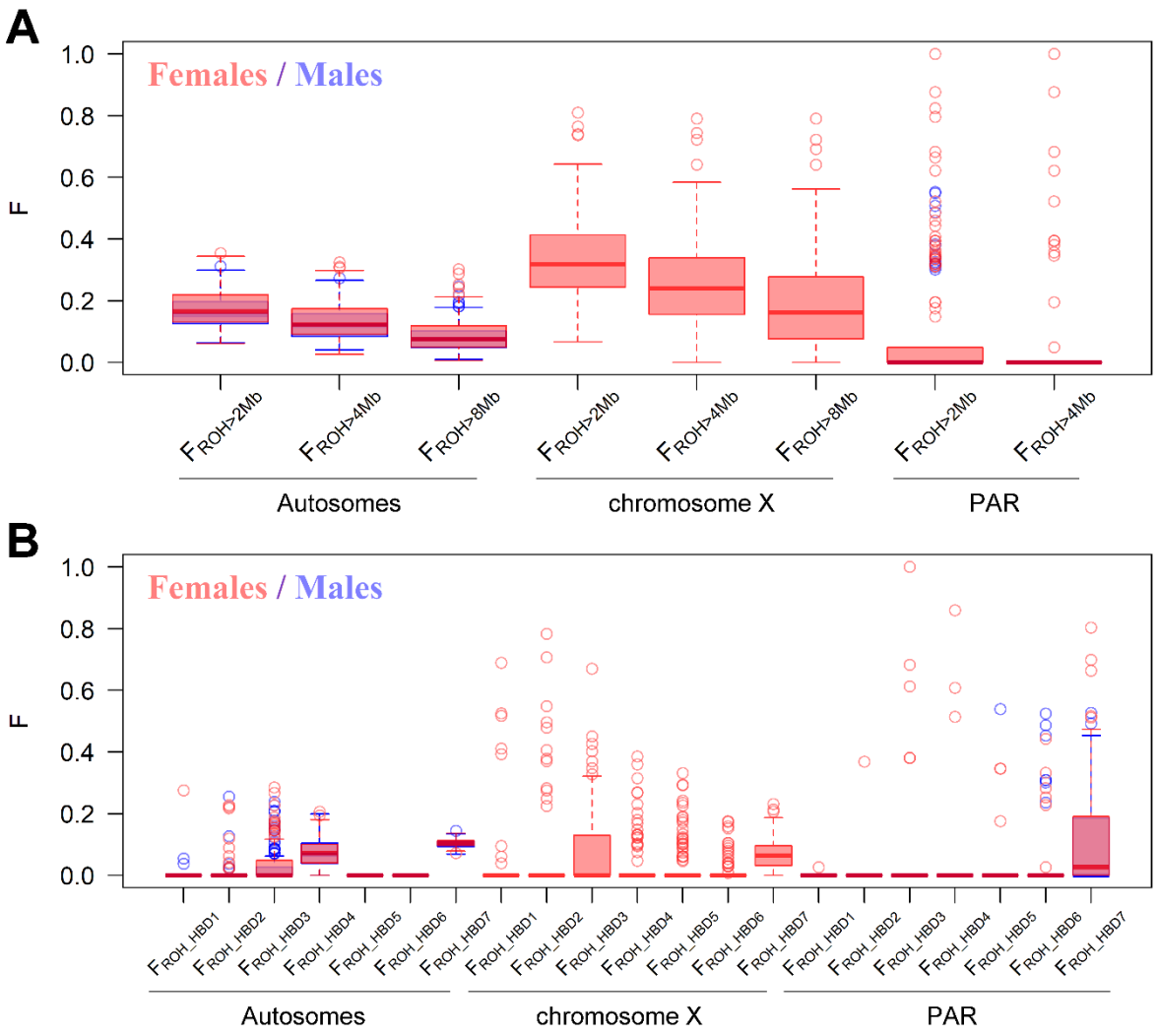
Appendix 5. Distribution of F_{ROH} values estimated with SVS using different minimum ROH lengths **(A)** and RZooROH by each HBD category **(B)** on autosomes, X chromosome and PAR in metapopulation of native Croatian cattle breeds, separated by sex.



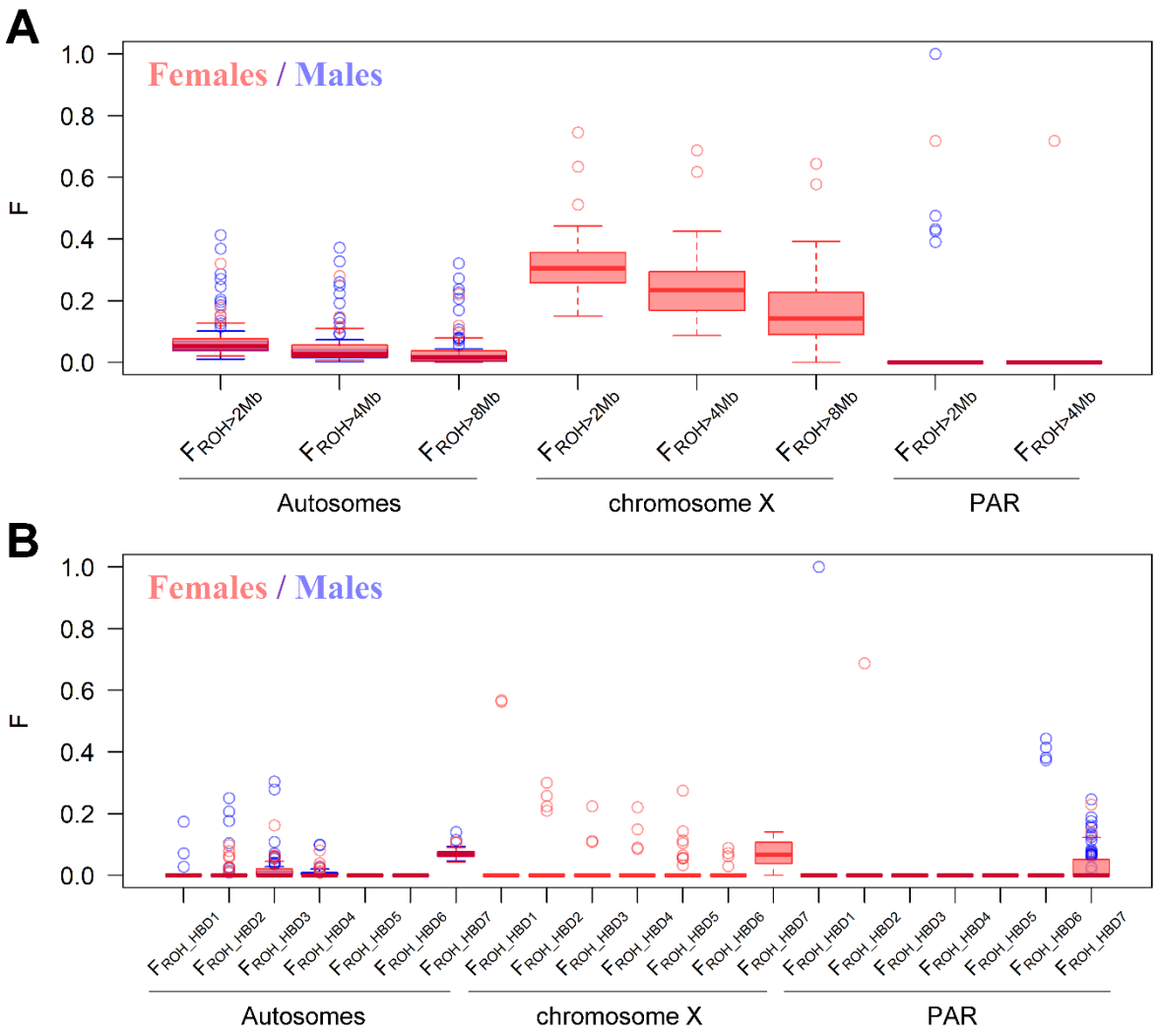
Appendix 6. Distribution of F_{ROH} values estimated with SVS using different minimum ROH lengths **(A)** and RZooROH by each HBD category **(B)** on autosomes, X chromosome and PAR in Nellore, separated by sex.



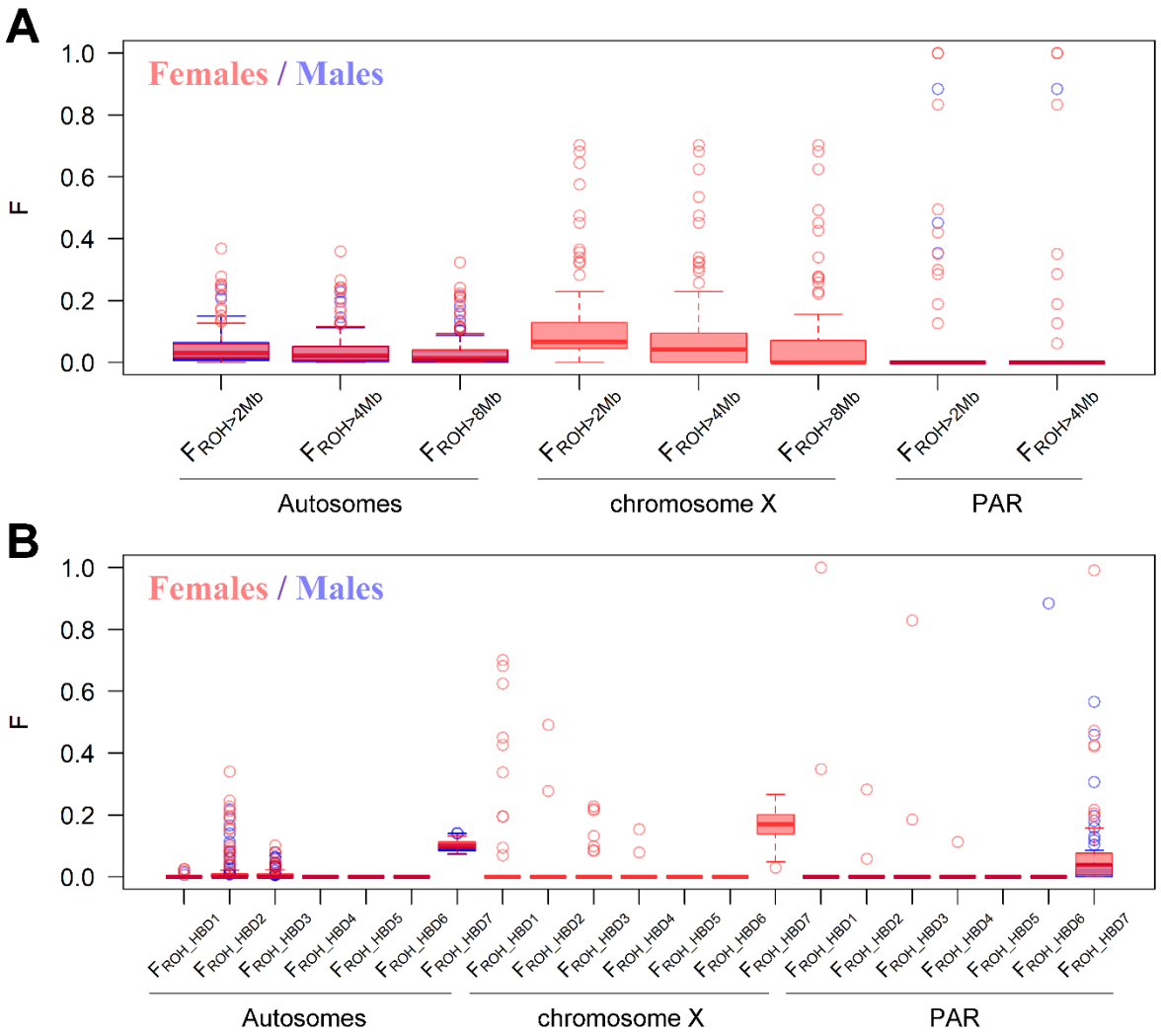
Appendix 7. Distribution of F_{ROH} values estimated with SVS using different minimum ROH lengths **(A)** and RZooROH by each HBD category **(B)** on autosomes, X chromosome and PAR in Labrador Retriever, separated by sex.



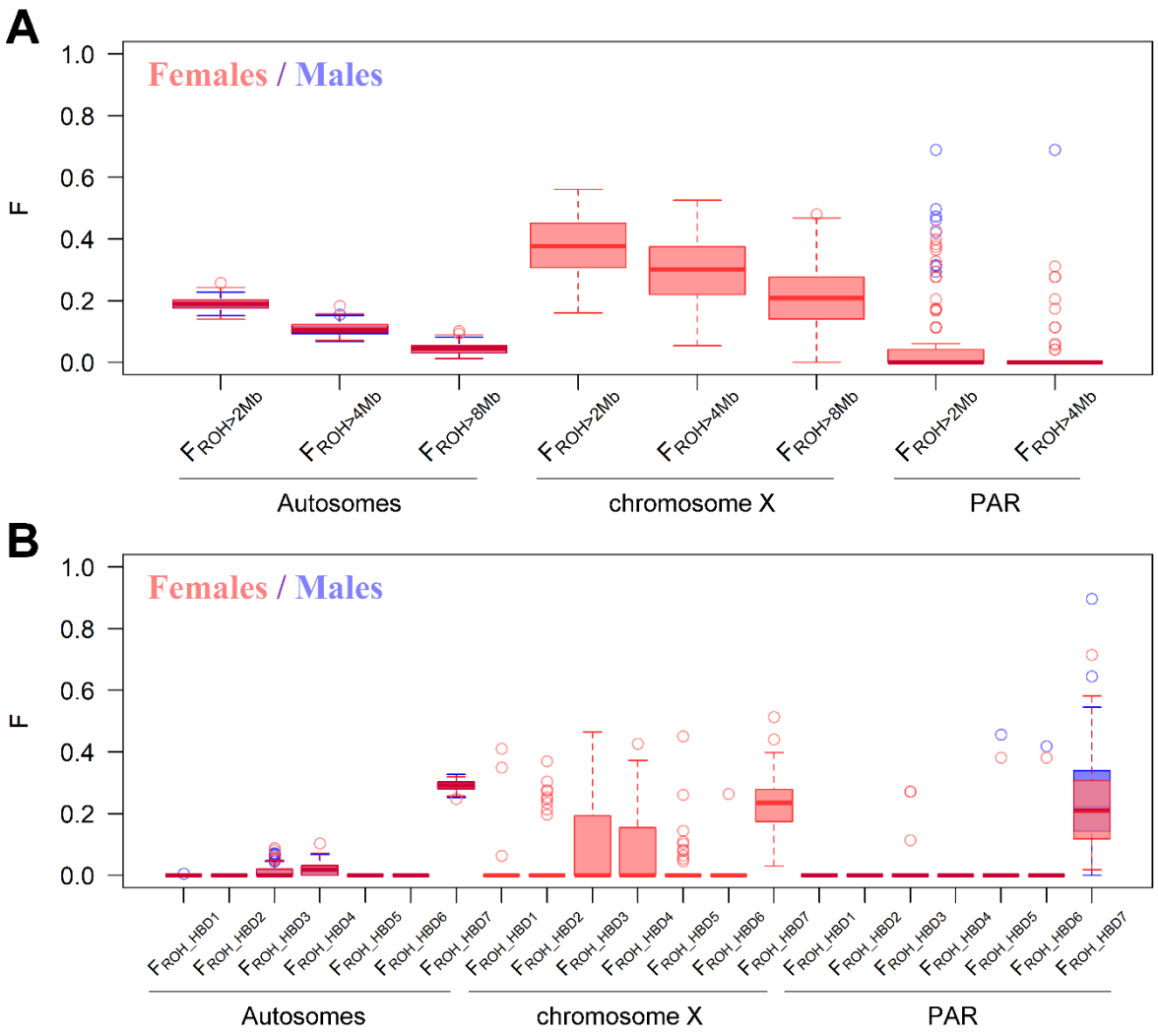
Appendix 8. Distribution of F_{ROH} values estimated with SVS using different minimum ROH lengths **(A)** and RZooROH by each HBD category **(B)** on autosomes, X chromosome and PAR in Patagonian Sheepdog, separated by sex.



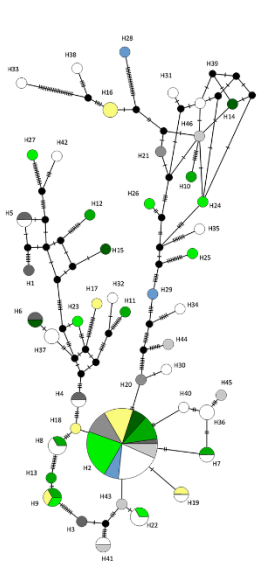
Appendix 9. Distribution of F_{ROH} values estimated with SVS using different minimum ROH lengths **(A)** and RZooROH by each HBD category **(B)** on autosomes, X chromosome and PAR in metapopulation of native Croatian sheep breeds, separated by sex.



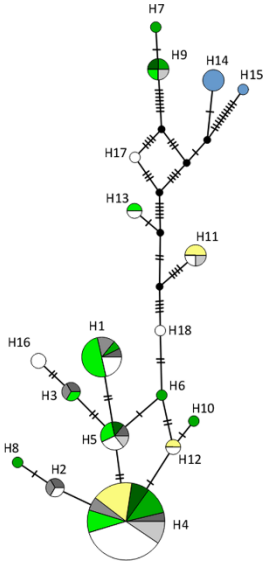
Appendix 10. Distribution of F_{ROH} values estimated with SVS using different minimum ROH lengths **(A)** and RZooROH by each HBD category **(B)** on autosomes, X chromosome and PAR in Soay, separated by sex.



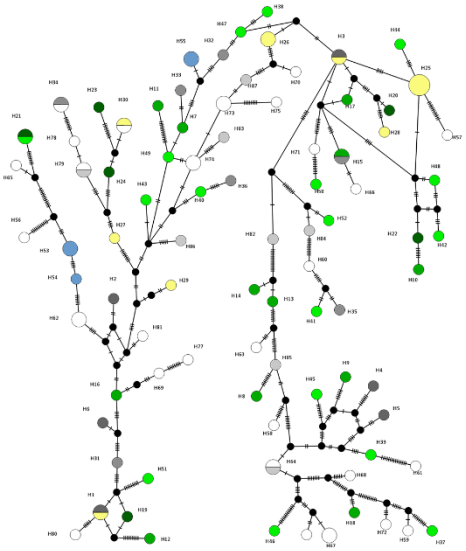
Appendix 11. MJNs showing the phylogenetic relationship between haplotypes for the first five mapped candidate regions with respect to their breed origin. The size of the haplotypes (number of SNPs) is indicated in parentheses.



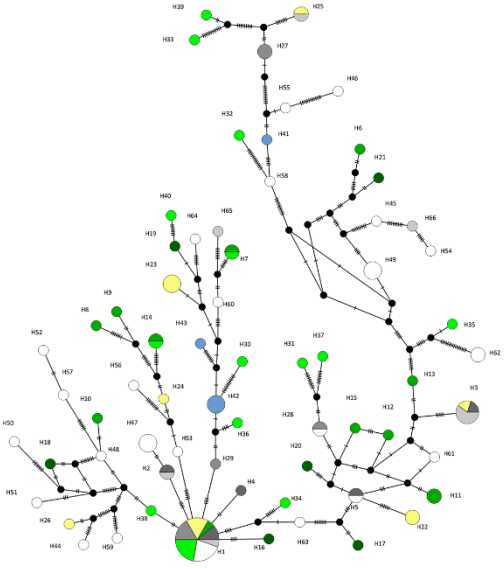
13.04 - 13.69 Mb (78)



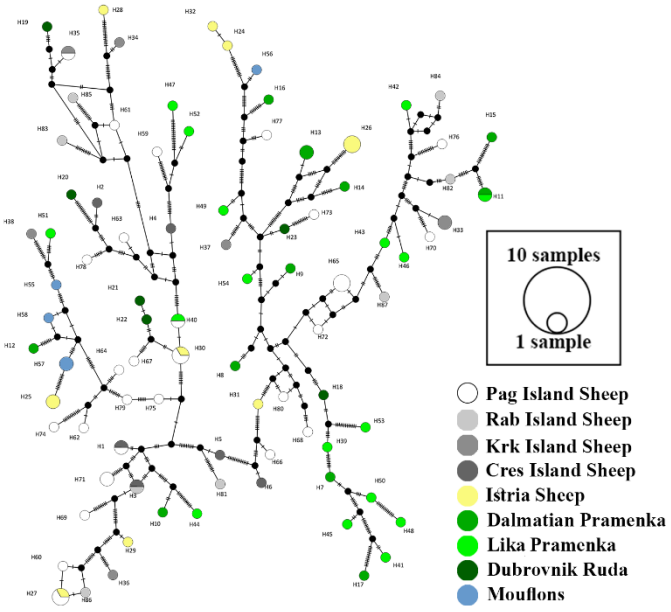
21.96 - 22.26 Mb (34)



32.20 - 32.80 Mb (77)

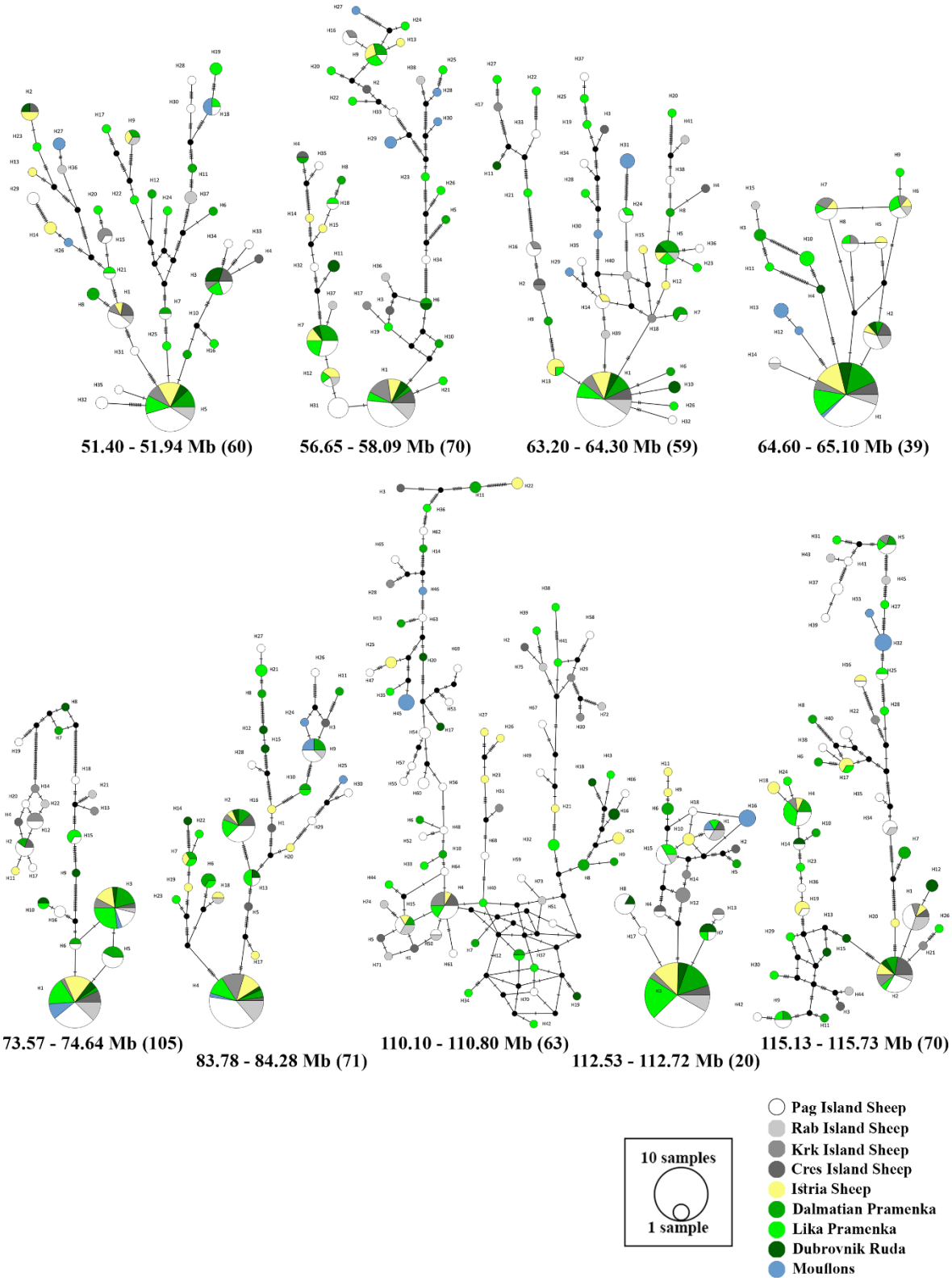


41.00 - 41.50 Mb (74)



42.50 - 43.00 Mb (76)

Appendix 12. MJNs showing the phylogenetic relationship between haplotypes for the last six mapped candidate regions with respect to their breed origin. The size of the haplotypes (number of SNPs) is indicated in parentheses.



Appendix 13. Comparison of selection candidate regions mapped at the X chromosome in metapopulation of Croatian sheep breeds with other sheep studies.

Candidate region (Mb)	Overlapped signal (Mb)	Candidate genes	Reference
13.04−13.69	13.20−13.60	*CA5B*, *ZRSR2*, *AP1S2*, *GRPR*	Chen *et al*. (2018)
21.96−22.26	/		
32.20−32.80	31.98−32.38	/	Zhu *et al*. (2015)
	32.40−32.60	/	Chen *et al*. (2018)
	32.10−32.40	/	Liu *et al*. (2016)
41.00−43.00	41.36−41.76	*EFHC2*	Zhu *et al*. (2015)
	41.20−41.40	*EFHC2*	Chen *et al*. (2018)
	38.94−44.20	*NDP, EFHC2*	Cesarani *et al*. (2022)
51.40−51.94	51.17−51.91	*SHROOM4*, *DGKK*, *CCNB3*	Zhu *et al*. (2015)
	51.18−51.56	*SHROOM4*	Zhu *et al*. (2020)
56.64−58.09	57.4−58.00	*OPHN1*, *YIPF6*	Chen *et al*. (2018)
	56.50−58.00	*AR*, *OPHN1*, *YIPF6*	Liu *et al*. (2016)
	55.78−58.33	*AR*, *OPHN1*, *YIPF6*	Cesarani *et al*. (2022)
	55.79−58.35	*AR*, *OPHN1*, *YIPF6*	Manzari *et al*. (2019)
63.20−65.10	63.09−63.46	*RLIM*, *KIAA2022*, *ABCB7*, *MAGT1*, *ATRX*, *FGF16*	Zhu *et al*. (2020)
	64.80−65.00	*ATRX, FGF16*	Chen *et al*. (2018)
	64.95−65.05	*ATRX, FGF16*	Kardos *et al*. (2015)
73.57−74.54	73.69−74.09	*CHM*, *DACH2*	Zhu *et al*. (2015)
	73.44−73.94	*CHM*, *DACH2*	Zhu *et al*. (2020)
83.78−84.28	/		
110.10−110.80	110.20−110.40	*DOCK11*	Chen *et al*. (2018)
112.53−112.72	112.66−113.06	*PLS3*	Zhu *et al*. (2015)
115.30−115.73	/		