Federal University of Piauí
Center for Agricultural Sciences
Graduate Program in Tropical Animal Science

ADAPTATION HOTSPOTS REVEALED BY GENOMICS AND EXTINCTION RISK IN BRAZILIAN GOAT POPULATIONS

FRANCISCO DE ASSIS DINIZ SOBRINHO

Teresina-PI, August 2024

FRANCISCO DE ASSIS DINIZ SOBRINHO

ADAPTATION HOTSPOTS REVEALED BY GENOMICS AND EXTINCTION RISK IN BRAZILIAN GOAT POPULATIONS

Phd Tesis Graduate Program in Tropical Animal Science of Federal University of Piauí (UFPI), as a requirement for obtaining the degree of Doctor in Tropical Animal Science. Concentration Area: Animal Production in the Tropics

Federal University of Piauí – UFPI

Center for Agricultural Sciences

Graduate Program in Tropical Animal Science

Supervisor: Dra. Adriana de Mello Araújo

Co-supervisor: Dr. Miklos Maximiliano Bajay

Teresina-PI

August 2024

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Setorial CCA
Serviço de Representação da Informação

D585a Diniz Sobrino, Francisco de Assis.
Adaptation hotspots revealed by genomics and extinction risk in brazilian goat populations. / Francisco de Assis Diniz Sobrino. -- 2024.
86 f.

Tese (Doutorado) – Universidade Federal do Piauí, Centro de Ciências Agrárias, Programa de Pós-Graduação em Zootecnia Tropical, 2024.
"Supervisora: Profa. Dra. Adriana de Mello Atraújo."

1. Local goat populations. 2. Genomic regions. 3. Selection signatures. I. Araújo, Adriana de Mello. II. Título.

CDD 636.39

Bibliotecário: Rafael Gomes de Sousa - CRB3/1163

HOTSPOTS DE ADAPTAÇÃO REVELADOS PELA GENÔMICA E RISCO DE DESAPARECIMENTO NAS POPULAÇÕES CAPRINAS BRASILEIRAS

FRANCISCO DE ASSIS DINIZ SOBRINHO

Tese aprovada em: 20/08/2024

Banca Examinadora:

Documento assinado digitalmente
gov.br ADRIANA MELLO DE ARAUJO
Data: 28/08/2024 17:32:15-0300
Verifique em https://validar.iti.gov.br

Profa. Dra. Adriana Mello de Araújo (Presidente) / EMBRAPA

Documento assinado digitalmente
gov.br MIKLOS MAXIMILIANO BAJAY
Data: 22/08/2024 19:16:16-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Miklos Maximiliano Bajay (Externo) / UDESC

Documento assinado digitalmente
gov.br DANIELLE MARIA MACHADO RIBEIRO AZEVEDO
Data: 29/08/2024 12:29:08-0300
Verifique em https://validar.iti.gov.br

Profa. Dra. Danielle Maria Machado Ribeiro Azevedo (Externa) / EMBRAPA

Documento assinado digitalmente
gov.br FABIO BARROS BRITTO
Data: 30/08/2024 17:43:22-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Fábio Barros Britto (Interno) / CCN/UFPI

Documento assinado digitalmente
gov.br JEANE DE OLIVEIRA MOURA
Data: 02/09/2024 10:43:02-0300
Verifique em https://validar.iti.gov.br

Profa. Dra. Jeane de Oliveira Moura (Externa) / IFPI

Documento assinado digitalmente
gov.br LEONARDO CASTELO BRANCO CARVALHO
Data: 04/09/2024 08:48:19-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Leonardo Castelo Branco Carvalho (Externo) / UFRA

To God for the gift of life.

To everyone who always supports the achievement of my goals, especially my family. Also, to those who contributed to the completion of this work.

Acknowledgements

First and foremost, I thank God for the opportunity to complete this work and my family for their unconditional support. I am grateful for the learning experience and the ability to grow personally and professionally in the face of all the difficulties and challenges that have arisen.

also thank my advisor, Prof. Dr. Adriana Melo de Araújo, and my co-advisor, Dr. Miklos Maximiliano Bajay, for all their academic support, facilitation, constructive criticism, guidance, dedication, and availability. In this context, I also want to thank the project's direct collaborators, including Dr. Igor Ferreira do Nascimento, Dr. Danielle Maria M. Ribeiro Azevêdo, Dr. Ronaldo Cunha Coelho, Dr. Fábio Britto, Dr. Jeane de Oliveira Moura, Dr. José Lindbergh Sarmento, Dr. Giorgio Chessari and Dr. Salvatore Mastrangelo.

I extend my gratitude to the Federal Institute of Piauí (IFPI) for their support in completing this project, and to my friends, Pastors João Barbosa Lopes and Valdir Borges, for their prayers and support.

I offer a special thanks to the Graduate Program in Tropical Animal Science and the Center for Agricultural Sciences for all the support in completing this project. I emphasize the importance of the Federal University of Piauí (UFPI) in my academic life, where I have been a student since 1999. I started with a degree in Biological Sciences, completed a master's in Animal Science, and a PhD in Tropical Animal Science. I thank everyone who, in some way, contributed to the completion of this work.

Thank you!

"I am the light of the world; whoever follows me will never walk in darkness, but will have the light of life. "

(Jesus Christ)

Abstract

Autozygosity refers to the homozygous state of alleles that are identical by descent (IBD), which can occur due to various genetic phenomena. An important factor leading to autozygosity is increased inbreeding (F), which often results in negative effects such as reduced genetic diversity, decreased individual performance (inbreeding depression), and lower population viability. Among the available methods for estimating inbreeding, the one based on Runs of Homozygosity (ROH) is currently considered the most accurate. This study aims to provide a comprehensive analysis of the evolution of research in genetic diversity and ROH over time. To achieve this, we utilized the Web of Science (WoS) database to identify key publications, emerging research areas, and major trends in this field. The bibliometric analysis covered publications from 2009 to November 2023. We used the Bibliometrix package (version 4.1.3) in the R environment (version 4.3.1), supplemented by the Biblioshiny (version 4.1.3) graphical interface, to visualize and analyze these trends. The bibliometric study revealed an annual increase in the number of publications, with notable contributions from China, Italy, and the United States, particularly in prestigious journals. This growing interest reflects advancements in the functional analysis of ROH, driven by computational tools such as PLINK and detectRUNS, which apply sliding window methodologies for SNP analysis. However, challenges such as the loss of genetic diversity, habitat fragmentation, and limited variation in domesticated species persist. Research in ROH and genetic diversity has been propelled by technological advances in genomics and increasing international collaboration, with significant contributions from Asia, the Americas, and Europe. In addition to the bibliometric analysis, this study also focused on the genomic analysis of Runs of Homozygosity (ROH) and Heterozygosity-Rich Regions (HRR) in two goat breeds: the locally adapted Marota and the commercially important Anglo-Nubian. We used SNP data generated by the Illumina Goat SNP50 BeadChip to identify ROH and HRR and calculate the inbreeding coefficient based on ROH (FROH). For this analysis, we used the detectRUNS package in the R environment. Our genomic study revealed a total of 22,872 ROH, with the average number of ROH per individual varying between breeds—74.73 in Anglo-Nubian goats and 173.85 in Marota goats. Most of the detected ROH were short (<2Mb), constituting 65.60Additionally, we identified important candidate genes such as Zbtb11, IL18, and LPO, which overlapped with ROH and HRR regions and were significantly enriched in immune-related biological processes. Other candidate genes, including Tex12, Ypel5, Capn14, and Galnt14, were associated with traits related to embryonic development, body growth, lipid homeostasis, and brain functions. These findings provide valuable insights for breed conservation and the improvement of goat production systems.

Keywords: local goat populations; genomic regions; selection signatures; coefficient of inbreeding; scientific production

List of Figures

List of Tables

Contents

1 GENERAL INTRODUCTION

Among the livestock species of interest, caprine genetic resources play a crucial role in economic development, contributing significantly to large-scale production, especially of milk, and to rural subsistence, increasing the income of families in developing countries(Lin *et al.*, 2013; Periasamy *et al.*, 2017). In the Brazilian context, rural activity with small ruminants is divided between large producers and small breeders, with goats (*Capra hircus*) being a prominent presence, adapted to the semi-arid tropics (Moura *et al.*, 2019a; Moura *et al.*, 2019b).

Various goat breeds adapted to the Brazilian semi-arid tropical climate are valuable reservoirs of genes related to resistance to adverse conditions, parasites, and diseases, providing rich genetic diversity(Boettcher *et al.*, 2015; Diniz *et al.*, 2023). Therefore, the genomic characterization of these resources is essential to inform conservation and sustainable breeding policies(Chokoe *et al.*, 2022; Colli *et al.*, 2014; Persichilli *et al.*, 2021; Tolone *et al.*, 2022).

The Marota goat, raised in the semi-arid region, is a specific local type of small, rustic goats at risk of extinction maintained by governmental institutions and partners in the Brazilian Northeast. This variety constitutes a valuable reservoir of genetic diversity, as they can be considered a source of genes related to the ability to minimize the negative effects of climate and environmental changes, as well as resistance to parasites and diseases, attributable to their evolutionary history of adaptability to adverse environmental conditions (Araujo *et al.*, 2009; Moura *et al.*, 2019b; Diniz *et al.*, 2023). Despite their long history in Brazil, their molecular evolution and associated inbreeding processes still need to be better understood.

Therefore, the genomic characterization of these genetic resources is desirable to support and reference the development of conservation and sustainable breeding policies for this biological heritage. In this context, genomic scanning, especially focusing on runs of homozygosity (ROH), can provide crucial insights into molecular evolution, adaptability, and inbreeding processes in commercial and local goat populations (Marras *et al.*, 2015; Zavarez *et al.*, 2015; Ceballos *et al.*, 2018; Mulim *et al.*, 2022).

ROH becomes a vital tool for distinguishing between recent and ancient inbreeding and revealing selection signatures associated with adaptive and economic traits (Marras *et al.*, 2015; Zavarez *et al.*, 2015; Ceballos *et al.*, 2018; Mulim *et al.*, 2022). Knowing that ROH is not randomly distributed due to genetic recombination, the formation of ROH islands may result from selective pressure (Mulim *et al.*, 2022). Therefore, ROH analysis has been employed to explore selection signatures in various species, including ruminants

and non- ruminants(Marchesi *et al.*, 2018; Xu *et al.*, 2019; Cendron *et al.*, 2020; Mulim *et al.*, 2022).

Considering the evolutionary factors that impact population genomic structure, such as demography, recombination, and intraspecific selective pressure, this study reports a comprehensive genome-wide analysis with the characterization of runs of homozygosity (ROH) and adaptive selection signatures of two goat breeds from institutional herds, based on genotypes generated by the Goat SNP50 BeadChip.

Thus, this work offers a significant contribution to understanding the genetic diversity and evolutionary processes in Brazilian goat populations, highlighting the importance of ROH analysis and selection signatures for the conservation and sustainable management of these genetic resources. The main objectives of this study include providing a comprehensive understanding of the structure and dynamics of genomic diversity and the impact of publications on ROH over the past decade, as well as highlighting directions and opportunities for future research, characterizing ROH and Heterozygosity-Rich Regions (HRR) genome-wide, detecting positive selection signatures, and estimating inbreeding coefficients based on genomic information.

This Thesis adheres to the guidelines of the Manual for the Standardization of Monographs, Dissertations, and Theses of the Federal University of Piauí. Chapters I and II were prepared in accordance with the publication standards of the journals GENETICS and MOLECULAR BIOLOGY (ISSN 0100-8455) and PLOS ONE (ISSN 1932-6203), respectively. The study structure was segmented to optimize the approach to different topics. Initially, the literature review is highlighted, covering the themes explored in Chapters I and II, serving as the theoretical basis. The bibliographic references pertinent to this context were duly incorporated.

2 Genetic diversity and runs of homozygosity (ROH): A portrait of the quantitative academic publication dynamic and scientific metadata

Abstract

Autozygosity, the homozygous state of identical-by-descent alleles, arises from inbreeding (F) and leads to negative effects such as reduced genetic variance and inbreeding depression. Currently, among several alternative methods, F estimated from runs of homozygosity (ROH) is considered the most accurate. However, a comprehensive bibliometric analysis of the academic literature on ROH is lacking. To address this gap, this study analyzed papers published from 2009 to 2023, aiming to provide insights into the evolving research landscape on genetic diversity and ROH, identify global trends, and examine emerging directions in scientific research on this topic. Using specific search terms within the Web of Science (WoS) core collection, 406 publications were identified and analyzed using the R package Bibliometrix (version 4.1.3). Authors and institutions from various countries contributed to the literature, with China, Italy, and the United States leading in publication output. The analysis also highlighted the growing interest in the functional analysis of ROH, signatures of selection, and genetic diversity. This interest has been facilitated by advances in genomic technology and computational tools such as PLINK and detectRUNS, emphasizing the importance of increasing analytical precision in methodologies and parameters. Future research directions include refining analytical techniques and exploring heterozygosity-rich regions (HRR).

Keywords: genomic inbreeding; genetic diversity; bibliometric review; scientific production

2.1 Introduction

The evolutionary history of a species can leave marks on its genome. Since the 1970s, molecular genetics has made rapid progress with the development of tools and technologies for manipulating DNA in vitro. This has allowed geneticists to obtain genome sequences from model organisms, contributing to our understanding of the chemical structure, functions, and evolution of genetic material (Fan *et al.*, 2018).

Since then, the genomes of various living organisms have been sequenced, with decreasing costs and improved technologies such as polymerase chain reaction (PCR), next-generation sequencing (NGS), and advanced bioinformatics methods. Such advancements have enabled the acquisition of knowledge about the genomes of different species, opening up possibilities for applications in various fields including the economy, health, and species conservation, in ways that are both cost-effective and practical for routine use (Hunter-Zinck *et al.*, 2010; Fan *et al.*, 2018; Tan *et al.*, 2019).

Moreover, the greatest challenge for modern genetics is not only sequencing genomes but also transforming the data obtained into precise and accurate information. This enhancement in data quality improves our understanding of phylogenetic relationships and molecular evolution of the genomes across different taxa. It also enables the application of this knowledge as innovative technology in studies of life's diversity, with potential applications in veterinary medicine, genetic improvement, and conservation (Wang *et al.*, 2018; Tan *et al.*, 2019).

Identical-by-descent (IBD) haplotypes refer to segments of the genome inherited from a common ancestor without recombination. Advances in whole-genome sequencing technologies and the use of high-density SNP arrays for scanning the genome for runs of homozygosity (ROH) have been proposed as effective methods for identifying these IBD haplotypes. These technologies are already available for humans and species of agricultural interest (Gibson; Morton; Collins, 2006; Hunter-Zinck *et al.*, 2010; Wang *et al.*, 2010; Qiao *et al.*, 2017).

Advances in scientific knowledge have uncovered previously unknown aspects of the genome and biological systems, resulting in a surge of research in molecular genetics (Armenta-Medina *et al.*, 2022). In this context, whole-genome approaches have significantly improved genomic scans, revealing patterns of runs of homozygosity (ROH) that highlight genetic variation and enable the identification of associations between genomic regions and adaptive phenotypes.

When parents of an individual share a relatively recent common ancestor, their offspring may exhibit ROH for that segment (McQuillan *et al.*, 2008; Kirin *et al.*, 2010). This phenomenon provides a valuable, often untapped resource for studying genetic diversity and the evolutionary history of various animal and plant groups (Brito *et al.*,

2017; Fabbri *et al.*, 2021).

Population evolution is influenced by geographic effects, historical bottlenecks, isolation, and cultural factors such as consanguineous mating, which can lead to elevated levels of homozygosity or autozygosity at both individual and population levels (McQuillan *et al.*, 2008; Pemberton *et al.*, 2012; Bosse *et al.*, 2012).

Longer regions of homozygosity increase the risk of recessive genes causing disabilities or fitness loss, which can reduce the viability of the organism and even the population (McQuillan *et al.*, 2008; Bosse *et al.*, 2012; Pemberton *et al.*, 2012). For this reason, understanding genetic variation within and between populations is crucial for assessing the risk of population decline, as inbreeding is a major cause of adaptive reduction in various populations (Pemberton *et al.*, 2012; Marras *et al.*, 2015; Mable, 2019).

Therefore, characterizing the genetic resources of various genomes is critical for formulating guidelines to support the conservation and sustainable management of biological diversity. In this context, genomic analyses, particularly those focusing on runs of homozygosity (ROH), provide essential data on molecular evolution, adaptability, and inbreeding coefficients based on homozygosity regions (FROH) (Curik; Ferenčaković; Sölkner, 2014; Marras *et al.*, 2015; Purfield *et al.*, 2017).

To deepen our understanding of genetic diversity and ROH, this study offers a visual summary of existing research and proposes future analytical directions. Utilizing the Web of Science (WoS), we identified publications that illuminate the frontiers, hotspots, and emerging trends in the field. This approach sheds light on the evolving research landscape concerning genetic diversity and ROH, thereby facilitating a comprehensive understanding of key developments and identifying potential avenues for further investigation.

2.2 Materials and methods

2.2.1 Bibliographic data recovery

The Web of Science (WoS) platform indexes over 9,000 academic institutions and more than 12,000 conferences, providing a comprehensive database for shaping future research strategies (Birkle *et al.*, 2020). On February 21, 2024, we conducted a bibliometric study of documents related to genetic diversity and runs of homozygosity (ROH), published between 2009 and 2023, using the WoS Core Collection database.

We employed the search string TS = (genetic diversity AND runs of homozygosity) to retrieve relevant publications. Only articles and reviews were included in the analysis; meeting abstracts, editorial materials, and proceedings papers were excluded. Documents were imported in BibTeX format and analyzed using the Bibliometrix R package. The

analysis was further supported by the Biblioshiny web app[1] (Aria; Cuccurullo, 2017; Armenta-Medina *et al.*, 2022).

To ensure accuracy and completeness, we manually reviewed the articles to identify the species analyzed and the software used, such as PLINK and detectRUNS. We measured publication trends, productivity, country contributions, organizations, journals, keywords, and key references (Banchi *et al.*, 2022).

For each article, we collected the institutional address of the first author, including the geographical location of the institutions. If the first author's address was unavailable, the address of the second author's institution was used (Adriaanse; Rensleigh, 2013; Moura; Dawson; Nogueira, 2017; Armenta-Medina *et al.*, 2022).

2.2.2 Thematic evolution

The primary objective of this methodology is to elucidate the conceptual structure of the field by employing a term co-occurrence network to map and cluster keywords, subheadings, or abstracts within a bibliographic compilation (Aria; Cuccurullo, 2017; Moral-Muñoz *et al.*, 2020). To analyze the evolution of the field, the raw data were segmented into consecutive periods. This thematic evolution is represented through bibliometric maps, which are constructed by linking themes from one subperiod to the next through conceptual connections. A communication pathway between themes is established when links, such as shared keywords, connect them (Cobo *et al.*, 2011; Armenta-Medina *et al.*, 2022). (Armenta-Medina *et al.*, 2022).

2.3 Results

2.3.1 Quantitative publications

According to the data presented in the statistical graph of the research area, scientific production related to ROH and its genetic diversity has varied over time. It began with modest production in 2009, followed by a slight increase in 2015, and from 2017 onward, there has been considerable growth in relevant studies in this field. It is important to note that the initial applications of genomic technologies to ROH were recent, as indicated by the number of articles published in the early years, as shown in Fig. 2.1.

The increase in the number of publications from 2017 onward was likely due to the implementation and maturation of nucleic acid sequencing, as well as the high availability of information in public and private databases. Finally, the number of publications increased annually between 2014 and 2022. As illustrated in Fig. 2.1, studies of ROH and genetic diversity have increasingly focused on molecular genetics in recent years.

[1] version 3.1.4; accessed February 21, 2024, https://www.bibliometrix.org/home/

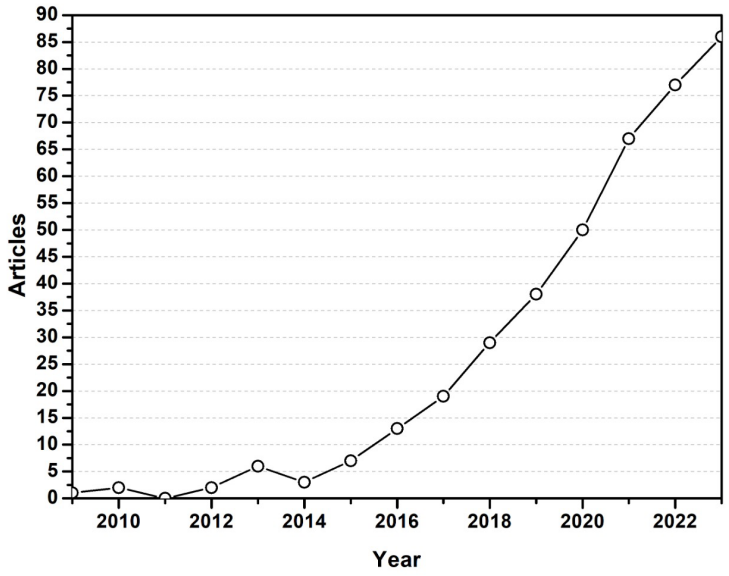

Figure 2.1 – Distribution of publications on ROH and diversity in molecular genetics by year.

The main data collected for a review of the literature on ROH and genomic diversity are presented in Table 2.1. The results show that interest in this field is growing, with an annual growth rate of 12.69%. The findings showed that ROH and genetic diversity are active and relevant fields, with each article receiving an average of 17.08 citations, demonstrating the importance of ROH to the scientific community.

Table 2.1 – Main information on the survey carried out

Description	Results
Timespan	2009:2023
Sources (journals)	87.00
Documents	406.00
Annual growth rate %	12.69
Document average age	3.62
Average citations per doc	17.08
References	13.55
Autor's keywords	834.00
Authors	2317.00
Authors of single-authored docs	1.00
Coauthors per doc	7.64
International coauthorships %	51.23

2.3.2 International Collaboration

The relevant literature on genomic diversity and ROH has been published across a diverse range of sources and journals, highlighting significant academic interest in these

topics. This growing focus is particularly evident in recent years, with research on genetic diversity and ROH attracting scientists from around the globe. A bibliometric analysis identified substantial academic output from 95 countries, totaling 406 relevant publications from 2009 to 2023.

Research institutes in Asia, the Americas, and Europe have been the primary contributors to this scientific production. Table 2.2 presents the distribution of publications by country, underscoring the global and multidisciplinary engagement with this complex subject. ROH research is globally relevant and actively pursued by researchers from a wide array of countries. From a networking perspective, the Single Country Publication (SCP) metric measures the level of within-country collaboration, while the Intra-Country Collaboration Index (i.e., multiple-country publication, MCP) monitors scientific collaboration between authors from different nations.

China's SCP metric is significantly more influential than its MCP, a phenomenon attributed to the scale and diversity of sinology. In contrast, the U.S. ranks as the third-largest academic contributor with 38 articles, but it is notable that the proportion of internationally collaborative documents is higher among U.S. researchers.

Table 2.2 – The 10 most prolific countries in genetic diversity and ROH, corresponding author's affiliation considered.

Country	N° of Articles	total frequency	*SCP	*MCP
China	79	0.302	56	23
Italy	40	0.153	28	12
United States	38	0.145	16	22
Spain	21	0.08	7	14
Germany	19	0.073	10	9
Brazil	14	0.053	8	6
Canada	14	0.053	8	6
Iran	14	0.053	3	11
Netherlands	13	0.050	3	10
France	10	0.038	4	6
Total	262	1	143	119

*SCP= single country publications; *MCP= multiple country publications.

The Table 2.3 presents the affiliations by field of study. The University of Palermo (Italy), Wageningen University and Research (Netherlands), and the Institute of Animal Science (China) emerge as leaders with 36, 33, and 32 publications, respectively. The University of Edinburgh (United Kingdom) and Copenhagen University (Sino-Denmark) also make significant contributions, with 26 and 21 articles, respectively. The range of contributors continues remarkably, with associations from many countries, including the United States, Spain, and Brazil.

Table 2.3 – Top 10 affiliations in the publications involving genetic diversity and ROH.

Affiliations	Number of articles	Articles (%)	Country
Univ Palermo	36	14.9%	Italy
Wageningen Univ and Res	33	13.7	Nettherlands
Inst Anim Sci	32	13.3%	China
University Edinburgh	26	10.8%	United Kingdom
University Copenhagen	21	8.7%	China
Cornel University	19	7.9%	United State
Uniy Calif Los Angeles	19	7.9%	United State
Univ Zagreb	19	7.9%	Croatia
China Agricultural University	18	7.5%	China
Swedish Univ Agr Sci	18	7.5%	Sweden
Total	241	100%	

2.3.3 Most relevant document

The 10 most cited articles had a total of 1,858 citations (Figure 2.2). These articles were distributed across 9 different scientific journals. The most cited article is the paper titled " Genomic Patterns of Homozygosity in Worldwide Human Populations", anthored By Pemberton *et al.* (2012) and published in The American Journal of Human Genetics (SJR factor 9,8), with a total of 314 citations.

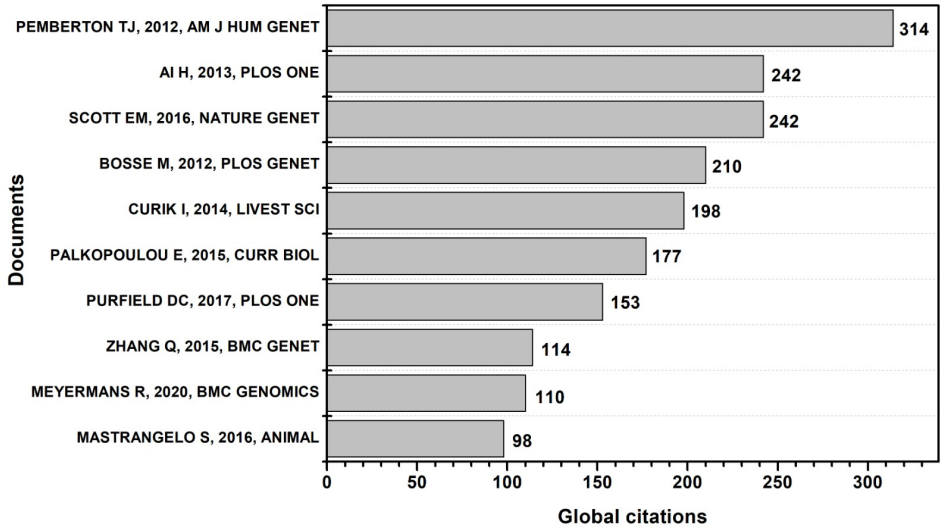

Figure 2.2 – The 10 most cited articles on genetic diversity and ROH

Although the Nature Genetics journal is not the main resource, it is noteworthy that this article received a significant number of citations. Interestingly, two authors of this article are among the ten most productive researchers in the field.

Notably, eight of the ten most cited articles were published after 2013, confirming

the trend observed in annual scientific production (Fig. 2.1). The growing interest in research into genetic polymorphisms and ROH, particularly since 2017, demonstrates the heterogeneity of population genomics studies and the detection of genetic hybridization through ROH.

Researchers are exploring new approaches, such as how evolutionary factors can affect a population's genomic structure (i.e., demography, assembly, and specific selection pressures).

2.3.4 Most frequent keywords

From a quantitative analysis, 3,121 keywords were identified as the most used in the literature under study. These terms, endowed with specificity, effectively represent the content described in the texts. When evaluating the primary meanings of these words, 'homozygosity,' 'run,' and 'genetic diversity' emerged as central pillars of genetic diversity and ROH research. These concepts have evolved remarkably over the years, revealing progressive interest and a deepening understanding (Fig. 2.3).

Over time, the information demonstrated a continuous and significant increase in keywords. Over time, the information demonstrated a continuous and significant increase in keywords. In the period analyzed, for example, the term 'Homozygosity' was registered a significant total of 154 times, while 'Runs' reached 143 occurrences.

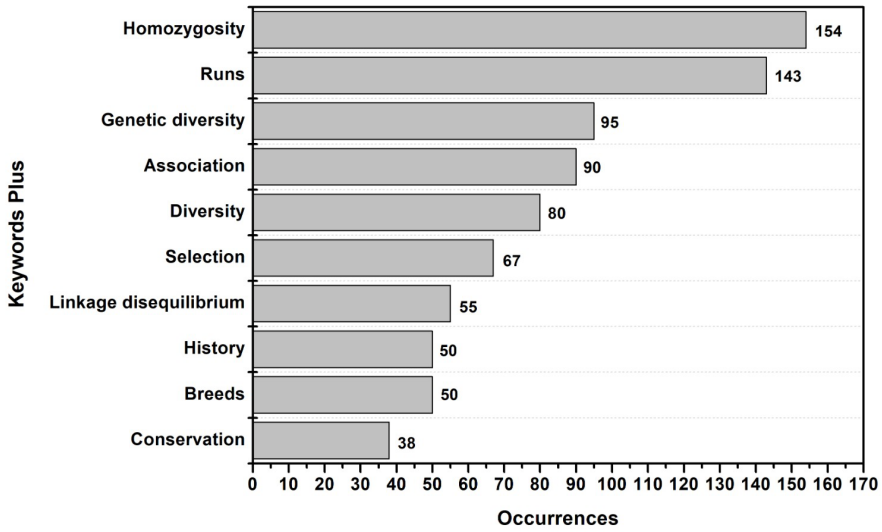

Figure 2.3 – Competition statistics for the main keywords from 2009 to 2023.

In addition, the term "genetic diversity" appeared 95 times, indicating strong interest in using ROH to elucidate biodiversity. The analysis also revealed the relevance of the terms "association" and "selection", with 90 and 67 occurrences, respectively. Notably,

the frequency of "linkage disequilibrium" also emerged significantly. Additionally, the goal of "conservation" was mentioned 38 times, and "history" appeared 50 times, signaling an increase in its occurrence over the years (Fig 2.3).

2.3.5 Thematic map

The use of thematic maps becomes particularly relevant when co-word analysis is applied to trace the structure of a specific discipline. A quantitative analysis revealed the emergence of four distinct thematic regions, each located in a specific quadrant (Fig. 2.4). In the upper right quadrant, themes such as "diversity",' "hybridization", "whole-genome sequence", "demography", and "linkage" stand out due to their high density and centrality, playing a fundamental role in structuring research fields.

Conversely, the topics in the upper left quadrant, such as "whole genome sequencing", "genetic variability", "genomic inbreeding coefficient", and "homozygosity", maintain intrinsic links and correspond to rapidly developing areas. This is reflected in their higher density despite more modest centrality, indicating they are specialized yet peripheral (Fig. 2.4).

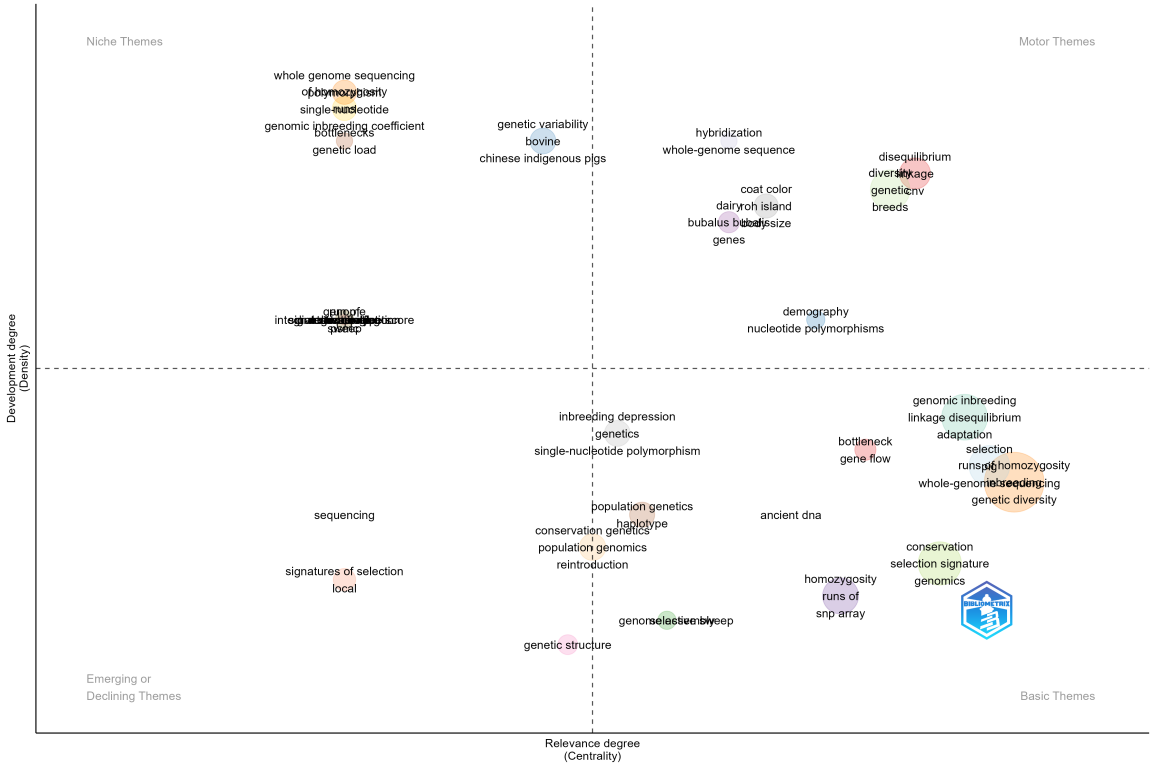

Figure 2.4 – Thematic map of genetic diversity and runs of homozygosity corresponding to the period 2013–2023.

In the lower left quadrant, marginalized and underdeveloped topics can be observed. While these topics are relatively sound, their low centrality and decreasing density indicate

either incipient growth or potential for expansion. This quadrant includes links to topics such as 'sequencing' and 'signatures of selection.

Finally, the lower right quadrant features substantial, wide-ranging, and cross-cutting topics characterized by high centrality but relatively low density. Topics such as "adaptation", "conservation", "runs of homozygosity", "genomic inbreeding", "genetics", and "linkage disequilibrium" in this quadrant suggest convergence in the study of genetic diversity and ROH across various species, from humans to different pets. However, these topics are still under development and are expected to form the basis for future research.

2.3.6 Applied Methodologies

The functional analysis of ROH is supported by computational tools that play a crucial role in detecting and characterizing these genomic regions. Among the tools used for analyzing SNP microarray data, PLINK software is extensively utilized (Peripolli *et al.*, 2017; Meyermans *et al.*, 2020). A total of 66 articles published between 2013 and 2018 explored the use of PLINK, with 56 of these studies highlighting the software's utility in analyzing ROH (Fig. 2.5).

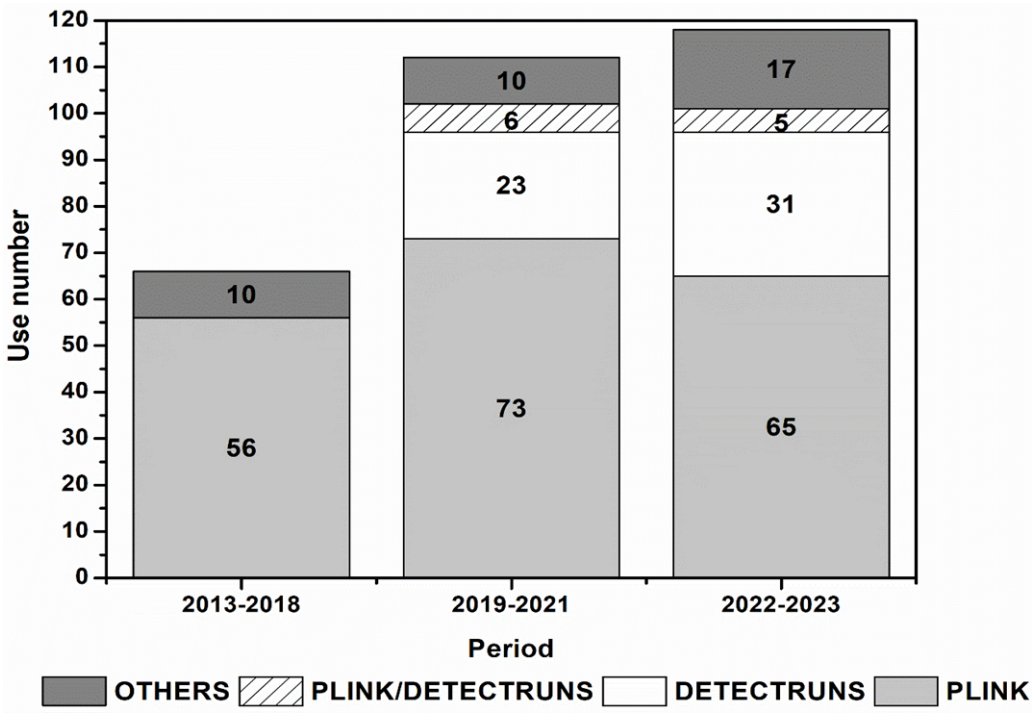

Figure 2.5 – Statistical software associated with runs of homozygosity (ROH) over a 15-year period.

However, significant changes have occurred in the landscape of software applications. Before 2013, the results were minimal. From 2019 to 2021, PLINK remained the most-used software, cited in 73 studies. Concurrently, the rise of the R package detectRUNS emerged

as a viable option, used in 23 studies during this period, with six studies using both PLINK and detectRUNS, and another ten studies employing different software programs.

Between 2022 and 2023, there was a significant increase in the variety of software used. PLINK continued to be the most frequently used, featured in 65 studies, followed by detectRUNS v. 0.9.5, which appeared in 31 articles. Additionally, five studies utilized both PLINK and detectRUNS, while 17 studies explored other software programs, such as Golden Helix SNP & Variation Suite and ZANARDI.

2.3.7 Publications Analysis Based on Animal Diversity

The analysis of approximately 275 documents reveals that ROH research necessitates a careful sampling methodology to assess the distribution of genotypes in both domestic and wild animal populations. The data indicates that the sample encompasses a diverse set of species, including humans (25), cattle (68), sheep (54), pigs (35), goats (23), chickens (25), and other domestic and wild animals (59) (Fig. 2.6).

This research demonstrates a strong interest in population genetics, focusing on genomic inbreeding in both human and animal populations, particularly emphasizing recessive diseases, animal husbandry, and the conservation of endangered species. While the U.S. concentrates on human and non-domestic animal populations, China and many other countries focus their studies on the breeding and diversity of domestic animals.

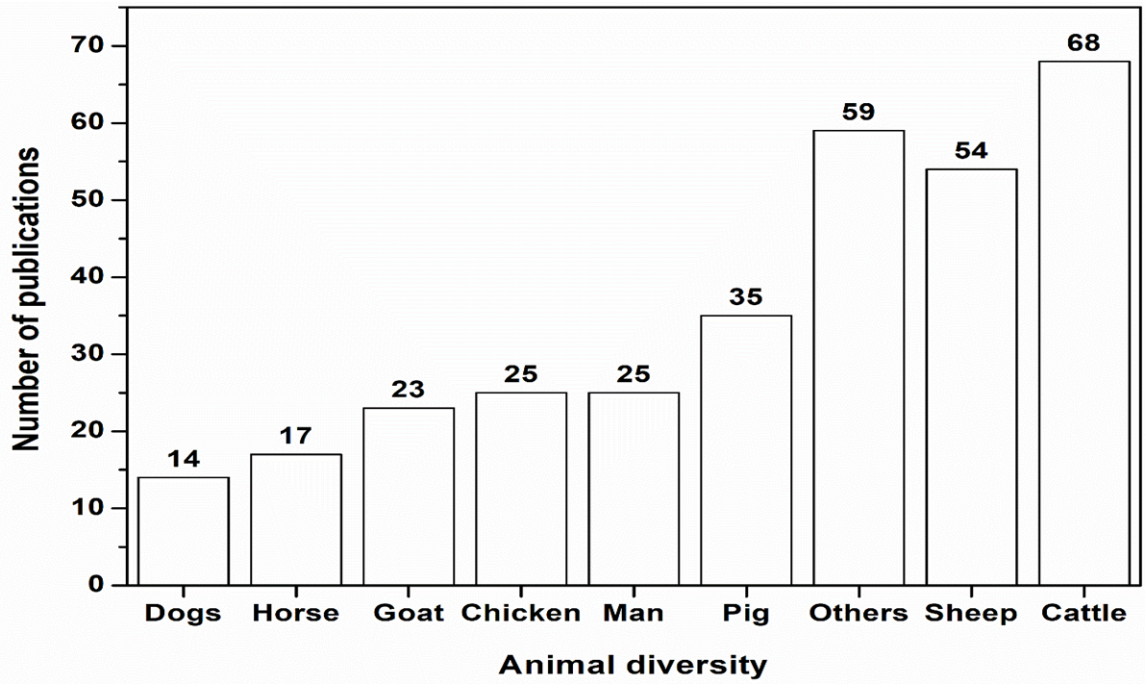

Figure 2.6 – Distribution of animal species used in the study of genomic diversity and ROH.

2.4 Discussion

2.4.1 Quantitative publications per year

Quantifying genetic diversity and levels of kinship in humans, plants, and animals is a key element for evolutionary biologists and agricultural scientists. This quantification enables the development of new methods aimed at revealing knowledge applicable to genetic improvement and the conservation of populations at risk of extinction (Curik; Ferenčaković; Sölkner, 2014; Selli *et al.*, 2021).This importance is reflected in the increase in the number of publications, which may be directly related to the growing availability of biological information, making the application of genomic technologies more accessible. As a result, these technologies can now be adopted with relatively less investment in research and have facilitated the creation of larger laboratories (Armenta-Medina *et al.*, 2022).

This development represents a significant milestone, as the significant improvement in the availability of genomic sequences enables the creation of genetic maps for animals of farming interest. These advances have been documented by several experienced authors (Curik; Ferenčaković; Sölkner, 2014; Brito *et al.*, 2017; Periasamy *et al.*, 2017; Selli *et al.*, 2021), who have exploited the potential of the data to improve animal management, genetic conservation, and selection.

Thus, broadening access to genome sequencing data not only expands the frontiers of scientific knowledge but also encourages the development of more effective management strategies, benefiting sectors such as livestock production. This field continues to play a key role in promoting sustainability and preserving crops, thus enriching evolutionary and agricultural research (Lin *et al.*, 2013; Curik; Ferenčaković; Sölkner, 2014; Periasamy *et al.*, 2017).

2.4.2 Main countries and institutions linked to publications involving ROHs

The Republic of China ranks first in the number of publications, but most of these publications are from SCP. Meanwhile, MCP represents more than half of all publications, with the U.S. ranking first with multiple contributions. This suggests that the results are associated with the availability of qualified human resources mixed with technological capacity and public and private investment in scientific research.

It should also be noted that China, Italy, and the United States are key collaborators in the development and dissemination of genomic technologies, genetic studies, and bioinformatics related to polymorphisms, ROH, and inbreeding (Mastrangelo *et al.*, 2016; Scott *et al.*, 2016; Wang *et al.*, 2018; Bertolini *et al.*, 2018a; Li *et al.*, 2022). Brazil in South America and Iran in Asia are highlighted as emerging significant contributors in this specific field of research, indicating their growing influence and scientific contributions.

This reflects the global expansion of the study area, with a diversification of scientific output beyond the traditionally dominant countries (Nosrati *et al.*, 2021; Mulim *et al.*, 2022).

China, the USA, and Italy have played significant roles in scientific research related to animal production. More developed countries with higher incomes typically possess greater resources for scientific research, which elucidates the prominence of these nations in the analyses. China's economic importance is also noteworthy, with relevant contributions to the development of technologies and investigative efforts applied to agriculture and livestock (Andrews *et al.*, 2018; Deng *et al.*, 2021; Mörschbächer; Granada, 2022).

2.4.3 Strategic areas and thematic maps

The top right quadrant of Figure 4 contains topics related to diversity, hybridization, whole-genome sequencing, demography, and genes. The relevant principles and concepts related to these topics are extensively described within the frameworks of population structure, genetic conservation, and the molecular basis of genetic diversity. Such studies reflect the broad examination of gene diversity and ROH in human, domesticated, and wild populations (Bosse *et al.*, 2012; Marras *et al.*, 2015; Brito *et al.*, 2017; Zhang *et al.*, 2023; Chessari *et al.*, 2024).

Topics related to diversity, hybridization, and whole-genome sequencing also encompass concepts such as genomics and genetics. These issues are highlighted due to their association with nucleic acid sequencing technologies, which continue to generate large volumes of data, which enable comparative studies of different biological species and populations worldwide using bioinformatics tools (Schneider; Orchard, 2011; Fan *et al.*, 2018; Clark *et al.*, 2019).

In the top left quadrant of Figure 4, there is a concentration of specific topics with associated field density. This quadrant highlights topics such as whole genome sequencing, genetic variability, genomic inbreeding coefficient, and polymorphism. These topics are linked to genomic identification, species characterization, and population decline. The high density of these topics suggests rapid development, although their centrality is relatively modest, indicating that they are specialized and still peripheral compared to other research areas (Banchi *et al.*, 2022; Marino; Petrera; Abeni, 2023).

Furthermore, the thematic maps offer a panoramic view of the controversies surrounding racial similarity and genetic diversity. These visual resources highlight both the surrounding areas and the main elements of the research, revealing the complexity of the relationships between topics. Understanding these dynamics is crucial for advancing the field and fostering significant discoveries in genetics and evolution. (Cobo *et al.*, 2011; Aria; Cuccurullo, 2017).

2.4.4 Software and statistical approaches

The conversation about software tools used to analyze ROH and heterozygosity-rich regions is comprehensive and essential in current genomic research. In this study, a quantitative assessment was made of the use of programs capable of identifying and characterizing ROH in the genome. The results make it possible to visualize different programs. PLINK software, which uses a sliding window (SW) to manage and analyze data on individual genotypes, has a high frequency of use (Purcell *et al.*, 2007; Chang *et al.*, 2015; Meyermans *et al.*, 2020).

Additionally, detectRUNS is a well-known R package designed to identify both ROH and stretches of heterozygosity (BISCARINI et al., 2018: detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes). This software utilizes consecutive-windows and sliding-windows approaches, offering features to summarize findings and calculate inbreeding coefficients using ROH data (Purcell *et al.*, 2007; Marras *et al.*, 2015).

In addition to detecting runs, the package offers several convenient functions for summarizing and plotting results, and for computing FROH. Runs of heterozygosity, also referred to as heterozygosity-rich regions, are a relatively novel concept introduced by Williams *et al.* (2016). One of the objectives of this project is to develop methods and code, and to explore parameters for the detection of 'runs of heterozygosity' in diploid genomes, particularly in animals.

Besides the sliding window and consecutive run methods employed by detectRUNS, there are other software tools and methodologies suitable for detecting ROH. These alternative tools can provide complementary insights and enhance the robustness of genetic analyses (Meyermans *et al.*, 2020).

2.4.5 Consecutive run Method

The concept implemented by Marras *et al.* (2015), initiates a search for a ROH by checking the homozygosity of the first SNP; if this SNP is homozygous, it is considered to be potentially in a ROH. The next SNP is then evaluated: if it is heterozygous, the process stops; if homozygous, a potential ROH may be detected. This continues until the minimum number of SNP in a ROH is met: a ROH is thus detected and will span until a heterozygous SNP is encountered.

This method does not use overlapping windows. Parameters based on heterozygous and/or missing SNP loci can be tuned to relax the identification of ROH. This method can be adapted to detect 'runs of heterozygosity' (ROHet) by simply inverting the detection from homozygous loci to heterozygous loci (and adjusting parameters). The Python code from the Zanardi software package (https://github.com/bioinformatics-ptp/Zanardi) was

expanded and translated into R for implementing the consecutive method in detectRUNS.

2.4.6 An alternative method: the sliding windows.

The software Plink(Purcell *et al.*, 2007) employs a sliding-window algorithm to identify ROH in diploid genomes, such as the human genome. This method includes the following steps (Bjelland *et al.*, 2013):

1. Starting from the first SNP, a sliding window of n SNPs scans the genome, moving from one chromosome to the next.

2. When a stretch of m homozygous SNPs is found (where m is less than or equal to n), the number of sliding windows covering this stretch is counted for each SNP locus.

3. This count is divided by the total number of windows covering each locus. Initially, this count increases from 1 to n, then remains constant at n, and finally decreases from n to 1.

4. SNP loci for which this ratio is 5% or higher are classified as being in a ROH.

5. Flexibility parameters can be introduced during step 2, such as allowing a certain number of heterozygous SNP loci or missing SNP loci within a ROH..

This procedure was translated into R code and adapted for the detection of 'runs of heterozygosity' (ROHet). The key modification involves counting stretches of heterozygous instead of homozygous SNPs, taking into account the following parameters: the minimum number of SNPs included in the heterozygosity-rich regions (HRR); the number of missing or opposite genotypes allowed; the maximum gap between consecutive SNPs; and the minimum length of an HRR. The R code to implement this algorithm was developed in the detectRUNS package.

R was primarily used to write functions in detectRUNS. The R package format was chosen to facilitate documentation, organization, distribution, and reproducibility of the code (Williams *et al.*, 2016; Chessari *et al.*, 2024). TThe detectRUNS package has experienced a significant surge in utilization and is likely one of the most widely adopted programs for analyzing SNP genotypes for ROH and HRR across various animal species (Gaspar *et al.*, 2023; Visser *et al.*, 2023; Vostry *et al.*, 2023; Zhang *et al.*, 2023).

Numerous studies have investigated ROH and HRR in domestic and wild animals (Mulim *et al.*, 2022; Hendricks *et al.*, 2022), encompassing a wide variety of species. These studies have been fundamental in advancing the understanding of genetic structure and the evolutionary history of various animal populations. Analyzing these regions provides

valuable insights into the processes of inbreeding, natural selection, and adaptation to environmental pressures (Marras *et al.*, 2018; Selli *et al.*, 2021; Chessari *et al.*, 2024).

However, research on ROH and HRR faces significant methodological challenges. The definition of parameters for identifying these regions remains an open question, with no precise guidelines established. The lack of consensus in the literature regarding the ideal criteria for detecting ROH and HRR can lead to variations in results and interpretations, complicating comparisons between studies (Meyermans *et al.*, 2020; Selli *et al.*, 2021; Falchi *et al.*, 2024).

The discrepancies observed in the results of various studies can be attributed to variations in detection parameters, such as the minimum number of SNPs, the minimum length of the regions, the number of consecutive heterozygous markers, and the number of missing SNPs. Different SNP densities of the array low, medium, and high also influence the results. Other methodological factors, such as quality control through minor allele frequency (MAF) pruning and linkage disequilibrium (LD), can also significantly impact the results (Meyermans *et al.*, 2020; Selli *et al.*, 2021; Falchi *et al.*, 2024).

To conduct analyses of ROH and HRR, it is necessary to use genotype files in PED and MAP formats. It is crucial that these files have unique and distinct acronyms to precisely identify the different study groups. Correct labeling of the files facilitates the organization and analysis of genetic data, ensuring that the results are accurately attributed to their respective populations.

Since parameter configuration for these analyses is still an uncertain issue without precise guidelines, it is recommended to start with a broad screening. According to Purfield *et al.* (2012), to reduce the occurrence of ROH by chance, the minimum number of SNPs needed to constitute an ROH was estimated using the following formula:

$$l = \frac{\ln(\frac{\alpha}{n_s \times n_i})}{\ln(1 - het^{-})} \tag{2.1}$$

In formula 2.1, n_s represents the number of SNPs within an individual, n_i is the number of individuals within a population i, α is the percentage of false positive ROH (set at 0.05), and (het^{-}) is the mean heterozygosity across all SNPs for population i. This initial approach allows for better calibration of analysis parameters, ensuring greater accuracy in the results obtained.

According to Falchi *et al.* (2024), to account for the lower number of heterozygote genotypes in the genome, the number of opposite SNPs allowed in a ROHet was calculated as follows:

$$maxOpp = \frac{nSNP_{ROHom}}{nSNP_{ROHet}} \tag{2.2}$$

Where $nSNP_{ROHom}$ is the minimum number of SNPs in a ROHom and $nSNP_{ROHet}$ is the

minimum number of SNPs in a ROHet, both computed using the previous equation(Purfield *et al.*, 2012; Purfield *et al.*, 2017; Meyermans *et al.*, 2020; Selli *et al.*, 2021; Falchi *et al.*, 2024).

The variability in methodologies underscores the necessity for standardizing ROH and HRR studies. Without consistent guidelines, direct comparisons between studies can be hindered, potentially compromising the reliability of conclusions. Therefore, it is crucial for future research to adopt uniform methodologies, leveraging tools like the R package detectRUNS, and validate detection parameters to ensure the reproducibility and comparability of findings.

2.5 Conclusions

Bibliometric analysis has demonstrated considerable progress in studies related to ROH and genetic diversity over the years. Initially modest in 2009, there was a significant increase from 2017 onwards, driven by advancements in genomic technologies and greater availability of information in public and private databases. Many relevant articles were identified across 87 sources or journals, demonstrating the growing interest in this topic. Furthermore, the bibliometric analysis highlighted considerable international cooperation in this field, with researchers from 95 countries contributing to academic research. Study centers in Asia, America, and Europe stood out as the main drivers of scientific production in this area.

The research also highlighted advancements in the methods and technology employed in the study of ROH. The PLINK software was commonly utilized for examining SNP microarray data, demonstrating its efficacy in detecting and describing homozygous genomic regions. Additionally, the emerging detectRUNS tool in R is a potential alternative for more thorough and precise analytical approaches. The analysis of keywords unveiled central new topics such as 'homozygosity,' 'migration,' 'genetic variability,' 'correlation,' and 'ancestry,' showing an increasing depth and curiosity in exploring the patterns of ROH and how they influence population structure across generations. These themes demonstrate a variety of perspectives, ranging from research on human communities to both domestic and wild fauna.

The steady advancement of research into ROH and genetic diversity indicates a continued and important interest in this field. However, challenges remain, such as the need to develop more advanced analytical techniques to handle complex and diverse datasets. Future investigations should delve into new areas, such as heterozygosity-rich regions, which are capable of detecting differences between chromosomal regions, providing a more detailed picture of genetic diversity.

Bibliography

ADRIAANSE, L. S.; RENSLEIGH, C. Web of science, scopus and google scholar: A content comprehensiveness comparison. *The Electronic Library*, Emerald Group Publishing Limited, v. 31, n. 6, p. 727–744, 2013. Cited on page 7.

AL-MAMUN, H. A. *et al.* Genome-wide linkage disequilibrium and genetic diversity in five populations of australian domestic sheep. *Genetics Selection Evolution*, Springer, v. 47, p. 1–14, 2015. Cited on page 36.

ANDREWS, K. R. *et al.* A bioinformatic pipeline for identifying informative snp panels for parentage assignment from rad seq data. *Molecular ecology resources*, Wiley Online Library, v. 18, n. 6, p. 1263–1281, 2018. Cited on page 16.

ARAUJO, A. M. de *et al.* Crescimento e mortalidade em um rebanho de conservação de caprinos marota no brasil. Revista Cientfica de Producão Animal, v. 11, n. 2, p. 103-109, 2009., 2009. Cited 2 on page 1 and 33.

ARIA, M.; CUCCURULLO, C. bibliometrix: An r-tool for comprehensive science mapping analysis. *Journal of informetrics*, Elsevier, v. 11, n. 4, p. 959–975, 2017. Cited 2 on page 7 and 16.

ARIAS, K. D. *et al.* Approaching autozygosity in a small pedigree of gochu asturcelta pigs. *Genetics Selection Evolution*, Springer, v. 55, n. 1, p. 74, 2023. Cited 2 on page 36 and 37.

ARMENTA-MEDINA, D. *et al.* A bibliometric analysis of mexican bioinformatics: a portrait of actors, structure, and dynamics. *Biology*, MDPI, v. 11, n. 1, p. 131, 2022. Cited 3 on page 5, 7, and 15.

BANCHI, P. *et al.* Trends in small animal reproduction: a bibliometric analysis of the literature. *Animals*, MDPI, v. 12, n. 3, p. 336, 2022. Cited 2 on page 7 and 16.

BANFI, C. *et al.* Prenylcysteine oxidase 1 is a key regulator of adipogenesis. *Antioxidants*, MDPI, v. 12, n. 3, p. 542, 2023. Cited on page 45.

BATZIOS, A. *et al.* Technical indicators and economic performance of dairy goat farms in greece: An efficiency analysis. *Livestock Science*, Elsevier, v. 271, p. 105210, 2023. Cited on page 33.

BERTOLINI, F. *et al.* Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. *Genetics Selection Evolution*, Springer, v. 50, p. 1–12, 2018. Cited 2 on page 15 and 43.

BERTOLINI, F. *et al.* Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genetics Selection Evolution*, Springer, v. 50, p. 1–24, 2018. Cited on page 44.

BIRKLE, C. *et al.* Web of science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , v. 1, n. 1, p. 363–376, 2020. Cited on page 6.

BISCARINI, F. *et al.* Detect runs of homozygosity and runs of heterozygosity in diploid genomes. CRAN (The Comprehensive R Archive Network), 2019. Cited on page 35.

BISCARINI, F. *et al.* Insights into genetic diversity, runs of homozygosity and heterozygosity-rich regions in maremmana semi-feral cattle using pedigree and genomic data. *Animals*, MDPI, v. 10, n. 12, p. 2285, 2020. Cited 2 on page 37 and 44.

BJELLAND, D. *et al.* Evaluation of inbreeding depression in holstein cattle using whole-genome snp markers and alternative measures of genomic inbreeding. *Journal of dairy science*, Elsevier, v. 96, n. 7, p. 4697–4706, 2013. Cited on page 18.

BOETTCHER, P. J. *et al.* Genetic resources and genomics for adaptation of livestock to climate change. *Frontiers in genetics*, Frontiers Media SA, v. 5, p. 461, 2015. Cited 2 on page 1 and 33.

BOSSE, M. *et al.* Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS genetics*, Public Library of Science San Francisco, USA, v. 8, n. 11, p. e1003100, 2012. Cited 3 on page 6, 16, and 34.

BRITO, L. F. *et al.* Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide snp markers. *BMC genomics*, Springer, v. 18, p. 1–20, 2017. Cited 5 on page 6, 15, 16, 43, and 44.

CARRARA, E. *et al.* Assessment of runs of homozygosity, heterozygosity-rich regions and genomic inbreeding estimates in a subpopulation of guzerá (bos indicus) dual-purpose cattle. *Journal of Animal Breeding and Genetics*, Wiley Online Library, v. 141, n. 2, p. 207–219, 2024. Cited on page 37.

CEBALLOS, F. C. *et al.* Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics*, Nature Publishing Group UK London, v. 19, n. 4, p. 220–234, 2018. Cited on page 1.

CEDERLUND, M. *et al.* A1m/α1-microglobulin is proteolytically activated by myeloperoxidase, binds its heme group and inhibits low density lipoprotein oxidation. *Frontiers in physiology*, Frontiers Media SA, v. 6, p. 11, 2015. Cited on page 46.

CENDRON, F. *et al.* Genome-wide snp analysis reveals the population structure and the conservation status of 23 italian chicken breeds. *Animals*, MDPI, v. 10, n. 8, p. 1441, 2020. Cited 2 on page 2 and 34.

CHANG, C. C. *et al.* Second-generation plink: rising to the challenge of larger and richer datasets. *Gigascience*, Oxford University Press, v. 4, n. 1, p. s13742–015, 2015. Cited 3 on page 17, 35, and 36.

CHESSARI, G. *et al.* Characterization of heterozygosity-rich regions in italian and worldwide goat breeds. *Scientific Reports*, Nature Publishing Group UK London, v. 14, n. 1, p. 3, 2024. Cited 6 on page 16, 18, 19, 34, 44, and 45.

CHOKOE, T. C. *et al.* Conservation status and historical relatedness of south african communal indigenous goat populations using a genome-wide single-nucleotide polymorphism marker. *Frontiers in Genetics*, Frontiers Media SA, v. 13, p. 909472, 2022. Cited 2 on page 1 and 44.

CLARK, D. W. *et al.* Associations of autozygosity with a broad range of human phenotypes. *Nature communications*, Nature Publishing Group UK London, v. 10, n. 1, p. 4957, 2019. Cited on page 16.

COBO, M. J. *et al.* An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of informetrics*, Elsevier, v. 5, n. 1, p. 146–166, 2011. Cited 2 on page 7 and 16.

COLLI, L. *et al.* Assessing the spatial dependence of adaptive loci in 43 european and western asian goat breeds using aflp markers. *PLoS One*, Public Library of Science San Francisco, USA, v. 9, n. 1, p. e86668, 2014. Cited on page 1.

CORTELLARI, M. *et al.* Runs of homozygosity in the italian goat breeds: Impact of management practices in low-input systems. *Genetics Selection Evolution*, Springer, v. 53, p. 1–11, 2021. Cited on page 44.

CURIK, I.; FERENČAKOVIĆ, M.; SÖLKNER, J. Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock science*, Elsevier, v. 166, p. 26–34, 2014. Cited 3 on page 6, 15, and 42.

DENG, H. *et al.* Impact of public research and development and extension on agricultural productivity in china from 1990 to 2013. *China Economic Review*, Elsevier, v. 70, p. 101699, 2021. Cited on page 16.

DENISKOVA, T. E. *et al.* Snp-based genotyping provides insight into the west asian origin of russian local goats. *Frontiers in genetics*, Frontiers Media SA, v. 12, p. 708740, 2021. Cited 4 on page 36, 42, 43, and 44.

DINIZ, F. d. A. *et al.* Capra hircus outliers markers in brazil: Searching for genomic regions under the action of natural selection. *Genetics and molecular biology*, SciELO Brasil, v. 46, n. 3, p. e20230084, 2023. Cited 3 on page 1, 33, and 44.

DUNCE, J. M.; SALMON, L. J.; DAVIES, O. R. Structural basis of meiotic chromosome synaptic elongation through hierarchical fibrous assembly of syce2-tex12. *Nature structural & molecular biology*, Nature Publishing Group US New York, v. 28, n. 8, p. 681–693, 2021. Cited on page 45.

DZOMBA, E. *et al.* Runs of homozygosity analysis of south african sheep breeds from various production systems investigated using ovinesnp50k data. *BMC genomics*, Springer, v. 22, p. 1–14, 2021. Cited 2 on page 36 and 43.

FABBRI, M. C. *et al.* Genetic diversity and population history of eight italian beef cattle breeds using measures of autozygosity. *PLoS One*, Public Library of Science San Francisco, CA USA, v. 16, n. 10, p. e0248087, 2021. Cited 2 on page 6 and 36.

FALCHI, L. *et al.* Effect of genotyping density on the detection of runs of homozygosity and heterozygosity in cattle. *Journal of Animal Science*, Oxford University Press US, p. skae147, 2024. Cited 2 on page 19 and 20.

FAN, H. *et al.* Conservation genetics and genomics of threatened vertebrates in china. *Journal of genetics and genomics*, Elsevier, v. 45, n. 11, p. 593–601, 2018. Cited 2 on page 5 and 16.

GASPAR, D. *et al.* Genome-wide assessment of the population structure and genetic diversity of four portuguese native sheep breeds. *Frontiers in Genetics*, Frontiers Media SA, v. 14, p. 1109490, 2023. Cited on page 18.

GIBSON, J.; MORTON, N. E.; COLLINS, A. Extended tracts of homozygosity in outbred human populations. *Human molecular genetics*, Oxford University Press, v. 15, n. 5, p. 789–795, 2006. Cited on page 5.

GJERSTORFF, M. F. *et al.* Gage cancer-germline antigens are recruited to the nuclear envelope by germ cell-less (gcl). Public Library of Science San Francisco, USA, 2012. Cited on page 45.

HENDRICKS, S. A. *et al.* Genomic assessment of cancer susceptibility in the threatened catalina island fox (urocyon littoralis catalinae). *Genes*, MDPI, v. 13, n. 8, p. 1496, 2022. Cited on page 18.

HOSONO, K. *et al.* Ypel5 protein of the ypel gene family is involved in the cell cycle progression by interacting with two distinct proteins ranbpm and ranbp10. *Genomics*, Elsevier, v. 96, n. 2, p. 102–111, 2010. Cited on page 45.

HUNTER-ZINCK, H. *et al.* Population genetic structure of the people of qatar. *The American Journal of Human Genetics*, Elsevier, v. 87, n. 1, p. 17–25, 2010. Cited on page 5.

IGOSHIN, A. V. *et al.* Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in siberian cattle populations. *BMC genetics*, Springer, v. 20, p. 5–14, 2019. Cited on page 46.

ISLAM, R. *et al.* Genome-wide runs of homozygosity, effective population size, and detection of positive selection signatures in six chinese goat breeds. *Genes*, MDPI, v. 10, n. 11, p. 938, 2019. Cited 3 on page 36, 43, and 44.

KEIGHTLEY, M.-C. *et al.* The pu. 1 target gene zbtb11 regulates neutrophil development through its integrase-like hhcc zinc finger. *Nature communications*, Nature Publishing Group UK London, v. 8, n. 1, p. 14911, 2017. Cited on page 45.

KIRIN, M. *et al.* Genomic runs of homozygosity record population history and consanguinity. *PloS one*, Public Library of Science San Francisco, USA, v. 5, n. 11, p. e13996, 2010. Cited on page 5.

LI, G. *et al.* Genome-wide estimates of runs of homozygosity, heterozygosity, and genetic load in two chinese indigenous goat breeds. *Frontiers in Genetics*, Frontiers Media SA, v. 13, p. 774196, 2022. Cited 2 on page 15 and 43.

LIN, B. Z. *et al.* Genetic diversity and structure in a sian native goat analyzed by newly developed snp markers. *Animal science journal*, Wiley Online Library, v. 84, n. 8, p. 579–584, 2013. Cited 3 on page 1, 15, and 33.

LIN, W.-R. *et al.* Galnt14 genotype predicts postoperative outcome of stage iii colorectal cancer with oxaliplatin as adjuvant chemotherapy. *Medicine*, LWW, v. 95, n. 17, p. e3487, 2016. Cited on page 45.

LIU, X. X.; CAI, L.; LIU, F. J. An in silico analysis of human sperm genes associated with asthenozoospermia and its implication in male infertility. *Medicine*, LWW, v. 97, n. 49, p. e13338, 2018. Cited on page 45.

MABLE, B. K. Conservation of adaptive potential and functional diversity: integrating old and new approaches. *Conservation Genetics*, Springer, v. 20, n. 1, p. 89–100, 2019. Cited on page 6.

MANUNZA, A. *et al.* Discovering novel clues of natural selection on four worldwide goat breeds. *Scientific Reports*, Nature Publishing Group UK London, v. 13, n. 1, p. 2110, 2023. Cited on page 44.

MARCHESI, J. *et al.* Relationship of runs of homozygosity with adaptive and production traits in a paternal broiler line. *Animal*, Cambridge University Press, v. 12, n. 6, p. 1126–1134, 2018. Cited 2 on page 2 and 34.

MARINO, R.; PETRERA, F.; ABENI, F. Scientific productions on precision livestock farming: An overview of the evolution and current state of research based on a bibliometric analysis. *Animals*, MDPI, v. 13, n. 14, p. 2280, 2023. Cited on page 16.

MARRAS, G. *et al.* Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in italy. *Animal genetics*, Wiley Online Library, v. 46, n. 2, p. 110–121, 2015. Cited 6 on page 1, 6, 16, 17, 33, and 34.

MARRAS, G. *et al.* Characterization of runs of homozygosity and heterozygosity-rich regions in a commercial turkey (meleagris gallopavo) population. In: *Proceedings of the 11th World Congress of Genetics Applied to Livestock Production, Auckland, New Zealand*. [S.l.: s.n.], 2018. p. 11–16. Cited on page 19.

MARTÍNEZ-ROCHA, R. *et al.* Assessment of genetic diversity, runs of homozygosity, and signatures of selection in tropical milking criollo cattle using pedigree and genomic data. *Genes*, MDPI, v. 13, n. 10, p. 1896, 2022. Cited on page 37.

MASTRANGELO, S. *et al.* Genome-wide patterns of homozygosity reveal the conservation status in five italian goat populations. *Animals*, MDPI, v. 11, n. 6, p. 1510, 2021. Cited 3 on page 42, 43, and 44.

MASTRANGELO, S. *et al.* Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. *Animal*, Cambridge University Press, v. 10, n. 5, p. 746–754, 2016. Cited on page 15.

MCQUILLAN, R. *et al.* Runs of homozygosity in european populations. *The American Journal of Human Genetics*, Elsevier, v. 83, n. 3, p. 359–372, 2008. Cited 4 on page 5, 6, 34, and 36.

MEYERMANS, R. *et al.* How to study runs of homozygosity using plink? a guide for analyzing medium density snp data in livestock and pet species. *BMC genomics*, Springer, v. 21, p. 1–14, 2020. Cited 6 on page 13, 17, 19, 20, 35, and 36.

MICHAILIDOU, S. *et al.* Analysis of genome-wide dna arrays reveals the genomic population structure and diversity in autochthonous greek goat breeds. *PloS one*, Public Library of Science San Francisco, CA USA, v. 14, n. 12, p. e0226179, 2019. Cited on page 43.

MILLER, D. E. *et al.* Genetic, inflammatory, and epithelial cell differentiation factors control expression of human calpain-14. *G3: Genes, Genomes, Genetics*, Oxford University Press, v. 9, n. 3, p. 729–736, 2019. Cited on page 45.

MORAL-MUÑOZ, J. A. *et al.* Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional de la Información*, v. 29, n. 1, 2020. Cited on page 7.

MÖRSCHBÄCHER, A. P.; GRANADA, C. E. Mapping the worldwide knowledge of antimicrobial substances produced by lactobacillus spp.: A bibliometric analysis. *Biochemical Engineering Journal*, Elsevier, v. 180, p. 108343, 2022. Cited on page 16.

MOURA, J. d. O. *et al.* Population genomics and gene introgression in goat herds naturally adapted to brazil. *Revista Ciência Agronômica*, SciELO Brasil, v. 50, n. 3, p. 476–483, 2019. Cited 5 on page 1, 33, 35, 43, and 44.

MOURA, J. O. *et al.* Diversidade genética em caprino marota e manejo genético para rebanhos de conservação. *Archivos de zootecnia*, v. 68, n. 261, p. 138–145, 2019. Cited 3 on page 1, 33, and 35.

MOURA, R. F.; DAWSON, D. A.; NOGUEIRA, D. M. The use of microsatellite markers in neotropical studies of wild birds: a literature review. *Anais da Academia Brasileira de Ciências*, SciELO Brasil, v. 89, p. 145–154, 2017. Cited on page 7.

MULIM, H. A. *et al.* Characterization of runs of homozygosity, heterozygosity-enriched regions, and population structure in cattle populations selected for different breeding goals. *BMC genomics*, Springer, v. 23, n. 1, p. 209, 2022. Cited 9 on page 1, 2, 16, 18, 33, 34, 37, 44, and 45.

NAKAJIMA, M. *et al.* Claudin-1 binder enhances epidermal permeability in a human keratinocyte model. *Journal of Pharmacology and Experimental Therapeutics*, ASPET, v. 354, n. 3, p. 440–447, 2015. Cited on page 45.

NANDOLO, W. *et al.* Timing and extent of inbreeding in african goats. *Frontiers in Genetics*, Frontiers Media SA, v. 10, p. 537, 2019. Cited on page 43.

NOSRATI, M. *et al.* The pattern of runs of homozygosity and genomic inbreeding in world-wide sheep populations. *Genomics*, Elsevier, v. 113, n. 3, p. 1407–1415, 2021. Cited on page 16.

ONZIMA, R. B. *et al.* Genome-wide characterization of selection signatures and runs of homozygosity in ugandan goat breeds. *Frontiers in Genetics*, Frontiers Media SA, v. 9, p. 318, 2018. Cited on page 44.

PEMBERTON, T. J. *et al.* Genomic patterns of homozygosity in worldwide human populations. *The American Journal of Human Genetics*, Elsevier, v. 91, n. 2, p. 275–292, 2012. Cited 3 on page 6, 10, and 34.

PERIASAMY, K. *et al.* Mapping molecular diversity of indigenous goat genetic resources of asia. *Small Ruminant Research*, Elsevier, v. 148, p. 2–10, 2017. Cited 3 on page 1, 15, and 33.

PERIPOLLI, E. *et al.* Runs of homozygosity: current knowledge and applications in livestock. *Animal genetics*, Wiley Online Library, v. 48, n. 3, p. 255–271, 2017. Cited on page 13.

PERIPOLLI, E. *et al.* Assessment of runs of homozygosity islands and estimates of genomic inbreeding in gyr (bos indicus) dairy cattle. *BMC genomics*, Springer, v. 19, p. 1–13, 2018. Cited on page 37.

PERSICHILLI, C. *et al.* Genome-wide diversity of pagliarola sheep residual population and its conservation implication. *Italian Journal of Animal Science*, Taylor & Francis, v. 20, n. 1, p. 1695–1705, 2021. Cited on page 1.

PURCELL, S. *et al.* Plink: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, Elsevier, v. 81, n. 3, p. 559–575, 2007. Cited 2 on page 17 and 18.

PURFIELD, D. C. *et al.* Runs of homozygosity and population history in cattle. *BMC genetics*, Springer, v. 13, p. 1–11, 2012. Cited 2 on page 19 and 20.

PURFIELD, D. C. *et al.* The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PloS one*, Public Library of Science San Francisco, CA USA, v. 12, n. 5, p. e0176780, 2017. Cited 4 on page 6, 20, 33, and 43.

QIAO, H. *et al.* Using the kde method to model ecological niches: A response to blonder et al.(2017). *Global Ecology and Biogeography*, John Wiley & Sons Ltd, v. 26, n. 9, 2017. Cited on page 5.

RAMOS-ONSINS, S. E. *et al.* Mining the pig genome to investigate the domestication process. *Heredity*, Nature Publishing Group, v. 113, n. 6, p. 471–484, 2014. Cited on page 36.

SANTOS, W. Bizarria dos *et al.* Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the mangalarga marchador horse breed. *Journal of Animal Breeding and Genetics*, Wiley Online Library, v. 138, n. 2, p. 161–173, 2021. Cited on page 45.

SARAVANAN, K. *et al.* Genome-wide analysis of genetic diversity and selection signatures in three indian sheep breeds. *Livestock Science*, Elsevier, v. 243, p. 104367, 2021. Cited on page 36.

SARAVANAN, K. *et al.* Genomic scans for selection signatures revealed candidate genes for adaptation and production traits in a variety of cattle breeds. *Genomics*, Elsevier, v. 113, n. 3, p. 955–963, 2021. Cited on page 37.

SARAVANAN, K. *et al.* Genome-wide assessment of genetic diversity, linkage disequilibrium and haplotype block structure in tharparkar cattle breed of india. *Animal biotechnology*, Taylor & Francis, v. 33, n. 2, p. 297–311, 2022. Cited on page 36.

SCHNEIDER, M. V.; ORCHARD, S. Omics technologies, data and bioinformatics principles. *Bioinformatics for Omics Data: Methods and Protocols*, Springer, p. 3–30, 2011. Cited on page 16.

SCOTT, E. M. *et al.* Characterization of greater middle eastern genetic variation for enhanced disease gene discovery. *Nature genetics*, Nature Publishing Group US New York, v. 48, n. 9, p. 1071–1076, 2016. Cited on page 15.

SELLI, A. *et al.* Detection and visualization of heterozygosity-rich regions and runs of homozygosity in worldwide sheep populations. *Animals*, MDPI, v. 11, n. 9, p. 2696, 2021. Cited 6 on page 15, 19, 20, 34, 35, and 36.

SILVA, A. A. *et al.* Gwas and gene networks for milk-related traits from test-day multiple lactations in portuguese holstein cattle. *Journal of Applied Genetics*, Springer, v. 61, n. 3, p. 465–476, 2020. Cited on page 45.

SUDHAKAR, D. V. *et al.* Exome sequencing and functional analyses revealed cetn1 variants leads to impaired cell division and male fertility. *Human Molecular Genetics*, Oxford University Press, v. 32, n. 4, p. 533–542, 2023. Cited on page 45.

TAN, M. P. *et al.* Applications of next-generation sequencing technologies and computational tools in molecular evolution and aquatic animals conservation studies: A short review. *Evolutionary Bioinformatics*, SAGE Publications Sage UK: London, England, v. 15, p. 1176934319892284, 2019. Cited on page 5.

TOLONE, M. *et al.* Genomic tools for the characterization of local animal genetic resources: application in mascaruna goat. *Animals*, MDPI, v. 12, n. 20, p. 2840, 2022. Cited on page 1.

TOMINAGA, K. *et al.* Il-12 synergizes with il-18 or il-1β for ifn-γ production from human t cells. *International immunology*, Oxford University Press, v. 12, n. 2, p. 151–160, 2000. Cited on page 45.

TONG, S.-f. *et al.* Genome-wide detection for runs of homozygosity analysis in three pig breeds from chinese taihu basin and landrace pigs by slaf-seq data. *Journal of Integrative Agriculture*, Elsevier, v. 21, n. 11, p. 3293–3301, 2022. Cited on page 36.

VALLEE, R. B.; VARMA, D.; DUJARDIN, D. L. Zw10 function in mitotic checkpoint control, dynein targeting, and membrane trafficking: is dynein the unifying theme? *Cell Cycle*, Taylor & Francis, v. 5, n. 21, p. 2447–2451, 2006. Cited on page 45.

VISSER, C. *et al.* Pedigree and genome-based patterns of homozygosity in the south african ayrshire, holstein, and jersey breeds. *Frontiers in Genetics*, Frontiers, v. 14, p. 1136078, 2023. Cited 2 on page 18 and 43.

VOSTRY, L. *et al.* Genomic diversity and population structure of the czech holstein cattle. *Livestock science*, Elsevier, v. 273, p. 105261, 2023. Cited on page 18.

WANG, L.-S. *et al.* Population-based study of genetic variation in individuals with autism spectrum disorders from croatia. *BMC Medical Genetics*, Springer, v. 11, p. 1–19, 2010. Cited on page 5.

WANG, Q. *et al.* Whole-genome resequencing of chinese pangolins reveals a population structure and provides insights into their conservation. *Communications Biology*, Nature Publishing Group UK London, v. 5, n. 1, p. 821, 2022. Cited on page 37.

WANG, X. *et al.* Genetic diversity, population structure and phylogenetic relationships of three indigenous pig breeds from jiangxi province, china, in a worldwide panel of pigs. *Animal genetics*, Wiley Online Library, v. 49, n. 4, p. 275–283, 2018. Cited 2 on page 5 and 15.

WIJKSTROM-FREI, C. *et al.* Lactoperoxidase and human airway host defense. *American journal of respiratory cell and molecular biology*, American Thoracic Society, v. 29, n. 2, p. 206–212, 2003. Cited on page 46.

WILLIAMS, J. L. *et al.* Inbreeding and purging at the genomic level: the chillingham cattle reveal extensive, non-random snp heterozygosity. *Animal genetics*, Wiley Online Library, v. 47, n. 1, p. 19–27, 2016. Cited 3 on page 17, 18, and 45.

XU, Z. *et al.* Assessment of autozygosity derived from runs of homozygosity in jinhua pigs disclosed by sequencing data. *Frontiers in Genetics*, Frontiers Media SA, v. 10, p. 274, 2019. Cited 2 on page 2 and 34.

ZAVAREZ, L. B. *et al.* Assessment of autozygosity in nellore cows (bos indicus) through high-density snp genotypes. *Frontiers in genetics*, Frontiers Media SA, v. 6, p. 5, 2015. Cited 2 on page 1 and 33.

ZHANG, F.; YU, X. Wac, a functional partner of rnf20/40, regulates histone h2b ubiquitination and gene transcription. *Molecular cell*, Elsevier, v. 41, n. 4, p. 384–397, 2011. Cited on page 46.

ZHANG, L. *et al.* Population genetic analysis of six chinese indigenous pig meta-populations based on geographically isolated regions. *Animals*, MDPI, v. 13, n. 8, p. 1396, 2023. Cited 3 on page 16, 18, and 36.

3 Genomic analysis revealed genes hotspots on adaptation and risk of deseapparecence of local Marota goat breed

Abstract

The preservation of genetic resources is a major global concern to ensure genetic diversity and sustainable comprehensive identification of Runs of Homozygosity (ROH) and Heterozygosity-Rich Regions (HRR), as well as inbreeding estimation, are essential for the effective management of farm animal genetic resources and for preventing genetic erosion in livestock. Marota and Anglo-Nubian goats are conserved in three different institutional and private herds in Brazil, but their genomic information regarding ROH, HRR, and inbreeding levels is unknown. In this study, we analyzed the characteristics of ROH and HRR in goats using the Illumina Goat SNP50 BeadChip and calculated the inbreeding coefficient based on ROH (FROH). Additionally, we identified candidate genes in genomic regions associated with ROH and HRR. A total of 22,872 runs of homozygosity (ROH) were identified. The average number of ROH per individual between the two goat breeds ranged from 74.73 (Anglo-Nubian commercial goat) to 173.85 (Local Marota goat). Analysis of the distribution of ROH according to their size showed that, for both breeds, the majority of ROH detected were in the short (<2Mb) category (65.60%). ROH-based inbreeding estimates in Marota goats generally range from low to high values, between 0.040 and 0.283, mainly due to the rapid decline in effective population size in recent generations. The annotated genes (*Zbtb11, IL18* and *LPO*) that overlapped with ROH and HRR islands were significantly enriched in immune-related biological processes. Furthermore, we identified several candidate genes (*Tex12, Ypel5, Capn14* and *Galnt14*) that may be associated with adaptations related to diverse traits, such as embryonic development, body growth, lipid homeostasis, and brain functions. This study reveals levels of genomic inbreeding and associates genes in ROH and HRR regions with environmental adaptation pathways in goats maintained in institutional herds. These results are valuable for prioritizing populations for conservation and developing management practices for the genetic improvement of these Brazilian goats

Keywords: local goat populations; genomic regions; inbreeding; heterozygosity.

3.1 Introduction

Among livestock species, goats *(Capra hircus)* play a crucial role in economic development, contributing significantly to rural production, particularly in milk production, and to rural subsistence by providing additional income to families in developing countries (Lin *et al.*, 2013; Periasamy *et al.*, 2017; Batzios *et al.*, 2023). In the Brazilian context, rural activity with small ruminants is divided between large producers and small breeders, with goats being a striking presence, adapted to the semi-arid tropics (Moura *et al.*, 2019a; Moura *et al.*, 2019b).

Indigenous populations appear to be more genetically diverse than the commercial breeds, as they have been improved and established through a long breeding history, by processes remarkably different from those used for commercial breeds. Several goat breeds adapted to Brazil's semi-arid tropical climate represent valuable reservoirs of genes related to resistance to adverse conditions, parasites and diseases, providing a rich genetic diversity (Boettcher *et al.*, 2015; Diniz *et al.*, 2023).

It is therefore important to identify these genetic resources in order to provide support and guidance in drawing up policies to guarantee the conservation and sustainable reproduction of this biological heritage. The sustainable management of farm animal genetic resources is an urgent and complex task needed to protect their significant contribution to the collection of genetic Taking two Brazilian goat breeds as examples: the Anglo-Nubian (commercial) and Marota (local) breeds (Diniz *et al.*, 2023; Moura *et al.*, 2019a).

The Marota goat, bred in the semi-arid region, is a specific local type of small, hardy goat at risk of extinction, maintained by government institutions and partners in northeastern Brazil (Araujo *et al.*, 2009; Moura *et al.*, 2019a). This population constitutes a valuable reservoir of genetic diversity, as they can be considered a source of genes related to the ability to minimize negative effects linked to climate and environmental changes, resistance to parasites and diseases, attributable to their evolutionary history of adaptability to adverse environmental conditions (Moura *et al.*, 2019a; Boettcher *et al.*, 2015; Araujo *et al.*, 2009). Despite its long history in Brazil, the genetic diversity and population structure of the Marota breed are still poorly understood (Moura *et al.*, 2019b; Diniz *et al.*, 2023).

Thus, the genomic characterization of these genetic resources is desirable in order to serve as a support or reference to help draw up policies for the conservation and sustainable reproduction of this biological heritage. In this context, genomic scanning, especially with a focus on runs of homozygosity (ROH), can provide crucial insights into molecular evolution, adaptability and ROH-based inbreeding coefficient (FROH) in commercial and local goat populations (Zavarez *et al.*, 2015; Purfield *et al.*, 2017; Marras *et al.*, 2015; Mulim *et al.*, 2022).

Autozygosity is the homozygous state of identical-by-descent (IBD) alleles, which can result from several phenomena. Increased inbreeding (F) leads to various negative effects such as reduced genetic variance, lower individual performance (inbreeding depression), and decreased population viability (McQuillan *et al.*, 2008; Pemberton *et al.*, 2012). Nowadays, among several alternative methods, F estimated from ROH is considered the most precise (Marras *et al.*, 2015; Selli *et al.*, 2021). The ROH approach can also be used to identify genomic regions potentially under selection and involved in defining population-specific traits (Bosse *et al.*, 2012; Pemberton *et al.*, 2012; Selli *et al.*, 2021).

Considering gene recombination, genome regions of large amount of homozygosity are supposed to be a result from selection, leading to favorable alelles in surrounding regions. Therefore, ROH analysis has been employed to explore signatures of selection in several species, including both ruminants and non-ruminants (Marchesi *et al.*, 2018; Xu *et al.*, 2019; Cendron *et al.*, 2020; Mulim *et al.*, 2022).

Runs of heterozygosity, described as heterozygosity-rich regions (HRRs), can also offer insights into population structure and demographic history. These HRRs may contain loci harboring genes associated with adaptive traits, such as response to heat stress, immune response, survival rate, fertility, and other fitness-related traits (Selli *et al.*, 2021; Chessari *et al.*, 2024).

In this study, we analyzed the characteristics of ROH and HRR in the locally adapted Marota breed and the commercial Anglo-Nubian breed using SNP data. We calculated the inbreeding coefficient based on ROH (FROH) and identified candidate genes associated with ROH and HRR. This analysis aims to provide a detailed understanding of genetic variability and population structure, contributing to the development of conservation programs and the enhancement of production systems.

3.2 Methods

3.2.1 Ethics statement

The database for this study, titled “Population genomics, genetic diversity, and gene introgression in naturally adapted goat conservation flocks in Brazil”, was approved by the Ethics and Research Committee of the Federal University of Piauí, adhering to the established guidelines, with registration number 058/14.

3.2.2 Location of sampled data

The state of Piauí, located in the northeast region of Brazil, borders the states of Maranhão, Ceará, Bahia, Tocantins and Pernambuco (Fig 3.1). It spans an area of 251,529 km^2 and harbors the third-largest goat population in Brazil (IBGE, 2017).

The data used in this study included 192 animals from two distinct breeds found in the semi-arid biome: the rare landrace Marota (n = 86), and the standardized breed Anglo-Nubian (n = 106). Individuals from the Anglo-Nubian breed were acquired from the municipality of Teresina, PI, Brazil (05°02'39.95" S, 42°47'03.70" W), while those from the Marota landrace were provided the National Goat Conservation Program at Embrapa (05°19'20" S, 41°33'09" W) and a private farm in Elesbão Veloso, PI, Brazil (06°12'07" S, 42°08'25" W).

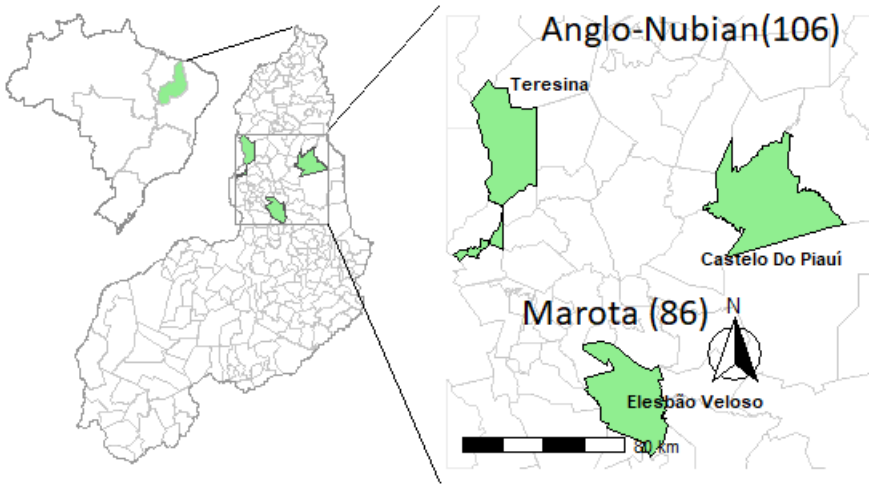

Figure 3.1 – Geographical and phenotypic distribution of two goat populations in Piauí, Brazil.

3.2.3 Genotyping

Genotyping was performed using the 50K Illumina Goat Bead Chip, which contains 53,347 evenly spaced SNPs. This chip employs Illumina™ Infinium technology and the iScan platform. The genotyping protocol adhered to the guidelines provided by the manufacturer, available at link (Moura *et al.*, 2019a; Moura *et al.*, 2019b).

3.2.4 Quality control

The quality of the genotypic data was assessed using PLINK version 1.9 (Chang *et al.*, 2015). The chromosomal coordinates of SNPs were aligned to the ARS1 genome assembly (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_001704415.2). Markers assigned to unmapped contigs and to sex chromosomes were excluded from further analysis. Quality control metrics included a call-rate threshold of ≥ 0.95 (Meyermans *et al.*, 2020; Selli *et al.*, 2021), yielding a final dataset of 43,133 SNPs.

3.2.5 Detection of ROH and HRR

The detectRUNS package, as proposed by Biscarini *et al.* (2019) for the R environment, was used to detect both ROH and HRR using the sliding window and consecutive

runs methods (Chang *et al.*, 2015; Arias *et al.*, 2023). The following sliding window constraints were imposed for ROH detection: (i) window size of 20 SNPs across the genome (ii) the minimum number of SNPs included in a sliding window was set at 20 (minSNP); (iii) the maximum number of heterozygotes allowed in a run was 1 (maxOppWindow); (iii) the number of missing SNPs allowed per window was 1 (maxMissWindow). Run-related parameters included: (i) the maximum gap between consecutive SNPs was set to 250 kb (maxGap = 250,000 bps); (ii) the minimum ROH length was established at 1000 kb (minLengthBps = 1,000,000); (iii) the minimum density of one SNP every 70 kb (minDensity-1/70); (iv) the proportion of homozygous overlapping windows was set at 0.05 (Ramos-Onsins *et al.*, 2014; Islam *et al.*, 2019; Meyermans *et al.*, 2020; Saravanan *et al.*, 2021a; Saravanan *et al.*, 2022; Selli *et al.*, 2021; Zhang *et al.*, 2023).

The percentage of chromosomes covered by ROH was estimated following the method proposed by Al-Mamun *et al.* (2015): the sum of all ROHs on a chromosome divided by the number of individuals with ROHs on that chromosome gives the average length of the ROHs; the average length of the ROH on that chromosome is then divided by the length of the corresponding chromosome and multiplied by 100 to convert it into a percentage (Al-Mamun *et al.*, 2015; Saravanan *et al.*, 2021a; Tong *et al.*, 2022).

We estimated the ROH for each goat and categorized them into the following length classes (1–2, 2–4, 4–8, 8–16, and >16 Mb). The frequency and average length of the ROH for each length category were calculated. We determined the total number of ROH identified for each length category in each individual of each breed. The sum and mean of ROH were calculated by summing the length of all ROH for each animal in the goat populations and then averaging the results by breed population (Islam *et al.*, 2019; Deniskova *et al.*, 2021).

The genomic inbreeding coefficient (FROH) for each breed was calculated using the following method (Equation 3.1), as described by McQuillan *et al.* (2008):

$$F_{ROH} = \frac{L_{ROH}}{L_{AUTO}} \tag{3.1}$$

where LROH represents the total length of ROH of each individual in the genome, and LAUTO is the length of the autosomal genome of the goat, set at 2,464.80 Mb (McQuillan *et al.*, 2008; Dzomba *et al.*, 2021; Fabbri *et al.*, 2021).

3.3 Selection signatures and detection of ROH islands

To identify the genomic regions that were most commonly associated with ROH, the percentage of occurrences of SNP in ROH was estimated by counting the number of times that each SNP appeared in a ROH and dividing that number by the number of

animals for each breed. The SNP frequencies (%) within detected ROH were assessed for each breed and plotted against the SNP positions across autosomes. ROH islands were defined as clusters of runs that were ≥ Mb. with a minimum of 12 SNPs and foundin more than ≥ 50% samples (Peripolli *et al.*, 2018; Saravanan *et al.*, 2021b; Martínez-Rocha *et al.*, 2022; Wang *et al.*, 2022; Carrara *et al.*, 2024).

3.4 Heterozygosity-rich regions detection (HRR)

HRR detection employed the consecutive runs (CS) method, with the following parameters: (i) a minimum of 10 SNPs in a run; (ii) a maximum of five homozygous SNPs in a run; (iii) a maximum of five missing SNPs in a run; (iv) the minimum HRR length was set to 1 Mb; (v) the maximum gap between consecutive SNPs was set to 1 Mb (Arias *et al.*, 2023). HRR segments are predicted to be less frequent and shorter (<1 Mb) than ROH segments. The primary criterion for their identification is the number of allowed homozygous SNPs within a segment (Biscarini *et al.*, 2020; Mulim *et al.*, 2022; Arias *et al.*, 2023). HRR must present a frequency of 45% in individuals in the population (Mulim *et al.*, 2022).

3.5 Gene Annotation

Information on the annotated genes within the highly homozygous (ROH islands) and heterozygous (HRRt islands) genomic regions detected in Marota and Anglo-Nubian goats was obtained from the Genome Data Viewer tool provided by NCBI and GeneCards[1]. The *Capra hircus* genome assembly GCA_000317765.1 was used as a reference. Additionally, we conducted an extensive literature search.

3.6 Results

3.6.1 Detection of ROH

A total of 22,872 ROH were identified in the 192 samples analyzed. The number of ROHs on each goat chromosome is displayed in Figure 3.2 which illustrates the plausible correlation between the number of ROH fragments and the length of each chromosome.

ROH segments were identified in all breeds, with mean lengths ranging from 2.01 Mb in the Marota breed to 2.07 Mb in the Anglonubiana breed. The average number of ROH varied from 74.73 to 173.85 in the Anglonubiana and Marota breeds, respectively. The maximum individual ROH length and number were observed in the Marota breed

[1] Last access 31.05.2024

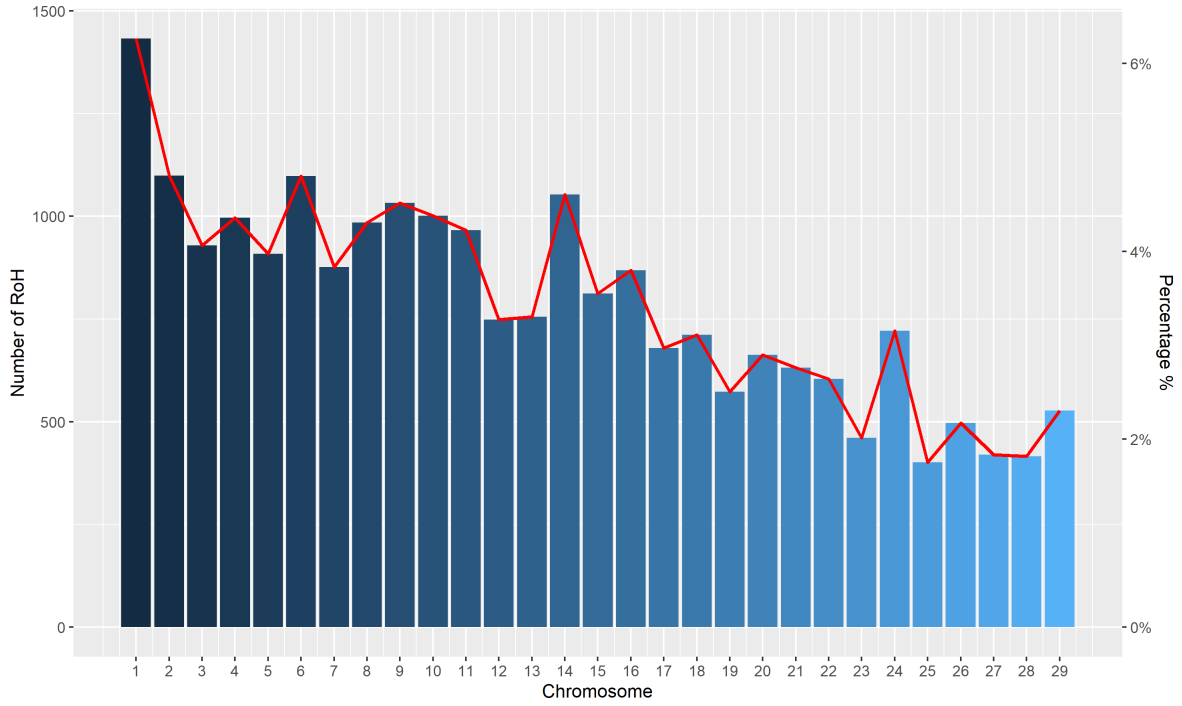

Figure 3.2 – Number of ROH per chromosome and average percentage of each chromosome covered by ROH for all goats.

(14.55 Mb and 173.85, respectively), while the Anglonubiana breed exhibited the lowest values (1.0 Mb and 4.0 Mb, respectively), see (Table 3.1).

Table 3.1 – ROH length and number in goat populations from the Brazilian semi-arid region.

		ROH length			ROH number		
Breed	n	Mean	Min	Max	Mean	Min	Max
Marota	86	2.01 ± 1.13	1.00	14.55	173.85 ± 49.46	56	269
Anglo-Nubian	106	2.07 ± 1.12	1.00	12.87	74.73 ± 45.9	4	270

n: number of individuals; Min: minimum values observed in estimations of individual animals within each breed; Max: maximum values observed within each breed.

3.6.2 Genomic inbreeding (FROH) coefficients

The average genomic inbreeding coefficient (FROH) for the Anglo-Nubian breed was 0.0627, with a range from 0.003 to 0.323. In contrast, the Marota breed exhibited higher FROH values, with an average of 0.1419 and a range from 0.040 to 0.283 (Fig 3.3). To investigate recent and past inbreeding in each breed, the distribution of different ROH size classes was established (Table 3.2).

The percentage of ROH in the length classes for the populations is shown in Table 3.2. The two populations presented more ROH between 1 and 2 Mb than in the other classes (Table 3.2), with Marota having the highest proportion (65.60 %), and a very low proportion of ROH above 8 Mb (0.4%).

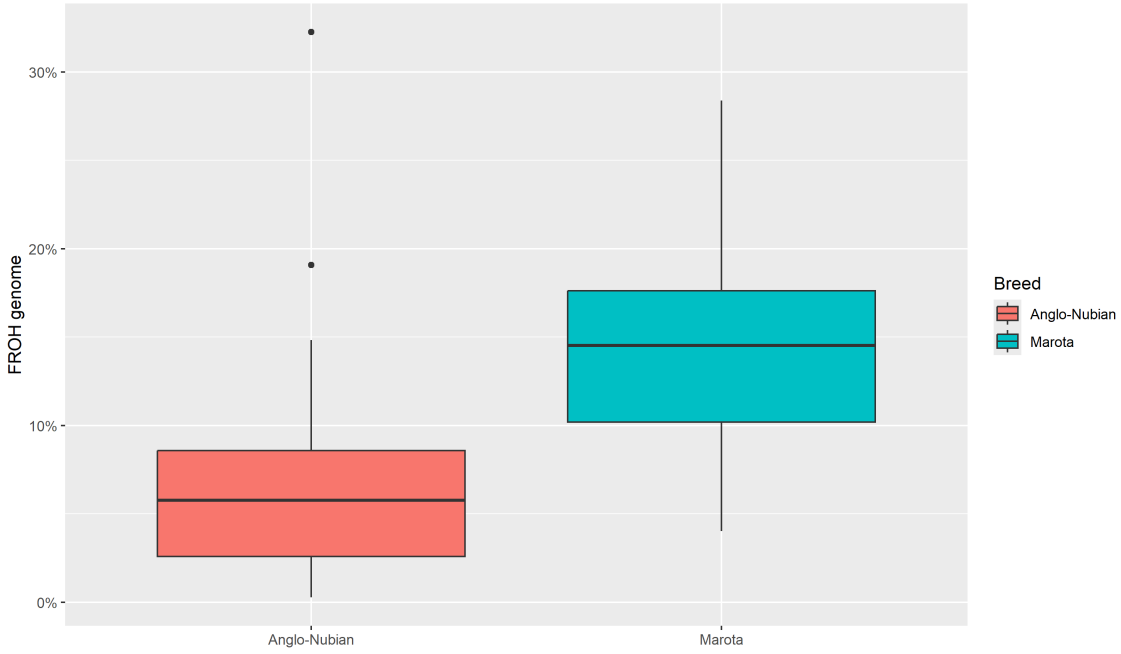

Figure 3.3 – Boxplot of FROH distribution within the Anglo-Nubian and Marota goat breeds.

The Anglonubian breed had exactly the same proportion of ROH above 8 Mb (0.4%).

Table 3.2 – Summary of number of runs of homozygosity (ROH) by size class in Marota and Anglo-Nubian goats.

Class (Mb)	Marota		Anglo-Nubian	
1-2	9,813	65.60%	4,26	62.60%
2-4	4,962	28.50%	2,473	31.20%
4-8	818	5.50%	456	5.80%
8-16	60	0.40%	30	0.40%
> 16	0	0.00%	0	0.00%
Total	14,951	100%	7,921	100%

Mb= Million base pairs

3.6.2.1 Run of homozygosity islands

To identify genomic regions potentially candidates for selection and/or conservation, the frequency of SNPs contained in the runs was plotted on the autosomes of each breed (Fig 3.4). Generally, adjacent SNPs with a proportion of ROH occurrences above the adopted threshold form genomic regions called ROH islands.

As shown in Fig. 3.4, a total of seven genomic regions characterized by a high occurrence of ROH, potentially important for selection, were identified. Applying the

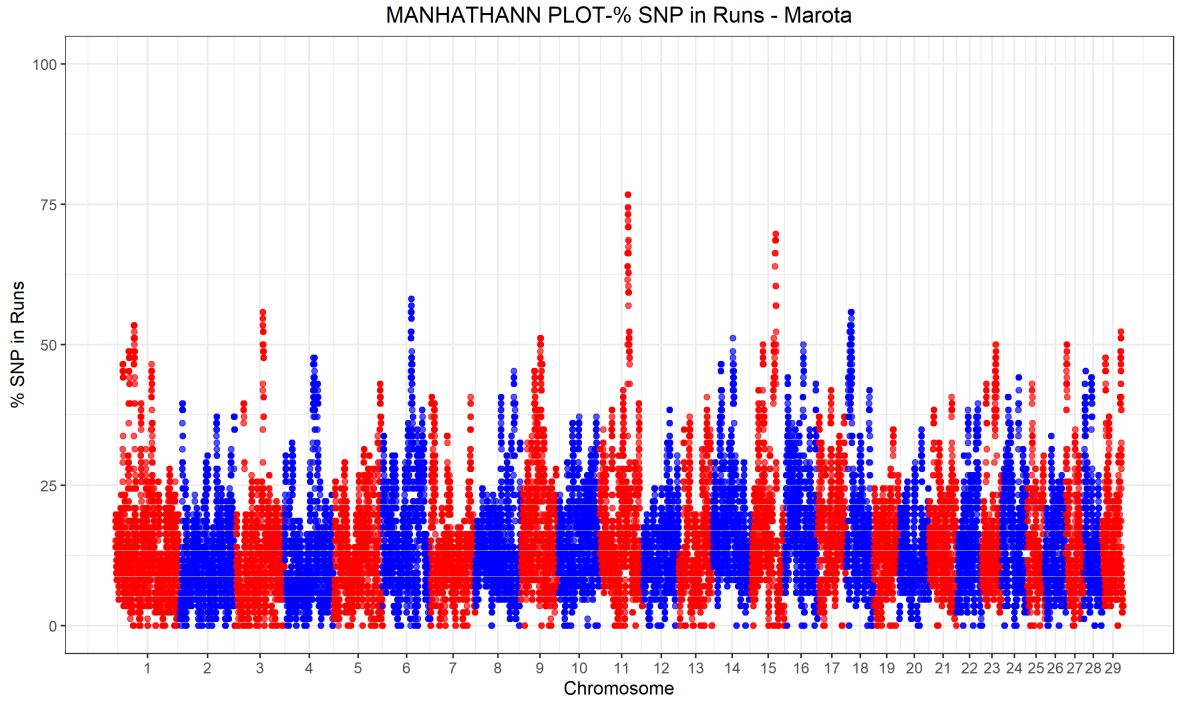

Figure 3.4 – Manhattan plot of the proportion of times each SNP falls within a ROH in the Marota goat breed.

aforementioned 50% threshold, seven ROH segments were detected (Table 3.3), with the Marota breed exhibiting the largest number of these regions.

The autosomes that consistently displayed significant regions with high ROH occurrence were chromosomes 01, 03, 06, 11, 15, and 18. Interestingly, the Anglonubian breed was an exception, with no ROH islands detected in its genomic composition (Table 3.3).

3.6.3 Heterozygosity-rich regions

The analysis of HRR was conducted utilizing the investigative method (CR), yielding a total of 17,278 HRR. Table 3.4 provides descriptive statistics for HRR categorized by breed.

Given the analysis, regions identified in at least 45% of the animals were deemed HRR islands and were subsequently scrutinized for candidate genes and pathways. The specific HRR islands discovered in the Marota and Anglo-Nubian goat breeds are reported in Table 3.5.

Table 3.3 – Gene annotation of runs of homozygosity (ROH islands) found in the autosomal genome of the Marota goat population.

*Chr.	Start *(BP)	End *(BP)	*n	Genes
1	45,076,527	45,783,017	12	*ABI3BP, SENP7, IMPG2, TRMT10C,PCNP, ZBTB11*
6	69,566,353	70,525,889	17	*LNX1, GSX2, CHIQUE2, PDGFRA*
11	70,903,218	72,142,096	21	*PLB1, FOSL2, BRE, TRNAC-GCA, RBK5, MRPL33, SLC4A1AP, SUPT7L, GPN1, CCDC121, ZFN512, GCKR, FNDC4, IFT172, KRTCAP3, NRBP1, PCPM1G, ZFN513, SNX17, FEI2B4, GTF3C2, MPV17, UCN, TRIM54, DNAJ5G, SLC3AOA3, ATRAID*
11	67,889,566	70,479,268	48	*ANX4, GMCL1, SRNP27, MXD1, ASPRV1, PCBP1, C11H2ofr42, TIA1, PCYOX1, SNRPG, EHD3, CAPN14, GALNT14, CAPN13, LCLAT1, LBH, ALK, CLIP4, GMCL1, SRNP27, MXD1, ASPRV1, PCBP1, C11H2ofr42, TIA1, SNRPG, EHD3, TRNAG-CCC, CAPN14, GALNT14, TRNAG-UCC, CAPN13, YPEL5*
15	57,947,094	60,035,602	34	*NCAM1, C15H11orf57, USP28, CLDN25, ZW10, DRD2, HTR3B, TMPRSS5, ANKK1, TTC12, PT, BCO2, TEX12, IL18, PIH1D2, DLAT, DIXDC1*
18	11,115,584	12,091,788	18	*CDH13, HSBP1, MLYCD, OSGIN1, NECAB2, SLC38A8, MBTPS1, HSDL1, DNAAF1, TAF1C, ADAD2, KCNG4, WFDC1, TLDC1*
18	8,548,476	9,214,650	15	*CDIL2, DYNLRB2, CMC2, CENPN, ATMIN, C18H16ofr46, GCSH, PKD1L2*

* Chr.: chromosome, *BP: base pairs, *n: number of SNPs in the ROH island.

Table 3.4 – HRR length and number in goat populations from the Brazilian semi-arid region.

		HRR length			HRR number		
Breed	n	Mean	Min	Max	Mean	Min	Max
Marota	86	1.18 ± 0.2	1.00	2.98	67.24 ± 15.49	41	108
Anglo-Nubian	106	1.18 ± 0.18	1.00	2.63	108.44 ± 34.59	48	360

n: number of individuals; Min: minimum values observed in estimations of individual animals within each breed; Max: maximum values observed within each breed.

Table 3.5 – Gene annotation of the heterozygosity-rich regions (HRR islands) found in the autosomal genome of Marota and Anglo-Nubian goats.

Breed	*Chr.	Start *(BP)	End *(BP)	*n	Genes
Marota	19	7,937,845	9,069,843	10	MSI2, MRPS23, CUEDC1, VEZF1, SRSF1, DYNLL2, MKS1, LPO, MPO, TSPOAPI, SUPT4H1, RNF43, HSF5.
Marota	13	35,082,374	35,788,791	8	BAMBI, WAC, MPP7
Anglo-Nubian	3	164,779	1,669,699	18	DUSP28, OTOS, ANKMY1, GPC1, COPS9, NDUFA10
Anglo-Nubian	27	18,840,355	19,909,422	11	GTF2E2, SMIM18, RBPMS, DCTN6, MBOAT4, LEPROLTL1, SARAF, TRNAQCUG, TRNAE-UUC
Anglo-Nubian	21	22,153,516	23,030,023	10	HOMER2, CPEB1, AP3B2, FSD2, UAU, FAM103A1,
Anglo-Nubian	15	80,435,546	81,371,836	9	MRE11A, GRIA4, MSANTD4, KBTBD3, AASDHPPT, GPR83, TRNAG-UCC, IZUMO1R, ANKRD49

* Chr.: chromosome, *BP: base pairs, *n: number of SNPs in the ROH island.

3.7 Discussion

3.7.1 Identified Homozygosity Patterns (ROH)

The analysis of patterns of homozygosity (ROH) has emerged as a relevant informative tool for elucidating the demographic history of populations, as it provides precise information on genetic diversity, while the evaluation of the length of ROH makes it possible to detect past and recent genomic inbreeding in animals (Curik; Ferenčaković; Sölkner, 2014; Deniskova *et al.*, 2021; Mastrangelo *et al.*, 2021).

It was observed that the average number of ROH in local goat breeds from Northeast Brazil ranges between 74.73 (Anglonubian goats) and 173.85 ROH (Marota goats), respectively. These values are higher compared to Turkish (60 ROH), Russian

(approximately 73 ROH), and Central Asian (90 ROH) breeds (Bertolini *et al.*, 2018a; Deniskova *et al.*, 2021). In the context of American goats, with an average of 133 ROH, only Anglonubian goats have lower values (Bertolini *et al.*, 2018a).

In general, differences in the number and characteristics of ROH may indicate historical disparities in genomic structure. Highly selected breeds often exhibit a greater abundance of ROH and broader coverage compared to local breeds (Purfield *et al.*, 2017; Bertolini *et al.*, 2018a; Dzomba *et al.*, 2021; Visser *et al.*, 2023). In our study, the local Marota breed stands out for its elevated number of ROH when contrasted with the Anglonubian goat breed. Bertolini *et al.* (2018a) identified a trend of increased homozygosity in local breeds, attributing this phenomenon to their small population size and geographical isolation.

3.7.2 Genomic inbreeding of homozygosity (FROH)

This study presents information on the level of genomic inbreeding in goat breeds in the Brazilian semi-arid region. Knowledge about the level of endogamy in a population has practical applications in management and improving the understanding of relationships between individuals. This helps to accurately calculate inbreeding levels, correct errors often found in family trees, and identify inbreeding of ancestors not documented in pedigrees. This method is especially suitable for studying genetic variation in small, at-risk populations (Nandolo *et al.*, 2019; Michailidou *et al.*, 2019).

The inbreeding coefficient was calculated from the genomic data of all tested individuals. The average FROH in Brazilian local Marota goats (0.1419) iexceeded that of other goat breeds worldwide (FROH= 0.12). In contrast, Anglo-Nubian goats exhibited lower FROH values (0.0627) and demonstrated a lower degree of inbreeding compared to several goat breeds (Michailidou *et al.*, 2019; Li *et al.*, 2022).

In the local breed, this situation could encourage reproduction between closely related individuals, leading to a significant prevalence of recessive genes in homozygous conditions due to inbreeding and genetic alterations. Consequently, they would become more susceptible to selective pressures (Brito *et al.*, 2017; Bertolini *et al.*, 2018a; Islam *et al.*, 2019; Mastrangelo *et al.*, 2021). Therefore, closely monitoring the population of this goat breed is essential to prevent the loss of genetic resources and minimize the negative effects of harmful mutations, inbreeding depression, and reduced genetic diversity (Moura *et al.*, 2019a).

In the context of analyzing goat populations in institutional herds, a significant prevalence of short ROH segments was observed, specifically 21,508 segments (<4 Mb), representing 94% of all ROH detected. In the present study, 1,364 ROH larger than 4 Mb were detected in the genomes of the Marota and Anglonubian breeds, with an average of

0.06 ROH per animal.

Generally, comparing ROH results is not a simple task due to the variety of criteria used in different studies. Our research revealed that in the Marota and Anglonubian breeds, the majority of ROH are short or medium in size. Therefore, our findings suggest that animals experienced ancestral inbreeding events due to the absence of longer ROH (Mulim *et al.*, 2022; Islam *et al.*, 2019; Brito *et al.*, 2017; Onzima *et al.*, 2018). Deniskova *et al.* (2021) similar results for ROH <4 using a 50K panel for Russian goat.

The detection of Anglo-Nubian gene introgression in Marota herds (Moura *et al.*, 2019a), along with the prevalence of short-length ROH and a relatively high inbreeding coefficient (FROH = 0.1419), highlights the complexity of genetic variability in this breed. While the implemented management strategies are helping to preserve the genetic characteristics of the Marota breed, the high level of inbreeding indicates that challenges remain. Ongoing monitoring and adjustments to these strategies are essential to ensure the genetic diversity needed for adaptation and the integrity of local goat (Moura *et al.*, 2019a; Cortellari *et al.*, 2021; Mastrangelo *et al.*, 2021).

3.7.3 Runs of homozygosity(ROH) and heterozygosity-rich regions (HRR)

Livestock farming is a human practice, influenced by various environmental factors. When environmental conditions are unfavorable, animals can adapt in various ways to cope with stressors. Therefore, biological adaptation is essential to ensure that animals are healthy and productive. In their DNA, it is possible to find marks of natural selection that arise in response to environmental pressures, and that can be identified through genomic and bioinformatics techniques (Cortellari *et al.*, 2021; Mulim *et al.*, 2022; Manunza *et al.*, 2023).

In this context, some goat breeds introduced to the Brazilian semi-arid region demonstrate adaptive evolution aimed at survival, reproduction, and production across different ecological zones. It is likely that their genomes have developed unique genetic characteristics, evolving to adapt to remarkably heterogeneous environments (Moura *et al.*, 2019a; Diniz *et al.*, 2023). It is therefore expected that artificial and natural selection will act on animal genomic diversity, leaving positive or deleterious adaptive signatures (Brito *et al.*, 2017; Bertolini *et al.*, 2018b).

Analysis of ROH and HRR in the Marota and Anglo-Nubian breeds has revealed markers potentially under selection for adaptation to the semi-arid environment of Brazil. Our discoveries have led to the identification of target genes related to adaptation, specifically concerning the immune/inflammatory response, energy homeostasis, reproductive and production traits, and heat stress (Biscarini *et al.*, 2020; Cortellari *et al.*, 2021; Chokoe *et al.*, 2022; Mulim *et al.*, 2022; Manunza *et al.*, 2023; Chessari *et al.*, 2024).

Research into the specific functions of genes has unveiled an intricate panorama of biological processes, wherein genes play distinct and essential roles in cellular functions, contributing to a wide range of biological processes. For example, the *GMCL1* and *TEX12* genes are known to be involved in the regulation and/or participation of reproductive processes, specifically in sperm production (Gjerstorff *et al.*, 2012; Liu; Cai; Liu, 2018; Dunce; Salmon; Davies, 2021; Sudhakar *et al.*, 2023).

The*Zbtb11* gene, with its integrase-like HHCC zinc finger, serves as a crucial regulator in neutrophil differentiation, contributing to the immune defense system. Similarly, the*IL18* gene expresses the IL-18 protein, a pro-inflammatory cytokine pivotal for immune response regulation, activating immune system cells like T lymphocytes and natural killer (NK) cells, thereby enhancing the production of inflammatory cytokines (Tominaga *et al.*, 2000; Keightley *et al.*, 2017).

The*LBH* and*YPEL5* genes play specific roles in embryonic development, RNA processing, and cell cycle regulation, thereby influencing cell growth (Hosono *et al.*, 2010). The data indicate that certain genes, subject to natural selection, are involved in a wide range of cellular metabolic processes, including protein cleavage and lipid homeostasis. These genes include *PCYOX1, CAPN14, GALNT14, TRNAG-UCC, CAPN13* and *LCLAT1* (Lin *et al.*, 2016; Miller *et al.*, 2019; Silva *et al.*, 2020; Banfi *et al.*, 2023). Genes such as *CLDN25* and *ZW10* actively participate in regulating cellular permeability and chromosomal segregation during cell division and may play roles in cell growth (Vallee; Varma; Dujardin, 2006; Nakajima *et al.*, 2015).

Heterozygosity-rich regions present a new concept introduced by Williams *et al.* (2016), being much less characterized than ROH in livestock. They are identified as heterozygous genomic regions potentially associated with disease resistance, immunity, and adaptive processes, serving as a valuable genetic reservoir and providing additional insights into goat genomes (Chessari *et al.*, 2024).

Heterozygosity loci are clustered in islands throughout the genome and are significantly rarer and shorter compared to ROH (Chessari *et al.*, 2024). Several studies have reported divergent means of HRR per chromosome (Santos *et al.*, 2021; Mulim *et al.*, 2022; Chessari *et al.*, 2024). Variations in methods, animal populations, and SNP arrays may account for differences from our results, a trend similarly observed by Chessari *et al.* (2024).

After applying a threshold of 45%, there were 66 SNPs. These SNPs formed six regions across six chromosomes. They are located on chromosomes 3, 13, 15, 19, 21, and 27, determined by a single region. In these regions, we identified 46 gene. Based on genetic functions obtained from http://www.genecards.org (last accessed on July 4, 2024), we conclude that the genes found in the HRR are important in adaptive biological processes.

Of the 46 gene annotations, three genes with known and well-described functions are particularly notable. The *MPO* gene, located on chromosome 19, is responsible for encoding the enzyme myeloperoxidase. In humans, myeloperoxidase is involved in generating free radicals and hypochlorite, released by neutrophils during the inflammatory response to bacterial infections (Cederlund *et al.*, 2015).

Similarly, the*LPO* gene encodes the enzyme lactoperoxidase, which plays a fundamental role in immune defense against bacterial infections by catalyzing the production of HOSCN from H2O2 and SCN-, a component of the antibacterial defense system present in the respiratory tract (Wijkstrom-Frei *et al.*, 2003). There is also the*WAC* gene is a vital component of the molecular mechanism that coordinates the cell's response to genotoxic stress, ensuring genome integrity and maintenance of cellular homeostasis through transcriptional regulation of p53 target genes (Zhang; Yu, 2011).

Our findings are in line with previous research involving Russian cattle, particularly the identification of a single candidate region on chromosome 15 of the Anglo-Nubian goat. Several genes are located within this region, namely *MSANTD4* and *GRIA4* . Both genes are potential contributors to the climate stress resistance phenotype, given their indirect functions in heat shock response (*MSANTD4*) and body thermoregulation (*GRIA4*)(Igoshin *et al.*, 2019).

3.8 Conclusions

Genomic diversity and ROH parameters reveal well-defined differences that correspond to the geographical distribution of goats, as well as their breeding history and population size.

The ROH models revealed that the high FROH values in Marota goats could seriously affect their overall biological fitness. High levels of inbreeding are responsible for the fixation of genomic regions carrying deleterious mutations, which could push the population toward an extinction vortex

Genes found on ROH islands and HRR are associated with relevant pathways of environmental adaptation. Thus, the presence of ROH and HRR islands in the genome of landraces highlights the evolutionary forces that drive adaptation.

Finally, ROH and HRR approaches underscores the importance of utilizing applied genome marker-based data to mitigate the loss of diversity in the future and to inform the development of more effective breeding and conservation programs.

Bibliography

ADRIAANSE, L. S.; RENSLEIGH, C. Web of science, scopus and google scholar: A content comprehensiveness comparison. *The Electronic Library*, Emerald Group Publishing Limited, v. 31, n. 6, p. 727–744, 2013. Cited on page 7.

AL-MAMUN, H. A. *et al.* Genome-wide linkage disequilibrium and genetic diversity in five populations of australian domestic sheep. *Genetics Selection Evolution*, Springer, v. 47, p. 1–14, 2015. Cited on page 36.

ANDREWS, K. R. *et al.* A bioinformatic pipeline for identifying informative snp panels for parentage assignment from rad seq data. *Molecular ecology resources*, Wiley Online Library, v. 18, n. 6, p. 1263–1281, 2018. Cited on page 16.

ARAUJO, A. M. de *et al.* Crescimento e mortalidade em um rebanho de conservação de caprinos marota no brasil. Revista Cientfica de Producão Animal, v. 11, n. 2, p. 103-109, 2009., 2009. Cited 2 on page 1 and 33.

ARIA, M.; CUCCURULLO, C. bibliometrix: An r-tool for comprehensive science mapping analysis. *Journal of informetrics*, Elsevier, v. 11, n. 4, p. 959–975, 2017. Cited 2 on page 7 and 16.

ARIAS, K. D. *et al.* Approaching autozygosity in a small pedigree of gochu asturcelta pigs. *Genetics Selection Evolution*, Springer, v. 55, n. 1, p. 74, 2023. Cited 2 on page 36 and 37.

ARMENTA-MEDINA, D. *et al.* A bibliometric analysis of mexican bioinformatics: a portrait of actors, structure, and dynamics. *Biology*, MDPI, v. 11, n. 1, p. 131, 2022. Cited 3 on page 5, 7, and 15.

BANCHI, P. *et al.* Trends in small animal reproduction: a bibliometric analysis of the literature. *Animals*, MDPI, v. 12, n. 3, p. 336, 2022. Cited 2 on page 7 and 16.

BANFI, C. *et al.* Prenylcysteine oxidase 1 is a key regulator of adipogenesis. *Antioxidants*, MDPI, v. 12, n. 3, p. 542, 2023. Cited on page 45.

BATZIOS, A. *et al.* Technical indicators and economic performance of dairy goat farms in greece: An efficiency analysis. *Livestock Science*, Elsevier, v. 271, p. 105210, 2023. Cited on page 33.

BERTOLINI, F. *et al.* Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. *Genetics Selection Evolution*, Springer, v. 50, p. 1–12, 2018. Cited 2 on page 15 and 43.

BERTOLINI, F. *et al.* Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genetics Selection Evolution*, Springer, v. 50, p. 1–24, 2018. Cited on page 44.

BIRKLE, C. *et al.* Web of science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , v. 1, n. 1, p. 363–376, 2020. Cited on page 6.

BISCARINI, F. *et al.* Detect runs of homozygosity and runs of heterozygosity in diploid genomes. CRAN (The Comprehensive R Archive Network), 2019. Cited on page 35.

BISCARINI, F. *et al.* Insights into genetic diversity, runs of homozygosity and heterozygosity-rich regions in maremmana semi-feral cattle using pedigree and genomic data. *Animals*, MDPI, v. 10, n. 12, p. 2285, 2020. Cited 2 on page 37 and 44.

BJELLAND, D. *et al.* Evaluation of inbreeding depression in holstein cattle using whole-genome snp markers and alternative measures of genomic inbreeding. *Journal of dairy science*, Elsevier, v. 96, n. 7, p. 4697–4706, 2013. Cited on page 18.

BOETTCHER, P. J. *et al.* Genetic resources and genomics for adaptation of livestock to climate change. *Frontiers in genetics*, Frontiers Media SA, v. 5, p. 461, 2015. Cited 2 on page 1 and 33.

BOSSE, M. *et al.* Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS genetics*, Public Library of Science San Francisco, USA, v. 8, n. 11, p. e1003100, 2012. Cited 3 on page 6, 16, and 34.

BRITO, L. F. *et al.* Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide snp markers. *BMC genomics*, Springer, v. 18, p. 1–20, 2017. Cited 5 on page 6, 15, 16, 43, and 44.

CARRARA, E. *et al.* Assessment of runs of homozygosity, heterozygosity-rich regions and genomic inbreeding estimates in a subpopulation of guzerá (bos indicus) dual-purpose cattle. *Journal of Animal Breeding and Genetics*, Wiley Online Library, v. 141, n. 2, p. 207–219, 2024. Cited on page 37.

CEBALLOS, F. C. *et al.* Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics*, Nature Publishing Group UK London, v. 19, n. 4, p. 220–234, 2018. Cited on page 1.

CEDERLUND, M. *et al.* A1m/α1-microglobulin is proteolytically activated by myeloperoxidase, binds its heme group and inhibits low density lipoprotein oxidation. *Frontiers in physiology*, Frontiers Media SA, v. 6, p. 11, 2015. Cited on page 46.

CENDRON, F. *et al.* Genome-wide snp analysis reveals the population structure and the conservation status of 23 italian chicken breeds. *Animals*, MDPI, v. 10, n. 8, p. 1441, 2020. Cited 2 on page 2 and 34.

CHANG, C. C. *et al.* Second-generation plink: rising to the challenge of larger and richer datasets. *Gigascience*, Oxford University Press, v. 4, n. 1, p. s13742–015, 2015. Cited 3 on page 17, 35, and 36.

CHESSARI, G. *et al.* Characterization of heterozygosity-rich regions in italian and worldwide goat breeds. *Scientific Reports*, Nature Publishing Group UK London, v. 14, n. 1, p. 3, 2024. Cited 6 on page 16, 18, 19, 34, 44, and 45.

CHOKOE, T. C. *et al.* Conservation status and historical relatedness of south african communal indigenous goat populations using a genome-wide single-nucleotide polymorphism marker. *Frontiers in Genetics*, Frontiers Media SA, v. 13, p. 909472, 2022. Cited 2 on page 1 and 44.

CLARK, D. W. *et al.* Associations of autozygosity with a broad range of human phenotypes. *Nature communications*, Nature Publishing Group UK London, v. 10, n. 1, p. 4957, 2019. Cited on page 16.

COBO, M. J. *et al.* An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of informetrics*, Elsevier, v. 5, n. 1, p. 146–166, 2011. Cited 2 on page 7 and 16.

COLLI, L. *et al.* Assessing the spatial dependence of adaptive loci in 43 european and western asian goat breeds using aflp markers. *PLoS One*, Public Library of Science San Francisco, USA, v. 9, n. 1, p. e86668, 2014. Cited on page 1.

CORTELLARI, M. *et al.* Runs of homozygosity in the italian goat breeds: Impact of management practices in low-input systems. *Genetics Selection Evolution*, Springer, v. 53, p. 1–11, 2021. Cited on page 44.

CURIK, I.; FERENČAKOVIĆ, M.; SÖLKNER, J. Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock science*, Elsevier, v. 166, p. 26–34, 2014. Cited 3 on page 6, 15, and 42.

DENG, H. *et al.* Impact of public research and development and extension on agricultural productivity in china from 1990 to 2013. *China Economic Review*, Elsevier, v. 70, p. 101699, 2021. Cited on page 16.

DENISKOVA, T. E. *et al.* Snp-based genotyping provides insight into the west asian origin of russian local goats. *Frontiers in genetics*, Frontiers Media SA, v. 12, p. 708740, 2021. Cited 4 on page 36, 42, 43, and 44.

DINIZ, F. d. A. *et al.* Capra hircus outliers markers in brazil: Searching for genomic regions under the action of natural selection. *Genetics and molecular biology*, SciELO Brasil, v. 46, n. 3, p. e20230084, 2023. Cited 3 on page 1, 33, and 44.

DUNCE, J. M.; SALMON, L. J.; DAVIES, O. R. Structural basis of meiotic chromosome synaptic elongation through hierarchical fibrous assembly of syce2-tex12. *Nature structural & molecular biology*, Nature Publishing Group US New York, v. 28, n. 8, p. 681–693, 2021. Cited on page 45.

DZOMBA, E. *et al.* Runs of homozygosity analysis of south african sheep breeds from various production systems investigated using ovinesnp50k data. *BMC genomics*, Springer, v. 22, p. 1–14, 2021. Cited 2 on page 36 and 43.

FABBRI, M. C. *et al.* Genetic diversity and population history of eight italian beef cattle breeds using measures of autozygosity. *PLoS One*, Public Library of Science San Francisco, CA USA, v. 16, n. 10, p. e0248087, 2021. Cited 2 on page 6 and 36.

FALCHI, L. *et al.* Effect of genotyping density on the detection of runs of homozygosity and heterozygosity in cattle. *Journal of Animal Science*, Oxford University Press US, p. skae147, 2024. Cited 2 on page 19 and 20.

FAN, H. *et al.* Conservation genetics and genomics of threatened vertebrates in china. *Journal of genetics and genomics*, Elsevier, v. 45, n. 11, p. 593–601, 2018. Cited 2 on page 5 and 16.

GASPAR, D. *et al.* Genome-wide assessment of the population structure and genetic diversity of four portuguese native sheep breeds. *Frontiers in Genetics*, Frontiers Media SA, v. 14, p. 1109490, 2023. Cited on page 18.

GIBSON, J.; MORTON, N. E.; COLLINS, A. Extended tracts of homozygosity in outbred human populations. *Human molecular genetics*, Oxford University Press, v. 15, n. 5, p. 789–795, 2006. Cited on page 5.

GJERSTORFF, M. F. *et al.* Gage cancer-germline antigens are recruited to the nuclear envelope by germ cell-less (gcl). Public Library of Science San Francisco, USA, 2012. Cited on page 45.

HENDRICKS, S. A. *et al.* Genomic assessment of cancer susceptibility in the threatened catalina island fox (urocyon littoralis catalinae). *Genes*, MDPI, v. 13, n. 8, p. 1496, 2022. Cited on page 18.

HOSONO, K. *et al.* Ypel5 protein of the ypel gene family is involved in the cell cycle progression by interacting with two distinct proteins ranbpm and ranbp10. *Genomics*, Elsevier, v. 96, n. 2, p. 102–111, 2010. Cited on page 45.

HUNTER-ZINCK, H. *et al.* Population genetic structure of the people of qatar. *The American Journal of Human Genetics*, Elsevier, v. 87, n. 1, p. 17–25, 2010. Cited on page 5.

IGOSHIN, A. V. *et al.* Genome-wide association study and scan for signatures of selection point to candidate genes for body temperature maintenance under the cold stress in siberian cattle populations. *BMC genetics*, Springer, v. 20, p. 5–14, 2019. Cited on page 46.

ISLAM, R. *et al.* Genome-wide runs of homozygosity, effective population size, and detection of positive selection signatures in six chinese goat breeds. *Genes*, MDPI, v. 10, n. 11, p. 938, 2019. Cited 3 on page 36, 43, and 44.

KEIGHTLEY, M.-C. *et al.* The pu. 1 target gene zbtb11 regulates neutrophil development through its integrase-like hhcc zinc finger. *Nature communications*, Nature Publishing Group UK London, v. 8, n. 1, p. 14911, 2017. Cited on page 45.

KIRIN, M. *et al.* Genomic runs of homozygosity record population history and consanguinity. *PloS one*, Public Library of Science San Francisco, USA, v. 5, n. 11, p. e13996, 2010. Cited on page 5.

LI, G. *et al.* Genome-wide estimates of runs of homozygosity, heterozygosity, and genetic load in two chinese indigenous goat breeds. *Frontiers in Genetics*, Frontiers Media SA, v. 13, p. 774196, 2022. Cited 2 on page 15 and 43.

LIN, B. Z. *et al.* Genetic diversity and structure in a sian native goat analyzed by newly developed snp markers. *Animal science journal*, Wiley Online Library, v. 84, n. 8, p. 579–584, 2013. Cited 3 on page 1, 15, and 33.

LIN, W.-R. *et al.* Galnt14 genotype predicts postoperative outcome of stage iii colorectal cancer with oxaliplatin as adjuvant chemotherapy. *Medicine*, LWW, v. 95, n. 17, p. e3487, 2016. Cited on page 45.

LIU, X. X.; CAI, L.; LIU, F. J. An in silico analysis of human sperm genes associated with asthenozoospermia and its implication in male infertility. *Medicine*, LWW, v. 97, n. 49, p. e13338, 2018. Cited on page 45.

MABLE, B. K. Conservation of adaptive potential and functional diversity: integrating old and new approaches. *Conservation Genetics*, Springer, v. 20, n. 1, p. 89–100, 2019. Cited on page 6.

MANUNZA, A. *et al.* Discovering novel clues of natural selection on four worldwide goat breeds. *Scientific Reports*, Nature Publishing Group UK London, v. 13, n. 1, p. 2110, 2023. Cited on page 44.

MARCHESI, J. *et al.* Relationship of runs of homozygosity with adaptive and production traits in a paternal broiler line. *Animal*, Cambridge University Press, v. 12, n. 6, p. 1126–1134, 2018. Cited 2 on page 2 and 34.

MARINO, R.; PETRERA, F.; ABENI, F. Scientific productions on precision livestock farming: An overview of the evolution and current state of research based on a bibliometric analysis. *Animals*, MDPI, v. 13, n. 14, p. 2280, 2023. Cited on page 16.

MARRAS, G. *et al.* Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in italy. *Animal genetics*, Wiley Online Library, v. 46, n. 2, p. 110–121, 2015. Cited 6 on page 1, 6, 16, 17, 33, and 34.

MARRAS, G. *et al.* Characterization of runs of homozygosity and heterozygosity-rich regions in a commercial turkey (meleagris gallopavo) population. In: *Proceedings of the 11th World Congress of Genetics Applied to Livestock Production, Auckland, New Zealand*. [S.l.: s.n.], 2018. p. 11–16. Cited on page 19.

MARTÍNEZ-ROCHA, R. *et al.* Assessment of genetic diversity, runs of homozygosity, and signatures of selection in tropical milking criollo cattle using pedigree and genomic data. *Genes*, MDPI, v. 13, n. 10, p. 1896, 2022. Cited on page 37.

MASTRANGELO, S. *et al.* Genome-wide patterns of homozygosity reveal the conservation status in five italian goat populations. *Animals*, MDPI, v. 11, n. 6, p. 1510, 2021. Cited 3 on page 42, 43, and 44.

MASTRANGELO, S. *et al.* Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in three local dairy cattle breeds. *Animal*, Cambridge University Press, v. 10, n. 5, p. 746–754, 2016. Cited on page 15.

MCQUILLAN, R. *et al.* Runs of homozygosity in european populations. *The American Journal of Human Genetics*, Elsevier, v. 83, n. 3, p. 359–372, 2008. Cited 4 on page 5, 6, 34, and 36.

MEYERMANS, R. *et al.* How to study runs of homozygosity using plink? a guide for analyzing medium density snp data in livestock and pet species. *BMC genomics*, Springer, v. 21, p. 1–14, 2020. Cited 6 on page 13, 17, 19, 20, 35, and 36.

MICHAILIDOU, S. *et al.* Analysis of genome-wide dna arrays reveals the genomic population structure and diversity in autochthonous greek goat breeds. *PloS one*, Public Library of Science San Francisco, CA USA, v. 14, n. 12, p. e0226179, 2019. Cited on page 43.

MILLER, D. E. *et al.* Genetic, inflammatory, and epithelial cell differentiation factors control expression of human calpain-14. *G3: Genes, Genomes, Genetics*, Oxford University Press, v. 9, n. 3, p. 729–736, 2019. Cited on page 45.

MORAL-MUÑOZ, J. A. *et al.* Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional de la Información*, v. 29, n. 1, 2020. Cited on page 7.

MÖRSCHBÄCHER, A. P.; GRANADA, C. E. Mapping the worldwide knowledge of antimicrobial substances produced by lactobacillus spp.: A bibliometric analysis. *Biochemical Engineering Journal*, Elsevier, v. 180, p. 108343, 2022. Cited on page 16.

MOURA, J. d. O. *et al.* Population genomics and gene introgression in goat herds naturally adapted to brazil. *Revista Ciência Agronômica*, SciELO Brasil, v. 50, n. 3, p. 476–483, 2019. Cited 5 on page 1, 33, 35, 43, and 44.

MOURA, J. O. *et al.* Diversidade genética em caprino marota e manejo genético para rebanhos de conservação. *Archivos de zootecnia*, v. 68, n. 261, p. 138–145, 2019. Cited 3 on page 1, 33, and 35.

MOURA, R. F.; DAWSON, D. A.; NOGUEIRA, D. M. The use of microsatellite markers in neotropical studies of wild birds: a literature review. *Anais da Academia Brasileira de Ciências*, SciELO Brasil, v. 89, p. 145–154, 2017. Cited on page 7.

MULIM, H. A. *et al.* Characterization of runs of homozygosity, heterozygosity-enriched regions, and population structure in cattle populations selected for different breeding goals. *BMC genomics*, Springer, v. 23, n. 1, p. 209, 2022. Cited 9 on page 1, 2, 16, 18, 33, 34, 37, 44, and 45.

NAKAJIMA, M. *et al.* Claudin-1 binder enhances epidermal permeability in a human keratinocyte model. *Journal of Pharmacology and Experimental Therapeutics*, ASPET, v. 354, n. 3, p. 440–447, 2015. Cited on page 45.

NANDOLO, W. *et al.* Timing and extent of inbreeding in african goats. *Frontiers in Genetics*, Frontiers Media SA, v. 10, p. 537, 2019. Cited on page 43.

NOSRATI, M. *et al.* The pattern of runs of homozygosity and genomic inbreeding in world-wide sheep populations. *Genomics*, Elsevier, v. 113, n. 3, p. 1407–1415, 2021. Cited on page 16.

ONZIMA, R. B. *et al.* Genome-wide characterization of selection signatures and runs of homozygosity in ugandan goat breeds. *Frontiers in Genetics*, Frontiers Media SA, v. 9, p. 318, 2018. Cited on page 44.

PEMBERTON, T. J. *et al.* Genomic patterns of homozygosity in worldwide human populations. *The American Journal of Human Genetics*, Elsevier, v. 91, n. 2, p. 275–292, 2012. Cited 3 on page 6, 10, and 34.

PERIASAMY, K. *et al.* Mapping molecular diversity of indigenous goat genetic resources of asia. *Small Ruminant Research*, Elsevier, v. 148, p. 2–10, 2017. Cited 3 on page 1, 15, and 33.

PERIPOLLI, E. *et al.* Runs of homozygosity: current knowledge and applications in livestock. *Animal genetics*, Wiley Online Library, v. 48, n. 3, p. 255–271, 2017. Cited on page 13.

PERIPOLLI, E. *et al.* Assessment of runs of homozygosity islands and estimates of genomic inbreeding in gyr (bos indicus) dairy cattle. *BMC genomics*, Springer, v. 19, p. 1–13, 2018. Cited on page 37.

PERSICHILLI, C. *et al.* Genome-wide diversity of pagliarola sheep residual population and its conservation implication. *Italian Journal of Animal Science*, Taylor & Francis, v. 20, n. 1, p. 1695–1705, 2021. Cited on page 1.

PURCELL, S. *et al.* Plink: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, Elsevier, v. 81, n. 3, p. 559–575, 2007. Cited 2 on page 17 and 18.

PURFIELD, D. C. *et al.* Runs of homozygosity and population history in cattle. *BMC genetics*, Springer, v. 13, p. 1–11, 2012. Cited 2 on page 19 and 20.

PURFIELD, D. C. *et al.* The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PloS one*, Public Library of Science San Francisco, CA USA, v. 12, n. 5, p. e0176780, 2017. Cited 4 on page 6, 20, 33, and 43.

QIAO, H. *et al.* Using the kde method to model ecological niches: A response to blonder et al.(2017). *Global Ecology and Biogeography*, John Wiley & Sons Ltd, v. 26, n. 9, 2017. Cited on page 5.

RAMOS-ONSINS, S. E. *et al.* Mining the pig genome to investigate the domestication process. *Heredity*, Nature Publishing Group, v. 113, n. 6, p. 471–484, 2014. Cited on page 36.

SANTOS, W. Bizarria dos *et al.* Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the mangalarga marchador horse breed. *Journal of Animal Breeding and Genetics*, Wiley Online Library, v. 138, n. 2, p. 161–173, 2021. Cited on page 45.

SARAVANAN, K. *et al.* Genome-wide analysis of genetic diversity and selection signatures in three indian sheep breeds. *Livestock Science*, Elsevier, v. 243, p. 104367, 2021. Cited on page 36.

SARAVANAN, K. *et al.* Genomic scans for selection signatures revealed candidate genes for adaptation and production traits in a variety of cattle breeds. *Genomics*, Elsevier, v. 113, n. 3, p. 955–963, 2021. Cited on page 37.

SARAVANAN, K. *et al.* Genome-wide assessment of genetic diversity, linkage disequilibrium and haplotype block structure in tharparkar cattle breed of india. *Animal biotechnology*, Taylor & Francis, v. 33, n. 2, p. 297–311, 2022. Cited on page 36.

SCHNEIDER, M. V.; ORCHARD, S. Omics technologies, data and bioinformatics principles. *Bioinformatics for Omics Data: Methods and Protocols*, Springer, p. 3–30, 2011. Cited on page 16.

SCOTT, E. M. *et al.* Characterization of greater middle eastern genetic variation for enhanced disease gene discovery. *Nature genetics*, Nature Publishing Group US New York, v. 48, n. 9, p. 1071–1076, 2016. Cited on page 15.

SELLI, A. *et al.* Detection and visualization of heterozygosity-rich regions and runs of homozygosity in worldwide sheep populations. *Animals*, MDPI, v. 11, n. 9, p. 2696, 2021. Cited 6 on page 15, 19, 20, 34, 35, and 36.

SILVA, A. A. *et al.* Gwas and gene networks for milk-related traits from test-day multiple lactations in portuguese holstein cattle. *Journal of Applied Genetics*, Springer, v. 61, n. 3, p. 465–476, 2020. Cited on page 45.

SUDHAKAR, D. V. *et al.* Exome sequencing and functional analyses revealed cetn1 variants leads to impaired cell division and male fertility. *Human Molecular Genetics*, Oxford University Press, v. 32, n. 4, p. 533–542, 2023. Cited on page 45.

TAN, M. P. *et al.* Applications of next-generation sequencing technologies and computational tools in molecular evolution and aquatic animals conservation studies: A short review. *Evolutionary Bioinformatics*, SAGE Publications Sage UK: London, England, v. 15, p. 1176934319892284, 2019. Cited on page 5.

TOLONE, M. *et al.* Genomic tools for the characterization of local animal genetic resources: application in mascaruna goat. *Animals*, MDPI, v. 12, n. 20, p. 2840, 2022. Cited on page 1.

TOMINAGA, K. *et al.* Il-12 synergizes with il-18 or il-1β for ifn-γ production from human t cells. *International immunology*, Oxford University Press, v. 12, n. 2, p. 151–160, 2000. Cited on page 45.

TONG, S.-f. *et al.* Genome-wide detection for runs of homozygosity analysis in three pig breeds from chinese taihu basin and landrace pigs by slaf-seq data. *Journal of Integrative Agriculture*, Elsevier, v. 21, n. 11, p. 3293–3301, 2022. Cited on page 36.

VALLEE, R. B.; VARMA, D.; DUJARDIN, D. L. Zw10 function in mitotic checkpoint control, dynein targeting, and membrane trafficking: is dynein the unifying theme? *Cell Cycle*, Taylor & Francis, v. 5, n. 21, p. 2447–2451, 2006. Cited on page 45.

VISSER, C. *et al.* Pedigree and genome-based patterns of homozygosity in the south african ayrshire, holstein, and jersey breeds. *Frontiers in Genetics*, Frontiers, v. 14, p. 1136078, 2023. Cited 2 on page 18 and 43.

VOSTRY, L. *et al.* Genomic diversity and population structure of the czech holstein cattle. *Livestock science*, Elsevier, v. 273, p. 105261, 2023. Cited on page 18.

WANG, L.-S. *et al.* Population-based study of genetic variation in individuals with autism spectrum disorders from croatia. *BMC Medical Genetics*, Springer, v. 11, p. 1–19, 2010. Cited on page 5.

WANG, Q. *et al.* Whole-genome resequencing of chinese pangolins reveals a population structure and provides insights into their conservation. *Communications Biology*, Nature Publishing Group UK London, v. 5, n. 1, p. 821, 2022. Cited on page 37.

WANG, X. *et al.* Genetic diversity, population structure and phylogenetic relationships of three indigenous pig breeds from jiangxi province, china, in a worldwide panel of pigs. *Animal genetics*, Wiley Online Library, v. 49, n. 4, p. 275–283, 2018. Cited 2 on page 5 and 15.

WIJKSTROM-FREI, C. *et al.* Lactoperoxidase and human airway host defense. *American journal of respiratory cell and molecular biology*, American Thoracic Society, v. 29, n. 2, p. 206–212, 2003. Cited on page 46.

WILLIAMS, J. L. *et al.* Inbreeding and purging at the genomic level: the chillingham cattle reveal extensive, non-random snp heterozygosity. *Animal genetics*, Wiley Online Library, v. 47, n. 1, p. 19–27, 2016. Cited 3 on page 17, 18, and 45.

XU, Z. *et al.* Assessment of autozygosity derived from runs of homozygosity in jinhua pigs disclosed by sequencing data. *Frontiers in Genetics*, Frontiers Media SA, v. 10, p. 274, 2019. Cited 2 on page 2 and 34.

ZAVAREZ, L. B. *et al.* Assessment of autozygosity in nellore cows (bos indicus) through high-density snp genotypes. *Frontiers in genetics*, Frontiers Media SA, v. 6, p. 5, 2015. Cited 2 on page 1 and 33.

ZHANG, F.; YU, X. Wac, a functional partner of rnf20/40, regulates histone h2b ubiquitination and gene transcription. *Molecular cell*, Elsevier, v. 41, n. 4, p. 384–397, 2011. Cited on page 46.

ZHANG, L. *et al.* Population genetic analysis of six chinese indigenous pig meta-populations based on geographically isolated regions. *Animals*, MDPI, v. 13, n. 8, p. 1396, 2023. Cited 3 on page 16, 18, and 36.

4 CONCLUDING REMARKS

This work aimed to analyze autozygosity and runs of homozygosity (ROH) in sheep and goat populations in Brazil, focusing on the genetic implications of inbreeding and genetic diversity related to increased adaptation and fitness. In the first study, we conducted a bibliometric analysis of 406 publications on ROH over the past decade, revealing a growing interest and significant advancements in the field, driven by computational tools like PLINK and detectRUNS.

Countries such as China, Italy, and the United States stood out in contributions, reflecting the importance of international collaboration and technological development in genomics. Persistent challenges like genetic diversity loss and habitat fragmentation in domesticated species highlight the need to refine analytical techniques and explore regions of high heterozygosity to better understand ROH patterns and their evolutionary implications.

In the second study, we utilized a 50K SNP genotype dataset from two Brazilian goat breeds, Anglo-Nubian and Marota, to analyze the abundance and length of ROH. We identified 22,872 ROH, with an average ranging from 74.73 (Anglo-Nubian) to 173.85 (Marota) per individual. Most ROH were short (<2.0 Mb), suggesting older inbreeding events. ROH-based inbreeding (FROH) was low in the Anglo-Nubian breed (0.0627) and high in the Marota breed (0.1419), reflecting a reduction in the effective population size of Marota over generations. We identified regions of high heterozygosity and genes related to various traits, such as embryonic development, body growth, lipid homeostasis, and brain functions, indicating that these regions are under selective pressure due to adaptation and selection for different purposes in the studied breeds.

These studies converge on the importance of understanding the genetic structure and variability within populations. Identifying regions of high heterozygosity and detailed ROH analysis provide valuable insights into the adaptation history and evolutionary processes shaping sheep and goat populations in Brazil. The findings advance scientific knowledge and have practical implications for breeding programs and conservation, helping to mitigate the negative inbreeding effects and promoting the health and genetic diversity.

In conclusion, this work provides a robust foundation for future research in evolutionary genetics, emphasizing the importance of international collaboration and advanced technologies in addressing complex issues of genetic diversity and adaptation. It is expected that the concepts and methodologies developed here can be applied in other contexts, significantly contributing to the evolution of the field and the preservation of sheep and goat populations.

Appendix

APPENDIX A – R script

```
library(openxlsx)
library(detectRUNS)
library(tidyr)
library(ggplot2)
library(dplyr)
library(reshape2)
library(Rcpp)
library(gridExtra)
library(data.table)
library(plyr)
library(tidyverse)

slidingRuns_hom <- slidingRUNS.run(
genotypeFile = 'plink.ped',
mapFile = 'plink.map',
windowSize = 20,
minSNP = 20,
maxOppWindow = 1,
maxMissWindow = 1,
maxGap = 250000, #bps
minLengthBps = 1000000,#bps
minDensity = 1/70, # SNP/kbps
threshold = 0.05,
ROHet = FALSE,
maxOppRun = NULL,
maxMissRun = NULL
)
summaryList <- summaryRuns(
runs = slidingRuns_hom, mapFile = 'plink.map',
genotypeFile = 'plink.ped',
Class = 2, snpInRuns = TRUE)

slidingRuns_hom %>%
dplyr::summarise(N =length(lengthBps),
media=mean(lengthBps/(summaryList$result_Froh_genome_wide$sum/
summaryList$result_Froh_genome_wide$Froh_genome)[1])*100,
desvio=sd(lengthBps/(summaryList$result_Froh_genome_wide$sum/
summaryList$result_Froh_genome_wide$Froh_genome)[1])*100,
cv = desvio/media)


slidingRuns_hom %>%
dplyr::group_by(group) %>%
dplyr::summarise(N =length(lengthBps),
media=mean(lengthBps/(summaryList$result_Froh_genome_wide$sum/
summaryList$result_Froh_genome_wide$Froh_genome)[1])*100,
desvio=sd(lengthBps/
(summaryList$result_Froh_genome_wide$sum/
summaryList$result_Froh_genome_wide$Froh_genome)[1])*100,
cv = desvio/media)
```

```
tab1 <- slidingRuns_hom %>%
dplyr::group_by(group) %>%
dplyr::summarise(N =length(lengthBps),
valor_length = gsub(".",",",
x=paste0(round(mean(lengthBps/10^6),2)," + ",
round(sd(lengthBps/10^6),2)),fixed = T),
minimo_length=round(min(lengthBps/10^6),2),
maximo_length=round(max(lengthBps/10^6),2))
tab2 <- slidingRuns_hom %>%
dplyr::group_by(group,id) %>%
dplyr::summarise(lengthBps =length(lengthBps)) %>%
dplyr::group_by(group) %>%
dplyr::summarise(valor_number = gsub(".",",",
x=paste0(round(mean(lengthBps),2)," + ",
round(sd(lengthBps),2)),fixed = T),
minimo_number=round(min(lengthBps),2),
maximo_number=round(max(lengthBps),2))
tab_final <- tab1 %>%
left_join(tab2)
getwd()
write.xlsx(tab_final,"estatistica_roh.xlsx")
#### ILHAS DE HOMOZIGOSIDADE, PERCENTIS
x <- summaryList$SNPinRun %>%
as.data.table()
x <- x[,top99 :=quantile(COUNT,.99), by=BREED]
x_select <- x %>%
filter(top99<COUNT)
x <- x %>%
dplyr::group_by(CHR,BREED) %>%
dplyr::summarise(x=sum(top99<COUNT)) %>%
filter(x>30) %>%
spread(key="BREED",value="x",fill = 0)
write.xlsx(x,"top_30.xlsx")
write.xlsx(x_select,"homozigose_99.xlsx")
library(base)
## calculating the average runs of homozygosity per chromosome by breed
summaryList$summary_ROH_mean_chr
media_RoH_cromo <-slidingRuns_hom %>%
dplyr::group_by(chrom=as.numeric(chrom),group) %>%
dplyr::summarise(N=length(id))
ggplot(data=media_RoH_cromo) +
geom_bar(aes(x=chrom,fill=group,y=N),
stat = "identity",position = "dodge")
## Calculating the average number of runs per chromosome
summaryList$summary_ROH_mean_chr
media_RoH_cromo <-slidingRuns_hom %>%
dplyr::group_by(chrom=as.numeric(chrom)) %>%
dplyr::summarise(N=length(id))
media_RoH_cromo$prop <- media_RoH_cromo$N/sum(media_RoH_cromo$N)
ylim.prim <- range(media_RoH_cromo$N)
ylim.sec <- range(media_RoH_cromo$prop)
b <- diff(ylim.prim)/diff(ylim.sec)
a <- ylim.prim[1] - b*ylim.sec[1]

tiff("figuras/Figure 2.tiff",width=25,height=15,res=300,units="cm")
ggplot(data=media_RoH_cromo) +
geom_bar(aes(x=chrom,y=N,
```

```
fill=chrom),stat = "identity",position = "dodge") +
geom_line(aes(x=chrom,y=N),col="red",size=1, show.legend = FALSE) +
xlab("Chromosome") +
ylab("Number of RoH")+
scale_x_continuous(
breaks = 1:29)+
scale_color_gradient2(midpoint=15, low="lightblue", mid="blue",
high="darkblue" )+
theme(legend.position = "none")+
scale_y_continuous(sec.axis = sec_axis(~ (. - a)/b,
name = "Percentage %",
labels = scales::label_percent()))
dev.off()

## Statistics - NUMBER AND LENGHT ROH FOR BREED
library(dplyr)
library(plyr)
detach(package:plyr)
number=slidingRuns_hom %>% group_by(group,id) %>% tally() %>%
select(-one_of(setdiff(names(slidingRuns_hom), "group"))) %>%
summarise(mean = mean(n), sd = sd(n), min = min(n), max = max(n))
write.table(number,'number.txt')
length=slidingRuns_hom %>% group_by(group) %>%
summarise(mean = mean(lengthBps), sd = sd(lengthBps),
min = min(lengthBps), max = max(lengthBps))
write.table(length,'length.txt')

summaryList$summary_ROH_count_chr
summaryList$result_Froh_chromosome_wide
res <- summaryList$summary_ROH_count
write.table(summaryList$summary_ROH_count,'resumo_F_ROH.txt1')
res$cat <- rownames(res)
res <-  res %>% gather(key = "raca",value = "ncorrida",-cat)

# Mean of FROH by race
froh <- Froh_inbreeding(runs = slidingRuns_hom, mapFile = 'plink.map')
write.table(froh,'froh.txt')
inbreeding_statistics <- froh %>%
dplyr::group_by(group) %>%
dplyr::summarise(N = length(Froh_genome),
mean=paste0(round(mean(Froh_genome)*100,2),"%"),
std=paste0(round(sd(Froh_genome)*100,2),"%"),
cv = paste0(round(sd(Froh_genome)/
mean(Froh_genome)*100,2),"%"),
percent_ind_inf_10percent =
paste0(round(mean(Froh_genome<0.10)*100,2),"%"),
percent_ind_inf_15percent =
paste0(round(mean(Froh_genome<0.15)*100,2),"%"))
write.xlsx(inbreeding_statistics,"inbreeding_statistics_homo.xlsx")


dados_grafico1 <- summaryList$summary_ROH_count_chr
dados_grafico2 <- summaryList$summary_ROH_percentage_chr
dados_grafico1$chr <- rownames(dados_grafico1)
dados_grafico <- dados_grafico1 %>%
left_join(dados_grafico2,by=c("chr"="chrom"))
dados_grafico_t  <- dados_grafico %>%
```

```
gather(key = "variavel",value="valor",-chr) %>%
mutate(raca=gsub(pattern= "(.*)\\.\\w",replacement="\\1",variavel),
tipo=gsub(pattern="(.*)\\.(\\w)",replacement="\\2",variavel),
chr = as.numeric(chr)) %>%
select(-variavel) %>%
spread(key="tipo",value="valor")
scaleFactor <-dados_grafico_t %>%
dplyr::group_by(raca) %>%
dplyr::summarise(mx = max(x) ,
my = max(y)) %>%
mutate(scl = mx/my)
a <-summaryList$result_Froh_genome_wide
table(a$sum/a$Froh_genome)
2464792553
dados_grafico_t$tamanho <- dados_grafico_t$x/dados_grafico_t$y

# write.table (summaryList,'resumo geral.txt')
summaryList$result_Froh_genome_wide
res <- summaryList$summary_ROH_count
write.table (summaryList$summary_ROH_count,'resumo_F_ROH.txt1')
res$cat <- rownames(res)
res <-  res %>% gather(key = "raca",value = "ncorrida",-cat)
# write.table (summaryList$summary_ROH_mean_chr,'resumo_F_ROH.txt')
# # res$cat <- rownames(res)
# res <-  res %>% gather(key = "raca",value = "ncorrida",-chrom)

a <- summaryList$summary_ROH_mean_chr
a <- a %>% gather(-chrom,key="raca",value="media")
library(ggplot2)
gplot(a) +
geom_density(aes(x=media,group=raca,col=raca))

library(tidyverse)
tiff("Figure3_Class.tiff",width=20,height=10,res=300,units="cm")
ggplot(data=res) +
geom_bar(aes(x = fct_inorder(cat),weight =ncorrida,fill=raca),
position = "dodge") +
labs(fill="Breed")+
xlab("ROH length category")+
ylab("Frequency")
dev.off()

tiff("Figure 4a.tiff",width=25,height=15,res=300,units="cm")
Figure3<-plot_manhattanRuns(
runs =slidingRuns_hom[slidingRuns_hom$group=="Marota",],
genotypeFile = 'plink.ped', mapFile = 'plink.map',
plotTitle = "MANHATHANN PLOT-% SNP in Runs",)
dev.off()
tiff("Figure 4b.tiff",width=25,height=15,res=300,units="cm")
Figure3<-plot_manhattanRuns(
runs =slidingRuns_hom[slidingRuns_hom$group=="Anglonubiana",],
genotypeFile = 'plink.ped', mapFile = 'plink.map',
plotTitle = "MANHATHANN PLOT-% SNP in Runs",)
dev.off()

## boxplot
tiff("Figure 5 - BOXPLOT.tiff",width=25,height=15,res=300,units="cm")
```

```
ggplot(froh) +
geom_boxplot(aes(x=group,y=Froh_genome, fill=group))+
labs(fill="Breed")+
xlab("") + ylab("FROH genome") +
scale_y_continuous(labels = scales::percent)
dev.off()
tiff("Figure 6.tiff",width=25,height=15,res=300,units="cm")
plot_InbreedingChr(runs = slidingRuns_hom, mapFile = 'plink.map', style='All',
savePlots = TRUE,outputName = "inbreeding")
dev.off()
tiff("Figure 7.tiff",width=25,height=15,res=300,units="cm")
plot_InbreedingChr(runs = slidingRuns_hom, mapFile = 'plink.map', style='All',
savePlots = TRUE,outputName = "inbreeding")
dev.off()

##savedRunFile<- system.file("extdata","diniz2023_goat-subset.sliding.csv",
package = "detectRUNS")
runs<-readExternalRuns(inputFile = saveddeRunFile,
program = "detectRuns")
head(runs)
# gerando manual
a11 <-  slidingRuns_hom[slidingRuns_hom$chrom==11,]
ggplot(a11) +
geom_line(aes(x=((from+to)/2)/1000000,y=nSNP,col=group))
tiff("Figure 7a (chromo 11).tiff",width=25,
height=15,res=300,units="cm")
plot_SnpsInRuns(
runs = slidingRuns_hom[slidingRuns_hom$chrom==11,],
genotypeFile = 'plink.ped',
mapFile = 'plink.map')
dev.off()
tiff("Figure 7b (chromo 15).tiff",width=25,height=15,res=300,units="cm")
plotSnpsInRuns(
runs = slidingRuns_hom[slidingRuns_hom$chrom==15,],
genotypeFile = 'plink.ped',
mapFile = 'plink.map')
dev.off()
## Identify top RUNS
topRuns <- tableRuns(
runs =  slidingRuns_hom,
genotypeFile = 'plink.ped',
mapFile = 'plink.map',
threshold = 0.50)
write.table (topRuns,'topRuns.txt')
res$cat <- rownames(res)
res <-  res %>% gather(key = "raca",
value = "ncorrida",-cat)
consecutiveRuns_het <- consecutiveRUNS.run(
genotypeFile ='plink.ped',
mapFile = 'plink.map',
minSNP = 10,
ROHet = TRUE,
maxGap = 10^6,
minLengthBps = 1000000,
maxOppRun = 5,
maxMissRun = 5
)
```

```
summaryList <- summaryRuns(
runs = consecutive_het, mapFile = 'plink.map',
genotypeFile = 'plink.ped',
Class = 2, snpInRuns = TRUE)
tiff("Figure 18.tiff",width=25,height=15,res=300,units="cm")
plot_manhattanRuns(
runs = consecutiveRuns_het[consecutiveRuns_het$group=="Marota",],
genotypeFile = 'plink.ped', mapFile = 'plink.map',
plotTitle = "Consecutive HETERO")
dev.off()

hist(froh$Froh_genome,main=NULL,
xlab="inbreeding coefficients of each individual sample")
dev.off()
summaryList <- summaryRuns(
runs = consecutiveRuns_het, mapFile = 'plink.map',
genotypeFile = 'plink.ped',
Class = 2, snpInRuns = TRUE)
res <- summaryList$summary_ROH_count
res$cat <- rownames(res)
res <-  res %>% gather(key = "raca",value = "ncorrida",-cat)
tiff("Figure 19.tiff",width=25,height=15,res=300,units="cm")
ggplot(data=res) +
geom_bar(aes(x = cat,weight =ncorrida,fill=raca),
position = "dodge") +
labs(fill="Breed")+
xlab("ROH length category")+
ylab("Frequency")
dev.off()
summaryList$summary_het_count
summaryList$summary_het_mean_chr
plot_Runs(runs = consecutiveRuns_het,
savePlots = TRUE,outputName = "x")
plot_SnpsInRuns(
runs = slidingRuns_het[slidingRuns_hom$chrom==7,],
genotypeFile = 'plink.ped',
mapFile = 'plink.map')
plot_manhattanRuns(
runs = slidingRuns_het[slidingRuns_het$group=="MarotaEMBR",],
genotypeFile = 'plink.ped',
mapFile = 'plink.map',plotTitle = "TITUTLO")
plot_InbreedingChr(runs = consecutiveRuns_het,
mapFile = 'plink.map', style='All',
savePlots = TRUE,outputName = "inbreeding")
topRuns <- tableRuns(
runs =  consecutiveRuns_het, genotypeFile = 'plink.ped',
mapFile = 'plink.map',
threshold = 0.5)

##CONSECUTIVERUNS HETEROZIGOSE - used 17.278
consecutiveRuns_het <- consecutiveRUNS.run(
  genotypeFile ='plink.ped',
  mapFile = 'plink.map',
  minSNP = 10,
  ROHet = TRUE,
  maxGap = 10^6,
  minLengthBps = 1000000,
```

```
  maxOppRun = 5,
  maxMissRun = 5
)
summaryList <- summaryRuns(
  runs = consecutiveRuns_het, mapFile = 'plink.map', genotypeFile = 'plink.ped',
  Class = 2, snpInRuns = TRUE)
consecutiveRuns_het <- consecutiveRuns_het %>% as.data.table()

## Identify top RUNS
topRuns <- tableRuns(
  runs =  consecutiveRuns_het,
  genotypeFile = 'plink.ped',
  mapFile = 'plink.map',
  threshold = 0.45)


## calculating the average runs of homozygosity per chromosome by breed
summaryList$summary_HET_mean_chr
media_HET_cromo <-consecutiveRuns_het %>%
  dplyr::group_by(chrom=as.numeric(chrom),group) %>%
  dplyr::summarise(N=length(id))

ggplot(data=media_HET_cromo) +
  geom_bar(aes(x=chrom, fill=group,y=N),stat = "identity",position = "dodge")


## Calculating the average number of runs per chromosome
summaryList$summary_HET_mean_chr
media_HET_cromo <-consecutiveRuns_het %>%
  dplyr::group_by(chrom=as.numeric(chrom)) %>%
  dplyr::summarise(N=length(id))

media_HET_cromo$prop <- media_HET_cromo$N/sum(media_HET_cromo$N)

ylim.prim <- range(media_HET_cromo$N)
ylim.sec <- range(media_HET_cromo$prop)

b <- diff(ylim.prim)/diff(ylim.sec)
a <- ylim.prim[1] - b*ylim.sec[1]
number=consecutiveRuns_het %>% group_by(group,id) %>% tally() %>%
select(-one_of(setdiff(names(consecutiveRuns_het), "group"))) %>%
summarise(mean = mean(n), sd = sd(n), min = min(n), max = max(n))
write.table(number,'number.txt')
length=consecutiveRuns_het %>% group_by(group) %>%
 summarise(mean = mean(lengthBps), sd = sd(lengthBps),
 min = min(lengthBps), max = max(lengthBps))
write.table(length,'length.txt')


summaryList$summary_HET_count_chr
summaryList$result_Fhet_chromosome_wide
res <- summaryList$summary_HET_count
write.table(summaryList$summary_HET_count,'resumo_F_HET.txt1')
res$cat <- rownames(res)
res <-  res %>% gather(key = "raca",value = "ncorrida",-cat)

# Mean of FHET by race
```

```
fhet <- Fhet_inbreeding(runs = consecutiveRuns_het, mapFile = 'plink.map')
write.table(Fhet,'fhet.txt')
inbreeding_statistics <- fhet %>%
  dplyr::group_by(group) %>%
  dplyr::summarise( N = length(Fhet_genome),
mean=paste0(round(mean(Fhet_genome)*100,2),"%"),
std=paste0(round(sd(Fheth_genome)*100,2),"%"),
cv = paste0(round(sd(Froh_genome)/mean(Fhet_genome)*100,2),"%"),
percent_ind_inf_10percent = paste0(round(mean(Fhet_genome<0.10)*100,2),"%"),
percent_ind_inf_15percent = paste0(round(mean(Fhet_genome<0.15)*100,2),"%"))

write.xlsx(inbreeding_statistics,"inbreeding_statistics_het.xlsx")
```