Acta Universitatis Agriculturae Sueciae
Doctoral Thesis No. 2025:1

Exploring the Genetic Regulation of Ability and Quality of Gaits in Icelandic Horses

Heiðrún Sigurðardóttir

Landbúnaðarháskóli Íslands
Agricultural University of Iceland

SLU SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES

Exploring the Genetic Regulation of Ability and Quality of Gaits in Icelandic Horses

Heiðrún Sigurðardóttir

PhD thesis submitted in complete fulfilment of a
Doctor Philosophiae joint degree education in
Agricultural Science
Land use & cultivation
Faculty of Agricultural Sciences
Agricultural University of Iceland, Hvanneyri
and
Animal Science
Department of Animal Biosciences
Faculty of Veterinary Medicine and Animal Science
Swedish University of Agricultural Sciences, Uppsala

Main academic supervisors
Dr. Susanne Eriksson, Department of Animal Biosciences, SLU
Dr. Þorvaldur Kristjánsson, The Agricultural University of Iceland

Co-supervisors
Dr. Gabriella Lindgren, Department of Animal Biosciences, SLU
Dr. Elsa Albertsdóttir , Independent researcher
Dr. Marie Rhodin, Department of Animal Biosciences, SLU

Hvanneyri / Uppsala 2025

Acta Universitatis Agriculturae Sueciae

2025:1

Icelandic title: Erfðafræðilegur grunnur gangtegunda íslenska hestsins

English title: Exploring the Genetic Regulation of Ability and Quality of Gaits in Icelandic Horses

ISSN 1652-6880

ISBN (print version) 978-91-8046-436-9

ISBN (electronic version) 978-91-8046-486-4

ISBN (AUI) 978-9935-512-52-9

https://doi.org/10.54612/a.74nb3gs77n

Print: Háskólaprent, Reykjavík 2025

Cover artwork by Rán Flygenring

Agricultural Science

Land use & cultivation

Faculty of Agricultural Sciences

Agricultural University of Iceland, Hvanneyri

Animal Science

Department of Animal Biosciences

Faculty of Veterinary Medicine and Animal Science

Swedish University of Agricultural Sciences, Uppsala

The research presented in this doctoral thesis was conducted under the joint auspices of the Agricultural University of Icelandic and the Swedish University of Agricultural Sciences.

Clarification of contribution

I hereby declare that the writing of the following thesis and the four accompanying papers is my work, done under the supervision and with the assistance of Dr. Susanne Eriksson, Dr. Þorvaldur Kristjánsson, Dr. Gabriella Lindgren, Dr. Elsa Albertsdóttir and Dr. Marie Rhodin. The research was initiated by Dr. Susanne Eriksson and Dr. Gabriella Lindgren and was further developed in cooperation with Sigurðardóttir, Kristjánsson, Albertsdóttir and Rhodin.

The contribution of Heiðrún Sigurðardóttir to the papers included in this thesis was as follows:

Paper I Performed the experiments and data analysis together with co-first-author Rosengren. Interpreted the results together with the co-authors. Drafted the manuscript together with Rosengren and Solé, further revised by co-authors.

Paper II Collected half of the samples. Performed the experiments and data analysis. Interpreted the results together with the co-authors. Drafted the manuscript, further revised by co-authors. Responsible for correspondence with the scientific journal.

Paper III Selected the horses for the study together with Albertsdóttir and Kristjánsson. Organised and collected the blood samples along with certified veterinarians. Performed the bioinformatic analysis with assistance from Niazi. Interpreted the results together with the co-authors. Drafted the manuscript, further revised by co-authors. Responsible for correspondence with the scientific journal.

Paper IV Collected half of the samples. Performed the experiments and data analysis. Interpreted the results together with the co-authors. Drafted the manuscript, further revised by co-authors. Responsible for correspondence with the scientific journal.

_____

Heiðrún Sigurðardóttir

Abstract

Sigurðardóttir, H. (2025). *Exploring the Genetic Regulation of Ability and Quality of Gaits in Icelandic Horses*. Doctoral thesis.

The Icelandic horse is celebrated for its unique gaits and versatility as a riding horse, yet the genetic basis of these characteristics is not fully understood. Its diverse gait range, the availability of extensive phenotypic records, and the application of advanced genomic technologies make the Icelandic horse an invaluable model for exploring the genetic foundations of gait traits and performance.

This thesis aimed to identify novel genetic factors affecting gaits and performance in the Icelandic horse by integrating genomic approaches to analyse genetic variation through genome-wide patterns of homozygosity and polymorphic variation. The study utilized high-density SNP genotype data, generated using a 670k genotype array, from 380 horses to perform genome-wide association (GWA) studies and runs of homozygosity (ROH) analyses. Additionally, 39 horses were included for whole-genome sequencing (WGS) analysis.

Through GWA studies, quantitative trait loci (QTLs) affecting gait traits were identified. One QTL was associated with breeding scores for back and croup conformation, which significantly influenced the quality of the lateral gaits tölt and pace. Two additional QTLs were directly linked to breeding scores for pace. Haplotype analysis revealed two frequent haplotypes for each identified QTL that had significant effects on the associated trait. The haplotypes associated with back and croup harboured genes with known roles in muscle and skeletal development in humans. The haplotypes affecting pace scores were located within the candidate genes *STAU2* and *RELN* which are expressed in neural tissue. In addition to the observed effects on pace, the *STAU2* haplotypes were found to significantly influence the quality of trot and gallop, while the *RELN* haplotypes impacted the quality of tölt, trot, canter, and gallop, and may potentially also influence the trainability and precocity of young horses. Moreover, interactions between *STAU2*, *RELN*, and the previously identified *DMRT3* gene were observed, involving both additive and compensatory effects. Sequence analysis in these regions revealed a partial loss of the *STAU2* protein caused by a one-base-pair frameshift insertion. This variant was proposed to be causative for the observed effects of the *STAU2* haplotypes. In the *RELN* region, potential regulatory elements were identified and suggested as candidate variants influencing the effects of the *RELN* haplotypes.

The enduring significance of the genomic region on ECA23, harbouring the *DMRT3* gene, was confirmed through ROH analysis, while additional regions of potential relevance to performance traits were identified. Analyses of genomic inbreeding and genetic diversity indicated that recent breeding practices in the Icelandic horse breed have been sustainable, maintaining adequate genetic diversity.

In conclusion, this thesis expands current knowledge of the genetic basis of gait and performance in the Icelandic horse, by demonstrating that genetic factors beyond *DMRT3* contribute to the regulation of gait ability and quality. The findings warrant further investigations and may ultimately benefit the breeding of the Icelandic horse and assist breeders in making more informed breeding decisions.

Keywords: Equine genomics, Back and croup conformation, Genome-wide association, *RELN*, *STAU2*, *DMRT3*, Whole-genome sequencing, Regulatory elements, Frameshift mutation, Genetic diversity, Signatures of selection

Ágrip

Íslenski hesturinn er þekktur víða um heim sem fjölhæfur reiðhestur vegna hans einstöku ganghæfileika. Erfðafræðilegur grunnur þessara eiginleika er þó enn að miklu leyti óþekktur. Ganghæfileikar hans ásamt aðgengilegum og víðtækum svipfarsmælingum, gera íslenska hestinn að afar verðmætu viðfangi til rannsókna á þessu sviði.

Meginmarkmið þessarar rannsóknar var að bera kennsl á nýja erfðaþætti sem hafa áhrif á gangtegundir og hæfileika íslenska hestsins. Það var gert með því að samþætta aðferðir sem annars vegar kanna arfhrein svæði á erfðamenginu vegna úrvals, og hins vegar svæði á erfðamenginu sem einkennast af breytileika arfgerða í tengslum við ákveðna eiginleika. Í rannsókninni var notast við háþéttni SNP-arfgerðagreiningar, byggðar á 670k flögu, fyrir 380 hross með kynbótadóma, til að framkvæma víðtæka erfðamengisleit og greiningu á samfelldum röðum arfhreinna svæða. Þar að auki voru erfðamengi 39 hrossa heilraðgreind til frekari greininga.

Með víðtækri erfðamengisleit fundust áður óþekkt tengsl milli ganghæfileika og breytileika á ákveðnum svæðum á erfðamenginu (*e. quantitative trait locus,* QTL). Eitt QTL hafði tengsl við einkunnir fyrir bak og lend og hafði marktæk áhrif á gæði tölts og skeiðs. Tvö önnur QTL höfðu bein tengsl við einkunn fyrir skeið. Setraðagreining leiddi í ljós tvær algengar setraðir í öllum þessum QTL sem höfðu marktæk áhrif. Setraðirnar sem höfðu tengsl við einkunn fyrir bak og lend voru staðsettar innan gena sem hafa þekkt hlutverk í vöðva- og beinabyggingu í mannfólki. Setraðirnar sem höfðu áhrif á einkunnir fyrir skeið voru staðsettar í genunum *STAU2* og *RELN*, sem bæði eru tjáð í taugavef. Umfram tengslin við skeiðeinkunn þá höfðu setraðirnar í *STAU2*-geninu einnig marktæk áhrif á einkunnir fyrir brokk og greitt stökk. Að sama skapi höfðu setraðirnar í *RELN*-geninu marktæk áhrif á einkunnir fyrir tölt, brokk, hægt og greitt stökk, umfram skeiðeinkunn. Vísbendingar bentu einnig til að *RELN*-setraðirnar hefðu áhrif á bráðþroska og hversu fljót hross væru til í þjálfun. Þar að auki voru víxlverkunum milli *STAU2*, *RELN* og hins áður þekkta *DMRT3*-gens lýst í formi samleggjandi- og uppbótaáhrifa. Frekari greining með heilraðgreiningagögnum sýndi fram á að hliðrunarstökkbreyting (*e. frameshift mutation*) í *STAU2*-geninu, sem olli skertri próteinframleiðslu, væri líkleg orsök fyrir þeim áhrifum sem setraðirnar sýndu á skeið, brokk og greitt stökk. Sambærileg greining leiddi í ljós stýriþætti í *RELN*-geninu sem mögulega útskýra þau áhrif sem *RELN*-setraðirnar höfðu á skeið, tölt, brokk, hægt og greitt stökk.

Með greiningu á samfelldum röðum arfhreinna svæða var mikilvægi hluta litnings númer 23 staðfest gagnvart ganghæfni og getu íslenska hestsins, en þar er *DMRT3*-genið staðsett. Greiningin leiddi einnig í ljós áður óþekkt

gen með mögulegar tengingar við afköst og getu. Útreikningar á skyldleikarækt og erfðabreytileika sem byggðu á erfðagögnum gáfu til kynna að ræktun íslenska hestsins síðastliðna áratugi hafi verið sjálfbær þar sem tap erfðabreytileika var í lágmarki.

Rannsóknin eykur þekkingu á erfðafræðilegum grunni gangtegunda og sýnir fram á að erfðaþættir umfram *DMRT3*-genið hafa áhrif á ganghæfni íslenska hestsins. Niðurstöðurnar styðja við frekari rannsóknir á þessu sviði, ásamt því að nýtast ræktendum íslenska hestsins við að taka upplýstari ákvarðanir í ræktunarstarfinu.

Lykilorð: Erfðamengjagreiningar, Bak og lend, Víðtæk erfðamengisleit, *RELN*, *STAU2*, *DMRT3*, Heilraðgreining, Stýriþættir, Hliðrunarstökkbreyting, Erfðabreytileiki, Merki úrvals

Sammanfattning

Islandshästen är uppskattad för sina unika gångarter och mångsidighet som ridhäst, men den genetiska grunden för dessa egenskaper är inte helt klarlagd. Dess olika gångarter, tillgången till omfattande egenskapsregister och möjligheten att använda avancerad genomisk teknologi gör islandshästen till en ovärderlig modell för att utforska de genetiska grunderna för gångarter och prestation.

Denna avhandling syftade till att identifiera nya genetiska faktorer som påverkar gångarter och prestationsförmåga hos islandshästen genom att integrera olika genomiska metoder för att analysera genetisk variation: dels genomomfattande mönster av homozygositet och dels polymorf variation. Studien använde SNP-genotypdata med hög densitet, genererad med hjälp av en 670k-array, från 380 hästar för att utföra genomomfattande associationsstudier (genome wide association study, GWA) och studier av sammanhängande homozygota regioner (runs of homozygosity, ROH). Dessutom inkluderades 39 hästar för analys av helgenomsekvensering (whole genome sequencing, WGS).

Genom GWA-studier identifierades loci för kvantitativa egenskaper (quantitative trait loci, QTL) som påverkar gångegenskaper. En QTL var associerad med avelsbedömningspoäng för rygg- och korskonformation, vilket signifikant påverkade kvaliteten på de laterala gångarterna tölt och pass. Ytterligare två QTL:er var direkt associerade med avelsbedömningspoäng för pass. Haplotypanalys avslöjade två vanligt förekommande haplotyper för varje identifierad QTL som hade signifikanta effekter på den associerade egenskapen. Haplotyperna som var associerade med rygg och kors innehöll gener med kända roller i muskel- och skelettutveckling hos människor. Haplotyperna som påverkar passpoängen var lokaliserade inom kandidatgenerna *STAU2* och *RELN* som uttrycks i nervvävnad. Utöver de observerade effekterna på pass visade sig *STAU2*-haplotyperna signifikant påverka kvaliteten på trav och galopp, medan *RELN*-haplotyperna påverkade kvaliteten på tölt, trav, samlad galopp och galopp, och potentiellt också hur lätträdade och tidigt utvecklade unga hästar är. Dessutom observerades interaktioner mellan *STAU2*, *RELN* och den tidigare identifierade ”gångartsgenen” *DMRT3*, som involverade både additiva och kompensatoriska effekter. Sekvensanalys i dessa regioner avslöjade en partiell förlust av *STAU2*-proteinet orsakad av en så kallad ”frameshift insertion” med ett baspar. Denna variant föreslogs vara orsak till de observerade effekterna av *STAU2*-haplotyperna. I *RELN*-regionen identifierades potentiella regulatoriska element vilka föreslogs som kandidatvarianter bakom effekterna av *RELN*-haplotyperna.

Betydelsen av den genomiska regionen på ECA23, som inkluderar *DMRT3*-genen, bekräftades genom ROH-analys, medan ytterligare regioner av potentiell relevans för prestationsegenskaper identifierades. Analyser av genomisk inavel och genetisk mångfald indikerade att den senaste tidens avelsmetoder inom islandshästarsen har varit hållbara vad beträffar bevarande av genetisk mångfald.

Sammanfattningsvis bidrar denna avhandling med ytterligare kunskap om den genetiska grunden för gångarts- och prestationsförmåga hos islandshästen, genom att visa att genetiska faktorer bortom *DMRT3* bidrar till regleringen av gångartsförmåga och -kvalitet. Resultaten motiverar ytterligare undersökningar och kan i slutändan gynna aveln av islandshästen och hjälpa uppfödare att fatta mer välgrundade avelsbeslut.

Nyckelord: Hästgenomik, Rygg- och korskonformation, Helgenomstudie, *RELN*, *STAU2*, *DMRT3*, Helgenomsekvensering, Regulatoriska element, Frameshift-mutation, Genetisk variation, Spår av selektion

Preface

It always seems impossible until it's done

– Nelson Mandela

Dedication

To my dearest Hrafnhildur

Acknowledgements

The research presented in this thesis was conducted at the Faculty of Agricultural Sciences, Agricultural University of Iceland (AUI) and the Department of Animal Biosciences, Swedish University of Agricultural Sciences (SLU). This work would not have been possible without the invaluable contributions and support of many colleagues, friends, and family members.

First and foremost, I am endlessly grateful to my supervisors, **Susanne Eriksson, Þorvaldur Kristjánsson, Gabriella Lindgren**, **Elsa Albertsdóttir** and **Marie Rhodin**. **Susanne**, for being a remarkable mentor with seemingly endless hours in the day, for your constant availability and insightful advice, and for being a rock-solid support and a good friend through it all. Thank you for turning my wordy marathons into elegant sprints, proving that sometimes less is more, and for believing in me, even though my use of < and > symbols has made you question whether I skipped a crucial day in elementary math class! **Þorvaldur**, for sharing your endless knowledge of the Icelandic horse, for your ever-positive and supportive attitude, and for being a good friend. **Gabriella**, for your unwavering encouragement to aim higher, contagious energy, and enthusiasm that have inspired me throughout this journey. **Elsa**, for your friendship and longtime guidance from my undergraduate studies and up to this point. **Marie**, for sharing your expertise, insightful guidance, and positive encouragement along the way.

Special thanks to my collaborators, whose expertise significantly enriched this work: **Henrik Boije,** for sharing your knowledge of the nervous system and locomotor networks at the cellular level, **Michela Ablondi,** for your expertise in population genomics and invaluable insights, and **Adnan Niazi,** for helping advance this research through bioinformatic approaches.

I would also like to thank **Maria Rosengren** for your collaboration and support in the early stages of my studies and for sharing your data on Icelandic horses, and **Rakan Naboulsi** for your advice and assistance with bioinformatics. Furthermore, I would like to thank **Tytti Vanhala** for coming to my rescue during the pandemic cutting an endless amount of hair for me, and for teaching me the ropes in the lab, and offering great advice.

I am sincerely thankful to **Diana Ekman**, my supervisor at the NBIS Swedish Bioinformatics Advisory Program, for your expertise, support, understanding, and easy-going attitude.

A special acknowledgement to **Þorvaldur Árnason** for your support, interest in my work, and the fruitful discussions we shared – they meant a lot to me. My gratitude also extends to **Heimir Gunnarsson** and **SIF Avel**, for showing interest in my research and providing a platform to present my findings to Icelandic horse enthusiasts in Sweden. Furthermore, I would like to thank **Kristinn Hugason** and **Óðinn Örn Jóhannsson** for their support.

Many thanks to my colleagues at SLU and AUI – it has been a privilege to know all of you and to be part of such inspiring academic environments. Special thanks to my fellow PhD students; particularly **Paulina**, for your good company during courses and conferences. My gratitude also goes to **Ragnheiður**, **Gunnhildur**, **Þórunn Edda**, **Hlynur**, **Bjarni Diðrik** and **Álfheiður M.** for your administrative support and for ensuring a smooth journey through the joint education program. Thank you, **Åsa Gelinder Viklund**, for your interest in my work and for providing opportunities to present it to your students. I am also grateful to **Sofia Mikko**, for your inspiration and support as the director of graduate studies.

Special thanks to the **IT department at AUI** and **Valdimar** at Tölvuþjónustan for addressing my server-related requests swiftly and effectively. A sincere thank you to **Cano** for keeping my computer and data secure and for always having the friendliest attitude.

A heartfelt acknowledgement to all the **Icelandic horses and their owners** who contributed to this research – without you, this work would not have been possible – literally! Special thanks to **Hulda**, for embarking on a cross-Iceland road trip to collect blood samples – proving that no distance is too great when science calls.

Last but not least, I am endlessly grateful to my friends and family for their unwavering support and encouragement throughout this journey. To my dearest **Tjörvi**, for always being there and being my number one fan, for all the laughs and the endless encouragement. To my sweet **Hrafnhildur**, who showed me the meaning of strength and courage – you will forever be in my heart and mind. To my **mother and father**, for your love and always believing in me. To **Margrét Rós**, **Guðni Páll**, **Katla**, **Jökull**, **Embla**, **BÓKI and co**, thank you all for your love and support. To my besties, **Arna Huld**, **Áslaug**, **Freyja**, **Guðrún**, and **Lára**, for keeping me "relatively" sane. To my friends, **Vala** and **Gummi**, for taking such good care of my horses in Iceland while I was away, and to **Gísli** for all the delightfully nerdy horse discussions.

The work in this thesis was funded by grants from the Agricultural University of Iceland's Doctoral Fund, the Conservation Fund for the Icelandic Horse Breed, the Swedish Icelandic Horse Association (SIF Avel), Blikastaðir Fund, the Icelandic Genetic Resource Council in Agriculture (AGROGEN), the Swedish Research Council (VR) and the Swedish Foundation for Strategic Research.

Table of contents

List of original papers

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I Rosengren, M. K.*, **Sigurðardóttir, H.***, Eriksson, S., Naboulsi, R., Jouni, A., Novoa-Bravo, M., Albertsdóttir, E., Kristjánsson, Th., Rhodin, M., Viklund, Å., Velie, B. D., Negro, J. J., Solé, M. & Lindgren, G. (2021). A QTL for conformation of back and croup influences lateral gait quality in Icelandic horses. *BMC Genomics, 22*(1), 267. https://doi.org/10.1186/s12864-021-07454-z

II **Sigurðardóttir, H.**, Boije, H., Albertsdóttir, E., Kristjánsson, Th., Rhodin, M., Lindgren, G., & Eriksson, S. (2023). The genetics of gaits in Icelandic horses goes beyond *DMRT3*, with *RELN* and *STAU2* identified as two new candidate genes. *Genetics Selection Evolution, 55*(1), 89. https://doi.org/10.1186/s12711-023-00863-6

III **Sigurðardóttir, H.**, Eriksson, S., Niazi, A., Rhodin, M., Albertsdóttir, E., Kristjánsson, Th., & Lindgren, G. (2025). Genetic influence of a *STAU2* frameshift mutation and *RELN* regulatory elements on performance in Icelandic horses. (submitted)

IV **Sigurðardóttir, H.**, Ablondi, M., Kristjánsson, Th., Lindgren, G., & Eriksson, S. (2024). Genetic diversity and signatures of selection in Icelandic horses and Exmoor ponies. *BMC Genomics, 25*(1), 772. https://doi.org/10.1186/s12864-024-10682-8

Papers I, II and IV are open-access and reproduced with the permission of the publishers.

*Shared first authorship

List of tables

List of figures

1. Introduction

The Icelandic horse breed is renowned for its unique gaits and performance as a riding horse and is characterized by a distinct set of traits. Yet, within the breed's overall uniformity lies a remarkable diversity of phenotypes, with significant variation in each trait among the horses. What genetic factors contribute to the breed's uniformity and its phenotypic variation? To uncover the genetic uniqueness of the Icelandic horse, we must investigate both its defining features and the genetic mechanisms driving its diversity.

1.1 Origin of gaited horses

A gait refers to the way a horse moves its legs while walking or running. In that sense, all horses exhibit gaits. However, no other species demonstrates the same degree of innate variation in locomotion traits as horses (Barrey, 2013). As a result, equine gaits are classified in more detail based on differences in speed and footfall patterns.

All horse breeds are capable of performing the walk, trot, canter, and gallop, which are considered the basic gaits. The walk is the slowest gait, featuring an even four-beat, symmetrical pattern with no suspension phase, meaning two or three hooves are always on the ground (Barrey, 2013; Clayton, 1995). The trot is an intermediate-speed gait, though it can vary significantly in speed. It is a two-beat, diagonally symmetrical gait that includes a suspension phase where all hooves are off the ground simultaneously (Barrey, 2013; Clayton, 1994b). Both canter and gallop are asymmetrical gaits that also include a suspension phase. The canter is a three-beat gait with a medium speed, while the gallop is a four-beat, high-speed gait (Barrey, 2013; Clayton, 1994a; Deuel & Lawrence, 1986). Horses that are restricted to these basic gaits are commonly classified as 'non-gaited' horses.

Other gaits, such as pace and various four-beat ambling gaits, including tölt, running walk, foxtrot and rack, are collectively referred to as alternate gaits. Pace is a laterally symmetrical gait, commonly performed at high speeds, and is characterized by a suspension phase where all four feet are off the ground simultaneously (Barrey, 2013; Wilson et al., 1988). Four-beat ambling gaits vary in speed and are defined by their symmetry, ranging from diagonal to lateral movement. Unlike the pace, these gaits have no suspension phase, meaning one or two hooves are always in contact with the ground (Barrey, 2013; Nicodemus & Clayton, 2003; Zips et al., 2001). Horses that can perform any of the alternate gaits are classified as 'gaited' horses.

The genetic basis for gait diversity in horses is not yet fully understood, but a major breakthrough came with the discovery of a key mutation that distinguishes gaited from non-gaited horses, paving the way for further genetic research on equine gaits. This mutation, known as the 'gait keeper' mutation (DMRT3_Ser301STOP, chr23: g.22,391,254C>A), is a single-base nonsense mutation in the *doublesex and*

mab-3 related transcription factor 3 (*DMRT3*) gene (Andersson et al., 2012). It results in a premature stop codon, leading to a truncated *DMRT3* protein, which alters the horse's locomotion pattern and enables development of pace (Andersson et al., 2012). This mutation is, therefore, common in gaited breeds, while most non-gaited breeds retain the wild-type allele (Promerová et al., 2014). Also in trotting breeds used for harness racing, the frequency of the 'gait keeper' mutation is relatively high (Andersson et al., 2012; Jäderkvist Fegraeus et al., 2015; Jäderkvist, Andersson, et al., 2014; Promerová et al., 2014; Ricard, 2015). It is suggested that the mutation's secondary effect extends the speed range of symmetrical gaits like trot and pace, preventing a switch to asymmetrical gaits like canter or gallop. Evidence also indicates that the mutation may negatively impact the quality of trot and canter in some breeds (Jäderkvist et al., 2015; Kristjansson et al., 2014).

Given the significant impact of the 'gait keeper' mutation on gaiting ability in horses, it is likely that the emergence of this mutation marks the origin of gaited horses. A study by Staiger et al. (2017) suggested that the mutation likely arose soon after horse domestication (5-6,000 years ago) and then spread through human activity and horse trade. However, definitive evidence pinpointing the geographic origin of the 'gait keeper' mutation has not yet been established, though several hypotheses have been proposed. Staiger et al. (2017) suggested that the mutation was likely under strong positive selection from the start due to its beneficial effects on speed and rider comfort. This may have facilitated its rapid spread across regions, making it difficult to pin down its exact geographic origin. In contrast, Wutke et al. (2016) hypothesized that the mutation originated in the British Isles during the medieval period (850–900 AD) before spreading to Iceland and eventually continental Europe. However, this hypothesis has not gained widespread acceptance due to limited sampling. Furthermore, historical records suggest the existence of gaited Asturian horses in Northern Spain approximately 2,000 years ago (Hendricks, 2007), challenging the proposed timeline and location of the mutation's origin. The geographic origin of the 'gait keeper' mutation remains a mystery, but it is evident that the mutation eventually reached Iceland, where it became well-established and contributed to the gait diversity that defines the Icelandic horse.

1.2 Origin and evolution of the Icelandic horse breed

Little is known for certain about the origin of the Icelandic horse. However, it is generally believed that the breed descends from a selection of horses brought to Iceland by Norse settlers around 1100 years ago (Adalsteinsson, 1981). This belief is somewhat supported by molecular genetic studies, which have shown a moderate genetic relationship between the Icelandic horse and Scandinavian breeds such as the Norwegian Fjord horse, the Nordland horse, the North Swedish horse, Coldblooded trotters, and the Finnhorse, but also with the Shetland pony, Miniature horse, and the Mongolian horse (Bjørnstad et al., 2003; Bjørnstad & Røed, 2001; McCue et al., 2012; Petersen, Mickelson, Cothran, et al., 2013).

Figure 1. The Icelandic horse was used for transport and labour prior to Iceland's industrialization.The photo on the left shows a caravan of horses transporting bales of hay home from the fields (Árnesinga Municipal Archives (Héraðsskjalasafn Árnesinga), photo no. 2007 33 GÓ 03209, photographer unknown). The photo on the right depicts another caravan of horses delivering mail, led by the postman Hans Hannesson (Photo by Magnús Ólafsson).

The exact number of horses originally brought to Iceland is unknown, but it is widely believed that the breed emerged from a relatively small gene pool of horses carefully selected for specific purposes (Björnsson & Sveinsson, 2006, pp. 18-33). Early in the breed's development, it is likely that some unfavourable alleles were naturally purged from the population as the horses adapted to Iceland's harsh environment. At the same time, humans likely applied some selective pressures, as horses were used for both transportation and labour (Björnsson & Sveinsson, 2006, pp. 76-101; Hugason, 1994). Traits such as comfortable gaits, endurance and a forward-going temperament would have been advantageous for horses used in transportation, while strength, robustness and calmness were likely preferred in labour horses. Additionally, historical records suggest that certain coat colours were selectively bred for (Björnsson & Sveinsson, 2006, pp. 18-33).

In addition to adapting to Iceland's tough environment, marked by extreme weather, poor feed and possible periods of starvation, the Icelandic horse also endured several genetic bottlenecks. One of the most significant occurred following the 18th-century *Skaftáreldar* volcanic eruption and the subsequent *Móðuharðindi (e. Famine of the mists)* period, during which over 70% of Iceland's livestock perished. The horse population plummeted from approximately 30,000 to about 8,000 (Hreiðarsdóttir et al., 2014). Another major bottleneck occurred in the mid-20th century when industrialization replaced horses with machines for labour (Figure 1). As a result, horse breeding shifted to prioritize riding horses, leading to the decline of horses better suited for labour, and the population decreased significantly during this period (Hreiðarsdóttir et al., 2014). However, in the wake of selective breeding and the increased global popularity of the breed during the late 20th century, particularly in northern Europe, the population size surged. In 1959, the population counted approximately 30,000 horses, but to date, approximately 300,000 horses are registered across 31 countries according to WorldFengur, the official studbook of origin for the Icelandic horse (Lorange, 2011).

Although there are a few historical accounts of importing individuals of other horse breeds into Iceland, the scale of these imports was not enough to leave a lasting impact on the breed's genome (Guðlaugsson, 2006). In fact, laws enacted in the late 19th century banned the importation of live animals and genetic material – a ban that remains in place today (Animal Importation Act 1990). Consequently, the Icelandic horse has remained largely genetically isolated over the past few centuries. Moreover, the modern breeding of riding horses follows a closed studbook system (FEIF – International Federation of Icelandic Horse Associations, 2024), helping to preserve the genetic distinctiveness of the breed.

1.3 Rise of the modern Icelandic horse

At the turn of the 20th century, some advances were made toward systematic breeding and a studbook for the breed was established in 1923. However, it was not until the 1950s that the first official breeding goal was introduced (Arnórsson, 2006). This goal emphasised the breeding of a versatile riding horse with five gaits. A selection index was implemented, including three conformation traits and eight traits related to riding abilities, and systematic phenotypic data collection began. These milestones marked the beginning of the modern breeding system used today.

The official breeding goal and selection index have evolved over the decades. Today, the primary objective is to breed healthy, fertile, and durable riding horses, with an emphasis on their suitability for leisure riding, travel, and various types of competitions (FEIF, 2024). The Icelandic horses should possess a calm, friendly and cooperative character, while also being courageous and reliable. Their conformation should be functional, supporting their health, durability, and natural ability to perform gaits. The conformation should, furthermore, allow the horse to move effortlessly under a rider while maintaining balance and correct posture. The aim is to breed a high-quality five-gaited horse that maintains the correct beat and posture across all gaits, moving with fluidity in an even rhythm. Additionally, the horse should demonstrate suppleness, lightness, long strides, and the ability to vary speed across all gaits (FEIF, 2024).

The breeding goal is further defined by the traits included in the selection criteria. Currently, the selection criteria consist of 16 traits, whereof eight related to conformation and eight to riding abilities (FEIF,

Figure 2. The modern Icelandic horse bred for functional conformation and versatile gait capacity. Photos by Óðinn Örn Jóhannsson.

2024). Conformation traits contribute 35% of the total index weight, with *Neck, withers and shoulders* (8%), *Proportions* (7%) and *Back and croup* (5.5%) being the most important. Riding ability traits account for 65% of the total index weight, where *Tölt* (16%), *Pace* (10%) and the composite trait *Form under rider* (10%) are the most valuable traits.

All the selection traits are subjectively assessed at standardised breeding field tests by a panel of certified breeding judges (FEIF, 2024). Traits are scored on a scale of 5.0 to 10.0, with increments of 0.5. A score of 5.0 indicates that a trait was not shown, while scores of 9.5-10.0 represent the specific breeding goal for that trait. In addition, 11 morphological traits are objectively measured and used to substantiate the subjective scoring of the selection traits. Horses participating in the breeding field tests must be at least four years old and can be reassessed in subsequent years. Each horse must be registered, individually identified and have its parentage confirmed through DNA analysis to participate in the tests (FEIF, 2024).

The *International Federation of Icelandic Horse Associations* (FEIF) oversees the breeding field tests, ensuring standardized procedures across all member countries for consistent assessment (FEIF, 2024). All data collected from these tests are available in WorldFengur, as well as information on pedigree, offspring, estimated breeding values, owners, breeders, competition results, and more for Icelandic horses worldwide (Lorange, 2011).

The breeding assessments form the basis for estimating breeding values (EBV) in the population. As early as the 1980s, the breeding program adopted the multi-trait *Best Linear Unbiased Prediction* (BLUP) animal model to estimate breeding values (Árnason, 1983, 1984). Since then, the evaluation process has expanded to include all assessments from standardised breeding field tests across FEIF member countries, incorporating individual and related animals' assessments, including progeny.

Studies have shown considerable genetic progress in most of the selection traits, particularly those with the highest weighting (Sigurðardóttir, 2012). Heritability estimates for the traits ranged from 0.15 (leg stance) to 0.56 (mane and tail) for conformation traits and from 0.18 (walk) to 0.60 (pace) for riding ability traits (Albertsdóttir et al., 2008; Albertsdóttir et al., 2011). The heritability estimates thus suggest a level of complexity, indicating that these traits are influenced by both genetic and environmental factors. Genetic correlations between the traits ranged from -0.22 to 0.92, with the strongest correlations observed within the riding ability traits (Albertsdóttir et al., 2008).

A study on the functional aspect of conformation traits found moderate to high heritability for the morphological measurements, ranging from 0.20 to 0.85 (Kristjansson et al., 2016). The study confirmed that high withers and an uphill conformation were advantageous for all gaits, while a forward (downhill) sloping back or swayback was found to be disadvantageous. Additionally, the results of the study indicated that the length and shape of the croup were distinguishing factors between high-quality and lower-quality horses in trot, tölt, and pace (Kristjansson et al., 2016).

1.4 Genetic foundation of gaits and performance

1.4.1 Genetic complexity of quantitative traits

We are still in the early stages of understanding the genetic basis of complex quantitative traits in horses, such as gaits and performance. These traits are generally governed by numerous genetic factors with additive effects spread across the genome, each with a relatively small contribution, and are further influenced by environmental factors. Interactions between loci and genes, known as epistasis, may also contribute to the variation observed in these traits (Carlborg & Haley, 2004; Mackay, 2014). Accurately predicting complex trait phenotypes from genotypes requires understanding epistasis, as interactions between major genes can significantly influence the trait and cannot be simply explained by adding the effects of individual genes (Boucher & Jenna, 2013; Carlborg & Haley, 2004; Fang et al., 2019; Forsberg et al., 2017; Sackton & Hartl, 2016). Despite the growing number of studies on epistasis, identifying genetic interactions on a genome-wide scale remains a significant challenge (Fang et al., 2019).

Additionally, traits such as conformation and temperament significantly influence a horse's performance, not only due to the potential pleiotropic effects of certain genes but, more importantly, because performance is shaped by the complex interaction between physical structure and mental aptitude. These traits jointly contribute to a horse's ability to succeed across different disciplines, making them important factors in overall performance outcomes.

1.4.2 Identified genes affecting gaits and locomotion

As outlined in Chapter 1.1, the discovery of the 'gait keeper' mutation in the *DMRT3* gene (Andersson et al., 2012) marked a significant breakthrough in the field of gait genetics. This discovery spurred a wave of subsequent studies exploring the mutation's effects across various horse breeds (Chandra Paul et al., 2020; Jäderkvist Fegraeus et al., 2015; Jäderkvist Fegraeus, Lawrence, et al., 2017; Jäderkvist, Andersson, et al., 2014; Jäderkvist et al., 2015; Jäderkvist, Kangas, et al., 2014; Kristjansson et al., 2014; Novoa-Bravo et al., 2018; Pereira et al., 2016; Promerová et al., 2014; Regatieri et al., 2017; Ricard, 2015; Sonali et al., 2023), rapidly advancing our genetic understanding of gaits in a relatively short time.

In Icelandic horses, a single copy of the 'gait keeper' mutation (CA genotype) enhances the natural ability to tölt, while two copies (AA genotype) enable the development of pace (Andersson et al., 2012; Kristjansson et al., 2014). Additionally, horses with the *DMRT3* AA genotype have been shown to achieve significantly higher breeding scores for tölt compared to those with the CA genotype. However, the AA genotype has also been associated with lower scores for the basic gaits (walk, trot, canter, and gallop) in breeding field tests (Andersson et al., 2012; Kristjansson et al., 2014). The frequency of the 'gait keeper' mutation in the Icelandic horse breed is high, estimated at 0.75-0.94 (Kristjansson et al., 2014; Promerová et al., 2014). This closely matches the 0.95 frequency observed in the datasets used in this study. This

shows that the mutation is not fixed within the breed, which explains some of the variation in pace ability. Interestingly, more than 30% of horses with the AA genotype do not perform pace (Andersson et al., 2012; Jäderkvist et al., 2015; Kristjansson et al., 2014) and are referred to as four-gaited in spite of their genetic precondition to develop pace. Additionally, genetic variation in pace scores remains among horses that perform this gait during breeding field tests (Albertsdóttir et al., 2011), suggesting that factors beyond the 'gait keeper' mutation influence pace ability and quality. A pilot study on four- and five-gaited Icelandic horses with the AA genotype at the *DMRT3* gene yielded inconclusive results, largely due to the small sample size (20 four-gaited and 35 five-gaited horses) (Jäderkvist Fegraeus, Hirschberg, et al., 2017). Studies on other gaited horse breeds have explored genetic differences between horses exhibiting different gait patterns despite sharing the same *DMRT3* genotype (Amano et al., 2018; McCoy et al., 2019; Staiger, Al Abri, Silva, et al., 2016). In Tennessee Walking Horses, a genome-wide association (GWA) study identified markers on equine chromosome (ECA) 11 and ECA19 that are predictive of gait type (lateral gaits only or both lateral and diagonal gaits) (Staiger, Al Abri, Silva, et al., 2016). These markers were located within the *SMTNL2*, *FBXO40* and *ARGFX* genes. In Standardbred horses, which perform either a diagonal or lateral gait during harness racing, an algorithm based on seven single-nucleotide polymorphisms (SNPs) with large effects was developed (McCoy et al., 2019). This algorithm, which differentiates between pacers and trotters, achieved a prediction accuracy of over 99%. It has, however, not been validated in other breeds.

1.4.3 Genetic factors impacting racing performance

Another important discovery in horse performance research was the identification of a sequence variation at the equine *myostatin* (*MSTN*) gene on ECA18, which has a significant genomic influence on optimum race distance aptitude in Thoroughbred racehorses. Specifically, one SNP (g.66493737C>T) has proven to be a strong predictor of the horse's suitability for different racing distances (Binns et al., 2010; Hill et al., 2011; Hill, Gu, et al., 2010; Hill, McGivney, et al., 2010; Hill et al., 2019; McGivney et al., 2012; Tozaki et al., 2012). Further studies on Thoroughbred racehorses have identified variations in the expression of the *PDK4* and the *COX4I2* genes that influence race performance (Gu et al., 2010; Hill, Eivers, et al., 2010; Hill, Gu, et al., 2010), as well as predictive markers associated with genes involved in myogenesis and muscle maintenance (Shin et al., 2015).

Studies on the effects of the *MSTN* gene in Icelandic horses suggest it has no clear impact on performance (Bas Conn, 2018; François et al., 2016). Frequency estimates of the predictive SNP (g.66493737C>T) from the same studies show a high prevalence of the T allele (0.89-0.90) in Icelandic horses, consistent with long-term selection for riding horses used for transportation and travel, where endurance and stamina are key traits. One study did find a significant positive association between the T allele and the EBV for the conformation trait *Neck, withers and shoulders* (François et al., 2016). However, due to the composite nature of this trait, this finding is difficult to interpret without further investigation.

In Coldblooded Trotters and Standardbreds, both primarily bred for harness racing, a genomic region near the *endothelin 3* (*EDN3*) gene was associated with racing performance (Jäderkvist Fegraeus et al., 2018). This region contains an enhancer cluster active in endothelial cells, interacting with genes relevant to blood pressure regulation, such as *GNAS* and *SPO11* (Fegraeus et al., 2024).

The *dedicator of cytokinesis 8* (*DOCK8*) gene has also been associated with harness racing success in Coldblooded Trotters (Velie et al., 2018). This gene plays a role in cell signalling, as well as in regulating cell movement and shape. While its specific effects on performance are not well understood, its close proximity to the *DMRT3* gene on ECA23 makes it particularly intriguing. Furthermore, studies in humans have suggested potential overlapping or shared effects between the *DMRT3* and *DOCK8* genes (Glessner et al., 2017; Kang et al., 2010). It is, therefore, interesting that sequence variation in *DOCK8* is associated with performance in a breed where the 'gait keeper' mutation has been shown to enhance racing ability (Jäderkvist, Andersson, et al., 2014). The possibility of shared functional influence between these two genes can thus not be ruled out. The effects of the *DOCK8* gene were investigated in a small cohort of 131 Icelandic horses with records from pace racing (Bas Conn, 2018). According to the results, the *DOCK8* polymorphism was indicated to be associated with pace speed and to potentially segregate between elite pace racers and other horses. This needs further validation, though.

In Arabian horses, a breed renowned for athleticism, sequence variations in the *SORCS3, SLC39A12* and *SLC16A1* genes have been linked to success in endurance racing (Ricard et al., 2017; Ropka-Molik et al., 2019), with one specific marker serving as a predictor for optimal race distance (Ropka-Molik et al., 2019).

1.4.4 Genetic links between conformation and locomotion

Genetic and phenotypic associations between conformation, locomotion and performance have been documented in multiple horse breeds, including the Icelandic horse (Albertsdóttir et al., 2008; Bonow et al., 2024; Holmström & Back, 2013; Jönsson et al., 2014; Koenen et al., 1995; Kristjansson et al., 2016; Novotna et al., 2022; Rustin et al., 2009; Sánchez-Guerrero et al., 2016). However, knowledge of candidate genes affecting these traits is limited, except for height at withers.

Height at withers has been shown to have a moderately favourable genetic correlation with dressage performance in some European warmblood breeds (Rustin et al., 2009; Viklund et al., 2008), while the genetic correlations to gaits in Icelandic horses are low (0.03-0.14), though still favourable (Albertsdóttir et al., 2008). These findings suggest that the genetic basis for height at withers may influence performance in certain breeds. In warmblood horses, height at the withers is predominantly governed by the *LCORL/NCAPG* locus on ECA3 (Makvandi-Nejad et al., 2012; Metzger et al., 2013; Reich et al., 2024; Ricard et al., 2023; Sevane et al., 2017; Tetens et al., 2013; Vosgerau et al., 2022). This locus has also been associated

with skeletal traits across multiple horse breeds (Frischknecht et al., 2016; He et al., 2015; Signer-Hasler et al., 2012; Staiger, Al Abri, Pflug, et al., 2016), though no such relationship has been observed in the Icelandic horse.

Regions on ECA1 have also been shown to significantly contribute to height variation in horses (Metzger et al., 2018; Signer-Hasler et al., 2012). Other genes associated with height at withers across various breeds include *ANKRD1*, *ZFAT*, *HMGA2*, *LASP1*, *ADAMTS17*, *OSTN*, *GH1,* and *TBX3* (Al Abri et al., 2018; Liu et al., 2020; Liu et al., 2022; Makvandi-Nejad et al., 2012; Metzger et al., 2018; Sevane et al., 2017; Signer-Hasler et al., 2012). The *ZFAT* gene on ECA9 has furthermore been associated with length of both back and croup (Signer-Hasler et al., 2012) and the *HMGA2* gene on ECA6 has been associated with conformation scores for front and hind parts in Spanish Purebred horses (Sevane et al., 2017).

Back conformation is widely regarded as one of the most critical traits for performance in riding horses (Kristjansson et al., 2016; Wolschrijn et al., 2013), prompting an increased focus on this trait in genomic studies of various riding horse breeds (Cook et al., 2010; Frischknecht et al., 2016; Gmel et al., 2023; Sevane et al., 2017; Signer-Hasler et al., 2012). These studies have identified quantitative trait loci (QTLs) linked to back traits such as lordosis, swayback, and back length on ECA2, 9, 10, 20, and 21, involving several candidate genes. While these QTLs show predictive potential, they have yet to be validated through functional analysis, highlighting the need for further research to solidify the genetic foundation of this critical trait, which impacts both performance and the overall health of riding horses.

In addition to back conformation, neck conformation is equally critical for proper equine function (Bonow et al., 2024; Wolschrijn et al., 2013). However, genetic understanding of this complex trait is limited. Two recent studies employing two-dimensional shape data and joint angle measurements have identified genomic regions linked to this trait. One on ECA16 was found to be significantly associated with neck shape and width, distinguishing between a slender neck and a 'cresty' neck (Gmel et al., 2023), while another on ECA28 was linked to the poll angle (Gmel et al., 2019). Advances in using two- and three-dimensional shape data, along with other objective measurements, for phenotyping conformation traits (Druml et al., 2015; Gmel et al., 2023; Gmel et al., 2022; Gmel et al., 2019; Kristjansson et al., 2016; Kristjansson et al., 2013; Ricard et al., 2023), offers promising opportunities to improve accuracy and address the challenges posed by the subjective nature of conformation assessments. Greater objectivity in these phenotypes, combined with machine learning and rapidly evolving artificial intelligence, will hopefully lead to more robust results and significantly enhance the outcomes of future trait mapping studies.

1.4.5 Genetic influences of temperament on gaits and performance

The success of a performing horse in any discipline depends greatly on its mentality (McBride & Mills, 2012), along with functional conformation and genetic ability to meet the demands placed on it.

However, the challenges in defining and assessing highly subjective temperament traits continue to be a barrier (Sigurðardóttir et al., 2017), limiting the likelihood of significant findings in studies such as GWA. Despite this, Jäderkvist Fegraeus et al. (2017) and Velie et al. (2018) identified the same QTL on ECA6 in two separate GWA studies on different breeds – one examining pace ability in Icelandic horses and the other focusing on harness racing success in Nordic Coldblooded trotters. Both studies pinpointed the *GRIN2B* gene, known for its role in memory and learning in laboratory species (Sahin et al., 2023; Tang et al., 1999), as a candidate gene. A larger cohort of Icelandic pace-racing horses was later genotyped for the identified variants, providing supportive results of the gene's impact on both speed and focus/motivation (Bas Conn, 2018). The *GRIN2B* gene has, furthermore, been found to be a target of selection in French Trotters and Gidran horses (Grilz-Seger, Neuditschko, et al., 2019) and Swedish Warmblood sport horses (Ablondi et al., 2019).

Additionally, the *NTM* gene, also associated with learning and memory (Mazitov et al., 2017; Pan et al., 2011), has been proposed as a key contributor to the gallop racing phenotype (Han et al., 2022; McGivney et al., 2019) and was one of the top genes selected for during horse domestication (Schubert et al., 2014). In relation to performance, one could argue that to better understand the genetics of a 'winner temperament' and the variants distinguishing elite from non-elite horses, performance phenotypes might offer more valuable insights than subjective assessments of personality traits.

Nevertheless, studies focused on direct temperament traits have also successfully identified candidate genes linked to personality, such as the *dopamine D4 receptor* (*DRD4*) gene (Hori et al., 2013; Momozawa et al., 2005; Ninomiya et al., 2013). Holtby et al. (2023) also identified genes associated with coping mechanisms during early training and cortisol levels the first time a horse is mounted. Another study, based on owner-reported temperament traits, discovered five candidate regions potentially influencing traits like 'anxiousness,' 'tractability,' and 'agonistic behaviour' in Tennessee Walking horses (Staiger, Albright, et al., 2016). Interestingly, none of these genes, nor the *GRIN2B* or *NTM* genes, appear on a list of candidate genes for horse temperament derived from studies on human personality traits (Momozawa et al., 2007; Song et al., 2017; Yokomori et al., 2023). Only a few genes from that list have shown direct associations with horse temperament, with variants significantly segregating between docile and aggressive horses (Song et al., 2017).

1.5 Dissecting the Icelandic horse genome

The Icelandic horse has attracted considerable scientific interest due to its unique ability to perform six gaits, but the genetic basis for this versatility remains only partially understood. With the diverse gait capacity and extensive phenotypic records, combined with advanced genomic technologies, the Icelandic horse offers a valuable model for further research into the equine genome, particularly the genetics of gait.

1.5.1 Methods for detecting genetic variation

Rapid advances in genomic techniques and resources in the equine field now enable researchers to analyse entire genomes from various perspectives (Raudsepp et al., 2019). A key aspect of this progress is the ongoing refinement of the horse reference genome's quality, completeness, and functional annotation (Kalbfleisch et al., 2018; Wade et al., 2009). One approach involves utilizing genetic variation to investigate associations between SNPs and specific phenotypes. This method, known as a GWA study, is statistically robust when sufficient data size and quality are provided (Risch & Merikangas, 1996), and has facilitated the identification of numerous genetic variants associated with phenotypic variation across multiple species (Hu et al., 2021; Sollis et al., 2023). When combined with mixed models, GWA analysis corrects for population stratification, commonly seen in admixed or related groups, thus minimizing statistical overestimation and reducing false positives (Shin & Lee, 2015). GWA studies commonly use DNA arrays that include a carefully selected subset of SNPs, chosen through stringent filtering based on e.g. LD, to ensure that the fewest possible SNPs are used while still providing comprehensive genomic coverage (Jorgenson & Witte, 2006; Schaefer et al., 2017). This approach effectively covers a significant portion of the genome while reducing the costs of genotyping all genomic variants. However, in species with diverse breeds, such as horses, DNA arrays are typically designed based on reference genomes from one or a few individuals of a specific breed. This can create challenges when studying distantly related breeds, as sequences that differ significantly from the reference genome may be less effectively detected. The rapid advancement of pan-genome development, however, holds promise for improving detection efficiency across diverse breeds (Gong et al., 2023).

Most variants identified in GWA studies are common, with minor allele frequency (MAF) typically set to $\geq$ 5%, and tend to have modest or small effects on traits (Chen et al., 2009). Importantly, these variants typically do not represent the actual causal variants; instead, they point to genomic regions of interest, commonly referred to as QTLs. To pinpoint the candidate variant within these regions, whole-genome sequencing (WGS) can be employed to generate a more comprehensive set of variants for further exploration, as done in Paper III in this thesis.

Despite the dramatic reductions in DNA sequencing costs over the past decades, researchers still face critical decisions on how to allocate sequencing resources between coverage depth and sample size. Consequently, careful selection of animals is required to ensure they represent the desired phenotypes while minimizing sample numbers to achieve sufficient coverage. Following comprehensive bioinformatic analysis using variant calling pipelines, such as the Genome Analysis Toolkit (GATK) best practices (Van der Auwera & O'Connor, 2020), along with other variant analysis tools, researchers can make significant progress in predicting and interpreting the effects of candidate variants. In this context, genome browsers developed by the University of California Santa Cruz (UCSC Genome Browser) (Nassar et al., 2023), Ensembl (Harrison et al., 2024), and the Functional Annotation of Animal Genomes (FAANG) (Giuffra et al., 2019) play a pivotal role. However, functional studies remain essential to confirm and validate their biological impact.

1.5.2 Exploring the genomic landscape of a population

Another approach for analysing the entire genome involves examining genome-wide patterns of homozygosity to assess population genomic measures. Selective breeding practices often reduce genetic diversity in targeted regions of the genome, leading to increased homozygosity. These patterns, known as signatures of selection, can be studied using modern genomic methods, including the estimation of continuous homozygous segments, referred to as runs of homozygosity (ROH) (Curik et al., 2014; Peripolli et al., 2017). This method provides valuable insights into the genetic consequences of selection as well as serves as a measure of inbreeding, as ROH can result from the mating of related animals (Curik et al., 2014; Peripolli et al., 2017). In general, short ROHs indicate distant inbreeding, while longer ROHs (> 5.0 Mb) suggest more recent inbreeding, occurring approximately within the last 10 generations (Curik et al., 2014). The genomic inbreeding coefficient, F_{ROH}, is defined as the proportion of the autosomal genome that falls within ROH above a specified length (McQuillan et al., 2008).

The Icelandic horse has been under rigorous artificial selection in recent decades. Furthermore, historical population bottlenecks and genetic drift may have contributed to a loss of genetic variability, given the small genetic pool from which the breed emerged and its lengthy history of isolation. Inbreeding based on pedigree data has been monitored in the breeding program for the last generations, and the most recently estimated mean pedigree-based inbreeding coefficient (F_{PED}) for all Icelandic horses born in Iceland in 2020 was reported to be 0.03 (Árnadóttir, 2022). The effective population size (Ne) for the same cohort was estimated to range from 95 to 103 horses depending on the pedigree completeness index (Árnadóttir, 2022). The mean inbreeding coefficient based on microsatellite data for 442 Icelandic horses has furthermore been reported to be close to 0.04, and the corresponding Ne estimate was 215 (Hreiðarsdóttir et al., 2014). Other studies using smaller datasets of Icelandic horses (N = 18-25) and medium-density SNP data (~50,000 SNPs) have reported inbreeding coefficients to be 0.08-0.09 (McCue et al., 2012; Petersen, Mickelson, Cothran, et al., 2013) and Ne estimation based on LD analysis to be 555 individuals (Petersen, Mickelson, Cothran, et al., 2013). More recent studies using ROH analysis with medium-density SNP data have reported average F_{ROH} values for Icelandic horses to range from 0.03-0.13 (Colpitts et al., 2022; Cosgrove et al., 2020; Meyermans et al., 2020). Furthermore, studies using medium-density SNP data have reported signatures of selection on ECA3, 10, 11, 15 and 23 in the Icelandic horse (Colpitts et al., 2022; Gorssen et al., 2021; Petersen, Mickelson, Rendahl, et al., 2013). However, directly comparing ROH analyses is challenging because results are highly influenced by QC settings and ROH definitions, and no standard guidelines exist for these.

1.5.3 Holistic perspective to genome-wide investigation

By integrating genomic approaches that analyse genetic variation through genome-wide patterns of homozygosity and polymorphic variation, researchers can enhance the understanding of the studied genome, as demonstrated in this thesis for the Icelandic horse. ROH analysis identifies genomic regions influenced by positive or negative selection, shedding light on the genetic basis of traits that offer adaptive

advantages or result from artificial selection. This approach provides insights into the evolutionary pressures shaping the population's genetic diversity. In contrast, GWA studies pinpoint polymorphic variation directly linked to specific phenotypic traits, a process that can be further refined using WGS. While GWA studies exclude homozygous regions with MAF < 0.05, ROH analysis specifically focuses on these regions, offering complementary perspectives. Together, these methods offer insights into the genetic background, combining evolutionary perspectives with precise trait associations.

2. Aims of the thesis

The overall aim of this thesis was to provide novel insight into the genetic background of gaiting ability and quality of gaits in Icelandic horses.

The specific aims were to:

- Identify genomic regions associated with conformation of back and croup in Icelandic horses and investigate their effects on riding ability traits assessed at breeding field tests (Paper I).

- Identify novel genetic factors that influence pacing ability and quality and investigate their interactions with the 'gait keeper' mutation in the *DMRT3* gene (Paper II).

- Explore candidate causal mutations responsible for the observed effects of the *STAU2* and *RELN* haplotypes on gaits and performance (Paper III).

- Study the genomic population structure of the Icelandic horse by analysing runs of homozygosity and identifying genomic regions under putative selection, as well as estimating genomic inbreeding and genetic diversity (Paper IV).

3. Summary of investigations

This thesis presents the work conducted across four papers. In Paper I, a GWA analysis was performed using SNP genotype data from 190 Icelandic horses to investigate associations between specific SNPs and breeding field test scores for the conformation trait *Back and croup*. In Paper II, an additional 190 SNP-genotyped samples were included in the dataset, and a second GWA study was conducted, this time focusing on the gait trait *Pace*. As a sequel to this paper, WGS data from 39 breeding horses were utilized to further investigate potential causal mutations within the identified QTLs found to be associated with pace. Results from this analysis are depicted in Paper III. Finally, the complete dataset of 380 SNP-genotyped samples was further used to analyse ROHs and estimate genetic diversity, with the results published in Paper IV.

3.1 Material

The thesis included a total of 419 DNA samples from 406 individual horses. All horses were privately owned, and informed consent was obtained from the owners prior to sampling. Hair samples were collected from the tails of 380 horses during breeding field tests or visits to trainers and breeders in Iceland and Sweden, while blood samples were collected by certified veterinarians in both countries after ethical permits were achieved. Half of the individuals selected for SNP genotyping were chosen based on mane growth characteristics (used in Papers I, II and IV), while the other half was randomly chosen at breeding field tests (used in Papers II and IV). Close relationships and stratifications were avoided in both datasets. Horses that were selected for sequencing (Paper III) were carefully curated based on criteria designed to categorize them into elite and non-elite groups. The selection process and criteria are detailed in Paper III. A summary of the datasets used, comprising SNP genotype data and WGS data, is presented in Table 1.

Phenotypic data for the thesis were retrieved from the WorldFengur database, consisting of morphological measurements and breeding scores for traits subjectively assessed during standardized breeding field tests. The primary traits studied were the conformation trait *Back and croup*, and the gait traits *Pace*, *Tölt*, *Slow tölt*, *Trot*, *Canter,* and *Gallop*. Detailed descriptions of these phenotypes are provided in Paper I and II. Additionally, morphological measurements, such as *Height at front*, *Height of withers*, and *Back inclination*, were included, as well as sub-traits describing *Croup type* and *Backline*, with descriptions available in Paper I, II, and III.

Pedigree data, also obtained from the WorldFengur database, covered individuals born between 1860 and 2023, although the earliest records represent only a small portion of the population from that time. The pedigree file encompassed a total of 548,779 individuals.

Table 1. Summary of the genomic datasets analysed in Papers I-VI.

	SNP genotype dataset 1	SNP genotype dataset 2	Whole-genome sequence data
Number of horses	190	190	39
Sample type	Hair	Hair	Blood
Year of sample collection	2018	2020	2022-2023
Data was used in papers	I, II and IV	II and IV	III
Gender			
Male	90	76	19
Female	100	114	20
Birth year			
Min	1993	2005	2006
Max	2014	2016	2017
Average	2009 ± 5	2013 ± 2	2012 ± 3
Available breeding assessments			
Never assessed	11	0	0
Conformation only	5	0	0
Conformation & riding ability	174	190	39
Assessment year			
Min	1999	2013	2012
Max	2022	2022	2023
Average	2015 ± 5	2020 ± 1	2019 ± 2
Age when assessed			
Min	4	4	4
Max	15	13	14
Mean	6.8 ± 1.9	6.5 ± 1.5	7.5 ± 2.5
Country of assessment			
Iceland	81	190	27
Sweden or other	98	0	12
***DMRT3* 'gait keeper' genotype**			
CC	1	0	0
CA	21	11	4
AA	165	179	35
Not known	3	0	0
Genetic relationship based on pedigree			
Min	0.01	0.01	0.01
Max	0.30	0.19	0.12
Mean	0.04 ± 0.02	0.04 ± 0.02	0.04 ± 0.02

3.2 Methods

3.2.1 Analysing variation with SNP genotype data

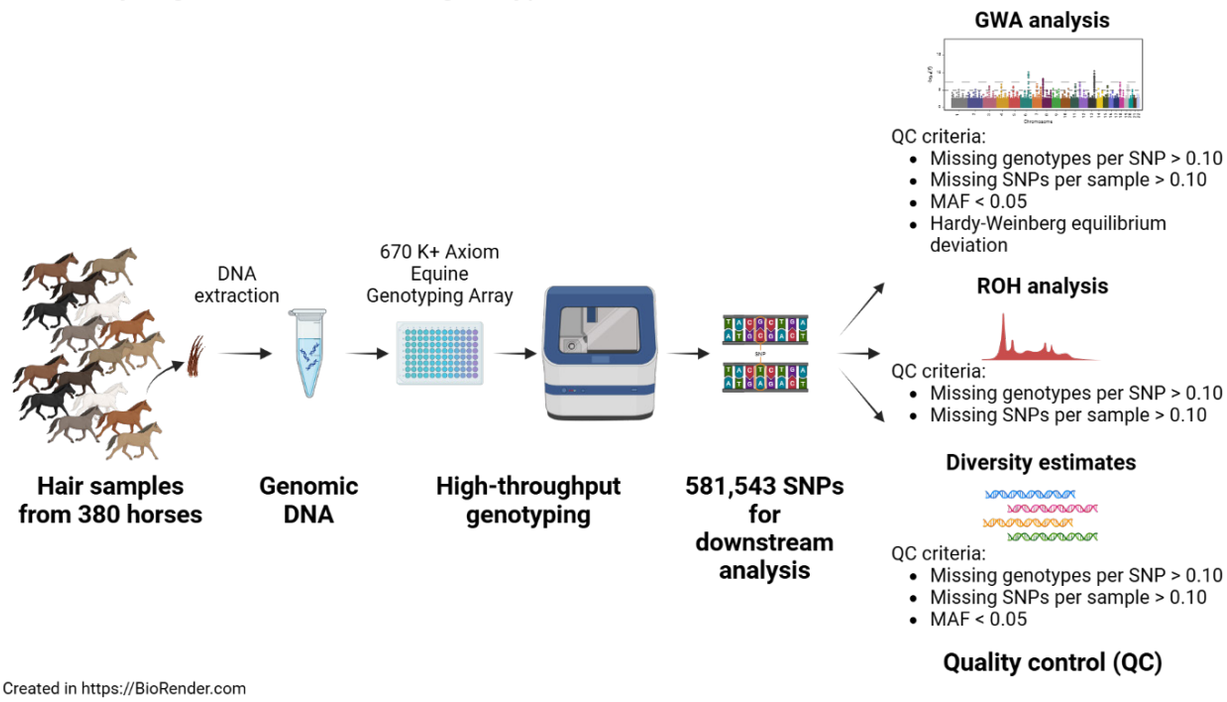


Figure 3. Processing of SNP genotype data: from horse hair samples to the application of three QC approaches for different analyses.

Genomic DNA was extracted from the follicles of the sampled hairs using protocols detailed in Papers I and II. The DNA samples were genotyped using the 670 K+ Axiom Equine Genotyping Array, yielding 581,543 SNPs recommended by the genotyping facilities for further analysis. SNP positions were mapped to the EquCab3.0 reference genome. Quality control (QC) was performed using three criteria depending on the downstream analysis (Figure 3).

GWA studies in Papers I and II were conducted using the GenABEL package (Aulchenko et al., 2007) in R (v.3.6.1) (R Core Team, 2019). This package allows an estimation of a polygenic model using a hierarchical generalized linear model (Rönnegård et al., 2010) and implements a mixed model-structured association approach. Additionally, GenABEL enables the computation of a genomic kinship matrix between studied individuals, which can be used to identify potential stratification in the sample as well as to adjust for the population structure.

In Paper I, genotype data, breeding scores for back and croup, the fixed effect of sex, and a genomic kinship matrix were used as inputs for the relevant GenABEL functions. In Paper II, the fixed effect of the *DMRT3* 'gait keeper' genotype was added to the model, with breeding scores for pace used for association analysis. Fixed effects such as age at assessment, country, and year of assessment were tested

but found not to be significant ($p > 0.05$) in either model. Genome-wide significance was determined by Bonferroni correction (Duggal et al., 2008; Lander & Kruglyak, 1995) (Paper I: $p \leq 1.3 \times 10^{-7}$, Paper II: $p \leq 1.4 \times 10^{-7}$), with a suggestive significance threshold set at $p \leq 1.0 \times 10^{-5}$.

SNPs surpassing the suggestive significance threshold were further analysed for LD ($r^2 \geq 0.8$), followed by haplotype analysis, where frequencies and effects of haplotypes on back and croup (Paper I) and pace (Paper II) were estimated. Additionally, haplotype effects on other traits from breeding field tests were assessed by comparing the scores of individuals homozygous for different haplotypes.

In Paper I, the allele frequency of the top SNP associated with back and croup was examined across several gaited and non-gaited breeds. In Paper II, haplotype frequencies were compared among groups of Icelandic horses with varying levels of pacing ability, and the phenotypic variance explained by significant haplotypes was quantified. Interactions between identified haplotypes and the 'gait keeper' mutation in the *DMRT3* gene were also explored in Paper II. Detailed descriptions of the statistical methods are provided in Papers I and II.

3.2.2 Whole genome sequence analysis

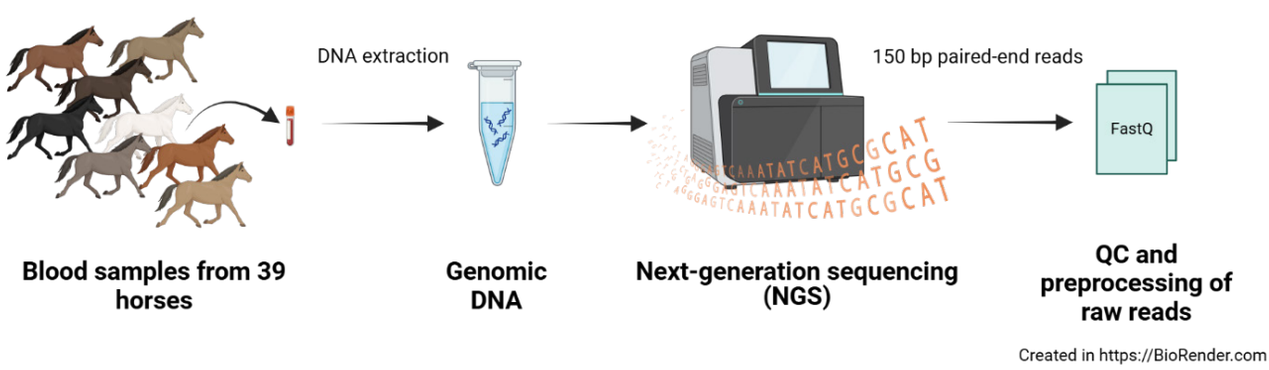


Figure 4. Processing of whole-genome sequence data: from blood samples collected from horses to the generation of raw sequence files.

For paper III, genomic DNA from blood samples was extracted using the QIAsymphony instrument (QIAGEN, Hilden, Germany), following a standard protocol. These samples were sequenced on an S4 flowcell using the Illumina NovaSeq 6000 system with a 150 bp paired-end read length (Figure 4).

The quality control and variant calling pipeline for the sequence data in Paper III is described in detail in the paper and outlined in Figure 5. Processed reads were mapped to the EquCab3.0 reference genome (Ensembl build 110). The haplotypes identified in Paper II were redefined due to the higher variant density in the sequence data. Variants in LD ($r^2 \geq 0.8$) with the previously identified top SNPs were considered for further analysis with a bioinformatics approach. This included variant annotation,

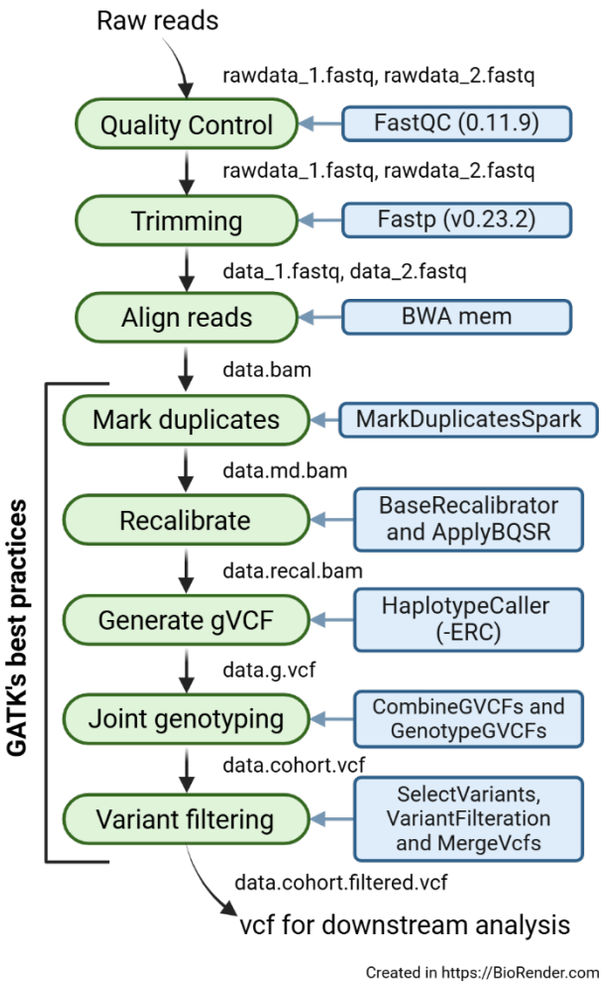


Figure 5. QC and variant calling pipeline used in Paper III.

ranking, and extraction of Genomic Evolutionary Rate Profiling (GERP) scores, along with examining overlaps with known histone modification marks from the FAANG project.

Protein analysis was conducted using bioinformatics tools described in Paper III to translate coding sequences into protein sequences and for protein modelling. A general linear model was then applied to assess phenotypic differences between individuals homozygous for the reference haplotype and those homozygous for the alternative haplotype in the regions of interest. Additionally, publicly available genomic data was used to estimate allele frequencies of the top SNPs in the two target regions in other gaited and non-gaited breeds.

3.2.3 Analysing homozygosity with SNP genotype data

In Paper IV, ROHs were analysed where the ROH parameter settings were optimised according to the recommendations of Meyermans *et al.* (2020) to achieve maximum genome coverage. These optimized parameter settings, detailed in Paper IV, enabled ROH detection for 99.4% of the autosomal genome, indicating high validity of the analysis. The ROH length setting did not affect genome coverage and was therefore chosen based on the correlation between the F_{ROH} and F_{PED} values. The highest correlation ($r = 0.57$, $p < 2.2 \times 10^{-16}$) was observed when the minimum ROH length was set equal to 100 kb. The identified ROHs were divided into length classes ($0.1 < ROH \leq 1$ Mb; $1 < ROH \leq 2$ Mb; $2 < ROH \leq 4$ Mb; $4 < ROH \leq 8$ Mb; and $ROH > 8$ Mb) to enable approximation of their origins over time.

ROH islands shared by more than 70% of the horses were determined as signatures of selection. This 70% threshold, which is conservative compared to values reported in the literature (Ablondi et al., 2020; Ablondi et al., 2019; Amano et al., 2023; Bizarria dos Santos et al., 2021; Colpitts et al., 2022; Grilz-Seger, Druml, Neuditschko, Dobretsberger, et al., 2019; Grilz-Seger, Druml, Neuditschko, Mesarič, et al., 2019; Grilz-Seger et al., 2018; Grilz-Seger, Neuditschko, et al., 2019; Mousavi et al., 2023; Nazari et al., 2022), was chosen to minimise false positive signatures of selection caused by population history events, such as genetic bottlenecks.

Using the same ROH analysis settings, signatures of selection were also examined in the Exmoor pony, an ancient native breed adapted to harsh conditions similar to the Icelandic horse but not bred for gait performance in recent times. This comparison provided a valuable opportunity to differentiate between signatures of selection related to performance traits and those resulting from environmental adaptations. The genetic relationships and clustering patterns in the two breeds were further investigated by applying principal component analysis (PCA).

To estimate genomic inbreeding, F_{ROH} was calculated by summing the total length of ROH for each individual and dividing it by the autosomal genome length (McQuillan et al., 2008), set at 2281 Mb based on the genome length covered by SNPs. For comparison with other studies, F_{ROH} values were also calculated using a minimum ROH length of 500 kb instead of 100 kb. The pedigree-based inbreeding coefficient F_{PED} was also estimated for the horses included in the dataset.

Genetic diversity in the Icelandic horse population was estimated by analysing the effective population size (Ne) using genomic data from horses born between 2006 and 2016 (N = 342), representing approximately one generation. Genetic diversity was further assessed by calculating observed (H_O) and expected heterozygosity (H_E) for all 380 horses. Detailed analysis descriptions are provided in Paper IV.

3.2.4 Functional annotation of candidate genes and variants

In all four papers, the NCBI genome data viewer (Rangwala et al., 2021) and the Ensembl genome browser (Harrison et al., 2024) were used to screen for candidate genes based on the EquCab3.0 reference genome. The Horse QTLdb (Hu et al., 2021) was also consulted to identify overlaps with previously reported QTL in horses. Functional annotation of potential candidate genes was conducted using the GeneCards database (Safran et al., 2021; Stelzer et al., 2016).

Additionally, in Paper III, a more extensive bioinformatics approach was employed. In addition to the abovementioned resources, the UCSC Genome Browser (Nassar et al., 2023) and JBrowse for the Equine FAANG project (Giuffra et al., 2019) were used for the functional analysis of the variants of interest.

In Paper IV, candidate genes located within identified signatures of selection were subjected to gene ontology (GO) analysis using PANTHER v18.0 (Thomas et al., 2022) to determine significantly enriched biological processes and molecular functions under positive selection.

4. Main findings

4.1 QTL associated with breeding score for back and croup

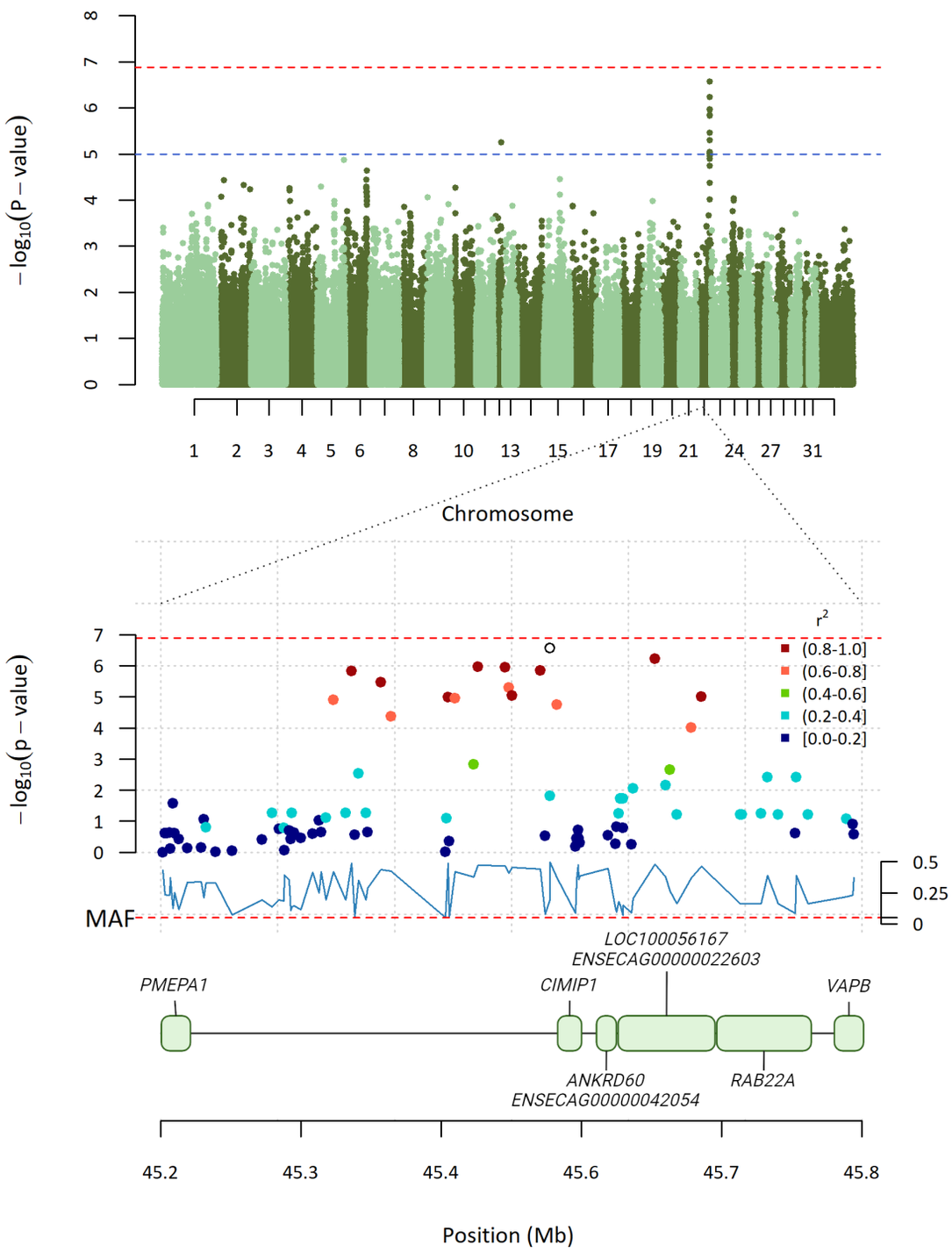


Figure 6. GWA results for the breeding score of back and croup. On top is the Manhattan plot from the mixed model association analysis showing a QTL on ECA22. The red horizontal line indicates the Bonferroni significance threshold ($p \leq 1.3 \times 10^{-7}$), and the blue horizontal line indicates the suggestive genome-wide significance level ($p \leq 1.0 \times 10^{-5}$). In the middle, is a focused LD Manhattan plot on ECA22 with the top SNP as an open circle. Thirteen SNPs reached the suggestive threshold, of which ten were in LD ($r^2 \geq 0.8$). Below the LD Manhattan plot is an overview of the genes that are located in this region. All positions refer to the EquCab3.0 genome assembly.

The first GWA study identified a distinct QTL on ECA22 (45,347,522-45,662,708) associated with breeding scores for back and croup (Figure 6). This QTL contained 13 SNPs that met the suggestive significance threshold ($p \leq 1.0 \times 10^{-5}$), with 10 of these SNPs in LD ($r^2 \geq 0.8$) (45,363,022-45,662,708; ~300 kb). Two significant haplotypes ($p < 0.001$) affecting back and croup scores were identified. The more frequent ($f_{hap} = 0.48$) one was associated with a score increase (+0.3) and was therefore called the 'favourable' haplotype. The other one had a lower frequency ($f_{hap} = 0.38$) and was associated with a score decrease (-0.3) and was therefore referred to as the 'unfavourable' haplotype. The two haplotypes are depicted in Figure 7.

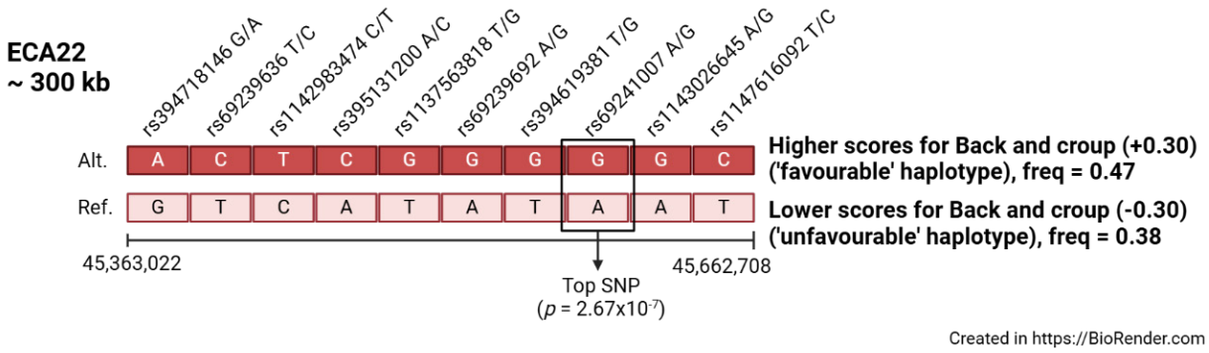


Figure 7. The significant haplotypes ($p < 0.001$) that were associated with breeding scores for back and croup.

The haplotypes overlapped with three genes; the *ciliary microtubule inner protein 1* gene (*CIMIP1*, synonym to *C22H20orf85*), the *ankyrin repeat domain 60* (*ANKRD60*) gene and *LOC100056167* (ENSECAG00000022603) gene, annotated as *serine/threonine-protein phosphatase 4 regulatory subunit 1*.

Among the horses studied, 34 were homozygous for the favourable haplotype, and 28 were homozygous for the unfavourable haplotype. Comparisons between these groups based on other scores and measurements from breeding field tests indicated that the favourable haplotype was also associated with improved scores for tölt and pace, and more favourable morphological characteristics, such as depth of the breast, forelimb length, width of the tuber coxae and width of the tuber ischii, backline and croup type. The differences were statistically significant ($p < 0.05$) for all these traits (Table 2, Paper I).

The alternative allele for the top SNP (rs69241007) on ECA22, which was part of the favourable haplotype, had a frequency of $f_{alt\_al} = 0.50$ in this dataset of 190 horses. Examination of this allele in other breeds, both gaited and non-gaited, showed generally lower frequencies, although gaited breeds tended to exhibit a slightly higher frequency than non-gaited breeds.

4.2 QTLs associated with pace and other gaits

The second GWA study identified two QTLs linked to breeding scores for pace. The first QTL was located on ECA9 (11,533,922-13,457,268) and included seven SNPs meeting the suggestive significance threshold ($p \leq 1.4 \times 10^{-7}$). Within this QTL, four SNPs were in LD across approximately 171 kb (13,198,591-13,370,069). The second QTL was located on ECA4 (4,222,615-4,228,914) and contained three SNPs that reached the suggestive significance threshold ($p \leq 1.0 \times 10^{-5}$), with two SNPs in LD ($r^2 \geq 0.8$) over a region of approximately 6 kb (4,222,615-4,228,503) (Figure 8).

In the QTL on ECA9, two significant haplotypes ($p < 0.001$) were observed. The more frequent haplotype ($f_{hap} = 0.90$) had a strong positive effect on pace scores (+0.89), while the less frequent haplotype ($f_{hap} = 0.08$) was linked to lower scores (-0.89). Similarly, within the QTL on ECA4, two haplotypes were detected that significantly affected pace scores ($p < 0.001$). The more common haplotype ($f_{hap} = 0.54$) was linked to higher pace scores (+0.51), while the less common haplotype ($f_{hap} = 0.45$) was associated with lower scores (-0.51). The two significant haplotypes on each QTL are depicted in Figure 9. The haplotypes on ECA9 and ECA4 individually accounted for 4.4% of the phenotypic variance in pace scores, and when combined with the *DMRT3* 'gait keeper' mutation, they explained 23% of the variance.

The haplotypes on ECA9 overlapped with the *staufen double-stranded RNA binding protein 2* (*STAU2*) gene (Figure 8) and were labelled as the *STAU2* haplotypes, with the favourable haplotype for pace designated 'S' and the unfavourable one 's'. Similarly, the ECA4 haplotypes overlapped with the *reelin* (*RELN*) gene (Figure 8) and labelled as the *RELN* haplotypes: the favourable one as the 'R' haplotype and the unfavourable as the 'r' haplotype.

Among the horses studied, 289 were homozygous for the 'S' haplotype (SS genotype), while only five horses were homozygous for the 's' haplotype (ss haplotype). Due to the limited number of ss individuals, a combined group, including all the other horses in the dataset, was compared against the SS genotype group. This analysis revealed significant differences ($p < 0.01$) in scores for trot, gallop, and pace, where horses with the SS genotype had higher scores for pace, but lower for trot and gallop.

For the *RELN* haplotypes, 99 horses were homozygous for the 'R' haplotype (RR genotype), and 65 were homozygous for the 'r' haplotype (rr genotype). Comparing the phenotypes for the breeding field test traits between these two groups showed that horses with the RR genotype had higher scores for pace, while those with the rr genotype scored significantly higher for other gaits (tölt, slow tölt, trot, canter, and gallop; $p \leq 0.001$). Horses with the rr genotype were also measured significantly higher at front and with higher withers than those with the RR genotype ($p < 0.05$).

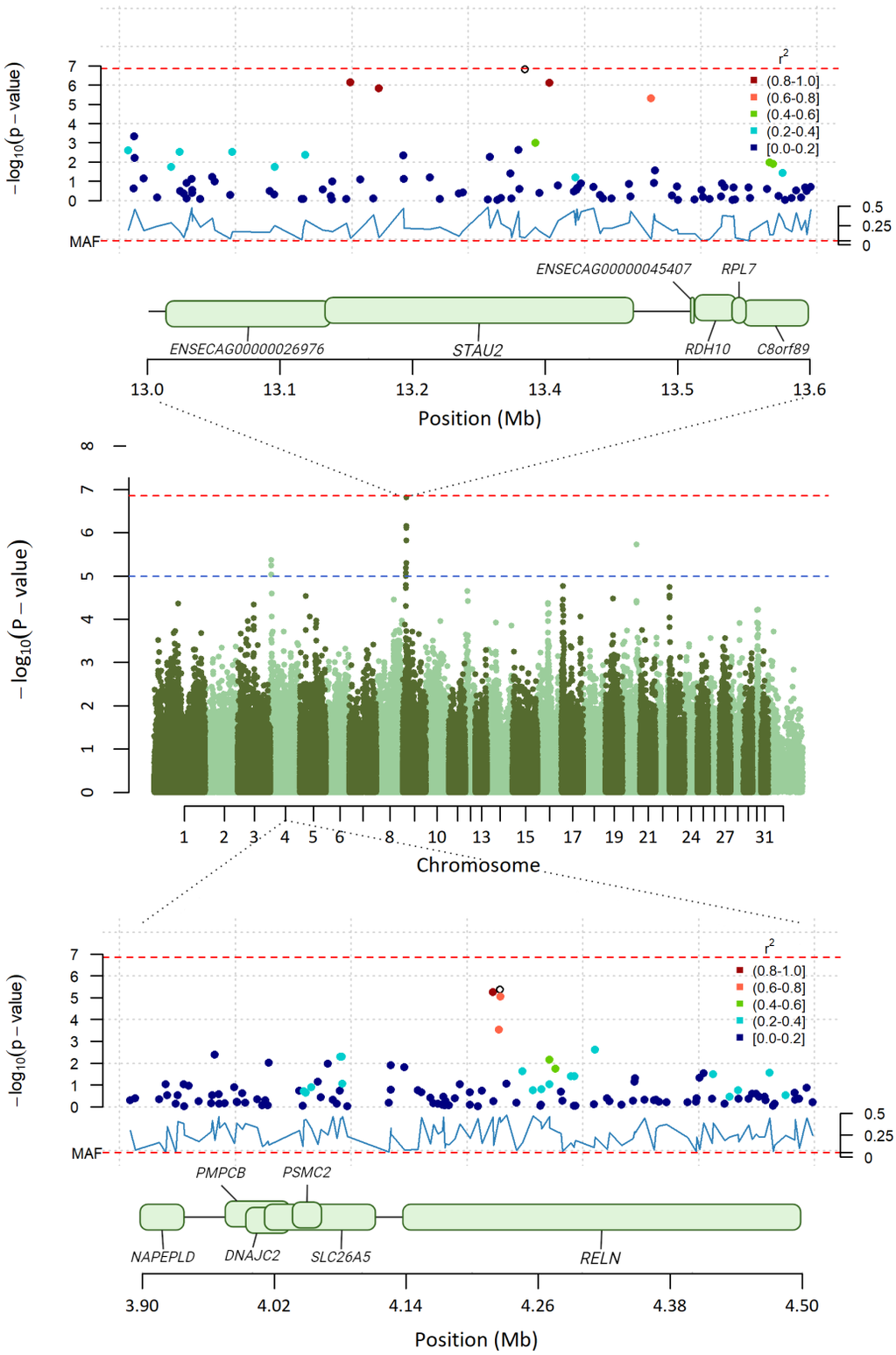


Figure 8. GWA results for the breeding score for pace. In the middle is the Manhattan plot from the mixed model association analysis showing QTLs on ECA4 and ECA9. The red horizontal line indicates the Bonferroni significance threshold ($p \leq 1.4 \times 10^{-7}$), and the blue horizontal line indicates the suggestive genome-wide significance level ($p \leq 1.0 \times 10^{-5}$). Above the GWA Manhattan plot is a focused LD Manhattan plot on ECA9 with the top SNP as an open circle. Seven SNPs reached the suggestive threshold of which four were in LD ($r^2 \geq 0.8$). Below the GWA Manhattan plot is a focused LD Manhattan plot on ECA4 with the top SNP as an open circle. Three SNPs reached the suggestive threshold of which two were in LD ($r^2 \geq 0.8$). Below each LD Manhattan plot is an overview of the genes that are located in that region. All positions refer to the EquCab3.0 genome assembly.

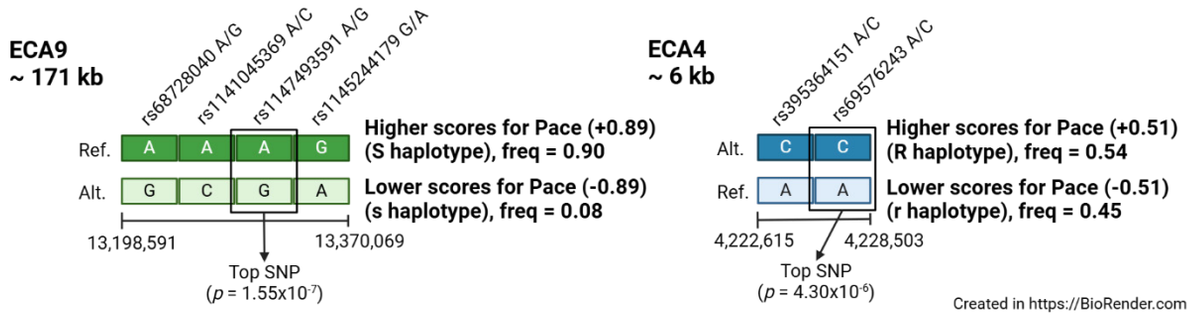


Figure 9. The significant haplotypes ($p < 0.001$) on ECA9 (to the left) and ECA4 (to the right) that were associated with breeding scores for pace.

Given the contrasting effects of the haplotypes, genotype frequencies were examined in horses with varying levels of pace ability. Four-gaited horses with the AA genotype at the *DMRT3* 'gait keeper' locus showed significantly lower frequencies of both the SS and RR genotypes (SS: 63% and RR: 15%) than those five-gaited with the same 'gait keeper' genotype (SS: 88% and RR: 31%) (Table 2). Furthermore, the rr genotype was absent among the small group of pace racers, while the ss genotype was absent among horses with the CA genotype in the 'gait keeper' locus. The prevalence of the SS genotype was also observed to be highest in this group (90%).

Table 2. Genotype frequencies for the *STAU2* and *RELN* haplotypes in groups of horses with different levels of pace ability. Significant differences from a pairwise comparison ($p < 0.001$) are shown in bold fonts. AA horses refer to horses homozygous for the *DMRT3* 'gait keeper' mutation while CA horses refer to those heterozygous for this mutation. Furthermore, 5g refers to five-gaited horses and 4g refers to four-gaited horses.

	N	***STAU2* haplotypes**			***RELN* haplotypes**		
		SS	**Ss***	**ss**	**RR**	**Rr***	rr
All horses	372	82%	17%	1%	27%	55%	17%
All AA horses	340	81%	17%	1%	27%	56%	17%
5g AA horses (pace>5.0)	248	**88%**	**12%**	**1%**	**31%**	56%	**13%**
All CA horses	31	90%	10%	*0%*	29%	48%	23%
4g AA horses (pace=5.0)	82	**63%**	**33%**	**4%**	**15%**	54%	32%
Pace racers (AA horses)	10	70%	30%	*0%*	30%	70%	*0%*

N number of horses. *The frequency of the heterozygous genotypes (Ss and Rr) also included the rare haplotypes according to the haplotype analyses (2% of the sample size on ECA9 and 1% on ECA4).

The genetic interactions between the identified haplotypes and the *DMRT3* 'gait keeper' mutation were analysed, and the findings aligned with the effects seen from each haplotype on its own. The most favourable genetic combination for pace was the SS:RR genotype combination, along with the AA genotype for the 'gait keeper' mutation. Conversely, the Ss:rr genotype combination was the least advantageous for pace, yet it was the most beneficial for other gait traits, including tölt, slow tölt, trot, gallop, and canter.

4.3 Candidate causal variant and regulatory elements

The genomic regions of the *STAU2* and *RELN* genes were further analysed with sequence data. The data had an average coverage depth of 9.5x, ranging from 6.9x to 15.6x in the *STAU2* region (chr9:13,155,196-13,453,483) and 6.8x to 15.2x in the *RELN* region (chr4:4,127,654-4,586,670). Details of *STAU2* and *RELN* gene structures and transcript variants are provided in Paper III.

Redefining the haplotypes with additional LD estimates ($r^2 \geq 0.8$) showed that the refined *STAU2* haplotype included 39 variants in LD and was expanded to cover approximately 250 kb (chr9:13,198,591-13,448,796), or 84% of the *STAU2* gene. The refined *RELN* haplotype contained 24 variants while retaining the same boundaries and spanning approximately 6 kb (chr4:4,222,615-4,228,503). A detailed list of the variants comprising the refined haplotypes is provided in Paper III.

VEP analysis of the refined *STAU2* haplotype identified a frameshift variant (rs3431723252) in exon 14 (ENSECAE00000277920) within the predicted *STAU2-206* transcript (ENSECAT00000063627.2). Further examination showed that the frameshift mutation caused a change in amino acid encoding and a premature stop codon, reducing the encoded protein by approximately 4% (Figure 10). The locus of the mutation exhibited a low GERP score (0.05), indicating minimal selective constraint at this position or insufficient alignment depth to get a meaningful estimate of constraint.

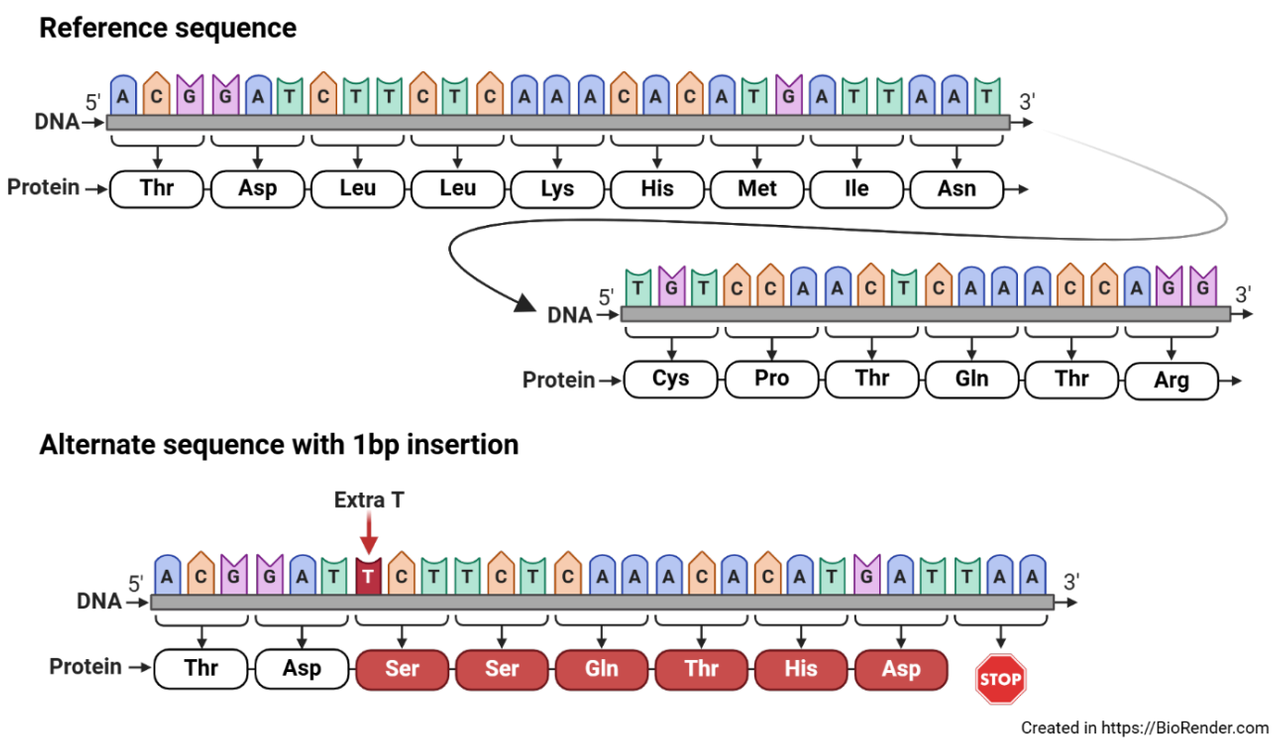


Figure 10. Partial protein alignment of the predicted *STAU2* protein in horses carrying the reference 'S' haplotype and those carrying the alternative 's' haplotype.

Further analysis of the *STAU2* haplotype variants using FAANG data revealed that two intronic variants (rs1149068012 and rs396678054) overlapped with histone H3 lysine 4 monomethylation (H3K4me1) marks in equine brain, adipose, and ovarian tissues.

VEP analysis of the refined *RELN* haplotype identified one synonymous variant (rs393806000) in exon 28 (ENSECAE00000163903). Additionally, small insertions (rs3101690850, rs3434195117, and rs3430280667) were identified flanking this exon 28. GERP scores for these flanking insertion loci indicated weak to moderate evolutionary conservation (0.89-2.72), suggesting these sites may be under relaxed selective pressure or subject to some purifying selection. In contrast, the synonymous variant locus had a negative GERP score (-2.48), indicating potential neutrality or weak negative selection.

Further analysis with FAANG data revealed that over half of the variants within the refined *RELN* haplotype overlapped with a broad histone H3 lysine 27 trimethylation (H3K27me3) modification mark in equine brain tissue.

Among the 39 horses sampled, 29 were homozygous for the 'S' haplotype (SS genotype), while none was homozygous for the 's' haplotype; however, ten horses were heterozygous carriers (Ss genotype). For the *RELN* haplotype, 14 horses were homozygous for the 'R' haplotype (RR genotype), and nine were homozygous for the 'r' haplotype (rr genotype). In evaluating phenotypic differences between genotypes in this relatively small dataset, the significant associations found for pace scores in Paper II were no longer evident for either haplotype region. In fact, no significant differences for any trait were found between the SS genotype group and other horses in the dataset.

However, significant differences were observed between the genotype groups for the refined *RELN* haplotype for tölt, slow tölt, trot, gallop, and canter, with the rr genotype group displaying significantly higher scores for these gait traits compared to the RR genotype group, as also reported in Paper II. Additionally, comparisons of the conformation trait back and croup, and measurements of height at front and back inclination, were also favourable for the rr genotype group. Moreover, horses with the rr genotype were presented at breeding field tests at a significantly younger age than those with the RR genotype, a finding not observed in Paper II.

Allele frequencies for the top SNPs from the GWA study were similar across both Icelandic horse datasets (SNP genotype and sequence data) in this thesis. In other breeds analysed, the *STAU2* frameshift variant and the *STAU2* top SNP showed a pattern consistent with the Icelandic horse – a high frequency of the reference allele and nearly identical allele frequency estimates of the frameshift variant and the top SNP. No notable differences in allele frequencies were found between gaited and non-gaited breeds. In contrast, the top *RELN* SNP exhibited greater variability in allele frequencies across breeds, though this variability seemed to be unrelated to gaiting abilities.

4.4 Population genomic findings

ROH analysis identified several signatures of selection across the Icelandic horse genome (Figure 11). The most notable ROH island, present in over 90% of the individuals, was found on ECA23 in the region where the *DMRT3* gene is located. Additional ROH islands, possibly linked to performance traits, were identified on ECA17 and ECA1.

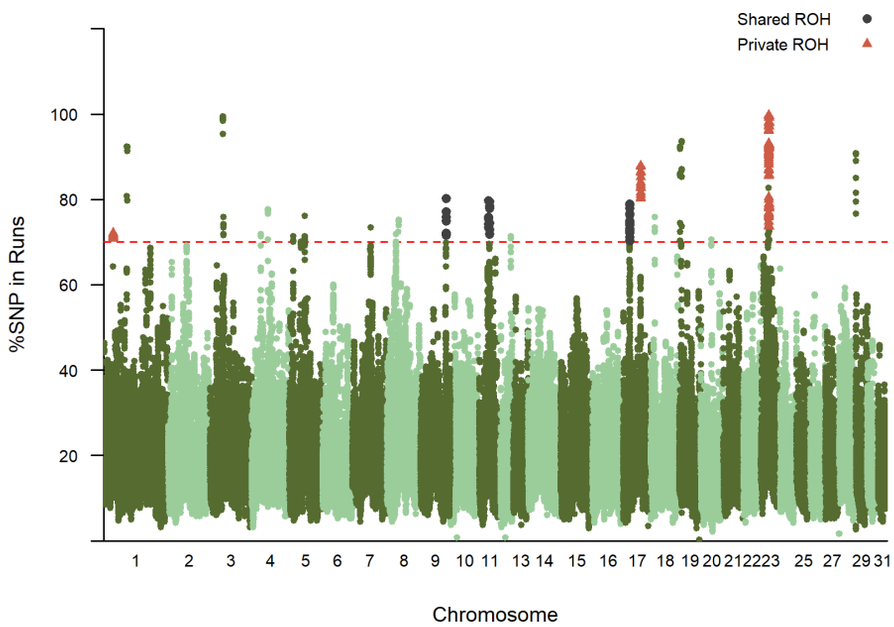


Figure 11. Manhattan plot showing ROH islands across all autosomes in the Icelandic horse. The x-axis represents the chromosome numbers, and the y-axis represents the proportion of horses sharing a ROH. The ROH islands exceeding the 70% threshold (red dotted line) were considered signatures of selection.

Table 3. List of ROH islands in the Icelandic horses with annotated genes located within each ROH island and traits possibly associated with the genes or regions.

ECA	Start-to-end position[a]	Length (kb)	nSNP	Annotated genes within ROH island	Suggested associated trait(s)[b]
1	26,821,929-26,922,176	100.2	19	*SH3PXD2A, NEURL1*	Performance; learning and memory
9	73,198,547-73,286,648	88.1	14	*ENSECAG00000023276, CYRIB*	Immune system
11	25,277,282-25,336,522	59.2	9	*ABI3, ZNF652*	Coat type and body size
	29,120,371-29,172,299	51.9	8	//	
	29,181,352-29,237,132	55.8	5	//	
17	18,706,560-18,829,942	123.4	17	*FOXO1*	Insulin resistance
17	50,290,519-50,523,326	232.8	36	*SLAIN1, EDNRB*	Performance; physiology
23	21,584,553-21,696,531	112	21	*PGM5, ENSECAG00000003227*	Performance; gaits
	21,771,215-21,877,092	105.9	17	*DOCK8*	
	22,117,843-22,706,518	588.7	117	*KANK1, DMRT1, DMRT3, DMRT2*	

ECA = equine chromosome, Length (kb) = length of a ROH in kilobase, nSNP = number of SNP in a ROH. [a] Positions are according to genomic coordinates in EquCab3.0 reference genome. [b] Based on the HorseQTL database and functional annotations

ROH islands, potentially indicative of adaptations to challenging environments, were identified in the Icelandic horse genome and found to overlap partially or entirely with those in the Exmoor pony breed. These shared regions included ROH islands on ECA17, ECA11 and ECA9. Table 3 presents the ROH islands identified in the Icelandic horse genome, the genes found within those ROHs and potential traits associated with these ROHs.

The average F_{ROH} for the studied horses was estimated at 0.20 when including ROH segments as short as 100 kb. Individual F_{ROH} values ranged from 0.07 to 0.30. The highest mean F_{ROH} among the chromosomes was observed on ECA23, while the lowest was found on ECA12 and ECA20. F_{ROH} estimates for different ROH length classes showed that most inbreeding could be traced to the high proportion of ROH (96.4%) found in the shortest length class (0.1 to 1 Mb). In contrast, longer ROH segments (> 4 to ≤ 8 Mb, and > 8 Mb) accounted for only a small portion (0.2%) of the inbreeding. This trend was also observed when limiting the analysis to ROH segments down to at least 500 kb, which reduced the average F_{ROH} estimate to 0.08.

Based on genomic data, the Ne for the most recent generation of Icelandic horses was estimated at approximately 125 individuals. The trend in Ne showed a general decline over the past 60 generations, with more marked reductions occurring approximately 18-23 and 7-8 generations ago (Figure 12). In contrast, the Ne trend stabilized in the most recent 3-4 generations, fluctuating within a narrow range of 123 to 127 individuals. Additionally, the average H_O and H_E were both estimated at 0.34 across the studied dataset, with individual H_O values ranging from 0.30 to 0.38.

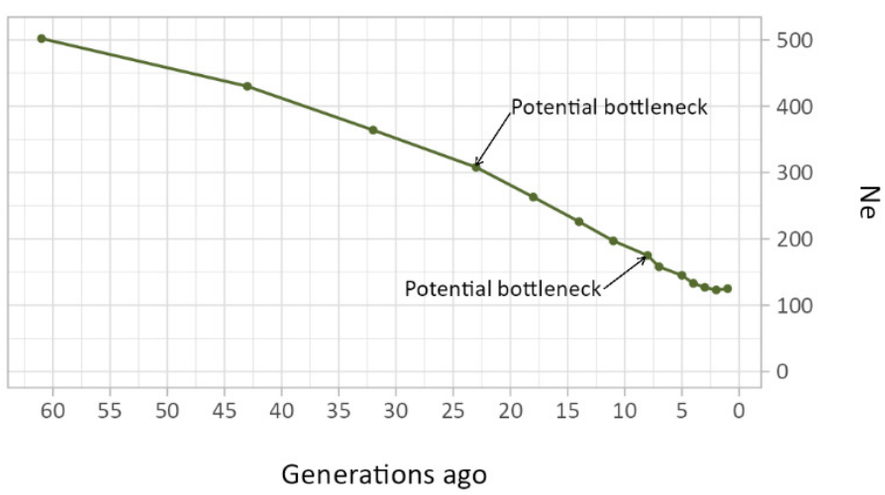


Figure 12. Trends in the effective population size of the Icelandic horse breed over the last 60 generations based on genomic information.

5. Discussion

This thesis sought to identify novel genetic factors with direct and indirect effects on gait and performance in the Icelandic horse, by integrating genomic approaches to examine genetic variation through both genome-wide patterns of homozygosity and polymorphic variation. The findings demonstrate that additional genetic factors beyond the *DMRT3* gene contribute to the regulation of gait ability and quality in Icelandic horses. Through GWAS, QTLs with direct or indirect effects on gait traits were identified. One QTL was associated with breeding scores for back and croup conformation, indirectly influencing breeding scores for the lateral gaits tölt and pace, while two QTLs were directly associated with breeding scores for pace. Further analysis of these three QTLs identified significant haplotypes, where those two detected for pace scores were located within the candidate genes *RELN* and *STAU2*. Sequence data analysis in the latter regions revealed a candidate causal variant in *STAU2* and implicated potential regulatory elements in *RELN*. Additionally, the enduring significance of the genomic region on ECA23, where the *DMRT3* gene is located, was confirmed through genome-wide patterns of homozygosity and additional regions potentially relevant to performance were identified.

5.1 Genomic region associated with back and croup, influencing lateral gait quality

Breeding scores for back and croup reflect a highly complex phenotype that combines subjective assessments of multiple anatomical features, including back inclination, width and muscularity, loin length and width, as well as croup length, slope, form, and muscularity. This complexity makes it challenging to interpret the association identified on ECA22 with this trait. To gain further insight, differences between haplotype groups were examined for related traits, including sub-traits and specific morphological measurements describing some of these anatomical features. This approach enabled a more thorough interpretation, suggesting that the identified QTL may be more specifically associated with backline shape and croup structure. Both of these features are known to significantly influence riding ability in Icelandic horses (Kristjansson et al., 2016), which may partly explain the haplotypes' effects on lateral gaits like tölt and pace.

A well-balanced backline, which is neither sloped forward nor swaybacked, may enhance the back flexibility and the horse's capacity to support a rider's weight, thereby improving both performance and health. Similarly, a well-formed, evenly shaped croup can facilitate greater muscle development, enhancing the horse's pushing power as well as enabling it to carry more weight on its hindquarters (Clayton, 2017). Pushing power is particularly important for speed-related gaits (Crook et al., 2008), such as pace, while carrying ability is considered crucial for gaits like tölt, where it is believed to support proper posture and enhance overall balance. This balance is likely even more critical for Icelandic horses, which are often expected to carry riders who are relatively large compared to the horses' size (Gunnarsson et al., 2017; Stefánsdóttir et al., 2017).

Comparing haplotype groups also revealed differences in breast depth and forelimb length. These two measurements are directly connected, as forelimb length is estimated based on the difference between breast depth and height at withers. While these findings are not directly related to back and croup conformation, they are generally associated with skeletal structure. These traits may also impact the quality of tölt and pace, as longer forelimbs enhance stride length (Hoyt et al., 2000; Pontzer, 2007) and may influence the horse's overall height at front, both of which are important considerations in scoring tölt and pace at breeding field tests (FEIF, 2024).

The haplotypes identified on ECA22 encompass three genes: *CIMIP1* (C22H20orf85), *ANKRD60* and *LOC100056167*. The gene *LOC100056167* is unannotated in the horse genome but shows 84.2% identity to the human pseudogene *PPP4R1L*. Both *CIMIP1* and *ANKRD60* are associated with adolescent idiopathic scoliosis in humans and, together with *PPP4R1L*, have been linked to human body height. This suggests that these genes may play a role in skeletal structure, supporting our findings.

Scoliosis, defined as an abnormal lateral curvature of the spine, is known to cause generalized skeletal muscle weakness, respiratory impairment, and exercise limitations in humans (Martínez-Llorens et al., 2010). As previously noted, breeding scores for back and croup conformation in horses assess both muscular and skeletal characteristics, suggesting that the back and croup phenotype may share certain features with the scoliosis phenotype. Despite these similarities, determining the direct effects of the identified QTL on back and croup requires further investigation. Nonetheless, the effect of this QTL is more likely related to functional advantages or disadvantages affecting movement and back and croup strength in horses, rather than to more severe dysfunctions. This is supported by the relatively high frequency of the 'unfavourable' haplotype among the Icelandic horses in this study. Additionally, when examining allele frequencies of the top SNP in other breeds, the reference allele associated with the 'unfavourable' haplotype shows a generally higher frequency, indicating it may not be unfavourable in other breeds. This suggests that selection for the alternative allele may have occurred in Icelandic horses, possibly due to its positive impact on tölt and pace, traits that are particularly valued in this breed.

5.2 Genomic regions associated with gaits and performance

The haplotypes identified on ECA9 and ECA4, which significantly influence breeding scores for pace, overlapped with the *STAU2* and *RELN* genes. Both genes are expressed in neural tissue, making them likely candidates for influencing the development or the function of the locomotor network and, therefore, affecting gait scores as observed. The involvement of regulatory elements that govern the expression of nearby genes cannot be ruled out, although other genes in the vicinity of the identified regions are either all involved in general cellular processes or not expressed in neural tissue.

5.2.1 The STAU2 gene and its implications with gaits

The *STAU2* gene encodes a protein involved in mRNA transport (Cockburn et al., 2012) and contributes to processes that influence neuron proliferation and fate determination (Chowdhury et al., 2021). Loss of *STAU2* protein affects neuron survival, and mutant mice display reduced motor coordination but enhanced motor learning abilities (Pernice et al., 2019). This suggests that while *STAU2* deficiency impairs basic motor output, it may facilitate learning of movement patterns.

The identified *STAU2* frameshift variant appears to reduce protein production, potentially leading to *STAU2* deficiency. This aligns with our findings, as the 'S' haplotype, representing the reference sequence without the insertion that causes the frameshift and premature stop codon, seems to be favourable for pace, possibly due to enhanced motor coordination. However, it's unclear why enhanced motor coordination would negatively affect scores for trot and gallop, as shown in Paper II, unless the coordination is specific to lateral movements. This is, however, not supported by allele frequencies in other gaited- and non-gaited breeds studied in this thesis.

The negative effects on trot and gallop, paired with positive effects on pace, might suggest a similar mechanism to the influence of the *DMRT3* 'gait keeper' mutation or possibly additive effects between the two genetic factors. However, no clear differences in the *STAU2* allele frequency were found between gaited and non-gaited breeds investigated in this thesis. In fact, the reference allele is common across all other breeds ($A \geq 0.65$), suggesting selection pressure for the reference allele across breeds, regardless of gaiting ability. Given its strong impact on pace, the high frequency of this variant in Icelandic horses is understandable, though its prevalence in other breeds is less straightforward. The comparison with other breeds in this thesis would benefit from a large number of breeds and larger datasets per breed to draw conclusions from this.

The *STAU2* frameshift variant locus appears not to be conserved across species, and it is not annotated in the human (hg38) and mouse (mm39) genomes, pointing to a phenotype that may be somewhat exclusive to equine biology. Given the uncertain role of the *STAU2* frameshift variant as well as the mere predicted existence of the transcript it is annotated in, functional experiments are needed to elucidate the variant's precise function and influence.

5.2.2 The RELN gene's complex role in horse performance

The *RELN* gene encodes an extracellular matrix protein that facilitates cell-to-cell interactions essential for the migration and positioning of neurons during development (Hattori & Kohno, 2021; Lakatosova & Ostatnikova, 2012; Nimura et al., 2019; Vaswani & Blaess, 2016). The *RELN* gene has been widely explored, and several mutations have been reported in mice and rats that are associated with the 'reeler' phenotype (Andersen et al., 2002; Caviness et al., 1972; Falconer, 1951; Kikkawa et al., 2003; Miao et al., 1994), which results in abnormal locomotor behaviours. Furthermore, loss of reelin, the protein produced

by *RELN*, can disrupt neuronal positioning, leading to severe brain abnormalities and neurological disorders in humans (Chen et al., 2017; Folsom & Fatemi, 2013; Ishii et al., 2016).

The *RELN* gene is also implicated in synaptic plasticity and epigenetic regulation of DNA methylation within the adult central nervous system in rats, particularly in the hippocampus in response to environmental stimuli during associative learning (Miller & Sweatt, 2007; Sui et al., 2012). Additionally, *RELN* is highly co-expressed with the *Drd1* gene in mice, which encodes the dopamine receptor D1 in the dorsomedial striatum (de Guglielmo et al., 2022). Equine temperament has previously been linked to a related gene, the dopamine D4 receptor gene (Momozawa et al., 2005).

Based on the findings in Paper II, we hypothesized that the effects of the *RELN* haplotype were related to gaits and locomotion patterns, consistent with the known expression of the *RELN* gene in the developing spinal cord (Krzyzanowska et al., 2020). The two haplotypes had evident contrasting effects on gaits, the 'R' haplotype was associated with higher scores for pace, while the 'r' haplotype was linked to higher scores for tölt, trot, gallop, and canter. Additionally, significant differences in genotype frequencies were observed between five-gaited and four-gaited horses (all with the AA genotype in the *DMRT3* variant), further suggesting that this region may partly explain why some horses with the genetic predisposition to pace lack the capacity to perform pace with ease.

However, it is important to consider the potential bias in the data due to the preselection of horses presented at breeding field tests (Albertsdóttir et al., 2012). The preselection may be even stricter for four-gaited horses, particularly for tölt and trot, to compensate for the absence of a pace score. This is further supported by significant differences found for height of withers and height at front – traits that are known to strongly influence riding ability in Icelandic horses (Kristjansson et al., 2016) – between horses with different genotypes (RR vs. rr genotypes). Given that *RELN* appears to be unlikely to directly influence these musculoskeletal traits, these findings may suggest an indirect effect related to selection practices.

The horses selected for sequencing in Paper III were chosen based on criteria distinguishing elite from non-elite horses, including conformation traits such as back and croup, height at front, and back inclination, traits shown in Paper I to be associated with tölt and pace quality. The selection was, furthermore, based solely on phenotypes, with no prior knowledge of the horses' genotypes.

Interestingly, all horses with the rr genotype (9 horses) were classified as elite, though this may be coincidental given the small dataset of 39 horses. This resulted in significant differences between genotype groups for tölt, slow tölt, trot, gallop, and canter, as well as for the conformation traits used for selection. However, as before, it is likely that these conformation differences identified arise from indirect effects rather than direct effects of the *RELN* gene, suggesting that these distinctions may not be entirely due to

the different *RELN* genotypes. Nevertheless, these are interesting findings, and the effects of the *RELN* gene on the gaits partly support our previous findings in Paper II.

Unfortunately, the significant effect of the *RELN* haplotype on pace was not observed in the dataset containing sequence data. This is most likely due to the small sample which also included four-gaited horses. In total, only five horses with the rr genotype and ten horses with the RR genotype had pace scores above 5.0. Furthermore, all the rr genotype horses were elite, while most of the RR horses were non-elite. These factors likely contributed to the non-significant results for pace, though they were expected given the limited dataset.

Among the sequenced horses in this thesis, rr genotype horses were found to be significantly younger at the time of their breeding field test, a difference not found within the SNP genotype data. However, because all rr horses were elite and, therefore, had favourable conformation, this suggests that *RELN* is not solely responsible for these effects. Still, given *RELN*'s known roles in associative learning through synaptic plasticity and the epigenetic regulation of DNA methylation (Miller & Sweatt, 2007; Sui et al., 2012), along with its co-expression with the *Drd1* gene (de Guglielmo et al., 2022), it is possible that *RELN* affects overall performance through trainability and precocity rather than specific locomotion patterns. This hypothesis is further supported by the overlap of a wide H3K27me3 modification mark, typically associated with gene downregulation, across a large portion of the *RELN* haplotype. However, the observed age differences associated with these genotypes need validation in a larger dataset.

Additionally, if the rr genotype significantly impacted age at assessment, one would expect a higher frequency of the alternative allele (0.55-0.58), included in the r haplotype, unless the favourable effects of the R haplotype on pace counterbalance this trend. Given *RELN*'s diverse biological roles, its influence on performance in Icelandic horses may well involve multiple pathways.

5.2.3 Gene interactions

The findings presented in this thesis suggest interactions between the novel genes *STAU2* and *RELN* and the previously identified candidate gene *DMRT3*. Notably, the similar effects of *STAU2* and *DMRT3*, as discussed before, may indicate additive effects of the two genes. Additionally, *STAU2* seems to partially compensate for the loss of one or two copies of the *DMRT3* 'gait keeper' mutation with respect to lateral movements in tölt and pace, indicating potential epistatic interactions between these genes. The *RELN* gene further demonstrates additive effects with *STAU2* and *DMRT3*, as detailed in Paper II, supported by the joint substantial proportion of phenotypic variance in pace explained by these three genes.

In a thesis focused on identifying genetic factors associated with gaits, the consistent significance of conformation traits – particularly backline and height at the front – on the gait traits and performance merits attention. This emerges both through the impact of QTL associated with back and croup

on lateral gaits and through the ambiguous influence of the *RELN* gene on these conformation traits, despite no clear evidence of interactions between these genetic factors. These findings not only underscore the importance of conformation in horse functionality but also highlight the intricate network of genes influencing horse performance. Furthermore, they emphasize the need to consider these indirect effects of conformation when interpreting direct genetic contributions to gaits and performance.

5.3 Population genomic insights

5.3.1 Performance-related signatures of selection

The most prominent ROH island hot spot in the Icelandic horse was located on ECA23, a region containing key genes like *DMRT3* and *DOCK8*, both of which are known to be causative or associated with gait and performance across various horse breeds, including the Icelandic horse (Andersson et al., 2012; Bas Conn, 2018; Chandra Paul et al., 2020; Jäderkvist Fegraeus et al., 2015; Jäderkvist, Andersson, et al., 2014; Jäderkvist et al., 2015; Jäderkvist, Kangas, et al., 2014; Kristjansson et al., 2014; Novoa-Bravo et al., 2018; Promerová et al., 2014; Ricard, 2015; Velie et al., 2018). The ROH containing *DMRT3* was the longest identified in this study, spanning 589 kb and shared by over 90% of the studied horses. This extensive ROH indicates relatively recent selective pressure in this region, underscoring its significance to the unique gait performance characteristic of the Icelandic breed. This ROH overlaps the previously characterized 'Gait keeper' haplotype, which contains the *KANK1* and *DMRT1-3* genes (Andersson et al., 2012; Staiger et al., 2017). Studies have proposed overlapping or interactive genetic effects between the *DMRT1-3* and *DOCK8* genes (Glessner et al., 2017; Kang et al., 2010; Velie et al., 2018), suggesting that *DOCK8* may contribute to gait performance in conjunction with *DMRT3*.

Another notable ROH island was detected on ECA17, containing the *EDNRB* gene, part of the endothelin gene family. This gene may interact with family members such as *EDN3*, previously associated with blood supply regulation in high-performance racing horses (Fegraeus et al., 2024; Jäderkvist Fegraeus et al., 2018). This ROH island may, therefore, reflect selection for performance-related traits. Another potential performance-linked ROH island was found on ECA1, harbouring the *NEURL1* gene, which is associated with learning and memory processes in mice (Pavlopoulos et al., 2011). This gene could potentially influence training responsiveness in horses.

These ROH islands on ECA23, ECA17, and ECA1 were not present in the Exmoor pony genome, underscoring their potential association with performance traits in Icelandic horses. A more detailed discussion of additional genes located within these ROH islands can be found in Paper IV.

5.3.2 Genomic inbreeding and genetic diversity

The F_{ROH} estimations were significantly influenced by parameter settings, particularly the minimum ROH length. When a 100 kb minimum was used, the average F_{ROH} was estimated at 0.20, while a 500 kb minimum resulted in a much lower average of 0.08, aligning more closely with previous studies, which report values between 0.03 and 0.13 (Colpitts et al., 2022; Cosgrove et al., 2020; Meyermans et al., 2020). However, we ultimately chose the 100 kb threshold for our analysis, as this produced F_{ROH} values with a stronger correlation to F_{PED} values derived from pedigrees with high completeness. Including shorter ROHs thus yielded a more comprehensive estimate, though estimates based on longer ROHs may better capture recent inbreeding levels relevant to current breeding practices (Stoffel et al., 2021; Sumreddee et al., 2021).

To assess the timing of inbreeding events, we divided ROHs into length classes. Most inbreeding appeared in shorter ROHs (0.1 to ≤ 1 Mb), which comprised over 96% of all identified ROHs, likely due to a limited founder population, genetic drift, and purging as the breed adapted. In contrast, ROHs longer than 4 Mb indicated minimal inbreeding, each with F_{ROH} values of 0.01 for the > 4 to ≤ 8 Mb and > 8 Mb classes. The relatively small number of ROHs longer than 4 Mb suggests no evidence of recent excessive inbreeding.

Our findings showed that observed heterozygosity (mean H_O = 0.34) matched the level of expected heterozygosity, suggesting that the genome has remained relatively stable across recent generations, with signs of efforts made to avoid close-relative matings. This stability is consistent with our genomic Ne estimates over the past 3-4 generations, which ranged from 123 to 127 individuals.

Previous pedigree-based estimates for the Icelandic horse have shown a more rapid decrease in Ne over the generations (Kristjansson, 2007; Árnason et al., 1995), not entirely in concordance with the trend of our prediction model seen in Figure 12. While the pedigree completeness for the Icelandic horse is high, pedigree-based estimates rely on probability estimates based on documented ancestry, whereas genomic data – although derived from a genotyped subset of the population – offers a more detailed and comprehensive view of the population's genetic structure and historical dynamics. Some variation between these estimates is, therefore, expected. Nevertheless, the most recent pedigree-based Ne estimate of 100 individuals (Árnadóttir, 2022) closely aligns with our genomic estimate of 125 individuals. Additionally, the average F_{PED} calculated for horses in this study closely matches recent estimates for all horses born in Iceland between 2011 and 2020 (Árnadóttir, 2022). This suggests that our study sample, predominantly consisting of preselected Icelandic breeding horses likely to contribute to future generations, provides a fairly accurate representation of the current generation of Icelandic horses.

Together, these findings – the low occurrence of long ROHs indicating limited recent inbreeding, the alike level of H_E and H_O, and stable Ne estimates over recent generations – suggest a sustainable breeding strategy during the last decades in the modern Icelandic horse breed. However, the increased use of specific breeding animals within the gene pool (Árnadóttir, 2022) underscores the importance of active monitoring of inbreeding and genetic diversity to maintain the sustainability of the breed.

6. Conclusions

- This thesis extends current knowledge on the Icelandic horse genome by demonstrating that additional genetic factors play a role in regulating gait ability and quality beyond the influence of the previously identified *DMRT3* 'gait keeper' mutation.

- A novel QTL associated with the conformation of the back and croup, with a significant impact on the quality of pace and tölt, was identified.

- A novel QTL harbouring the *STAU2* gene and associated with pace quality and ability, as well as the quality of trot and gallop, was identified. A partial loss of the *STAU2* protein, caused by a one-base-pair frameshift insertion in the gene, was suggested to be the causative variant for the observed effects.

- A novel QTL harbouring the *RELN* gene and associated with the quality of pace, tölt, trot, canter, and gallop and potential contribution to precocity, was identified. Regulatory elements located within the region may serve as the causative factors for the observed effects.

- The *RELN* and the *STAU2* genes may partly segregate between four-gaited or five-gaited horses with the same *DMRT3* 'gait keeper' genotype.

- Genetic interactions between *STAU2, RELN* and the *DMRT3* 'gait keeper' mutation were observed.

- The importance of conformation traits – particularly backline and height at front – in regard to riding ability in Icelandic horses was further reinforced.

- ROH analysis confirmed the critical role of the *DMRT3* gene region in governing gaits in the Icelandic horse, while also identifying additional candidate genes that may impact performance.

- Analyses of genomic inbreeding and genetic diversity indicated that breeding practices within the breed are sustainable though continued monitoring is recommended to preserve and maintain the diversity.

- The findings in this thesis may ultimately benefit the breeding of the Icelandic horse and assist breeders in making more informed breeding decisions.

7. Future work

To gain a comprehensive understanding of the biological function of the QTL associated with back and croup conformation and its influence on gait quality, further research is necessary. Future studies should focus on identifying candidate causal variants by integrating sequence data with advanced bioinformatic analyses, as effectively demonstrated for the gait-associated QTLs in this thesis. Additionally, the known functions of the genes within the QTL associated with back and croup are related to muscle and skeletal features in humans, resembling the back and croup phenotype in horses, thereby warranting further functional investigations.

To address the complexities surrounding the roles of the *STAU2* and *RELN* genes, further functional studies are essential. For *STAU2*, particular attention should be given to validating the predicted transcript containing the identified frameshift variant. For *RELN*, epigenetic analyses would be especially valuable to explore its potential role in horse learning behaviour and locomotion. Additionally, gene-gene interactions should be explored further to shed light on the nature and extent of the genetic interplay between the *RELN*, *STAU2*, and *DMRT3* genes. Investigating the compensatory interactions between *STAU2* and *DMRT3* would be of particular interest.

Moreover, special attention should be given to examining the effects of these variants in other breeds, to deepen our understanding and uncover potential patterns in how these variants influence gaits and performance beyond the Icelandic horse.

From a practical perspective, validating the predictive capacity of *STAU2* and *RELN* variants in distinguishing four- and five-gaited horses would provide valuable insights. Furthermore, exploring the *STAU2* and *RELN* profiles in a larger cohort of pace racers could help elucidate the effects of these genes on speed performance. Consequently, incorporating these predictive markers into routine breeding evaluations for the Icelandic horse should be explored to improve the accuracy of selection based on estimated breeding values. As further functional studies are conducted, these findings may also eventually enable the development of genetic tests for causal mutations and variants, providing valuable pre-training insights into horses' biological capabilities.

The data generated in this thesis further provides valuable opportunities for further research. Notably, several intriguing results from the GWA study, including insights into front leg movements, remain unexplored and unpublished. Additionally, the sequence data offers an excellent resource for investigating further the genetic differences in contrasting groups defined by specific traits, such as elite horses vs. non-elite horses and four-gaited vs. five-gaited horses.

Despite the significant progress presented in this thesis, our understanding of the complex genetic basis of gaits in the Icelandic horse is still in its early stages. The large volume of gait and performance data collected for the breed offers wide opportunities for future genetic studies. However, these datasets are limited by a lack of standardized recordings for "average horses" and those with lower-quality traits, due to the stringent preselection of horses attending the breeding field tests. This is particularly true for high-value traits like tölt. Expanding studies to include a more diverse range of horses, including those at the extremes of the quality spectrum, could enhance the identification of additional QTLs and improve statistical power. While recording such data can be costly, advancements in machine learning and large-scale phenotyping technologies may soon make this process more efficient and affordable, ultimately benefitting scientists and the Icelandic horse breeding industry by providing more accessible data for genetic analyses.

In conclusion, this thesis underscores the need for further functional studies to clarify the precise roles of the identified genetic variants, their influence on biological processes, and their broader impact on horse performance beyond the Icelandic horse. These findings may also pave the way for comparative functional genomics, offering potential links between gait and performance traits in horses and similar traits in other mammals.

8. References

Ablondi, M., Dadousis, C., Vasini, M., Eriksson, S., Mikko, S., & Sabbioni, A. (2020). Genetic Diversity and Signatures of Selection in a Native Italian Horse Breed Based on SNP Data. *Animals*, *10*(6), 1005. https://doi.org/10.3390/ani10061005

Ablondi, M., Viklund, Å., Lindgren, G., Eriksson, S., & Mikko, S. (2019). Signatures of selection in the genome of Swedish warmblood horses selected for sport performance. *BMC Genomics*, *20*(1), 717. https://doi.org/10.1186/s12864-019-6079-1

Adalsteinsson, S. (1981). Origin and conservation of farm animal populations in Iceland. *Zeitschrift für Tierzüchtung und Züchtungsbiologie*, *98*(1-4), 258-264.

Al Abri, M. A., Posbergh, C., Palermo, K., Sutter, N. B., Eberth, J., Hoffman, G. E., & Brooks, S. A. (2018). Genome-wide scans reveal a quantitative trait locus for withers height in horses near the ANKRD1 gene. *J Equine Vet Sci*, *60*, 67-73. https://doi.org/10.1016/j.jevs.2017.05.008

Albertsdóttir, E., Eriksson, S., Näsholm, A., Strandberg, E., & Árnason, T. (2008). Genetic correlations between competition traits and traits scored at breeding field-tests in Icelandic horses. *Livest Sci*, *114*(2-3), 181-187. https://doi.org/10.1016/j.livsci.2007.04.022

Albertsdóttir, E., Eriksson, S., Sigurdsson, Á., & Árnason, T. (2011). Genetic analysis of 'breeding field test status' in Icelandic horses. *J Anim Breed Genet*, *128*(2), 124-132. https://doi.org/10.1111/j.1439-0388.2010.00902.x

Albertsdóttir, E., Árnason, T., Eriksson, S., Sigurdsson, Á., & Fikse, W. F. (2012). Effects of integrated genetic evaluations for Icelandic horses on predictive ability, accuracy and selection bias. *J Anim Breed Genet*, *129*(1), 41-49. https://doi.org/10.1111/j.1439-0388.2011.00940.x

Amano, T., Onogi, A., Yamada, F., Kawai, M., Shirai, K., & Ueda, J. (2018). Genome-wide association mapping and examination of possible maternal effect for the pace trait of horses. *Anim Genet*, *49*(5), 461-463. https://doi.org/10.1111/age.12711

Amano, T., Yokawa, H., Masuda, Y., Tozaki, T., Kawai, M., & Shirai, K. (2023). Genome-wide search reveals the uniqueness of DNA regions associated with coat color and innate immunity in Hokkaido Native Horse. *Anim Sci J*, *94*(1), e13884. https://doi.org/10.1111/asj.13884

Andersen, T. E., Finsen, B., Goffinet, A. M., Issinger, O-G., & Boldyreff, B. (2002). A reeler mutant mouse with a new, spontaneous mutation in the reelin gene. *Mol Brain Res*, *105*(1), 153-156. https://doi.org/10.1016/S0169-328X(02)00389-3

Andersson, L. S., Larhammar, M., Memic, F., Wootz, H., Schwochow, D., Rubin, C-J., …Hjälm, G. (2012). Mutations in *DMRT3* affect locomotion in horses and spinal circuit function in mice. *Nature*, *488*(7413), 642-646. https://doi.org/10.1038/nature11399

Arnórsson, K. (2006). Ræktunin (e. The Breeding). In G. B. Björnsson & H. J. Sveinsson (Eds.), Íslenski hesturinn *(e. The Icelandic Horse)* (2nd ed., pp. 202-247). Mál og menning.

Aulchenko, Y. S., Ripke, S., Isaacs, A., & Van Duijn, C. M. (2007). GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, *23*(10), 1294-1296. https://doi.org/10.1093/bioinformatics/btm108

Barrey, E. (2013). Gaits and interlimb coordination. In W. Back & H. M. Clayton (Eds.), *Equine locomotion* (2nd ed., pp. 85-97). Elsevier.

Bas Conn, L. (2018, Autumn). *The role of polymorphisms of the MSTN, GRIN2B and DOCK8 genes in the performance of pace-racing Icelandic horses* [Master's thesis, Swedish University of Agricultural Sciences]. Epsilon Archive for Student Projects. http://urn.kb.se/resolve?urn=urn:nbn:se:slu:epsilon-s-10183

Binns, M. M., Boehler, D. A., & Lambert, D. H. (2010). Identification of the myostatin locus (*MSTN*) as having a major effect on optimum racing distance in the Thoroughbred horse in the USA. *Anim Genet*, *41*(s2), 154-158. https://doi.org/10.1111/j.1365-2052.2010.02126.x

Bizarria dos Santos, W., Pimenta Schettini, G., Fonseca, M. G., Pereira, G. L., Loyola Chardulo, L. A., Rodrigues Machado Neto, O., …Abdallah Curi, R. (2021). Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the Mangalarga Marchador horse breed. *J Anim Breed Genet*, *138*(2), 161-173. https://doi.org/10.1111/jbg.12508

Björnsson, G. B., & Sveinsson, H. J. (2006). Landneminn (e. The Settler). In G. B. Björnsson & H. J. Sveinsson (Eds.), Íslenski hesturinn *(e. The Icelandic Horse)* (2nd ed., pp. 18-33). Mál og menning.

Björnsson, G. B., & Sveinsson, H. J. (2006). Á spjöldum sögunnar (e. The history of the Icelandic horse). In G. B. Björnsson & H. J. Sveinsson (Eds.), Íslenski hesturinn *(e. The Icelandic Horse)* (2nd ed., pp. 76-101). Mál og menning.

Bjørnstad, G., Nilsen, N., & Røed, K. (2003). Genetic relationship between Mongolian and Norwegian horses? *Anim Genet*, *34*(1), 55-58. https://doi.org/10.1046/j.1365-2052.2003.00922.x

Bjørnstad, G., & Røed, K. (2001). Breed demarcation and potential for breed allocation of horses assessed by microsatellite markers. *Anim Genet*, *32*(2), 59-65. https://doi.org/10.1046/j.1365-2052.2001.00705.x

Bonow, S., Eriksson, S., Strandberg, E., Thorén Hellsten, E., & Gelinder Viklund, Å. (2024). Phenotypic associations between linearly scored traits and sport performance in the Swedish Warmblood horse population. *Livest Sci*, *282*, 105438. https://doi.org/10.1016/j.livsci.2024.105438

Boucher, B., & Jenna, S. (2013). Genetic interaction networks: better understand to better predict. *Front Genet*, *4*, Article 290. https://doi.org/10.3389/fgene.2013.00290

Carlborg, Ö., & Haley, C. S. (2004). Epistasis: too often neglected in complex trait studies? *Nat Rev Genet*, *5*(8), 618-625. https://doi.org/10.1038/nrg1407

Caviness, V. S., Jr., So, D. K., & Sidman, R. L. (1972). The hybrid Reeler mouse. *J Hered*, *63*(5), 241-246. https://doi.org/10.1093/oxfordjournals.jhered.a108286

Chandra Paul, R., Ba Nguyen, T., Okuda, Y., Nu Anh Le, T., Mosese Dau Tabuyaqona, J., Konishi, Y., …Kunieda, T. (2020). Distribution of the mutant allele of the *DMRT3* gene associated with ambling gaits in Japanese native horse populations. *Anim Sci J*, *91*(1), e13431. https://doi.org/10.1111/asj.13431

Chen, N., Bao, Y., Xue, Y., Sun, Y., Hu, D., Meng, S., …Shi, J. (2017). Meta-analyses of RELN variants in neuropsychiatric disorders. *Behav Brain Res*, *332*, 110-119. https://doi.org/10.1016/j.bbr.2017.05.028

Chen, X., Jorgenson, E., & Cheung, S. T. (2009). New tools for functional genomic analysis. *Drug Discov Today*, *14*(15), 754-760. https://doi.org/10.1016/j.drudis.2009.05.005

Chowdhury, R., Wang, Y., Campbell, M., Goderie, S. K., Doyle, F., Tenenbaum, S. A., …Temple, S. (2021). STAU2 binds a complex RNA cargo that changes temporally with production of diverse intermediate progenitor cells during mouse corticogenesis. *Development*, *148*(15). https://doi.org/10.1242/dev.199376

Clayton, H. M. (1994a). Comparison of the collected, working, medium and extended canters. *Equine Vet J*, *26*(S17), 16-19. https://doi.org/10.1111/j.2042-3306.1994.tb04866.x

Clayton, H. M. (1994b). Comparison of the stride kinematics of the collected, working, medium and extended trot in horses. *Equine Vet J*, *26*(3), 230-234. https://doi.org/10.1111/j.2042-3306.1994.tb04375.x

Clayton, H. M. (1995). Comparison of the stride kinematics of the collected, medium, and extended walks in horses. *Am J Vet Res*, *56*(7), 849-852. https://doi.org/10.2460/ajvr.1995.56.07.849

Clayton, H. M. (2017, March 20). *The anatomy of the dressage horse hindquarters*. Equisearch. Retrieved December 1, 2024, from https://equisearch.com/articles/the-anatomy-of-dressage-horse-hindquarters/

Cockburn, D. M., Charish, J., Tassew, N. G., Eubanks, J., Bremner, R., Macchi, P., & Monnier, P. P. (2012). The double-stranded RNA-binding protein Staufen 2 regulates eye size. *Mol Cell Neurosci*, *51*(3), 101-111. https://doi.org/10.1016/j.mcn.2012.08.008

Colpitts, J., McLoughlin, P. D., & Poissant, J. (2022). Runs of homozygosity in Sable Island feral horses reveal the genomic consequences of inbreeding and divergence from domestic breeds. *BMC Genomics*, *23*(1), 501. https://doi.org/10.1186/s12864-022-08729-9

Cook, D., Gallagher, P. C., & Bailey, E. (2010). Genetics of swayback in American Saddlebred horses. *Anim Genet*, *41*(s2), 64-71. https://doi.org/10.1111/j.1365-2052.2010.02108.x

Cosgrove, E. J., Sadeghi, R., Schlamp, F., Holl, H. M., Moradi-Shahrbabak, M., Miraei-Ashtiani, S. R., …Brooks, S. A. (2020). Genome diversity and the origin of the Arabian horse. *Sci Rep-UK*, *10*(1), 9702. https://doi.org/10.1038/s41598-020-66232-1

Crook, T. C., Cruickshank, S. E., McGowan, C. M., Stubbs, N., Wakeling, J. M., Wilson, A. M., Payne, R. C. (2008). Comparative anatomy and muscle architecture of selected hind limb muscles in the Quarter Horse and Arab. *J Anat, 212*(2), 144-152. https://doi.org/10.1111/j.1469-7580.2007.00848.x

Curik, I., Ferenčaković, M., & Sölkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest Sci*, *166*, 26-34. https://doi.org/10.1016/j.livsci.2014.05.034

de Guglielmo, G., Iemolo, A., Nur, A., Turner, A., Montilla-Perez, P., Martinez, A.,…Telese, F. (2022). Reelin deficiency exacerbates cocaine-induced hyperlocomotion by enhancing neuronal activity in the dorsomedial striatum. *Genes Brain Behav*, *21*(7), e12828. https://doi.org/10.1111/gbb.12828

Deuel, N., & Lawrence, L. (1986). Gallop velocity and limb contact variables of quarter horses. *J Equine Vet Sci*, *6*(3), 143-147. https://doi.org/10.1016/S0737-0806(86)80064-8

Druml, T., Dobretsberger, M., & Brem, G. (2015). The use of novel phenotyping methods for validation of equine conformation scoring results. *animal*, *9*(6), 928-937. https://doi.org/10.1017/S1751731114003309

Duggal, P., Gillanders, E. M., Holmes, T. N., & Bailey-Wilson, J. E. (2008). Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome-wide association studies. *BMC Genomics*, *9*(1), 1-8. https://doi.org/10.1186/1471-2164-9-516

Falconer, D. S. (1951). Two new mutants, 'trembler' and 'reeler', with neurological actions in the house mouse (Mus musculus L.). *J Genet*, *50*(2), 192-205. https://doi.org/10.1007/BF02996215

Fang, G., Wang, W., Paunic, V., Heydari, H., Costanzo, M., Liu, X., …Myers, C. L. (2019). Discovering genetic interactions bridging pathways in genome-wide association studies. *Nat Commun*, *10*(1), 4274. https://doi.org/10.1038/s41467-019-12131-7

Fegraeus, K., Rosengren, M. K., Naboulsi, R., Orlando, L., Åbrink, M., Jouni, A., …Lindgren, G. (2024). An endothelial regulatory module links blood pressure regulation with elite athletic performance. *PLoS Genet*. https://doi.org/10.1371/journal.pgen.1011285

FEIF – International Federation of Icelandic Horse Associations. (2024, April). *FEIF General Rules and Regulations*. https://www.feiffengur.com/documents/FEIF%20Breeding%202024_web.pdf

Folsom, T. D., & Fatemi, S. H. (2013). The involvement of reelin in neurodevelopmental disorders. *Neuropharmacology*, *68*, 122-135. https://doi.org/10.1016/j.neuropharm.2012.08.015

Forsberg, S. K. G., Bloom, J. S., Sadhu, M. J., Kruglyak, L., & Carlborg, Ö. (2017). Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. *Nat Genet*, *49*(4), 497-503. https://doi.org/10.1038/ng.3800

François, L., Fegraeus, K. J., Eriksson, S., Andersson, L. S., Tesfayonas, Y. G., Viluma, A., …Velie, B. D. (2016). Conformation traits and gaits in the Icelandic horse are associated with genetic variants in Myostatin (*MSTN*). *J Hered*, *107*(5), 431-437. https://doi.org/10.1093/jhered/esw031

Frischknecht, M., Signer-Hasler, H., Leeb, T., Rieder, S., & Neuditschko, M. (2016). Genome-wide association studies based on sequence-derived genotypes reveal new QTL associated with conformation and performance traits in the Franches-Montagnes horse breed. *Anim Genet*, *47*(2), 227-229. https://doi.org/10.1111/age.12406

Giuffra, E., Tuggle, C. K., & Consortium, F. (2019). Functional Annotation of Animal Genomes (FAANG): current achievements and roadmap. *Annu Rev Anim Biosci*, *7*, 65-88. https://doi.org/10.1146/annurev-animal-020518-114913

Glessner, J. T., Li, J., Wang, D., March, M., Lima, L., Desai, A., …the Janssen-CHOP Neuropsychiatric Genomics Working Group (2017). Copy number variation meta-analysis reveals a novel duplication at 9p24 associated with multiple neurodevelopmental disorders. *Genome Med*, *9*(1), 106. https://doi.org/10.1186/s13073-017-0494-1

Gmel, A. I., Brem, G., & Neuditschko, M. (2023). New genomic insights into the conformation of Lipizzan horses. *Sci Rep-UK*, *13*(1), 8990. https://doi.org/10.1038/s41598-023-36272-4

Gmel, A. I., Burren, A., & Neuditschko, M. (2022). Estimates of genetic parameters for shape space data in Franches-Montagnes horses. *Animals*, *12*(17).

Gmel, A. I., Druml, T., von Niederhäusern, R., Leeb, T., & Neuditschko, M. (2019). Genome-wide association studies based on equine joint angle measurements reveal new QTL affecting the conformation of horses. *Genes*, *10*(5).

Gong, Y., Li, Y., Liu, X., Ma, Y., & Jiang, L. (2023). A review of the pangenome: how it affects our understanding of genomic variation, selection and breeding in domestic animals? *J Anim Sci Biotechnol*, *14*(1), 73. https://doi.org/10.1186/s40104-023-00860-1

Gorssen, W., Meyermans, R., Janssens, S., & Buys, N. (2021). A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species. *Genet Sel Evol*, *53*, 1-10. https://doi.org/10.1186/s12711-020-00599-7

Grilz-Seger, G., Druml, T., Neuditschko, M., Dobretsberger, M., Horna, M., & Brem, G. (2019). High-resolution population structure and runs of homozygosity reveal the genetic architecture of complex traits in the Lipizzan horse. *BMC Genomics*, *20*(1), 174. https://doi.org/10.1186/s12864-019-5564-x

Grilz-Seger, G., Druml, T., Neuditschko, M., Mesarič, M., Cotman, M., & Brem, G. (2019). Analysis of ROH patterns in the Noriker horse breed reveals signatures of selection for coat color and body size. *Anim Genet*, *50*(4), 334-346. https://doi.org/10.1111/age.12797

Grilz-Seger, G., Mesarič, M., Cotman, M., Neuditschko, M., Druml, T., & Brem, G. (2018). Runs of homozygosity and population history of three horse breeds with small population size. *J Equine Vet Sci*, *71*, 27-34. https://doi.org/10.1016/j.jevs.2018.09.004

Grilz-Seger, G., Neuditschko, M., Ricard, A., Velie, B., Lindgren, G., Mesarič, M., …Druml, T. (2019). Genome-wide homozygosity patterns and evidence for selection in a set of European and Near Eastern horse breeds. *Genes*, *10*(7), 491. https://doi.org/10.3390/genes10070491

Gu, J., MacHugh, D. E., McGivney, B. A., Park, S. D. E., Katz, L. M., & Hill, E. W. (2010). Association of sequence variants in *CKM* (creatine kinase, muscle) and *COX4I2* (cytochrome c oxidase, subunit 4, isoform 2) genes with racing performance in Thoroughbred horses. *Equine Vet J*, *42*(s38), 569-575. https://doi.org/10.1111/j.2042-3306.2010.00181.x

Gunnarsson, V., Stefánsdóttir, G. J., Jansson, A., & Roepstorff, L. (2017). The effect of rider weight and additional weight in Icelandic horses in tölt: part II. Stride parameters responses. *animal*, *11*(9), 1567-1572. https://doi.org/10.1017/S1751731117000568

Guðlaugsson, Þ. (2006). Á erlendri grundu (e. On foreign grounds). In G. B. Björnsson & J. H. Sveinsson (Eds.), Íslenski hesturinn *(e. The Icelandic horse)* (2nd ed., pp. 285-323). Mál og menning.

Han, H., McGivney, B. A., Allen, L., Bai, D., Corduff, L. R., Davaakhuu, G., …Hemmings, A. J. (2022). Common protein-coding variants influence the racing phenotype in galloping racehorse breeds. *Commun Biol*, *5*(1), 1320. https://doi.org/10.1038/s42003-022-04206-x

Harrison, P. W., Amode, M. R., Austine-Orimoloye, O., Azov, A. G., Barba, M., Barnes, I.,…Yates, A. D. (2024). Ensembl 2024. *Nucleic Acids Res*, *52*(D1), D891-D899. https://doi.org/10.1093/nar/gkad1049

Hattori, M., & Kohno, T. (2021). Regulation of Reelin functions by specific proteolytic processing in the brain. *J Biochem*, *169*(5), 511-516. https://doi.org/10.1093/jb/mvab015

He, S., Zhang, L., Li, W., & Liu, M. (2015). BIEC2-808543 SNP in the LCORL gene is associated with body conformation in the Yili horse. *Anim Biotechnol*, *26*(4), 289-291. https://doi.org/10.1080/10495398.2014.995303

Hendricks, B. L. (2007). *International encyclopedia of horse breeds* (1st ed.). University of Oklahoma Press.

Hill, E. W., Eivers, S. S., McGivney, B. A., Fonseca, R. G., Gu, J., Smith, N. A.,…Katz, L. M. (2010). Moderate and high intensity sprint exercise induce differential responses in *COX4I2* and *PDK4* gene expression in Thoroughbred horse skeletal muscle. *Equine Vet J*, *42*(s38), 576-581. https://doi.org/10.1111/j.2042-3306.2010.00206.x

Hill, E. W., Fonseca, R. G., McGivney, B. A., Gu, J., MacHugh, D. E., & Katz, L. M. (2011). *MSTN* genotype (g.66493737C/T) association with speed indices in Thoroughbred racehorses. *J Appl Physiol*, *112*(1), 86-90. https://doi.org/10.1152/japplphysiol.00793.2011

Hill, E. W., Gu, J., Eivers, S. S., Fonseca, R. G., McGivney, B. A., Govindarajan, P., …MacHugh, D. (2010). A sequence polymorphism in *MSTN* predicts sprinting ability and racing stamina in Thoroughbred horses. *PLoS One*, *5*(1), e8645. https://doi.org/10.1371/journal.pone.0008645

Hill, E. W., Gu, J., McGivney, B. A., & MacHugh, D. E. (2010). Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. *Anim Genet*, *41*(s2), 56-63. https://doi.org/10.1111/j.1365-2052.2010.02104.x

Hill, E. W., McGivney, B. A., Gu, J., Whiston, R., & MacHugh, D. E. (2010). A genome-wide SNP-association study confirms a sequence variant (g.66493737C>T) in the equine myostatin (*MSTN*) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. *BMC Genomics*, *11*(1), 552. https://doi.org/10.1186/1471-2164-11-552

Hill, E. W., McGivney, B. A., Rooney, M. F., Katz, L. M., Parnell, A., & MacHugh, D. E. (2019). The contribution of myostatin (*MSTN*) and additional modifying genetic loci to race distance aptitude in Thoroughbred horses racing in different geographic regions. *Equine Vet J*, *51*(5), 625-633. https://doi.org/10.1111/evj.13058

Holmström, M., & Back, W. (2013). The effects of conformation. In W. Back & H. M. Clayton (Eds.), *Equine locomotion* (2nd ed., pp. 229-244). Elsevier.

Holtby, A. R., Hall, T. J., McGivney, B. A., Han, H., Murphy, K. J., MacHugh, D. E., …Hill, E. W. (2023). Integrative genomics analysis highlights functionally relevant genes for equine behaviour. *Anim Genet*, *54*(4), 457-469. https://doi.org/10.1111/age.13320

Hori, Y., Ozaki, T., Yamada, Y., Tozaki, T., Kim, H.-S., Takimoto, A.,…Fujita, K. (2013). Breed differences in dopamine receptor D4 gene (*DRD4*) in horses. *J Equine Sci*, *24*(3), 31-36. https://doi.org/10.1294/jes.24.31

Hoyt, D. F., Wickler, S. J., & Cogger, E. A. (2000). Time of contact and step length: the effect of limb length, running speed, load carrying and incline. *J Exp Biol*, *203*(2), 221-227. https://doi.org/10.1242/jeb.203.2.221

Hreiðarsdóttir, G. E., Árnason, Þ., Svansson, V., & Hallsson, J. H. (2014). Analysis of the history and population structure of the Icelandic horse using pedigree data and DNA analyses. *Icel Agric Sci*, *27*, 63-79.

Hu, Z.-L., Park, C. A., & Reecy, J. M. (2021). Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. *Nucleic Acids Res*, *50*(D1), D956-D961. https://doi.org/10.1093/nar/gkab1116

Hugason, K. (1994). Breeding of Icelandic toelter horses: an overview. *Livest Prod Sci*, *40*(1), 21-29. https://doi.org/10.1016/0301-6226(94)90262-3

Ishii, K., Kubo, K-i., & Nakajima, K. (2016). Reelin and neuropsychiatric disorders. *Front Cell Neurosci*, *10*. https://doi.org/10.3389/fncel.2016.00229

Jorgenson, E., & Witte, J. S. (2006). Coverage and power in genome-wide association studies. *Am J Hum Genet*, *78*(5), 884-888. https://doi.org/10.1086/503751

Jäderkvist Fegraeus, K., Hirschberg, I., Árnason, T., Andersson, L., Velie, B., Andersson, L., & Lindgren, G. (2017). To pace or not to pace: a pilot study of four-and five-gaited Icelandic horses homozygous for the *DMRT3* 'Gait Keeper'mutation. *Anim Genet*, *48*(6), 694-697. https://doi.org/10.1111/age.12610

Jäderkvist Fegraeus, K., Johansson, L., Mäenpää, M., Mykkänen, A., Andersson, L. S., Velie, B. D., …Lindgren, G. (2015). Different *DMRT3* genotypes are best adapted for harness racing and riding in Finnhorses. *J Hered*, *106*(6), 734-740. https://doi.org/10.1093/jhered/esv062

Jäderkvist Fegraeus, K., Lawrence, C., Petäjistö, K., Johansson, M. K., Wiklund, M., Olsson, C., …Ihler, C-F. (2017). Lack of significant associations with early career performance suggest no link between the *DMRT3* "Gait Keeper" mutation and precocity in Coldblooded trotters. *PLoS One*, *12*(5), e0177351. https://doi.org/10.1371/journal.pone.0177351

Jäderkvist Fegraeus, K., Velie, B. D., Axelsson, J., Ang, R., Hamilton, N. A., Andersson, L., …Lindgren, G. (2018). A potential regulatory region near the EDN3 gene may control both harness racing performance and coat color variation in horses. *Physiol Rep*, *6*(10), e13700. https://doi.org/10.14814/phy2.13700

Jäderkvist, K., Andersson, L. S., Johansson, A. M., Árnason, T., Mikko, S., Eriksson, S., …Lindgren, G. (2014). The DMRT3 'gait keeper' mutation affects performance of Nordic and Standardbred trotters. *J Anim Sci*, *92*(10), 4279-4286. https://doi.org/10.2527/jas.2014-7803

Jäderkvist, K., Holm, N., Imsland, F., Árnason, T., Andersson, L., Andersson, L. S., & Lindgren, G. (2015). The importance of the *DMRT3* 'Gait keeper' mutation on riding traits and gaits in Standardbred and Icelandic horses. *Livest Sci*, *176*, 33-39. https://doi.org/10.1016/j.livsci.2015.03.025

Jäderkvist, K., Kangas, N., Andersson, L. S., & Lindgren, G. (2014). Gaitedness is associated with the *DMRT3* 'Gait keeper' mutation in Morgan and American Curly horses. *Anim Genet*, *45*(6), 908-909. https://doi.org/10.1111/age.12228

Jönsson, L., Egenvall, A., Roepstorff, L., Näsholm, A., Dalin, G., & Philipsson, J. (2014). Associations of health status and conformation with longevity and lifetime competition performance in young Swedish

Warmblood riding horses: 8,238 cases (1983–2005). *J Am Vet Med Assoc*, *244*(12), 1449-1461. https://doi.org/10.2460/javma.244.12.1449

Kalbfleisch, T. S., Rice, E. S., DePriest Jr, M. S., Walenz, B. P., Hestand, M. S., Vermeesch, J. R., …Saremi, N. F. (2018). Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun Biol*, *1*(1), 197. https://doi.org/10.1038/s42003-018-0199-z

Kang, J. U., Koo, S. H., Kwon, K. C., & Park, J. W. (2010). Frequent silence of chromosome 9p, homozygous DOCK8, DMRT1 and DMRT3 deletion at 9p24.3 in squamous cell carcinoma of the lung. *Int J Oncol*, *37*(2), 327-335. https://doi.org/10.3892/ijo_00000681

Kikkawa, S., Yamamoto, T., Misaki, K., Ikeda, Y., Okado, H., Ogawa, M., …Terashima, T. (2003). Missplicing resulting from a short deletion in the reelin gene causes reeler-like neuronal disorders in the mutant shaking rat Kawasaki. *J Comp Neurol*, *463*(3), 303-315. https://doi.org/10.1002/cne.10761

Koenen, E. P. C., van Veldhuizen, A. E., & Brascamp, E. W. (1995). Genetic parameters of linear scored conformation traits and their relation to dressage and show-jumping performance in the Dutch Warmblood Riding Horse population. *Livest Prod Sci*, *43*(1), 85-94. https://doi.org/10.1016/0301-6226(95)00010-I

Kristjansson, T. (2007). *Erfðaafjölbreytileiki íslenska hrossastofnsins og verndun hans* [Conference session]. Íslensk búfjárrækt: Málstofa til heiðurs Hjalta Gestssyni níræðum, Reykjavík, Iceland.

Kristjansson, T., Bjornsdottir, S., Albertsdóttir, E., Sigurdsson, A., Pourcelot, P., Crevier-Denoix, N., & Arnason, T. (2016). Association of conformation and riding ability in Icelandic horses. *Livest Sci*, *189*, 91-101. https://doi.org/10.1016/j.livsci.2016.05.010

Kristjansson, T., Bjornsdottir, S., Sigurdsson, A., Andersson, L., Lindgren, G., Helyar, S., …Arnason, T. (2014). The effect of the 'Gait keeper' mutation in the *DMRT3* gene on gaiting ability in Icelandic horses. *J Anim Breed Genet*, *131*(6), 415-425. https://doi.org/10.1111/jbg.12112

Kristjansson, T., Bjornsdottir, S., Sigurdsson, A., Crevier-Denoix, N., Pourcelot, P., & Arnason, T. (2013). Objective quantification of conformation of the Icelandic horse based on 3-D video morphometric measurements. *Livest Sci*, *158*(1-3), 12-23. https://doi.org/10.1016/j.livsci.2013.09.019

Krzyzanowska, A., Cabrerizo, M., Clascá, F., & Ramos-Moreno, T. (2020). Reelin immunoreactivity in the adult spinal cord: a comparative study in rodents, carnivores, and non-human primates. *Front Neuroanat*, *13*. https://doi.org/10.3389/fnana.2019.00102

Lakatosova, S., & Ostatnikova, D. (2012). Reelin and its complex involvement in brain development and function. *Int J Biochem Cell B*, *44*(9), 1501-1504. https://doi.org/10.1016/j.biocel.2012.06.002

Lander, E., & Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet*, *11*(3), 241-247.

Liu, S., Jiang, S., Dong, X. G., Cui, R., Ling, Y., & Zhao, C. (2020). Novel variants in the *HMGA2* gene are associated with withers height in Debao pony. *J Equine Vet Sci*, *88*, 102948. https://doi.org/10.1016/j.jevs.2020.102948

Liu, X., Zhang, Y., Liu, W., Li, Y., Pan, J., Pu, Y., …Jiang, L. (2022). A single-nucleotide mutation within the *TBX3* enhancer increased body size in Chinese horses. *Curr Biol*, *32*(2), 480-487. https://doi.org/10.1016/j.cub.2021.11.052

Lorange, J. B. (2011). WorldFengur – the studbook of origin for the Icelandic horse. *Acta Vet Scand*, *53*(1), S5. https://doi.org/10.1186/1751-0147-53-S1-S5

Mackay, T. F. C. (2014). Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet*, *15*(1), 22-33. https://doi.org/10.1038/nrg3627

Makvandi-Nejad, S., Hoffman, G. E., Allen, J. J., Chu, E., Gu, E., Chandler, A. M.,…Brooks, S. A. (2012). Four loci explain 83% of size variation in the horse. *PLoS One*, *7*(7), e39929. https://doi.org/10.1371/journal.pone.0039929

Martínez-Llorens, J., Ramírez, M., Colomina, M., Bagó, J., Molina, A., Cáceres, E., & Gea, J. (2010). Muscle dysfunction and exercise limitation in adolescent idiopathic scoliosis. *Eur Respir J*, *36*(2), 393-400. https://doi.org/10.1183/09031936.00025509

Mazitov, T., Bregin, A., Philips, M.-A., Innos, J., & Vasar, E. (2017). Deficit in emotional learning in neurotrimin knockout mice. *Behav Brain Res*, *317*, 311-318. https://doi.org/10.1016/j.bbr.2016.09.064

McBride, S. D., & Mills, D. S. (2012). Psychological factors affecting equine performance. *BMC Vet Res*, *8*(1), 180. https://doi.org/10.1186/1746-6148-8-180

McCoy, A. M., Beeson, S. K., Rubin, C.-J., Andersson, L., Caputo, P., Lykkjen, S., …McCue, M. E. (2019). Identification and validation of genetic variants predictive of gait in standardbred horses. *PLoS Genet*, *15*(5), e1008146. https://doi.org/10.1371/journal.pgen.1008146

McCue, M. E., Bannasch, D. L., Petersen, J. L., Gurr, J., Bailey, E., Binns, M. M., …Mickelson, J. R. (2012). A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genet*, *8*(1), e1002451. https://doi.org/10.1371/journal.pgen.1002451

McGivney, B., Hernandez, B., Katz, L., MacHugh, D., McGovern, S., Parnell, A., …Hill, E. (2019). A genomic prediction model for racecourse starts in the Thoroughbred horse. *Anim Genet*, *50*(4), 347-357. https://doi.org/10.1111/age.12798

McGivney, B. A., Browne, J. A., Fonseca, R. G., Katz, L. M., MacHugh, D. E., Whiston, R., & Hill, E. W. (2012). *MSTN* genotypes in Thoroughbred horses influence skeletal muscle gene expression and racetrack performance. *Anim Genet*, *43*(6), 810-812. https://doi.org/10.1111/j.1365-2052.2012.02329.x

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., …Tenesa, A. (2008). Runs of homozygosity in European populations. *Am J Hum Genet*, *83*(3), 359-372. https://doi.org/10.1016/j.ajhg.2008.08.007

Metzger, J., Rau, J., Naccache, F., Bas Conn, L., Lindgren, G., & Distl, O. (2018). Genome data uncover four synergistic key regulators for extremely small body size in horses. *BMC Genomics*, *19*(1), 492. https://doi.org/10.1186/s12864-018-4877-5

Metzger, J., Schrimpf, R., Philipp, U., & Distl, O. (2013). Expression levels of *LCORL* are associated with body size in horses. *PLoS One*, *8*(2), e56497. https://doi.org/10.1371/journal.pone.0056497

Meyermans, R., Gorssen, W., Buys, N., & Janssens, S. (2020). How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics*, *21*(1), 1-14.

Miao, G. G., Smeyne, R. J., D'Arcangelo, G., Copeland, N. G., Jenkins, N. A., Morgan, J. I., & Curran, T. (1994). Isolation of an allele of reeler by insertional mutagenesis. *P Natl Acad Sci USA*, *91*(23), 11050-11054. https://doi.org/10.1073/pnas.91.23.11050

Miller, C. A., & Sweatt, J. D. (2007). Covalent modification of DNA regulates memory formation. *Neuron*, *53*(6), 857-869. https://doi.org/10.1016/j.neuron.2007.02.022

Momozawa, Y., Takeuchi, Y., Kusunose, R., Kikusui, T., & Mori, Y. (2005). Association between equine temperament and polymorphisms in dopamine D4 receptor gene. *Mamm Genome*, *16*(7), 538-544. https://doi.org/10.1007/s00335-005-0021-3

Momozawa, Y., Takeuchi, Y., Tozaki, T., Kikusui, T., Hasegawa, T., Raudsepp, T., …Mori, Y. (2007). SNP detection and radiation hybrid mapping in horses of nine candidate genes for temperament. *Anim Genet*, *38*(1), 81-83. https://doi.org/10.1111/j.1365-2052.2006.01541.x

Mousavi, S. F., Razmkabir, M., Rostamzadeh, J., Seyedabadi, H.-R., Naboulsi, R., Petersen, J. L., & Lindgren, G. (2023). Genetic diversity and signatures of selection in four indigenous horse breeds of Iran. *Heredity*, *131*(2), 96-108. https://doi.org/10.1038/s41437-023-00624-7

Nassar, L. R., Barber, G. P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., …Lee, B. T. (2023). The UCSC genome browser database: 2023 update. *Nucleic Acids Res*, *51*(D1), D1188-D1195. https://doi.org/10.1093/nar/gkac1072

Nazari, F., Seyedabadi, H.-R., Noshary, A., Emamjomeh-Kashan, N., & Banabazi, M-H. (2022). A genome-wide scan for signatures of selection in Kurdish horse breed. *J Equine Vet Sci*, *113*, 103916. https://doi.org/10.1016/j.jevs.2022.103916

Nicodemus, M. C., & Clayton, H. M. (2003). Temporal variables of four-beat, stepping gaits of gaited horses. *Appl Anim Behav Sci*, *80*(2), 133-142. https://doi.org/10.1016/S0168-1591(02)00219-8

Nimura, T., Itoh, T., Hagio, H., Hayashi, T., Di Donato, V., Takeuchi, M., …Hibi, M. (2019). Role of Reelin in cell positioning in the cerebellum and the cerebellum-like structure in zebrafish. *Dev Biol*, *455*(2), 393-408. https://doi.org/10.1016/j.ydbio.2019.07.010

Ninomiya, S., Anjiki, A., Nishide, Y., Mori, M., Deguchi, Y., & Satoh, T. (2013). Polymorphisms of the dopamine D4 receptor gene in stabled horses are related to differences in behavioral response to frustration. *Animals*, *3*(3), 663-669. https://doi.org/10.3390/ani3030663

Novoa-Bravo, M., Jäderkvist Fegraeus, K., Rhodin, M., Strand, E., García, L. F., & Lindgren, G. (2018). Selection on the Colombian Paso horse's gaits has produced kinematic differences partly explained by the *DMRT3* gene. *PLoS One*, *13*(8), e0202584. https://doi.org/10.1371/journal.pone.0202584

Novotna, A., Birovas, A., Vostra-Vydrova, H., Vesela, Z., & Vostry, L. (2022). Genetic parameters of performance and conformation traits of 3-year-old warmblood sport horses in the Czech Republic. *Animals*, *12*(21), 2957. https://doi.org/10.3390/ani12212957

Pan, Y., Wang, K.-S., & Aragam, N. (2011). NTM and NR3C2 polymorphisms influencing intelligence: Family-based association studies. *Prog Neuro-Psychoph*, *35*(1), 154-160. https://doi.org/10.1016/j.pnpbp.2010.10.016

Pavlopoulos, E., Trifilieff, P., Chevaleyre, V., Fioriti, L., Zairis, S., Pagano, A., …Kandel, E. R. (2011). Neuralized1 activates CPEB3: a function for nonproteolytic ubiquitin in synaptic plasticity and memory storage. *Cell*, *147*(6), 1369-1383.

Pereira, G. L., de Matteis, R., Regitano, L. C., Chardulo, L. A. L., & Curi, R. A. (2016). *MSTN*, *CKM*, and *DMRT3* gene variants in different lines of Quarter horses. *J Equine Vet Sci*, *39*, 33-37. https://doi.org/10.1016/j.jevs.2015.09.001

Peripolli, E., Munari, D. P., Silva, M. V. G. B., Lima, A. L. F., Irgang, R., & Baldi, F. (2017). Runs of homozygosity: current knowledge and applications in livestock. *Anim Genet*, *48*(3), 255-271. https://doi.org/10.1111/age.12526

Pernice, H. F., Schieweck, R., Jafari, M., Straub, T., Bilban, M., Kiebler, M. A., & Popper, B. (2019). Altered glutamate receptor ionotropic delta subunit 2 expression in Stau2-deficient cerebellar Purkinje cells in the adult brain. *Int J Mol Sci*, *20*(7), 1797. https://doi.org/10.3390/ijms20071797

Petersen, J. L., Mickelson, J. R., Cothran, E. G., Andersson, L. S., Axelsson, J., Bailey, E., …McCue, M. E. (2013). Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One*, *8*(1), e54997. https://doi.org/10.1371/journal.pone.0054997

Petersen, J. L., Mickelson, J. R., Rendahl, A. K., Valberg, S. J., Andersson, L. S., Axelsson, J., …McCue, M. E. (2013). Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genet*, *9*(1), e1003211. https://doi.org/10.1371/journal.pgen.1003211

Pontzer, H. (2007). Effective limb length and the scaling of locomotor cost in terrestrial animals. *J Exp Biol*, *210*(10), 1752-1761. https://doi.org/10.1242/jeb.002246

Promerová, M., Andersson, L. S., Juras, R., Penedo, M. C. T., Reissmann, M., Tozaki, T., …Andersson, L. (2014). Worldwide frequency distribution of the *'Gait keeper'* mutation in the *DMRT3* gene. *Anim Genet*, *45*(2), 274-282. https://doi.org/10.1111/age.12120

Rangwala, S. H., Kuznetsov, A., Ananiev, V., Asztalos, A., Borodin, E., Evgeniev, V., …Rudnev, D. (2021). Accessing NCBI data using the NCBI sequence viewer and genome data viewer (GDV). *Genom Res*, *31*(1), 159-169. https://doi.org/10.1101/gr.266932.120

Raudsepp, T., Finno, C. J., Bellone, R. R., & Petersen, J. L. (2019). Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era. *Anim Genet*, *50*(6), 569-597. https://doi.org/10.1111/age.12857

Regatieri, I. C., Pereira, G. L., Neto, A. R. T., Ferraz, G. C., Curi, R. A., & Queiroz-Neto, A. (2017). Polymorphisms in *MCT1*, *CD147*, *PDK4*, and *DMRT3* genes in Arabian and Quarter Horses. *J Equine Vet Sci*, *48*, 161-165. e161. https://doi.org/10.1016/j.jevs.2016.06.077

Reich, P., Möller, S., Stock, K. F., Nolte, W., von Depka Prondzinski, M., Reents, R., …Falker-Gieske, C. (2024). Genomic analyses of withers height and linear conformation traits in German Warmblood horses using imputed sequence-level genotypes. *Genet Sel Evol*, *56*(1), 45. https://doi.org/10.1186/s12711-024-00914-6

Ricard, A. (2015). Does heterozygosity at the *DMRT3* gene make French trotters better racers? *Genet Sel Evol*, *47*(1), 10. https://doi.org/10.1186/s12711-015-0095-7

Ricard, A., Crevier-Denoix, N., Pourcelot, P., Crichan, H., Sabbagh, M., Dumont-Saint-Priest, B., & Danvy, S. (2023). Genetic analysis of geometric morphometric 3D visuals of French jumping horses. *Genet Sel Evol*, *55*(1), 63. https://doi.org/10.1186/s12711-023-00837-8

Ricard, A., Robert, C., Blouin, C., Baste, F., Torquet, G., Morgenthaler, C., …Barrey, E. (2017). Endurance exercise ability in the horse: a trait with complex polygenic determinism. *Front Genet*, *8*. https://doi.org/10.3389/fgene.2017.00089

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, *273*(5281), 1516-1517. https://doi.org/10.1126/science.273.5281.15

Ropka-Molik, K., Stefaniuk-Szmukier, M., Szmatoła, T., Piórkowska, K., & Bugno-Poniewierska, M. (2019). The use of the *SLC16A1* gene as a potential marker to predict race performance in Arabian horses. *BMC Genetics*, *20*(1), 73. https://doi.org/10.1186/s12863-019-0774-4

Rustin, M., Janssens, S., Buys, N., & Gengler, N. (2009). Multi-trait animal model estimation of genetic parameters for linear type and gait traits in the Belgian warmblood horse. *J Anim Breed Genet*, *126*(5), 378-386. https://doi.org/10.1111/j.1439-0388.2008.00798.x

Rönnegård, L., Shen, X., & Alam, M. (2010). hglm: A package for fitting hierarchical generalized linear models. *R J*, *2*(2), 20-28.

Sackton, T. B., & Hartl, D. L. (2016). Genotypic context and epistasis in individuals and populations. *Cell*, *166*(2), 279-287. https://doi.org/10.1016/j.cell.2016.06.04

Safran, M., Rosen, N., Twik, M., BarShir, R., Stein, T. I., Dahary, D., …Lancet, D. (2021). The GeneCards Suite. In I. Abugessaisa & T. Kasukawa (Eds.), *Practical Guide to Life Science Databases* (1st ed., pp. 27-56). Springer Nature Singapore.

Sahin, L., Keloglan Müsüroglu, S., Selin Cevik, O., Cevik, K., & Orekici Temel, G. (2023). Hyperthyroidism leads learning and memory impairment possibly via GRIN2B expression alterations. *Brain Res*, *1802*, 148209. https://doi.org/10.1016/j.brainres.2022.148209

Schaefer, R. J., Schubert, M., Bailey, E., Bannasch, D. L., Barrey, E., Bar-Gal, G. K.,…McCue, M. E. (2017). Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds. *BMC Genomics*, *18*(1), 565. https://doi.org/10.1186/s12864-017-3943-8

Schubert, M., Jónsson, H., Chang, D., Der Sarkissian, C., Ermini, L., Ginolhac, A., …Petersen, B. (2014). Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *P Natl Acad Sci-Biol*, *111*(52), E5661-E5669. https://doi.org/10.1073/pnas.1416991111

Sevane, N., Dunner, S., Boado, A., & Cañon, J. (2017). Polymorphisms in ten candidate genes are associated with conformational and locomotive traits in Spanish Purebred horses. *J Appl Genet*, *58*, 355-361. https://doi.org/10.1007/s13353-016-0385-y

Shin, D.-H., Lee, J. W., Park, J.-E., Choi, I.-Y., Oh, H.-S., Kim, H. J., & Kim, H. (2015). Multiple genes related to muscle identified through a joint analysis of a two-stage genome-wide association study for racing performance of 1,156 Thoroughbreds. *Asian-Australas J Anim Sci*, *28*(6), 771-781. https://doi.org/10.5713/ajas.14.0008

Shin, J., & Lee, C. (2015). A mixed model reduces spurious genetic associations produced by population stratification in genome-wide association studies. *Genomics*, *105*(4), 191-196. https://doi.org/10.1016/j.ygeno.2015.01.006

Signer-Hasler, H., Flury, C., Haase, B., Burger, D., Simianer, H., Leeb, T., & Rieder, S. (2012). A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PLoS One*, *7*(5), e37282. https://doi.org/10.1371/journal.pone.0037282

Sigurðardóttir, H. (2012, May). *Athugun á úrvalsstyrkleika og erfðaframförum í einstökum eiginleikum í íslenskri hrossarækt (e. Investigation of selection intensity and genetic gain within breeding field test traits in Icelandic Horse breeding)* [Bachelor's thesis, Agricultural University of Iceland]. Skemman. http://hdl.handle.net/1946/12115

Sigurðardóttir, H., Albertsdóttir, E., & Eriksson, S. (2017). Analysis of new temperament traits to better understand the trait *spirit* assessed in breeding field tests for Icelandic horses. *Acta Agr Scand A-An*, *67*(1-2), 46-57. https://doi.org/10.1080/09064702.2017.1383507

Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., …Hayhurst, J. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic acids res*, *51*(D1), D977-D985. https://doi.org/10.1093/nar/gkac1010

Sonali, Bhardwaj, A., Unnati, Nayan, V., Legha, R. A., Bhattacharya, T. K.,…Giri, S. K. (2023). Identification and characterization of single nucleotide polymorphisms in *DMRT3* gene in Indian horse (*Equus caballus*) and donkey (*Equus asinus*) populations. *Anim Biotechnol*, *34*(9), 4910-4920. https://doi.org/10.1080/10495398.2023.2206866

Song, S., Oh, D.-Y., Cho, G.-J., Kim, D. H., Park, Y.-S., & Han, K. (2017). Targeted next-generation sequencing for identifying genes related to horse temperament. *Genes Genom*, *39*(12), 1325-1333. https://doi.org/10.1007/s13258-017-0597-5

Staiger, E. A., Al Abri, M. A., Pflug, K. M., Kalla, S. E., Ainsworth, D. M., Miller, D., …Brooks, S. A. (2016). Skeletal variation in Tennessee Walking horses maps to the *LCORL/NCAPG* gene region. *Physiol Genomics*, *48*(5), 325-335. https://doi.org/10.1152/physiolgenomics.00100.2015

Staiger, E. A., Al Abri, M. A., Silva, C. A. S., & Brooks, S. A. (2016). Loci impacting polymorphic gait in the Tennessee Walking horse. *J Anim Sci*, *94*(4), 1377-1386. https://doi.org/10.2527/jas.2015-9936

Staiger, E. A., Albright, J. D., & Brooks, S. A. (2016). Genome-wide association mapping of heritable temperament variation in the Tennessee Walking Horse. *Genes Brain Behav*, *15*(5), 514-526. https://doi.org/10.1111/gbb.12290

Staiger, E. A., Almén, M. S., Promerová, M., Brooks, S., Cothran, E. G., Imsland, F., …Andersson, L. (2017). The evolutionary history of the *DMRT3* '*Gait keeper*' haplotype. *Anim Genet*, *48*(5), 551-559. https://doi.org/ https://doi.org/10.1111/age.12580

Stefánsdóttir, G. J., Gunnarsson, V., Roepstorff, L., Ragnarsson, S., & Jansson, A. (2017). The effect of rider weight and additional weight in Icelandic horses in tölt: part I. Physiological responses. *animal*, *11*(9), 1558-1566. https://doi.org/10.1017/S1751731117000556

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., …Mazor, Y. (2016). The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinform*, *54*(1), 1.30. 31-31.30. 33. https://doi.org/10.1002/cpbi.5

Stoffel, M. A., Johnston, S. E., Pilkington, J. G., & Pemberton, J. M. (2021). Mutation load decreases with haplotype age in wild Soay sheep. *Evol Lett*, *5*(3), 187-195. https://doi.org/10.1002/evl3.229

Sui, L., Wang, Y., Ju, L.-H., & Chen, M. (2012). Epigenetic regulation of reelin and brain-derived neurotrophic factor genes in long-term potentiation in rat medial prefrontal cortex. *Neurobiol Learn Mem*, *97*(4), 425-440. https://doi.org/10.1016/j.nlm.2012.03.007

Sumreddee, P., Hay, E. H., Toghiani, S., Roberts, A., Aggrey, S. E., & Rekaya, R. (2021). Grid search approach to discriminate between old and recent inbreeding using phenotypic, pedigree and genomic information. *BMC Genomics*, *22*, 1-17. https://doi.org/10.1186/s12864-021-07872-z

Sánchez-Guerrero, M., Molina, A., Gómez, M., Peña, F., & Valera, M. (2016). Relationship between morphology and performance: signature of mass-selection in Pura Raza Español horse. *Livest Sci*, *185*, 148-155. https://doi.org/10.1016/j.livsci.2016.01.003

Tang, Y-P., Shimizu, E., Dube, G. R., Rampon, C., Kerchner, G. A., Zhuo, M., …Tsien, J. Z. (1999). Genetic enhancement of learning and memory in mice. *Nature*, *401*(6748), 63-69. https://doi.org/10.1038/43432

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

Tetens, J., Widmann, P., Kühn, C., & Thaller, G. (2013). A genome-wide association study indicates *LCORL/NCAPG* as a candidate locus for withers height in German Warmblood horses. *Anim Genet*, *44*(4), 467-471. https://doi.org/10.1111/age.12031

Thomas, P. D., Ebert, D., Muruganujan, A., Mushayahama, T., Albou, L-P., & Mi, H. (2022). PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Sci*, *31*(1), 8-22. https://doi.org/10.1002/pro.4218

Tozaki, T., Hill, E. W., Hirota, K., Kakoi, H., Gawahara, H., Miyake, T., …Kurosawa, M. (2012). A cohort study of racing performance in Japanese Thoroughbred racehorses using genome information on ECA18. *Anim Genet*, *43*(1), 42-52. https://doi.org/10.1111/j.1365-2052.2011.02201.x

Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the cloud: using Docker, GATK, and WDL in Terra* (1st ed.). O'Reilly Media.

Vaswani, A. R., & Blaess, S. (2016). Reelin signaling in the migration of ventral brain stem and spinal cord neurons. *Front Cell Neurosci*, *10*. https://doi.org/10.3389/fncel.2016.00062

Velie, B. D., Fegraeus, K. J., Solé, M., Rosengren, M. K., Røed, K. H., Ihler, C-F., …Lindgren, G. (2018). A genome-wide association study for harness racing success in the Norwegian-Swedish coldblooded trotter reveals genes for learning and energy metabolism. *BMC Genetics*, *19*(1), 80. https://doi.org/10.1186/s12863-018-0670-3

Viklund, Å., Hellsten, E. T., Näsholm, A., Strandberg, E., & Philipsson, J. (2008). Genetic parameters for traits evaluated at field tests of 3-and 4-year-old Swedish Warmblood horses. *animal*, *2*(12), 1832-1841. https://doi.org/10.1017/S1751731108003030

Vosgerau, S., Krattenmacher, N., Falker-Gieske, C., Seidel, A., Tetens, J., Stock, K. F., …Thaller, G. (2022). Genetic and genomic characterization followed by single-step genomic evaluation of withers height in German Warmblood horses. *J Appl Genet*, *63*(2), 369-378. https://doi.org/10.1007/s13353-021-00681-w

Wade, C. M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., …Lindblad-Toh, K. (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science*, *326*(5954), 865-867. https://doi.org/doi:10.1126/science.1178158

Wilson, B., Neal, R., Howard, A., & Groenendyk, S. (1988). The gait of pacers 1: kinematics of the racing stride. *Equine Vet J*, *20*(5), 341-346. https://doi.org/10.1111/j.2042-3306.1988.tb01542.x

Wolschrijn, C., Audigié, F., Wijnberg, I. D., Johnston, C., Denoix, J-M., & Back, W. (2013). The neck and back. In W. Back & H. M. Clayton (Eds.), *Equine locomotion* (2nd ed., pp. 199-227). Elsevier.

Wutke, S., Andersson, L., Benecke, N., Sandoval-Castellanos, E., Gonzalez, J., Hallsson, J. H., …Ludwig, A. (2016). The origin of ambling horses. *Curr Biol*, *26*(15), R697-R699. https://doi.org/10.1016/j.cub.2016.07.001

Yokomori, T., Ohnuma, A., Tozaki, T., Segawa, T., & Itou, T. (2023). Identification of personality-related candidate genes in Thoroughbred racehorses using a bioinformatics-based approach involving functionally annotated human genes. *Animals*, *13*(4), 769. https://doi.org/10.3390/ani13040769

Zips, S., Peham, C., Scheidl, M., Licka, T., & Girtler, D. (2001). Motion pattern of the toelt of Icelandic horses at different speeds. *Equine Vet J*, *33*(S33), 109-111. https://doi.org/10.1111/j.2042-3306.2001.tb05371.x

Árnadóttir, E. (2022, May). *Erfðafjölbreytileiki íslenska hrossastofnsins (e. Genetic diversity of the Icelandic horse population)* [Bachelor's thesis, Agricultural University of Iceland]. Skemman. http://hdl.handle.net/1946/41594

Árnason, T. (1983). *Genetic studies on conformation and performance of Icelandic toelter horses* [PhD dissertation, Swedish University of Agricultural University]. SLU University Library, Uppsala.

Árnason, T. (1984). Genetic studies on conformation and performance of Icelandic Toelter horses: IV. Best linear unbiased prediction of ten correlated traits by use of an "Animal Model". *Acta Agr Scand*, *34*(4), 450-462. https://doi.org/10.1080/00015128409435413

Árnason , T., Klemetsdal, G., & Sigurssson, Á. (1995, August 11-13). *Nordiska hästraser - gamla genresurser* ägnade for framtidens behov [Conference session]. International Symposium on Horse Breeding and Production in Cold Climatic Regions, Reykjavík, Iceland.

I

Rosengren et al. BMC Genomics (2021) 22:267
https://doi.org/10.1186/s12864-021-07454-z

BMC Genomics

RESEARCH ARTICLE Open Access

A QTL for conformation of back and croup influences lateral gait quality in Icelandic horses

Maria K. Rosengren[1*†], Heiðrún Sigurðardóttir[1,2†], Susanne Eriksson[1], Rakan Naboulsi[1], Ahmad Jouni[1], Miguel Novoa-Bravo[1,3], Elsa Albertsdóttir[4], Þorvaldur Kristjánsson[2], Marie Rhodin[5], Åsa Viklund[1], Brandon D. Velie[6], Juan J. Negro[7], Marina Solé[1†] and Gabriella Lindgren[1,8†]

Abstract

Background: The back plays a vital role in horse locomotion, where the spine functions as a spring during the stride cycle. A complex interaction between the spine and the muscles of the back contribute to locomotion soundness, gait ability, and performance of riding and racehorses. Conformation is commonly used to select horses for breeding and performance in multiple horse breeds, where the back and croup conformation plays a significant role. The conformation of back and croup plays an important role on riding ability in Icelandic horses. However, the genes behind this trait are still unknown. Therefore, the aim of this study was to identify genomic regions associated with conformation of back and croup in Icelandic horses and to investigate their effects on riding ability. One hundred seventy-seven assessed Icelandic horses were included in the study. A genome-wide association analysis was performed using the 670 K+ Axiom Equine Genotyping Array, and the effects of different haplotypes in the top associated region were estimated for riding ability and additional conformation traits assessed during breeding field tests.

Results: A suggestive quantitative trait loci (QTL) for the score of back and croup was detected on *Equus caballus* (ECA) 22 (p-value = 2.67×10^{-7}). Haplotype analysis revealed two opposite haplotypes, which resulted in higher and lower scores of the back and croup, respectively (p-value < 0.001). Horses with the favorable haplotype were more inclined to have a well-balanced backline with an uphill conformation and had, on average, higher scores for the lateral gaits tölt (p-value = 0.02) and pace (p-value = 0.004). This genomic region harbors three genes: *C20orf85*, *ANKRD60* and *LOC100056167*. *ANKRD60* is associated with body height in humans. *C20orf85* and *ANKRD60* are potentially linked to adolescent idiopathic scoliosis in humans.

(Continued on next page)

* Correspondence: maria.rosengren@slu.se
Maria K Rosengren & Heiðrún Sigurðardóttir share first authorship.
†Marina Solé & Gabriella Lindgren share senior authorship.
[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden
Full list of author information is available at the end of the article

(Continued from previous page)
Conclusions: Our results show that the detected QTL for conformation of back and croup is of importance for quality of lateral gaits in Icelandic horses. These findings could result in a genetic test to aid in the selection of breeding horses, thus they are of major interest for horse breeders. The results may also offer a gateway to comparative functional genomics by potentially linking both motor laterality and back inclination in horses with scoliosis in humans.

Keywords: Back, Backline, Conformation, Croup, High-density genome scan, Icelandic horse, Lateral gait quality, Novel QTL

Background

Associations of body measurements with locomotor health and sports performance have been reported in many different breeds, including Icelandic horses [1–11]. Discriminant analyses have shown that several morphological features distinguish with high accuracy between low-class and high-class Icelandic horses with respect to different riding ability traits [3]. The most important features for gait ability in Icelandic horses are the height of the horse at front compared to hind (uphill conformation) with well-balanced backline, croup proportions and width of chest [1, 3]. The analyses also indicated the disadvantage of a forward inclination in the back or a sway back [3]. Conformation of the back and croup thus play a major role on riding ability in Icelandic horses.

The Icelandic horse official breeding goal promotes five-gaited horses with a functional and aesthetically pleasing conformation [12]. Zoometric measurements and subjective scores for conformation and riding ability traits are recorded at breeding field tests. Genetic correlations between conformation of back and croup, and gait qualities have been estimated as moderate to high (0.19–0.54) [1]. Furthermore, moderate heritabilities (0.29–0.31) have been estimated for the subjectively scored back and croup trait [1, 13] and the objectively measured zoometric traits pertaining to conformation of back and croup (0.20–0.25) [3]. For the subjectively scored riding ability traits, the heritability estimates range from 0.18 (walk) to 0.60 (pace) [1, 13].

Despite conformation traits being moderately heritable in the Icelandic horse, only mutations in the Myostatin gene have previously been associated with conformation traits, i.e. estimated breeding values of neck, withers and shoulders [14]. In other horse breeds, as well as other species, many different genes have been shown to influence body size. *LCORL*, *NCAPG* and *HMGA2* are major genes known to regulate body size in mammals including humans, cattle, sheep, dogs and horses [15–23]. These genes, along with other genes such as *ZFAT* and *LASP1*, affect not only the body size of the horse but more specifically the height at withers [15, 24, 25]. Three novel missense variants located in the *ADAMTS17*, *OSTN* and *GH1* genes explained 61% of the variance of withers height in Shetland pony-related breeds [26]. Other additional quantitative trait loci have also shown significant associations with morphometric angular measurements, with regions on chromosomes ECA28 and ECA29 associated with poll angle in horses [27]. However, the genes behind many other conformation traits are still unknown.

Considering the heritability of conformation of back and croup and its genetic correlation with riding ability, we hypothesized that major genetic factors of importance for back and croup also influence gait quality in Icelandic horses. Therefore, the aim of this study was to identify genomic regions associated with conformation of back and croup in Icelandic horses and investigate their effects on riding ability traits assessed at breeding field tests.

Results

Genome-wide association analysis for conformation of back and croup

In total, 383,896 SNPs (373,041 autosomal and 10,855 X chromosomal) and 177 horses passed QC and were included in the GWA analysis. Thirteen SNPs located on ECA22: 45347522–45,662,708 reached the suggestive threshold ($p < 1.0 \times 10–5$) of which ten were in LD ($r2 \geq 0.8$) (Fig. 1). Additionally, one single SNP reached the suggestive threshold on ECA12 (Fig. 1). A summary of the GWA results for the 50 top SNPs is presented in Additional file 1.

Haplotype analysis

The haplotype analysis revealed two opposite haplotypes which resulted in higher and lower scores for back and croup (p-value < 0.001) (Table 1). Thirty-four horses were homozygous for the haplotype associated with a higher score and 28 horses homozygous for the haplotype associated with a lower score of back and croup. Five different haplotypes were estimated (Table 1). Haplotypes determined to be too rare to estimate their specific regression coefficients were pooled into a separate group with a frequency of 0.07 (results not presented).

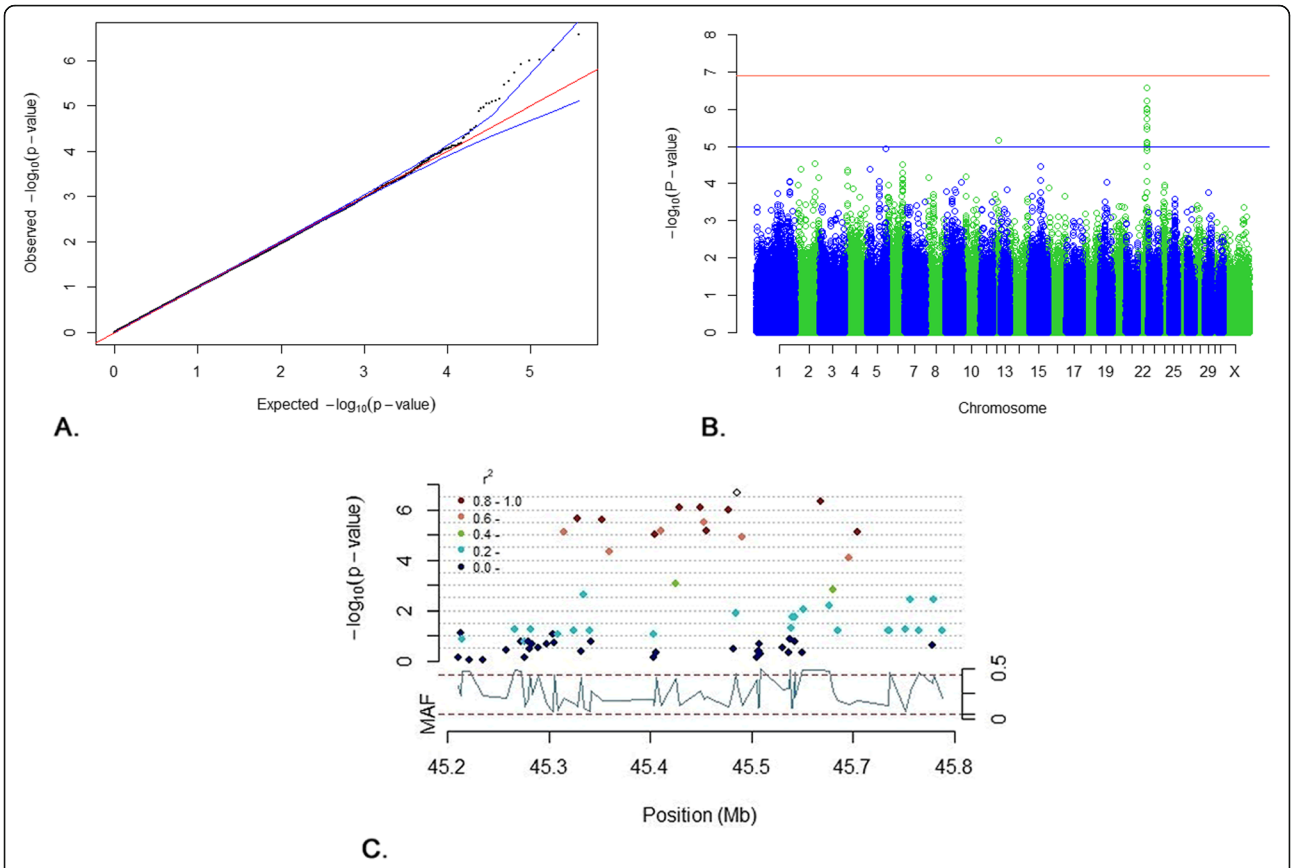


Fig. 1 GWA results for the score of back and croup. **a.** QQ plot where the blue lines represent the 0.05–0.95 confidence interval. The estimated lambda value was 0.98 (se 2.55×10^{-5}). **b.** Manhattan plot from the mixed model association analysis. The red horizontal line indicates Bonferroni significance threshold ($p < 6.9 \times 10^{-8}$) and the blue horizontal line indicates the suggestive genome-wide significance level ($p < 1.0 \times 10^{-5}$). **c**. LD Manhattan plot on ECA22 with the top SNP as an open circle. Thirteen SNPs reached the suggestive threshold of which ten were in LD. All positions refer to EquCab3.0

Table 1 Results from haplotype analysis for the score of back and croup

Haplotypes (SNPs numbers[a])										Coef	Freq	*p*-value	Sim. *p*-value
1	2	3	4	5	6	7	8[a]	9	10				
G	T	C	A	T	A	T	A	A	T	−0.300	0.383	**< 0.001**	**< 0.001**
G	T	C	A	T	A	T	A	G	C	0.090	0.021	0.657	0.718
G	T	C	A	G	G	G	A	A	T	0.119	0.027	0.518	0.889
G	C	T	C	T	A	G	A	A	T	0.090	0.025	0.626	0.963
A	C	T	C	G	G	G	G	G	C	0.300	0.474	**< 0.001**	**< 0.001**

Sim. *p*-value = *p*-value adjusted by using 100,000 permutations
Significant results in bold
Coef. coefficient, estimated effect of the haplotype on the score of back and croup from the glm model in the haplotype analysis
Freq. frequencies
[a]SNP numbers in bp position order with top SNP as number 8 with reference allele A and alternate allele G

Phenotype association of the haplotypes with a significant effect on the score of back and croup

The t-test analyses revealed that several traits in addition to back and croup significantly differed in mean scores between horses with the favourable and unfavorable haplotype. The two haplotype groups differed significantly in mean scores (p-value ≤0.05) for the gait traits tölt and pace (Table 2). The two haplotype groups also differed significantly in means for the zoometric measurements of depth at breast, width of hips and thigh bones, and length of the forelimbs. In addition to this, there were significant differences between the two haplotype groups for the sub-traits backline and the croup type.

Allele frequency of top SNP and *DMRT3* in different breeds

Comparing allele frequencies of the top SNP identified from GWA analysis between different breeds revealed a higher frequency of the alternate allele (the favorable

Table 2 Significant results from *t*-test comparing phenotypes in horses with different haplotypes

Trait	Favorable haplotype		Unfavorable haplotype				
	N	Mean	N	Mean	*t*-value	df	*p*-value
Back and croup	34	8.29	28	7.71	4.05	58.08	< 0.001
Tölt[a]	33	8.41	27	7.96	2.52	45.79	0.015
Pace[a]	33	7.18	27	6.09	2.99	50.24	0.004
Slow tölt[a]	33	8.14	26	7.73	2.14	45.19	0.038
Depth at breast (M4)[b]	33	63.2	28	64.6	−3.52	56.22	0.001
Width of the hips (M7)[b]	23	47.0	20	48.1	−2.21	37.54	0.033
Width between thigh bones (M8)[b]	23	43.0	20	44.2	−2.23	38.86	0.031
Length of forelimbs (M1-2xM4)[b]	33	15.2	28	12.1	3.22	40.81	0.003
Backline[c]	34	1.79	28	2.25	−2.69	58.91	0.009
Croup type[c]	34	1.85	28	2.18	−2.31	53.23	0.025

N Number of horses
[a]Subjectively assessed traits (scale 5–10)
[b]Zoometric measurements (cm)
[c]Subjectively assessed sub-traits (scale 1–3)

allele) in the Icelandic breed compared with all other investigated breeds (Table 3).

Functional annotation of genes in the region associated with the score of back and croup

The detected QTL ECA22: 45347522–45,662,708 harbors the genes *Chromosome 22 C20orf85 homolog* (*C22H20orf85*), *Ankyrin repeat domain 60* (*ANKRD60*) and *LOC100056167* described as *serine/threonine-protein phosphatase 4 regulatory subunit 1*. The SNP on ECA12 (position 26,756,656–26,756,656) was located close to the gene *solute carrier family 22 member 8* (*SLC22A8*). None of the significant SNPs (on ECA12 and 22) overlapped any known QTL for conformation in horses [39].

Discussion

Conformation of the back and croup plays an important role for riding ability, gait ability, welfare, and longevity of the horse [1, 3, 13, 40]. The present study was performed to identify genomic regions associated with conformation of the back and croup in Icelandic horses and investigate their effects on riding ability. A novel QTL was detected on ECA22 with candidate genes associated with scoliosis and anthropometric traits in humans [41, 42]. Our results show that this QTL is of importance not only for conformation of back and croup, but also for riding ability traits, especially lateral gait quality, in Icelandic horses.

Possible links between scoliosis, motor laterality and lateral gaits

The detected QTL for the trait back and croup harbors the genes *C22H20orf85* and *ANKRD60,* both of which are potentially linked to adolescent idiopathic scoliosis (AIS) in humans [41]. Scoliosis is defined as a lateral curvature of the spine and it is the most common vertebral disorder in children and adolescents [43]. In humans, scoliosis can be caused by muscular dystrophy or cerebral palsy, but the cause is usually unknown and therefore referred to as idiopathic [43]. AIS in humans has been shown to result in a generalized skeletal muscle weakness, respiratory impairment and exercise limitation [44]. Studies on scoliosis in humans have also shown correlation between handedness and truncal asymmetry [45–49] and that molecular basis of handedness are more likely formed by spinal gene expression asymmetries rather than in the motor cortex [50]. Symptoms of scoliosis in horses has been described as an S-shaped bend of the caudal thoracic vertebral column, resulting in restricted movements of the hind limbs and inflexibility of the back [51]. Another report described symptoms as a lateral deviation of the head and cervical and cranial thoracic vertebral column to one side, and associated rotation of the thoracic vertebrae. These deviations result in difficulties for a horse to walk in a straight line [52]. However, severe thoracic vertebral malformations in horses are infrequent, and mild to moderate forms of scoliosis may go undetected as the strong dorsal spinal musculature can mask subtle deviations of the vertebral column [52]. Scores for conformation of back and croup in horses involve both muscular and skeletal assessments, which may indicate that the back and croup phenotype shares some features with mild forms of scoliosis. It is well known that horses commonly demonstrate motor laterality (handedness) [53–55] and some even have difficulties walking in a straight line at the beginning of training. The latter often need more time in training to improve their balance and straightness.

In general, disorders of the back appear to be relatively common in horses and lead to pain and decreased performance [51]. However, to our knowledge, there are no

Table 3 Allele frequency of top SNP for back and croup and *DMRT3*

Breed	Top SNP			*DMRT3*		
	N	AF alt	Source	N	AF alt	Source
Icelandic horses included in present study[a]	177	0.50	Array genotyping	177	0.94	Array genotyping
Icelandic horses unassessed[b]	49	0.51	SNP genotyping	49	0.90	SNP genotyping
Other gaited breeds						
Rocky-Mountain	36	0.33	SNP genotyping	27	1	SNP genotyping
Colombian paso horses Colombian trocha	37	0.24	Array genotyping	37	0.0	SNP genotyping
Colombian trot and gallop	11	0.23	Array genotyping	11	0.0	SNP genotyping
Colombian paso fino	38	0.29	SNP genotyping	28	1	[28]
Partly gaited breeds						
American Curly	27	0.32	SNP genotyping	101	0.70	[29]
American Saddlebred	42	0.29	SNP genotyping	89	0.28	[30]
Morgan	30	0.44	SNP genotyping	59	0.14	[29]
Non- gaited breeds						
Exmoor	279	0.01	[31]	27	0.0	[31]
Connemara Pony	40	0.05	[32]	35	0.0	[30]
Swedish Warmblood	379	0.26	[33]	64	0.0	[30, 34]
Thoroughbred racehorses	370	0.14	[35]	55	0.0	[30, 34]
Persian-Arabian horses	101	0.32	[36]	69	0.0	[30]
North-Swedish draught	25	0.38	[37]	34	0.0	[30, 34]
Harness racing breeds						
Coldblooded trotters	565	0.13	[38]	306	0.45	[30]
Standardbred	40	0.29	SNP genotyping	270	0.97	[30, 34]

N number of horses included in dataset
Top SNP the top SNP identified from the GWA analysis for back and croup
AF alt frequency of alternate allele
DMRT3 AF alt allele frequency of the alternate allele A in the DMRT3 gene known as the "Gait Keeper" mutation
[a]The 177 Icelandic horses included in the present study
[b]Icelandic horses used for riding but that had not attended breeding field test

studies reporting the prevalence of back problems or scoliosis in Icelandic horses, and it is generally hard to diagnose back pain in horses. The effect of the QTL is more likely related to functional advantage or disadvantage for movements and strength of the back and croup in horses rather than the result of more severe dysfunctions and pain. This is supported by the relatively high frequency of the unfavorable haplotype among the Icelandic horses in the present study.

Top SNP allele frequency in other breeds

Icelandic horses had a higher frequency of the alternate allele (the favorable allele) of the top SNP for back and croup compared with all other investigated breeds, including the other gaited and partly gaited breeds. In addition, the Icelandic horses with the favorable haplotype had on average higher scores for the lateral gaits tölt and pace. Therefore, it is likely that the quality of the lateral gaits rather than the ability to perform the gaits is affected by the QTL. Almost all Icelandic horses carry at least one copy of the mutant allele A in the *DMRT3* gene known as the "Gait Keeper" mutation [30, 34]. This mutation is known to affect the pattern of locomotion in horses and the ability to perform lateral gaits [34]. The Icelandic horses in the present study had a high frequency of the *DMRT3* "Gait Keeper" mutation (0.94), 157 of the 177 horses were homozygous AA. The *DMRT3* genotype was taken into account in the phenotype association analysis. Pace scores in horses with the CA genotype were considered as a missing value. Despite this, the Icelandic horses with the favorable haplotype had higher scores for pace. This further supports our hypothesis that the detected QTL affects the quality and not the ability of lateral gaits. The genotyped gaited breed Rocky-Mountain Horse is known to be fixed for the *DMRT3* "Gait Keeper" mutation [30]. The other genotyped gaited breeds American Curly, American Saddlebred and Morgan horses have a moderate high frequency of the *DMRT3* "Gait Keeper" mutation [30, 34, 56]. These breeds are considered as partly gaited as not all horses within the breed perform ambling gaits. Trotters are also known to

perform lateral gaits, and the reported frequency of the *DMRT3* mutation is high in Standardbreds (0.97–1.00) [30, 34] and relatively high in Coldblooded trotters (0.45) [30]. All of these gaited and partly gaited breeds had a higher frequency of the reference allele than the alternate allele for the top SNP of back and croup. The genotyped Colombian paso horses (CPH) included a group of horses that perform trocha and one group that only perform trot and gallop. The trocha gait is defined as a four-beat gait that includes a lateral step but it is diagonally coupled and therefore not considered a lateral gait [28, 57]. The allele frequency of the top SNP did not differ between these two groups. A group of CPH that perform the lateral gait paso fino was also genotyped. However, like all the other genotyped breeds, this group had a lower frequency of the alternate allele of the top SNP for back and croup compared to the Icelandic horses. None of the other genotyped breeds in this study segregates for the *DMRT3* mutation [30, 34], nor do they perform lateral gaits.

The 49 unassessed Icelandic horses had a similar allele frequency of the top SNP for back and croup as well as for the *DMRT3* mutation as the 177 assessed Icelandic horses included in the present study. The unassessed group included riding school horses and horses used for hobby riding. It could be argued that balance and straightness is even more essential for the training of Icelandic horses as they carry relatively heavy (adult) riders, relative to their size, in lateral gaits such as tölt and pace with strong focus on the gait quality. In addition, the Icelandic horses with the favorable haplotype had higher average scores for the lateral gaits tölt and pace, which are highly valued traits in the breed. It is likely that there has been selection for the alternate allele of the top SNP in Icelandic horses.

Genes within the QTL associated with musculoskeletal traits

The gene *ANKRD60* is associated with body height in humans [42] and a recent study in American Miniature Horses reported a QTL for withers height close to another Ankyrin Repeat Domain gene *ANKRD1* [58]. The QTL region on ECA22 harbors the gene *LOC100056167* that is not well annotated in horses. The gene is described as serine/threonine-protein phosphatase 4 regulatory subunit 1 and appears to blast with the pseudogene *PPP4R1L* in humans with 84.17% identity [59]. The pseudogene *PPP4R1L* is transcribed in humans and *LOC100056167* has exons. *PPP4R1L* has a potential effect on bone mineral density as it has a protein phosphatase regulator activity [60]. *PPP4R1L* is regulated by an enhancer (Genehancer ID GH20J058887) with potential implications on body height and BMI-adjusted waist circumference in humans [61, 62]. Therefore, it is possible that the detected QTL effects both the muscular and skeletal system.

The horses with the favorable haplotype in the present study had longer forelimbs than those with the unfavorable haplotype. This may be explained, at least to some extent, by the effects of the genes *ANKRD60* and *LOC100056167*. According to a previous study, high-class Icelandic horses are distinguished from low-class horses by an uphill conformation [3]. High-class horses have higher withers and higher set neck and back, compared to height at croup and tuber coxae [3]. Uphill conformation is believed to facilitate ease of collection and lightness in the front part, features that are taken into account when gait quality is subjectively assessed at breeding field tests [12]. Stride length is associated with limb length in horses and other species [63–65] and stride length is also taken into account when assessing the gait quality at breeding field tests [12]. Consequently, stride length and uphill conformation are important factors for higher gait quality scores, both of which may be connected to longer forelimbs. This further supports the results from this study as the horses with the favorable haplotype had both longer forelimbs and higher scores for tölt and pace. In line with this, the horses with the unfavorable haplotype also had a deeper breast and more negative standardized marks for the sub-trait backline compared with the ones with the favorable haplotype. This indicates that a downhill conformation is more common in horses with the unfavorable haplotype. It is possible that a downhill inclination creates an imbalance between the front and back of the horse, causing difficulties for the horse to stretch the hind legs forward, thus losing the ability for self-carriage and collection. This may also result in a shorter stride length, causing lower scores for tölt and pace.

Length and form of the croup are also known to discriminate between high-class and low-class Icelandic horses [3]. In the present study, horses with the favorable haplotype had more positive standardized marks for the sub-trait croup type. This trait is defined as how evenly the croup is shaped and suggests that the haplotype does not influence the length or inclination of the croup, but only the shape of it. The difference between the two haplotype groups for the width of hips (M7) and width between the thighbones (M8) suggest that horses with the favorable haplotype may have a slimmer framed croup than horses with the unfavorable haplotype.

Complexity of the phenotype

Until around year 2010, a soft, lower backline was considered to be favorable for the assessment of back and croup of Icelandic horses, as a low position of the back was assumed desirable for tölt [12]. A study in American Saddlebred horses detected a region on ECA20 associated with extreme lordosis (swayback) [66]. However, in the present study no significant association with back

and croup was detected on ECA20. Horses with the haplotype associated with lower score of back and croup were more inclined to have a forward sloping and/or swayback backline.

The back and croup is a complex trait, with muscular as well as skeletal features of both the back and the croup subjectively assessed and scored together as a single trait. Our results show that the novel detected QTL associated with back and croup conformation influences various riding ability and conformation traits. It should be noted that the complex conformation and riding ability traits are likely to be influenced by many different genes as well as environmental factors such as feeding and training. Therefore, further studies are needed to determine the effects of this newly discovered QTL.

Conclusions

This study provides valuable information about the genetics of conformation of the back and croup in Icelandic horses. A novel QTL for the trait back and croup was detected on ECA22: 45347522–45,662,708. The QTL is associated with the back inclination, the form of the croup, and length of limbs as well as the quality of the lateral gaits pace and tölt. These findings could result in the offering of a genetic test to aid in the selection of breeding horses, thus they are of major interest for horse breeders. The genomic region harbors genes associated with scoliosis and anthropometric traits in humans. The findings could serve as a platform to study any potential link between scoliosis and motor laterality in horses and other species. Further analyses are needed to fully understand the biological function of this genomic region on the conformation of back and croup and its influence on gait quality.

Methods

Animals

In total, 177 Icelandic horses (77 males and 100 females) born between 1993 and 2014 were included in the study. Hair samples were collected at breeding field tests and by visiting trainers and breeders in Iceland and Sweden. A few samples were also sent in by horse owners after personal contact and posting on social media. Only privately owned horses participated in the study and the horses were not specifically selected based on conformation of back and croup. Pedigree data were obtained from the international Icelandic horse database Worldfengur [67]. Maximum relatedness between horses was limited to half-siblings.

Phenotyping

Phenotype data were obtained from the international Icelandic horse database Worldfengur [67]. The phenotype used for the genome-wide association (GWA) analysis consisted of the subjectively assessed score for back and croup recorded at breeding field tests between 1999 and 2018. Additional conformation and riding ability traits assessed at breeding field tests were used to investigate the effects of genomic regions detected from GWA analysis. Of the 177 horses had 115 attended more than one breeding field test. For these horses, information from the latest assessment was used. The majority of horses were assessed in year 2018 ($n = 89$). The horses were assessed in Iceland ($n = 81$), Sweden ($n = 87$), Germany ($n = 3$), Denmark ($n = 2$) and Norway ($n = 4$). Icelandic horses can attend breeding field test from when they are 4 years old. The age of assessment was on average 6.7 years and ranged from 4 to 15 years. In our sample, 173 horses were assessed for both conformation and riding ability traits, and 4 horses were only assessed for conformation traits as the ridden test is optional. Pace scores for horses with the CA genotype for the DMRT3 gene ($n = 20$) were treated as missing values.

Back and croup

Back and croup, along with other conformation and riding ability traits assessed at breeding field tests, were subjectively scored on a scale from 5 to 10 with 0.5 intervals, where a score of 5 was only given if a trait was not presented. Assessment of the trait back and croup comprises several aspects of the conformation of the back, croup and loins. The slope and shape of the backline, which is defined as the line from the base of withers to the lumbosacral joint, were assessed. Length and slope of the croup were also assessed, as well as the width and muscularity of the back, the length and width of the loins and the form and muscularity of the croup [12]. A high score for back and croup represents a strong, well-balanced backline and a well-muscled wide back. The croup should be long, evenly formed, well-muscled and adequately sloping. A low score is associated with a swayback, stiff or forward sloping backline, a too short or too long and/or unevenly formed croup and poorly muscled back and croup [12]. When the judging panel has reached a consensus on a score for back and croup according to the judging scale, they have the possibility to use standardized marks to describe the most prominent positive and/or negative attributes of the trait.

Pictures with examples of horses representing high and low score for back and croup are presented in Fig. 2. The 177 horses in the study had a score of back and croup that ranged from 6.5 to 9.0 with a mean value of 8.1 (SD 0.56) (Fig. 3). The distribution of the scores for back and croup was slightly negatively skewed (coefficient of skewness − 0.36). Transformation of the raw data to increase normality was tested but was found to not affect the results. Moreover, the residuals from the

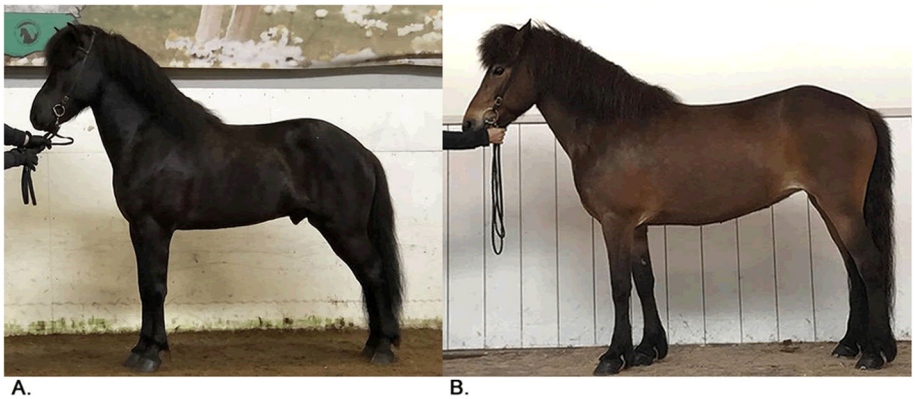


Fig. 2 Examples of Icelandic horses representing high and low score of back and croup. **a.** Icelandic horse that represents high score of back and croup. The backline is well-balanced and the back is wide and well-muscled. (Photo: Hrefna María Ómarsdóttir). **b.** Icelandic horse that represents a low score of back and croup with a forward sloping backline and a less muscled croup. (Photo: The Swedish Icelandic Horse Association, SIF). Images used in the figure are privately owned and thus were not taken from previously published sources requiring written permission for use

linear models were normally distributed (results not presented).

Sub-traits based on standardized marks for back and croup

For the purpose of more detailed analysis of the score for back and croup, the standardized marks used to describe prominent positive and negative attributes of the trait were defined as two different sub-traits; backline and croup type. These sub-traits were analysed on a linear scale ranging from 1 to 3, where 1 represented a positive mark, 3 represented a negative mark and 2 represented no mark and was interpreted as an intermediate description of the trait (not positive or negative). A positive mark for the sub-trait backline was given for good backline (well-balanced backline) and the options for negative marks were forward sloping back, straight back, sway back and/or stiff loins. For the sub-trait croup type, a positive mark was given for evenly formed croup and the options for negative marks were rounded croup, narrowing croup, roof-shaped croup and/or coarse croup.

Additional trait assessment scores from breeding field tests

Besides the conformation trait back and croup, scores for the gait traits tölt, slow tölt, trot, pace, gallop, canter and walk and the trait form under rider were included in this study. Features of each gait such as beat, suppleness, stride length, leg-action, speed capacity, collection and lightness were taken into account when assessing the

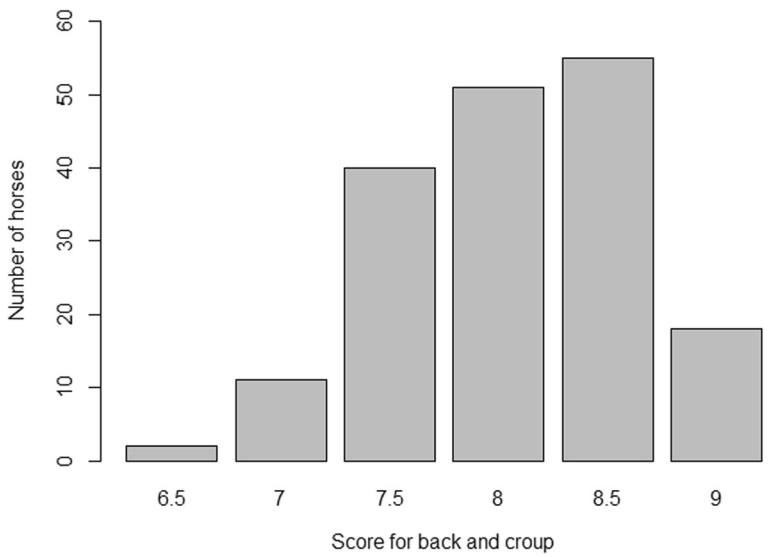


Fig. 3 Distribution of scores for back and croup in the 177 horses. (Picture generated from the data analysis in R)

gaits [12]. Scores of all these traits were included to investigate the effects of the detected regions from GWA analysis on the trait back and croup.

Zoometric traits measured at breeding field tests

Zoometric measurements are traditionally recorded at breeding field tests to corroborate the subjective conformation assessments [12]. All these measurements were included to investigate the effects of the detected genomic regions from GWA analysis for the trait back and croup. The measurements consisted of height at withers (M1), height at lowest point of back (M2), height at croup (M3), depth of breast (M4), length of body from the point of shoulder to tuber ischii (M5), width of chest between the points of the shoulders (M6), width of the hips between the tuber coxae (M7) and width of the hips between the hip joints (M8) (Fig. 4). Length of forelimbs is traditionally assessed from calculation of the difference between height at withers and depth at breast times two (M1-2xM4), as it gives better comparison of the leg length to consider the variation in breast depth between different horses. Other calculated measurements used for conformation assessments were difference between height at withers and height at back (M1-M2), difference between height at withers and height at croup (M1-M3), difference between height at croup and height at back (M3-M2), difference between length of the horse and height at withers (M5-M1), difference between length of the horse and height at croup (M5-M3) and difference between width of hips and width between thigh bones (M7-M8).

DNA isolation

DNA was extracted from hair roots using a standard procedure of hair preparation. One hundred eighty-six microlitre of 5% Chelex® 100 Resin (Bio-Rad Laboratories, Hercules, CA) and 14 μL of proteinase K (20 mg/mL; Merck KgaA, Darmstadt, Germany) were added to each sample. This mix was incubated at 56 °C for 2 h at 600 rpm and proteinase K was inactivated for 10 min at 95 °C.

Genotyping and quality control

The 177 Icelandic horses were genotyped on the 670 K+ Axiom Equine Genotyping Array. Quality control (QC) was performed with the package GenABEL [69] in R [70] to remove poorly genotyped and noisy data based on the following thresholds: missing genotypes per single nucleotide polymorphism (SNP) (> 0.10), missing SNPs per sample (> 0.10), minor allele frequency (MAF) (< 0.05) and Hardy-Weinberg equilibrium (p-value 1e-^{10}).

Genome-Wide Association Study (GWAS)

GWA analyses were performed using the package GenABEL [69] in R [70]. Possible fixed effects were tested in a linear model using anova as a post hoc test. The tested fixed effects were sex (male or female), age at assessment in age classes (4, 5, 6 or ≥ 7 years old), age at assessment in years as a linear regression, country of assessment in two classes (Iceland or Sweden/other countries) and year of assessment in five classes (< 2010, 2010–2015, 2016, 2017 or 2018). The division of year of assessment classes was based on change in how the back and croup phenotype was to be assessed, and number of horses in the data from different years. The DMRT3 genotype was also tested as an effect. None of these fixed effects were found to be significant ($p \leq 0.05$) for the trait back and croup and were thus not included in the GWA models. To investigate potential stratification, a multidimensional scaling (MDS) plot was constructed based on a genomic relationship matrix using the GenABEL package and ibs() function [69]. No outliers were apparent on the MDS plot and no stratification of horses with low and high score of back and croup was detected. A

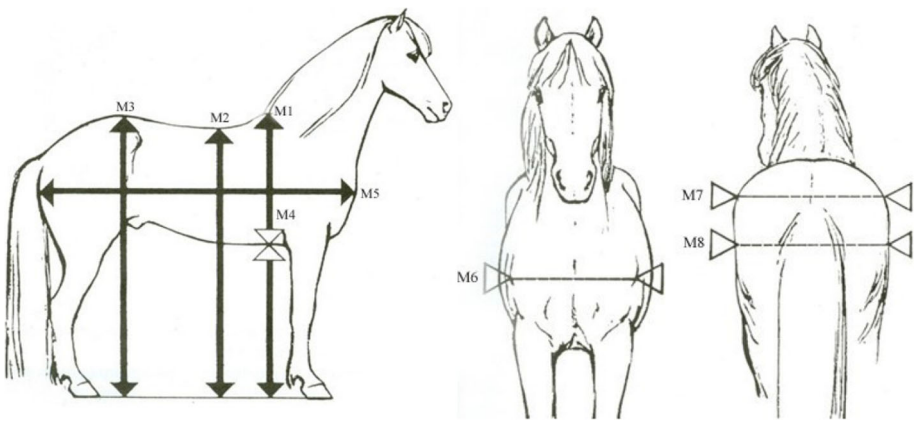


Fig. 4 Zoometric measurements recorded at standardized breeding field tests for Icelandic horses [68]. Original images created by Pétur Behrens. Images used in the figure are privately owned and thus were not taken from previously published sources requiring written permission for use

visualization of the genomic-kinship matrix using MDS is shown in Fig. 5.

The genomic-kinship matrix together with the phenotype of back and croup were passed to the polygenic_hglm function using family gaussian in GenABEL [69, 71]. To account for any population stratification, the GWA analysis was performed using a mixed model-structured association approach with the mmscore function in GenABEL [69]. Genome-wide significance was determined by Bonferroni correction and a suggestive genome-wide significance threshold was set at 1.0×10^{-5} [72, 73]. QQ and linkage disequilibrium (LD) manhattan plots were performed using the package cgmisc 2.0 [74].

Haplotype analysis

Haplotype analysis was performed with the haplo.stats package in R [70]. A linkage disequilibrium plot was constructed and the ten significant SNPs in LD ($r^2 \geq 0.8$) were used in the function haplo.em to estimate haplotypes. The haplotype effect on the score of back and croup was estimated by a gereralized linear model (glm) with the function haplo.glm.. The most frequent haplotype was used as a reference and only haplotypes with frequencies greater than 0.02 were included. A simulated *p*-value was estimated by using 100,000 permutations considering an additive effect.

Phenotype association of significant haplotypes

Phenotype association of the horses homozygous for the haplotypes that had a significant effect on the conformation of back and croup was performed using a two-tailed Student's t-test in R [70]. Significance level was set at p-value ≤0.05. Traits tested were all the zoometric traits, the subjectively scored riding ability traits and the subjectively assessed sub-traits.

Genotyping of the top SNP and *DMRT3* in other gaited and partly gaited breeds

Horses of other gaited breeds (Rocky-Mountain: 36 horses, Colombian paso fino horses: 38 horses) and partly gaited breeds (American Curly: 27 horses, American Saddlebred: 42 horses, Morgan: 30 horses and Standardbred: 40 horses) were genotyped for the top SNP using StepOnePlus Real-Time PCR System (Life Technologies) with a custom TaqMan SNP genotyping assay (Applied Biosystems). A group of 49 Icelandic horses used for riding but that had not attended breeding field test was also genotyped. The sequence of the primers and probes was designed as follows: forward primer: 5′-GGAAGTTTCTAAACATTTTTGAAGGC TTTT-3′; reverse primer: GGAGGGAAGTCAATTGAC AAACG; mutant probe (FAM): 5′-CCTCCACGGC ATCA-3′; reference probe (VIC): 5′-TCCCTCCACA GCATCA-3′. The reaction volume of 15 μl contained: 1.5 μl DNA, 0.38 μl Genotyping Assay 40X, 7.50 μl Genotyping Master Mix 2X, and 5.62 μl deionized water. The thermal cycle included 95 °C for 10 min, 40 cycles of 95 °C for 15 s, and 60 °C for 1 min.

SNP genotyping of the *DMRT3_Ser301STOP* marker known as the "Gait Keeper" mutation was performed using custom designed TaqMan SNP Genotyping Assays (Applied Biosystem) as described previously [30, 34].

Functional annotation

The bioinformatics database NCBI was used to screen for candidate genes based on the EquCab3.0 reference

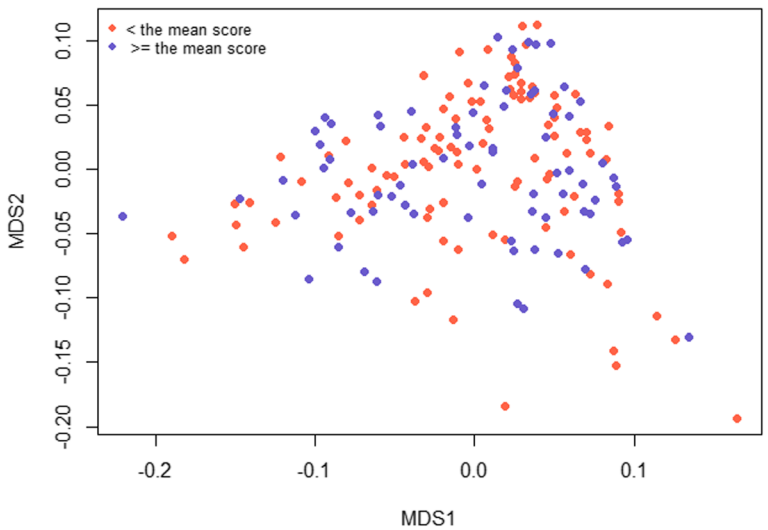


Fig. 5 MDS plot for the score back and croup. Visualization of population stratification across the 177 Icelandic horses that passed the QC for the score back and croup. Red represents horses that had a score lower than the mean 8.1 and blue represents horses that had a score higher or equal to 8.1

genome and annotation release 103 [75] and HorseQTLdb release 41 to search for known quantitative trait loci (QTLs) for conformation in horses [39]. Functional annotation of possible candidate genes was performed using the GeneCards database [76]. All positions refer to the EquCab3.0 reference genome.

Abbreviations

ADAMTS17: *ADAM Metallopeptidase With Thrombospondin Type 1 Motif 17*; AF: Allele frequency; AIS: Adolescent idiopathic scoliosis; *ANKRD1*: *Ankyrin Repeat Domain 1*; *ANKRD60*: *Ankyrin repeat domain 60*; BLAST: Basic Local Alignment Search Tool; *C22H20orf85*: *Chromosome 22 C20orf85 homolog*; CPH: Colombian paso horses; *DMRT3*: *Doublesex And Mab-3 Related Transcription Factor 3*; ECA: *Equus caballus* chromosome; *GH1*: *Growth Hormone 1*; glm: Generalized linear model; GWA: Genome-wide association; *HMGA2*: *High Mobility Group AT-Hook 2*; *LASP1*: *LIM And SH3 Protein 1*; *LCORL*: *Ligand Dependent Nuclear Receptor Corepressor Like*; LD: Linkage disequilibrium; MAF: Minor allele frequency; MDS: Multidimensional scaling; *NCAPG*: *Non-SMC Condensin I Complex Subunit G*; *OSTN*: *Osteocrin*; *PPP4R1L*: *Protein Phosphatase 4 Regulatory Subunit 1 Like*; QC: Quality control; QTL: Quantitative trait loci; SD: Standard deviation; SIF: The Swedish Icelandic Horse Association; *SLC22A8*: *Solute carrier family 22 member 8*; SNP: Single nucleotide polymorphism; VR: The Swedish Research Council; *ZFAT*: *Zinc Finger And AT-Hook Domain Containing*

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-07454-z.

Additional file 1.

Acknowledgements

The authors would like to acknowledge all horse owners and The Array and Analysis Facility, Department of Medical Sciences, Uppsala University. Heimir Gunnarsson breeding leader of the Icelandic horse breeding association in Sweden for sharing valuable information and thoughts. Tytti Vanhala at SLU for performing genotyping and Anna Johansson at the National Bioinformatics Infrastructure Sweden at SciLifeLab for bioinformatics advice.

Authors' contributions

GL and SE initiated and designed the study. MKR, MS and JJN collected the samples. MKR extracted the DNA. EA provided the phenotypic data from WorldFengur. MKR and HS performed the experiments and data analysis. MS, MKR and HS drafted the manuscript. AJ performed genotyping in other gaited breeds. RN and AJ performed the allele frequency analysis. MNB contributed with samples from CPH. SE and MS advised with the statistical analysis. MKR, HS, SE, MNB, EA, TK, MR, ÅV, BDV, JJN, MS and GL contributed to the interpretation of the results. All authors read, suggested modifications and approved the final manuscript.

Funding

The Swedish Research Council (VR), The Swedish Research Council for Environment, Agricultural Science and Spatial Planning (Formas), The Swedish Icelandic Horse Association SIF, The Swedish-Norwegian Foundation for Equine Research, The Agricultural University of Iceland's Doctoral Fund, The Stock Protection Fund for the Icelandic Horse Breed and Blikastaðir Fund. Open Access funding provided by Swedish University of Agricultural Sciences.

Availability of data and materials

The datasets generated and analyzed during the current study are not publicly available since the study was performed in collaboration with the Icelandic horse breeding industry and has a commercial value for them. However, data is available from the corresponding author on reasonable request and with permission of the Icelandic horse association.

Declaration

Ethics approval and consent to participate

Hair samples were collected following owner's informed written consent and according to ethical approval by the Ethics Committee for Animal Experiments in Uppsala, Sweden (number: 5.8.18–15453/2017).

Consent for publication

Not applicable.

Competing interests

The authors declare competing interest concerning commercial applications of the current study. GL is a co-inventor on a patent application concerning commercial testing of the DMRT3 mutation. The stated patents does not restrict research applications of the method.

Author details

[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden. [2]The Agricultural University of Iceland, Borgarnes, Iceland. [3]Genética Animal de Colombia Ltda, Bogotá, Colombia. [4]The Icelandic Agricultural Advisory Centre, Reykjavík, Iceland. [5]Department of Anatomy, Physiology and Biochemistry, Swedish University of Agricultural Sciences, Uppsala, Sweden. [6]School of Life & Environmental Sciences, University of Sydney, Sydney, Australia. [7]Department of Evolutionary Ecology, Doñana Biological Station, CSIC, Seville, Spain. [8]Livestock Genetics, Department of Biosystems, KU Leuven, Leuven, Belgium.

Received: 17 September 2020 Accepted: 19 February 2021
Published online: 14 April 2021

References

1. Albertsdóttir E, Eriksson S, Näsholm A, Strandberg E, Árnason T. Genetic correlations between competition traits and traits scored at breeding field-tests in Icelandic horses. Livest Sci. 2008;114(2–3):181–7.
2. Solé M, Gómez MD, Galisteo AM, Santos R, Valera M. Kinematic characterization of the Menorca horse at the walk and the trot: influence of hind limb pastern angle. J Equine Vet Sci. 2013;33(9):726–32.
3. Kristjansson T, Bjornsdottir S, Albertsdóttir E, Sigurdsson A, Pourcelot P, Crevier-Denoix N, et al. Association of conformation and riding ability in Icelandic horses. Livest Sci. 2016;189:91–101.
4. Holmström M, Back W. The effects of conformation. In: Equine Locomotion. 2nd ed. London: WB Saunders; 2013. p. 229–43.
5. Janczarek I, Wilk I, Strzelec K. Correlations between body dimensions of young trotters and motion parameters and racing performance. Pferdeheilkunde. 2017;33(2):139–45.
6. Sánchez-Guerrero MJ, Molina A, Gómez MD, Peña F, Valera M. Relationship between morphology and performance: signature of mass-selection in Pura Raza Español horse. Livest Sci. 2016;185:148–55.
7. Sánchez-Guerrero MJ, Cervantes I, Molina A, Gutiérrez JP, Valera M. Designing an early selection morphological linear traits index for dressage in the Pura Raza Español horse. Animal. 2017;11(6):948–57.
8. Jönsson L, Egenvall A, Roepstorff L, Näsholm A, Dalin G, Philipsson J. Associations of health status and conformation with longevity and lifetime competition performance in young Swedish Warmblood riding horses: 8,238 cases (1983-2005). J Am Vet Med Assoc. 2014 Jun 15;244(12):1449–61.
9. Jönsson L, Näsholm A, Roepstorff L, Egenvall A, Dalin G, Philipsson J. Conformation traits and their genetic and phenotypic associations with health status in young Swedish warmblood riding horses. Livest Sci. 2014 May 1;163(1):12–25.
10. Koenen EPC, Van Veldhuizenb AE, Brascamp EW. Genetic parameters of linear scored conformation traits and their relation to dressage and show-jumping performance in the Dutch Warmblood Riding Horse population. Livest Sci. 1995;43:85–94.
11. Rustin M, Janssens S, Buys N, Gengler N. Multi-trait animal model estimation of genetic parameters for linear type and gait traits in the Belgian warmblood horse. J Anim Breed Genet. 2009;126(5):378–86.
12. FEIF. FEIF General Rules and Regulations. Available from: https://www.feiffengur.com/documents/FEIF_Rules_Regulations2020_complete.pdf [cited 2020 Aug 6].

13. Albertsdóttir E, Eriksson S, Sigurdsson Á, Árnason T. Genetic analysis of "breeding field test status" in Icelandic horses. J Anim Breed Genet. 2011; 128(2):124–32.
14. François L, Fegraeus KJ, Eriksson S, Andersson LS, Tesfayonas YG, Viluma A, et al. Conformation traits and gaits in the icelandic horse are associated with genetic variants in Myostatin (MSTN). J Hered. 2016;107(5):431–7.
15. Makvandi-Nejad S, Hoffman GE, Allen JJ, Chu E, Gu E, Chandler AM, et al. Four loci explain 83% of size variation in the horse. PLoS One. 2012;7(7):1–6.
16. Weedon MN, Lango H, Lindgren CM, Wallace C, David M, Mangino M, et al. Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet. 2008;40(5):575–83.
17. Metzger J, Schrimpf R, Philipp U, Distl O. Expression levels of LCORL are associated with body size in horses. PLoS One. 2013;8(2):56497.
18. Frischknecht M, Jagannathan V, Plattet P, Neuditschko M, Signer-Hasler H, Bachmann I, et al. A non-synonymous HMGA2 variant decreases height in shetland ponies and other small horses. PLoS One. 2015;10(10):1–11.
19. Hayward JJ, Castelhano MG, Oliveira KC, Corey E, Balkman C, Baxter TL, et al. Complex disease and phenotype mapping in the domestic dog. Nat Commun. 2016;7.
20. Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, et al. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat Genet. 2018;50(3):362–7.
21. Mostafavi A, Asadi Fozi M, Esmailizadeh Koshkooieh A, Mohammadabadi M, Ivanivna Babenko O, Ihorivna Klopenko N. Effect of LCORL gene polymorphism on body size traits in horse populations. Acta Sci Anim Sci. 2020;42. https://doi.org/10.4025/actascianimsci.v42i1.47483.
22. Abri MA Al Holl HM, Kalla SE, Sutter NB, Brooks SA. Whole genome detection of sequence and structural polymorphism in six diverse horses. PLoS One 2020;15(4). https://doi.org/10.1371/journal.pone.0230899.
23. Posbergh CJ, Huson HJ. All sheeps and sizes: a genetic investigation of mature body size across sheep breeds reveals a polygenic nature. Anim Genet. 2020;52:99–107.
24. Signer-Hasler H, Flury C, Haase B, Burger D, Simianer H, Leeb T, et al. A genome-wide association study reveals loci influencing height and other conformation traits in horses. PLoS One. 2012;7(5):3–8.
25. Tetens J, Widmann P, Kühn C, Thaller G. A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses. Anim Genet. 2013;44(4):467–71.
26. Metzger J, Rau J, Naccache F, Bas Conn L, Lindgren G, Distl O. Genome data uncover four synergistic key regulators for extremely small body size in horses. BMC Genomics. 2018;19(1):1–15.
27. Gmel AI, Druml T, von Niederhäusern R, Leeb T, Neuditschko M. Genome-wide association studies based on equine joint angle measurements reveal new QTL affecting the conformation of horses. Genes (Basel). 2019;10(5):370.
28. Novoa-Bravo M, Jäderkvist Fegraeus K, Rhodin M, Strand E, García LF, Lindgren G. Selection on the Colombian paso horse's gaits has produced kinematic differences partly explained by the DMRT3 gene. PLoS One. 2018; 13(8):e0202584.
29. Jäderkvist K, Kangas N, Andersson LS, Lindgren G. Gaitedness is associated with the DMRT3 'gait keeper' mutation in Morgan and American curly horses. Anim Genet. 2014;45(6):908–9.
30. Promerová M, Andersson LS, Juras R, Penedo MCT, Reissmann M, Tozaki T, et al. Worldwide frequency distribution of the ' *Gait keeper* ' mutation in the *DMRT3* gene. Anim Genet. 2014;45(2):274–82.
31. Velie BD, Shrestha M, François L, Schurink A, Tesfayonas YG, Stinckens A, et al. Using an Inbred Horse Breed in a High Density Genome-Wide Scan for Genetic Risk Factors of Insect Bite Hypersensitivity (IBH). PLoS One. 2016; 11(4):e0152966 Barendse W, editor.
32. Finno CJ, Stevens C, Young A, Affolter V, Joshi NA, Ramsay S, et al. SERP INB11 Frameshift Variant Associated with Novel Hoof Specific Phenotype in Connemara Ponies. PLoS Genet. 2015;11(4):e1005122 Barsh GS, editor.
33. Ablondi M, Viklund Å, Lindgren G, Eriksson S, Mikko S. Signatures of selection in the genome of Swedish warmblood horses selected for sport performance. BMC Genomics. 2019;20(1):717.
34. Andersson LS, Larhammar M, Memic F, Wootz H, Schwochow D, Rubin CJ, et al. Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. Nature. 2012;488(7413):642–6.
35. Fawcett JA, Sato F, Sakamoto T, Iwasaki WM, Tozaki T, Innan H. Genome-wide SNP analysis of Japanese Thoroughbred racehorses. PLoS One. 2019; 14(7). https://doi.org/10.1371/journal.pone.0218407.
36. Sadeghi R, Moradi-Shahrbabak M, Reza Miraei Ashtiani S, Schlamp F, Cosgrove EJ, Antczak DF, et al. Genetic Diversity of Persian Arabian Horses and Their Relationship to Other Native Iranian Horse Breeds. J Hered. 2019; 110:173–82.
37. Velie BD, Lillie M, Fegraeus KJ, Rosengren MK, Solé M, Wiklund M, et al. Exploring the genetics of trotting racing ability in horses using a unique Nordic horse model. BMC Genomics. 2019;20(1):104.
38. Velie BD, Fegraeus KJ, Solé M, Rosengren MK, Røed KH, Ihler C-F, et al. A genome-wide association study for harness racing success in the Norwegian-Swedish coldblooded trotter reveals genes for learning and energy metabolism. BMC Genet. 2018;19(1). https://doi.org/10.1186/s12863-018-0670-3.
39. Horse QTL Database. Available from: https://www.animalgenome.org/cgi-bin/QTLdb/EC/index. [cited 2020 Aug 6]
40. Jönsson L. Orthopaedic health, conformation and longevity in riding horses. Doctoral thesis. SLU; 2013.
41. Liu J, Zhou Y, Liu S, Song X, Yang XZ, Fan Y, et al. The coexistence of copy number variations (CNVs) and single nucleotide polymorphisms (SNPs) at a locus can result in distorted calculations of the significance in associating SNPs to disease. Hum Genet. 2018;137(6–7):553–67.
42. Kim JJ, Lee HI, Park T, Kim K, Lee JE, Cho NH, et al. Identification of 15 loci influencing height in a Korean population. J Hum Genet. 2010 Jan 6;55(1): 27–31.
43. Konieczny MR, Senyurt H, Krauspe R. Epidemiology of adolescent idiopathic scoliosis. J Child Orthop. 2013;7:3–9 Springer.
44. Martínez-Llorens J, Ramírez M, Colomina MJ, Bagó J, Molina A, Cáceres E, et al. Muscle dysfunction and exercise limitation in adolescent idiopathic scoliosis. Eur Respir J 2010;36:393–400.
45. Milenkovic SM, Kocijancic RI, Belojevic GA. Left handedness and spine deformities in early adolescence. Eur J Epidemiol. 2004;19:969–72.
46. Grivas T, Mihas K, Vasiliadis E, Maziotou C, Karathanou S, Polyzois V. Handedness and laterality of trunk Rotation in children screened for scoliosis. Proceedings of the International Research Society of Spinal Deformities conference; 2004 June 10-12; Biennial Meeting. 2004:64–7.
47. de Yang Z, Li M. There may be a same mechanism of the left-right handedness and left-right convex curve pattern of adolescent idiopathic scoliosis. Med Hypotheses. 2011;76(2):274–6.
48. Catanzariti JF, Guyot MA, Agnani O, Demaille S, Kolanowski E, Donze C. Eye-hand laterality and right thoracic idiopathic scoliosis. Eur Spine J. 2014;23(6): 1232–6.
49. Grivas TB, Vasiliadis ES, Polyzois VD, Mouzakis V. Trunk asymmetry and handedness in 8245 school children. Pediatr Rehabil. 2006;9(3):259–66.
50. Ocklenburg S, Schmitz J, Moinfar Z, Moser D, Klose R, Lor S, et al. Epigenetic regulation of lateralized fetal spinal gene expression underlies hemispheric asymmetries. Elife. 2017;6. https://doi.org/10.7554/eLife.22784.
51. Jeffcott LB. Disorders of the thoracolumbar spine of the horse — a survey of 443 cases. Equine Vet J. 1980;12(4):197–210.
52. Wong D, Miles K, Sponseller B. Congenital Scoliosis In A Quarter Horse Filly. Vet Radiol Ultrasound. 2006;47(3):279–82.
53. Lucidi P, Bacco G, Sticco M, Mazzoleni G, Benvenuti M, Bernabò N, et al. Assessment of motor laterality in foals and young horses (Equus caballus) through an analysis of derailment at trot. Physiol Behav. 2013;109(1):8–13.
54. McGreevy PD, Thomson PC. Differences in motor laterality between breeds of performance horse. Appl Anim Behav Sci. 2006;99(1–2):183–90.
55. Cully P, Nielsen B, Lancaster B, Martin J, McGreevy P. The laterality of the gallop gait in Thoroughbred racehorses. PLoS One. 2018;13(6):e0198545 Bartos L, editor.
56. Regatieri IC, Eberth JE, Sarver F, Lear TL, Bailey E. Comparison of *DMRT3* genotypes among American Saddlebred horses with reference to gait. Anim Genet. 2016 Oct 1;47(5):603–5.
57. Nicodemus MC, Clayton HM. Temporal variables of four-beat, stepping gaits of gaited horses. Appl Anim Behav Sci. 2003;80(2):133–42.
58. Al Abri MA, Posbergh C, Palermo K, Sutter NB, Eberth J, Hoffman GE, et al. Genome-Wide Scans Reveal a Quantitative Trait Locus for Withers Height in Horses Near the ANKRD1 Gene. J Equine Vet Sci. 2018;60:67–73 e1.
59. BLAST: Basic Local Alignment Search Tool. Available from: https://blast.ncbi.nlm.nih.gov/Blast.cgi. [cited 2020 Aug 7]
60. Gaudet P, Livstone MS, Lewis SE, Thomas PD. Phylogenetic-based propagation of functional annotations within the gene ontology consortium. Brief Bioinform. 2011;12(5):449–62.
61. Tachmazidou I, Süveges D, Min JL, Ritchie GRS, Steinberg J, Walter K, et al. Whole-genome sequencing coupled to imputation discovers genetic signals for anthropometric traits. Am J Hum Genet. 2017;100(6):865–84.

62. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155(4):934.
63. Hoyt DF, Wickler SJ, Cogger EA. Time of contact and step length: the effect of limb length, running speed, load carrying and incline. J Exp Biol. 2000; 203(2):221–7.
64. Sparrow LM, Pellatt E, Yu SS, Raichlen DA, Pontzer H, Rolian C. Gait changes in a line of mice artificially selected for longer limbs. PeerJ. 2017;2017(2): e3008.
65. Pontzer H. Effective limb length and the scaling of locomotor cost in terrestrial animals. J Exp Biol. 2007;210(10):1752–61.
66. Cook D, Gallagher PC, Bailey E. Genetics of swayback in American Saddlebred horses. Anim Genet. 2010;41(SUPPL. 2):64–71.
67. WorldFengur The Studbook of Origin for the Icelandic horse. Available from: https://www.worldfengur.com/. [cited 2020 Aug 6]
68. The Icelandic Agricultural Advisory Centre. General Information on Breeding Field Tests – Measurements (body measures). Available from: https://www.rml.is/static/files/Hrossaraekt_RML/skrokkmal.pdf. [cited 2020 Aug 6]
69. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007;23(10):1294–6.
70. R Core Team. R: a language and environment for statistical computing. Vienna: Foundation for Statistical Computing; 2015.
71. Rönnegård L, Shen X, Alam M. Hglm: a package for fitting hierarchical generalized linear models. R J. 2010;2(2):20–8.
72. Lander E, Kruglyak L. Genetic dissection of complex traits : reporting linkage results. Nat Genet. 1995;11:241–7.
73. Duggal P, Gillanders EM, Holmes TN, Bailey-Wilson JE. Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. BMC Genomics. 2008;9:1–8.
74. Kierczak M, Jabłońska J, Forsberg SKG, Bianchi M, Tengvall K, et al. cgmisc: enhanced genome-wide association analyses and visualization. Bioinformatics. 2015;31:3830–1.
75. National Center for Biotechnology Information. Available from: https://www.ncbi.nlm.nih.gov/. [cited 2020 Aug 6]
76. Stelzer G, Rosen R, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. Current Protocols in Bioinformatics. 2016;54:1.30.1–1.30.33. Available from: https://www.genecards.org/. [cited 2020 Aug 6]

Publisher's Note

II

Sigurðardóttir *et al.*
Genetics Selection Evolution (2023) 55:89
https://doi.org/10.1186/s12711-023-00863-6

Genetics Selection Evolution

RESEARCH ARTICLE

Open Access

The genetics of gaits in Icelandic horses goes beyond *DMRT3*, with *RELN* and *STAU2* identified as two new candidate genes

Heiðrún Sigurðardóttir[1,2*], Henrik Boije[3], Elsa Albertsdóttir[4], Thorvaldur Kristjansson[2], Marie Rhodin[5], Gabriella Lindgren[1,6] and Susanne Eriksson[1]

Abstract

Background In domesticated animals, many important traits are complex and regulated by a large number of genes, genetic interactions, and environmental influences. The ability of Icelandic horses to perform the gait 'pace' is largely influenced by a single mutation in the *DMRT3* gene, but genetic modifiers likely exist. The aim of this study was to identify novel genetic factors that influence pacing ability and quality of the gait through a genome-wide association study (GWAS) and correlate new findings to previously identified quantitative trait loci (QTL) and mutations.

Results Three hundred and seventy-two Icelandic horses were genotyped with the 670 K+ Axiom Equine Genotyping Array, of which 362 had gait scores from breeding field tests. A GWAS revealed several SNPs on *Equus caballus* chromosomes (ECA) 4, 9, and 20 that were associated ($p < 1.0 \times 10^{-5}$) with the breeding field test score for pace. The two novel QTL on ECA4 and 9 were located within the *RELN* and *STAU2* genes, respectively, which have previously been associated with locomotor behavior in mice. Haplotypes were identified and the most frequent one for each of these two QTL had a large favorable effect on pace score. The second most frequent haplotype for the *RELN* gene was positively correlated with scores for tölt, trot, gallop, and canter. Similarly, the second most frequent haplotype for the *STAU2* gene had favorable effects on scores for trot and gallop. Different genotype ratios of the haplotypes in the *RELN* and *STAU2* genes were also observed in groups of horses with different levels of pacing ability. Furthermore, interactions ($p < 0.05$) were detected for the QTL in the *RELN* and *STAU2* genes with the *DMRT3* gene. The novel QTL on ECA4, 9, and 20, along with the effects of the *DMRT3* variant, were estimated to account jointly for 27.4% of the phenotypic variance of the gait pace.

Conclusions Our findings provide valuable information about the genetic architecture of pace beyond the contribution of the *DMRT3* gene and indicate genetic interactions that contribute to the complexity of this trait. Further investigation is needed to fully understand the underlying genetic factors and interactions.

*Correspondence:
Heiðrún Sigurðardóttir
heidrun.sigurdardottir@slu.se; heidruns@lbhi.is
Full list of author information is available at the end of the article

Background

Many important traits in domesticated animals are quantitative, complex traits, including performance traits in horses. Complex traits are generally regulated by a large number of genes and influenced by environmental factors. Interactions between loci and between genes, known as epistasis, may also contribute to the phenotypic variation of a trait [1, 2]. Previous studies suggest that accurate prediction of complex trait phenotypes based on genotypes requires knowledge about the existence of epistasis that may influence the trait [3, 4]. In the case of an interaction between major genes, the phenotype cannot be predicted simply by adding the effects of each locus [2, 5, 6]. In spite of an increasing number of studies on epistasis, as exemplified by the aforementioned studies, the discovery of genetic interactions on a genome-wide scale remains a major challenge [6].

In this study, we further investigated the genetic background of the economically valuable gait trait 'pace' in Icelandic horses. This is a complex trait that has previously been shown to be largely influenced by a single mutation in the *doublesex and mab-3 related transcription factor 3* (*DMRT3*) gene, and to be potentially influenced by epistatic effects [7].

Gait versatility is a well-known characteristic of the Icelandic horse breed. Its unique ability to perform five gaits, including the lateral gaits pace and tölt, is one of the hallmarks of the breed. Pace is a gait with a suspension phase and should ideally be ridden at high speed, making it a popular gait for racing. However, the ability to pace varies between individuals in the breed, and some individuals seem to lack it and only perform the four gaits walk, trot, canter/gallop, and tölt. Because of this, it is common to refer to Icelandic horses as being either four- or five-gaited.

The discovery that a single base change in the *DMRT3* gene influences gait variability in the Icelandic horse breed has played a key role in understanding its underlying mechanisms [7]. The 'gait keeper' mutation alters the pattern of locomotion and has a predominant effect on gaiting ability, in that one copy of the mutant allele [allele *A* at the *DMRT3*_Ser301STOP marker at nucleotide position 22,999,655 on *Equus caballus* chromosome (ECA) 23] enhances the natural ability to tölt, and two copies enable the development of pace [7, 8]. It has also been noted that scores from breeding field tests for the basic gaits (walk, trot, and canter/gallop) are negatively impacted by the *AA* genotype of the *DMRT3*-variant [7, 8]. In 2012, the frequency of the *A*-allele in a selected group of Icelandic horses was estimated to be 0.94 [8] and the fact that this allele is not fixed within the breed explains some of the variation in pace ability. However, there is a high proportion (> 30%) of homozygous *AA* horses that do not perform pace [7–9], and these are referred to as four-gaited in spite of their genetic precondition to develop pace. In addition, genetic variation in scores for pace among horses that did perform this gait at breeding field tests was detected in Icelandic horses [10]. Estimates of the heritability for gait scores in Icelandic horses, which are subjectively assessed by certified judges at standardized breeding field tests [11], have been reported to range from 0.18 (walk) to 0.60 (pace) [10, 12], which indicates that genetic factors as well as environmental factors influence gait quality. This is supported by the considerable genetic improvement that has been obtained in Icelandic horses during the last decades, from selection on routinely estimated breeding values for these assessed gait scores [13].

Conformation traits have been shown to discriminate between four- and five-gaited horses that are homozygous *AA* for *DMRT3* in the Icelandic horse breed to some extent [14]. Rosengren et al. [15] identified a quantitative trait locus (QTL) on ECA22 that is associated with an inclination of the backline, form of the croup, and an uphill conformation in the Icelandic horse breed. This QTL was also shown to have a considerable effect on the quality of the lateral gaits tölt and pace, which further supports the effects of conformation on gaits and highlighting the complexity of the gait traits.

Studies on other gaited horse breeds have investigated genetic differences between horses with different gait patterns but the same *DMRT3* genotype [16–18]. A pilot study of four- and five-gaited Icelandic horses with the *AA* genotype at the *DMRT3* gene did not reach conclusive results due to a rather small number of horses (20 four-gaited and 35 five-gaited horses) [19]. Thus, knowledge about the genetic background of gaiting ability and the quality of gaits in Icelandic horses is still incomplete. The aim of the present study was to further investigate the genetic background of pace in Icelandic horses, using a larger dataset to identify novel genetic factors that influence pacing ability and quality through a genome-wide association study (GWAS). A second aim was to investigate how the new identified QTL interacted with previously reported QTL and mutations. This is a first step towards understanding the complex genetic background that underlies the phenotype of pace beyond the *DMRT3* gene, or possibly interacting with it.

Methods

Animals

The study included 372 privately owned Icelandic horses born between 1993 and 2016, among which 160 were stallions or geldings, and 212 were mares. All horses had been shown at a breeding field test, except for 10 geldings that were specialized in pace racing. Hair samples

from some of the horses were originally collected for an unrelated study with selection of horses based on the mane and tail characteristics. The remaining horses were selected from breeding field tests in 2020 with inclusion of both individuals with high and low assessment scores from the tests. Hair samples were collected from the horses' tails, with informed consent from the owners. Collection was done at breeding field tests and during visits to trainers and breeders in Iceland and Sweden. Pedigree data were obtained from the international Worldfengur database [20]. The maximum relatedness between individuals in the sample was a half-sib relationship, and efforts were made to balance the contributions from different families and avoid stratification in the data. The number of unique sires and dams for the individuals in the study was 244 and 362, respectively.

Phenotype data

The phenotype data used were retrieved from the Worldfengur database [20] and consisted of scores for pace recorded at standardized breeding field tests for Icelandic horses between 1999 and 2022 [11]. Scores for other gaits assessed at breeding field tests were also used to further investigate the effects of the genomic regions that were found to be associated with pace. In cases where horses had been assessed at more than one breeding field test, the highest assessment score was used. Most horses were assessed in Iceland (N = 269) and the others were assessed in Denmark (N = 1), Germany (N = 6), Norway (N = 1), and Sweden (N = 85). The age of the horses when they attended a breeding field test ranged from 4 to 15 years (mean = 6.7 years).

Pace

Pace is described as a symmetrical, two-beat gait in which ipsilateral legs move nearly synchronously back and forth with a brief moment of suspension. Pace is an energetic gait ridden at high speed, where the horse lengthens its strides, lifts its back, and extends the head and neck forward. Pure pace is characterized by a visible suspension phase where none of the hoofs touch the ground and divergence from synchronous movements of the ipsilateral legs is not noticeable [21].

The quality of pace is subjectively assessed by a panel of certified judges at standardized breeding field tests and scored on a scale from 5 to 10 with 0.5 intervals [11]. To receive the highest score for pace, the horse should be able to pace steadily in good balance, with long strides, elegant and light movements, good suspension, and excellent speed. Correct body function and a long, strong topline where the horse extends its head and neck forward are of great importance for the higher scores. A faulty and/or uneven beat in pace, lack of speed, stiff movements, short strides, and a concave topline are considered poor qualities of pace and entail lower scores. A score of 5 is given only if the horse does not present the pace trait [11].

The scores for pace that the horses received in this study ranged from 5.0 to 10.0, with a mean of 7.0 [standard deviation (SD) 1.55]. The distribution of the scores for pace higher than 5.0 was slightly negatively skewed (−0.13) (Fig. 1). Transformation of the data to increase normality was tested but did not affect the results. The distributions after rank- and log transformation of the pace scores are presented in Additional file 1: Fig. S1.

Other traits assessed at breeding field tests

In addition to pace, scores for the gait traits tölt, trot, canter, gallop, and walk were used to investigate the effects of the detected genomic regions associated with pace. As for pace, the quality of these gaits was subjectively assessed by the judging panel and scored on a scale from 5 to 10 with 0.5 intervals [11]. Tölt is a symmetrical four-beat ambling gait with an ipsilateral sequence of footfall and a large speed variation but without a suspension phase. Trot is a symmetrical two-beat, diagonal gait with a moment of suspension. Both canter and gallop are asymmetrical gaits with a suspension phase but canter is characterized as a three-beat, medium-speed gait, whereas gallop is defined as a four-beat, high-speed gait. Walk is a symmetrical, four-beat stepping gait with an ipsilateral movement and without suspension [21].

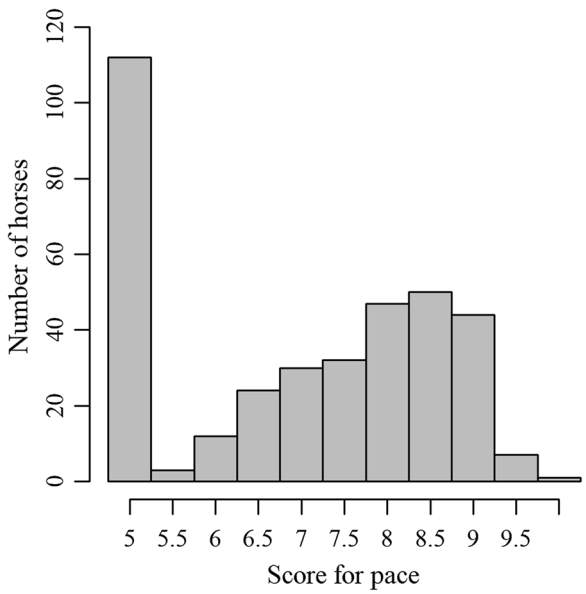


Fig. 1 Distribution of scores for pace. Distribution of scores for pace in the sample of 362 horses assessed at a breeding field test

Descriptive statistics and distributions of the scores for these traits are in Additional file 2: Table S1 and Additional file 3: Fig. S2. The distribution of these scores did not differ significantly from a normal distribution based on a Jarque–Bera normality test ($p > 0.05$), except the distribution of scores for tölt. However, transformation of the data to increase normality did not affect the results. The distributions after rank- and log transformation of the scores for tölt are in Additional file 4: Fig. S3.

Morphological measurements are also recorded at breeding field tests [22]. These measurements, made with a rod, include height at withers (M1), height at the lowest point of the back (M2), height at croup (M3), and depth of breast (M4) (Fig. 2), among others. These measurements are used to assess the longitudinal balance of the horse in standstill position and its level of uphill conformation by comparing the height of the front to that of the hind part (M1-M3) and by calculating height of the withers (M1-M2).

DNA preparation, genotyping, and quality control

DNA was extracted from approximately 25 hair roots from each horse, which were digested in 200 µl of 5% Chelex 100 Resin (Bio-Rad Laboratories, Hercules, CA, US) and 14 µl of proteinase K (20 mg/ml; Merck KgaA, Darmstadt, Germany) by incubation for 2 h at 56 °C at 900 rpm, followed by heat inactivation of the proteinase K at 95 °C for 10 min. After centrifugation, 150 µl from the top of each sample were withdrawn, leaving the Chelex and hair roots in the tube; to obtain more DNA, some of the samples were submitted to a second round of digestion, i.e. 100 µl of 5% Chelex 100 Resin and 7 µl of proteinase K were added and the sample were incubated again for 2 h at 56 °C. DNA concentration was determined using a NanoDrop-2000 (Thermo Fisher Scientific, Wilmington, DE, USA) and a Qubit 3.0 fluorometer (Life Technologies).

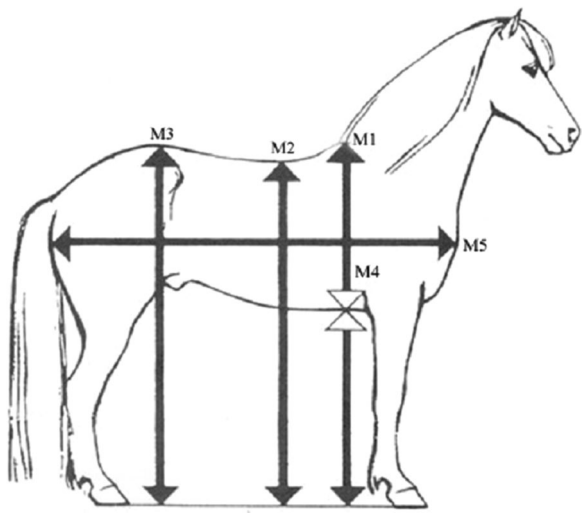


Fig. 2 Morphological measurements recorded at standardized breeding field tests for Icelandic horses [22]. Original image created by Pétur Behrens

The 372 DNA samples were genotyped with the 670 K+ Axiom Equine Genotyping Array. Quality control (QC) was performed using the GenABEL package [23, 24] in R (v.3.6.1) [25]. Poorly genotyped samples and noisy data were removed based on the following criteria: if the rate of missing genotypes per single nucleotide polymorphism (SNP) was higher than 0.10, the rate of missing SNPs per sample was higher than 0.10, the minor allele frequency (MAF) was lower than 0.05, and if there was deviation from Hardy-Weinberg equilibrium (p-value $\leq 1.0 \times 10^{-10}$). Genotyping of the *DMRT3*_Ser-301STOP SNP, known as the 'gait keeper' mutation [7], was performed manually on 25 samples that yielded a low call rate for this marker in the 670 K+ Axiom Equine Array genotyping, using custom-designed TaqMan SNP Genotyping Assays (Applied Biosystem). The 10-µl reaction volume contained: 1 µl DNA (concentration 150–170 ng/µl), 0.25 µl Genotyping Assay 40X, 5 µl Genotyping Master Mix 2X, and 3.75 µl deionized water and the following PCR conditions: 95 °C for 10 min, 40 cycles of 95 °C for 15 s, and 60 °C for 1 min. Results revealed that most of the sampled horses, i.e. 340, carried two copies of the mutant allele (*AA* genotype), 31 individuals carried one copy (*CA* genotype), and one individual carried none (*CC* genotype).

Genome-wide association study

A multidimensional scaling (MDS) plot was constructed based on a genomic-kinship matrix obtained with the ibs() function in the GenABEL package [23, 24] to identify potential stratification among the individuals. Different fixed effects were tested with the lm() function for a linear model in R (v.3.6.1) [25] using analysis of variance (ANOVA) as a post-hoc test, including sex (male or female), *DMRT3* genotype, age at assessment (4, 5, 6, and ≥ 7 years old), age in years at assessment as a covariate, country of assessment (Iceland versus other countries), and year of assessment (≤ 2016, 2017–2019, 2020, and 2021–2023). None of these effects were significant ($p \leq 0.05$) for the score for pace, except the effects of sex and of *DMRT3* genotype, and these were included in further analyses.

The GenABEL package [23, 24] in R (v.3.6.1) [25] was used to perform the GWAS for pace score. The genomic-kinship matrix obtained with the ibs() function, together with sex and *DMRT3* genotype as fixed effects, were used in a hierarchical generalized linear model for pace score with the polygenic_hglm() function [26]. This was

done to estimate residuals and the inverse of the variance-covariance matrix, to be subsequently used in a mixed model-structured association approach with the mmscore() function in the GenABEL package [23, 24]. Genome-wide significance was determined by Bonferroni correction ($p < 6.9 \times 10^{-8}$), with a suggestive genome-wide significance threshold set at 1.0×10^{-5} [27, 28].

An additional analysis was done using only pace scores for horses with the *AA* genotype for *DMRT3*. This did not alter the general results in terms of the detected genomic regions and, therefore, to retain more observations in the analysis, we chose to include all *DMRT3* genotypes and account for the fixed effects of *DMRT3* genotype in the GWAS.

Functional annotation of candidate genes

The genome browser Ensembl (release 108, Oct 2022) [29] was used to screen for candidate genes based on the EquCab3.0 reference genome and QTL annotation in the HorseQTLdb (release 49, Dec 2022) [30] was used to search for known QTL for gait and performance traits in horses. Functional annotation of possible candidate genes was performed using the GeneCards database (version 5.14, Jan 2023) [31, 32]. All positions refer to the EquCab3.0 reference genome.

Haplotype analyses

To visualize linkage disequilibrium (LD) between the significant SNPs from the GWAS, an LD Manhattan plot was constructed using the package cgmisc 2.0 [33] in R (v.3.6.1) [25]. Calculations for pairwise LD were also performed in PLINK (v.1.9) [34, 35] using the --r2 command. A haplotype analysis was performed with the haplo.stats package [36] in R (v.3.6.1) [25]. The function haplo.em() was used to estimate the frequency of the different haplotypes derived from significant SNPs that were in significant LD ($r^2 \geq 0.8$). The haplotype effect on the pace score was estimated using a generalized linear model with the function haplo.glm(). The most frequent haplotype was used as a reference and only haplotypes with frequencies higher than 0.02 were included. An empirical *p*-value was estimated by using 100,000 permutations based on an additive effect.

Phenotypic variation of the scores for pace was analyzed with the polygenic_hglm() function in the GenABEL package [23, 24] in R (v.3.6.1) [25] to quantify the variation explained by significant haplotypes and top SNPs. A phenotype association analysis of the horses that were homozygous for the haplotypes that had a significant effect on pace was conducted using a one-way ANOVA based on a general linear model procedure in R (v.3.6.1) [25]. The significance level was set at a *p*-value ≤ 0.05. All gaits assessed at breeding field tests were tested. Gait scores were corrected for the fixed effects of sex and *DMRT3* genotype.

The horses were divided into five groups based on their ability to pace and the haplotype frequency within each group was investigated. The groups comprised (1) all horses with the *AA* genotype at the *DMRT3*-variant (*AA* horses, N = 340), (2) horses with the *AA* genotype at the *DMRT3*-variant and that had shown pace at a breeding field test (5-gaited *AA* horses, N = 248), (3) horses with the *AA* genotype at the *DMRT3*-variant and that had been shown at a breeding field test but had not shown pace during the test (4-gaited *AA* horses, N = 82), (4) horses with the *CA* genotype at the *DMRT3*-variant (*CA* horses, N = 31), and (5) horses with the *AA* genotype at the *DMRT3*-variant that had raced in a pace competition but had not attended a breeding field test (pace racers, N = 10). In addition, the whole sample was entered as a reference group, including the single horse with the *CC* genotype at the *DMRT3*-variant (All horses, N = 372). Significance of differences in genotype frequencies between the groups was estimated with the pairwise_prop_test() function in the rstatix package [37] in R (v.3.6.1) [25], a post-hoc test following a significant chi-square test of homogeneity.

Interactions between the detected significant haplotypes and the 'gait keeper' mutation in the *DMRT3* gene [7], as well as the previously identified QTL on ECA22 associated with the conformation of the back and croup [15], were investigated for all gaits. Interaction effects were investigated by constructing boxplots with the ggplot2 package [38] in R (v.3.6.1) [25], using ANOVA and Tukey's test as a post-hoc test. The significance level was set at *p*-value ≤ 0.05.

Results

Genotyping and quality control

In total, 361,333 SNPs (358,655 on autosomes and 2678 on the X chromosome) and 372 horses passed QC and were included in the GWAS. One horse was excluded from the dataset because of too high identity-by-state (IBS) with another horse, i.e. greater than 0.95. No apparent outliers or stratification in the data was detected in the MDS plot (Fig. 3).

Genome-wide association analysis

Three SNPs, located on ECA4 between 4,222,615 and 4,228,914 bp, reached the suggestive threshold ($p < 1.0 \times 10^{-5}$) and two of them had high LD ($r^2 \geq 0.8$) (Fig. 4b, c). Seven SNPs on ECA9 between 11,533,922 and 13,457,268 bp also reached the suggestive threshold, of which one almost reached the Bonferroni significance threshold ($p < 6.9 \times 10^{-8}$) (Fig. 4b). Four of the top SNPs on ECA9 were in LD (Fig. 4d). One SNP reached

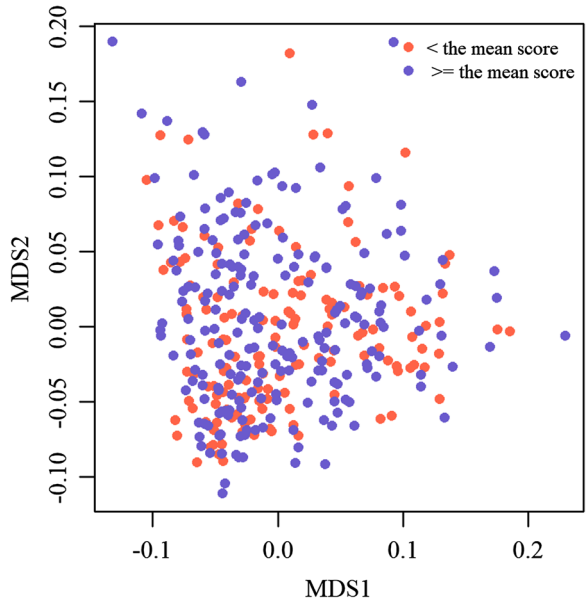


Fig. 3 MDS plot for pace score. Visualization of population stratification across the horses that passed quality control and had a record for the phenotype. Red represents horses that had a score lower than the mean 7.0 and blue represents horses with a score higher or equal to 7.0

the suggestive threshold on ECA20 at 52,057,378 bp, and was not in LD with any other SNP in the region (Fig. 4b). A summary of the GWA results for the 50 top SNPs is presented in Additional file 5: Table S2.

Candidate genes related to pace

The region detected on ECA4: 4,222,615–4,228,914 bp was located within the *reelin* (*RELN*) gene and the locations of the SNPs relative to the *RELN* gene's exon and intron positions are presented in Additional file 6: Fig. S4). Ten other genes were found in the vicinity of the detected region on ECA4 (see Additional file 7: Table S3). Regarding the region detected on ECA9: 11,533,922–13,457,268 bp, the five SNPs with the lowest *p*-value within this region were located within or in near proximity (~4 kb) of the annotated *staufen double-stranded RNA binding protein 2* (*STAU2*) gene; their locations relative to the gene's exon and intron positions are presented in Additional file 8: Fig. S5. Nine other genes were found in the vicinity of this region (see Additional file 7: Table S3). The region located on ECA20: 52,057,378 bp contained a single SNP that was located within a long non-coding RNA gene (*ENSECAG00000046047*) and 12 protein-coding genes and 12 other long non-coding RNA genes were found in the vicinity of this region (see Additional file 7: Table S3). None of the suggestive SNPs (on ECA4, 9, and 20) overlapped with any known QTL for gait traits in horses [30].

Haplotype analyses

Haplotype analysis for the identified region on ECA4 revealed two haplotypes that resulted in higher and lower scores for pace, respectively (p-value < 0.001) (Table 1). Ninety-nine horses were homozygous for the haplotype associated with a higher score for pace (R haplotype) and 65 horses were homozygous for the haplotype associated with a lower score for pace (r haplotype). The pooled group of rare haplotypes (haplotype frequency < 0.02) had a frequency of 0.01 (results not presented).

The haplotype analysis for the region identified on ECA9 also revealed two haplotypes which resulted in higher and lower scores for pace, respectively (p-value < 0.001) (Table 1). In total, 298 horses were homozygous for the haplotype associated with a higher score for pace (S haplotype) and five horses were homozygous for the haplotype associated with a lower score for pace (s haplotype). The pooled group of rare haplotypes had a frequency of 0.02 (results not presented).

No haplotype analysis was performed for the SNP identified on ECA20 since it was not in LD with any other SNP in the region. The allele frequencies of this SNP were 0.90 for allele *A* and 0.10 for allele *C*.

The proportion of phenotypic variation of the pace score explained by the significant haplotypes in the *RELN* and *STAU2* genes was 4.4% for each (Table 1). The proportion of phenotypic variation explained by the top SNP on ECA20 was 4.2%. The effects of *DMRT3* genotype combined with the effects of the significant haplotypes of the *RELN* and *STAU2* genes and the top SNP on ECA20 accounted for 27.4% of the phenotypic variance of the pace score. The *DMRT3*-variant explained 13.7% of the phenotypic variance.

(See figure on next page.)

Fig. 4 GWA results for the score for pace. **a** Quantile-quantile (QQ) plot where the blue lines represent the 0.05–0.95 confidence interval. The estimated lambda value was 0.99 (se 2.84×10^{-5}). **b** Manhattan plot from the mixed model association analysis. The red horizontal line indicates the Bonferroni significance threshold ($p < 6.9 \times 10^{-8}$) and the blue horizontal line indicates the suggestive genome-wide significance level ($p < 1.0 \times 10^{-5}$). **c** LD Manhattan plot on ECA4 with the top SNP as an open circle. Three SNPs reached the suggestive threshold of which two were in LD ($r^2 \geq 0.8$). **d** LD Manhattan plot on ECA9 with the top SNP as an open circle. Seven SNPs reached the suggestive threshold of which four were in LD ($r^2 \geq 0.8$). All positions refer to the EquCab3.0 genome assembly

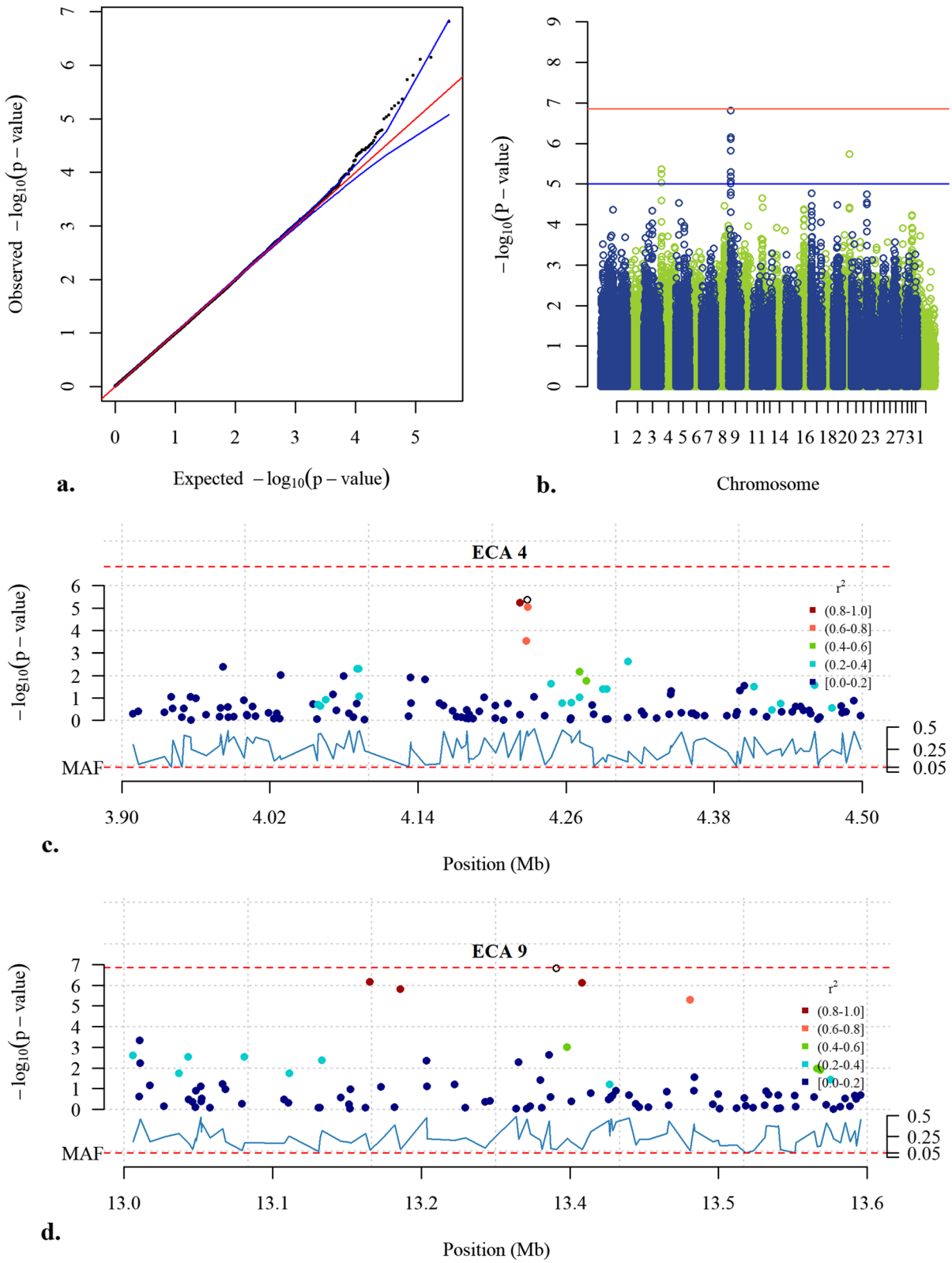


Fig. 4 (See legend on previous page.)

Table 1 Results from the haplotype analysis for SNPs on ECA4 and 9 associated with pace score

Chr	Hap. ID	Haplotypes (SNP number[a])				Freq	Coeff	*p*-value	Sim. *p*-value	σ_p^2 ratio
ECA4		1		2[b]						
	R	*C*		*C*		0.54	0.51	< 0.001	< 0.001	0.044
	r	*A*		*A*		0.45	−0.51	< 0.001	< 0.001	
ECA9		1	2	3[c]	4					
	S	*A*	*A*	*A*	*G*	0.90	0.89	< 0.001	< 0.001	0.044
	s	*G*	*C*	*G*	*A*	0.08	−0.89	< 0.001	< 0.001	

Chr: chromosome; Hap. ID: haplotype identification letters to distinguish between different haplotypes; Freq: haplotype frequency; Coeff: coefficient, estimated effect of the haplotype on the pace score from a GLM regression in the haplotype analysis; Sim. *p*-value: *p*-value adjusted by using 100,000 permutations; σ_p^2 ratio: ratio of phenotypic variance explained by the haplotypes

[a] SNP numbers are in bp positional order

[b] The top SNP on ECA4 is number 2 in the positional order, where the reference allele is *A* and alternate allele is *C*

[c] The top SNP on ECA9 is number 3 in the positional order, where the reference allele is *G* and alternate allele is *A*

Table 2 Results from the ANOVA that compared phenotypes in horses with different QTL genotypes in the *RELN* gene

	ECA4-RR			ECA4-rr			*t*-value	*p*-value	df
	N	LSmeans	se	N	LSmeans	se			
Gait traits[a]									
Tölt	99	8.30	±0.07	65	8.62	±0.10	3.39	0.001	356
Slow tölt	97	8.06	±0.07	64	8.39	±0.10	3.25	0.001	351
Trot	99	7.95	±0.07	65	8.34	±0.10	3.83	< 0.001	356
Pace	99	7.45	±0.16	65	6.46	±0.22	−4.49	< 0.001	356
Gallop	99	7.99	±0.06	65	8.27	±0.08	3.33	0.001	356
Canter	97	7.66	±0.06	64	8.08	±0.09	4.62	< 0.001	350
Walk	99	7.79	±0.07	65	7.87	±0.08	0.75	0.455	356
Morphological traits[b]									
Height of withers (M1–M2)	90	10.21	±0.19	63	10.82	±0.26	2.32	0.021	334
Height at front (M1–M3)	98	4.27	±0.21	64	4.90	±0.30	2.06	0.040	353

N number of horses, *LSmeans* least squares means, *se* standard error, *df* degrees of freedom

[a] Subjectively assessed traits (judging scale 5–10)

[b] Morphological measurements (cm)

ANOVA revealed that the group of individuals that were homozygous for the R haplotype (RR genotype) differed significantly in mean scores (p-value ≤ 0.05) from those that were homozygous for the r haplotype (rr genotype) for all the gait phenotypes assessed at breeding field tests, except for walk (Table 2). These two groups differed significantly for the morphological measurements of height of the withers (from the lowest point on the back to the highest point on the withers) and of height at the front compared to the hind part (height at the croup to height at the withers). While horses with the RR genotype had a higher mean score for pace, horses with the rr genotype had higher mean scores for all the other gaits, height of the withers, and height at the front compared to the hind part (more uphill conformation).

Due to the high frequency of homozygotes for the S haplotype (SS genotype) at the *STAU2* gene and the low frequency of homozygotes for the s haplotype (ss genotype), the ANOVA yielded no significant results when comparing the means for these two groups. However, a comparison of horses with the SS genotype with the rest of the sample, including horses with the ss or Ss genotype, and horses carrying other rare haplotypes (2% of the sample) revealed significant differences (p-value ≤ 0.05) for pace, trot, and gallop (Table 3). The horses with the SS genotype had a higher mean score for pace but lower mean scores for all the other gaits, compared to all other horses. Morphological measurements were not significantly different (p-value > 0.05) between these two groups of horses.

The ANOVA revealed that the pace score for homozygotes for the *A* allele (*AA*) at the top SNP on ECA20 (N = 293, LSmeans = 7.15, se = ± 0.10) differed significantly (p-value < 0.001) from that of horses with the *AC*

Table 3 Results from the ANOVA that compared phenotypes in horses with different QTL genotypes in the *STAU2* gene

Gait traits[a]	ECA9-SS			ECA9-Ss + ss			*t*-value	*p*-value	df
	N	LSmeans	se	N	LSmeans	se			
Tölt	298	8.35	± 0.05	64	8.49	± 0.08	1.70	0.091	357
Slow tölt	295	8.14	± 0.05	62	8.24	± 0.09	1.11	0.268	352
Trot	298	8.05	± 0.05	64	8.28	± 0.09	2.62	0.009	357
Pace	298	7.15	± 0.10	64	6.13	± 0.19	-5.39	< 0.001	357
Gallop	298	8.08	± 0.04	64	8.29	± 0.07	3.01	0.003	357
Canter	294	7.82	± 0.04	62	7.97	± 0.08	1.93	0.054	351
Walk	298	7.81	± 0.05	64	7.84	± 0.09	0.29	0.770	357

N number of horses, *LSmeans* least squares means, *se* standard error, *df* degrees of freedom

[a] Subjectively assessed traits (judging scale 5–10)

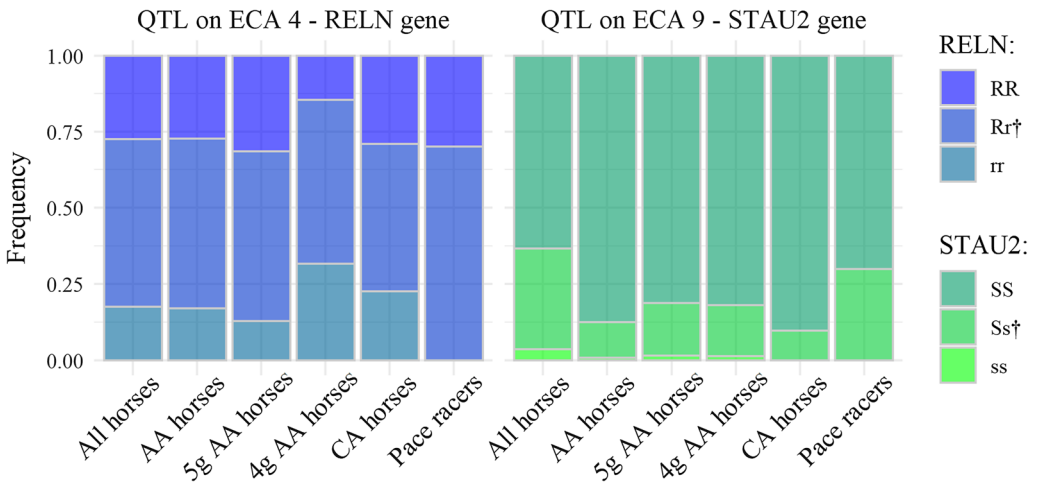


Fig. 5 Frequencies of the genotypes at the *RELN* and *STAU2* genes in groups of horses with variable pacing ability. The number of horses within each group is All horses N = 372, *AA* horses N = 340, 5-gaited *AA* horses N = 248, 4-gaited *AA* horses N = 82, *CA* horses N = 31, and Pace racers N = 10. †The frequency of the heterozygous genotypes (Rr and Ss) also included the rare haplotypes according to the haplotype analyses (1% of the sample size on ECA4 and 2% on ECA9)

or *CC* genotype (N = 69, LSmeans = 6.26, se = ± 0.19). Mean scores for the other gaits assessed at breeding field tests were not significantly different between these two groups.

We found that the genotype frequencies for the QTL detected in the *RELN* and *STAU2* genes differed between the horses that had different levels of pacing ability (Fig. 5). The group of horses with a genetic precondition to pace that had shown pace at breeding field tests (5-gaited *DMRT3 AA* horses) had higher frequencies of the RR and SS genotype at the *RELN* and *STAU2* genes, respectively, and a lower frequency of the rr genotype compared to all horses with the *AA* genotype at the *DMRT3* gene. On the contrary, the group of horses with a genetic precondition to pace that had not shown pace at breeding field tests (4-gaited *DMRT3 AA* horses) had the highest frequency of both the rr and ss genotypes of all groups of horses and the lowest frequency of the RR and SS genotypes.

The group of horses without the favorable genetic variant for pace (*DMRT3 CA* horses) had the highest frequency of the SS genotype among all groups, but a lower frequency of the heterozygous genotypes (Rr and Ss) and lacked the ss genotype altogether. The group of horses that had raced in pace competitions but that had not attended a breeding field test (pace racers) had a higher frequency of the heterozygous genotypes (Rr and Ss) compared to the other groups and lacked the rr and ss genotypes altogether.

For both QTL, genotype frequencies in the group of four-gaited *DMRT3 AA* horses differed significantly from those in the group of five-gaited *DMRT3 AA* horses ($p < 0.001$), the group of all *DMRT3 AA* horses ($p = 0.015$), and the group with all horses in the dataset ($p = 0.015$

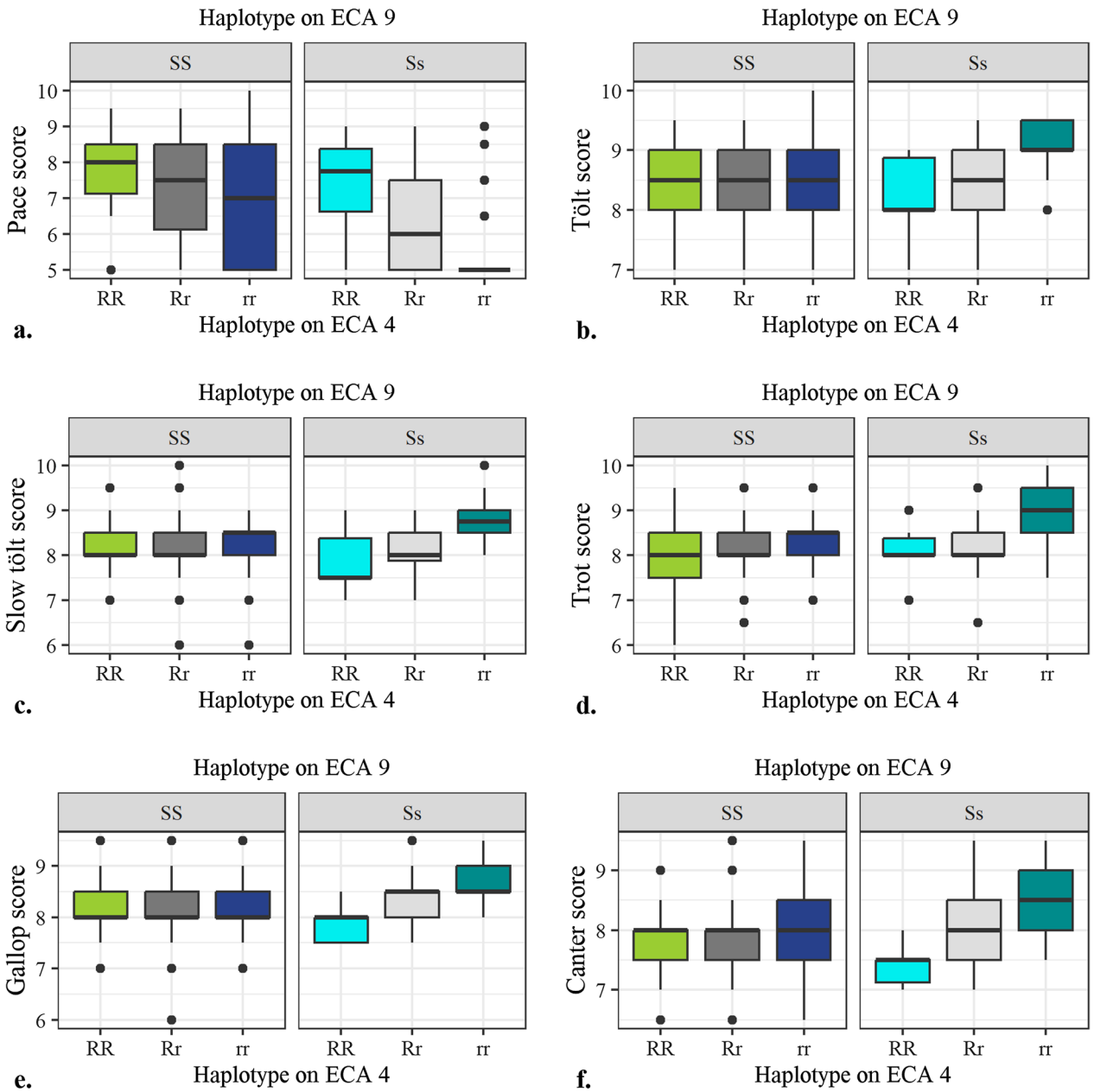


Fig. 6 Interactions between the significant haplotypes on ECA4 and ECA9. Interactions between the significant haplotypes on ECA4 and ECA9 in a group of horses with the *AA* genotype at the *DMRT3*-variant on the subjective scores for **a** pace, **b** tölt, **c** slow tölt, **d** trot, **e** gallop and **f** canter. Number of horses with each combination is SS:RR = 78, SS:Rr = 150, SS:rr = 41, Ss:RR = 10, Ss:Rr = 29, Ss:rr = 17. The groups referred to as heterozygous individuals (Rr and Ss) also included horses possessing the rare haplotypes detected in the haplotype analyses (1% of the sample size with the rare haplotypes on ECA4 and 2% of the sample size with the rare haplotypes on ECA9)

(*RELN*), $p = 0.008$ (*STAU2*)). Pairwise comparison of genotype frequencies between other groups were non-significant ($p > 0.05$). The number of horses with different genotypes in each group is presented in Additional file 9: Table S4.

Interactions between the significant haplotypes in the *RELN* gene on ECA4 and the *STAU2* gene on ECA9 were investigated within the group of horses with the *AA* genotype at the *DMRT3*-variant (Fig. 6). Horses with the SS and RR genotypes (SS:RR) had significantly higher mean scores for pace than horses with the Ss:Rr and Ss:rr genotype combinations. Horses with the Ss:rr genotype combination also had significantly lower mean scores

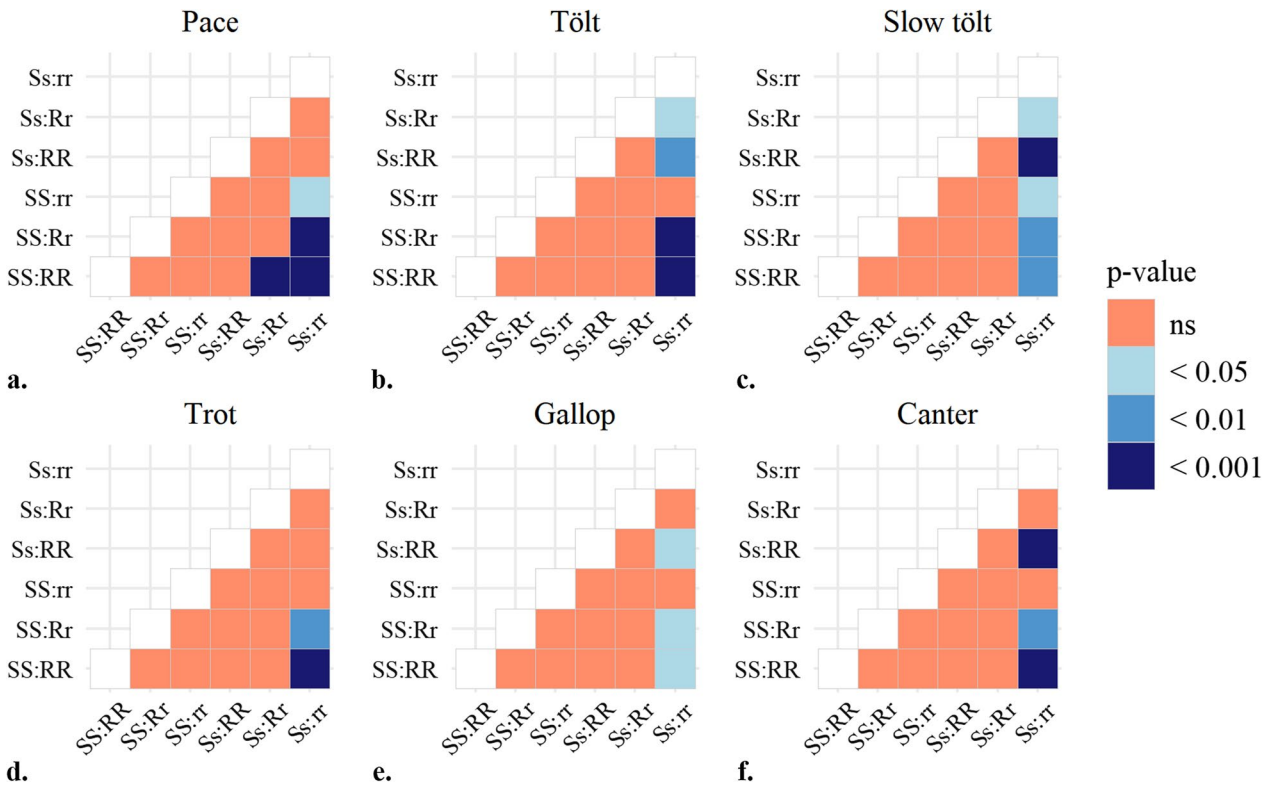


Fig. 7 Significance levels of the differences in mean scores between different haplotype interactions for different gaits: **a** pace, **b** tölt, **c** slow tölt, **d** trot, **e** gallop, and **f** canter. Number of horses with each combination is SS:RR = 78, SS:Rr = 150, SS:rr = 41, Ss:RR = 10, Ss:Rr = 29, Ss:rr = 17

for pace than those with the SS:Rr and SS:rr genotype combinations.

For the other gaits investigated (except walk), horses with the Ss:rr genotype combination had higher mean assessment scores than horses with the SS:RR and SS:Rr genotype combinations. Furthermore, horses with the Ss:rr genotype combination had higher mean scores than horses with the Ss:RR genotype combination for tölt, slow tölt, gallop, and canter. Horses with the Ss:rr genotype combination also had higher mean scores than horses with the Ss:Rr genotype combination for tölt and slow tölt, and than horses with the SS:rr genotype combination for slow tölt.

Results for horses with the ss:RR and ss:Rr genotype combinations (only 2 and 3 individuals, respectively) are not presented in Fig. 6, since the mean scores for these few horses did not differ significantly from the mean scores for horses with different genotype combinations. The significance levels of the differences in mean scores for other genotype interactions are presented in Fig. 7.

Analysis of the significant haplotypes in the *RELN* gene on ECA4 and the *STAU2* gene on ECA9 was also performed for all gaits within the group of 31 horses with the *CA* genotype at the *DMRT3*-variant, but these results were not significant (results not shown). In addition, we did not find clear indications of interactions between the haplotypes in the *RELN* and *STAU2* genes and the QTL previously detected on ECA22 for back and croup (results not shown).

Discussion

The present study was performed as a first step towards a better understanding of the genetic architecture of pace beyond the contribution of the *DMRT3* gene by identifying other genomic regions that are relevant for this gait. A single associated SNP on ECA20 was located within a long non-coding RNA gene, in the vicinity of several other genes and its potential influence on pace is still unclear. Two novel QTL were detected on ECA4 and 9, which were located within the *RELN* and *STAU2* genes, respectively. Our results suggest that these QTL may be important not only for pace but also for other gaits that are assessed at breeding field tests and, to some extent, they seem to differentiate horses with different levels of pacing ability and quality. Furthermore, there seem to be some interactions between these QTL and the *DMRT3* gene, and genetic compensation may potentially play a role for this trait.

Potential causative genes

While the two identified QTL on ECA4 and 9 are located in different genes, to date, we cannot rule out an effect on regulatory elements that govern the expression of nearby genes. However, as discussed below, both the *RELN* and *STAU2* genes appear to be likely candidates, whereas the genes in the vicinity of the identified regions (see Additional file 7: Table S3) are either all involved in general cellular processes or not expressed in neural tissue, making it unlikely that they affect development or the function of the locomotor network.

The QTL detected on ECA4 is located within the *RELN* gene, which encodes an extracellular matrix protein that mediates cell to cell interactions that are critical for migration and positioning of neurons during development [39–42]. The *RELN* gene is expressed in the developing spinal cord of zebrafish and mice but its expression has also been described in the mature spinal cord of primates, which indicates a vital and conserved function during spinal cord development [43]. The *RELN* gene has been widely explored and several mutations have been reported in mice and rats that are associated with the "reeler phenotype" [44–48], which results in abnormal locomotor behaviors such as tremors, dystonia, and ataxia [48]. Abnormal migration of motor neurons has been observed in mutant mice, which likely affects their synaptic input and might be responsible for the reeler phenotype [41, 49]. Given that neurons expressing the *DMRT3* gene are known to synapse onto motor neurons [7, 50], the normal inhibition provided by these neurons could be disturbed due to aberrant innervation, thus explaining the cumulative effect of the 'gait keeper' mutation and the RR genotype.

The QTL on ECA9 is located within the *STAU2* gene, which encodes a protein involved in mRNA transport [51]. The *STAU2* gene is expressed in the developing zebrafish and in the mouse spinal cord, and contributes to the asymmetric distribution of mRNA in dividing intermediate progenitor cells [52], which is a process known to affect the proliferation and fate assignment of neurons. Loss of *STAU2* affects the survival of neurons and mutant mice display reduced motor coordination but enhanced motor learning abilities [53]. This suggests that, while negatively affecting early motor output, it may facilitate the learning of new gaits, as in the case of five-gaited horses. Thus, modification of the expression of *STAU2* may improve the ability of Icelandic horses to learn to pace, which may explain the higher quality pace in horses that are homozygous for the 'gait keeper' mutation and that also carry the SS genotype.

Effects of haplotypes in the *RELN* gene on gaits and topline conformation

Breeding horses are preselected to attend breeding field tests based on their assumed potential to get high scores, where the emphasis on tölt and pace is strong, in parallel to their importance in the breeding goal [10, 11]. Therefore, the somewhat balanced frequency of the opposite haplotypes at the *RELN* gene (0.54 for R and 0.45 for r) is not unexpected, considering their opposite effects on pace and tölt. Furthermore, horses that are preselected to be shown at breeding field tests despite a poor or absent ability to pace tend to have higher average quality for the other gaits to compensate for the missing score for pace. This could partly explain the higher average scores for all gaits except pace of horses with the rr genotype, since the four-gaited horses with the *CA* or *AA* genotype at the *DMRT3*-variant were more likely to have the rr genotype than the other horses in the study. However, favorable effects of the r haplotype on tölt, trot, gallop, and canter cannot be completely excluded. The relatively high frequency of the R haplotype in horses with the *AA* genotype at the *DMRT3*-variant that do perform pace, either at breeding field tests or in pace competitions, further supports the cumulative effect of the R haplotype and the 'gait keeper' mutation for pacing ability and quality.

The QTL in the *RELN* gene seems to affect not only the gait traits but also morphological traits. Horses with the rr genotype had on average higher withers and were higher at the front compared to horses with the RR genotype. The advantage of high withers and an uphill conformation for riding ability in Icelandic horses has been previously reported [14]. However, the relationship between score for pace and height of withers, as well as height at the front, was curvilinear, which indicates that there may be an optimum height of withers and height at the front for the pace score. This may contribute to the opposite effects of the identified QTL in the *RELN* gene on pace score and scores for the other gaits assessed at breeding field tests. On the other hand, the preselection of individuals for a higher quality of tölt, trot, gallop, and canter among the horses shown without pace at breeding field tests, may also promote bias in the data in that horses performing as four-gaiters have higher withers and a more uphill conformation as they have been shown to be favorable attributes for those gaits.

Haplotypes in the *STAU2* gene may influence locomotive learning

The high frequency of the S haplotype at the *STAU2* gene indicates a strong selection for this genomic region. This is supported by the large favorable effect of the S haplotype on pace, which has been strongly emphasized in the breeding goal since 1950 [54]. The other lateral gait,

tölt, has also been emphasized in the breeding goal and has been reported to have a favorable, but not especially high, genetic correlation (0.38) with pace [12]. However, the results from our present study suggest rather an unfavorable effect of this haplotype on all gaits except pace, particularly on trot and gallop. These effects of the S haplotype concur to some extent with the observed effects of the *A* allele at the *DMRT3* gene [7, 8]. This may indicate a similar reinforcement by the S haplotype to that of the *A* allele in *DMRT3* for the coordination of the ipsilateral legs and subsequent negative effect on the synchronized movement of diagonal legs, or possibly cumulative effects of the two genetic factors.

The high frequency of the SS genotype (90%) in the group of horses with the *CA* genotype at the *DMRT3*-variant suggests that the S haplotype may be important for *CA* horses to perform well at breeding field tests. The S haplotype may improve motor learning abilities, similar to the effects of a variant in the *STAU2* gene in mice [53]. In our dataset, based only on their *DMRT3* genotype, few of the horses that carried the SS genotype would have been expected not to receive the scores for tölt or pace that they got at breeding field tests. One horse with the *CC* genotype at the *DMRT3*-variant received a score of 7.5 for tölt and two horses with the *CA* genotype received scores of 6.0 and 6.5 for pace (results not shown). The number of such horses is too small to draw firm conclusions, but further studies on whether the S haplotype could compensate for the lack of the *A*-allele (*DMRT3*) would be interesting.

Genetic difference between five-gaited horses and pace racers

Both the R haplotype at the *RELN* gene and the S haplotype at the *STAU2* gene showed large favorable effects on pace performance at breeding field tests. Therefore, it is interesting to compare the group of five-gaited horses with the more specialized pace racers. Good pace racers need to have the ability to pace in balance at high speed, but in contrast to the five-gaited horses shown at breeding field tests, the quality of the other gaits is less important for racing performance. Therefore, not all pace racers are good five-gaited horses and vice versa, which may be traced back to different genetic makeups of the two groups, in addition to different training methods. The most prominent genetic difference between these two groups that was observed in the current study was the absence of the rr genotype among the pace racers and the relatively high frequency of this genotype among the five-gaiters. This may further indicate the importance of the R haplotype for pacing ability, but the r haplotype may also enhance the quality of the other gaits, which is more important for the five-gaited horse. However, there were only ten pace racers included in this study, making the comparison somewhat uncertain. Thus, further research with a larger number of pace racers is needed to validate the genetic difference between pace racers and five-gaited horses used for other purposes.

Interactions between haplotypes in the *RELN* and *STAU2* genes

The RR:SS genotype combination was the most favorable for pace score, which indicates that this genotype combination has additive effects. In contrast, the rr:Ss genotype combination was the most favorable for scores for tölt, slow tölt, trot, gallop, and canter. This is consistent with the individual effects of the haplotypes, assuming that the r haplotype enhances the quality of the gaits other than pace and the S haplotype enhances motor learning.

Data quality

In general, genetic studies rely heavily on the quality of the phenotypes used in the analysis. Objective measurements are usually less affected by environmental factors and therefore tend to yield higher heritability estimates. However, for routine large-scale recordings in field tests of horses, gait traits are, for practical and economic reasons, still more often subjectively than objectively assessed. The moderate to high heritability estimates for gait scores in Icelandic horses obtained in several studies [10, 12], in particular for pace, as well as results from previous genomic studies [7, 15] show that the standardized assessments by trained judges do provide useful information. In this study, we used the effects accounted for in the current statistical model for the routine genetic evaluation of Icelandic horses as a starting point, as well as some additional environmental factors such as the *DMRT3* genotype and the country where the horse was assessed. In our dataset of genotyped horses, only the effects of sex and *DMRT3* genotype were significant.

The quality of our data is furthermore subject to the relationships between the individuals in our sample, since the Icelandic horse population is a relatively small and closed population [55]. However, according to the MDS plot (Fig. 3), which was based on the genomic-kinship matrix from our dataset, no stratification or outliers were detected. Furthermore, relationships between individuals were accounted for by including the genomic-kinship matrix in the GWA model. Therefore, we do not believe that the results were affected by bias due to relationships between individuals.

Conclusions

This study provides valuable information about the genetic background of the gait pace in Icelandic horses. Two new candidate genes, *RELN* and *STAU2*, were identified through GWAS. Both genes are known to be associated with locomotor behavior in mice, which may indicate analogous functions in horses. Furthermore, the results from this study indicate genetic interactions between these novel genes and the previously identified candidate gene *DMRT3*. The novel genes and the identified top SNP on ECA20, along with the effects of *DMRT3* genotype were estimated to account for 27.4% of the phenotypic variance of the gait. Opposite haplotypes were identified for both these genes that appear to influence quality and ability to pace. Furthermore, these haplotypes appear to have opposing effects on the other gaits, especially trot, gallop, and canter. This suggests that these genetic factors may partly contribute to whether a horse is trained and shown as a four- or five-gaited horse at a breeding field test. These findings may aid in the selection of breeding and competition horses and are thus of major interest to horse breeders. Further investigation is needed to fully understand the underlying genetic factors and the nature of their interactions that contribute to the complexity of the pace trait's phenotype.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12711-023-00863-6.

Additional file 1: Figure S1. Distributions of scores for pace after rank- and log-transformation. (a) Rank- (W = 0.98) and (b) log-transformed (W = 0.90) distribution of scores for pace in the sample of 362 horses assessed at a breeding field test. Neither of these transformations resulted in a normal distribution of the scores ($p > 0.05$)

Additional file 2: Table S1. Descriptive statistics of the gait scores (other than pace). Number of assessments, mean score, standard deviation (sd), range, skewness, kurtosis, and a *p*-value from a Jarque–Bera Normality Test for the gaits (other than pace) assessed at breeding field tests for horses included in the dataset.

Additional file 3: Figure S2. Distribution of the gait scores (other than pace). Distribution of scores for (a) tölt, (b) slow tölt, (c) trot, (d) gallop, (e) canter, and (f) walk in the sample of 362 horses assessed at a breeding field test.

Additional file 4: Figure S3. Distributions of the scores for tölt after rank- and log-transformation. (a) Rank- (W = 0.98) and (b) log-transformed (W = 0.95) distribution of scores for tölt in the sample of 362 horses assessed at a breeding field test. Neither of these transformations resulted in a normal distribution of the scores ($p > 0.05$).

Additional file 5: Table S2. Top 50 SNPs from the GWAS. A summary of the GWAS results for the 50 top SNPs.

Additional file 6: Figure S4. Relative location of the identified SNPs in the *RELN* horse gene.

Additional file 7: Table S3. List of genes located in the vicinity of the regions identified on ECA4, 9, and 20. A list of all the genes located in the vicinity (± 500.00 kb) of the regions identified on ECA4: 4,222,615–4,228,914 bp [39–42, 56–64], ECA9: 11,533,922–13,457,268 bp [31, 32, 51, 52, 65–72] and ECA20: 52,057,378–52,057,378 bp.

Additional file 8: Figure S5. Relative location of the identified SNPs in the *STAU2* horse gene. Relative location of the identified SNPs in the *STAU2* horse gene. Two of the identified SNPs, that were located ~ 1600 kb away from the gene, were not included in this figure.

Additional file 9: Table S4. Number of horses within each group of horses with variable pacing ability. The number of horses with different haplotypes on ECA4 and ECA9 in groups of horses with variable pacing ability.

Acknowledgements

The authors would like to acknowledge all the horse owners for their cooperation and contribution. The authors would also like to acknowledge Tytti Vanhala at the The Swedish University of Agricultural Sciences (SLU) for extracting DNA and performing genotyping and Diana Ekman in the Swedish Bioinformatics Advisory Program at the Science for Life Laboratory (SciLifeLab) for advice on bioinformatics.

Author contributions

HS, EA, TK, GL, and SE initiated and designed the study. HS collected half of the samples. EA provided the phenotypic data from WorldFengur. HS performed the experiments and data analysis and drafted the manuscript. HB advised and contributed to drafting interpretations of gene functions. SE advised with the statistical analysis. HS, HB, EA, TK, MR, GL, and SE contributed to the interpretation of the results. All authors suggested modifications to the manuscript. All authors read and approved the final manuscript.

Funding

Open access funding provided by Swedish University of Agricultural Sciences. The Agricultural University of Iceland's Doctoral Fund, The Stock Protection Fund for the Icelandic Horse Breed, The Swedish Icelandic Horse Association (SIF), Blikastaðir Fund, The Icelandic Genetic Resource Council in Agriculture (AGROGEN), The Swedish Research Council (VR) and the Swedish Foundation for Strategic Research. Open Access funding is provided by the Swedish University of Agricultural Sciences.

Availability of data and materials

The datasets generated and analyzed during the current study are not publicly available since the study was performed in collaboration with the Icelandic horse breeding industry and has commercial value for them. However, data are available from the corresponding author on reasonable request and with the permission of the Icelandic Horse Association.

Declarations

Ethics approval and consent to participate

Hair samples were collected according to the ethical approval by the Ethics Committee for Animal Experiments in Uppsala, Sweden (number: 5.8.18-15453/2017) and the animal experiment license by the Icelandic Food and Veterinary Authority in Iceland (number: 2020-04-02/2003120).

Consent for publication

Not applicable.

Competing interests

The authors declare competing interests concerning the commercial applications of the current study. GL is a co-inventor of a patent application concerning commercial testing of the *DMRT3* mutation. The stated patent does not restrict research applications of the method.

Author details

[1]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, P.O. Box 7023, 75007 Uppsala, Sweden. [2]Faculty of Agricultural Sciences, Agricultural University of Iceland, Borgarbyggð, 311 Hvanneyri, Iceland. [3]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden. [4]The Icelandic Agricultural Advisory Centre, Hagatorgi 1, 107 Reykjavik, Iceland. [5]Department of Anatomy, Physiology, and Biochemistry, Swedish University of Agricultural Sciences, P.O. Box 7011, 75007 Uppsala, Sweden. [6]Department of Biosystems, Center for Animal

Breeding and Genetics, KU Leuven, Kasteelpark Arenberg 30, 3001 Leuven, Belgium.

Received: 17 May 2023 Accepted: 1 December 2023
Published online: 11 December 2023

References

1. Mackay TFC. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. Nat Rev Genet. 2014;15:22–33.
2. Carlborg Ö, Haley CS. Epistasis: too often neglected in complex trait studies? Nat Rev Genet. 2004;5:618–25.
3. Sackton TB, Hartl DL. Genotypic context and epistasis in individuals and populations. Cell. 2016;166:279–87.
4. Forsberg SKG, Bloom JS, Sadhu MJ, Kruglyak L, Carlborg Ö. Accounting for genetic interactions improves modeling of individual quantitative trait phenotypes in yeast. Nat Genet. 2017;49:497–503.
5. Boucher B, Jenna S. Genetic interaction networks: better understand to better predict. Front Genet. 2013;4:290.
6. Fang G, Wang W, Paunic V, Heydari H, Costanzo M, Liu X, et al. Discovering genetic interactions bridging pathways in genome-wide association studies. Nat Commun. 2019;10:4274.
7. Andersson LS, Larhammar M, Memic F, Wootz H, Schwochow D, Rubin C-J, et al. Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. Nature. 2012;488:642–6.
8. Kristjansson T, Bjornsdottir S, Sigurdsson A, Andersson L, Lindgren G, Helyar S, et al. The effect of the 'Gait keeper' mutation in the DMRT3 gene on gaiting ability in Icelandic horses. J Anim Breed Genet. 2014;131:415–25.
9. Jäderkvist K, Holm N, Imsland F, Árnason T, Andersson L, Andersson LS, et al. The importance of the DMRT3 'Gait keeper' mutation on riding traits and gaits in Standardbred and Icelandic horses. Livest Sci. 2015;176:33–9.
10. Albertsdóttir E, Eriksson S, Sigurdsson Á, Árnason T. Genetic analysis of 'breeding field test status' in Icelandic horses. J Anim Breed Genet. 2011;128:124–32.
11. General rules and regulations: breeding rules and regulations. 2023. https://www.feiffengur.com/documents/FEIF%20Breeding_2023.pdf/. Accessed 12 May 2023.
12. Albertsdóttir E, Eriksson S, Näsholm A, Strandberg E, Árnason T. Genetic correlations between competition traits and traits scored at breeding field-tests in Icelandic horses. Livest Sci. 2008;114:181–7.
13. Sigurðardóttir H. Athugun á úrvalsstyrkleika og erfðaframförum í einstökum eiginleikum í íslenskri hrossarækt (investigation of selection intensity and genetic gain within breeding field test traits in Icelandic horse breeding). BSc. thesis, Agricultural University of Iceland. 2012.
14. Kristjansson T, Bjornsdottir S, Albertsdóttir E, Sigurdsson A, Pourcelot P, Crevier-Denoix N, et al. Association of conformation and riding ability in Icelandic horses. Livest Sci. 2016;189:91–101.
15. Rosengren MK, Sigurðardóttir H, Eriksson S, Naboulsi R, Jouni A, Novoa-Bravo M, et al. A QTL for conformation of back and croup influences lateral gait quality in Icelandic horses. BMC Genomics. 2021;22:267.
16. McCoy AM, Beeson SK, Rubin C-J, Andersson L, Caputo P, Lykkjen S, et al. Identification and validation of genetic variants predictive of gait in standardbred horses. PLoS Genet. 2019;15: e1008146.
17. Amano T, Onogi A, Yamada F, Kawai M, Shirai K, Ueda J. Genome-wide association mapping and examination of possible maternal effect for the pace trait of horses. Anim Genet. 2018;49:461–3.
18. Staiger EA, Abri MA, Silva CAS, Brooks SA. Loci impacting polymorphic gait in the Tennessee walking horse. J Anim Sci. 2016;94:1377–86.
19. Jäderkvist Fegraeus K, Hirschberg I, Árnason T, Andersson L, Velie B, Andersson L, et al. To pace or not to pace: a pilot study of four- and five-gaited Icelandic horses homozygous for the DMRT3 'Gait Keeper'mutation. Anim Genet. 2017;48:694–7.
20. Worldfengur—the studbook of origin for the Icelandic horse. Farmers Association of Iceland and FEIF International Federation of Icelandic Horse, Reykjavik. 2001. https://www.worldfengur.com/. Accessed 19 July 2022.
21. Kristjansson T, Bardarson S, Aevarsson S, Reynisson G. The gaits of the Icelandic horse. 2014. https://horsesoficeland.cdn.prismic.io/horsesoficeland/3f1b003e-5e7d-404f-875c-a286be86925b_gangteg_ens_p.pdf/. Accessed 12 May 2023.
22. General information on breeding field tests—measurements (body measures). 2018. https://www.rml.is/static/files/Hrossaraekt_RML/skrokkmal.pdf/. Accessed 12 May 2023.
23. Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007;23:1294–6.
24. GenABEL project developers. GenABEL: genome-wide SNP association analysis. 2013. https://CRAN.R-project.org/package=GenABEL/. Accessed 12 May 2023.
25. R Core team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019.
26. Rönnegård L, Shen X, Alam M. Hglm: a package for fitting hierarchical generalized linear models. R J. 2010;2:20–8.
27. Duggal P, Gillanders EM, Holmes TN, Bailey-Wilson JE. Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. BMC Genomics. 2008;9:519.
28. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995;11:241–7.
29. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al. Ensembl 2022. Nucleic Acids Res. 2022;50:D988–95.
30. Hu Z-L, Park CA, Reecy JM. Bringing the animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Res. 2022;50:D956–61.
31. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinform. 2016;54:1.30.1–33.
32. Safran M, Rosen N, Twik M, BarShir R, Stein TI, Dahary D, et al. The GeneCards suite. In: Abugessaisa I, Kasukawa T, et al., editors. Practical guide to Life science databases. Singapore: Springer Nature Singapore; 2021. p. 27–56.
33. Kierczak M, Jablonska J, Forsberg SKG, Bianchi M, Tengvall K, Pettersson M, et al. Cgmisc: enhanced genome-wide association analyses and visualization. Bioinformatics. 2015;31:3830–1.
34. Purcell S, Chang C. PLINK v1.90. 2021. https://www.cog-genomics.org/plink/1.9/. Accessed 12 May 2023.
35. Gaunt TR, Rodríguez S, Day INM. Cubic exact solutions for the estimation of pairwise haplotype frequencies: implications for linkage disequilibrium analyses and a web tool 'CubeX.' BMC Bioinformatics. 2007;8:428.
36. Sinnwell JP, Schaid DJ. Haplo.stats: statistical analysis of haplotypes with traits and covariates when linkage phase is ambiguous. 2022. https://CRAN.R-project.org/package=haplo.stats/. Accessed 12 May 2023.
37. Kassambara A. rstatix: pipe-friendly framework for basic statistical tests. 2023. https://CRAN.R-project.org/web/packages/rstatix/index.html/. Accessed 20 July 2023.
38. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2016.
39. Hattori M, Kohno T. Regulation of Reelin functions by specific proteolytic processing in the brain. J Biochem. 2021;169:511–6.
40. Lakatosova S, Ostatnikova D. Reelin and its complex involvement in brain development and function. Int J Biochem Cell Biol. 2012;44:1501–4.
41. Vaswani AR, Blaess S. Reelin signaling in the migration of ventral brain stem and spinal cord neurons. Front Cell Neurosci. 2016;10:62.
42. Nimura T, Itoh T, Hagio H, Hayashi T, Di Donato V, Takeuchi M, et al. Role of Reelin in cell positioning in the cerebellum and the cerebellum-like structure in zebrafish. Dev Biol. 2019;455:393–408.
43. Krzyzanowska A, Cabrerizo M, Clascá F, Ramos-Moreno T. Reelin immunoreactivity in the adult spinal cord: a comparative study in rodents, carnivores, and non-human primates. Front Neuroanat. 2020;13: 102.
44. Andersen TE, Finsen B, Goffinet AM, Issinger O-G, Boldyreff B. A reeler mutant mouse with a new, spontaneous mutation in the reelin gene. Brain Res Mol Brain Res. 2002;105:153–6.
45. Kikkawa S, Yamamoto T, Misaki K, Ikeda Y, Okado H, Ogawa M, et al. Missplicing resulting from a short deletion in the reelin gene causes reeler-like neuronal disorders in the mutant shaking rat Kawasaki. J Comp Neurol. 2003;463:303–15.
46. Miao GG, Smeyne RJ, D'Arcangelo G, Copeland NG, Jenkins NA, Morgan JI, et al. Isolation of an allele of reeler by insertional mutagenesis. Proc Natl Acad Sci USA. 1994;91:11050–4.
47. Caviness VS Jr, So DK, Sidman RL. The hybrid reeler mouse. J Hered. 1972;63:241–6.
48. Falconer DS. Two new mutants, 'trembler' and 'reeler', with neurological actions in the house mouse (*Mus musculus* L.). J Genet. 1951;50:192–205.

49. Phelps PE, Rich R, Dupuy-Davies S, Ríos Y, Wong T. Evidence for a cell-specific action of reelin in the spinal cord. Dev Biol. 2002;244:180–98.
50. Menelaou E, Kishore S, McLean DL. Mixed synapses reconcile violations of the size principle in zebrafish spinal cord. Elife. 2022;11: e64063.
51. Cockburn DM, Charish J, Tassew NG, Eubanks J, Bremner R, Macchi P, et al. The double-stranded RNA-binding protein staufen 2 regulates eye size. Mol Cell Neurosci. 2012;51:101–11.
52. Chowdhury R, Wang Y, Campbell M, Goderie SK, Doyle F, Tenenbaum SA, et al. STAU2 binds a complex RNA cargo that changes temporally with production of diverse intermediate progenitor cells during mouse corticogenesis. Development. 2021;148:dev199376.
53. Pernice HF, Schieweck R, Jafari M, Straub T, Bilban M, Kiebler MA, et al. Altered glutamate receptor ionotropic delta subunit 2 expression in Stau2-deficient cerebellar purkinje cells in the adult brain. Int J Mol Sci. 2019;20:1797.
54. Arnórsson K. Ræktunin [The breeding]. In: Björnsson GB, Sveinsson HJ, editors. Íslenski hesturinn [The Icelandic horse]. 1st ed. Reykjavík: Mál og menning; 2006. p. 202–47.
55. Hreiðarsdóttir GE, Árnason Þ, Svansson V, Hallsson JH. Analysis of the history and population structure of the Icelandic horse using pedigree data and DNA analyses. Icel Agric Sci. 2014;27:63–79.
56. Jin J, Cardozo T, Lovering RC, Elledge SJ, Pagano M, Harper JW. Systematic analysis and nomenclature of mammalian F-box proteins. Gene Dev. 2004;18:2573–80.
57. Kim D, LaQuaglia MP, Yang SY. A cDNA encoding a putative 37 kDa leucine-rich repeat (LRR) protein, p37NB, isolated from S-type neuroblastoma cell has a differential tissue distribution. Biochim Biophys Acta. 1996;1309:183–8.
58. Zhou X, Yang G, Huang R, Chen X, Hu G. SVH-B interacts directly with p53 and suppresses the transcriptional activity of p53. FEBS Lett. 2007;581:4943–8.
59. Tsuboi K, Okamoto Y, Ikematsu N, Inoue M, Shimizu Y, Uyama T, et al. Enzymatic formation of *N*-acylethanolamines from *N*-acylethanolamine plasmalogen through *N*-acylphosphatidylethanolamine-hydrolyzing phospholipase D-dependent and -independent pathways. Biochim Biophys Acta. 2011;1811:565–77.
60. Hölzen L, Syré K, Mitschke J, Brummer T, Miething C, Reinheckel T. Degradome-focused RNA interference screens to identify proteases important for breast cancer cell growth. Front Oncol. 2022;12: 960109.
61. Helary L, Castille J, Passet B, Vaiman A, Beauvallet C, Jaffrezic F, et al. DNAJC2 is required for mouse early embryonic development. Biochem Biophys Res Commun. 2019;516:258–63.
62. Sahu I, Glickman MH. Proteasome in action: substrate degradation by the 26S proteasome. Biochem Soc Trans. 2021;49:629–44.
63. Wang Z, Wang Q, Wu H, Huang Z. Identification and characterization of amphibian SLC26A5 using RNA-Seq. BMC Genomics. 2021;22:564.
64. Dillin A, Rine J. Separable functions of ORC5 in replication initiation and silencing in *Saccharomyces cerevisiae*. Genetics. 1997;147:1053–62.
65. Dou T, Fu M, Wang Y, Zhao Y, Wang Z, Bian Z, et al. Signatures of positive selection in LY96 gene in vertebrates. J Biosci. 2013;38:899–904.
66. Sánchez-Caballero L, Elurbe DM, Baertling F, Guerrero-Castillo S, van den Brand M, van Strien J, et al. TMEM70 functions in the assembly of complexes I and V. Biochim Biophys Acta Bioenergy. 2020;1861: 148202.
67. Andreou A, Yngvadottir B, Bassaganyas L, Clark G, Martin E, Whitworth J, et al. Elongin C (ELOC/TCEB1)-associated Von Hippel–Lindau disease. Hum Mol Genet. 2022;31:2728–37.
68. Mroczek S, Dziembowski A. U6 RNA biogenesis and disease association. Wiley Interdiscip Rev RNA. 2013;4:581–92.
69. Lei B, Xie L, Zhang S, Lv D, Shu F, Deng Y. *UBE2W* down-regulation promotes cell apoptosis and correlates with hypospermatogenesis. Andrologia. 2020;52:e13474.
70. Sandell LL, Lynn ML, Inman KE, McDowell W, Trainor PA. *RDH10* oxidation of vitamin A is a critical control step in synthesis of retinoic acid during mouse embryogenesis. PLoS One. 2012;7:e30698.
71. De Falco S, Russo G, Angiolillo A, Pietropaolo C. Human L7a ribosomal protein: sequence, structural organization, and expression of a functional gene. Gene. 1993;126:227–35.
72. Li JSZ, Miralles Fusté J, Simavorian T, Bartocci C, Tsai J, Karlseder J, et al. TZAP: a telomere-associated protein involved in telomere length control. Science. 2017;355:638–41.

Publisher's Note

IV

BMC Genomics

RESEARCH

Open Access

Genetic diversity and signatures of selection in Icelandic horses and Exmoor ponies

Heiðrún Sigurðardóttir[1,2*], Michela Ablondi[3], Thorvaldur Kristjansson[2], Gabriella Lindgren[1,4] and Susanne Eriksson[1]

Abstract

Background The Icelandic horse and Exmoor pony are ancient, native breeds, adapted to harsh environmental conditions and they have both undergone severe historic bottlenecks. However, in modern days, the selection pressures on these breeds differ substantially. The aim of this study was to assess genetic diversity in both breeds through expected (H_E) and observed heterozygosity (H_O) and effective population size (Ne). Furthermore, we aimed to identify runs of homozygosity (ROH) to estimate and compare genomic inbreeding and signatures of selection in the breeds.

Results H_O was estimated at 0.34 and 0.33 in the Icelandic horse and Exmoor pony, respectively, aligning closely with H_E of 0.34 for both breeds. Based on genomic data, the Ne for the last generation was calculated to be 125 individuals for Icelandic horses and 42 for Exmoor ponies. Genomic inbreeding coefficient (F_{ROH}) ranged from 0.08 to 0.20 for the Icelandic horse and 0.12 to 0.27 for the Exmoor pony, with the majority of inbreeding attributed to short ROHs in both breeds. Several ROH islands associated with performance were identified in the Icelandic horse, featuring target genes such as *DMRT3*, *DOCK8*, *EDNRB*, *SLAIN1*, and *NEURL1*. Shared ROH islands between both breeds were linked to metabolic processes (*FOXO1*), body size, and the immune system (*CYRIB*), while private ROH islands in Exmoor ponies were associated with coat colours (*ASIP*, *TBX3*, *OCA2*), immune system (*LYG1*, *LYG2*), and fertility (*TEX14*, *SPO11*, *ADAM20*).

Conclusions Evaluations of genetic diversity and inbreeding reveal insights into the evolutionary trajectories of both breeds, highlighting the consequences of population bottlenecks. While the genetic diversity in the Icelandic horse is acceptable, a critically low genetic diversity was estimated for the Exmoor pony, which requires further validation. Identified signatures of selection highlight the differences in the use of the two breeds as well as their adaptive trait similarities. The results provide insight into genomic regions under selection pressure in a gaited performance horse breed and various adaptive traits in small-sized native horse breeds. This understanding contributes to preserving genetic diversity and population health in these equine populations.

Keywords Runs of homozygosity, Heterozygosity, Effective population size, Genomic inbreeding, Performance, Adaptation, Metabolism, Immune system, Coat colours

*Correspondence:
Heiðrún Sigurðardóttir
heidrun.sigurdardottir@slu.se; heidruns@lbhi.is
[1]Department of Animal Biosciences, Swedish University of Agricultural Sciences, P.O. Box 7023, Uppsala 75007, Sweden
[2]Faculty of Agricultural Sciences, Agricultural University of Iceland, Hvanneyri, Borgarbyggð 311, Iceland
[3]Department of Veterinary Science, University of Parma, Parma 43126, Italy
[4]Center for Animal Breeding and Genetics, Department of Biosystems, KU Leuven, Leuven 3001, Belgium

Background

Monitoring the genetic diversity within populations is vital to ensure sustainable breeding and should be performed routinely within breeding programs [1, 2]. This especially applies to breeding programs involving closed populations, such as the Icelandic horse and the Exmoor pony breeds. Both breeds are ancient, native breeds adapted to harsh environmental conditions, and they have both undergone severe historic bottlenecks, albeit more pronounced in the Exmoor pony breed. However, in modern days, the selection pressures on these breeds differ substantially. The Icelandic horse has been bred primarily for its performance in five gaits, while conservation efforts have been the focus for the endangered Exmoor pony. Despite their similar starting points, the divergent breeding goals offer a unique opportunity to study the effects of artificial and natural selection by comparing the genomic selection signatures in these two breeds.

Little is known with certainty about the origin of the Icelandic horse breed, but it is generally believed to have descended from horses brought to the country by Norse settlers around 1100 years ago [3]. Since the settlement, the horses have remained isolated in Iceland and survived harsh weather conditions and natural disasters, such as volcanic eruptions, without significant introduction of foreign genetic material [3, 4]. Before selective breeding started in the 20th century, the horse was mainly used for labour and transportation and was primarily shaped by its harsh natural habitat. In the 1950s, the first official breeding goal, emphasising a versatile riding horse with five gaits, was introduced, resulting in a shift in the selection criteria for the breed [4]. Selective breeding became prevalent, and already in the 1980s, the official breeding program adopted the method of best linear unbiased prediction (BLUP) animal model to estimate breeding values [5, 6]. In the wake of selective breeding and the increased global popularity of the breed during the late 20th century, particularly in northern Europe, the population size surged. In 1959, the population counted approximately 30,000 horses [7], but to date, approximately 300,000 horses are registered across 31 countries [8].

The Exmoor pony, much like the Icelandic horse, is an ancient native breed adapted to harsh conditions. A stud book for the Exmoor pony was established in 1921 to promote the breeding of purebred Exmoor ponies and ensure they retain the traits and characteristics of their ancestors [9]. However, the breed faced a severe population bottleneck during World War II, dwindling to about 50 individuals. Consequently, conservation efforts have prioritized the Exmoor pony, implementing a breeding program specifically designed for its preservation [9]. The breed is named after the high moorland in north-western Somerset and northern Devon, England, where these ponies traditionally roam free. However, Exmoor ponies are also bred at other sites in the UK, Europe, and North America. Today, there are approximately 500 ponies on Exmoor and an additional 3500 Exmoor ponies in various locations across the UK and other countries [9]. About 500 breeding mares and 100 licensed, registered stallions globally produce between 100 and 150 foals annually. Each foal born to registered parents is inspected by trained inspectors to ensure that the Exmoor pony's characteristics and traits are maintained. The Exmoor pony breeding, therefore, focuses on maintaining breed standards, particularly regarding exterior features like coat colour and conformation [9].

Natural and artificial selection tends to reduce genetic variability within targeted genomic regions, resulting in increased homozygosity. These so-called signatures of selection in the genome can be studied using modern genomic methods, such as estimations of continuous homozygous segments called runs of homozygosity (ROH) [10, 11]. To date, estimates of ROH have been used to identify genomic regions potentially under artificial selection in multiple horse breeds. Several genomic regions associated with selection for athletic performance have been identified [12–16], and previously documented target gene (*DMRT3*) related to gait pattern has been confirmed [17, 18]. Furthermore, genomic regions associated with selection for complex traits such as temperament, disease susceptibility, and fertility have been suggested [12, 13, 17–21] as well as regions associated with coat pigmentation characteristics and morphological traits such as body size [17, 18, 21–23].

ROH can be caused by the mating of related animals and are, therefore, a measure of inbreeding [10, 11]. In general, short ROHs indicate distant inbreeding, but longer ROHs (>5.0 Mb) suggest more recent inbreeding where the common ancestor occurs approximately up to 10 generations back [10]. The genomic inbreeding coefficient F_{ROH} is defined as the proportion of the autosomal genome that lies within ROH above a specified length [24]. Recent studies on different horse breeds have reported F_{ROH} estimates to range from 0.10 to 0.29 in breeds with closed stud books [14, 18–23, 25, 26]. In contrast, much lower coefficients have been estimated in breeds with semi-open stud books, such as the Swedish Warmblood horse ($F_{ROH} = 0.006$) [12].

A recent estimate of the mean pedigree-based inbreeding coefficient (F_{PED}) for all Icelandic horses born in Iceland 2020, was reported to be 0.03 [27]. The effective population size (Ne) for the same cohort was estimated to range from 95 to 103 horses depending on the pedigree completeness index [27]. Inbreeding coefficients for the Icelandic horse population have also been estimated using genomic data. An estimate of the average genomic inbreeding coefficient based on microsatellite data was

0.04 [7], while those based on medium-density single-nucleotide polymorphism (SNP) data ranged from 0.08 to 0.13 [16, 18, 28–30]. For the Exmoor ponies, estimated genomic inbreeding coefficients have been reported to range from 0.17 to 0.25 [12, 18, 29].

Genomic data has furthermore been used to estimate the effective population size of both the Icelandic horse and Exmoor pony breeds. For the Icelandic horse, the Ne estimates varied depending on the type of genomic data used: ranging from 215 individuals based on microsatellite data from 442 horses [7] to 555 individuals based on SNP array data from 25 horses [29]. For the Exmoor pony, the Ne was estimated at 216 individuals based on a sample of 24 ponies with SNP array data [29]. Additionally, studies using medium-density SNP array data identified signatures of selection on equine chromosomes (ECA) 3, 10, 11, 15, and 23 in the Icelandic horse [18, 31, 32]. In contrast, a larger number of ROH islands were identified on ECA1-4, 6, 9, 11, 16, 18–19, 22–23, 28, and 30 in Exmoor ponies [17, 18, 31].

Due to inconsistencies in estimates between previous studies regarding especially genetic diversity in the Icelandic horse, as well as indications from pedigree analysis of a decreasing effective population size, updated estimations for this breed based on a larger data set and high-density SNP information are desired. The comparison with the Exmoor pony gives a valuable opportunity to distinguish between detected signatures of selection for performance, and signatures resulting from adaptations to harsh environment.

The aim of this study was therefore to assess genetic diversity and identify runs of homozygosity in the two breeds, and to estimate and compare genomic inbreeding and signatures of selection. We hypothesized that these breeds would share some signatures of natural selection for adaptation in their genomes, whereas signatures of artificial selection for performance would be specific for the Icelandic horse.

Methods

Sample collection

The study included 380 privately owned Icelandic horses born between 1993 and 2016, of which 166 were stallions or geldings and 214 were mares. Hair samples were collected from the horses' tails, and the collection was performed at breeding field tests and visits to trainers and breeders in Iceland and Sweden. The majority of horses were born in Iceland (N=299) and Sweden (N=72), while a few were born in Denmark, Germany, and Norway (N=9). According to previous studies Icelandic horses are well genetically connected within Iceland [7] and across country borders in continental Europe [33], indicating a comparable genetic background of horses in the sample. The sampled individuals were originally selected for different genome-wide association studies; half of the individuals were selected based on mane growth characteristics [34], while the other half was randomly chosen at breeding field tests [35]. All but ten geldings had been shown at a breeding field test and are therefore a part of preselected Icelandic horses more likely to contribute genetic material to future generations [36]. Based on pedigree data from the international Worldfengur database [8], the closest relatedness observed between individuals in the sample were two parent-offspring pairs. In addition, less than 1% of all possible relationships in the dataset were closer than half sibs but less related than full sibs. Efforts were made to balance the contributions from different families and avoid stratification in the data.

Genotype data for 280 Exmoor ponies was retrieved from a previous publication where details concerning data collection are described [37]. The Exmoor ponies were originally selected based on their insect bite hypersensitivity status, avoiding close relatedness as far as possible based on a complete pedigree data four generations back. Furthermore, three subpopulations were reported within the sample set [37].

Genotype data

The procedure of DNA extraction from the Icelandic horse samples was described in the aforementioned genome-wide association studies [34, 35]. The 380 DNA samples were genotyped with the 670 K+Axiom Equine Genotyping Array. Quality control (QC) was performed using PLINK v1.9 software [38, 39]. For the ROH analysis, poorly genotyped data was removed based on criteria of missing genotypes per SNP (>0.10) and missing SNPs per sample (>0.10). No pruning for low minor allele frequency (MAF), deviation from Hardy-Weinberg equilibrium (HWE) or strong LD was done for the ROH analysis as recommended by Meyermans et al. [30]. Criteria for MAF (<0.05) was however added when calculating the effective population size and observed and expected heterozygosity to be able to compare with similar studies. Only autosomal SNP markers were used for downstream analysis. The genotype data for the 280 Exmoor ponies was also derived from a 670 K Axiom Equine Genotyping Array. The same quality control criteria were used for the genotype data for the Exmoor ponies as for the Icelandic horse data described above. SNP positions were according to genomic coordinates in EquCab3.0 reference genome in both data sets.

After QC including pruning for MAF, the number of SNPs to be used for heterozygosity and Ne analyses was 360,755 and 322,209 for the Icelandic horses and Exmoor ponies, respectively. All the samples for the Icelandic horses passed QC, but six samples from the Exmoor pony group were discarded due to missing genotype data, leaving data for 274 Exmoor ponies for further analyses.

For the ROH analysis, QC was conducted without MAF pruning, resulting in 550,405 shared SNPs for downstream analysis using a combined dataset with information from both breeds.

Principal component analysis (PCA) was performed using the SNPRelate package [40] in R (version 4.3.1) [41] as a QC measure to identify outliers or sample mix-ups in the data as well as to visualize the genetic relationships and clustering patterns in the two breeds. The PCA plot, highlighting the distinct genetic signatures of the two breeds, is presented in Additional file 1: Fig. S1.

Pedigree analysis

Pedigree data for the Icelandic horse was obtained from the international Worldfengur database [8]. The pedigree data contained information about individuals born from 1860 to 2023, but the earliest records only included a small proportion of the population at that time. The quality of the pedigree data was estimated by calculating the pedigree completeness using the optiSel package [42] in R (version 4.3.1) [41]. The optiSel package was also used to estimate F_{PED} and Ne based on the pedigree data for the 380 Icelandic horses in this study. This package estimates Ne from the mean rate of increase in coancestry. Velie et al. [37] reported that the pedigree data for the Exmoor ponies was complete for four generations. This pedigree data, of lower depth than that for the Icelandic horses, was not available for pedigree analysis in the present study and therefore we focused on genomic analysis for the Exmoor ponies.

Heterozygosity and effective population size trajectory

Trends in recent Ne trajectories were determined for both breeds using the SNeP v1.1 software [43]. Only Icelandic horses born between 2006 and 2016 (342 horses), and Exmoor ponies born between 1999 and 2009 (148 ponies) were used for the Ne analysis, covering approximately one generation interval. The minimum and maximum distance between pairs of SNPs was set to 0.05 Mb and 40 Mb, respectively, and the alpha value for the formula by Corbin et al. [44] used by the software to estimate Ne from LD was set to 2.2. The recombination rate was furthermore set to 1.24×10^{-8}, and the Sved & Feldman approximation [45] was used as a recombination rate modifier. The default value of 0.05 was used as minimum MAF.

Observed (H_O) and expected heterozygosity (H_E) was estimated for all the Icelandic horses and all the Exmoor ponies using the --het command in PLINK v1.9 [38, 39].

Runs of homozygosity and genomic inbreeding

The detectRUNS package [46] in R (version 4.3.1) [41] was used for analysing ROH with a sliding windows approach. The scanning window size was set equal to 10 SNP loci, and the maximum number of heterozygous or missing SNP in the sliding window was set equal to 0. The ROH parameter settings were optimised following recommendations in Meyermans et al. [30]. The final definition of the settings was as follows: (i) maximum distance between consecutive SNPs equal to 100 kb, (ii) minimum SNP density equal to 0.05 SNP/kb, (iii) minimum number of SNP in a run equal to 10 and (iv) minimum length of a run equal to 100 kb. One missing and one heterozygous SNP was allowed per run. The settings allowed ROH detection for 99.4% of the autosomal genome, indicating high validity of the analysis [30]. The minimum length of a run did not affect the genome coverage. Therefore, it was chosen based on the correlations between the F_{ROH} and F_{PED} values for the Icelandic horses, which was highest ($r=0.57$, $p<2.2\times10^{-16}$) when the minimum ROH length was set equal to 100 kb. The identified ROH were divided into five length classes ($0.1<ROH\leq1$ Mb; $1<ROH\leq2$ Mb; $2<ROH\leq4$ Mb; $4<ROH\leq8$ Mb; and $ROH>8$ Mb).

The F_{ROH} was calculated by summing each individual's total length of ROH and dividing it by the autosomal genome length [24], which was set equal to 2281 Mb, based on the genome length covered by SNPs. F_{ROH} was calculated for each chromosome, length class, and as an average coefficient across the genome for each breed. Furthermore, to facilitate comparison with results from other similar studies, we also calculated the F_{ROH} values for both breeds when the minimum length of a run was set equal to 500 kb instead of 100 kb.

Signatures of selection and gene ontology

ROH islands that were shared by over 70% of the horses in each breed were determined as signatures of selection for that breed. A threshold of 70%, which is conservative compared to values found in the literature [12, 17, 18, 21–23, 25, 26, 47–49], was used to avoid false positive signatures of selection caused by population history events, such as genetic bottlenecks. The EquCab3.0 genomic coordinates of these regions were used to retrieve candidate gene lists from the genome browser Ensembl (release 110, July 2023) [50]. The candidate gene lists were subjected to a gene ontology (GO) analysis using PANTHER v18.0 (released Aug 2023) [51] to determine significantly enriched biological processes and molecular functions positively selected for in the breeds. Further functional annotation of possible candidate genes was performed using the GeneCards database (version 5.18, Oct 2023) [52, 53]. In addition, the Horse QTLdb (release 51, Aug 2023) [54 was used to identify any overlap with previously identified quantitative trait loci (QTL) in horses.

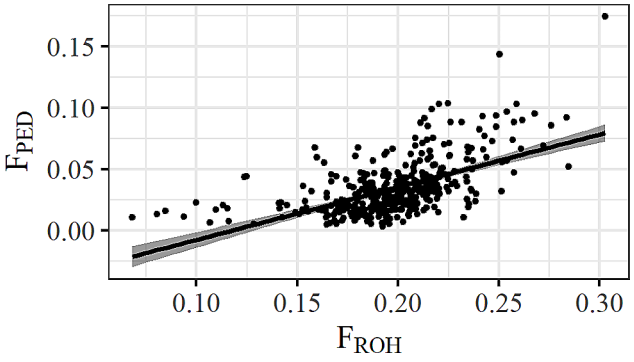


Fig. 1 Correlation between F_{ROH} and F_{PED}.
Pearson correlation between F_{ROH} and F_{PED} in the Icelandic horse, with a 95% confidence interval (grey area)

Results

Pedigree analysis

The evaluation of pedigree quality in the Icelandic horse dataset was based on the average number of discrete generation equivalents, resulting in a value of 8.21 (range: 6.15 to 10.1). This value signifies good pedigree completeness. The average F_{PED} for the 380 Icelandic horses in this study was estimated to be 0.03. The comparison between F_{ROH} and F_{PED} revealed a linear relationship (r=0.57, $p<2.2\times10^{-16}$) (Fig. 1).

Heterozygosity and effective population size trajectory

The overall mean H_O and H_E in the Icelandic horse were equally estimated to be 0.34. In the studied data, H_O values ranged from 0.30 to 0.38. Similarly, the H_E estimated for the Exmoor pony was 0.34 but the mean H_O was 0.33 and ranged between 0.19 and 0.41.

Based on genomic information, the Ne for the last generation of Icelandic horses was estimated to be approximately 125 individuals (Fig. 2). The trend exhibited an overall decline for the past 60 generations, with more pronounced decreases observed around 18–23 and 7–8 generations ago. However, in the most recent 3–4 generations, the Ne trend levelled off, fluctuating within the range of 123 to 127 individuals. In contrast, the genomic Ne estimate for the Exmoor pony was 42 individuals in the last generation. Furthermore, the trend observed for the Exmoor pony over the last 60 generations indicates a consistent, albeit gradual, decrease in Ne.

Runs of homozygosity and genomic inbreeding

A total of 573,746 and 548,302 ROH were identified for the Icelandic horse and the Exmoor pony, respectively (Table 1). In both cases, the majority of the identified ROH (≥96%) was categorised in the shortest length class (0.1 to ≤1 Mb) with an average ROH length of 0.24 Mb in the Icelandic horse and 0.26 Mb in the Exmoor pony. The average occurrence of the short ROH was 1455 per individual in the Icelandic horse data and 1921 per individual in the Exmoor pony data. Only 125 Icelandic horses, out of the 380, carried ROH islands categorised in the longest length class (>8.0 Mb), and on average they carried 2 such ROH islands with a mean length of 10.8 Mb. On the other hand, only one ROH in a single Exmoor pony was identified to belong to the longest length class.

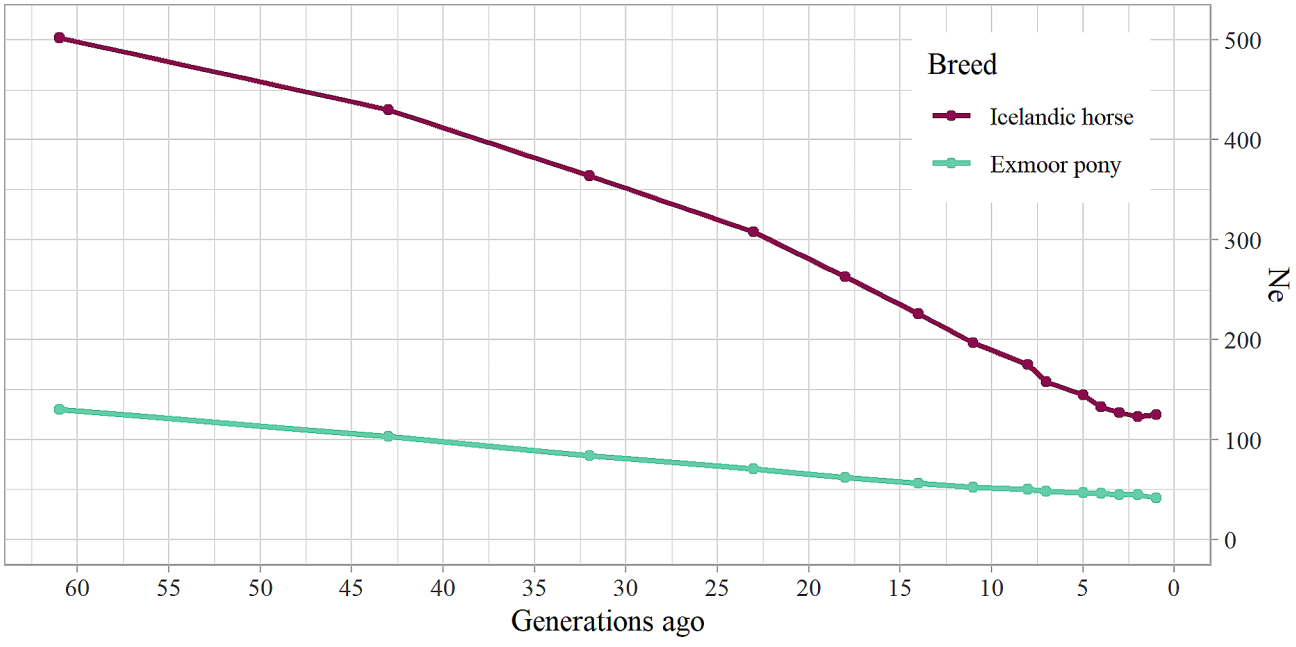


Fig. 2 Trends in effective population size
Changes in effective population size of the Icelandic horse breed and the Exmoor pony over the last 60 generations based on genomic information

Table 1 Descriptive variables from the ROH analysis of the Icelandic horse genome and the Exmoor pony genome

ROH length (Mb)	Icelandic horse					Exmoor pony				
	N_{ind}	N_{ROH}	ROH %	S_{ROH}	L_{ROH}	N_{ind}	N_{ROH}	ROH %	S_{ROH}	L_{ROH}
0.1 to ≤ 1	380	552,929	96.4%	1455	0.24	274	526,242	96.0%	1921	0.26
> 1 to ≤ 2	369	15,677	2.7%	43	1.34	255	19,052	3.5%	75	1.33
> 2 to ≤ 4	335	3,781	0.7%	11	2.66	201	2,879	0.5%	14	2.53
> 4 to ≤ 8	197	1,084	0.2%	5.5	5.40	56	128	0.0%	2.3	4.81
> 8	125	275	0.0%	2.2	10.8	1	1	0.0%	1.0	9.21

N_{ind} = Number of animals, N_{ROH} = total number of ROH, ROH % = relative percentage, S_{ROH} = average number of ROH per animal, L_{ROH} = average length of total number of ROH

ROH quantity, distribution, and average length were estimated per chromosome in both breeds (Fig. 3). The analysis revealed the highest number of ROH in both breeds on ECA1 (N_{ROH} = 47,793 in the Icelandic horse, N_{ROH} = 43,203 in the Exmoor pony), and the lowest on ECA31 (N_{ROH} = 6212 in the Icelandic horse, N_{ROH} = 6227 in the Exmoor pony). For Icelandic horses, ECA23 had the longest average ROH (L_{ROH} = 0.33 Mb), while the shortest (L_{ROH} = 0.25 Mb) were found on ECA12. In Exmoor ponies, the longest average ROH was on ECA22 (L_{ROH} = 0.36 Mb), and the shortest on ECA20 (L_{ROH} = 0.25 Mb).

The estimated mean F_{ROH} was relatively high in both breeds, with a total of 0.20 in the Icelandic horse and 0.27 for the Exmoor pony (Table 2) when including ROH lengths from 100 kb and higher. The individual F_{ROH} ranged from 0.07 to 0.30 for the Icelandic horses, and from 0.01 to 0.55 for the Exmoor ponies. The distribution of average F_{ROH} values across the genome in both breeds is shown in a violin plot in Fig. 4. F_{ROH} estimations for the different ROH length classes revealed that most of the inbreeding could be traced back to the high amount of ROH identified in the shortest length class (0.1 to ≤1 Mb) in both breeds. The length classes comprising longer ROH (>4 to ≤8 Mb, and >8 Mb) accounted for low amount of inbreeding in the Icelandic horse, with F_{ROH} values of 0.01 for each of these two ROH length classes, respectively. In the Exmoor pony, close to zero inbreeding was estimated based only on longer ROH (>4 Mb).

The analysis across chromosomes for the Icelandic horse revealed the highest mean F_{ROH}, including all ROHs (≥0.1 Mb), on ECA23 (F_{ROH} = 0.27±0.09, max=0.74, min=0.10) and the lowest mean F_{ROH} on ECA12 (F_{ROH} = 0.15±0.07, max=0.56, min=0.02) and ECA20 (F_{ROH} = 0.15±0.07, max=0.58, min=0.03). For the Exmoor pony, the highest mean F_{ROH} was identified on ECA22 (F_{ROH} = 0.34±0.17, max=0.97, min=0.01) and ECA23 (F_{ROH} = 0.35±0.17, max=0.92, min=0.00) and the lowest mean F_{ROH} on ECA12 (F_{ROH} = 0.22±0.13, max=0.67, min=0.00) and ECA20 (F_{ROH} = 0.21±0.13, max=0.78, min=0.00). A violin plot of mean genomic inbreeding across chromosomes within each breed is shown in Additional file 2: Fig. S2.

When the minimum length of ROH was set equal to 500 kb instead of 100 kb, and thus not including the shortest ROH (0.1–0.5 Mb), the average F_{ROH} for the Icelandic horse was 0.08 and it was 0.12 for the Exmoor pony. Details of F_{ROH} estimates within different length classes from this analysis are shown in Additional file 3: Table S1.

Signatures of selection and gene ontology

A total of 15 chromosomes (ECA1, ECA3-5, ECA7-9, ECA11-12, ECA17-20, ECA23, and ECA29) contained ROH islands that were shared by more than 70% of the individuals in the Icelandic horse sample (Fig. 5a), while a total of 23 chromosomes (ECA1-9, ECA11-12, ECA14-19, ECA21-24, and ECA30-31) contained ROH islands shared by more than 70% of the Exmoor ponies (Fig. 5b). The most prominent ROH island hot spot in the Icelandic horse, shared by over 90% of the individuals was located on ECA23 in the region where the *DMRT3* gene is located. On the other hand, the most prominent ROH island hot spot in the Exmoor pony, also shared by over 90% of the sampled individuals, was on ECA22. A complete list of all identified ROH islands for both breeds is shown in Additional file 4: Table S2. The two breeds had overlapping ROH islands on six chromosomes (ECA1, ECA3, ECA8, ECA9, ECA17 and ECA19). The list of annotated genes within and in the vicinity of these shared ROH islands is shown in Additional file 5: Table S3.

Overall, 37 annotated genes were located within the identified ROH islands in the Icelandic horse and 289 more in their ±500 kb vicinity. In the Exmoor pony, 181 annotated genes were identified within the ROH islands, and 645 more in their ±500 kb vicinity. Given the considerable number of ROH islands detected, we decided to concentrate on specific regions that could be associated with traits in either or both of the studied breeds in this article.

Table 3 presents the private ROH islands identified in the Icelandic horse that were linked to performance traits, and genes within those ROH. Similarly, Table 4 outlines the private ROH islands identified in the Exmoor ponies and associated genes that are related to coat colours, fertility, hypertension, and the immune system.

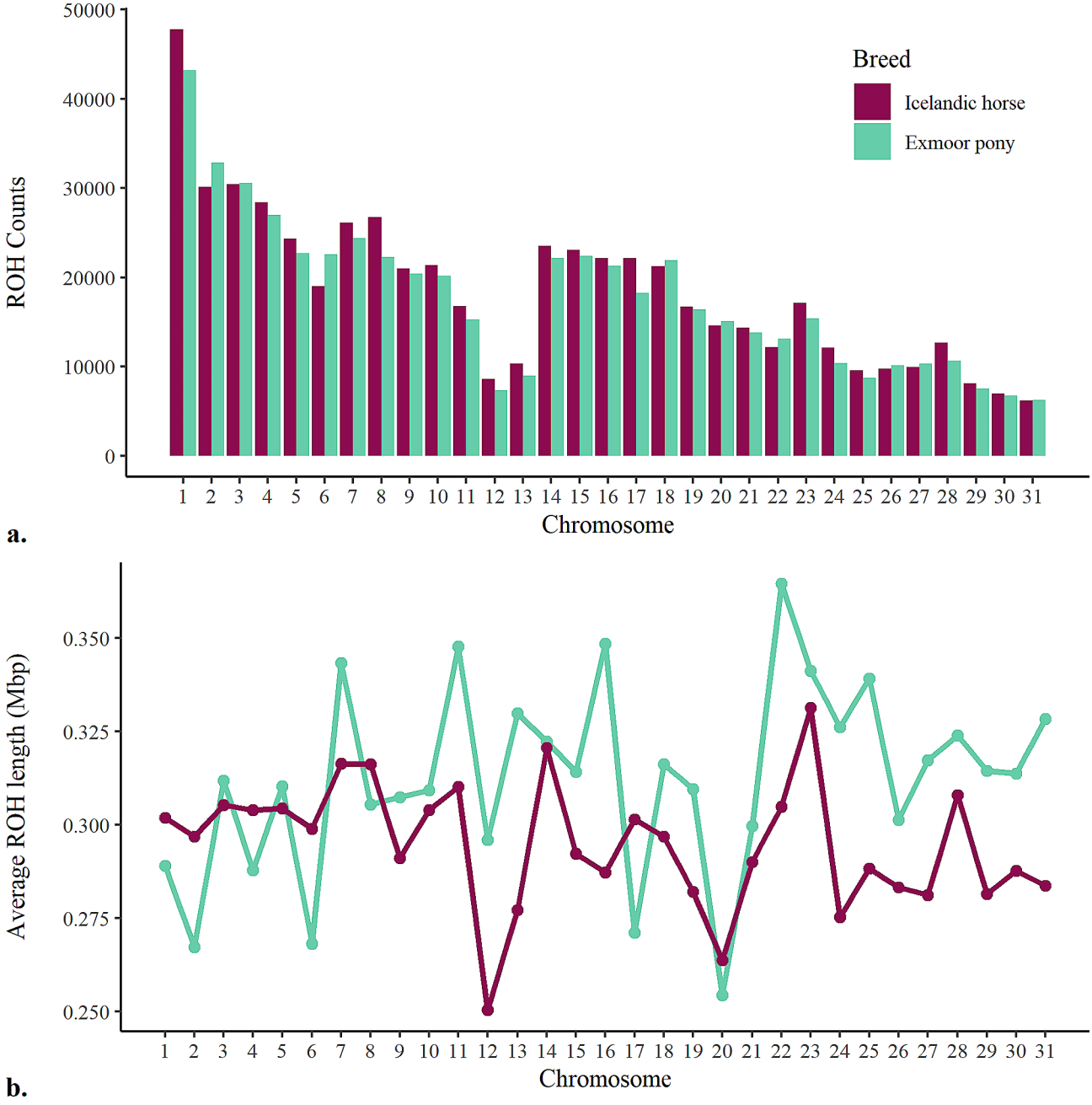


Fig. 3 Distribution and average length of ROH across chromosomes
(**a**) Distribution and (**b**) average length of ROH in Mb detected across the autosomal genome in the Icelandic horse and the Exmoor pony

Table 2 Results for each breed's mean F_{ROH} across the genome and the mean F_{ROH} across the five length classes

ROH length (Mb)	Icelandic horse F_{ROH}				Exmoor pony F_{ROH}			
	Mean	Min	Max	sd	Mean	Min	Max	sd
0.1 to ≤ 1	0.15	0.07	0.21	0.02	0.22	0.01	0.41	0.06
> 1 to ≤ 2	0.02	0.00	0.06	0.01	0.04	0.00	0.12	0.03
> 2 to ≤ 4	0.01	0.00	0.05	0.01	0.02	0.00	0.08	0.01
> 4 to ≤ 8	0.01	0.00	0.06	0.01	0.00	0.00	0.01	0.00
> 8	0.01	0.00	0.05	0.01	0.00	0.00	0.00	na
All ROH lengths	**0.20**	**0.07**	**0.30**	**0.03**	**0.27**	**0.01**	**0.55**	**0.08**

Mean = average F_{ROH} value, Min = minimum F_{ROH} value, Max = maximum F_{ROH} value, sd = standard deviation

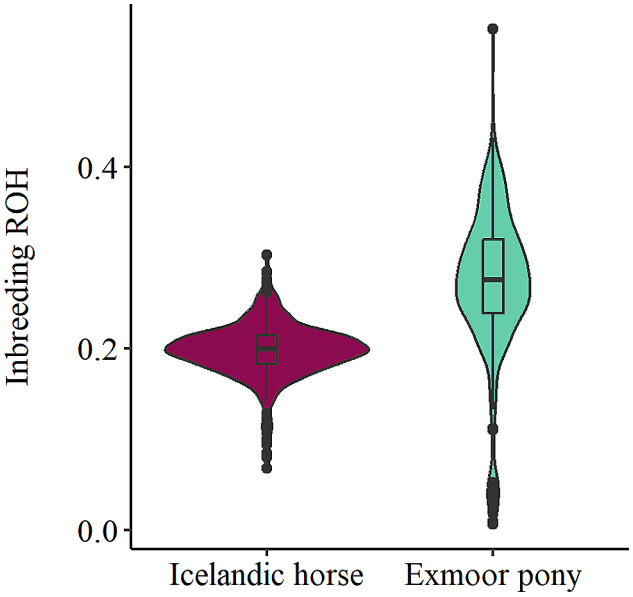


Fig. 4 Violin plot showing distribution of genome-wide F_{ROH} Distribution of average F_{ROH} across the genome for the Icelandic horse (to the left) and the Exmoor pony (to the right) represented with a violin plot including a box plot indicating the median, first and third quartile (Q1 and Q3) and the outliers

Additionally, Table 5 enumerates the shared ROH islands between the two breeds suggested to be associated with the immune system, metabolism, coat type, and body size.

Discussion

Disentangling genomic adaptation from natural and artificial selection within a genome is a challenging task. Some approaches include population genetic analysis, functional genomics, examination of historical data on breeding practices, and genomic comparisons. The Icelandic horse and Exmoor pony share many similarities, being ancient breeds of relatively small body size adapted to harsh conditions, but they also exhibit significant differences due to stringent selection for gait performance in Icelandic horses and emphasis on coat colour and conservation of Exmoor ponies. The comparison of the two breeds, therefore, gives a valuable opportunity to distinguish between detected signatures of selection for performance, and signatures resulting from adaptations to harsh environment. Furthermore, this study is the first to analyse genetic diversity and ROHs in the Icelandic horse using high-density SNP-marker data.

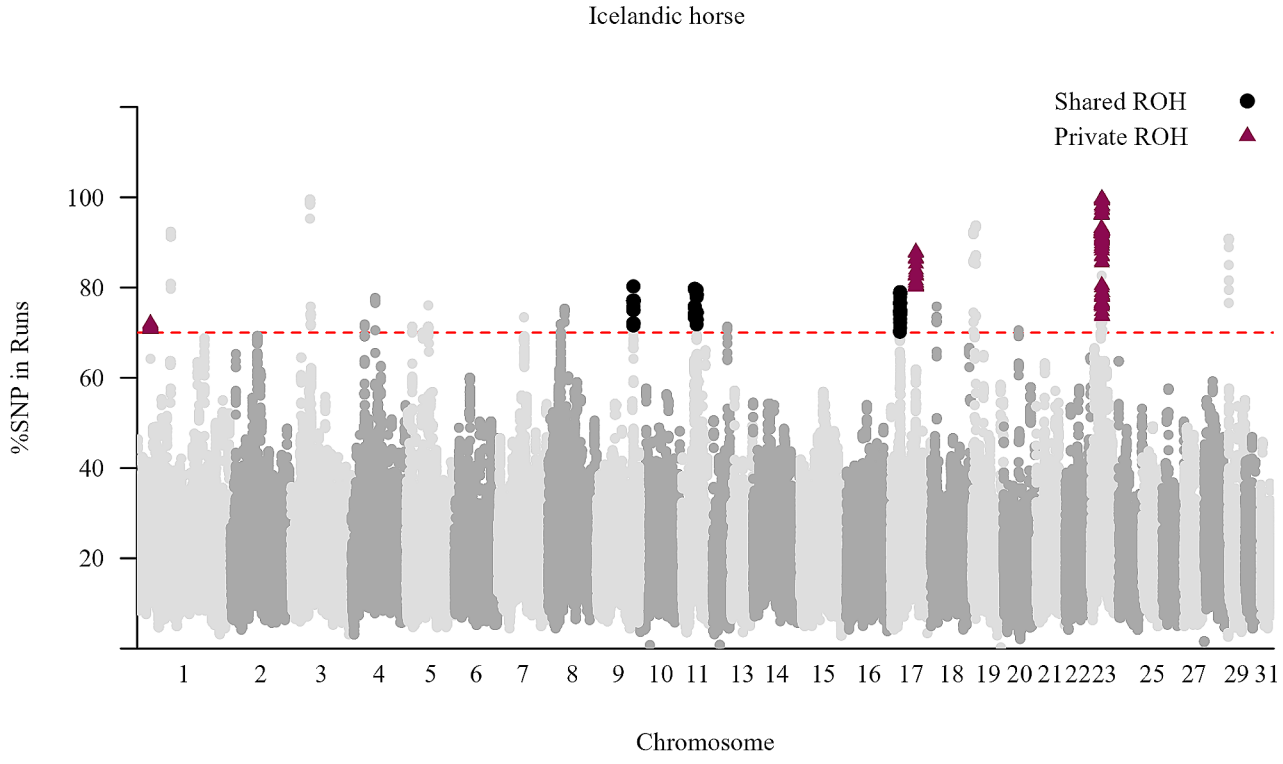
A substantial number of ROH islands, considered indicative of selection signatures, were identified in both breeds, and presented in Additional file 4: Table S2. However, in this discussion we focus on selected regions linked to specific traits relevant to the studied breeds. Shared ROH islands between both breeds were identified, associated with metabolic processes, body size, and the immune system. Additionally, private ROH islands linked to performance in the Icelandic horse and ROH islands associated with coat colours, hypertension, and fertility in the Exmoor pony were identified.

Exploring genetic diversity

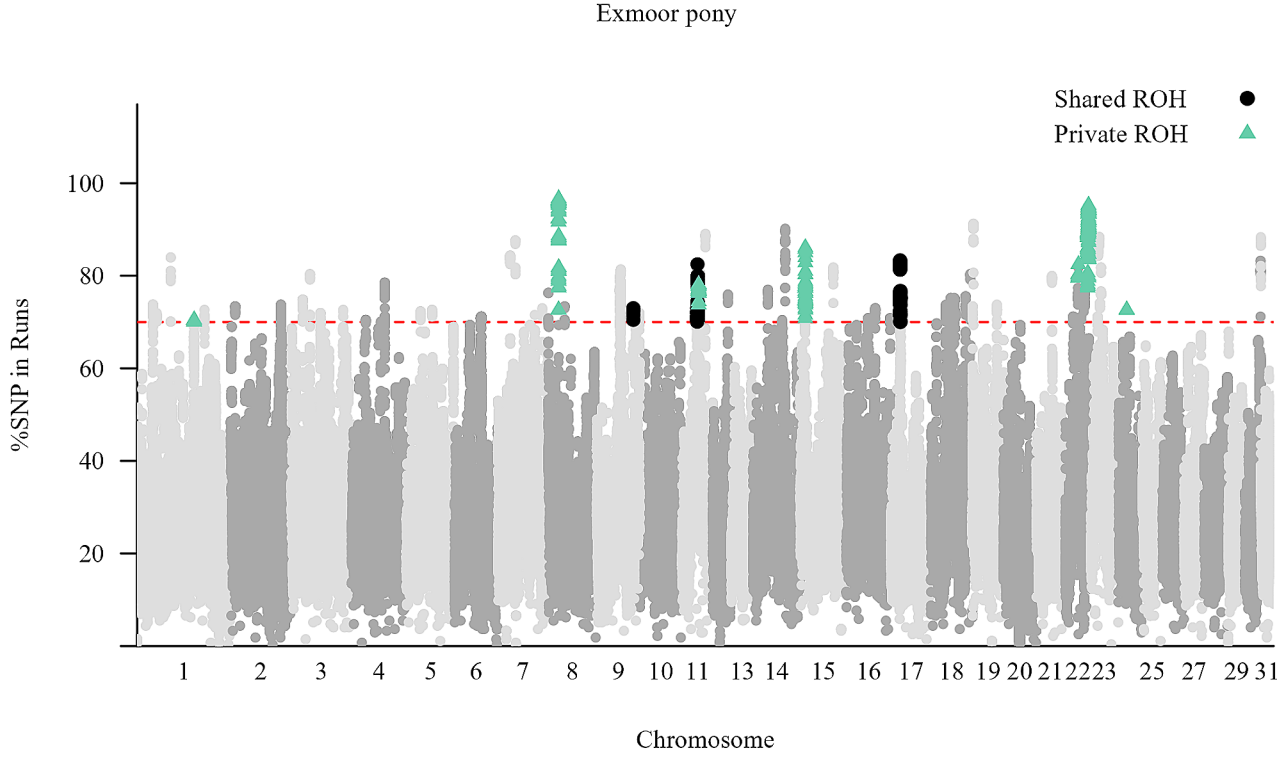
Our findings revealed similar heterozygosity estimates in both the Icelandic horse (mean H_O = 0.34) and the Exmoor pony (mean H_O = 0.33), which were also comparable to estimates reported in other breeds with closed populations. In a study by Cosgrove et al. [16], which estimated heterozygosity for various breeds in the development of the 670k genotyping array [55], reported average $H_O = 0.34 \pm 0.02$ for pony breeds, $H_O = 0.33 \pm 0.01$ for draft horse breeds, and $H_O = 0.32 \pm 0.01$ for the Icelandic horse.

Estimations of SNP based Ne of the Icelandic horse (125 individuals) further suggested an adequate genetic diversity within the breed, and a relative stability in Ne estimations over the last 3–4 generations. The equal estimations of H_E and H_O and the low number of longer ROH support the stability observed in the Ne trend in recent generations, and the absence of strong recent inbreeding. In contrast, the Ne estimate for the Exmoor pony (42 individuals) indicates severe loss of genetic diversity, posing a risk to the breed's sustainability. This decline is likely linked to the grave bottleneck experienced during the Second World War when the pony population diminished to about 50 individuals by the war's end [9]. While the bottleneck may not be explicitly evident in the Ne trend depicted in Fig. 2, the gradual, slow decrease suggests that conservation measures applied after the war may have been useful. However, there were a few Exmoor ponies with unexpectedly low F_{ROH} in the present study and the subpopulations that Velie et al. [37] identified in the Exmoor data, the relatively small sample size, and the lack of background information prevented us from drawing any definitive conclusions for the Ne trend.

Seen over a longer time span for the Icelandic horse, Ne estimations revealed a significant decrease at two historical events, resulting in a major decline in the population size of the Icelandic horse [7]. The first event occurred approximately 23 generations ago, aligning with the end of the 18th century considering a generation interval of roughly 10 years, coinciding with the *Skaftáreldar* volcanic eruption. This eruption had detrimental consequences for both humans and livestock in Iceland, and is said to have reduced the number of horses with 75% [56]. The second event took place around 8 generations ago, during the industrial revolution in Iceland. During this period, the role of the Icelandic horse as a working horse was superseded by machines, leading to a shift in the breeding goal towards breeding riding horses [4, 57].



a.



b.

Fig. 5 Manhattan plot showing ROH islands across chromosomes
ROH islands across all autosomes in (**a**) the Icelandic horse and (**b**) the Exmoor pony. The x-axis represents the chromosome numbers, and the y-axis represents the proportion of animals sharing a ROH. The ROH islands exceeding the 70% threshold (red dotted line) were considered signatures of selection

Table 3 List of selected private ROH islands in the Icelandic horses with annotated genes located within each ROH island and traits possibly associated with the genes or regions

ECA	Start to end position[1]	Length (kb)	nSNP	Annotated genes within ROH island	Suggested associated trait(s)[2]
1	26,821,929 – 26,922,176	100.2	19	*SH3PXD2A*, *NEURL1*	Learning and memory
17	50,290,519 – 50,523,326	232.8	36	*SLAIN1*, *EDNRB*	Performance
23	21,584,553 – 21,696,531	112.0	21	*PGM5*, *ENSECAG00000003227*	Gaits, performance
	21,771,215 – 21,877,092	105.9	17	*DOCK8*	
	22,117,843 – 22,706,518	588.7	117	*KANK1*, *DMRT1*, *DMRT3*, *DMRT2*	

ECA = equine chromosome, Length (kb) = length of a ROH in kilobase, nSNP = number of SNP in a ROH

[1]Positions are according to genomic coordinates in EquCab3.0 reference genome

[2]Based on HorseQTL database and functional annotations

Table 4 List of selected private ROH islands in the Exmoor pony with annotated genes located within each ROH island and traits possibly associated with the genes or regions

ECA	Start to end position[1]	Length (kb)	nSNP	Annotated genes within ROH island	Suggested associated trait(s)[2]
1	114,371,997 – 114,398,670	26.7	5	*OCA2*	Coat colour
	114,458,823 – 114,477,224	18.4	5		
8	20,588,303 – 20,896,604	308.3	50	*TBX3*, *TBX5*	Coat colour
11	33,050,441 – 33,198,699	148.3	22	*TEX14*	Fertility
15	10,154,743 – 10,399,931	245.2	38	*REV1*, *EIF5B*, *TXNDC9*, *LYG1*, *LYG2*	Immune system
22	25,912,035 – 26,060,652	148.6	12	*ENSECAG00000055038*, *ASIP*	Coat colour
22	44,748,478 – 44,900,296	151.8	41	*BMP7*, *ENSECAG00000038425*, *SPO11*	Hypertension and fertility
22	46,065,950 – 46,306,096	240.1	50	*NPEPL1*, *ENSECAG00000004696*, *ENSECAG00000040282*, *GNAS*	Hypertension
	46,313,381 – 46,480,417	167.0	41	*NELFCD*, *CTSZ*, *TUBB1*, *ENSECAG00000060119*, *PRELID3B*	
24	16,374,545 – 16,425,073	50.5	4	*ENSECAG00000031483*, *ENSECAG00000036606*, *MED6*, *ENSECAG00000041394*	Fertility

ECA = equine chromosome, Length (kb) = length of a ROH in kilobase, nSNP = number of SNP in a ROH.

[1]Positions are according to genomic coordinates in EquCab3.0 reference genome.

[2]Based on HorseQTL database and functional annotations.

Table 5 List of selected shared ROH islands in the Icelandic horses (ICE) and the Exmoor pony (EXM) with annotated genes located within each ROH island and traits possibly associated with the genes or regions

ECA	Breed	Start to end position[1]	Length (kb)	nSNP	Annotated genes within ROH island	Suggested associated trait(s)[2]
9	ICE	73,198,547 – 73,286,648	88.1	14	*ENSECAG00000023276*, *CYRIB*	Immune system
	EXM	73,072,557 – 73,156,479	83.9	5	*ENSECAG00000022588*, *ENSECAG00000046146*, *ENSECAG00000053909*	
11	ICE	25,277,282 – 25,336,522	59.2	9	*ABI3*, *ZNF652*	Coat type and body size
		29,120,371 – 29,172,299	51.9	8	//	
		29,181,352 – 29,237,132	55.8	5	//	
	EXM	30,266,841 – 30,334,358	67.5	12	*MMD*	
		30,371,968 – 30,440,645	68.7	8	*SMIM36*	
		30,471,631 – 30,600,164	128.5	21	*TMEM100*	
		30,722,954 – 30,853,756	130.8	16	*ENSECAG00000048512*	
		30,906,432 – 30,998,404	92.0	13	*ANKFN1*	
		31,320,686 – 31,495,910	175.2	26	*NOG*	
17	ICE	18,706,560 – 18,829,942	123.4	17	*FOXO1*	Metabolism
	EXM	18,735,602 – 18,829,942	94.3	14		

ECA = equine chromosome, Length (kb) = length of a ROH in kilobase, nSNP = number of SNP in a ROH

[1]Positions are according to genomic coordinates in EquCab3.0 reference genome

[2]Based on HorseQTL database and functional annotations

While the Ne estimates for the Icelandic horse exceed the generally recommended minimum size for sustaining genetic diversity in selectively bred populations, previous studies have shown a decline in pedigree-based Ne over generations. For instance, in 1989, the pedigree-based Ne was reported as 365 individuals [58], declining to 210 animals by 2000 [59], and currently estimated at around 100 individuals [27]. These figures differ somewhat with the Ne estimates from genomic data presented in Fig. 2 of this study. The pedigree completeness for the Icelandic horse is high, but pedigree measures are limited to probability estimates based on documented ancestry. In contrast, genomic data provides a more detailed and comprehensive view of the population's genetic structure and history but is based on a genotyped subset of the population. Therefore, some differences in pedigree-based and genomic estimates can be expected. However, it is reassuring that the differences were not substantial.

The average F_{PED} estimated in the present study closely aligns with recent calculations for all horses born in Iceland between 2011 and 2020 [27]. This suggests that our sample predominantly consisting of preselected Icelandic breeding horses likely to contribute to future generations fairly accurately represents the latest generation of Icelandic horses. A continued monitoring of relatedness and genetic contribution of breeding animals, and resulting inbreeding trends, is important to ensure a sustainable breeding program. This is especially important within the Icelandic horse population, because a large part of the population is geographically isolated in Iceland where importation of genetic material is prohibited according to the Animal Importation Act [60].

Interpreting genomic inbreeding

The quantity of detected ROH islands in this study was heavily influenced by the parameter settings, notably the minimum ROH length. Many equine studies using 670k SNP data set a minimum ROH length of 500 kb [17, 19, 21, 22, 26], excluding the shortest ROHs (<0.5 Mb). This exclusion further impacts F_{ROH} estimation, which is derived from the total genome length covered by ROHs. In our study, a 100 kb minimum ROH length resulted in the highest correlation between estimated F_{ROH} and F_{PED} values for the Icelandic horse. Using this setting, the mean F_{ROH} was higher (F_{ROH} = 0.20 for the Icelandic horse and F_{ROH} = 0.27 for the Exmoor pony) compared to previous studies. Previous reports for the Exmoor pony ranged from 0.17 to 0.25 [12, 18], while the disparity in F_{ROH} values for the Icelandic horse was more pronounced, with previous estimates ranging from 0.03 to 0.13 [16, 18, 30]. Excluding the shortest ROHs (<0.5 Mb), resulted in values closer to those previously reported. It can be argued that the F_{ROH} estimate based on the longer ROHs, reflecting more recent inbreeding, is more relevant for risk assessment of the current breeding practises [61, 62].

The high average F_{ROH} observed in the Icelandic horse when including the shortest ROHs, may be attributed to the breed's assumedly limited initial genetic pool, potential drift, and genetic purging during its adaptation process. The relatively small number of ROHs longer than 4 Mb suggests no evidence of recent excessive inbreeding. However, when compared with the near absence of long ROHs in the Exmoor pony, there is suggestive evidence of a stronger recent selection in the Icelandic horse. Additionally, there are indications of an increased contribution of a limited number of breeding animals in the Icelandic horse to the modern gene pool [27], emphasizing the importance of closely monitoring inbreeding and genetic diversity in the breed.

ROH island cold spots of different origin

ROH island cold spots were identified on ECA12 and ECA20 in both breeds, where the shortest average ROHs and the lowest mean F_{ROH} were found. The cold spot on ECA20 may be attributed to the major histocompatibility complex (MHC) covering a substantial portion of the chromosome [63, 64]. The MHC is a highly variable region associated with the immune system and benefits from heterozygosity [65], with the Icelandic horse, for instance, showing high MHC heterozygosity [66]. Furthermore, a possible cause of the cold spot identified on ECA12 is the higher percentage of the chromosome covered by copy number variation (CNV) gains and losses compared to other equine chromosomes [67–69]. CNV increases genetic diversity by varying the number of copies of genomic regions [70], indicating higher heterozygosity in the region on ECA12.

Performance-linked ROH islands on ECA23 and ECA17 in the Icelandic horse

The most prominent ROH island hot spot in the Icelandic horse was located on ECA23; a region harbouring genes such as the *DMRT3* and *DOCK8*, both known to be causative or highly associated with gaits and performance in many horse breeds [71–78]. A single mutation [DMRT3:Ser301STOP marker at nucleotide position 22,999,655 on ECA23] in the *DMRT3* gene, also referred to as the 'Gait keeper' mutation, alters the pattern of locomotion and has a predominant effect on gaiting ability in Icelandic horses [71, 72]. The identified ROH harbouring the *DMRT3* gene was the longest ROH (589 kb) identified in this study that was shared by over 70% of the Icelandic horses, indicating recent selection for this region.

This ROH also harbours the *DMRT1*, *DMRT2* and *KANK1* genes and overlaps the 'Gait keeper' haplotype previously identified [71, 79]. Furthermore, this region

overlaps a previously identified selection signature for the Icelandic horse in a study by Petersen et al. [31]. The *DOCK8* gene, located in another ROH (106 kb) on ECA23, has been shown to be associated with harness racing success in Nordic trotters [73]. Furthermore, in a small sample set of Icelandic horses, the *DOCK8* gene was found to be associated with pace racing success and to potentially segregate between elite pace racers and other horses [80]. Previous studies hypothesized overlapping or common gene effects of the *DMRT1-3* genes and the *DOCK8* gene [73, 81, 82].

Even though the 'Gait keeper' mutation has been shown to be a causative factor for gaiting ability, it is highly unlikely that it is the single cause as shown by multiple studies [35, 71, 72, 74, 76–78, 83]. It is therefore possible that the *DOCK8* gene contributes to the performance of gaits, alongside the *DMRT3* gene. The *PGM5* gene, located in the third ROH (112 kb) on ECA23, has no known association with performance in horses. It is predicted to enable metal ion binding activity and phosphoglucomutase activity, and to be associated with myofibril assembly and striated muscle tissue development in zebrafish [84], and may thus be a candidate to study further for performance in horses.

A relatively long ROH island (233 kb) was detected on ECA17 for the Icelandic horses, harbouring the genes *EDNRB* and *SLAIN1*. The *EDNRB* gene harbours the 'Overo allele', which has been shown to be the causative factor for the Overo coat colour in horses and the lethal white foal syndrome (LWFS) in homozygous form [85–87]. Since there are no reports of either the Overo coat colour or the LWFS in the Icelandic horse breed, the apparent selection intensity for this region is likely associated with another function of the gene *EDNRB*, that appears to have pleiotropic effects. The *EDNRB* gene is a part of the endothelin gene family, which plays a crucial role in regulating blood vessel tone and blood pressure [88, 89]. The *EDNRB* interacts with its family members, such as the *EDN3* gene [90–92] suggested to be associated with blood supply regulation in high-performing racing horses [93, 94]. Icelandic horses are trained for high intensity exercises [95, 96], indicating the importance of a robust regulatory system for the distribution of blood to the tissues. This ROH island may, therefore, be a product of selection for performance. The *SLAINI* gene has furthermore been associated with the developing nervous system in mouse embryos [97], indicating a possible importance for performance.

Another possible performance related ROH island was detected on ECA1, where the *NEURL1* and *SH3PXD2A* genes are located. One of the functions of the *NEURL1* gene is hippocampal-dependent synaptic plasticity, which affects learning and memory processes [98, 99]. This region could, therefore, be important for horses trained for performance. These ROH islands on ECA23, ECA17 and ECA1 were not identified in the Exmoor pony genome in this study, further underlining the possible association with performance.

ROH islands distinguished by coat colour genes in the Exmoor pony

The Exmoor pony is renowned for its distinctive bay coat colour and mealy markings. A ROH island hot spot was identified on ECA22, coinciding with the location of the *ASIP* gene which is responsible for the bay coat colour in horses [100]. Additionally, a prominent signature on ECA8, harbouring the *TBX3* and *TBX5* genes, was observed for this breed. While *TBX3* controls dun coat colour [101], the rarity of dun-coloured Exmoor ponies suggests that the signature likely reflects the high prevalence of the non-dun alleles in the gene.

The *OCA2* gene was identified in a ROH island on ECA1 for Exmoor ponies and is known to be one of the components of the mammalian pigmentary system [102–104]. The gene is a major determinant of brown and/or blue eye colour [103–105] and is hypothesized to be a key control point at which ethnic skin colour variation in humans is determined [106]. Efforts have been made to link this gene to horse colour phenotypes [107, 108] without success so far. The *OCA2* gene consistently emerges as a selection signature in the Exmoor pony genome [17], suggesting its potential association with some of their characteristics, such as the mealy markings. The mealy phenotype has previously been linked to the *EDN3* gene [93], which was also identified in this study, located near another ROH island on ECA22. Consequently, we recommend further exploration of these two candidate genes to ascertain their potential association with the mealy phenotype.

No ROH islands harbouring candidate genes for horse colour phenotypes were identified in the Icelandic horse genome. This absence may be attributed to the breeding goal for the Icelandic horse [109], which has consistently aimed at preserving a diverse range of coat colours, presumably leading to higher variability within the colour loci.

Signs of adaptation to limited feed supply

One of the most prominent ROH islands shared by the Icelandic horse and the Exmoor pony, harboured the *FOXO1* gene on ECA17. The *FOXO1* gene has been associated with insulin resistance [110–112] which is one of the key components of the equine metabolic syndrome (EMS) [113]. EMS is generally observed in breeds categorized as "easily fed," which typically require a lower nutritional intake to maintain body weight. These breeds, including the Icelandic horse and Exmoor pony, have historical backgrounds marked by poor feed availability

and periods of starvation. The hypothesis suggests that positive selection for this genomic region has historically contributed to the survival of these breeds in harsh winter conditions but may render them less adaptive to lush pastures and high-energy diets, and in some cases, low workload. Low insulin sensitivity, or even insulin resistance, has been reported in both breeds [114, 115].

Another key aspect of EMS involves a susceptibility to laminitis [113], which has been shown to be associated with hypertension in horses [115]. Moreover, hypertension arises from dysfunction in vascular endothelial cells in humans with type 2 diabetes [116], a syndrome considered closely related with EMS [117]. Additionally, the vascular endothelium plays a crucial role in preventing platelet activation and the adhesion of leukocytes to the vascular wall [115]. Within the Exmoor pony genome, a substantial homozygote region on ECA22 was identified, harbouring three distinct ROH islands. In the first ROH, the genes *BMP7* and *SPO11* were identified; the second contained the *NPEPL1* and *GNAS* genes, while the third encompassed the *NELFCD*, *CTSZ*, *TUBB1*, and *PRELID3B* genes. Notably, all genes in the third ROH are associated with the regulation of platelet properties [52, 53, 118, 119]. Moreover, research has linked the *BMP7* gene to diabetes and vascular calcification in humans [120], while the *SPO11* gene has been linked to endothelial dysfunction resulting from exercise-induced DNA damage in horses [94]. Furthermore, mutations in the *GNAS* gene have been established as causative for McCune-Albright syndrome in humans, a condition known to involve endocrinologic anomalies such as Cushing syndrome [121]. Equine Cushing's disease is recognized in many horse breeds and frequently leads to the development of laminitis [117]. At last, the aforementioned *EDN3* gene, which is a part of the endothelin gene family, is located in a close proximity (>165 kb) to the ROH islands.

The strong evidence of genes associated both directly and indirectly with vascular endothelin regulation in the specified ECA22 region suggests it could be a signature for positive selection, representing an adaptive trait in Exmoor ponies potentially related to varying feed supply. This study did not identify evidence of positive selection for the same region on ECA22 in the Icelandic horse, however.

Hot spot on ECA11 potentially linked to harsh climate adaptation

ROH islands were identified on ECA11 in both the Icelandic horse (25,277,282−29,237,132) and Exmoor pony (30,266,841−31,495,910) within a region that appears to be partly shared among various pony and draft horse breeds [17, 18, 21–23, 31, 47–49]. This region has been shown to have a low recombination rate in horses [122]. It ranges from approximately position 23 Mb to 32 Mb and has predominantly been associated with phenotypes such as a small to medium height at withers, and a compact, muscular body and robust bone structure, as observed in pony and draft horse breeds [123–125], and has also been suggested to be involved in hair and coat density and quality [48, 49, 126].

Whereas phenotypes such as limited height at withers and dense winter coat apply to the Icelandic horse and Exmoor pony, the genes in this wider genomic region previously suggested to be of importance for such traits were not within the ROH islands identified on ECA11 in the present study. However, we cannot exclude that selection has targeted other nearby genes, given the overall high gene density in the region, or that regulatory functions have been selected for. The low recombination rate in the region [122] suggests strong linkage and perhaps participation of many genes in similar processes.

Adapted antibacterial defence

A ROH island identified on ECA9 (73,072,557−73,286,648) shared by both horse breeds may play a crucial role in the immune system. Within this region lies the *CYRIB* gene, which has been shown to be associated with protection against Salmonella bacterial infections in humans and contribute to restricting infections mediated by Mycobacterium tuberculosis and Listeria monocytogenes [127].

Functional annotation analyses of genes found within a ROH island on ECA15 only in the Exmoor pony (10,154,743−10,399,931) revealed an enrichment in GO terms related to the "defence response to Gram-positive bacterium". Gram-positive bacteria include genera like Staphylococcus, Streptococcus, Clostridium, and Listeria, all known to cause diseases of varying severity in horses [128–131]. The genes identified within the ROH include *REV1*, *EIF5B*, *TXNDC9*, *LYG1*, and *LYG2*. The two last ones, *LYG1* and *LYG2*, have been reported to have a significant role in innate immunity in mammals [132]. As far as our knowledge extends, this specific region has not been recognized as a selection signature in other horse breeds, while the *LYG1* and *LYG2* genes have been identified as candidate genes for selection in sheep [133].

Male fertility related genomic regions in the exmoor pony

Three ROH islands detected in the Exmoor pony genome harbour genes related to male fertility traits. First, a ROH island was detected on ECA11, positioned at 33,050,441 to 33,198,699, containing the *TEX14* gene. Second, a ROH island was identified on ECA22 ranging from 44,748,478 to 44,900,296, harbouring the *SPO11* gene. At last, a ROH detected on ECA24 (16,374,545−16,425,073) harboured genes that, by a functional annotation analysis,

revealed an enrichment in the GO term related to "male gonad development".

The *TEX14* gene codes for a testis-specific protein and serves as a crucial element in the intercellular bridges of both male and female embryos. Adult male mice lacking *TEX14* mRNA are unable to reproduce (sterile), while females with the same genetic condition maintain their fertility [134, 135]. *TEX14* has further been suggested to have been targeted by selection for fertility in German warmblood horses [13] and was located within a ROH island detected in the Noriker horse genome [23]. The *SPO11* gene codes for an evolutionarily conserved topoisomerase-like protein that, in mammals, is functionally expressed in gonads during meiosis. It has been shown to be associated with male infertility in mice, humans, and cattle [136–140]. The ROH island on ECA24 contained the *MED6* gene and three novel genes (ENSECAG00000031483, ENSECAG00000036606, ENSECAG00000041394). ENSECAG00000041394 is an orthologue of the *ADAM20* mouse gene. The ADAM metallopeptidase domain 20 (*ADAM20*) gene is specifically expressed in testis and has been associated with male infertility in humans and mice [141–144]. The presence of these ROH islands associated with male fertility implies that this trait may have undergone positive selection in the Exmoor pony breed, as a survival trait in semiferal conditions. These ROHs were not detected in the Icelandic horse in this study.

Conclusions

This study provides insights into the genetic diversity and genomic ROH patterns in the Icelandic horse and Exmoor pony. Our assessments indicate that the genetic diversity in the Icelandic horse is on an acceptable level for a closed population undergoing artificial selection. Nevertheless, it is advisable to maintain ongoing monitoring to guarantee the preservation of genetic diversity and to support sustainable breeding practices for the Icelandic horse. In contrast, our results for the Exmoor pony indicates a critical state of genetic diversity. However, further research accounting for the population structure of the breed is needed to validate our findings.

The F_{ROH} estimates were significantly affected by the parameters employed in the ROH analysis, emphasizing the importance of considering these settings when comparing values across different studies. In our study, the high occurrence of short ROHs led us to attribute a larger extent of the identified inbreeding in both breeds to historical events like the breed's origin, bottlenecks, and adaptation, rather than recent and stringent selection practices.

Several ROH islands associated with performance were identified in the Icelandic horse, effectively distinguishing the breed from the Exmoor pony. The most prominent one on ECA23 featured the longest average ROHs and the highest mean F_{ROH} across all chromosomes, suggesting the most recent and stringent selection pressure. The shared ROH islands observed in both breeds were linked to traits associated with adapting to challenging environments with limited food resources, as well as to immune system function. Conversely, distinct ROH regions specific to Exmoor ponies were associated with their exterior characteristics such as coat colour, along with traits related to immune response and fertility.

In conclusion, this study provides knowledge contributing to preserving genetic diversity and population health in these two equine populations. Furthermore, the obtained results provide important insight into genomic regions shared by the two breeds, which are likely associated with adaptive traits shaped by natural selection. Genomic regions related to performance were identified only in the Icelandic horse, likely reflecting the artificial selection for gaits and performance that has occurred over the past few decades.

Abbreviations

BLUP	Best linear unbiased prediction
CNV	Copy number variation
DNA	Deoxyribonucleic acid
ECA	Equus caballus chromosome
EMS	Equine metabolic syndrome
GO	Gene ontology
HWE	Hardy-Weinberg equilibrium
LD	Linkage disequilibrium
LWFS	Lethal white foal syndrome
MAF	Minor allele frequency
MHC	Major histocompatibility complex
mRNA	Messenger RNA
PCA	Principal component analysis
QC	Quality control
QTL	Quantitative trait loci
RNA	Ribonucleic acid
ROH	Runs of homozygosity
SD	Standard deviation
SNP	Single nucleotide polymorphism

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10682-8.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Acknowledgements

The authors extend their gratitude to all horse owners for their cooperation and contributions. Special thanks are due to Elsa Albertsdóttir for providing valuable insights and pedigree data for the Icelandic horse, to Marie Rhodin for engaging discussions, to Tytti Vanhala at the Swedish University of Agricultural Sciences for her work in DNA extraction and genotyping, and to Diana Ekman of the Swedish Bioinformatics Advisory Program at SciLifeLab for her expertise and guidance in bioinformatics. The publication was made

with the contribution of the researcher Michela Ablondi which is employed with a research contract co-financed by the European Union – PON Research and Innovation 2014-2020 pursuant to art. 24, paragraph 3, lett. a) of Law 30 December 2010, n. 240 and subsequent amendments and of the D.M. 10 August 2021 no. 1062.

Author contributions
HS, TK, GL, and SE initiated and designed the study. HS collected half of the samples. MA provided scripts for the data analysis. HS performed the experiments and data analysis and drafted the manuscript. SE and MA advised with the statistical analysis. HS, MA, TK, GL, and SE contributed to the interpretation of the results. All authors read, suggested modifications, and approved the final manuscript.

Funding
Open access funding provided by Swedish University of Agricultural Sciences. The Agricultural University of Iceland's Doctoral Fund, The Stock Protection Fund for the Icelandic Horse Breed, The Swedish Icelandic Horse Association (SIF), Blikastaðir Fund, The Icelandic Genetic Resource Council in Agriculture (AGROGEN) and The Swedish Research Council (VR).

Data availability
This study did not generate new data; all data used were pre-existing. The Icelandic horse genotypes analysed during the study have been deposited in the European Variation Archive (EVA) [145] at EMBL-EBI under accession number PRJEB74212 (https://www.ebi.ac.uk/eva/?eva-study=PRJEB74212). The Exmoor pony genotypes are available via Figshare (DOI: https://doi.org/10.6084/m9.figshare.3145759).

Declarations

Ethics approval and consent to participate
This study is reported in accordance with ARRIVE guidelines [146]. Hair samples were collected according to ethical approval by the Ethics Committee for Animal Experiments in Uppsala, Sweden (number: 5.8.18–15453/2017) and an animal experiment license by the Icelandic Food and Veterinary Authority in Iceland (number: 2020-04-02/2003120). The study involved only privately owned horses, and informed consent was obtained from the horse owners for their participation in the study.

Consent for publication
Not applicable.

Competing interests
The authors declare competing interests concerning the commercial applications of the current study. GL is a co-inventor of a patent application concerning commercial testing of the *DMRT3* mutation. The stated patent does not restrict research applications of the method. None of the other authors have any competing interests.

Authors' information
Not applicable.

Footnotes
Not applicable.

Received: 27 March 2024 / Accepted: 1 August 2024
Published online: 08 August 2024

References

1. Eding H, Bennewitz J. Measuring genetic diversity in farm animals. In: Oldenbroek K, editor. Utilisation and conservation of farm animal genetic resources. 1st ed. Wageningen: Wageningen Academic; 2007. pp. 103–30.
2. Wolliams J, Berg P, Mäki-Tanila A, Meuwissen T, Fimland E. Sustainable management of animal genetic resources. Ås (NO): Nordisk Genbank Husdyr; 2005.
3. Adalsteinsson S. Origin and conservation of farm animal populations in Iceland. Z Tierz Züchtungsbio. 1981;98:258–64.
4. Arnórsson K. Ræktunin [The breeding]. In: Björnsson GB, Sveinsson HJ, editors. Íslenski hesturinn [The Icelandic horse]. 1st ed. Reykjavík: Mál og menning; 2006. pp. 202–47.
5. Árnason T. Genetic studies on conformation and performance of Icelandic toelter horses [dissertation]. Uppsala: Reklam & katalogtryck; 1983 [cited 2024 February 11].
6. Árnason T. Genetic studies on conformation and performance of Icelandic toelter horses: IV. Best linear unbiased prediction of ten correlated traits by use of an animal model. Acta Agr Scand. 1984;34:450–62.
7. Hreiðarsdóttir GE, Árnason Þ, Svansson V, Hallsson JH. Analysis of the history and population structure of the Icelandic horse using pedigree data and DNA analyses. Icel Agric Sci. 2014;27:63–79.
8. Worldfengur - The studbook of origin for the Icelandic horse. Farmers Association of Iceland and FEIF International Federation of Icelandic Horse, Reykjavik. 2001. https://www.worldfengur.com. Accessed 11 February 2024.
9. The Exmoor Pony Society. https://exmoorponysociety.org.uk. Accessed 30 May 2024.
10. Curik I, Ferenčaković M, Sölkner J. Inbreeding and runs of homozygosity: a possible solution to an old problem. Livest Sci. 2014;166:26–34.
11. Peripolli E, Munari DP, Silva MVGB, Lima ALF, Irgang R, Baldi F. Runs of homozygosity: current knowledge and applications in livestock. Anim Genet. 2017;48:255–71.
12. Ablondi M, Viklund Å, Lindgren G, Eriksson S, Mikko S. Signatures of selection in the genome of Swedish warmblood horses selected for sport performance. BMC Genomics. 2019;20:717.
13. Nolte W, Thaller G, Kuehn C. Selection signatures in four German warmblood horse breeds: tracing breeding history in the modern sport horse. PLoS ONE. 2019;14:e0215913.
14. Hill EW, Stoffel MA, McGivney BA, MacHugh DE, Pemberton JM. Inbreeding depression and the probability of racing in the Thoroughbred horse. P Roy Soc B-Biol Sci. 2022;289:20220487.
15. Santos WB, Schettini GP, Maiorano AM, Bussiman FO, Balieiro JCC, Ferraz GC, et al. Genome-wide scans for signatures of selection in Mangalarga Marchador horses using high-throughput SNP genotyping. BMC Genomics. 2021;22:737.
16. Cosgrove EJ, Sadeghi R, Schlamp F, Holl HM, Moradi-Shahrbabak M, Miraei-Ashtiani SR, et al. Genome diversity and the origin of the arabian horse. Sci Rep. 2020;10:9702.
17. Grilz-Seger G, Neuditschko M, Ricard A, Velie B, Lindgren G, Mesarič M, et al. Genome-wide homozygosity patterns and evidence for selection in a set of European and Near Eastern horse breeds. Genes-Basel. 2019;10:491.
18. Colpitts J, McLoughlin PD, Poissant J. Runs of homozygosity in Sable Island feral horses reveal the genomic consequences of inbreeding and divergence from domestic breeds. BMC Genomics. 2022;23:501.
19. Sadeghi R, Moradi-Shahrbabak M, Miraei Ashtiani SR, Schlamp F, Cosgrove EJ, Antczak DF. Genetic diversity of persian arabian horses and their relationship to other native Iranian horse breeds. J Hered. 2018;110:173–82.
20. Laseca N, Molina A, Ramón M, Valera M, Azcona F, Encina A et al. Fine-scale analysis of runs of homozygosity islands affecting fertility in mares. Front Vet Sci. 2022;9.
21. Ablondi M, Dadousis C, Vasini M, Eriksson S, Mikko S, Sabbioni A. Genetic diversity and signatures of selection in a native Italian horse breed based on SNP data. Animals. 2020;10:1005.
22. Grilz-Seger G, Druml T, Neuditschko M, Dobretsberger M, Horna M, Brem G. High-resolution population structure and runs of homozygosity reveal the genetic architecture of complex traits in the Lipizzan horse. BMC Genomics. 2019;20:174.
23. Grilz-Seger G, Druml T, Neuditschko M, Mesarič M, Cotman M, Brem G. Analysis of ROH patterns in the Noriker horse breed reveals signatures of selection for coat color and body size. Anim Genet. 2019;50:334–46.
24. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. Am J Hum Genet. 2008;83:359–72.
25. Bizarria dos Santos W, Pimenta Schettini G, Fonseca MG, Pereira GL, Loyola Chardulo LA, Rodrigues Machado Neto O, et al. Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the Mangalarga Marchador horse breed. J Anim Breed Genet. 2021;138:161–73.
26. Mousavi SF, Razmkabir M, Rostamzadeh J, Seyedabadi H-R, Naboulsi R, Petersen JL, et al. Genetic diversity and signatures of selection in four indigenous horse breeds of Iran. Heredity. 2023;131:96–108.

27. Árnadóttir E. Erfðafjölbreytileiki íslenska hrossastofnsins (Genetic diversity of the Icelandic horse population). BSc thesis, Agricultural University of Iceland. 2022.
28. McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM, et al. A high density SNP array for the domestic horse and extant perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. PLOS Genet. 2012;8:e1002451.
29. Petersen JL, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, Bailey E, et al. Genetic diversity in the modern horse illustrated from genome-wide SNP data. PLoS ONE. 2013;8:e54997.
30. Meyermans R, Gorssen W, Buys N, Janssens S. How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. BMC Genomics. 2020;21:94.
31. Petersen JL, Mickelson JR, Rendahl AK, Valberg SJ, Andersson LS, Axelsson J, et al. Genome-wide analysis reveals selection for important traits in domestic horse breeds. PLOS Genet. 2013;9:e1003211.
32. Gorssen W, Meyermans R, Janssens S, Buys N. A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species. Genet Sel Evol. 2021;53:2.
33. Shrestha M, Solé M, Ducro BJ, Sundquist M, Thomas R, Schurink A, et al. Genome-wide association study for insect bite hypersensitivity susceptibility in horses revealed novel associated loci on chromosome 1. J Anim Breed Genet. 2020;137:223–33.
34. Rosengren MK, Sigurðardóttir H, Eriksson S, Naboulsi R, Jouni A, Novoa-Bravo M, et al. A QTL for conformation of back and croup influences lateral gait quality in Icelandic horses. BMC Genomics. 2021;22:267.
35. Sigurðardóttir H, Boije H, Albertsdóttir E, Kristjansson T, Rhodin M, Lindgren G, et al. The genetics of gaits in Icelandic horses goes beyond DMRT3, with RELN and STAU2 identified as two new candidate genes. Genet Sel Evol. 2023;55:89.
36. Albertsdóttir E, Eriksson S, Sigurdsson Á, Árnason T. Genetic analysis of 'breeding field test status' in Icelandic horses. J Anim Breed Genet. 2011;128:124–32.
37. Velie BD, Shrestha M, François L, Schurink A, Tesfayonas YG, Stinckens A, et al. Using an inbred horse breed in a high density genome-wide scan for genetic risk factors of insect bite hypersensitivity (IBH). PLoS ONE. 2016;11:e0152966.
38. Purcell S, Chang C. PLINK v1.90. 2021. https://www.cog-genomics.org/plink/1.9/ Accessed 11 February 2024.
39. Purcell S. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J hum Genet. 2007;81:559–75.
40. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28:3326–8.
41. R Core team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2019.
42. Wellman R, optiSel. Optimum Contribution Selection and Population Genetics. 2023. https://CRAN.R-project.org/package=optiSel. Accessed 11 February 2024.
43. Barbato M, Orozco-terWengel P, Tapio M, Bruford MW. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front Genet. 2015;6:109.
44. Corbin LJ, Liu A, Bishop S, Woolliams J. Estimation of historical effective population size using linkage disequilibria with marker data. J Anim Breed Genet. 2012;129:257–70.
45. Sved J, Feldman M. Correlation and probability methods for one and two loci. Theor Popul Biol. 1973;4:129–32.
46. Biscarini F, Cozzi P, Gaspa G, Marras G, detectRUNS. Detect runs of homozygosity and runs of heterozygosity in diploid genomes. 2018. https://cran.r-project.org/package=detectRUNS. Accessed 11 February 2024.
47. Grilz-Seger G, Mesarič M, Cotman M, Neuditschko M, Druml T, Brem G. Runs of homozygosity and population history of three horse breeds with small population size. J Equine Vet Sci. 2018;71:27–34.
48. Nazari F, Seyedabadi H-R, Noshary A, Emamjomeh-Kashan N, Banabazi M-H. A genome-wide scan for signatures of selection in kurdish horse breed. J Equine Vet Sci. 2022;113:103916.
49. Amano T, Yokawa H, Masuda Y, Tozaki T, Kawai M, Shirai K. Genome-wide search reveals the uniqueness of DNA regions associated with coat color and innate immunity in Hokkaido native horse. Anim Sci J. 2023;94:e13884.
50. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al. Ensembl 2022. Nucleic Acids Res. 2022;50:D988–95.
51. Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou L-P, Mi H. PANTHER: making genome-scale phylogenetics accessible to all. Protein Sci. 2022;31:8–22.
52. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinform. 2016;54:1301–13033.
53. Safran M, Rosen N, Twik M, BarShir R, Stein TI, Dahary D, et al. The GeneCards suite. In: Abugessaisa I, Kasukawa T, editors. Practical guide to Life Science Databases. Singapore: Springer Nature Singapore; 2021. pp. 27–56.
54. Hu Z-L, Park CA, Reecy JM. Bringing the animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Res. 2022;50:D956–61.
55. Schaefer RJ, Schubert M, Bailey E, Bannasch DL, Barrey E, Bar-Gal GK, et al. Developing a 670k genotyping array to tag ~ 2 M SNPs across 24 horse breeds. BMC Genomics. 2017;18:565.
56. Björnsson GB, Sveinsson HJ. Á spjöldum sögunnar [The history of the Icelandic horse]. In: Björnsson GB, Sveinsson HJ, editors. Íslenski hesturinn [The Icelandic horse]. 1st ed. Reykjavík: Mál og menning; 2006. pp. 76–101.
57. Björnsson GB, Sveinsson HJ. Á tímamótum [At a turning point in time]. In: Björnsson GB, Sveinsson HJ, editors. Íslenski hesturinn [The Icelandic horse]. 1st ed. Reykjavík: Mál og menning; 2006. pp. 102–13.
58. Árnason T, Klemetsdal G, Sigurssson Á. Nordiska hästraser - gamla generesurser ägnade for framtidens behov. In: International Symposium on Horse Breeding and Production in Cold Climatic Regions; 11–13 August; Hotel Saga. Reykjavík, Iceland; 1993.
59. Kristjansson T. Erfðafjölbreytileiki íslenska hrossastofnsins og verndun hans. In: Dýrmundsson Ó, editor. Fjölrit Lbhí Nr. 14. Íslensk búfjárrækt: Málstofa til heiðurs Hjalta Gestssyni níræðum. 2006 November 17; Reykjavik, Iceland. Hvanneyri, Iceland: The Agricultural University of Iceland; 2007. pp. 95–102.
60. Animal Importation. Act 1990 (IS) s 2.
61. Sumreddee P, Hay EH, Toghiani S, Roberts A, Aggrey SE, Rekaya R. Grid search approach to discriminate between old and recent inbreeding using phenotypic, pedigree and genomic information. BMC Genomics. 2021;22:538.
62. Stoffel MA, Johnston SE, Pilkington JG, Pemberton JM. Mutation load decreases with haplotype age in wild Soay sheep. Evol Lett. 2021;5:187–95.
63. Ansari HA, Hediger R, Fries R, Stranzinger G. Chromosomal localization of the major histocompatibility complex of the horse (ELA) by in situ hybridization. Immunogenetics. 1988;28:362–4.
64. Gustafson A, Tallmadge RL, Ramlachan N, Miller D, Bird H, Antczak DF, et al. An ordered BAC contig map of the equine major histocompatibility complex. Cytogenet Genome Res. 2003;102:189–95.
65. Kelley J, Walter L, Trowsdale J. Comparative genomics of major histocompatibility complexes. Immunogenetics. 2005;56:683–95.
66. Holmes CM, Violette N, Miller D, Wagner B, Svansson V, Antczak DF. MHC haplotype diversity in Icelandic horses determined by polymorphic microsatellites. Genes Immun. 2019;20:660–70.
67. Solé M, Ablondi M, Binzer-Panchal A, Velie BD, Hollfelder N, Buys N, et al. Inter- and intra-breed genome-wide copy number diversity in a large cohort of European equine breeds. BMC Genomics. 2019;20:759.
68. Laseca N, Molina A, Valera M, Antonini A, Demyda-Peyrás S. Copy number variation (CNV): a new genomic insight in horses. Animals. 2022;12:1435.
69. Wang W, Wang S, Hou C, Xing Y, Cao J, Wu K, et al. Genome-wide detection of copy number variations among diverse horse breeds by array CGH. PLoS ONE. 2014;9:e86860.
70. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006;444:444–54.
71. Andersson LS, Larhammar M, Memic F, Wootz H, Schwochow D, Rubin C-J, et al. Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. Nature. 2012;488:642–6.
72. Kristjansson T, Bjornsdottir S, Sigurdsson A, Andersson L, Lindgren G, Helyar S, et al. The effect of the 'Gait keeper' mutation in the DMRT3 gene on gaiting ability in Icelandic horses. J Anim Breed Genet. 2014;131:415–25.
73. Velie BD, Fegraeus KJ, Solé M, Rosengren MK, Røed KH, Ihler C-F, et al. A genome-wide association study for harness racing success in the norwegian-swedish coldblooded trotter reveals genes for learning and energy metabolism. BMC Genet. 2018;19:80.
74. Jäderkvist K, Andersson LS, Johansson AM, Árnason T, Mikko S, Eriksson S, et al. The DMRT3 'gait keeper' mutation affects performance of nordic and standardbred trotters. J Anim Sci. 2014;92:4279–86.
75. Promerová M, Andersson L, Juras R, Penedo M, Reissmann M, Tozaki T, et al. Worldwide frequency distribution of the 'Gait keeper' mutation in the DMRT3 gene. Anim Genet. 2014;45:274–82.
76. Patterson L, Staiger E, Brooks S. DMRT3 is associated with gait type in Mangalarga Marchador horses, but does not control gait ability. Anim Genet. 2015;46:213–5.

77. Novoa-Bravo M, Jäderkvist Fegraeus K, Rhodin M, Strand E, García LF, Lindgren G. Selection on the Colombian paso horse's gaits has produced kinematic differences partly explained by the DMRT3 gene. PLoS ONE. 2018;13:e0202584.
78. Chandra Paul R, Ba Nguyen T, Okuda Y, Nu Anh Le T, Mosese Dau Tabuyaqona J, Konishi Y, et al. Distribution of the mutant allele of the DMRT3 gene associated with ambling gaits in Japanese native horse populations. Anim Sci J. 2020;91:e13431.
79. Staiger EA, Almén MS, Promerová M, Brooks S, Cothran EG, Imsland F, et al. The evolutionary history of the DMRT3 'Gait keeper' haplotype. Anim Genet. 2017;48:551–9.
80. Bas Conn L. The role of polymorphisms of the MSTN, GRIN2B and DOCK8 genes in the performance of pace-racing Icelandic horses. MSc thesis, Swedish University of Agricultural Sciences. 2018.
81. Kang JU, Koo SH, Kwon KC, Park JW. Frequent silence of chromosome 9p, homozygous DOCK8, DMRT1 and DMRT3 deletion at 9p24. 3 in squamous cell carcinoma of the lung. Int J Oncol. 2010;37:327–35.
82. Glessner JT, Li J, Wang D, March M, Lima L, Desai A, et al. Copy number variation meta-analysis reveals a novel duplication at 9p24 associated with multiple neurodevelopmental disorders. Genome Med. 2017;9:106.
83. Jäderkvist K, Holm N, Imsland F, Árnason T, Andersson L, Andersson LS, et al. The importance of the *DMRT3* 'Gait keeper' mutation on riding traits and gaits in Standardbred and Icelandic horses. Livest Sci. 2015;176:33–9.
84. Molt S, Bührdel JB, Yakovlev S, Schein P, Orfanos Z, Kirfel G, et al. Aciculin interacts with filamin C and Xin and is essential for myofibril assembly, remodeling and maintenance. J Cell Sci. 2014;127:3578–92.
85. Metallinos DL, Bowling AT, Rine J. A missense mutation in the endothelin-B receptor gene is associated with lethal white foal syndrome: an equine version of Hirschsprung disease. Mamm Genome. 1998;9:426–31.
86. Santschi EM, Purdy AK, Valberg SJ, Vrotsos PD, Kaese H, Mickelson JR. Endothelin receptor B polymorphism associated with lethal white foal syndrome in horses. Mamm Genome. 1998;9:306–9.
87. Yan GC, Croaker D, Zhang AL, Manglick P, Cartmill T, Cass D. A dinucleotide mutation in the endothelin-B receptor gene is associated with lethal white foal syndrome (LWFS); a horse variant of Hirschsprung disease (HSCR). Hum Mol Genet. 1998;7:1047–52.
88. Yanagisawa M, Kurihara H, Kimura S, Tomobe Y, Kobayashi M, Mitsui Y, et al. A novel potent vasoconstrictor peptide produced by vascular endothelial cells. Nature. 1988;332:411–5.
89. Inoue A, Yanagisawa M, Kimura S, Kasuya Y, Miyauchi T, Goto K, et al. The human endothelin family: three structurally and pharmacologically distinct isopeptides predicted by three separate genes. P Natl Acad Sci USA. 1989;86:2863–7.
90. Baynash AG, Hosoda K, Giaid A, Richardson JA, Emoto N, Hammer RE, et al. Interaction of endothelin-3 with endothelin-B receptor is essential for development of epidermal melanocytes and enteric neurons. Cell. 1994;79:1277–85.
91. Hosoda K, Hammer RE, Richardson JA, Baynash AG, Cheung JC, Giaid A, et al. Targeted and natural (piebald-lethal) mutations of endothelin-B receptor gene produce megacolon associated with spotted coat color in mice. Cell. 1994;79:1267–76.
92. Stanchina L, Baral V, Robert F, Pingault V, Lemort N, Pachnis V, et al. Interactions between *Sox10*, *Edn3* and *Ednrb* during enteric nervous system and melanocyte development. Dev Biol. 2006;295:232–49.
93. Jäderkvist Fegraeus K, Velie BD, Axelsson J, Ang R, Hamilton NA, Andersson L, et al. A potential regulatory region near the *EDN3* gene may control both harness racing performance and coat color variation in horses. Physiol Rep. 2018;6:e13700.
94. Fegraeus K, Rosengren MK, Naboulsi R, Orlando L, Åbrink M, Jouni A, et al. An endothelial regulatory module links blood pressure regulation with elite athletic performance. PLOS Genet. 2024;20:e1011285.
95. Stefánsdóttir G, Ragnarsson S, Gunnarsson V, Jansson A. Physiological response to a breed evaluation field test in Icelandic horses. Animal. 2014;8:431–9.
96. Stefánsdóttir G, Ragnarsson S, Gunnarsson V, Roepstorff L, Jansson A. A comparison of the physiological response to tölt and trot in the Icelandic horse. J Anim Sci. 2015;93:3862–70.
97. Hirst CE, Lim S-M, Pereira LA, Mayberry RA, Stanley EG, Elefanty AG. Expression from a betageo gene trap in the Slain1 gene locus is predominantly associated with the developing nervous system. Int J Dev Biol. 2010;54:1383–8.
98. Pavlopoulos E, Trifilieff P, Chevaleyre V, Fioriti L, Zairis S, Pagano A, et al. Neuralized1 activates *CPEB3*: a function for nonproteolytic ubiquitin in synaptic plasticity and memory storage. Cell. 2011;147:1369–83.
99. Taal K, Tuvikene J, Rullinkov G, Piirsoo M, Sepp M, Neuman T, et al. Neuralized family member *NEURL1* is a ubiquitin ligase for the cGMP-specific phosphodiesterase 9A. Sci Rep. 2019;9:7104.
100. Rieder S, Taourit S, Mariat D, Langlois B, Guérin G. Mutations in the agouti (*ASIP*), the extension (*MC1R*), and the brown (*TYRP1*) loci and their association to coat color phenotypes in horses (*Equus caballus*). Mamm Genome. 2001;12:450–5.
101. Imsland F, McGowan K, Rubin C-J, Henegar C, Sundström E, Berglund J, et al. Regulatory mutations in *TBX3* disrupt asymmetric hair pigmentation that underlies Dun camouflage color in horses. Nat Genet. 2016;48:152–8.
102. Lee S-T, Nicholls RD, Jong MT, Fukai K, Spritz RA. Organization and sequence of the human P gene and identification of a new family of transport proteins. Genomics. 1995;26:354–63.
103. Sturm RA, Frudakis TN. Eye colour: portals into pigmentation genes and ancestry. Trends Genet. 2004;20:327–32.
104. Sturm RA, Duffy DL, Zhao ZZ, Leite FP, Stark MS, Hayward NK, et al. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. Am J Hum Genet. 2008;82:424–31.
105. Duffy DL, Montgomery GW, Chen W, Zhao ZZ, Le L, James MR, et al. A three-single-nucleotide polymorphism haplotype in intron 1 of OCA2 explains most human eye-color variation. Am J Hum Genet. 2007;80:241–52.
106. Sturm RA. Molecular genetics of human pigmentation diversity. Hum Mol Genet. 2009;18:R9–17.
107. Kowalski EJA, Bellone RR. Investigation of HERC2 and OCA2 SNP for iris color variation in Puerto Rican Paso Fino horses. J Equine Vet Sci. 2011;31:319.
108. Bellone R, Lawson S, Hunter N, Archer S, Bailey E. Analysis of a SNP in exon 7 of equine OCA2 and its exclusion as a cause for Appaloosa spotting. Anim Genet. 2006;37:525.
109. General rules and regulations: Breeding rules and regulations. 2023. https://www.feiffengur.com/documents/FEIF Breeding_2023.pdf. Accessed 5 January 2024.
110. Zhang T, Kim DH, Xiao X, Lee S, Gong Z, Muzumdar R, et al. *FoxO1* plays an important role in regulating β-cell compensation for insulin resistance in male mice. Endocrinology. 2016;157:1055–70.
111. Puig O, Tjian R. Transcriptional feedback control of insulin receptor by dFOXO/FOXO1. Gene Dev. 2005;19:2435–46.
112. Matsumoto M, Han S, Kitamura T, Accili D. Dual role of transcription factor FoxO1 in controlling hepatic insulin sensitivity and lipid metabolism. J Clin Invest. 2006;116:2464–72.
113. Frank N, Geor RJ, Bailey SR, Durham AE, Johnson PJ. Equine metabolic syndrome. J Vet Intern Med. 2010;24:467–75.
114. Bröjer J, Lindåse S, Hedenskog J, Alvarsson K, Nostell K. Repeatability of the combined glucose-insulin tolerance test and the effect of a stressor before testing in horses of 2 breeds. J Vet Intern Med. 2013;27:1543–50.
115. Bailey SR, Habershon-Butcher JL, Ransom KJ, Elliott J, Menzies-Gow NJ. Hypertension and insulin resistance in a mixed-breed population of ponies predisposed to laminitis. Am J Vet Res. 2008;69:122–9.
116. Jansson P-A. Endothelial dysfunction in insulin resistance and type 2 diabetes. J Vet Intern Med. 2007;262:173–83.
117. Johnson P. The equine metabolic syndrome peripheral Cushing's syndrome. Vet Clin N Am-Equine. 2002;18:271–93.
118. Gieger C, Radhakrishnan A, Cvejic A, Tang W, Porcu E, Pistis G, et al. New gene functions in megakaryopoiesis and platelet formation. Nature. 2011;480:201–8.
119. Qayyum R, Snively BM, Ziv E, Nalls MA, Liu Y, Tang W, et al. A meta-analysis and genome-wide association study of platelet count and mean platelet volume in African americans. PLOS Genet. 2012;8:e1002491.
120. Freedman BI, Bowden DW, Ziegler JT, Langefeld CD, Lehtinen AB, Rudock ME, et al. Bone morphogenetic protein 7 (BMP7) gene polymorphisms are associated with inverse relationships between vascular calcification and BMD: the diabetes heart study. J Bone Min Res. 2009;24:1719–27.
121. Lumbroso S, Paris Fo, Sultan C. Activating Gsα mutations: analysis of 113 patients with signs of McCune-Albright syndrome - A European collaborative study. J Clin Endocr Metab. 2004;89:2107–13.
122. Szmatoła T, Gurgul A, Jasielczuk I, Oclon E, Ropka-Molik K, Stefaniuk-Szmukier M, et al. Assessment and distribution of runs of homozygosity in horse breeds representing different utility types. Animals. 2022;12:3293.
123. Makvandi-Nejad S, Hoffman GE, Allen JJ, Chu E, Gu E, Chandler AM, et al. Four loci explain 83% of size variation in the horse. PLoS ONE. 2012;7:e39929.

124. Bartholazzi Junior A, Quirino CR, Vega WHO, Rua MAS, David CMG, Jardim JG. Polymorphisms in the LASP1 gene allow selection for smaller stature in ponies. Livest Sci. 2018;216:160–4.
125. Skujina I, Winton CL, Hegarty MJ, McMahon R, Nash DM, Morel MCGD, et al. Detecting genetic regions associated with height in the native ponies of the British Isles by using high density SNP genotyping. Genome. 2018;61:767–70.
126. Thomer A, Gottschalk M, Christmann A, Naccache F, Jung K, Hewicker-Trautwein M, et al. An epistatic effect of *KRT25* on *SP6* is involved in curly coat in horses. Sci Rep. 2018;8:6374.
127. Yuki KE, Marei H, Fiskin E, Eva MM, Gopal AA, Schwartzentruber JA, et al. *CYRI/FAM49B* negatively regulates RAC1-driven cytoskeletal remodelling and protects against bacterial infection. Nat Microbiol. 2019;4:1516–31.
128. Van den Eede A, Martens A, Lipinska U, Struelens M, Deplano A, Denis O, et al. High occurrence of methicillin-resistant Staphylococcus aureus ST398 in equine nasal samples. Vet Microbiol. 2009;133:138–44.
129. Sweeney CR, Timoney JF, Newton JR, Hines MT. *Streptococcus equi* infections in horses: guidelines for treatment, control, and prevention of strangles. J Vet Intern Med. 2005;19:123–34.
130. Uzal FA, Navarro MA, Asin J, Henderson EE. Clostridial diseases of horses: a review. Vaccines. 2022;10:318.
131. Warner SL, Boggs J, Lee JK, Reddy S, Banes M, Cooley J. Clinical, pathological, and genetic characterization of *Listeria monocytogenes* causing sepsis and necrotizing typhlocolitis and hepatitis in a foal. J Vet Diagn Invest. 2012;24:581–6.
132. Irwin DM, Biegel JM, Stewart C-B. Evolution of the mammalian lysozyme gene family. BMC Evol Biol. 2011;11:166.
133. Mastrangelo S, Tolone M, Sardina MT, Sottile G, Sutera AM, Di Gerlando R, et al. Genome-wide scan for runs of homozygosity identifies potential candidate genes associated with local adaptation in Valle Del Belice sheep. Genet Sel Evol. 2017;49:84.
134. Greenbaum MP, Yan W, Wu M-H, Lin Y-N, Agno JE, Sharma M, et al. TEX14 is essential for intercellular bridges and fertility in male mice. P Natl Acad Sci-Biol. 2006;103:4982–7.
135. Greenbaum MP, Iwamori N, Agno JE, Matzuk MM. Mouse TEX14 is required for embryonic germ cell intercellular bridges but not female fertility. Biol Reprod. 2009;80:449–57.
136. Baudat F, Manova K, Yuen JP, Jasin M, Keeney S. Chromosome synapsis defects and sexually dimorphic meiotic progression in mice lacking Spo11. Mol Cell. 2000;6:989–98.
137. Carrell D, De Jonge C, Lamb D. The genetics of male infertility: a field of study whose time is now. Arch Androl. 2006;52:269–74.
138. Zhang J, Zhou D-x, Wang H-x, Tian Z. An association study of SPO11 gene single nucleotide polymorphisms with idiopathic male infertility in Chinese Han population. J Assist Reprod Gen. 2011;28:731–6.
139. Ghalkhani E, Sheidai M, Gourabi H, Noormohammadi Z, Bakhtari N, Malekasgar AM. Study of single nucleotide polymorphism (rs28368082) in SPO11 gene and its association with male infertility. J Assist Reprod Gen. 2014;31:1205–10.
140. Nicolini P, Amorín R, Han Y, Peñagaricano F. Whole-genome scan reveals significant non-additive effects for sire conception rate in Holstein cattle. BMC Genet. 2018;19:14.
141. Sha Y-W, Xu X, Ji Z-Y, Mei L-B, Qiu P-P, Ji H, et al. Sperm-egg fusion disorder in a Chinese male patient was associated with a rare ADAM20 variant. Oncotarget. 2017;9:2086–91.
142. Zhu G-Z, Lin Y, Myles DG, Primakoff P. Identification of four novel ADAMs with potential roles in spermatogenesis and fertilization. Gene. 1999;234:227–37.
143. vanH RH. ADAM 20 and 21; two novel human testis-specific membrane metalloproteases with similarity to fertilin-α. Gene. 1998;206:273–82.
144. Li Y, Chen Y, Wu W, Li N, Hua J. MMPs, ADAMs and ADAMTSs are associated with mammalian sperm fate. Theriogenology. 2023;200:147–54.
145. Cezard T, Cunningham F, Hunt SE, Koylass B, Kumar N, Saunders G, Shen A, Silva AF, Tsukanov K, Venkataraman S, Flicek P, Parkinson H, Keane TM. The European variation archive: a FAIR resource of genomic variation for all species. Nucleic Acids Res. 2021;50:D1216–20.
146. Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, et al. The ARRIVE guidelines 2.0: updated guidelines for reporting animal research. PLOS Biol. 2020;18:e3000410.

Publisher's Note