UNIVERSITÀ DI PARMA

UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN SCIENZE MEDICO-VETERINARIE

CICLO XXXIII

Pedigree and genomic information for horse breeding and genetic diversity conservation

Coordinatore:

Chiar.mo Prof. Gaetano Donofrio

Tutore:

Chiar.mo Prof. Alberto Sabbioni

Dottoranda: Michela Ablondi

Anni Accademici 2017/2018 – 2019/2020

Acknowledgment

I would like to express my deepest gratitude to Prof. Alberto Sabbioni for his support during the whole research. He has been a careful and dedicated supervisor, eager to help and encourage me every time I needed. Thanks to Prof. Paola Superchi for her caring and generous attitude, who welcomed me in the Department from the very first day. Many thanks to Prof. Andrea Summer, Prof. Massimo Malacarne, Claudio and Christos who recently introduced me into the very interesting genomic cattle sector. A big thank goes to all the members at the Department of Veterinary Science, in particular thanks to my colleagues Giorgia, Elena and Maria. Thanks Giorgia to kindly support me every time I needed and thanks Elena and Maria for making the working environment friendly and pleased.

I am grateful to the Libro Genealogico del Cavallo Bardigiano for providing very interesting data during my PhD. A special thank goes to Matteo Vasini who has been always very optimistic and enthusiastic about our research ideas and shared with us very nice opportunities.

A great thanks goes to my SLU co-supervisors: Prof. Sofia Mikko and Prof. Susanne Eriksson for involving me in extremely interesting research projects. I would also like to thank Sofia and Susanne for all the pleasant time spent together, I felt homey in every journey I made to Sweden. I really wish to keep this wonderful collaboration in the future. Thank also to all my colleagues at SLU and to Emma from the SWB association, a special thanks goes to Åsa, Gabriella, Marina and Suvi.

A big thank goes to my former colleague Lim from Wageningen, how has been a special friend and colleague, with whom I shared equally intense working and chilling moments.

Finally, thanks to all my family for their support during these years, a special thank goes to my mum Anita, who has been encouraging and inspiring me and to my grandmother Elvira who has always been my biggest fan. Thanks to my niece Sofia who shares the horse passion with me and with I have spent very enjoyable horse-riding sections! Many thanks to my uncle Roberto and my aunt Elisabetta for being constantly interested and curious about my research and for sharing their experiences in the academia. Thanks to Francesca and Beppe for their kind presence and to my beloved Jacopo, who has always been by my side – part of this PhD is yours.

Abstract

The horse industry has faced two main challenges in the last decades: I) how to cope with the loss of genetic diversity in native breeds and II) how to use the advance in genomics in the sport horse breeding sector to meet the demand for more innovative selection strategies.

Horses are nowadays mainly used for sport and leisure activities, and several local breeds, traditionally used in agriculture, have suffered a significant decrease in population size and loss of genetic diversity. Loss of genetic diversity determines in turn a reduction in individual fitness and population viability, thus endangering such local horse breeds.

This PhD research aimed at preventing loss of genetic diversity in local horse breeds by proposing a study on genetic diversity in a native Italian horse breed (Challenge #1, The Bardigiano case study). The Bardigiano horse is a native Italian breed living in rural areas, traditionally used in agriculture. The relatively small size and the closed status of the breed raise the issue of monitoring genetic diversity. In this PhD, the genetic diversity of the Bardigiano horse breed was studied based on pedigree and molecular data.

Pedigree data contained 9,469 horses, of which 3,416 are alive. Demographic and genetic parameters were estimated on subpopulations to evaluate potential genetic diversity differences among breeding and nonbreeding animals, horses showing different breeding values (EBVs) and horses bred in different areas. Based on pedigree data, from the first studied birth year cohort (1975–1983) inbreeding steadily increased, reaching in the last birth year cohort a mean value equal to 0.10. The rate of inbreeding per generation, assuming a mean generation interval of 8.74 years, was equal to 1.64%. Moreover, both average relatedness and inbreeding among horses with high and low EBVs showed significant differences.

Genotype data for 89 Bardigiano horses has permitted to study loss of genetic variability and to detect potential signatures of selection based on GGP Equine70k SNP data. The effective population size based on Linkage Disequilibrium (Ne) was equal to 39 horses, and it showed a decline over time. The average inbreeding based on runs of homozygosity (ROH) was equal to 0.17 (SD = 0.03). The majority of the ROH were

relatively short (91% were 2 Mbp long), thus highlighting the occurrence of past bottlenecks. A total of eight ROH islands, shared among more than 70% of the Bardigiano horses, were found. Four of them mapped to known quantitative trait loci related to morphological traits (e.g., body size and coat colour) and disease susceptibility. The work done in this PhD provided first insights into genetic diversity and selection signatures in an Italian native horse breed. This study highlighted that breeding strategies for the coming generations are needed to ensure long-term survival of the Bardigiano breed.

In addition, this PhD program contributed to further understand the performance traits' genetic background in sport horses by means of genomic data (Challenge #2, the Swedish Warmblood breed case study). Therefore, thanks to the collaboration with Prof. Sofia Mikko and Prof. Susanne Eriksson at the Swedish University of Agricultural Sciences (Uppsala, Sweden), high density Single Nucleotide Polymorphism (SNP) data of Swedish Warmblood Horses were examined with several methodologies to detect genomic regions potentially involved in sport performance traits.

The motivation behind this second study is the growing demand for improved physical skills and mental attitude in modern sport horses, which has led to intensive selection for performance in many warmblood studbooks. The aim was to detect genomic regions with low diversity, and therefore potentially under selection, in 380 Swedish Warmblood horses (SWB), by analysing high-density SNP data. To investigate if such signatures were the result of selection for equestrian sport performance, SWB SNP data were compared to those from Exmoor ponies, a horse breed not selected for sport performance traits. In addition, since several modern SWB breeders have specialized breeding strategies either towards show jumping or dressage disciplines, this PhD research explored the genomic structure of SWB horses to evaluate the presence of genomic subpopulations, and to search for signatures of selection in subgroups of SWB with high or low breeding values (EBVs) for show jumping .

The genomic scan for homozygous regions identified long runs of homozygosity (ROH) shared by more than 85% of the genotyped SWB individuals. Such ROH were located on ECA4, ECA6, ECA7, ECA10 and ECA17. Long ROH were instead distributed evenly across the genome of Exmoor ponies in 77% of the chromosomes. Two population

differentiation tests (F_{st} and XP-EHH) revealed signatures of selection on ECA1, ECA4, and ECA6 in SWB horses. Overall, genes related to behaviour, physical abilities and fertility, appear to be targets of selection in the SWB breed.

To evaluate potential population stratification within the SWB, Principal Coordinates Analysis and Discriminant Analysis of Principal Components were performed. In addition, the F_{st} and XP-EHH were used to scan the genome for potential signatures of selection. In accordance with current breeding practice, this study highlighted the development of two separate breed subpopulations with putative signatures of selection in eleven chromosomes. Such regions involve genes with known function in, e.g., mentality, endogenous reward system, development of connective tissues and muscles, motor control, body growth and development. This study showed genetic divergence due to specialization towards different disciplines in SWB horses. This latter evidence can be of interest for SWB and other horse studbooks encountering specialized breeding.

This second study provided a genome-wide map of selection signatures in SWB horses and showed potential divergence based on disciplines, and ground for further functional studies to unravel the biological mechanisms behind complex traits in horses.

Riassunto

Tra le numerose sfide che l'industria equina è stata chiamata ad affrontare negli ultimi decenni, due importanti questioni sono tuttora oggetto di dibattito: I) come affrontare la perdita di diversità genetica nelle razze autoctone a limitata diffusione e II) come introdurre l'innovazione portata dalle nuove tecnologie molecolari nel settore dell'allevamento equino.

La maggior parte delle razze equine sono oggigiorno utilizzate per attività sportive e amatoriali e diverse razze autoctone, tradizionalmente impiegate in agricoltura, sono a rischio di riduzione della loro numerosità, con conseguente perdita di diversità genetica. La perdita di diversità genetica provoca una diminuzione del fitness a livello individuale, ma presenta anche un effetto estremamente negativo sulla capacità di sopravvivenza di una popolazione nel lungo termine.

Questa ricerca di dottorato ha avuto perciò lo scopo di monitorare la perdita di diversità genetica nelle razze locali equine, proponendo uno studio sulla diversità genetica e prendendo come caso studio una razza equina autoctona italiana (Sfida # 1, Caso studio: Cavallo Bardigiano). Il cavallo Bardigiano è una razza autoctona italiana allevata principalmente in contesti rurali, e tradizionalmente utilizzata in agricoltura. Le dimensioni relativamente ridotte di questa popolazione e la presenza di un libro genealogico chiuso evidenziano l'importanza del monitoraggio della diversità genetica. In questo dottorato è stata pertanto studiata la diversità genetica nel cavallo Bardigiano sulla base di dati sia genealogici che molecolari.

I dati genealogici analizzati contenevano 9469 cavalli, di cui 3416 vivi. In questa ricerca di dottorato sono stati stimati parametri demografici e genetici su diverse sottopopolazioni all'interno della razza per valutare potenziali differenze genetiche tra, ad esempio, animali da riproduzione e non da riproduzione, cavalli con meriti genetici differenti (EBVs) oppure allevati in aree geografiche diverse. Sulla base dei dati genealogici, la consanguineità è aumentata costantemente dalla prima generazione analizzata in questo studio (1975-1983), raggiungendo nella più recente generazione un valore medio pari a 0,10. Il tasso di consanguineità per generazione, ipotizzando un intervallo medio di generazione di 8,74 anni, è risultato uguale all'1,64%. Inoltre, sono

state evidenziate differenze significative sia sulla parentela media che sulla consanguineità tra i cavalli con EBVs alti e bassi.

Nel corso di questo dottorato, sono stati inoltre genotipizzati 89 cavalli Bardigiani con un pannello a media densità (70k SNPs), che hanno permesso di comprendere ulteriormente la perdita di variabilità genetica nella razza e individuare potenziali segni di selezione. La numerosità effettiva della popolazione basata sul Linkage Disequilibrium (Ne) è risultata pari a 39 cavalli e ha mostrato una diminuzione costante nel tempo. La consanguineità media basata sulle cosiddette "Runs of Homozygosity" (ROH), è stata pari a 0,17 (SD = 0,03). La maggior parte delle ROH presenta una lunghezza relativamente breve (il 91% aveva lunghezza pari o inferiore a 2 Mbp), evidenziando la presenza di "colli di bottiglia" avvenuti nella popolazione nelle scorse generazioni. Sono state trovate in totale otto "ROH islands", ovvero ROH condivise in oltre il 70% dei cavalli Bardigiani testati. Quattro di queste "ROH islands" sono localizzate in prossimità di QTL (quantitative trait loci), conosciuti per la loro importanza riguardo caratteri morfologici (ad es. altezza al garrese e colore del mantello) e suscettibilità ad alcune patologie. Il lavoro svolto durante questo dottorato ha approfondito la diversità genetica nel cavallo Bardigiano sia attraverso informazioni di pedigree che genomiche, evidenziando la necessità di implementazione di nuove strategie per il monitoraggio della diversità genetica finalizzate a garantire la sopravvivenza a lungo termine di questa razza.

Questa ricerca di dottorato ha inoltre contribuito a comprendere il background genetico dei caratteri legati alle performance nei cavalli sportivi mediante l'analisi di dati genomici (Sfida # 2, caso studio: Cavallo a Sangue Caldo Svedese). Grazie alla collaborazione con la Prof.ssa Sofia Mikko e la Prof.ssa Susanne Eriksson dell'Università Svedese di Scienze Agrarie (Uppsala, Svezia), sono stati analizzati dati genomici di cavalli Svedesi a Sangue Caldo (SWB) al fine di identificare regioni genomiche coinvolte in caratteri legati alle prestazioni sportive.

In questo studio sono state rilevate regioni genomiche con bassa variabilità, e quindi potenzialmente sotto selezione, in 380 cavalli SWB analizzando dati genomici ad alta densità (670k). Per verificare se tali regioni genomiche fossero il risultato della selezione per prestazioni sportive, i risultati ottenuti nei cavalli SWB sono stati confrontati con quelli ottenuti in una razza non selezionata per performance sportive (Exmoor pony).

Inoltre, poiché è sempre più comune avere allevamenti specializzati nei confronti di una specifica disciplina equestre specifica (p.es: salto ostacoli o dressage), questa ricerca di dottorato ha esplorato la struttura genomica dei cavalli SWB per valutare la presenza di eventuali sottopopolazioni e cercare segni di selezione in sottogruppi di cavalli SWB con meriti genetici alti e bassi (EBV) per il salto ostacoli.

L'analisi delle ROH ha identificato ROH con lunghezza superiore a 500 kbp, condivise in più dell'85% degli individui SWB genotipizzati. Tali ROH sono localizzate nei seguenti cromosomi: ECA4, ECA6, ECA7, ECA10 e ECA17. Al contrario, le ROH identificate nei pony Exmoor sono distribuite uniformemente nel genoma nel 77% dei cromosomi. Due test di genetica delle popolazioni (F_{st} e XP-EHH) hanno rilevato potenziali segni di selezione su ECA1, ECA4 e ECA6 nei cavalli SWB. Nel complesso, da questo studio è emerso che geni relativi al comportamento, a capacità fisiche e alla fertilità, sembrano essere obiettivi di selezione nella razza SWB.

Per valutare la potenziale stratificazione e la presenza di eventuali sottogruppi legati alla tendenza alla specializzazione per una sola disciplina sportiva all'interno della popolazione SWB, sono state eseguite due analisi: PcoA e DAPC. Inoltre, F_{st} e XP-EHH sono stati usati per verificare la presenza di potenziali segni di selezione in due sottogruppi di cavalli creati sulla base degli EBV per la disciplina salto ostacoli. In accordo con l'attuale pratica di allevamento nel cavallo svedese, questo studio ha evidenziato la presenza di due sottopopolazioni con potenziali segni di selezione differenti in undici cromosomi. Queste regioni coinvolgono geni con funzioni legate al sistema endogeno di ricompensa, sviluppo dell'apparato muscolo scheletrico e tessuto connettivo e controllo motorio.

Questo studio ha fornito una prima mappa genomica dei segni di selezione nei cavalli SWB e ha presentato una potenziale divergenza nella popolazione causata dalla specializzazione in cavalli da salto ostacoli e da dressage. È possibile considerare questo studio il punto di partenza per ulteriori studi funzionali che contribuiranno a comprendere meglio i meccanismi biologici alla base della complessa natura dei caratteri di performance nei cavalli sportivi.

Contents

List of Figures

List of Tables

List of Abbreviations

ADCY1	Adenylate Cyclase 1
AI	Artificial insemination
AR	Average relationship
ARHGAP31	Rho GTPase activating protein 31
ARHGAP40	Rho GTPase activating protein 40
ARHGDIB	Rho GDP Dissociation Inhibitor Beta
ASIP	Peptide antagonist agouti-signalling-protein
B3GAT1	Beta-1,3-glucuronyltransferase 1
BIC	Bayesian Information Criterion
BLUP	Best linear unbiased prediction
BP	Breeding population
CACNA1A	Calcium voltage-gated channel subunit alpha1 A
CD80	CD80 molecular
CGE	Complete generation equivalent
cM	Centimorgan
CNV	Copy number variation
COL15A1	Collagen type XV alpha 1 chain
DAPC	Discriminant analysis of principal component
DBX1	Developing brain homeobox 1
DLGAP4	Disc large homolog associated protein 4
DMRT3	Doublesex and mab-3 related transcription factors
DNASE2	Deoxy-ribonuclease 2, lysosomal
EBV	Estimated breeding value
ECA	Equus Caballus autosome
ER	Endoplasmatic reticulum
F	Inbreeding coefficient of an individual
f_a	Effective number of ancestors
FAO	Food and Agriculture Organization
FDR	False discovery rate
f_e	Effective number of founders
f_{ge}	Founder genome equivalent
FIGNL1	Fidgetin Like 1
F_{is}	Inbreeding coefficient of an individual (I) relative to the subpopulation (S)
F_{it}	Inbreeding coefficient of an individual (I) relative to the total (T) population
FKBP14	FK506–binding protein 14
F_{ROH}	Genomic inbreeding based on ROH
F_{st}	Fixation index
GENO	Missing genotype per single SNP
GI	Generation Interval

GO	Gene Ontology
GPCRs	G-protein coupled receptors
GRIN2B	Glutamate Ionotropic Receptor NMDA Type Subunit 2B gene
GS	Glycogen synthase
GSK3B	Glycogen synthase kinase 3 beta
GUCY2C	Guanylate Cyclase 2C
GWAS	Genome Wide Association Study
GYS1	Gene glycogen synthase 1
HDAC9	Histone deacetylase 9
HWE	Hardy–Weinberg equilibrium
IBH	Insect bite hypersensitivity
IGF1	Insulin-like growth factor
IGFBP1	Insulin like growth factor binding protein 1
IGFBP3	Insulin like growth factor binding protein 3
IP	Inferior EBVs horses
kEDS	Kyphoscoliotic form of Ehlers–Danlos syndrome
LASP1	LIM and SH3 protein 1
LCORL	Ligand dependent nuclear receptor corepressor-like
LD	Linkage disequilibrium
lncRNA	Long non-coding RNAs
L_{ROH}	Average length of ROH
LRRC7	Leucin-rich-repeat-containing 7
MAF	Minor allele frequency
MC1R	Melanocortin receptor
MDS	Classical multidimensional scaling
miRNA	MicroRNA
miscRNA	Miscellaneous RNA
MN1	Meningioma 1
MYL9	Myosin light chain
NAV2	Neuron navigator 2
NCAPG	Non-SMC Condensin I Complex Subunit G
Ne	Effective population size
Ne_S	Ne slope analysis
NME7	NME/NM23 family member 7
NR1I2	Nuclear receptor subfamily 1 group I member 2
N_{ROH}	Total number of ROH
NS	Non-show jumping horses
OCS	Optimal contribution selection
OPRM1	Opioid receptor mu 1
P	P-value
PAK5	Activated kinase 5
PAPSS2	30-phosphoadenisine-50-phosphosulphate synthase 2
PCA	Principal component analysis

PCoA	Principal coordinates analysis
PCs	Principal components
PITPNB	Phosphatidylinositol transfer protein
PLCB1	Phospholipase C beta 1
PLOD1	Procollagen lysine, 2-oxoglutarate 5-dioxygenase 1
POPDC2	Popeye domain containing 2
PRKG1	Cyclic GMP–dependent protein kinase 1
PSD	Excitatory postsynaptic density
PSSM	Muscle disease polysaccharide storage myopathy
PTEN	Phosphatase and tensin homologue
PTMA	Prothymosin alpha
QC	Quality Control
QTL	Quantitative trait loci
RAC1	Rac Family Small GTPase 1
RBL1	Retinoblastoma 1
RERG	RAS Like Estrogen Regulated Growth Inhibitor
RGS17	Regulator of G protein signalling 17
RIA	Rate of inbred animals
ROH	Runs of homozygosity
RP	Reference population
SAMHD1	SAM and HD domain containing deoxynucleoside triphosphate triphosphohydrolase 1
SD	Standard deviation
SEC61B	Sec61 translocon beta subunit
SGMS1	Sphingomyelin synthase 1
SJ	Show jumping
snoRNA	Small nucleolar RNAs
SNP	Single-nucleotide polymorphism
SP	Superior EBVs horses
SPATA48	Spermatogenesis associated 48
SRC	SRC Proto–Oncogene/Non–Receptor Tyrosine Kinase
S_{ROH}	Average ROH number
STP	Selected tested population
SWB	Swedish Warmblood breed
TGF-beta	Transforming growth factor beta
TGFBR1	Transforming growth factor beta receptor 1
THSD7A	Thrombospondin, type I
TOP1	DNA Topoisomerase 1
TP	Total population
TTP	Total tested population
VSTM2L	V-set and transmembrane domain containing 2 like
WFFS	Warmblood Fragile Foal Syndrome
XP-EHH	Cross populations extended haplotype homozygosity
ZPBP	Zona pellucida binding protein
ΔF	Rate of increase in inbreeding

Chapter 1

1.1 Context

1.1.1 Horse industry worldwide

The horse family (Equidae) occupy a unique place in human history and civilisation. Horses have populated Earth for more than 50 million years, with fossils evidencing their origin in North America and the subsequent migrations towards Asia and Europe. The evolutionary lineage of the horse is one of the best documented in palaeontology. The history of Equidae began during the Eocene Epoch, with the presence of a hoofed, browsing mammal officially defined as “Hyracotherium” but commonly known as “Eohippus” [1]. Fossils of Eohippus, found in both North America and Europe, demonstrate how small-sized (42.7 to 50.8 cm tall) these animals were, if compared to modern horses. The line leading from Eohippus to the modern horse exhibits some major evolutionary trends: increased size, longer legs, fusion of independent bones, increased brain complexity, and teeth more suited for grazing (Figure 1).

The above-mentioned evolutionary changes did not progress steadily, with a gradual transition from Eohippus to modern horse. Some of the features, such as grazing dentition, appeared abruptly during the fossil record, rather than culminating after numerous, gradual changes [1]. Horse domestication most probably started in the Kazakh steppe during the Botai culture around 5,500 years ago [2]. Since then, horses have uninterruptedly accompanied any change in human society. Over the centuries, several cultures in Eurasia and Africa relied on horses for transportation, work, and military purposes. Thanks to horse riding, humans started to travel at a faster speed and to connect across vaster territories, thus revolutionizing warfare with chariotry and cavalry [3].

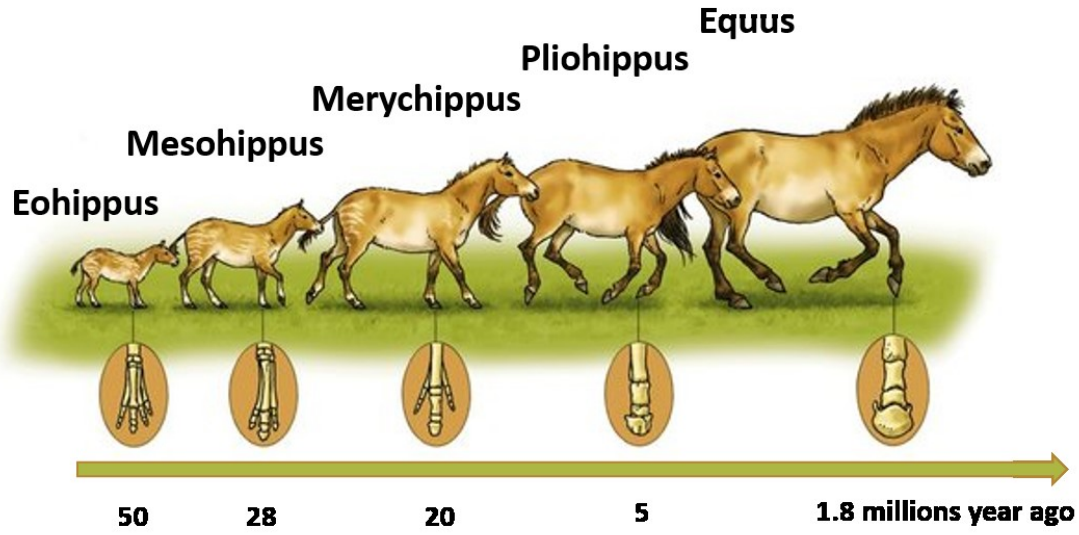

Figure 1 Evolution of the horse occurred over a geological time scale of 50 million years, from the small forest-dwelling Eohippus to the modern horse (Equus).

Without horses, the history of humanity would have been probably much different [4]. Domestication marked the first series of genetic improvements, which determined over 900 breeds present nowadays around the world [5]. In modern European breeds, three main introgression occurred: the Neapolitan wave (15th to 18th century), the Oriental wave (late 18th to late 19th century) where Arabian stallions were imported, and the English wave (early 19th century to the present) with the introgression of Thoroughbreds [6, 7].

Thanks to their versatile attitudes combined with power, agility, gracefulness and speed, horses are nowadays mostly bred for leisure and sport activities. To date, ten countries in the world present a horse population of more than a million individuals, with a global estimated equine population of 58 million, according to the Food and Agriculture Organization (FAO) of the United Nations [5]. The European Horse Network (EHN) estimates an overall impact of the horse industry in Europe of over 100 billion euros a year and similar figures have been calculated for the US [8]. An emerging trend has characterised the horse industry in the last decades: breed specialisation towards sport performance. The sector has experienced a growing interest to select horses for sport competitions, while guaranteeing elevated health standards. To date, roughly 18% of the horse population is composed of warmblood breeds mainly used for riding and sport discipline [9], where cross breeding and open studbook policies are common

practiced. On the other hand, over 80% of horse breeds are local, with tight connections with the territory of origin, and usually under a close studbook policy. Unfortunately, several living heritage breeds are threatened by the risk of extinction due to loss of original purpose and/or rivalry with specialized breeds. The second report on the state of the world's animal genetic resources published in 2015 stated that the risk status is unknown for 52% of the 905 recognised horse breeds, and 40% of the remainder is either at critical or endangered status [5].

1.1.2 Challenge #1: loss of genetic diversity in autochthonous breeds

Throughout history several evolutionary forces such as mutation, selective breeding, adaptation, isolation, bottlenecks and genetic drift have determined the formation of a broad variety of local populations. In the last centuries, this evidence culminated in the formation of well-defined breeds used for a vast range of purposes. Recently, due to the current demand for highly specialised breeds, the development of effective selection programs has accelerated genetic improvement. Meanwhile, artificial insemination and embryo transfer have facilitated the dissemination of genetic material in several breeds. Furthermore, progress in feed technology has allowed optimal nutrition, and enhanced transportation and communication systems have created uniform and strictly controlled production chains. As a result, a few highly productive and specialised breeds have replaced local ones across the world [10]. This growing global trend of using a limited number of highly specialised breeds is one of the most significant causes of genetic resources erosion [11]. According to FAO, roughly 17% of 8,800 breeds reported worldwide are at risk of extinction and nearly 100 breeds became extinct within the first 14 years of this century [5]. In the equine context, a decrease in population size for several breeds has been observed as a result of the above-mentioned changes, thus frightening their genetic variability [10].

Genetic variation (aka polymorphism) contributes to the populations' evolutionary potential, thus causing medium- to long-term implications in the proper management and future conservation of endangered species. Genetic diversity guarantees the capability to withstand environmental changes in the long term, while enabling breeds to respond to selection [10]. Particularly small populations are more prone to the

mating of related individuals, which results in inbred animals. Inbreeding causes a reduction in heterozygosity at population level, which might induce inability to adapt to changes in the coming generations. Unfortunately, the loss of genetic diversity determines a detrimental effect at individual level too. It has been shown that individuals with high homozygosity present lower fitness, if compared to non-inbred individuals [12]. Reduction in fitness due to inbreeding is referred to as "inbreeding depression" and it is defined as the reduction in fitness due to increased genome-wide homozygosity in offspring born from related parents [13]. It has been shown that inbreeding lowers fitness-related traits in several species. Inbreeding depression threatens inbred populations through loss of viability, fertility, disease resistance, and increased frequency of recessive monogenic diseases. Major abnormalities are more frequent in inbred animals than in outcrosses. These abnormalities include mutant phenotypes that are lethal early in the life, developmental complications and monogenic diseases. Inbreeding also results in a higher probability of fixation of deleterious alleles and loss of beneficial alleles which might in turn affect important economical traits [10]. To break this vicious circle, it is vital to study the current genetic variability in small populations and to develop tools and strategies to preserve them in the long term [12].

1.1.3 Challenge #2: scarce knowledge of the performance traits' genetic background in sport horses

In the last decade, the revolution in molecular genetics and bioinformatics have enabled the integration of genomic information in several animal breeding applications. Especially in dairy cattle, conventional genetic evaluations shifted to genomic evaluations, giving birth to a "new era of breeding" [14]. In dairy cattle, the concept of genomic selection (GS) was rapidly adopted and it radically changed breeding schemes and structures [15]. Today, GS is successfully used in all major dairy cattle populations worldwide. The reason behind this evidence can be motivated by the fact that GS improves the accuracy of early candidate selection and it substantially reduces the generation interval [16]. Thanks to the availability of genotype data, new approaches for the detection and management of genetic defects, and improved selection strategies that consider inbreeding were recently developed [17]. Although there is a shared consensus across sector experts that GS presents a promising potential in horses, significantly few genomic data are currently available and used in horses compared to dairy cattle [18, 19].

In the last decades, the horse breeding sector has experienced a growing interest to select horses for sport competitions, while guaranteeing elevated health standards. For this reason, breeding associations have activated intensive phenotypic recording systems and extensive research has been performed to estimate genetic parameters and breeding values for sport traits [20–22]. Given the role of high-quality phenotypes, recent initiatives improved objectivity of the assessments and the trait definitions via the use of linear recording system [23–25].

Several performance-related traits have been recognised as complex and polygenic traits. Complex traits are not determined by one or a few genomic variants but rather by several small contributions of variants across the genome. These traits generally present low-to-medium heritability, as the phenotypic outcome is determined by both the underlying genetic variation and interaction, and environmental factors (such as nutrition, training, farming system). The role of the environment complicates both the identification and the understanding of the genetic components driving those traits

[26]. Nevertheless, the majority of complex traits are economically relevant for the horse industry. Complex traits include measures of athletic performance such as jumping ability and gaits, but also welfare related traits, e.g. laminitis, equine asthma or recurrent airway obstruction and osteochondrosis [26]. To date, a comprehensive insight on the genetic background for athletic performance traits (e.g. the genomic regions/genes involved in specific horse's features) is still missing, to the best of the author's knowledge. Identifying such regions can increase our understanding of the biological mechanisms which underly those complex traits, thus propelling the use of genomic in horse breeding [18].

The knowledge of the genetic background behind sport traits could substantially improve the selection process. Given the long generation interval in horses, the ability to provide information as early as possible might be a key improvement. Recent advances in genome mapping provide great opportunities to exploit genetic markers or specific genes in the equine context as well.

1.2 Structure of the thesis

The present thesis is built upon the work that the PhD student has carried to analyse the two above-mentioned challenges as shown in Figure 2. To prevent loss of genetic diversity in local horse breed, an integrated approach applied to a local Italian horse breed is proposed and presented in Chapter 2. To fill the gap in the performance traits' genetic background in sport horses, high density Single Nucleotide Polymorphism (SNP) data of Swedish Warmblood Horses were examined with several methodologies and presented in Chapter 3. This latter chapter relied on the work the PhD student has done in collaboration with Prof. Sofia Mikko and Prof. Susanne Eriksson at the Swedish University of Agricultural Sciences (Uppsala, Sweden), as part of the internationalisation period for the Doctor Europaeus. Chapter 4 summarises the main outcomes from the preceding chapters and provides suggestions for future work.

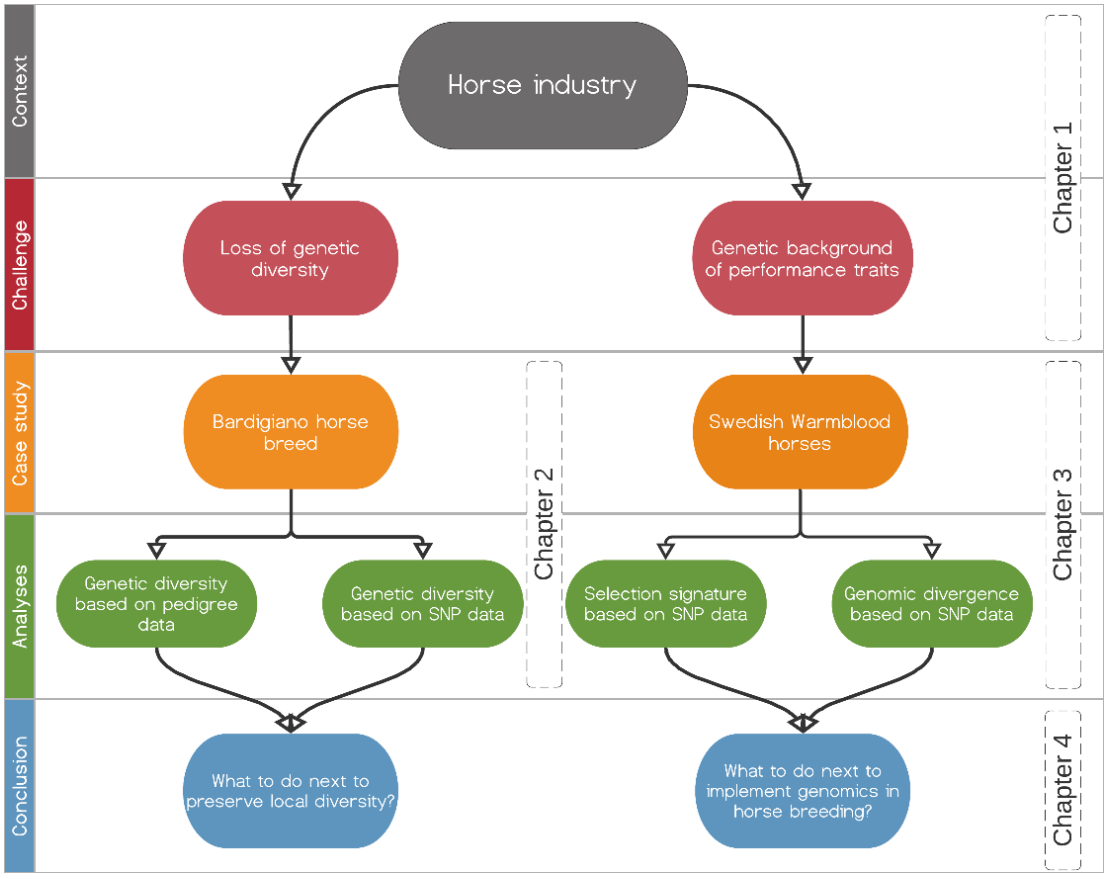

Figure 2 Visual flow chart of the PhD thesis.

Chapter 2

A combined study to comprehend the genetic architecture of an Italian local horse breed

This chapter is based on two published research studies:

1) Exploring genetic diversity in an Italian horse native breed to develop strategies for preservation and management by **Michela Ablondi[1]**, Matteo Vasini[2], Valentino Beretti[1], Paola Superchi[1], Alberto Sabbioni[1]. Published in the Journal of Animal Breeding and Genetics, J Anim Breed Genet. 2018; 135:450–459
2) Genetic diversity and signatures of selection in a native Italian horse breed based on SNP data by **Michela Ablondi[1]**, Christos Dadousis[1], Matteo Vasini[2], Susanne Eriksson[3], Sofia Mikko[3], Alberto Sabbioni[1]. Published in Animals, Animals 2020;10,1005

[1] Dipartimento di Scienze Medico-Veterinarie, Università di Parma, Parma, Italy

[2] Libro Genealogico Cavallo Bardigiano, Associazione Regionale Allevatori dell'Emilia-Romagna, Parma, Italy

[3] Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

2.1 INTRODUCTION

2.1.1 Introduction to the native breed

The Bardigiano is an Italian native horse breed with an old history. The origins of the breed can be traced back to the age of the Roman Empire in the Belgian Gaul province [27]. In 1864, the breed was already described as a homogeneous type reared from the Western Apennines to the Lunigiana valley. This equine breed is named after the area of origin: the town of Bardi, in the Parma Apennines. In the 19th century, Bardigiano horses covered a central role in human society, as they were used for both transportation in agriculture and meat production [28]. After the Second World War, the number of Bardigiano horses decreased dramatically down to just five stallions and 150 mares [29]. To protect the Bardigiano breed from extinction, the Bardigiano studbook was founded in 1977, and since then the Bardigiano is considered a purebred due to the closure policy applied by the studbook. The first action taken by the studbook was to register all horses raised within the autochthonous area of breeding showing typical Bardigiano breed features.

The distinctive morphological characteristic of this breed are: (a) meso-brachymorphic horse-type with a height ranging between 140 and 149 cm for males, and between 135 and 147 cm for females; (b) the only admitted coat colour is bay, with a preference for dark bay, while chestnuts and extremely light bays are not allowed; (c) limited white markings on the legs and face are allowed, although not preferred. An example of such breed is shown in Figure 3. More conformation related features include a light head, with a slightly concave back-nose profile, large, sparkling and expressive eyes, usually covered by a thick loose mane. Strong neckline with thick mane, low-medium withers, wide rump with well-developed transverse diameters are also representative traits in the Bardigiano horses. The robust skeleton, short pastoral and very hard hoof, make those horses especially suitable for living in rough terrain [28].

Figure 3 The Bardigiano horse Egea II, 24 years old (owned by the author). Bardigiano horses are often used in traditional exhibitions thanks to their quiet temperament, as shown on the right.

Thanks to their temperament, frugality and resistance, Bardigiano horses nicely adapt to several environmental conditions and uses, type of pasture and wild or semi-wild breeding system. This is the reason why they were bred in several areas in the North of Italy across extreme heterogeneous conditions. Only a few breeders are present outside Italy, such as in Germany, Switzerland and Hungary. Nowadays, the Bardigiano breed counts roughly 3000 live horses, mainly used for riding and light draft purposes, thanks to both their docile temperament and rustic attitude.

The breeding objective of the Bardigiano studbook is divided into three main goals: (a) ensure long term survival of the breed; (b) preserve Bardigiano's distinctive morphological features; and (c) increase the use in riding and draft activities. To preserve Bardigiano's distinctive characteristics, in 1987 (goal b), the Bardigiano studbook introduced an assessment to evaluate which animals could be officially registered as Bardigiano horses. This evaluation was based on both conformation measurements and subjective grading traits according to the breeding objective. To accomplish goal (c), in the '90s the Bardigiano studbook defined breeding values for

height at withers to help the conversion in taller horses more suitable for riding. In addition, to increase the objectivity and quality of phenotype recording, linear profiling was added to the traditional evaluation [23]. Therefore, despite actions were taken to achieve goals b) and c), little was accomplished to evaluate genetic diversity to ensure long-term survival of the breed (goal a). Due to the relatively small size and the closed status of the Bardigiano breed, avoiding the permanent loss of genetic diversity is a key matter of concern.

2.1.2 State of the art

Genetic diversity plays a key role in both adaptation and response to selection. The loss of genetic diversity causes negative effect at both individual and population levels. Inbreeding has harmful effects on individuals reducing their fitness and potential to survive. Inbreeding can reduce individual health and fertility [30, 31], and it might have a detrimental effect on the population lifespan in the long term. This is because inbreeding reduces genetic variability, thus creating a population more vulnerable and less adaptable to changes. Bottlenecks, limited introgression and limited number of individuals, can radically affect the rate of genetic diversity. To prevent the negative effects of the loss of genetic diversity, the FAO sets a maximum inbreeding rate of 1% per generation [11].

To monitor genetic diversity in a breed two main categories of data can be used: pedigree and genomic data. Inbreeding coefficients and average relationships (AR) over time have been used to evaluate how genetic diversity evolved during breed formation [30]. Demographic and genetic parameters, based on population structure by origin and based on founders and ancestors' contributions, have also been used to monitor the changes in a population over a short period of time [32]. These principles have been applied to characterize the genetic diversity in several horse breeds, including both sport and isolated breeds [33–42]. Those estimates can be in turn used to optimize breeding strategies to avoid loss of genetic diversity and enhance breeding programs. The above-mentioned analyses can be performed based both on pedigree and genomic data yet achieving different level of precision. Analyses at pedigree level are considered effective tools to evaluate the state of genetic diversity, especially in the context of small

and under-development populations. This latter aspect is related to the often-limited economic resources of those populations, causing the unavailability of more advanced technologies as SNP chip data. Exploring genetic diversity at pedigree level does not need extra economic resources, as from available data, we are able to identify indicators of genetic variability. However, it must be noticed that pedigree recording quality might be a limiting factor [43]. The latter aspect is especially important as the accuracy of inbreeding estimates relies on i) the correctness of pedigree data and ii) the pedigree depth, in other words the number of known generations, with higher the better.

Recent techniques in genomics permit researchers to gain more insights into genetic diversity and population history compared to pedigree-based methods [17, 44]. Several genome-wide population structure and genetic diversity studies have been performed in livestock species [45–48]. The first comprehensive insight into equine genetic diversity among a large breed cohort was published by Petersen et al. in 2013 [49]. This study used, among other methods, the fixation index (F_{st}) statistics, one of the most popular methods to capture between-breed divergence [46, 50–52]. Runs of homozygosity (ROH) are likewise a well-established method, widely used to detect within-breed loss of genetic diversity [53]. The ROHs are long consecutive homozygous segments distributed across the genome. Among other evolutionary forces, they arise by identical-by-descendent haplotypes which came from common ancestors [54]. Therefore, ROHs have been used as a valuable source of information to estimate genomic inbreeding (F_{ROH}) [55–58].

Compared to pedigree-based inbreeding, F_{ROH} is able to capture variation due to Mendelian sampling and linkage during gamete formation [17]. In addition, F_{ROH} does not rely on pedigree quality, which, as stated above, might be a limiting factor in inbreeding estimation based on genealogical data [59–63].

Overlapping homozygous regions, highly shared among individuals belonging to the same population, are called ROH islands. Since directional artificial selection reduces genomic variability, ROH islands are thought to be potential signs of selection around a target locus [53, 64]. Several recent examples of ROH and population structure analyses applied to European horse breeds show key aspects of history and selection pressure [59–63, 65]. However, few studies were conducted in the framework of European small

native horse breeds [62, 65–67] and, to the best of the PhD candidate's knowledge, none of them specifically analysed Italian autochthonous horse breeds.

2.1.3 Research questions

This study aims to investigate the genetic diversity of the Bardigiano breed by combining analyses based on both pedigree and SNP chip data. The former type of analysis was used to (a) assess past and current trends in the genetic structure; (b) check the up-to-date genetic variability based on pedigree and (c) provide a starting point to optimize the breeding program. Similarly, the analyses based on SNP chip data provided further insights on the genomic diversity in the Bardigiano breed as an example of reservoir for Italian native breeds. The main objectives based on SNP data were to d) estimate the effective population size, e) use ROH to estimate inbreeding coefficient based on genomic data, f) use ROH as signature of selection.

2.2 MATERIALS AND METHODS

2.2.1 Pedigree-based materials and methods

The ENDOG v4.8 software and the SAS 9.4 were used to carry out the below mentioned analyses based on pedigree data [68, 69].

2.2.1.1 Genealogical data description

The data were provided by the breeding association of the Bardigiano horse breed (Libro Genealogico Cavallo Bardigiano, Associazione Regionale Allevatori dell'Emilia-Romagna, Parma, Italy). Quality control (QC) of the data was performed on inconsistency of dam and sire registration number, birth date and sex identification. After QC of the data, the herd-book contained 9,469 horses (TP) 3,416 of which were currently alive: 2,575 females (75.38%) and 841 males (24.62%).

2.2.1.2 Evaluated reference populations

Currently alive horses were defined as the reference population (RP), and their pedigrees were traced back to the earliest recorded ancestors. The first ancestor recorded was traced back in 1939, whereas the youngest horses included in the study were born in 2017. In addition to the RP, breeding mares and stallions were distinguished from the RP to analyse the current breeding population (BP). To build the BP, two filters on age and number of offspring were applied: in the case of mares, we included only female horses younger than 20 years old with at least one offspring; in the case of stallions, we included only male horses younger than 20 years old with at least five offspring. The minimum number of offspring was designed by considering the distribution of progeny in the two genders separately. As differences in the selection strategies are expected between genetically superior and inferior animals, thus causing differences in their genetic variability, two additional RPs considering breeding values (EBVs) were defined. The EBVs are indexes combining 18 conformation traits. Percentile classes of EBVs were calculated from the 3,416 currently alive horses, and the 1st

percentile class and the 4th percentile class were considered as the two RPs for inferior (IP) and superior (SP) horses, respectively (Table 1).

Table 1 Description of the data available in the entire pedigree database (TP), in the RP, in the BP, in the SP and in the IP.

Parameters	TP	RP	BP	SP	IP
Number of horses	9,469	3,416	742	505	495
Number of males	2,113	841	66	119	35
Number of females	7,356	2,575	677	386	460
Number of horses with no progeny	5,928	2,252	0	236	215
Number of founder horses	1,698	648	35	14	61
Number of mating between full sibs	13	6	0	0	0
Number of mating between half sibs	259	103	16	8	16
Number of mating parent-offspring	162	51	10	2	8

TP = Total population; RP = Reference population; BP= Breeding population; SP = Superior population; IP = Inferior population.

2.2.1.3 Population structure based on demographic analyses

To characterize the Bardigiano population structure, several demographic parameters were calculated: number of registered horses, number of sires and dams per birth year and the total number of offspring per sire and dam. The geographical area of origin was known for 7,251 horses; thus, the population structure by origin was also considered. The term geographical area of origin is referred to the province of origin, which is the Italian administrative division between municipality and region. The F-statistics, F_{st}, F_{is} and F_{it}, were used to assess whether subpopulations' genetic differences were present between and within areas [70]. The following parameters were calculated within geographical area of breeding: number of individuals born, percentage of individuals with sire born in the same area, and percentage of individuals with sire born in a different one. The geographical areas of breeding were further classified by considering the altitude, to assess whether Bardigiano horses have been more present in mountains, hills or lowlands. The classification in altitude type was defined by following the guidelines from the Italian National Statistical Institute [71], where mountains are defined as lands higher than 700 m , hills from 300 to 700 m and lowlands lower than 300 m above sea level, respectively. To classify the altitude type within each province of breeding, the municipality of origin was used. Thus, per each province, three classes

of altitude were calculated. The completeness of pedigree information was investigated by using the equivalent complete generation (CGE), which is computed as $\sum (1/2)^n$, where n is the number of generations between individuals and each known ancestor [32]. The maximum number of fully traced generations was also considered for each individual. Generation intervals (GI) were calculated along the following pathways: sire to son, sire to daughter, dam to son and dam to daughter by using the average age of parents at the birth of their offspring.

2.2.1.4 Founder and ancestor evaluation

Both the effective number of founders (f_e) and the effective number of ancestors (f_a) were calculated as the minimum number of founders and ancestors explaining the observed genetic diversity in the RP [72]. To assess whether the population experienced bottlenecks, the ratio between f_e and f_a was calculated: when the ratio is close to unity, the population is stable in terms of the numbers of effectively contributing animals. In contrast, if the effective number of founders is larger than the effective number of ancestors, this means that some ancestors have played a major role in population formation [73]. The founder genome equivalent (f_{ge}), which is defined as the effective number of founders with non-random loss of founder alleles describing the observed genetic variability, was calculated in the two population clusters concerning EBVs percentile class (SP and IP) and compared to the result obtained from the RP.

2.2.1.5 Inbreeding and average relationship estimates

The individual inbreeding coefficient (*F*), defined as the probability of an individual to present two identical alleles by descendant, was computed by following Meuwissen and Luo, 1992 [74]. The rate of inbred Bardigiano horses within generation was calculated: if a horse showed an inbreeding coefficient higher than 0.05, then it was considered as inbred. The threshold of *F* ($F > 0.05$) was chosen by considering the distribution of the individual inbreeding coefficients in the RP. The AR defines the mean relationship of each individual with the remainder of the population. The AR can thus be interpreted as the contribution of the animal towards the whole pedigree. The trends of F and AR coefficients throughout the analysed generations were evaluated with the GLM

procedure in SAS [69]. The rate of increase in inbreeding (ΔF) was also calculated per generation, along with the AR. Average *F* and AR were calculated per each breeding province divided by altitude type (lowlands, hills and mountains).

2.2.1.6 Effective population size based on pedigree data

The effective population size (Ne) was estimated as follows: (a) computing the regression coefficient of the individual inbreeding coefficient in the TP over (i) the number of full generations traced; (ii) the maximum number of generations traced; and (iii) the CGEs, (b) as individual increase in inbreeding in the population for TP, RP and BP.

2.2.1.7 Degree of non-random mating in the breed

The degree of non-random mating practiced by breeders was assessed by comparing F and AR among sires and dams belonging to the BP with the results obtained from the RP. Differences in the number of sires building the superior and inferior populations (SP and IP), the inbreeding coefficient and the AR coefficients between those two subpopulations were tested by two sample t tests in SAS 9.4 [69].

2.2.2 Genomic-based materials and methods

2.2.2.1 Sample collection

The study also included genotype data from 89 Bardigiano horses (38 males and 51 females). The horses submitted to genotyping were carefully selected to represent most of the genetic variability from the pedigree information. The following criteria were applied in the selection procedure: (i) verification of pedigree depth, excluding animals with an equivalent complete generation (defined as $\sum(1/2)^n$, where n is the number of generations separating the individual from each of its known ancestors [32] lower than two; (ii) horses with available conformation measurements and linear and traditional evaluations were preferred [75]; (iii) no more than three animals descending

from the same stallion were allowed; (iv) animals had to belong to the last two generations, while considering an average generation interval of 8.74 years.

2.2.2.2 Genotyping

All samples were genotyped with the GGP Equine70k® (Illumina, San Diego, CA, USA) SNP chip, containing 65,157 SNPs separated on average by 40 kb.

2.2.2.3 Quality control of genotype data

The SNP physical positions were remapped from the former reference genome EquCab2 to EquCab3 [76] as described in [77]. Only SNPs located on the 31 Equus caballus autosome (ECA) chromosomes were retrieved and used in this study. The quality control (QC) was performed in PLINK v1.9 [78] by removing SNPs with a call rate lower than 0.90, a Hardy–Weinberg equilibrium (HWE) deviation with $p < 10^{-6}$ and minor allele frequencies (MAF) <0.01. Horses with a call rate lower than 0.90 were excluded.

2.2.2.4 Effective population size based on genotype data

Past and present effective population sizes (Ne) were estimated with the SNeP v.1.1 software [79] by using only horses in the last birth year cohort (2011–2019). The SNeP v.1.1 estimates the trends of the historical effective population size trajectories from SNP data based on linkage disequilibrium (LD). Recombination rate was calculated using the Sved and Feldman's mutation rate modifier [80], and sample size correction was applied for unphased genotypes. The cM unit in this study was set to 1.24 Mbp [81]. The minimum and maximum distance between pairs of SNPs was set to 0.5 Mbp and 26 Mbp, respectively. We also performed a Ne slope analysis (Ne_S) [82] to explore the rate and directionality of Ne changes occurring throughout generations. The Ne_S analysis identifies subtle changes in the inferred Ne curve which are normally not visible in the Ne plot. The slope of each segment, linking pairs of neighbouring Ne estimates, was first calculated and then normalized using the median of the two most proximal past Ne slope values.

2.2.2.5 Inbreeding calculation based on homozygosity

The ROH segments were detected via the *DetectRUNS* [83] package in R [84], and defined as follows: (i) at least 15 SNPs in a run, (ii) a minimum length of a run equal to 500 kb, (iii) a maximum distance between consecutive SNPs in a window 1000 kb, (iv) a lower density limit of 1 SNP per 100 kb [60] and (v) by allowing for a maximum of one missing and one heterozygous SNP in a run. The ROH segments were divided into the following five length classes: 0.5–1 Mbp, 1–2 Mbp, 2–4 Mbp, 4–8 Mbp and >8 Mbp. The total number of ROHs (N_{ROH}), average length of ROHs (L_{ROH}) and the average ROHs number (S_{ROH}) were reported according to each ROH length category. The genomic inbreeding coefficients (F_{ROH}) were calculated following the method described in [55]:

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

with L_{ROH} being the length of ROHs in each individual and L_{AUTO} the length of the autosomal genome (set to 2276 Gbp) based on the genome length covered by SNPs. Based on the hypothesis that the length of ROH reflects the chronological time points when inbreeding happened [57], genomic inbreeding was expressed for five separate ROH length categories (0.5–1 Mbp, 1–2 Mbp, 2–4 Mbp, 4–8 Mbp, >8 Mbp). By using the formula 1/2 g Morgan, with g being equal to generation, and the relationship between Centimorgan and Mb in horses [81], we estimated the time point of the inbreeding event based on ROH length. The inbreeding was also calculated per chromosome, and the type of distribution was evaluated based on skewness and kurtosis.

2.2.2.6 Candidate regions under selection based on runs of homozygosity

Putative ROH islands were determined based on overlapping homozygous regions within more than 70% of the horses. The EqCab3 genomic coordinates of these regions were used to retrieve candidate gene lists and annotations from the Biomart web interface in Ensembl [85] and from the UCSC genome browser platform [86]. Putative ROH islands were compared with quantitative trait loci (QTL) regions previously identified and reported in the Horse QTL database [87]. Functional analyses were performed on the genes potentially under selection, located in ROH shared in more than 85% of the animals or overlapping with known QTLs.

2.3. RESULTS

2.3.1 Pedigree-based results

2.3.1.1 Population structure and architecture

From the '70s, a gradual increase in the number of registered horses was recorded, with the highest peak of 389 horses listed in 2009. Since 2010, a reverse trend in the annual number of registered horses has been observed, from 336 in 2010 to 165 in 2016 (complete year) individuals. The average number of mares registered during the period of study was significantly higher than males ($p < 0.0001$) by a factor of three. The number of registered males over the 62 evaluated years ranged between 1 in 1939 and 178 in 2008. Until 1976, with the only exception of year of birth 1960, all male horses registered to the studbook were used for breeding. Since the foundation of the studbook in 1977, a steady decrease in the number of males used for breeding was observed, to a point where only 4.24% of the males born and registered in 2009 had offspring in the subsequent years. The number of sires born in the last birth cohort and used for breeding was equal to 50, which is 5.9% of the males currently present (Figure 4). The most popular stallion produced 222 offspring; half of the registered stallions produced more than 10 offspring and 13.8% of the sires produced one offspring only. The number of offspring per breeding mare ranged from 1 to 14, with an average of 2.60 offspring per mare.

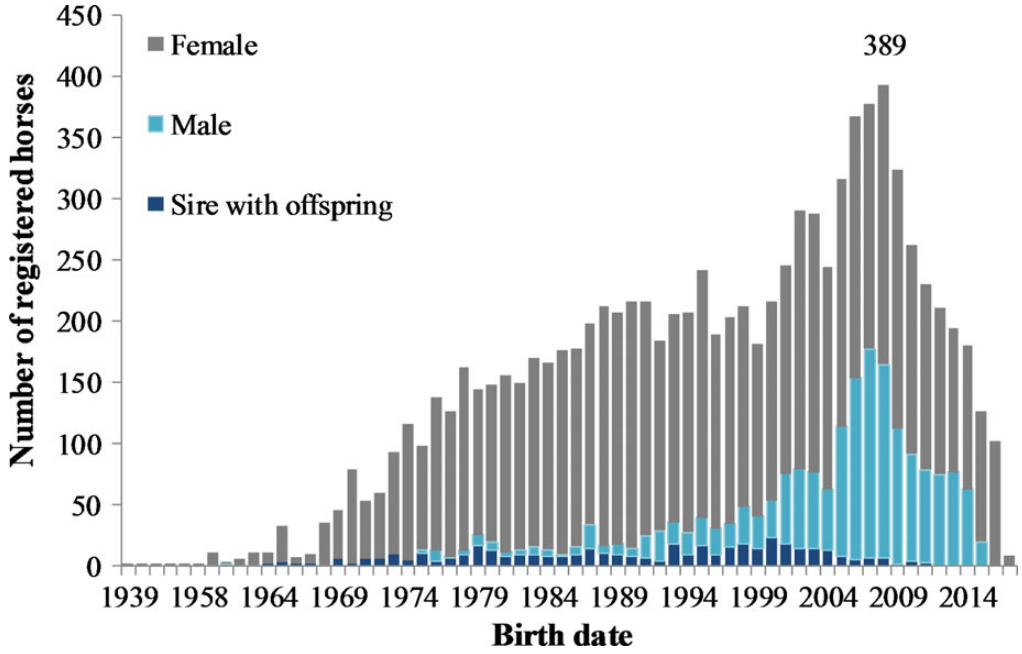

Figure 4 Animals registered per year of birth divided by sex: female, male and sire with offspring.

When considering all 7,251 horses with information available on the geographical area of breeding, we found that the Bardigiano was bred in 14 areas in Italy, and in three countries outside Italy. However, 93.5% of the horses were from four geographical areas: Parma, Piacenza, Genova and La Spezia, with 55.7% (4,041 horses) being present in the Parma province only. Breeding strategies differed between geographical areas of breeding: in the Parma territory, 64% of the stallions used for breeding were local stallions, whereas in Piacenza, Genova and La Spezia, the majority were foreign stallions. La Spezia counted the lowest number of breeding stallions born and exploited in their area of origin, with only 4% of the horses born from native stallions (Figure 5).

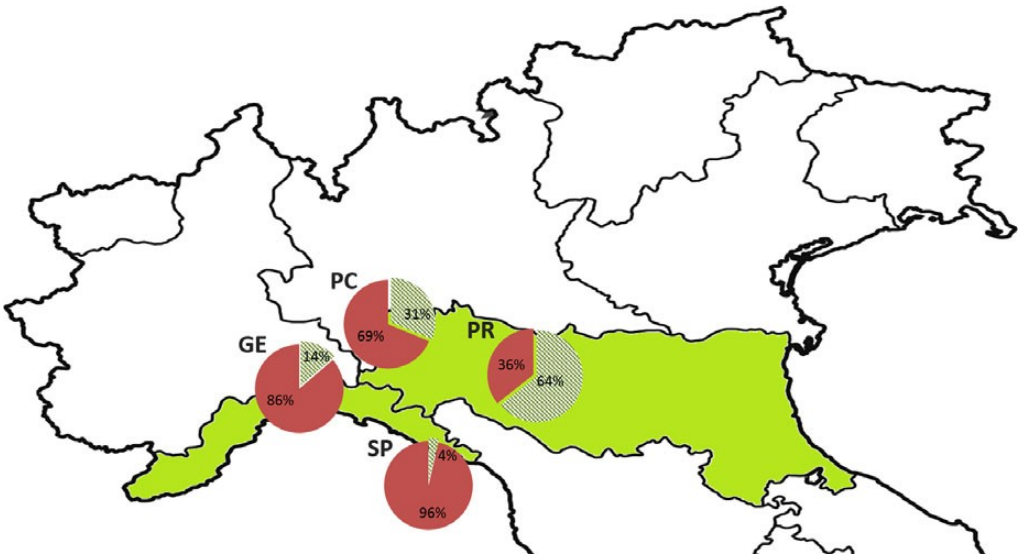

Figure 5 Bardigiano demographic distribution and population structure in different areas of Italy. Regions of Italy where the 96.5% of the Bardigiano horses have been bred are highlighted in chartreuse. Pie charts represent the percentage of stallion originally used in the breeding programme within each area. The percentage of breeding stallions born in the same area is represented with a striped hatch, whereas the percentage of foreign breeding stallions is shown with a filled pattern. GE = Genova; PC = Piacenza; PR = Parma and SP = La Spezia.

The F-statistics were calculated to evaluate the within-population genetic differences between Parma, Piacenza, Genova and La Spezia, and the values were as follows: F_{st} = 0.0028, F_{is} = 0.00941 and F_{it} = 0.0122. The classification in land types by altitude (mountains, hills and lowlands), showed that Bardigiano horses have been predominantly bred in mountainous (3,117 horses, 46%) and hilly (3,234 horses, 48%) areas.

The CGE investigates the completeness of the pedigree information, which resulted equal to 4.0 generations in the TP and 5.2 when considering only alive horses (RP). Percentage of ancestors known per ancestral generation differed between birth year cohorts. The maximum number of known ancestral generations ranged from 15 in the last birth year cohort (2011–2017), to 9 in the first birth year cohort after the foundation of the Bardigiano breeding association (1975–1983). Average CGEs differed between cohorts of horses born from different birth year cohort, for example, 2011–2017 was 6.1, from birth year cohort 1984–1992 was 2.9 and from birth year cohort 1966–1974 was 0.43. As expected, pedigree of horses born in the most recent years was deeper and presented a higher quality (Figure 6). The average GI calculated from all the pathways was equal to 8.74 years; the GI in the maternal lineages was higher than in

the paternal lineages: mother–daughter = 8.67 years and mother–son = 9.18 years, whereas father–daughter = 8.45 and father–son = 8.65 years.

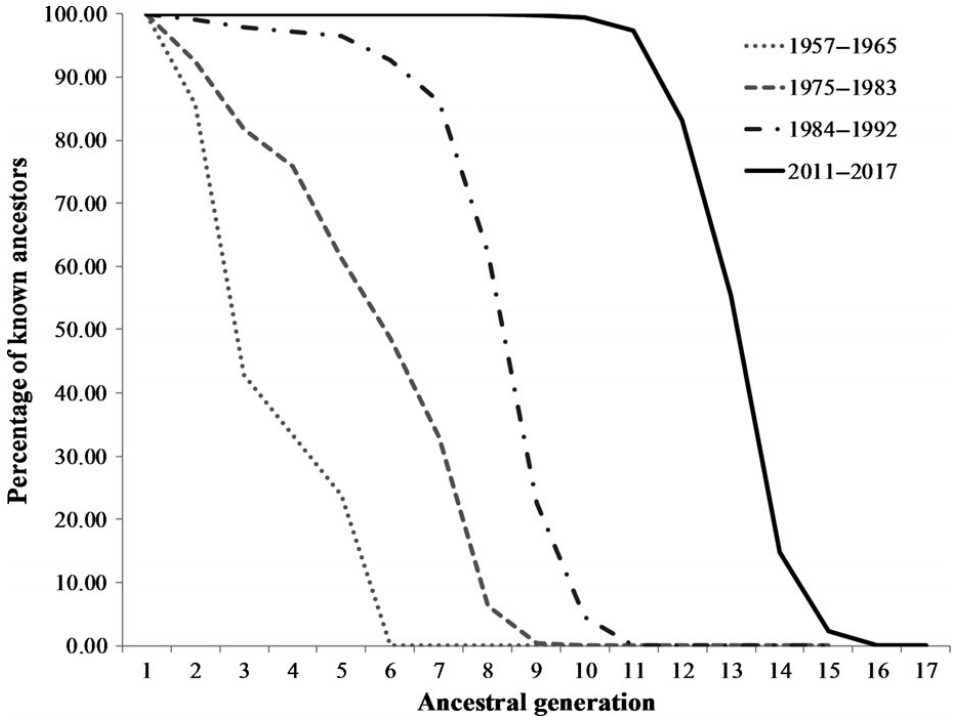

Figure 6 Percentage of known ancestors for horses born from 1957 to 2017 for four birth year cohorts where major differences were shown: 1957–1965,1975–1983, 1984–1992 and 2011–2017. Ancestral generation 1 corresponds to parents, ancestral generation 2 to grandparents, and so on.

2.3.1.2 Genetic variability based on pedigree

The effective number of founders (f_e) was equal to 20, and the effective number of ancestors (f_a) corresponded to 15 in the TP. When considering only alive animals, both f_e and f_a decreased to 18 and 14. The ratio between f_e and f_a was equal to 1.30 and 1.20 in the TP and RP. Eight ancestors explained half of the observed genetic diversity in the RP. Conversely, when considering only animals belonging to the SP cluster, the number of ancestors explaining the 50% of the observed genetic diversity decreased to 4. Differences were found in the numbers of founder genome equivalents (f_{ge}) between the two subpopulations defined on percentile class of EBVs: in the case of SP, the number of f_{ge} was 4.6, whereas in the case of IP, f_{ge} was equal to 11.0. The average inbreeding coefficient (F) was equal to 0.06 (SD = 0.06) in the whole population (TP). When considering alive animals (RP), F increased to 0.08 (SD = 0.06), ranging from 0 to 0.35 within population. Significant differences ($p < 0.05$) in the F between the groups of

superior and inferior animals were found: 0.10 (*SD* = 0.04) in the SP and 0.06 (*SD* = 0.04) in the IP. Both the *F* and the rate of inbred animals (RIA) calculated per birth year cohort increased steadily in the RP, reaching in the last cohort an *F* of 0.10 and an RIA equal to 86.5%. Horses, showing a level of inbreeding equal or higher than 0.25, were found in all the birth year cohorts (Table 2).

Table 2 Average inbreeding coefficient, rate of inbred animals (RIA) and highest inbreeding coefficient calculated for each generation from 1975–1983 to 2011–2017 in the RP.

Birth year cohort	Inbreeding*	RIA %**	Highest F***
1975–1983	0.03	20	0.25
1984–1992	0.04	29.2	0.25
1993–2001	0.06	56.3	0.32
2002–2010	0.08	77.3	0.34
2011–2017	0.1	86.5	0.35

*Inbreeding: mean inbreeding coefficient calculated per generation. **RIA: rate of inbred Bardigiano horses within generation in percentage; if a horse showed an inbreeding coefficient higher than 0.05, it was considered as inbred. ***Highest F: highest individual inbreeding coefficient found within generation.

The rate of inbreeding per generation (Δ*F*), assuming a mean GI of 8.74, was equal to 1.64% in the RP. In contrast, when considering only breeding animals, Δ*F* increased to 1.93%. Average relatedness (AR) within breeding animals did not differ significantly from the AR in the RP, and it was equal to 11.0% ($p > 0.05$) in both populations. In contrast, AR of stallions (13.0%) was higher, if compared to AR of mares (11.0%) ($p < 0.05$). A similar pattern was shown when comparing AR within superior and inferior animals with their EBV (AR_{SP} = 14.0% and AR_{IP} = 9.0%) ($p < 0.05$). Significant differences ($p < 0.05$) were shown for the comparison of *F* and AR between provinces in all cases except between La Spezia and Genova. Average *F* and AR, calculated for provinces of breeding per each altitude type (lowlands, hills and mountains), showed higher average inbreeding and relationship among individuals belonging to mountain and hill areas compared to that for lowland locations in all provinces. The only exception was La Spezia, which however did not show statistically significant differences ($p > 0.05$). Animals belonging to mountainous areas in the province of Piacenza showed the highest average *F* and AR, being equal to 0.10 and 0.13, respectively (Table 3).

Table 3 Population structure, inbreeding and relationship among classes of horses born in the most popular areas for the Bardigiano breeding: Genova, Piacenza, Parma and La Spezia, by considering the type of land (lowland, hill and mountain areas).

Location	N.[1]	Own Father[2]	Foreign Father[3]	F[4]	AR[5]	Highest F[6]	Highest AR[7]
GE: Low Land	103	4%	96%	0.03	0.06	0.25	0.15
GE: Hills	204	6%	94%	0.05	0.09	0.25	0.18
GE: Mountain	548	9%	91%	0.05	0.1	0.25	0.17
PC: Low Land	116	1%	99%	0.07	0.1	0.27	0.15
PC: Hills	956	24%	76%	0.09	0.13	0.35	0.19
PC: Mountain	598	12%	88%	0.1	0.13	0.21	0.18
PR: Low Land	201	11%	89%	0.08	0.11	0.27	0.17
PR: Hills	1899	24%	76%	0.08	0.1	0.31	0.18
PR: Mountain	1941	49%	51%	0.09	0.12	0.35	0.19
SP: Low Land	15	0%	100%	0.01	0.04	0.08	0.13
SP: Hills	175	5%	95%	0.06	0.09	0.22	0.15
SP: Mountain	30	0%	100%	0.01	0.03	0.01	0.07

Population structure parameters: [1]N. = number of foals born within location. [2]Own father = percentage of foals born from stallions belonging to the same area. [3]Foreign father = percentage of foals born from stallions belonging to different area. [4]Mean inbreeding (F) and [5]average relationship (AR). [6]Highest inbreeding (F) and [7]average relationship (AR) within location. GE, Genova; PC, Piacenza; PR, Parma and SP, La Spezia.

The effective population size (N_e) in the TP resulted equal to 60.81, 26.41 and 18.26 when using the increase in inbreeding by maximum generations, complete equivalent generations and full generations, respectively. The N_e calculated against the ΔF was equal to 30.67 in the RP and to 26.32 in the BP. In Figure 7, the breeding values (EBVs) of the 66 sires building the BP were plotted against their AR in the population. Sires with an AR lower than 13% and with an EBV higher than 50 were showed as rhombus representing the optimal in terms of potential breeding contribution. This result supports the hypothesis that differences in the selection strategies between genetically superior and inferior animals are expected, thus affecting their genetic variability. The 66.0% of the sires showing an EBV higher than average (EBV > 50) displayed also an AR higher than the average value in the sampled sire population (AR > 0.13).

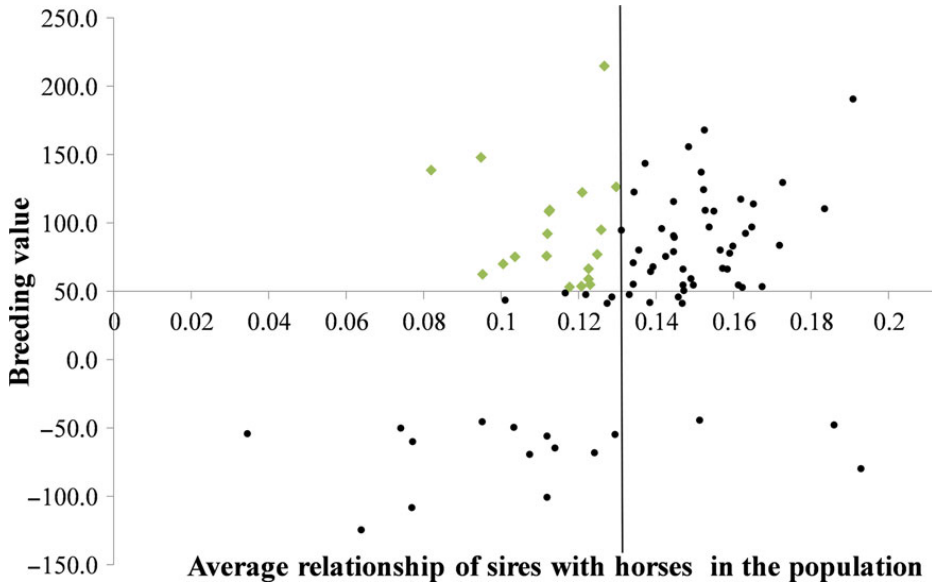

Figure 7 Breeding values plotted against the average relationship of sires to horses in the reference population. The mean breeding value in the sampled sires was equal to 50 and the average relationship to 0.13. The rhombuses represent the 20% best sires in term of optimal contribution.

2.3.2 Genomic-based results

All horses except one passed the genotype data quality control (37 males, 51 females), and 58,047 SNPs were retrieved. The average genotype call rate was 0.99 and the average pedigree depth based on complete generation equivalent of the selected horses was 6.28 (SD = 0.79). The horses descended from 54 stallions with an average of 1.65 offspring each (SD = 0.82) and 76 mares with 1–2 offspring each. In total, 80% of the horses were born between 2011 and 2019, whereas 18 horses were born between 2001 and 2010.

2.3.2.1 Trend of the effective population size

For the Ne calculation, only animals born between 2011 and 2019 were included, which numbered 70 horses. The Ne declined over time and was estimated to be ~39 horses one generation ago. Instead, the estimated Ne 20 generations ago was around 195 horses (Figure 8).

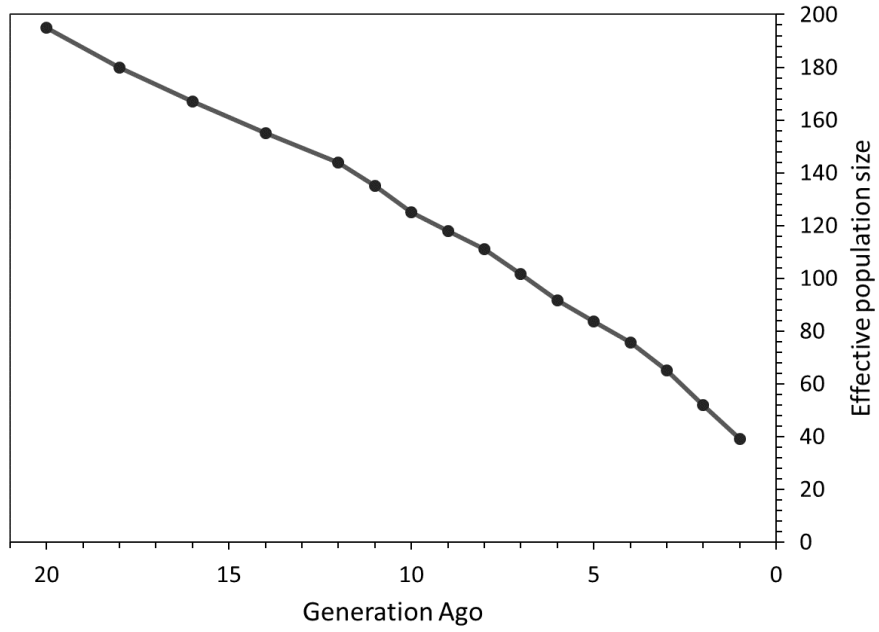

Figure 8 Plot of effective population size change between 20 generations ago and the last generation in the Bardigiano horse breed.

To investigate the change in slope of the inferred Ne obtained from LD-based estimation, we used the NeS method, which offers more detailed information about population changes 1–20 generations ago. A constant rate of change is shown as a flat line at 0 in the Y-axis, whereas deviations above and below 0 represent relative increases and reductions of the Ne change in slope compared to the previous two generations. This analysis highlighted two sharp reductions in Ne in the Bardigiano breed, being approximately ten and six generations ago (Figure 9).

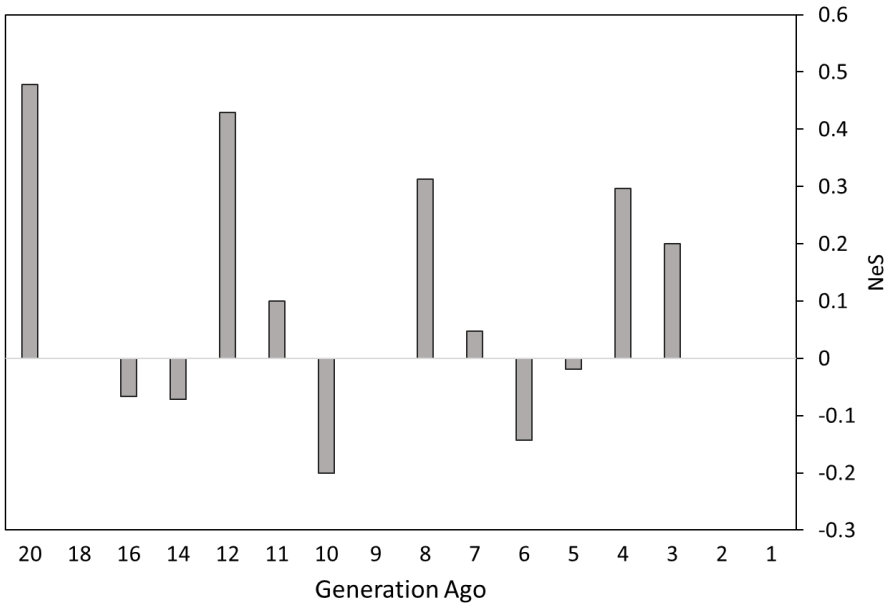

Figure 9 Ne Slope analysis (NeS) between 20 generations ago and the last generation in the Bardigiano horse breed. A constant rate of change in the Ne is shown as a flat line at 0 in the Y-axis.

2.3.2.2 Runs of homozygosity as a measure of inbreeding

A total of 28,423 ROHs were found among the 88 Bardigiano horses analysed in this study. An average of 323 ROHs were found per horse, with a maximum number equal to 365 and minimum of 285 ROHs. The average length of ROHs per individual was equal to 1215.8 Kb ± 173.7 Kbp (SD). The majority of ROHs were shorter than 2 Mbp, with 69.5% being shorter than 1 Mbp and 21.4% of ROHs between 1 and 2 Mbp length (Table 4). The proportion of ROHs longer than 8 Mbp was equal to 1.2%, with an average length of 13.4 Mbp. A total of 84 horses exhibited ROHs in the >8 Mbp length class, with an average of 4.2 ROHs per horse in this class and a minimum and maximum equal to 1 and 13 ROHs, respectively.

Table 4 Descriptive statistics of runs of homozygosity (ROH) per length class.

Length Class	N.[1]	N_{ROH}[2]	ROH Percentage[3]	S_{ROH}[4]	L_{ROH}[5]
0.5-1 Mbp	88	19,746	69.50%	224.4	0.7
1-2 Mbp	88	6,073	21.40%	69	1.33
2-4 Mbp	88	1587	5.60%	18	2.74
4-8 Mbp	88	664	2.30%	7.5	5.42
>8 Mbp	84	353	1.20%	4.2	13.4

[1] Number of animals, [2]Total number of ROH, [3] Relative percentages, [4] Average ROH number per animal, [5] Average length of ROHs.

The number of ROHs and length of ROH segments varied across chromosomes as shown in Figure 10. The highest number of ROHs was identified on ECA1 (2,386), while the lowest on ECA31 with 273 ROHs detected. Nevertheless, when normalizing for the chromosome length, the highest portion was found on ECA29 and ECA28. The highest chromosome length covered by ROHs was detected on ECA10, with 1.49 Mbp covered by ROHs followed by ECA5 (1.46Mbp), ECA1 (1.43 Mbp) and ECA3 (1.40 Mbp).

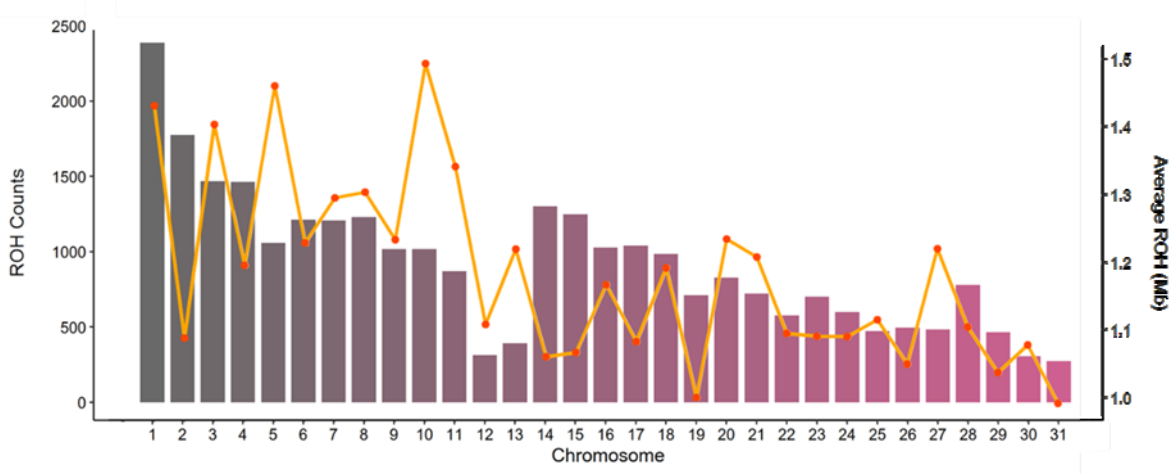

Figure 10 Distribution and average length of runs of homozygosity (ROH) in Mbp detected across the autosomal genome in the Bardigiano horses. The bar plots show the ROH counts per chromosome and the orange line the average ROH size (Mbp) per chromosome.

The inbreeding measured as the proportion of the genome covered by ROH (>0.5 Mbp) resulted in an average F_{ROH} equal to 0.17 ± 0.03 (SD). To differentiate ancient and recent inbreeding, we calculated F_{ROH} based upon five ROH length classes as presented in Table 5. For the ROH length class above 8 Mbp, four animals did not show ROHs and thus their inbreeding resulted equal to 0. The degree of inbreeding based on the longest ROH class showed the lowest result, with an average inbreeding equal to 0.02.

Table 5 Descriptive statistics of inbreeding based on runs of homozygosity (F_{ROH}) within each ROH length class. Number of animals which exhibits ROH within each length class, mean, minimum, maximum and standard deviation of F_{ROH}.

Length Class	Inbreeding based on ROH (F_{ROH})				
	N. [1]	Mean	Min. [2]	Max. [3]	SD [4]
0.5-1 Mbp	88	0.17	0.12	0.26	0.03
1-2 Mbp	88	0.10	0.06	0.20	0.03
2-4 Mbp	88	0.06	0.02	0.15	0.03
4-8 Mbp	88	0.04	0.01	0.13	0.02
>8 Mbp	88	0.02	0.00	0.10	0.02

[1] Number of animals, [2] Minimum, [3] Maximum [4] Standard deviation.

The inbreeding calculated per chromosome showed large variation among chromosomes both in terms of average values as well as distribution of inbreeding level per individual (Figure 11). The highest average inbreeding was found in ECA1 (mean = 0.21 ± 0.09 SD) and the lowest on ECA30 (0.12 ± 0.09 SD). For all chromosome except ECA3, ECA15 and ECA17, a highly skewed distribution towards higher values was found (distribution skewness above one). A kurtosis value above three was used to investigate the presence of outliers per each chromosome. Twenty-one out of the 31 chromosomes showed a kurtosis higher than three, highlighting the presence of some horses exhibiting excess of homozygosity.

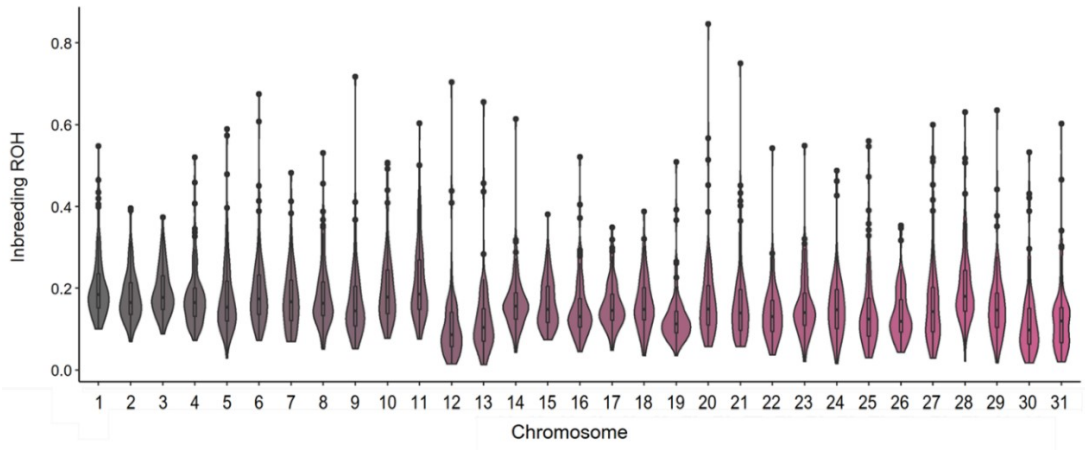

Figure 11 Violin plot of the inbreeding based on runs of homozygosity (F_{ROH}) calculated per chromosome in the Bardigiano horse breed.

2.3.2.3 Runs of homozygosity as potential signs of selection

A total of five chromosomes (ECA3, ECA10, ECA11, ECA15 and ECA19) contained ROH islands that were shared by more than 70% of individuals in the total sample. Three ROH islands were detected on ECA3, two on ECA10 and the remainders were equally distributed among ECA11, ECA15 and ECA19 (Figure 12).

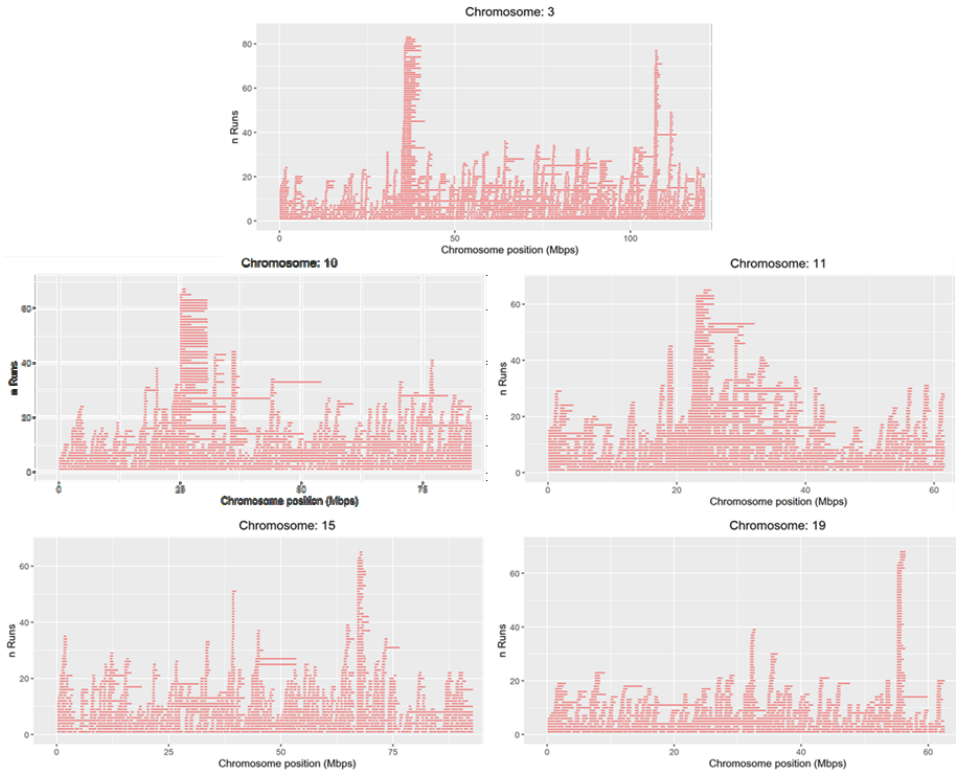

Figure 12 ROH islands on ECA3, ECA10, ECA11, ECA15 and ECA19 in the Bardigiano horse breed. In the x-axis the chromosome position in Mb and on the y-axis the number of horses showing a ROH in each chromosome position.

The three ROH islands located and ECA3 were shared in more than 85% of the animals, and in particular the ROH island located on ECA3 (ECA3: 35.48-36.01 Mbp) was shared in 93% of the animals (82 out of the 88). Table 6 reports the genomic coordinates of the ROH islands and the annotated genes. A total of 115 annotated genes, two miRNAs and three snoRNA were located within the eight ROH islands. The two ROH islands on ECA3 located between positions 35.48 and 37.59 Mbp overlapped with four known QTL regions, two related with insect bite hypersensitivity [88, 89], one with guttural pouch tympany [90] and one with hair pigmentation [91]. The ROH island on ECA3:106.77-107.71 Mbp coincided with a known QTL for body height at withers [92]. The ROH island on ECA10 did not overlap with known QTLs, whereas the ROH island on ECA11 coincided with four QTL regions related to insect bite hypersensitivity [88], hair density [93], overall body size [94] and height at withers [95].The ROH islands on ECA15 and ECA19 did not overlap with any known QTLs.

Table 6 Runs of homozygosity (ROH) islands shared in over 70% of the Bardigiano horses with genomic coordinates and list of annotated genes located within each ROH island.

ECA[1]	Start (bp)[2]	End (bp)[2]	Length (Kbp)	Annotated genes	% of horses
3	35,477,778	36,012,699	535	*ZNF469, ZFPM1, ZC3H18, IL17C, CYBA, MVD, SNAI3, RNF166, CTU2, PIEZO1*	*93%*
3	36,131,080	37,590,699	1460	*CBFA2T3, ACSF3, CDH15, SLC22A31, ANKRD11, SPG7, RPL13, CPNE7, DPEP1, CHMP1A, SPATA33, CDK10, SPATA2L, VPS9D1, ZNF276, FANCA, SPIRE2, TCF25, MC1R, TUBB3, DEF8, DBNDD1, GAS8, PRDM7, CENPE, BDH2, SLC9B2*	*92%*
3	106,769,095	107,709,841	941	*LCORL, NCAPG, DCAF16, FAM184B,*	*92%*
10	24,965,648	25,501,456	536	*IL11, TMEM190, TMEM238, RPL28, UBE2S, SHISA7, ISOC2, C19orf85, ZNF628, NAT14, SSC5D, SBK2, SBK3, ZNF579, FIZ1, ZNF524, ZNF865, ZNF784, ZNF580, ZNF581, CCDC106, ZNF835, U2AF2, EPN1, RFPL4A, EQUCABV1R902, EQUCABV1R903, NLRP4, NLRP13, NLRP5*	*76%*
10	25,636,022	26,106,403	470	*EDDM13, ZNF667, ZNF583, ZNF582, SMIM17, ZNF471, ZFP28, ZNF470, ZNF71*	*76%*
11	22,957,922	24,177,108	1219	*CDK12, MED1, STAC2, CACNB1, ARL5C, PLXDC1, FBXO47, LINC00672, LASP1, RPL23, C17orf98, CWC25, PIP4K2B, PSMB3, MLLT6 , PCGF2, CISD3, EPOP, SRCIN1, ARHGAP23, SOCS7, GPR179, MRPL45, NPEPPS, KPNB1, TBKBP1, TBX21, OSBPL7, MRPL10, LRRC46, SCRN2, SP6, SP2, PNPO*	*73%*
15	67,045,471	67,539,463	494	*LCLAT1, LBH, YPEL5*	*73%*
19	55,113,101	55,782,316	669	//	*77%*

[1] ECA: Equus caballus autosome chromosome; [2] Position in base pairs (bp) on EquCab3.

2.4 DISCUSSION

2.4.1 Comparison to "similar horse breeds" scenario based on pedigree data

The number of Bardigiano horses registered in the studbook showed two separate trends: from the beginning of breed formation to 2009, a steady increase was shown. On the other hand, from 2010 to current days, a reverse trend was observed. Similarly, the number of stallions used for breeding gradually decreased. The observed trend might be related to the decrease in the number of Bardigiano horses used for meat production. This reduction might be the consequence of three events. First, the opening of international markets of equine meat has increased the availability of meat from foreign countries, thus reducing local production. Second, due to the Eurozone crisis in the late 2009, a reduction in meat consumption has been generally observed in several European countries. This crisis might have affected horse meat consumption in Italy, thus dropping local horse meat production even further [96] . Finally, since the beginning of the 21st century in the equine market, the demand of horses for leisure activities has increased deeply. Consequently, the breeding goal of many Bardigiano breeders has changed from meat production to horses shaped for pleasure activities, which have a higher economic value, a longer lifespan and thus an adverse influence in the number of horses registered per year. In the last 13 years, the number of living animals has been fairly stable; from a previous study, this number resulted equal to 3,556 animals, which is 140 animals more than the current population [97].

Although the Bardigiano is not classified as at risk of extinction anymore, its population size is considerably smaller than other Italian native breeds. Currently, the Bardigiano has a registered population of 3,416 individuals which is lower than the Italian Haflinger, Murgese and the Italian Heavy Draft populations: all of them have a population size larger than 5,500 animals. A recent study on the Maremmano horse population showed a registered alive population of 5,705 horses and a BP of 1,532, therefore nearly two times larger than the Bardigiano breed [35]. Similarly, the Bardigiano population size is smaller than many international breeds; as an example, the number of Bardigiano

horses is roughly half of the current Lusitano population [41]. The number of breeding animals in the Bardigiano breed and in the endangered Old Kladruber horse population is very close: 742 and 612, respectively [42].

2.4.2 Proposed solutions to increase genetic variability in the coming generations

Selection intensity differed between mares and stallions: 26.3% of the females have been used for breeding versus only 7.8% for males. The proportion of breeding stallions to mares was equal to one stallion every 10 mares. In European horse breeds for sport performances, the Hanoverian breed presents 20 mares per stallion, while the Dutch harness population counts 1 stallion every 60 mares [36, 40]. Therefore, the Bardigiano breed presents significantly lower figures. Artificial insemination (AI) improves stallion reproduction efficiency, thus increasing the number of mated mares obtainable per stallion. Nowadays, AI is broadly used in most European horse breeds for sport performances, but not in the Bardigiano. The absence of AI in the Bardigiano might explain the lower mares/stallion ratio and a limited average number of offspring per stallion.

The number of sires born in the last birth cohort and used for breeding was equal to 50, which is 5.9% of all currently available male Bardigiano horses. Once a year, the pedigree database is updated by voluntary owner declaration of male horses undergone castration. Thus, this portion might be slightly higher as some of the male horses considered as potential breeding animals in reality have been gelded. Breeding strategies differed among areas where Bardigiano horses are commonly bred. Similarly to what previously stated, as the use of AI has not been implemented yet in this breed, the exchange of genetic material becomes challenging between more distant areas. The higher inbreeding and higher relationship among individuals belonging to the cluster of mountainous areas support this hypothesis. In addition, the F-statistics analyses revealed higher F_{st} and F_{is} than in the Maremmano breed, thus showing more genetic separation between areas of breeding and higher average inbreeding within area of breeding [35]. To allow for a more uniform genetic flow between areas, and to reduce

the risk of genetic drift and inbreeding within local area, we suggest the application of AI in this breed. However, if AI is introduced, breeding strategies for balancing stallion's contributions are needed to reduce the risk of critical use of only the most popular stallions.

In recent years, strategies to optimize the contribution of breeding animals have been implemented. Optimal contribution selection (OCS) was applied in the Norwegian and the North - Swedish coldblooded trotter, where the importance of annual mating quota per stallion was highlighted to monitor stallion's contribution in the population [98]. Hence, especially if AI is introduced in the Bardigiano breed, OCS might be an effective solution to balance stallion's contributions. Mating strategies to control future inbreeding were considered in the indigenous Swiss horse breed, where they combined the best linear unbiased prediction (BLUP) methodology with OCS theory [37]. Thus, OCS together with BLUP might be used to monitor the Bardigiano genetic diversity, while allowing for response to selection for economically relevant traits.

Even though in breeds potentially under endangered status the most urgent matter is to conserve genetic diversity, survival probability highly depends on their economic value. Thus, by improving economically relevant traits, such as conformation, gaits and attitude traits via the breeding program, profitability increases, thereby reducing the risk of extinction. Therefore, the simultaneous implementation of a selection program which optimizes the selection of breeding candidates in terms of both conservation and breeding values, is suggested to enhance the demand of Bardigiano horses in the equine sector.

The GI in the Bardigiano horse population was equal to 8.74 years. In draft horse breeds, the GI varies between 7.0 and 7.9 years for French Comtois draught horse population and for the Austrian Noriker draught horse, respectively [99, 100]. Horse breeds for sport purposes showed generally a longer GI: 10.3 for the Lusitano breed, 10.1 for the Andalusian horse population and 10 years for the Hanoverian horses [36, 41, 101]. The Dutch harness horse population presented a GI akin to what we found in the Bardigiano, i.e. 8.6 years [40]. Differences in GI among breeds might be related to the end use of the breed. Draught horse breeds mainly have a shorter GI, as this horse type is commonly used for meat production and it is rarely used in competitions. In contrast,

as horses belonging to sport breeds attend competitions postponing the reproduction activity, they tend to show longer GIs. The GI in the Bardigiano population fits in between sport and draft breed types.

Pedigree quality and depth appeared sufficient to calculate genetic diversity parameters. It is known that the absolute level of inbreeding is highly dependent on the quality of the data available [40, 102]. Hence, inbreeding rate appears to be a more appropriate parameter to assess genetic variability across time within a breed. The average inbreeding in the RP was 33.3% higher than in the whole population, thus showing a possible increase in the inbreeding over time. This assumption is confirmed by the rate of inbreeding per generation (ΔF), which resulted equal to 1.64% in RP. The Food and Agriculture Organization (FAO) stated that the value of ΔF should not exceed 1% to avoid substantial loss of genetic material over time [11]; unfortunately, this limit is not respected in the Bardigiano breed. In endangered horse populations, we found similar ΔFs: in the French Boulonnais draught horse population ΔF was equal to 1.38%, and in the Old Kladruber horse population equal to 1% [42, 100]. Although the Bardigiano is not included in the list of endangered breeds, it currently shows a critical increase in the inbreeding in the last generations.

From the comparison between clusters of EBVs, differences in the AR were found between superior and inferior animals: on average, horses with high breeding values were also more related to each other. Consequently, in order to keep the population free from extinction, it is important to include parameters for preserving the genetic diversity in the breeding program. One potential solution is to ensure that parents of future breeding animals are not closely related. Thus, including the AR as an extra parameter of selection in the breeding program might be an effective solution to avoid high relatedness among mating and consequently limiting the risk of severe loss of genetic diversity in the Bardigiano breed.

2.4.3 Comparison to similar horse breeds' scenario based on SNP chip data

The evaluation of genetic diversity with traditional methods based on pedigree analyses have been widely used to manage animals' genetic resources. In small livestock populations, where financial support is generally a limiting factor, pedigree-based methods are still commonly adopted due to their cost-effectiveness ratio [35, 42, 103, 104]. However, with novel molecular and bioinformatics approaches, genetic variability can be more proficiently evaluated by using genomic information, and it could lead to more precise and effective conservation programs. The SNP data analyses presented in this thesis can be considered as the first step in deepening our knowledge into the Bardigiano genomic diversity.

Effective population size (Ne) computed by using genealogical data could be biased if pedigree depth and quality are low [105]. To this end, molecular markers could be a valuable and alternative form of information to characterize Ne in livestock populations [106]. The Ne estimation based on LD information was approximately 39 horses in the last generation, which is lower than in the Persian Arabian Horses (Ne=113) [107], and slightly lower than in Saxon Thuringa Coldblood (Ne=48.1), and in the Rhenish German Draught Horse (Ne=46.1)[108]. Interestingly, the Ne estimate based on pedigree data (calculated as individual increases in inbreeding [109, 110] in the alive Bardigiano population) was lower (Ne=30.7), yet comparable, to what found by molecular data analyses. This evidence highlights that Ne estimate obtained from pedigree can be considered a viable approximation of molecular-based Ne in the case of Bardigiano horses. The estimation of Ne in the last 20 generations showed a steady decrease in effective population size, in line with the known history of the breed. The NeS analysis highlighted two sharp reductions in Ne in the Bardigiano breed, approximately around ten generations and six generations ago. During World War II, the number of Bardigiano horses decreased dramatically, with only five stallions and 150 mares surviving [29]. This evidence may justify the sharp reduction in the Ne about 10 generations ago. In contrast, the general increase and stabilization of the Ne from the 6^{th} generation and

onwards might reflect the institution of the Bardigiano studbook in 1977 which aimed to recover the breed.

The extent and frequency of ROHs have been widely used to infer ancestry at individual and breed level [53]. Long ROHs are generally considered to be a sign of recent inbreeding, whereas short ROHs can capture ancient inbreeding which derived from older ancestors and can capture population bottlenecks. Under the assumption that 1 cM equals to 1.24 Mbp, ROHs can be separated in length classes to express different points in time when inbreeding occurred [56]. In the Bardigiano horse, the majority of ROHs belonged to the shortest length class being equal to 0.5-1 Mb (69.5%). Similar results were observed in local breeds, as in the Bosnian Mountain Horse (61.7%) [62] and in the Noriker horse breed where the majority (60%–85%) of ROH segments was shorter than 4 Mbp [67]. This result might mainly originate from the small population size of the Bardigiano breed and the strong founder effect due to the experienced bottlenecks. In contrast, the lowest percentage of ROHs (1.2%) resided into the longest class (ROH>8Mbp), which highlights a rather small reduction in genetic variability occurred in the last generations. Likewise, the F_{ROH} ranged from 17.0%, considering all ROHs regardless of the size, to 2.0%, considering only the longest class (ROH > 8Mbp). This last result highlights the occurrence of older inbreeding, rather than inbreeding which happened in the last few generations. Nevertheless, the average F_{ROH} in the Bardigiano was higher if compared to other small-sized horse populations. For example, in the Croatian Posavje horse breed, F_{ROH} was equal to 8.59%; in the Bosnian mountain horse (originating from Bosnia and Herzegovina) this value reached a mean of 13.21%; and in the Austrian Noriker horse it ranged between 8.00% and 13.00%, depending on coat colour strains [66, 67]. Similar levels of F_{ROH} as in the Bardigiano horses were found in the Polish Koniks breed (mean=16.00%) which is a primitive breed closely related to the extinct wild Tarpan [111] and in Purebred Arabian horses (mean=17.70%), despite the larger population size [62]. In contrast, a higher level of inbreeding was found in Friesian horses (mean=22.30%) compared to the Bardigiano [112]. As expected from previous studies [17, 113, 114], the inbreeding based on molecular data in the Bardigiano tended to be higher than what reported above based on pedigree data.

2.4.5 Further evaluation of the role of the candidate genes potentially under selection

In general, the ROHs caused by inbreeding tend to be distributed unevenly across the genome, as we have found in the Bardigiano horse, with a different distribution of number and size of ROH for each chromosome [115]. However, ROHs that are located in specific genomic regions and shared among several individuals are thought to be potential signs of selection [60, 64, 115] . The reasons behind this theory are the driving forces of directional selection which increase homozygosity around a target locus (ROH island). As already mentioned, the breeding objective of the Bardigiano studbook can be divided into three main goals: (a) ensure long-term survival of the breed; (b) preserve Bardigiano distinctive morphological features; and (c) increase the use of this breed in riding and draft activities [75]. For the latter aspect, in the '90s, the Bardigiano studbook introduced a breeding value for the height at withers to help the conversion in taller horses, thus more suitable for leisure. It is noteworthy that in this study we found ROH island in two of the four loci previously found to explain 83% of size variation in horses [94]. Specifically, one ROH, located on ECA3 and shared among 93% of the Bardigiano horses, overlapped with the ligand dependent nuclear receptor corepressor-like (*LCORL*) gene [116]. The *LCORL* gene is a transcription factor that has been associated with human height [117–119]. In cattle, *LCORL* was identified in a screen for loci under selection [120] and the immediately adjacent gene, *NCAPG*, has been implicated in prenatal growth [121] as well as in body size in Franches-Montagnes horses [121]. The other ROH that overlapped with a locus among the four explaining 83% of size variation in horses, was on ECA11 where LIM and SH3 protein 1 (*LASP1*) gene is located. The *LASP1* gene mediates cell migration and survival and its expression is induced by insulin-like growth factor (*IGF1*) [122]. However, the ROH on ECA11 is located within a gene-dense region counting thirty-four genes; therefore, we cannot exclude that selection in this region might target other genes. Based on the overlap with known QTLs for body size, we hypothesize that the ongoing selection toward higher horses might have led to a reduction in variability in those two regions.

In addition, this latter region of ECA11 together with two ROH islands on ECA3 between 35.48 and 37.59 Mbp overlapped with known QTLs for equine insect bite hypersensitivity (IBH). The IBH is a pruritic skin allergy caused primarily by biting midges, Culicoides spp [88]. This skin disease has polygenic inheritance and occurs at high frequencies in several horse breeds worldwide, thus causing increased costs and reduced horse welfare. To the best of the author's knowledge, no cases of IBH have been hitherto reported in the Bardigiano breed, even if the animals are mostly farmed outside, and pasture in the mountain area is commonly practiced. The summer pasture exposure and the surrounding vegetation have been identified as risk factors associated with IBH in warmblood horses [123]. We therefore hypothesize that long and extensive exposure of insect bites might have shaped the genome of Bardigiano horses towards more resistant animals. Nevertheless, those three ROH islands located on ECA3 and ECA11 overlapped also with known QTLs for hair pigmentation [91] and hair density [93]. Interestingly, some of the distinctive features of the Bardigiano horses are related to hair type and pigmentation. The only coat colour accepted by the breeding association is bay, with a slight preference for dark bay, whereas chestnuts and extremely light bays are not allowed in the breed. In addition, only limited white markings on the legs and face are allowed and, in general, if present, they are scored as a defect in the horse evaluation. In the study by Haase et al. 2013 [91], seven QTLs were significantly associated with the white marking phenotype which explained ~54% of the total genetic variance [91, 124]. The ROH island identified on ECA3 (36.13-37.59 Mbp) in the Bardigiano horses overlapped with the melanocortin receptor gene (*MC1R*). The *MC1R* encoded by the Extension (E) locus controls, together with the peptide antagonist agouti-signalling-protein (*ASIP*), the amounts of melanin pigments in mammal [125]. We therefore cannot either rule out the hypothesis that this region might be under selection pressure because of morphological requirements in terms of coat colour in the Bardigiano breed.

Chapter 3

This chapter is based on the following published research studies:

1) Signatures of selection in the genome of Swedish warmblood horses selected for sport performance by **Michela Ablondi[1,2]**, Åsa Viklund[2], Gabriella Lindgren[2,3], Susanne Eriksson[2], Sofia Mikko[2]. Published in the BMC Genomics, BMC Genomics 2019; 20:717
2) Genomic Divergence in Swedish Warmblood Horses Selected for Equestrian Disciplines by **Michela Ablondi[1]**, Susanne Eriksson[2], Sasha Tetu[2], Alberto Sabbioni[1], Åsa Viklund[2], Sofia Mikko[2]. Published in Genes, Genes 2019; 10:976

[1] Dipartimento di Scienze Medico-Veterinarie, Università di Parma, Parma, Italy

[2] Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden

[3] Livestock Genetics, Department of Biosystems, Leuven, KU, Belgium.

3.1 INTRODUCTION

3.1.1 Introduction to the horse sport breed

The Swedish Warmblood (SWB) is a modern horse breed selected for equestrian sports, such as show jumping and dressage [126] (Figure 13). The origin of the breed dates back to the 18th century when the Royal Cavalry requested more agile and faster horses, thus leading to intensified breeding and selection of Swedish riding horses [127]. The SWB studbook was founded in 1928 with the initial aim to breed horses for multiple equestrian purposes. The current goal of the SWB studbook is to breed internationally competitive warmblood horses in terms of rideability, performance-oriented temperament, excellent gaits and/or jumping ability [126].

Figure 13 Examples of Swedish Warmblood horse (SWB). The photo on the top left corner shows Luxus CML (SWB) and was provided by Johannes Walter. The photos on the bottom left corner and on the right show Coreograph (SWB) and were provided by Prof. Sofia Mikko. The photographer of the picture on the left corner was Sofie Benneborg and on the right was Carin Wrange.

Since the demand for physical and mental abilities in sport horses has increased noticeably in the last decades, intensive selection of SWB horses for specific disciplines, namely dressage or show-jumping, is now common practice. Consequently, the SWB breeding program has evolved towards genetic improvement for specialized disciplines. Since 2002, breeding stallions are selected as either dressage or show jumping stallions [128], and since 2006 the SWB studbook has published a main estimated breeding value (EBV) for show jumping and another one for dressage performance based on scores at young horse tests and competition data [129]. However, all SWB horses tested at young horse evaluations and stallion performance tests are assessed for both gait and show jumping traits, regardless of the discipline they are bred for. The genetic trend based on EBVs has shown that genetic progress increased substantially in the mid-1980s for both disciplines, and has been considerably higher for show jumping than for dressage, mainly due to successful stallion selection and higher heritability for show jumping traits [129].

3.1.2 State of the art

Recent advances in genomic methodologies have paved the way to explore the effects of selection in the genome. Positive selection reduces genetic variability and it results in increased genomic homozygosity. As previously stated in Chapter 2, one common way to detect potential signs of selection is to analyse the genome to find stretches of homozygosity (ROH). Similarly, the degree of homozygosity at haplotype level can be used to detect signatures of positive selection [130]. The cross population extended haplotype homozygosity (XP-EHH) method estimates the length of extended haplotypes and evaluates differences between two populations. This method has previously been used to detect breed specific signatures of selection in domestic species [50, 131, 132]. The XP-EHH analysis effectively detects signals of differentiation across breeds and it has previously been used to detect putative loci affecting height in Shetland ponies [133]. Genomic regions associated with selection for racing performance were identified by a scan for selective sweeps using whole-genome sequencing data from Thoroughbreds and the native Jeju breed [134].

While the specialization for different disciplines within a breed implies some challenges for traditional genetic evaluation, it also provides possibilities to compare genomic information from horses originally from the same breed but selected for different purposes. Such knowledge can, in turn, help sharpen genetic selection tools for the future. Recent examples of genomic studies have shown the benefits of studying genetic subpopulations. In cattle and in horses, genetic stratification within breeds, resulting from selection for different purposes, was shown by using Principal Component Analysis (PCA) and Discriminant Analysis of Principal Component (DAPC) [50, 131, 135]. Petersen et al. (2014) used fixation index (F_{st}), PCA and haplotype analyses, to find genomic differences across six performance groups of Quarter Horse, caused by increased specialization over the past 75 years [136]. The presence of signatures of selection within Quarter Horse subpopulations was confirmed in a more recent study [137]. Principal Coordinates Analysis (PCoA) was used both to explore the origin of the feral horses in Theodore Roosevelt National Park, and to visualize genetic distances among horses belonging to an emergent breed in the Azores [138, 139].

3.1.3 Research questions

The aim of this study was to detect genomic regions under selection in SWB horses. In line with previous studies, where comparisons between breeds were used to highlight signatures of selection [64, 134], a genome scan for signatures of selection in SWB horses and Exmoor ponies was performed. The comparison with Exmoor ponies is justified as the Exmoor Pony Society primarily aims to protect the heritage of the breed by preserving its genetic diversity [140], and therefore selection for sport performance traits is not practiced [141]. The breeding mission of the Exmoor Pony Society is specifically to preserve this old native breed while maintaining all the traits of its ancestors (Figure 14). Moreover, a recent study showed that the Exmoor pony did not exhibit any common homozygous region with breeds intensively selected for sport performance, which supports its suitability for the purpose of this study [59].

Figure 14 Example of Exmoor pony. Photo taken from pikist free photo (https://www.pikist.com/free-photo-sivgv)

Based on the hypothesis that performance-oriented selective breeding increased genomic homozygosity in specific regions in SWB horses, two different approaches were used: 1) analysis of ROHs detected in SWB and Exmoor ponies, and 2) two population differentiation tests, F_{st} and XP-EHH analysis, comparing SWB and Exmoor ponies.

Since the current specialisation of SWB horses for specific disciplines, a further focused was applied on potential differences at genome level within subgroups of the SWB breed. Therefore 3) an exploration of the genomic structure in the SWB breed, 4) a detection of putative genomic subpopulations of SWB horses based on their EBVs for show jumping and 5) a detection potential signs of selection in the two subpopulations were perfomed. Two methods, PCoA and DAPC, were used to detect population stratification within the SWB breed (aims 3 and 4), while F_{st} and XP-EHH scanned the genome for potential signatures of selection (aim 5). To explain the biological importance of selection footprints, a functional classification of the genes identified within regions potentially under selection was performed.

3.2 MATERIALS AND METHODS

3.2.1 Sample collection

The study included 380 Swedish Warmblood horses born in 2010-2011 (selected tested population, STP). The horses (182 males and 198 females) were assessed at young horse evaluation tests at the age of three [20] and descended from 145 sires with 1-11 offspring each, and 372 mares with 1-2 offspring each. The selected horses were either: 1) horses with high scores for show jumping but lower ones for gaits (n=48), 2) horses with high scores for gaits but lower ones for show jumping (n=48), 3) horses with high scores for both show jumping and gaits (n=143), and 4) horses with low scores for both show jumping and gaits (n=141). Breeding values from the latest routine genetic evaluation (2018), estimated in a multi-trait animal model, and based on young horse tests, together with competition data [128], were available for all studied horses. A breeding value equal to 100 denotes the average for all tested horses between four and eighteen years of age in the SWB population. In this study, horses with EBVs for show jumping performance above 100 were classified as show jumping horses (SJ), and horses with EBVs less than 100 as non-show jumping horses (NS) (Figure 15). The majority, but not all, of the NS horses could be described as horses bred for the dressage discipline. Finally, the proportion of thoroughbred contribution was calculated based on four generations of each individual pedigree.

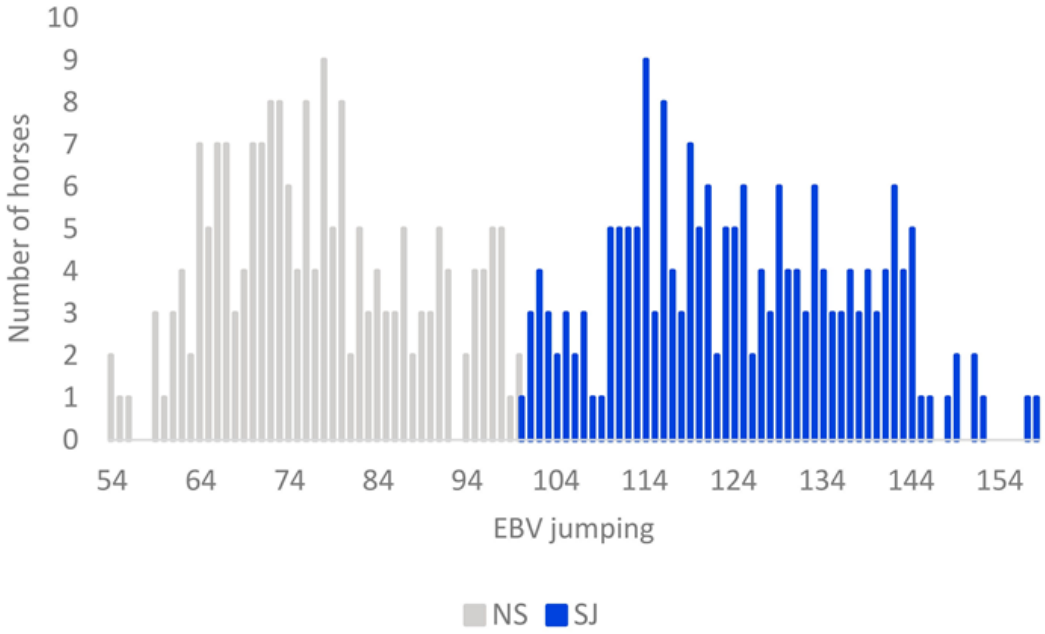

Figure 15 Distribution of estimated breeding values (EBV) for jumping performance in the 380 Swedish Warmblood horses included in this study. The distribution of EBVs for the horses assigned to the subpopulation non-show jumping horses (NS) are shown as grey bars, and, for show jumping horses (SJ), as blue bars.

Because some degree of preselection of horses shown at young horse tests can be expected, a comparison was made to assess if the 380 horses were representative of the SWB cohort at population level. The equality of mean EBV for show jumping between the STP, and all 1540 horses tested the same years was tested (2013–2014) (total tested population, TTP), as well as all 8273 horses born in the same year's cohort (reference population, RP). Likewise, the equality of mean EBV between the two subpopulations of STP horses (SJ and NS) was tested in SAS 9.4 [69].

3.2.2 DNA isolation

DNA was prepared from 20 hair roots, cut into 5% Chelex 100 Resin (Bio-Rad Laboratories, Hercules, CA, US), and 1.4 mg/ml Proteinase K (Merck KgaA, Darmstadt, Germany) in a total volume of 200 µl. The samples were incubated at 56°C, 1500 rpm for 2 hours, followed by heat inactivation of Proteinase K at 96°C for 10 minutes. DNA concentration was normalised, and the DNA was re-suspended in lowTE (1mM Tris, 0.1mM EDTA) at a concentration of 10 ng/µl.

3.2.3 Comparison between SWB and Exmoor ponies

3.2.3.1 Genotyping, quality control and inbreeding coefficients

All samples were genotyped by using the 670K Affymetrix® Axiom® Equine Genotyping Array (Thermo Fisher Scientific, Santa Clara, CA, USA) [142]. Genotype information (670k SNP array data) of 280 Exmoor ponies was retrieved for comparison from a previous publication, which includes details about the ponies' selection procedure for genotyping [141]. Individual inbreeding coefficients (f_i) were estimated from loss of heterozygosity using the PLINK–het command and horses with an f_i higher than 0.05 were excluded for further analyses. Quality Control (QC) was performed separately for each breed on the 31 autosomal chromosomes. The exclusion of poorly genotyped and faulty data was performed using PLINK v1.90 [78] based on the following criteria: minor allele frequency (MAF) (<0.01), missing genotype per single SNP (GENO) (>0.10), missing genotype per individual (>0.10) and Hardy-Weinberg equilibrium (HWE) (p<0.0001). A linkage disequilibrium pruning was applied for the ROH analyses. SNPs in linkage disequilibrium (LD) were excluded if the LD between each pair of SNPs was greater than 0.5 (r^2>0.5) in a window size of 50 SNPs moving 5 SNPs per window.

3.2.3.2 Definition of ROH

The ROHs were detected in SWB and Exmoor separately, by using a sliding-windows approach through the homozyg command in PLINK v1.90. As different signatures of selection are expected when using various ROH definitions, we classified ROH as short (<125kbp), medium (125kb to 500kbp) and long (>500kbp). The three-length classes were defined by the following criteria: minimum number of SNPs, minimum SNP density (SNP/Kbp), maximum gap between two SNPs (bp) and the minimum length to define a ROH (Kbp). The sliding window was defined by using the options homozyg-window-snp, homozyg-window-missing and homozyg-window-het in PLINK v 1.90 (Table 7). The last two options were only applied when detecting long stretches of homozygosity, thus allowing for one heterozygous and one missing SNP [53]. In total 249,395 SNPs were used for ROH analysis.

Table 7 Procedures used to define short, medium and long ROH and the sliding windows' parameters after QC.

Type	Criteria to define ROH				Sliding window		
	Num. SNPs	Density (SNP/kbp)	Max gap (bp)	Length (kbp)	Size (kbp)	Het.[1] SNPs	Miss.[2] SNPs
Short	3	1/100	5000	50-125	5	0	0
Medium	10	1/100	5000	125-500	12	0	0
Long	30	1/100	5000	>500	50	1	1

[1]Het: Number of heterozygous SNPs allowed in the sliding window. [2]Miss: Number of missing SNPs allowed in the sliding window.

The differences in the number of ROHs in each ROH length-class between the two breeds were tested by a one-way analysis of variance in R (v3.4.0) [84]. Proportional differences of ROH between breeds and uniformity of ROH over each respective chromosome were tested with the Chi square test (χ^2) for proportions and goodness of fit in R.

3.2.3.3 ROH as genomic signatures of selection within breed

A custom-made script in R was used to filter homozygous regions within long ROH shared by more than 85% of the studied individuals within breed. The use of 85% as threshold was chosen to identify regions containing fundamental loci for breed-type shared by most SWB horses, regardless of which equestrian sport discipline they were bred for, and thus enable detection of important loci for general sport performance. The EqCab2 genomic coordinates of these regions were used to retrieve candidate gene lists and annotations from the Biomart web interface in Ensembl release 94 [85]. Genetic co-expression, physical interactions or shared protein domains, were visualised by GeneMANIA in Cytoscape with human genes as reference [143]. Additionally, genes were compared with QTL regions previously identified and present in the Horse QTL database [87]. Statistical overrepresentation test of biological processes (GO terms) of candidate genes was conducted using PANTHER 14.0 (http://pantherdb. org/) [144].

The level of significance for the overrepresented biological processes was set as $p < 0.05$.

3.2.3.4 Population differentiation tests: SWB vs Exmoor ponies

The genetic differentiation between SWB and Exmoor ponies was verified by the fixation index (F_{st}) as defined by Nei (1987) [145]. To identify highly differentiated regions, we divided the genome into non-overlapping 500 Kbp windows. A F_{st} value was calculated for each SNP and the values were then averaged over the SNPs located in each window. Windows located at the extreme 1% of the empirical distribution of F_{st} values were considered as candidate regions [146].

For analysis of population differentiation at haplotype level, haplotypes were phased using Shape-it software [147] and filtered using *REHH* Package in R [148], resulting in 503,829 SNPs. We then used XP-EHH statistics to identify regions displaying significantly higher, or lower, extended haplotype homozygosity in one population compared to the other [130]. A position was considered under selection if the XP-EHH in two population-pairwise comparisons was above 4.34 or below -4.34 as suggested by Sabeti et al. (2007) [130]. To minimise the number of false positives and to account for large variation, selection signatures were averaged over windows. Only regions containing SNPs falling into the upper 99^{th} percentile, defined as 500 Kbp windows, and containing at least three SNPs were considered as putative signatures of selection [149]. A haplotype bifurcation diagram around a core marker was used to visualise signatures of selection indicated from the XP-EHH.

Furthermore, genes located in genomic regions containing significant SNPs based on both F_{st} and XP-EHH analyses were retrieved using Biomart, as previously described, and potential overlaps with QTLs, present in the Horse QTL database [87], were considered. Statistical overrepresentation test of biological processes (GO terms) of the candidate genes was conducted using PANTHER 14.0 as described above.

3.2.4 Comparison within SWB horses selected for different sport disciplines

3.2.4.1 Genotyping and quality control

Thanks to the availability of a new reference genome when this part of the study was performed, the 670K SNPs were remapped from the former reference genome EquCab2 to EquCab3 [150] using a Python script, as described in Beeson et al. (2019) [77]. Only SNPs located on the 31 autosome chromosomes were retrieved and used (606,287 SNPs). The quality control (QC) was performed in PLINK (v1.9) [78] and the applied filters were the same as for the comparison between SWB and Exmoor ponies: removing SNPs with a call rate lower than 0.90, MAF < 0.01 and Hardy–Weinberg equilibrium (HWE) deviation with $p < 0.0001$. All individuals had a call rate higher than 0.90. To determine population structure, a linkage disequilibrium (LD) pruning was applied, excluding SNPs if the LD between each pair of SNPs was greater than 0.5 ($r^2 > 0.5$) in a window size of 50 SNPs moving 5 SNPs per window [151].

3.2.4.2 Genomic structure of the SWB population

To assess and describe the potential genetic stratification in the STP of SWB horses, we used PCoA and DAPC. In the case of PCoA, the workflow was as follows: (i) Calculation of pairwise genomic distances between the horses were performed in PLINK (v1.9); (ii) The calculated genomic kinship coefficients were transformed to squared Euclidean distances, and the dissimilarities between the subjects within the matrix were captured in n–1 dimensional spaces of n observations (eigenvectors), via classical multidimensional scaling (MDS) [152] in R [153]; (iii) The proportion of variation captured by each eigenvector was calculated and the two eigenvectors explaining the largest proportion of the variance were plotted in R as a principal component plot.

The DAPC [154] was performed using the *adegenet* package [155] in R, to further evaluate the presence of two subpopulations in the STP of SWB horses. The Bayesian Information Criterion (BIC) analysis was used to determine the number of subpopulations (K) in the STP. The number of principal components (PCs) to retain in

the discriminant step in the DAPC was optimized using the cross-validation procedure in *adegenet*, where the dataset was divided into two sets selected by stratified random sampling: a training set (90% of the data) and a validation set (10% of the data). The optimal number of PCs was chosen based on the mean successful assignment of the predefined subpopulations (SJ and NS) and lowest root mean squared errors.

3.2.4.3 Divergent selection based on SNPs

The genetic divergence between SJ and NS was verified by the fixation index (F_{st}), as defined by Nei (1987) [145]. Measures of centrality and dispersion were used to compare F_{st} values for each SNP. Negative F_{st} values were set to zero, because negative values have no biological interpretation [156]. Values were interpreted by using the qualitative guidelines proposed by Wright (1978) [157], where an F_{st} value of 0.15–0.25 indicates large differentiation, 0.05–0.15 indicates moderate differentiation, and F_{st} < 0.05 indicates little differentiation among populations [158]. SNPs were plotted relative to their physical position within each autosome using a custom-made script in R. SNPs with an F_{st} value within the top 0.1% of the F_{st} distribution were considered as potential signatures of selection [159].

3.2.4.4 Divergent selection between subpopulations based on haplotypes

Haplotypes were phased using Shape-it software [147] and filtered using *REHH Package* in R to (1) discard haplotypes with missing data, (2) keep only fully genotyped markers [148]. Regions homozygous in one population but polymorphic in the other population were highlighted by the comparison of EHH score of the two subpopulations per SNP [130]. The XP-EHH score was computed for each autosomal SNP and it was defined and standardized according to [130, 148, 160].

The XP-EHH score was transformed into pXP-EHH, as shown by [160]:

$$p_{XP\text{-}EHH} = -log_{10}1 - 2\,\Phi(XP\text{-}EHH) - 0.5$$

where Φ(XP-EHH) is the Gaussian cumulative distribution function. The pXP-EHH can be interpreted as a two-sided p-value in a −log10 scale. Negative XP-EHH scores indicate that selection happened in SJ horses, while positive values indicate selection in the NS horses [50]. The False Discovery Rate (FDR) adjustment was performed, applying the Benjamini Hochberg method with FDR equal to 0.05 [161].

3.2.3.5 Candidate genes

A genomic region was considered as under potential selection if it contained significant SNPs based on both F_{st} and XP-EHH analyses. We used the Ensembl gene annotations EquCab3 to identify genes residing within regions extending 250 kb up - and downstream of significant SNP. This was done to include potential effects of regulatory changes on loci at some distance, and to reduce the risk of excluding the outermost parts of the selected haplotypes [85]. Functional classification, statistical overrepresentation of biological processes and pathways (GO terms) of candidate genes for each subpopulation were conducted using PANTHER 14.0 (http://pantherdb.org/) [144] on the Equine reference genome EquCab 3.0. Genes were further compared with previously identified QTL regions in the horse QTL database [87].

3.3 RESULTS

3.3.1 Comparison between SWB and Exmoor ponies

3.3.1.1 Genotyping, quality control, and inbreeding coefficients

The total genotyping call rate was 0.99 in SWB (n=380), and 0.97 in Exmoor ponies (n=274). The LD measured by the squared correlation coefficient (r^2) showed significant mean differences (p<0.005) between the two populations. In SWB horses, the average r^2 was 0.56 at 5 kb distance, compared to 0.60 in the Exmoor ponies. The estimated inbreeding coefficient based on loss of heterozygosity (f_i) in the 380 SWB ranged from -0.134 to 0.098 with an average value of 0.006. The f_i values in the 274 Exmoor ponies ranged from -0.378 to 0.530, with an average of 0.170. Overall, 129 Exmoor ponies and six SWB horses were excluded from further analyses as f_i exceeded the threshold of 5% set in this study.

3.3.1.2 ROH as genomic signatures of selection

The average number of both short (<125 Kbp), medium (125-500 Kbp) and long (>500 Kbp) ROH per individual was higher in the Exmoor ponies than in the SWB horses (p<0.0001); short ROHs were five times more abundant in the Exmoor ponies than in SWB horses (Table 8). The average number of short ROHs per animal was twice as high when including all the 274 Exmoor ponies that passed the QC, compared to when only including the 145 ponies with a f_i <5% in the analysis, which supported the use of the f_i threshold. Short and medium ROHs were distributed equally along the genome in both SWB and Exmoor ponies. Long ROHs were the rarest ones as shown in Table 8; nevertheless, such ROH covered, on average 286 Mbp in SWB horses and 1,182 Mbp in Exmoor ponies. While long ROHs in Exmoor ponies were spread across the genome, long ROHs in the SWB population were found in a few chromosomes only.

Table 8 Descriptive summary of ROH identified following three procedures (short, medium and long ROH detection) per breed: SWB horses and Exmoor ponies.

Breed	Short ROH			Medium ROH			Long ROH		
	Total N.[1]	N.[2]	L. (Kbp)[3]	Total N.	N.	L. (Kbp)	Total N.	N.	L. (Kbp)
SWB (N= 374)	257,419	688	76	374,665	1,001	445	184,846	494	580
Exmoor pony (N=145)	515,207	3,553	82	171,149	1,180	260	96,765	667	1773

[1] Total N.: Total number of ROHs detected in the population, [2] N.: average number of ROH per individual calculated as the Total N. divided by the number of individuals: 374 in the SWB horses and 145 in the Exmoor Ponies, [3] L.: average length of ROH expressed in Kbp.

The same pattern was discovered when filtering long homozygous regions shared by at least 85% of SWB horses and Exmoor ponies, respectively. The 65 long shared ROHs detected among SWB horses contained overlapping homozygous segments in five chromosomes: ECA4, ECA6, ECA7, ECA10 and ECA17 (Figure 16 a). In Exmoor ponies, 398 long shared homozygous segments were instead distributed over 24 out of 31 (77%) autosomal chromosomes (Figure 16 b).

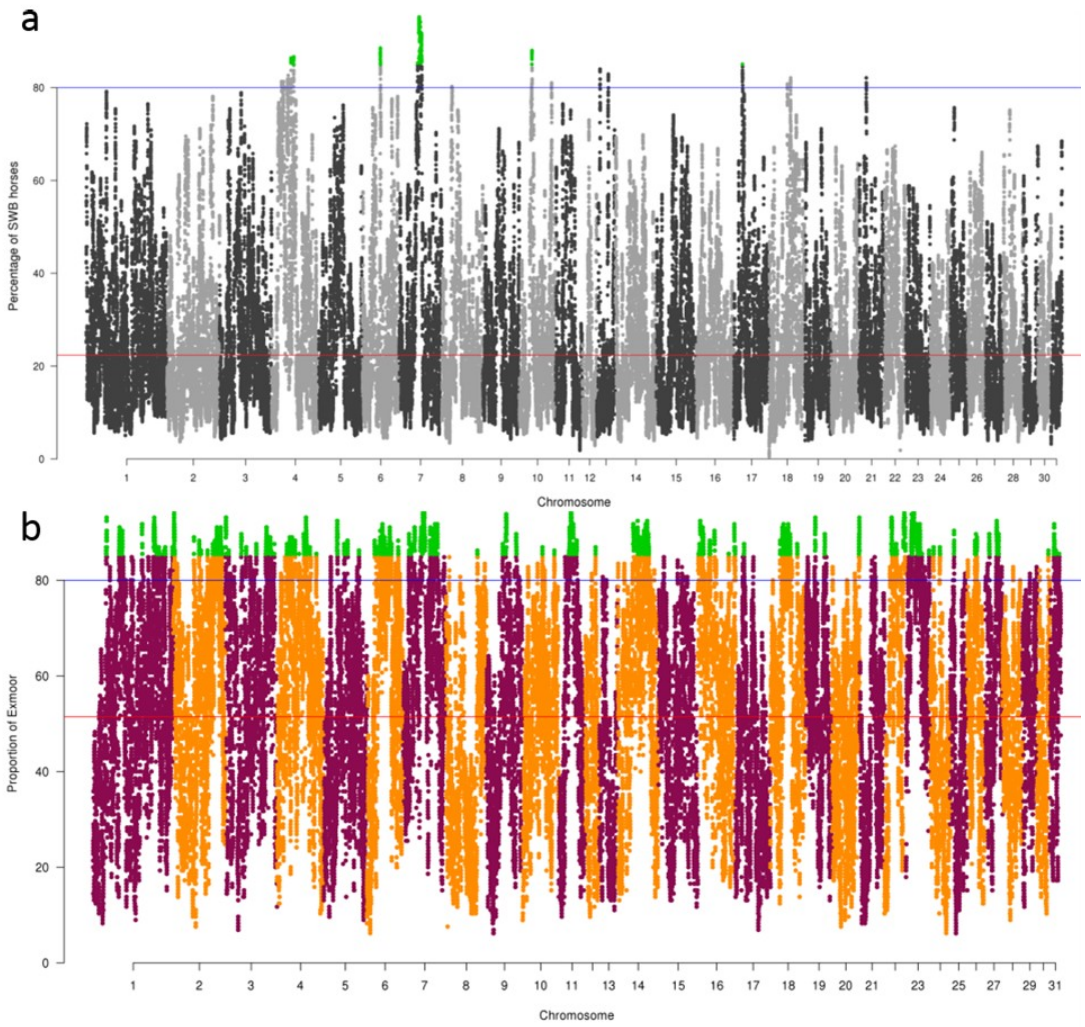

Figure 16 Incidence of each single nucleotide polymorphism (SNP) in ROH in the SWB horses and Exmoor ponies. Genomic positions highlighted in green represent SNPs in a homozygous segment shared in over 85% of the SWB horses (a) and Exmoor ponies (b). The blue line shows the threshold for SNPs present in more than 80% of the horses and the red line shows the average value (22%) of SNP incidence in homozygous segment in SWB horses and (51%) in Exmoor ponies.

The longest shared homozygous segment within ROH in SWB horses was 0.28 Mbp and was located on ECA7:42,688,962-42,905,689, whereas in Exmoor ponies the longest shared homozygous region was 0.37 Mbp on ECA22:46,333,459-46,708,156 (Figure 17).

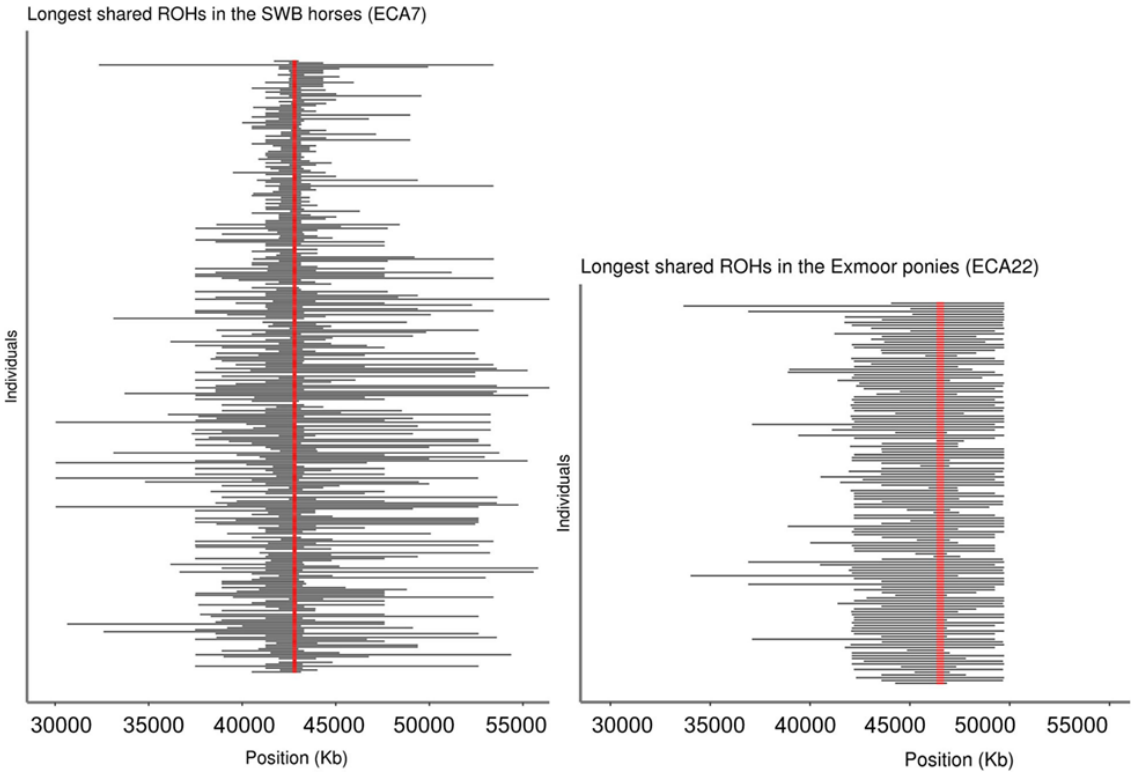

Figure 17 Interval of the longest shared homozygosity in SWB horses and Exmoor ponies. ROH length per each individual is indicated as horizontal black line. The red line indicates the shared interval among over 85% of the population.

The exact overlap in a ROH in ECA7 (7:49,160,767:49,212,921) coincided with a known QTL regions for body size (height at withers) [95]. The homozygous long ROH segments, shared by more than 85% of the studied Exmoor ponies, harboured 265 genes (Supplementary Table 1). In SWB, only 21 genes were located in the overlapping homozygous segments detected by long ROH analysis. Eighteen of them were annotated in the reference genome EquCab2 release 94 [144, 162], where ten had an annotated function, and eight were novel genes. The gene position, name and percentage of horses sharing an exact overlap in a ROH are listed in Table 9.

Table 9 Genes found in overlapping homozygous segments within long ROH in over 85% of the SWB horses.

Exact overlap in a ROH*	Gene symbol	Gene name	% of SWB
4:44,468,835:44,656,577	*THSD7A*	Thrombospondin type 1 domain containing 7A	86.0%
4:50,825,457:50,825,457	*HDAC9*	Histone deacetylase 9	86.6%
6:41,324,520:41,661,196	*GRIN2B*	Glutamate ionotropic receptor NMDA type subunit 2B	88.0%
7:42,688,962:42,905,689	*B3GAT1*	Beta-1,3-glucuronyltransferase 1	85.0%
	ENSECAG00000009503	Novel pseudogene	85.0%
	ENSECAG00000003683	Novel gene	85.0%
7:45,280,925:45,280,925	*IER2*	Immediate early response 2	91.0%
7:44,827,476:45,014,430	*CACNA1A*	Calcium voltage-gated channel subunit alpha1 A	93.6%
7:45,525,088:45,525,088	*DNASE2*	Deoxy-ribonuclease 2, lysosomal	94.0%
7:47,244,532:47,445,238	*ENSECAG00000010696*	Novel gene	90.4%
	ENSECAG00000012790	Novel gene	90.4%
	ENSECAG00000004301	Novel gene	91.2%
7:48,419,051:48,419,051	*ENSECAG00000020829*	Novel gene	91.2%
7:49,160,767:49,212,921	*SMARCA4*	SWI/SNF related, actin dependent regulator of chromatin	86.0%
7:50,250,257:50,286,784	*ENSECAG00000015373*	Novel gene	85.8%
	ZNF699	Zinc finger protein 699	89.0%
7:51,068,391:51,107,956	*ENSECAG00000002451*	Novel gene	91.4%
7:51,514,450:51,560,980	*ENSECAG00000002839*	Novel gene	91.2%
10:29,104,190:29,163,955	*ENSECAG00000003641*	Novel pseudogene	88.0%
	ENSECAG00000025061	Novel pseudogene	88.0%
17:19,157,873:19,157,873	*NEK5*	NIMA related kinase 5	85.1%

Only one of the ten annotated genes located in shared ROH in SWB were included in a network based on co-expression, physical interactions or shared protein domains (Figure 18).

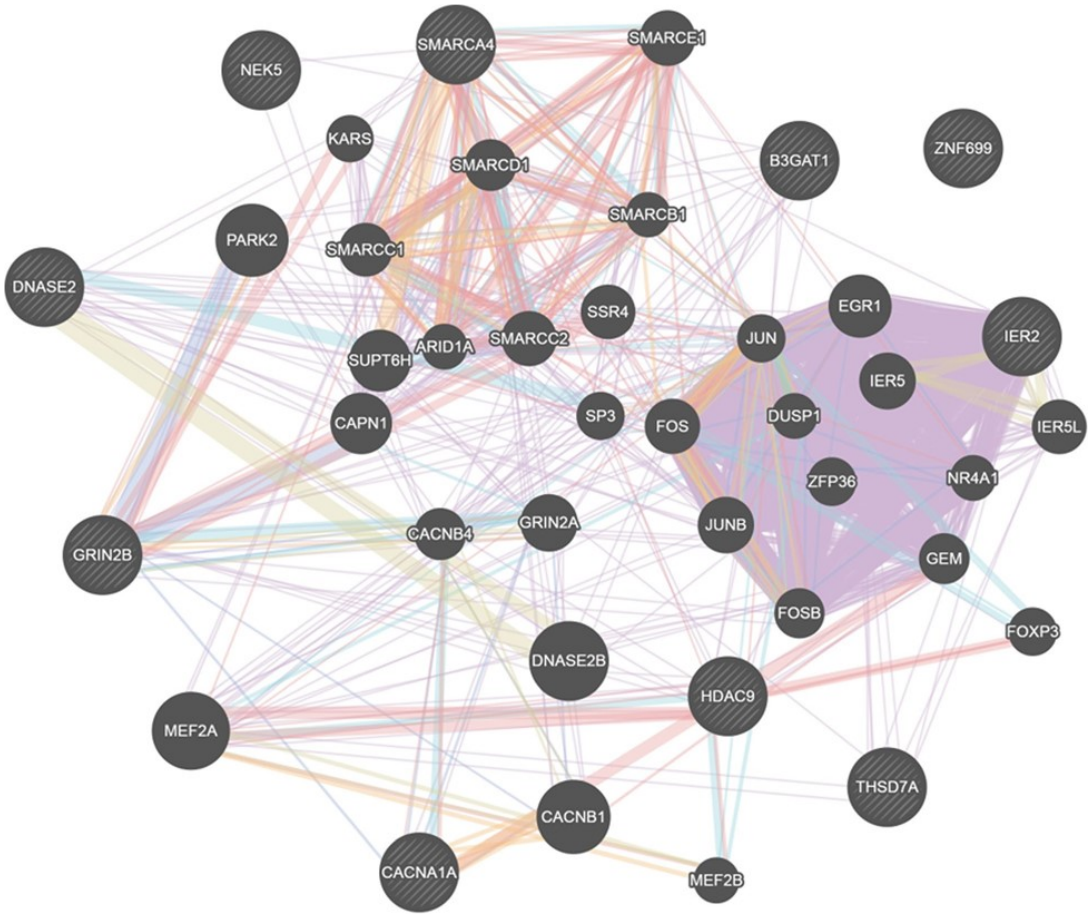

Figure 18 GeneMANIA representation of the genes found in the shared ROH in the SWB horses. The ten annotated genes are represented as stripped grey circles. Physical interactions are displayed as red lines, co-expressions as violet lines, predicted related genes as orange lines, shared pathways as light blues lines, co-localisations as blue lines and genetic interaction as green lines. The three most related genes with the potentially under selection ones are shown as plain circle.

From the set of candidate genes five significantly overrepresented biological processes were found; actin cytoskeleton reorganization (GO:0031532), cellular macromolecule catabolic process (GO:0044265), apoptotic process (GO:0006915), glycoprotein metabolic process (GO:0009100) and synaptic signalling (GO:0099536) (Supplementary Table 1).

3.3.1.3 Population differentiation tests: SWB vs Exmoor ponies

The average F_{st} for all windows between SWB and Exmoor ponies was equal to 0.11 with a standard deviation of 0.04. The distribution of F_{st} estimates per SNP is shown in Figure 19.

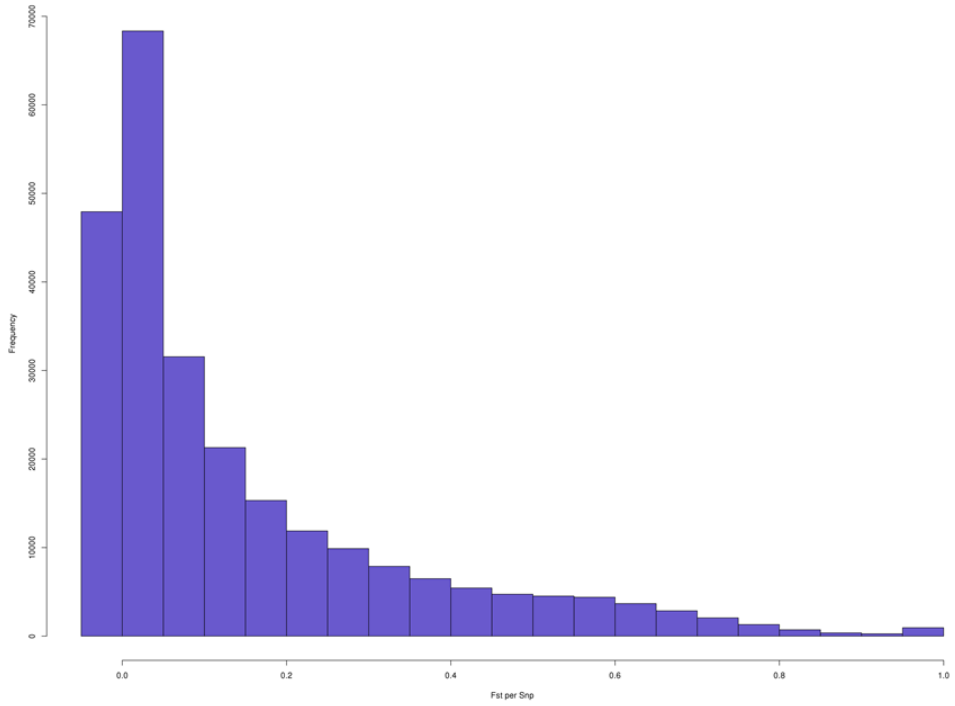

Figure 19 Distribution in frequency class of Fixation index (F_{st}) between SWB horses and Exmoor ponies.

Average F_{st} per window for a total of 126 windows exceeded the significance threshold, corresponding to an averaged F_{st} value ≥ 0.24 (Figure 20).

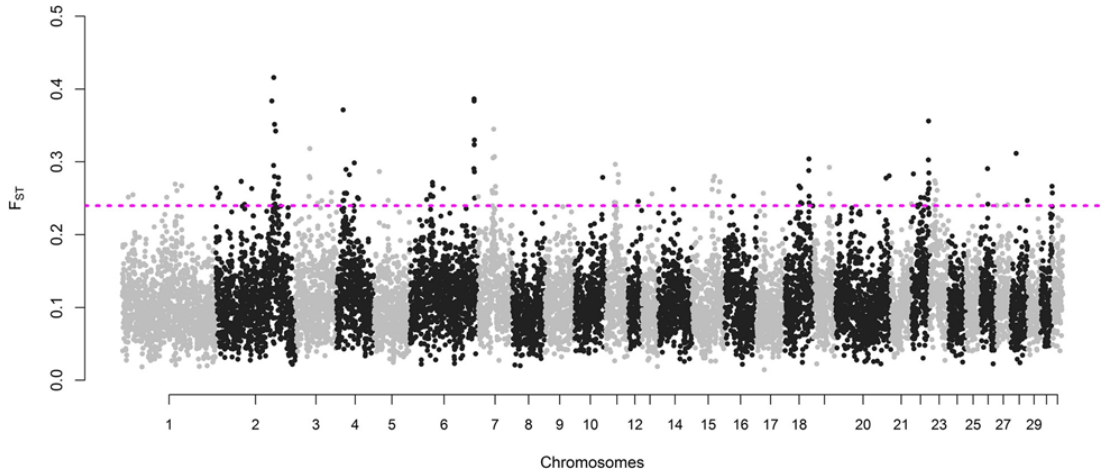

Figure 20 Genomic distribution of averaged F_{st} values in 500-kb windows plotted relative to their physical position within each autosomal chromosome. The cut-off to call a SNP as significant was defined as the highest 1% of the empirical distribution and is represented by the SNPs above the dotted pink line ($F_{st} \geq 0.24$).

The top 10 windows with highest F_{st} were located on ECA2, ECA4, ECA6, ECA7 and ECA22. Genomic regions with an XP-EHH value above 4.34 and thus potentially under positive selection in SWB horses were detected on ECA1, ECA2, ECA4, ECA6, ECA17 and ECA26. The complete list of markers underlining potential signatures of selection is presented in Supplementary Table 3. Eight signatures of selection located on ECA1, ECA4, and ECA6 in SWB horses remained significant after correcting for false discovery rate (Figure 21).

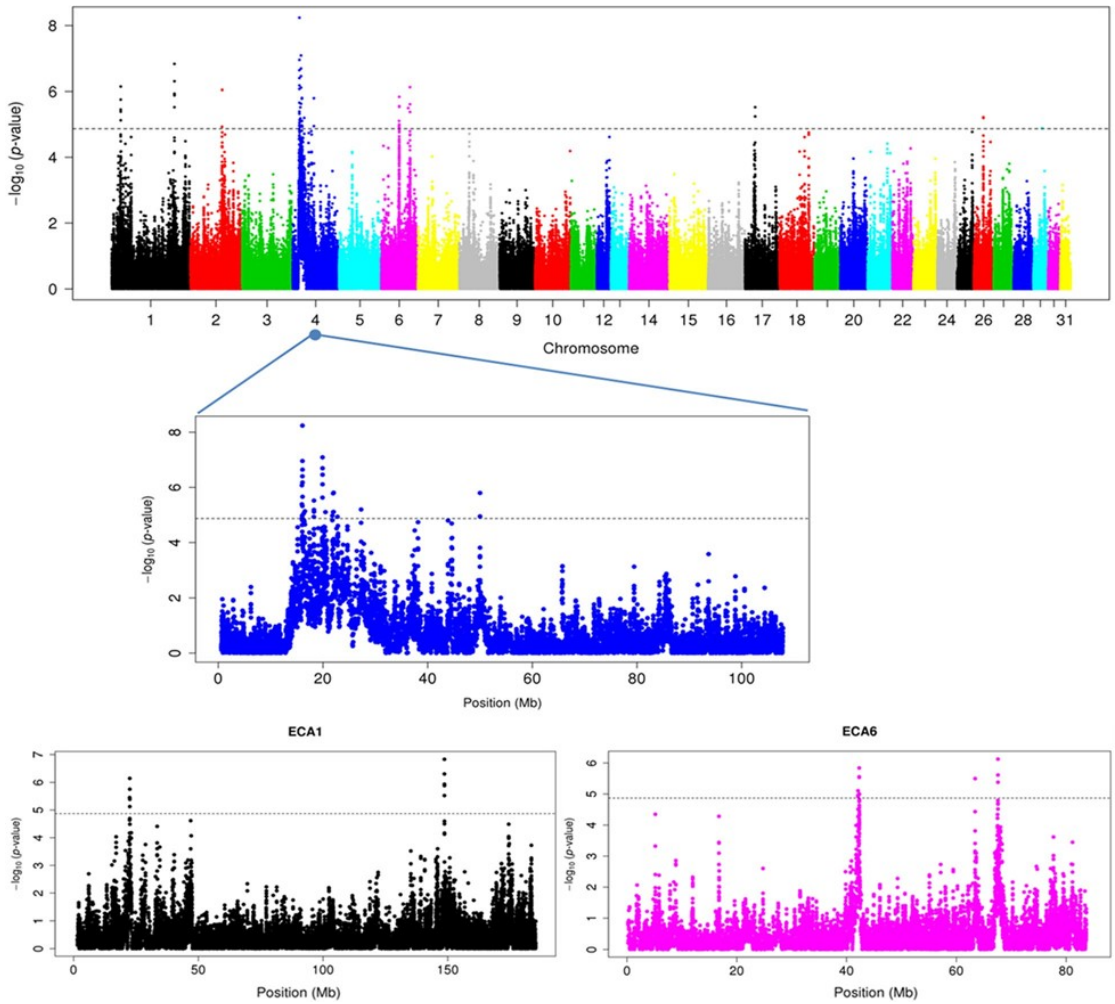

Figure 21 Genome scan of potential regions under selection detected by the cross populations extended haplotype homozygosity (XP-EHH) test. The log10(p-value) is plotted for each SNP per chromosome (top panel). The chromosomes ECA1, ECA4 and ECA6 contain significant SNPs, are zoomed in and displayed in the middle and lower panels.

The breakdown of linkage disequilibrium from the two SNPs with the highest XP-EHH estimate (XP-EHH=5.82 for ECA4:16,077,767 and XP-EHH=5.36 for ECA4: 19,918,093) is shown in Figure 22. No position with an XP-EHH value lower than -4.34 was found in Exmoor ponies, indicating a lack of recent selection.

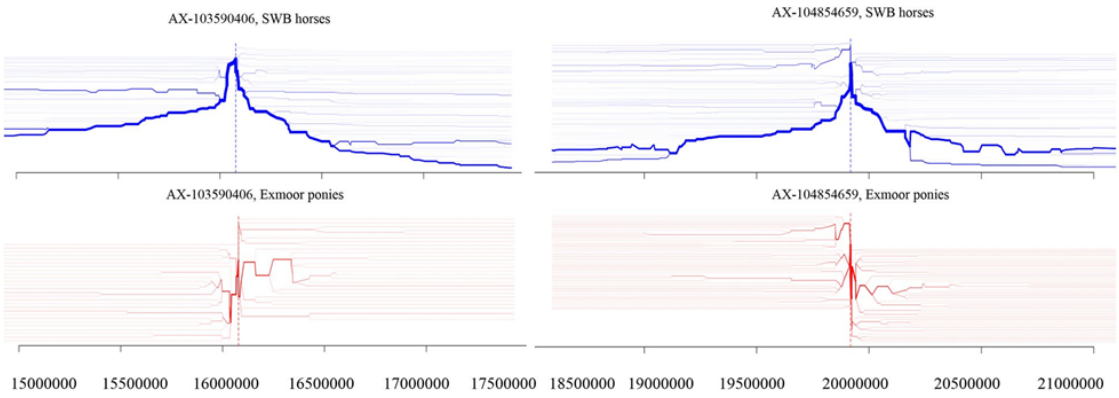

Figure 22 Bifurcation diagram for the two focal SNPs on chromosome 4: position 16,077,767 and 19,918,093. The two core SNPs were defined as the two with the highest XP-EHH result from the cross populations extended haplotype homozygosity test (XP-EHH=5.76 and XP-EHH=5.36). The two plots represent the LD breakdown at increasing distances from the core allele at the selected focal SNP in the SWB horses (on top and represented as blue lines) and in the Exmoor ponies (at the bottom and represented as red lines).

Four genomic regions were detected by both F_{st} and XP-EHH population differentiation tests. The genes located within those regions were considered as potentially under selection in SWB horses (Table 10).

Table 10 Genomic regions detected by both F_{st} and XP-EHH which were further considered as signs of selection.

ECA	Position (bp)			Average XP-EHH[1]	Average F_{st} [2]	Genes
1	148,660,078	:	149,160,078	4.67	0.26	*FSIP1, THBS1*
4	15,977,755	:	16,477,755	5.82	0.37	*ADCY1, IGFBP1, IGFBP3*
4	19,913,037	:	20,413,037	5.09	0.26	*ZPBP, SPATA48, IKZF1, FIGNL1, DDC, GRB10*
6	42,096,241	:	42,596,241	4.81	0.25	*GUCY2C, WBP11, SMCO3, C6H12orf60, ART4, MGP, ERP27, ARHGDIB, PDE6H, RERG*

[1]XP-EHH and [2]Fst averaged over 500-kb windows.

The region on ECA 6 (6:42,096,241:42,596,241) overlapped with a QTL associated with body conformation in horses [163]. Six overrepresented biological processes were found from the list of potentially-selected genes; cellular response to stimulus (GO:0051716), ribonucleotide biosynthetic process (GO:0009260), signal transduction (GO:0007165), regulation of cell communication (GO:0010646), peptidyl-arginine modification (GO:0018195) and mRNA cis splicing, via spliceosome (GO:0045292) (Supplementary Table 4) .

3.3.2 Comparison within SWB horses selected for different sport disciplines

3.3.2.1 Genotypic and phenotypic population structure of SWB horses

From the SNPs genotyped on the 670 K SNP-chip, 606,287 autosomal SNPs were retained after remapping to EquCab3. The total genotyping call rate was equal to 0.96 and no single individuals showed a call rate lower than 0.90, leaving all 380 SWB horses, and 456,381 SNPs in the analyses after QC. The LD pruning left 242,396 SNPs to be used in the PCoA and DAPC analyses. No significant differences in the mean show jumping EBV were found from the comparisons between STP and RP and STP and TTP ($p > 0.80$) (Figure 23).

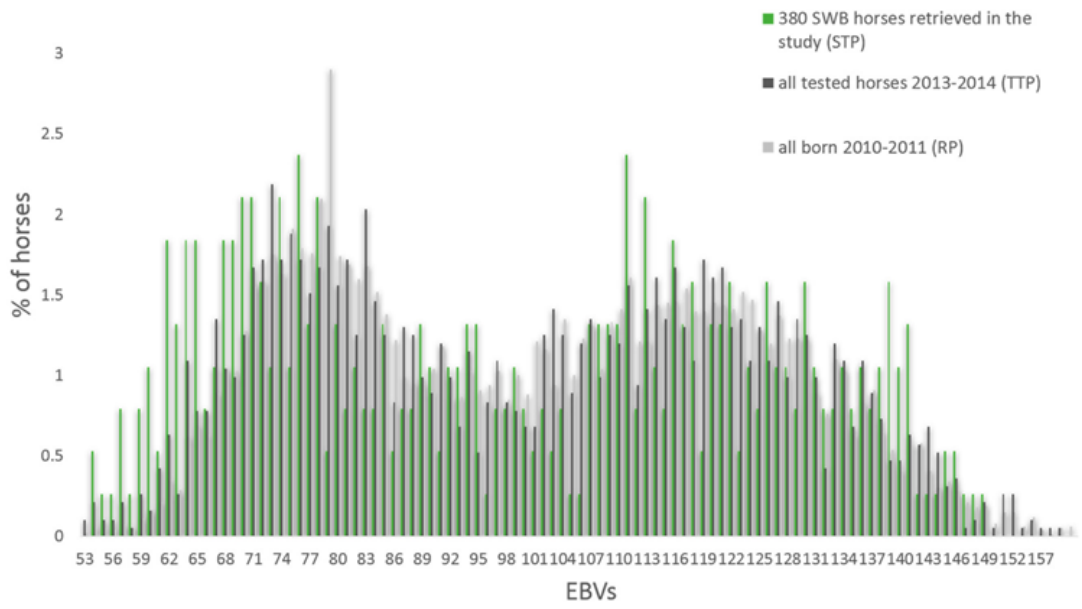

Figure 23 Percentage of horses with different estimated breeding values (EBVs) for show jumping in the studied horses (STP), all horses tested in young horse tests the same years (TTP), and all horses born in the same years' cohort (RP).

A total of 191 horses had an EBV for show jumping above 100 and were defined as show jumping horses (SJ), whereas the other 189 horses were defined as non-show jumping horses (NS). The average EBV for show jumping was significantly higher in SJ horses (125) compared to NS horses (77) ($p < 0.001$), corresponding to more than two genetic standard deviations (Table 11). Similarly, the average score for jumping technique at the young horse evaluation was significantly higher in SJ (8.3) than in NS horses (5.6) ($p < 0.001$). On the other hand, no significant difference between the two subpopulations was shown for mean test score for gaits, or for the conformation traits "type" and "head–neck–shoulder", which are traits not specifically related to show jumping performance. Until October 2018, 155 of the horses had competed in show jumping and 91 in dressage, whereas only 18 had competed in both disciplines. A significantly higher proportion of SJ horses (71%) had competed in show jumping classes compared to NS horses (11%) (Table 11).

Table 11 Comparison between breeding values (EBVs), scored points at young horse evaluation, and proportion of horses competing, in the two assigned subpopulations show jumping (SJ) and non-show jumping (NS) horses.

Trait	SJ Horses	NS Horses
Mean EBV for show jumping performance	125 (13.4) *	77 (11.3) *
Mean EBV for dressage performance	94 (10.2) *	119 (21.0) *
Mean score for jumping technique	8.3 (1.2) *	5.6 (1.2) *
Mean score for gaits **	7.2 (0.6) *	7.3 (0.8) *
Mean score for the conformation trait “type”	7.9 (0.7) *	7.8 (0.7) *
Mean score for the conformation trait “head–neck–body”	7.7 (0.6) *	7.7 (0.7) *
Percentage of horses competing in show jumping at least at regional level	71%	11%
Percentage of horses competing in dressage at least at regional level	10%	38%

* Standard deviation within brackets, ** Mean score of walk, trot and canter.

3.3.2.2 Population Substructure

The PCoA revealed some genetic stratification among SWB horses, with 5.2% of the total variance explained by the three main eigenvalues. The first two eigenvalues explained 2.63% and 1.29% of the total genetic variation of the data, as visualized in Figure 24. The SJ horses presented a more scattered cluster based on the second component (eigenvalue 2), indicating higher genetic distance and lower relatedness among individuals compared to the NS horses. Eleven horses, all sired by the same father, clustered as outliers in the PCoA. No offspring from this sire clustered among the other 369 individuals.

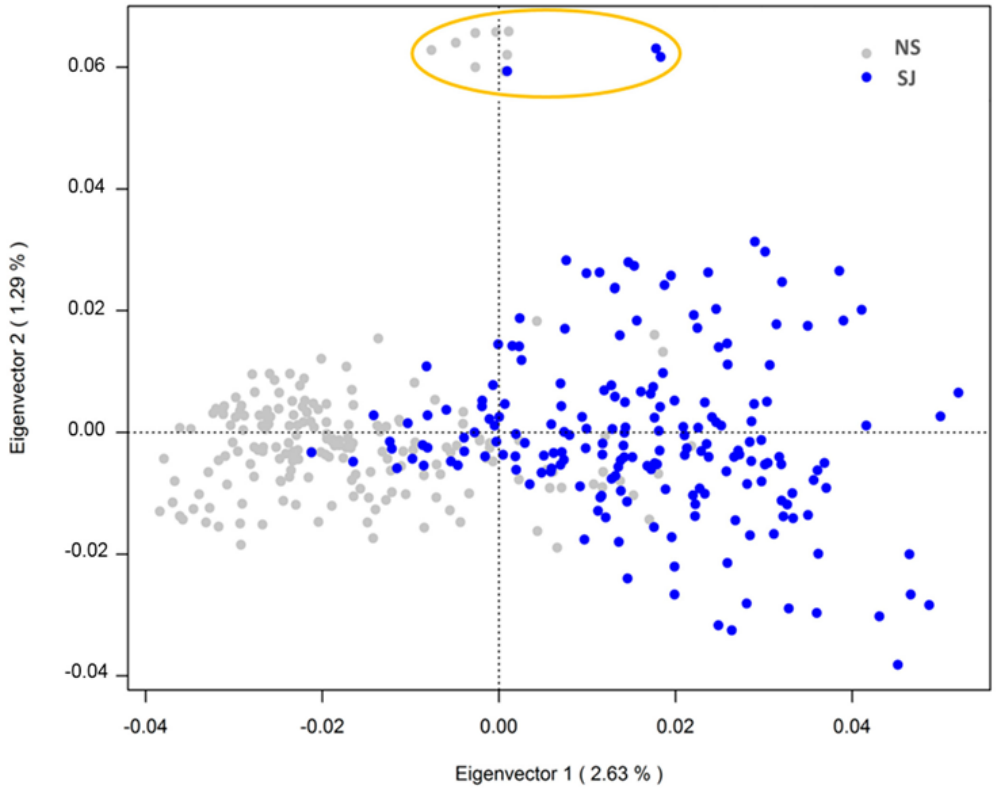

Figure 24 PCoA scatter plot of the first two Eigenvectors explaining genetic differentiation between non-show jumping (NS), which are grey coloured, and show jumping (SJ) horses (blue coloured) in the SWB population. The orange circle indicates the eleven outliers sired by the same stallion.

The BIC analyses based on the genotype data, setting the number of possible subpopulations (K = 1:5), indicated the most likely number of subpopulations was either two or three (Figure 25 a). The presence of two main subpopulations was further supported by the DAPC plot (Figure 25 b). The cross-validation test for the number of

PCs to retain, showed the highest proportion of the successful assignment to the predefined subpopulations (SJ and NS) and lowest root mean squared errors for 100 PCs. The retained 100 PCs explained ~47% of the total variation. The results obtained from the DAPC analysis supported those obtained from PCoA analysis, with most (96%) of the individuals correctly assigned to their predefined SJ and NS subpopulations based on EBVs (Figure 25 b). The 23 individuals in the overlapping area of the density DAPC plot were found to have more than twice as high average Thoroughbred proportion (22.4%) in their pedigree, compared with the remaining 357 horses (9.6%) in this study.

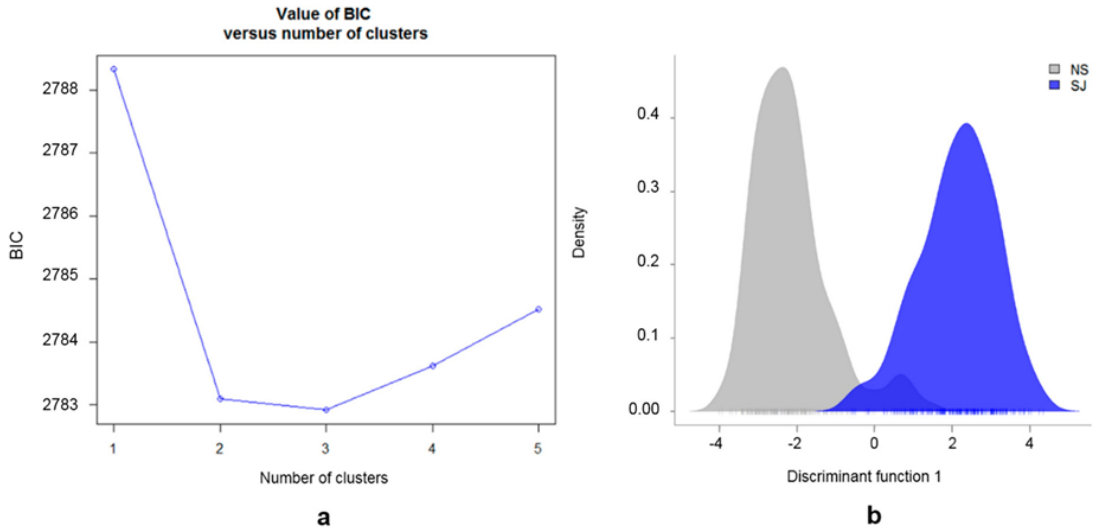

Figure 25 Genetic structure of SWB horses. (a) Inference of the number of clusters in SWB horses from analysis of genotype data based on K–means algorithm; (b) Density plot of individuals along the first discriminant function from the discriminant analysis of principal component (DAPC) for the two defined SWB subpopulations divided by EBVs for show jumping performance. The grey peak represents the non-show jumping horses (NS) and the blue peak represents the show jumping horses (SJ).

In the posterior distribution from the DAPC, the same 23 individuals showed less than 80% likelihood of being assigned to their predefined subpopulations (SJ or NS). All genotyped individuals are presented in Figure 26, where each line represents one individual, and the heat colour represents their membership probability, provided by DAPC, of being assigned to the predefined subpopulations (SJ or NS).

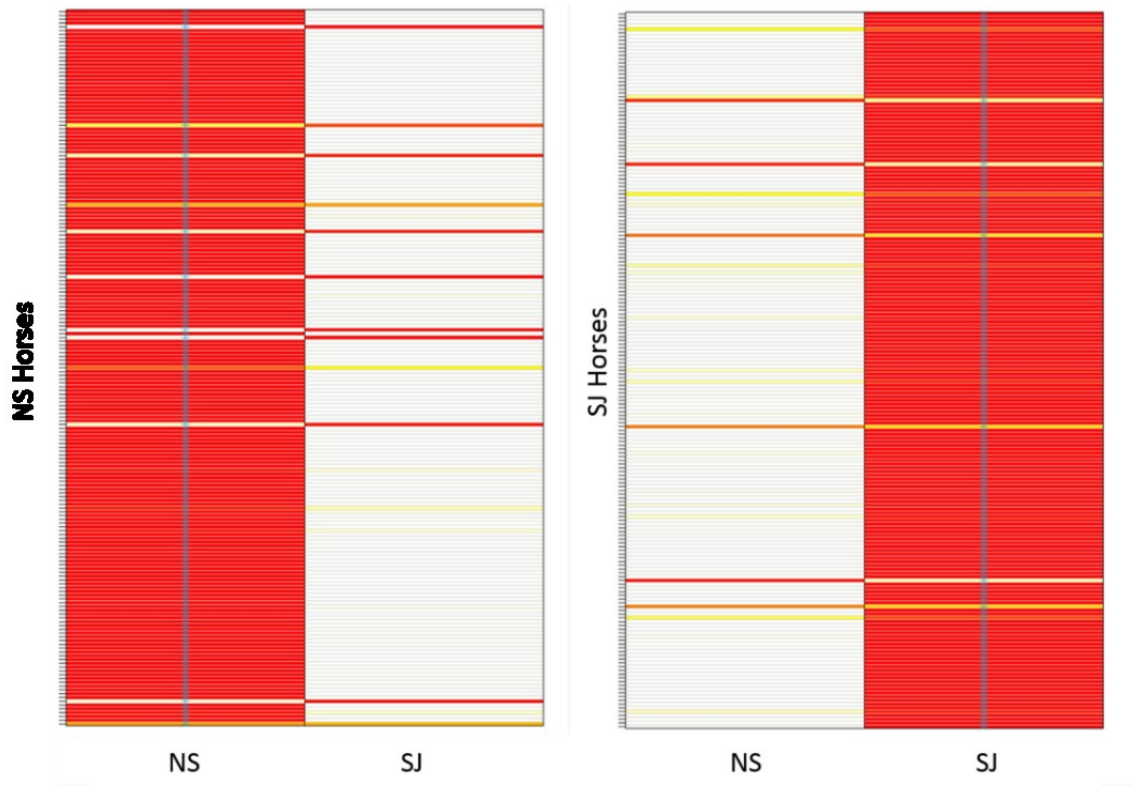

Figure 26 Heatmap showing the posterior probability of each individual being assigned to the predefined subpopulations based on EBVs where non-show jumping horses (NS) are shown in the left panel and show jumping horses (SJ) are shown in the right panel. A red bar represents a membership probability equal to one and a white bar represents a membership probability of zero.

3.3.2.3 Genomic divergence in SWB subpopulations

The average value of fixation index (F_{st}) between SJ and NS horses was 0.015 with a standard deviation of 0.017. A total of 347 SNPs exceeded the 0.1% threshold for significance, corresponding to F_{st} value ≥ 0.12, and were considered as putative signs of divergent selection. These SNPs were distributed in all chromosomes, except ECA24 and ECA29. A total of nine SNPs distributed on ECA1, ECA5 and ECA22, had F_{st} values higher than 0.20 (Supplementary Table 5). The highest number of significant SNPs was obtained for the ECA1 (n = 79, with F_{st} values ranging between 0.12–0.23); however, when normalizing for the number of SNPs present in the chromosomes, the highest number of outliers was found in ECA22 and ECA25, with 34 and 27 SNPs, respectively (Figure 27).

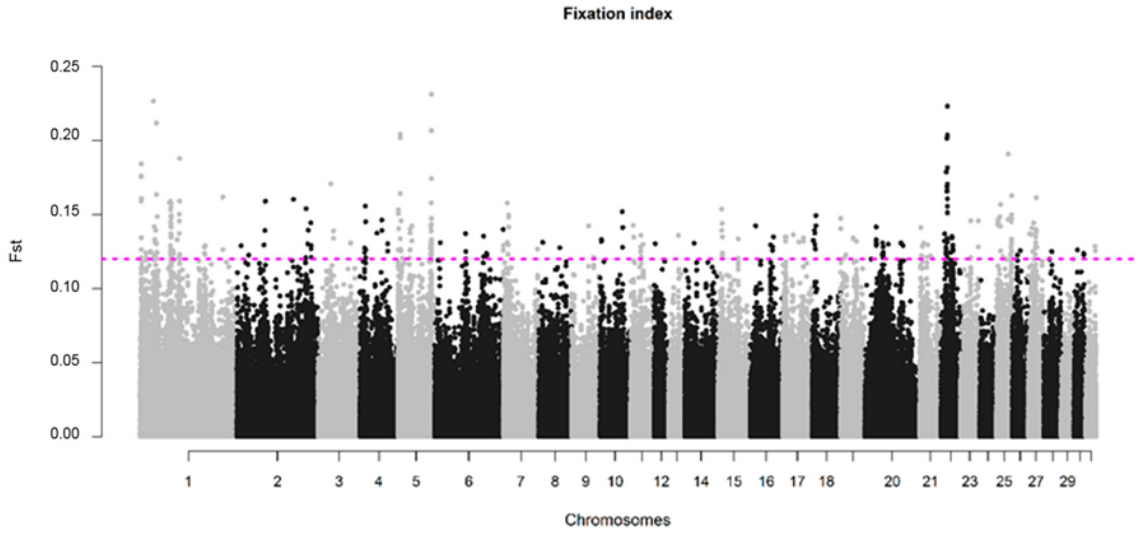

Figure 27 Genomic distribution of F_{st} values for SNPs plotted relative to their physical position within each autosomal chromosome. The cut-off to call a SNP as significant was defined as the highest 0.1% of the F_{st} values of the SNPs under analysis and is represented by the SNPs above the dotted pink line ($F_{st} \geq 0.12$).

3.3.2.4 Signatures of selection in SWB subpopulations

In SJ-horses, the XP-EHH-analyses detected positive selection in eight genomic regions distributed in seven chromosomes: one region per chromosome on ECA7, ECA8, ECA13, ECA19, ECA22, ECA31, and two regions on ECA12. In NS-horses, positive selection was detected in four genomic regions: one on ECA1, one on ECA23 and two on ECA25 (Figure 28).

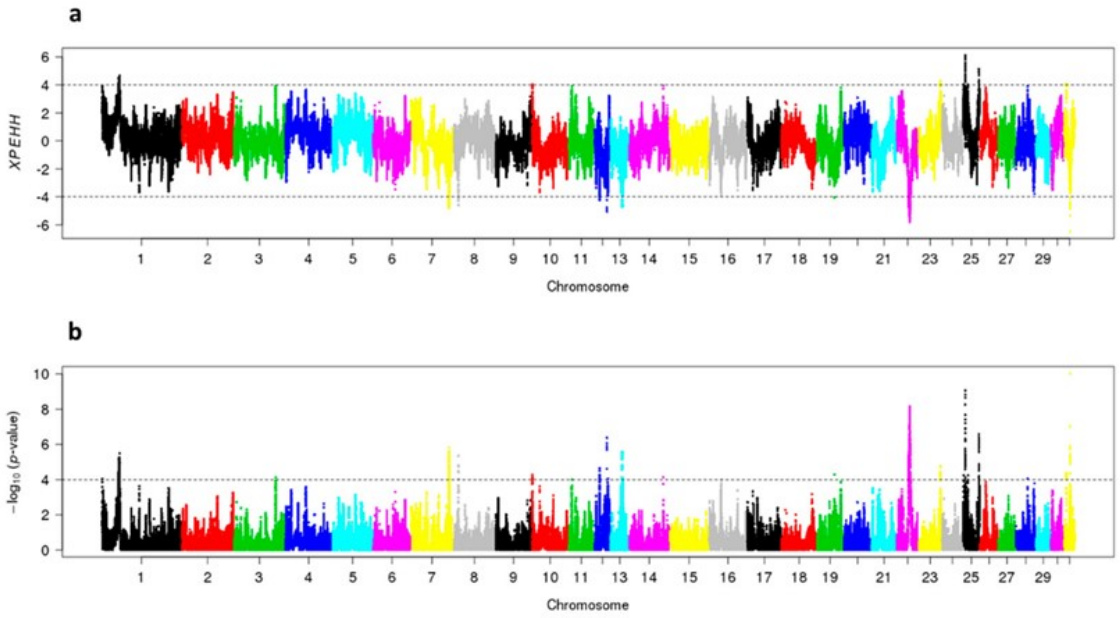

Figure 28 Cross population extended haplotype scores of the show jumping (SJ) and non-show jumping (NS) subpopulations: (a) XP-EHH where positive values indicate positive selection in the NS subpopulation, and negative values indicate positive selection in the SJ subpopulation. An XP-EHH value of +/−4 is considered significant and indicated by a dotted line; (b) log p-values of the XP-EHH values.

The genomic coordinates of the regions under selection, detected by XP-EHH in SJ horses, is shown in Table 12. In SJ horses, two significant regions on ECA8 and ECA31 overlapped with known QTLs for body growth and altitude adaptation. A total of six out of eight regions detected by the XP-EHH test overlapped with at least one significant SNP from the Fst analysis (75% concordance between the two tests). In total, 157 genomics elements were found within the six genomic regions under selection in SJ horses: 114 protein coding genes, one pseudogene, 22 lncRNAs, nine snoRNAs, seven miscRNAs, two miRNAs and two snRNA (Supplementary Table 6).

Table 12 Significant chromosomal regions found by the XP-EHH test in show jumping horses (SJ). If the region overlapped with a significant F_{st}, the highest overlapping F_{st} value within the significant XP-EHH region is shown.

Chromosome	Start Position (bp)	End Position (bp)	Length (kbp)	Lowest Adjusted *p*–Value	Highest F_{ST}
7	89,669,652	91,203,503	2918	0.0022	0.16
8	10,891,925	11,400,093	508	0.0040	0.13
12	11,601,991	12,104,261	502	0.0129	na
12	28,794,818	29,306,185	511	0.0009	0.13
13	27,280,931	27,830,947	550	0.0031	0.14
19	41,459,599	41,959,599	500	0.0247	0.13
22	27,935,550	31,789,708	3854	0.0008	0.14
31	13,778,365	14,321,247	543	0.00002	na

na = not applicable.

The PANTHER overrepresentation analysis recognized 74 genes out of the 114 protein coding genes located within the regions under selection in SJ horses. Eight GO terms were overrepresented for biological processes, as well as four pathways in SJ horses ($p < 0.01$) (Table 13). Sixteen of the 74 genes detected by PANTHER analysis were mainly related to development and regulatory functions within the eight overrepresented biological processes, while six genes were involved in overrepresented pathways.

Table 13 Overrepresented gene ontology (GO) term at p-value < 0.01 for the genes within regions indicated by both XP-EHH and F_{st} to be under selection in SJ horses. GO terms are presented for both for biological processes and pathways.

PANTHER GOs	Genes	Fold Enrichment	Raw *p*–Value
Slim Biological Process			
Cell differentiation (GO:0030154)	*RBL1, DBX1, SRC, NR1I2*	5.68	<0.001
Multicellular organism development (GO:0007275)	*DBX1, DLGAP4, NAV2, MYL9, KIAA1755, VSTM2L, GSK3B, NR1I2, POPDC2, DPF2*	3.33	0.002
Small GTPase mediated signal transduction (GO:0007264)	*ARHGAP40, ARHGAP31, CDC42EP2*	6.25	0.004
Cellular response to lipopolysaccharide (GO:1901700.)	*NR1I2, CD80*	9.25	0.005
DNA conformation change (GO:0071103)	*CDCA5, TOP1*	19.6	0.005
Regulation of protein complex assembly (GO:0043254)	*ARHGAP40*	13.1	0.011
Purine nucleoside triphosphate metabolic process (GO:0009144)	*SAMHD1*	>100	0.014
Regulation of sister chromatid segregation (GO:0033045)	*CDCA5*	>100	0.014
Regulation of exit from mitosis (GO:0007096)	*CDCA5*	>100	0.014
Protein K48-linked ubiquitination (GO:0070936)	*SYVN1*	71.9	0.018
Pathways			
De novo pyrimidine deoxyribonucleotide biosynthesis (P02739)	*VSTM2L*	37.7	<0.001
CCKR signalling map (P06959)	*GSK3B, SRC*	6.78	<0.001
DNA replication (P00017)	*TOP1*	17.3	<0.001
Angiogenesis (P00005)	*GSK3B, SRC, NR1I2*	4.78	<0.001

The genomic coordinates of the regions under selection detected by XP-EHH in NS horses are shown in Table 14. In total, 46 genomics elements were found within the four genomic regions under selection in NS horses: 31 protein coding genes, 13 lncRNAs, one miRNA and one snRNA (Supplementary Table 7).

Table 14 Significant chromosomal regions found by the XP-EHH test in non-show jumping horses (NS). If the region overlapped with a significant Fst, the highest overlapping Fst value within the significant XP-EHH region is shown.

Chromosome	Start Position (bp)	End Position (bp)	Length (Kbp)	Adjusted *p*-Value	Highest F_{st}
1	40,592,555	43,510,660	2918	0.023	0.21
23	51,321,303	51,844,470	523	0.0501	0.15
25	5,277,465	5,984,107	707	0.0002	0.15
25	37,351,887	37,886,254	534	0.0071	0.15

The significant region on ECA1 in NS horses comprised eleven genes and overlapped with a known QTL for free jumping. Five genes were found in the region on ECA23, seven genes on ECA25:5.3–5.8 Mbp, and eight genes on ECA25:37.4–37.9 Mbp. All significant regions found with the XP-EHH test overlapped with at least one significant SNP from the F_{st} estimate in NS horses (100% concordance between the two methods). The PANTHER overrepresentation analysis recognized 21 genes out of 31 protein coding genes found within the regions indicated to be under selection in NS horses. Five GO terms were overrepresented for biological processes, as well as four pathways in NS horses ($p < 0.01$) (Table 15). Among the biological processes and pathways, three genes were overrepresented in each of them.

Table 15 Overrepresented gene ontology term (GO) at p-value <0.01 for the genes within regions detected as potentially under selection in NS horses by the XP-EHH and overlapped with F_{st} significant SNPs. GO terms are presented for both for biological processes and pathways.

PANTHER GOs	Genes	Fold Enrichment	Raw *p*–Value
Slim Biological Process			
SRP-dependent translational protein targeting to membrane (GO:0006614)	*SEC61B*	100.0	0.007
Posttranslational protein targeting to endoplasmic reticulum membrane (GO:0006620)	*SEC61B*	100.0	0.009
Intracellular protein transmembrane transport (GO:0065002)	*SEC61B*	88.7	0.012
Activin receptor signalling pathway (GO:0032924)	*TGFBR1*	81.3	0.013
Phospholipid biosynthetic process (GO:0008654)	*SGMS1*	75.0	0.014
Pathways			
Insulin/IGF pathway–protein kinase B signalling cascade (P00033)	*PTEN*	35.0	<0.001
p53 pathway feedback loops 2 (P04398)	*PTEN*	27.1	<0.001
Sulphate assimilation (P02778)	*PAPSS2*	100.0	<0.001
TGF-beta signalling pathway (P00052)	*TGFBR1*	14.5	<0.001

3.4 DISCUSSION

3.4.1 Genomic traces of breed history

Selection in SWB horses for equestrian sport performance traits was highly intensified during the last decades. In 1973, a field test for young horses was introduced in the SWB breeding program to allow for a more accurate selection of the breeding animals and increased genetic gain [129]. The overlapping homozygous segments shared in most of SWB horses agree with the intensified selective breeding program applied by the SWB studbook in the last 40 years. This latter aspect might indicate that shared homozygous segments in SWB horses can be a result of recent selection for performance, rather than inbreeding. This is supported by the estimated inbreeding coefficient (f_i) that was below 5% in six SWB individuals only. This evidence is likely to be a result of a semi-open studbook allowing inflow of genetic material from other warmblood populations. In contrast, ROHs of all lengths were evenly spread across the genome of Exmoor ponies. Short ROHs were as much as five times more frequent in Exmoor ponies than in SWB horses, indicating historical inbreeding in line with the Exmoor's known demography. The Exmoor pony breed is listed as "endangered" by the Rare Breeds Survival Trust [164], and thus founder effect and genetic drift were expected in the breed. In agreement with such assumption, several Exmoor ponies showed high f_i, which indicates significant loss of heterozygosity. Therefore, we applied inbreeding correction to distinguish between homozygosity due to relatedness rather than selection. Our results support that long shared ROHs may have different origins due to different SWB horse and Exmoor pony population histories. In SWB horses, long ROHs may originate from recent intensive selection for sport traits, whereas in Exmoor ponies long ROHs may be the result of past bottlenecks. However, the difficulty to clearly define the origin of reduction in genetic diversity in small populations has been pointed out as they are vulnerable to genetic drift [165, 166]. Therefore, we did not further examine the results retrieved from ROH in the case of Exmoor ponies, but used the results as a comparison to the SWB. In the current study, we performed a correction for linkage disequilibrium (LD) which has not been common practice for less dense SNP

data for the ROH analysis. The LD measured by the squared correlation coefficient (r^2) was significantly higher in the Exmoor ponies if compared to SWB horses, most likely because of breed history. Therefore, LD pruning was performed to reduce strong LD between markers originating from population history rather than from positive selection [49, 167, 168].

3.4.2 Signatures of selection based on ROH in SWB horses

Performance traits are known to be complex traits caused by mutations in many genes and regulatory elements, generally connected in aggregated networks. Thus, it is not surprising that several of the genes indicated to be under selection in SWB horses share pathways, co-expression, co-localisation, and interact physically or genetically. Selection for muscle strength and function is indeed a favoured trait in the performing horse, and we found several genomic regions comprising genes involved in this. An especially interesting example is the Histone deacetylase 9 (*HDAC9*) with a known regulatory function in neuronal electrical activity of excitable skeletal muscle cells [169]. Two of the ten annotated genes detected in long ROHs share biological functions in synaptic signalling (GO:0099536); the Glutamate Ionotropic Receptor NMDA Type Subunit 2B gene (*GRIN2B*) and the Calcium Voltage-gated Channel Subunit 1 gene (*CACNA1A*). Both genes are involved in response to pain, synaptic transmission and receptor clustering. The protein encoded by *GRIN2B* is a NMDA receptor channel subunit critical for neuronal communication [170]. The *GRIN2B* gene has been pointed out as potentially important for performance in Icelandic horses and coldblooded trotters [171, 172] and as a target of selection in French trotters and Gidran horses [59]. Out of the 21 genes found in shared long ROHs of SWB horses, *GRIN2B* was the only one found also in shared long ROHs in Exmoor ponies. Therefore, this gene may have been under selection prior to horse breed formation. The gene *CACNA1A* encodes the protein Cav involved in muscle contraction, hormone and neurotransmitter release [173]. Behaviour analyses in mice suggested that Cav contributes to pain perception [174]. In humans loss-of-function mutations in the *CACNA1A* were implicated in neurological and psychiatric diseases [175]. The ROH analysis also detected a region comprising the Beta-1,3-glucuronyltransferase 1 (*B3GAT1*) which has been associated

with learning abilities in domesticated rats [176]. Genes involved in neurological control and signalling pathways were likewise found in shared ROH among Hanoverian sport horses [64]. The SWB breed is genetically connected with the Hanoverian breed, as well as with other European sport horse breeds [177]. This indicates that some of the presented findings are also applicable in other warmblood breeds used for sport and that cognitive reactions and functions may be important targets of selection in Warmblood breeds. This latter result agrees with previous studies where differences in temperament between breeds have been shown [178].

3.4.3 XP-EHH and F_{st} tests indicate recent selection in SWB horses

The average F_{st} of 0.11 between the two breeds was similar to the previously estimated average F_{st} of 0.10 between 37 horse populations across the globe [49], indicating that a comparison between SWB and Exmoor ponies is reasonable. The aim of the presented study was to detect signatures of selection for performance traits in SWB horses, comparing SWB horses with a breed not selected for sport performances. From the comparison with Exmoor ponies was possible to ensure that even the worst performing SWB horse in the study was genetically superior to the best Exmoor ponies for equestrian sports, thus excluding any potential confounding effects. The comparison with an ancestral breed of SWB horses would have been an ideal solution, however none of the remaining native Swedish breeds can be claimed as the ancestor of SWB horses. A comparison with different warmblood horse populations selected for sport performance, would also be interesting, but would answer a different research question. The XP-EHH analysis supports our hypothesis that recent selection took place in SWB horses, but not in Exmoor ponies, as significant regions were only found in the former. Since the Exmoor pony genome presents widespread homozygous regions, we probably failed to identify all regions under selection in the SWB by comparing the two breeds. However, we believe that this comparison has considerably reduced the risk of finding false positives. Concordance across statistics is generally used to support putative sweeps, thus, only regions detected by both population differentiation tests

were further analysed [132, 137, 179]. Selection for eight genes involved in cellular response to stimulus (GO:0051716) were found from the composite results of the F_{st} and XP-EHH tests. As an example, the Adenylate Cyclase 1 (*ADCY1*) gene encodes adenylyl cyclase (AC) primarily expressed in the brain, influences neuroplasticity, as long-term potentiation, depression and memory formation [180]. In mice, *ADCY1* and *ADCY8* play an important role in the formation and maintenance of fear memory, dopaminergic responses, and behavioural sensitisation [181, 182]. Additionally, we detected the Guanylate Cyclase 2C (*GUCY2C*) which is associated with human attention deficiency and hyperactive behaviour [183], along with the Rho GDP Dissociation Inhibitor Beta (*ARHGDIB*), and RAS Like Estrogen Regulated Growth Inhibitor (*RERG*) genes [184, 185]. These four genes (*ADCY1*, *GUCY2C*, *ARHGDIB*, and *RERG*) are all involved in G-protein coupled receptor, and GTPase mediated signal transduction, common in the central nervous system, suggesting an important role in learning and reactivity in the performing horse. Two other genes, represented in the cellular response to stimulus term, shared biological function related to signal transduction (GO:0007165) and regulation of cell communication (GO:0010646): the Insulin like growth factor binding protein 1 (*IGFBP1*) and the Insulin like growth factor binding protein 3 (*IGFBP3*). These were also detected in studies of selection signatures in German Warmblood breeds [60] and in French Trotters [59]. Both are members of the insulin-like growth factor binding protein (*IGFBP*) family that binds IGF-I and -II and regulates somatic growth with an important function in muscle growth distribution. The importance of growth traits was further supported by two detected regions overlapping with known QTLs related to conformation and morphology traits (body size and cannon bone circumference) [95, 163]. In agreements with other studies, the region on ECA4:19,913,037-20,413,037, found in this thesis, seems to be under selection in warmblood horses [60, 137]. This region contains several genes, for example Spermatogenesis associated 48 (*SPATA48*) and Zona pellucida binding protein (*ZPBP*), which play a role in equine fertility [60, 137]. Also, the most significant SNP in this region in the XP-EHH test was located within the *ZPBP* gene (Figure 5). The *SPATA48* also promotes osteoblast differentiation, which might have an important function in sport horses [186, 187]. In our study the overrepresentation test showed as significant gene

the Fidgetin Like 1 (*FIGNL1*) which has likewise a known function in osteoblast differentiation [188]. This gene was also found in the top ten enriched pathways from the iHS test in four German warmblood horse breeds [60]. In two other regions on ECA4 and ECA7, our ROH analysis detected the Thrombospondin, type I (*THSD7A*) and the Deoxy-ribonuclease 2, lysosomal (*DNASE2*), respectively, which are both involved in bone metabolism [189, 190]. These findings may indicate a link between fertility and bone metabolism. To further validate our findings, sport performing horses should be re-sequenced to find candidate causative mutations, and functional studies are needed to confirm biological effects of the mutations.

3.4.4 How to study SWB subpopulations based on discipline?

In the SWB studbook, EBVs are available for both show jumping and dressage performance. In the presented study, the show jumping EBV was used to divide SWB horses in two potentially genetically different groups, for several reasons. The main reasons were the higher heritability [191], and faster genetic progress for the show jumping rather than the dressage trait in SWB horses [129] , but also because horses competing in show jumping are more abundant [191]. Additionally, jumping performance traits are exclusive in the selection of show jumping horses, while traits related to conformation and gait are, to some extent, selected in both disciplines. For the above-mentioned reasons, we decided to use the EBV for show jumping to divide the sampled horses in two groups. The bimodal distribution of EBVs facilitated this division. Nevertheless, the mean EBVs for dressage was significantly higher in the NS group compared to the SJ horses. Thus, it is expected that most NS horses in the study were bred mainly for dressage performance, confirming the rationale behind the use of EBVs to divide the horses in this study. The horses were selected from the ones assessed at young horse tests in 2013 and 2014, with the aim of including horses with different pedigrees and performance levels. It was not possible to completely rule out the possibility that preselection may have influenced our results. However, the comparison between means and distributions of EBV in the sample (STP) and in the whole cohort of horses born in the same years (RP) did not show any obvious signs of preselection.

3.4.5 SWB breeding program and population substructure

Our results showed genomic substructures in the population in accordance with the breeding practice applied by most breeders. After a few decades of specialized breeding, the SWB population shows a clear tendency to divide into two main subpopulations with significantly different EBVs for show jumping. This process has probably been enhanced by the use of stallions imported from European studbooks with a focus on separate disciplines. The first eigenvalue in the PCoA analysis was unable to separate the individuals into two clearly distinct clusters. This result was expected as the process of specialization is still ongoing. The more scattered cluster, based on the second eigenvalue in the SJ subpopulation, indicated higher genetic distance and lower relatedness between individuals if compared to the NS subpopulation. This may be explained by a larger breeding cohort, and the extensive use of foreign stallions from several studbooks in the breeding of SWB horses for show jumping. The latter agrees with Thorén Hellsten et al. (2009)[192], where they showed that Holstein, Selle Français and KWPN breeds have had an extensive impact on show jumping performance in SWB horses. In contrast, no specific breed was shown to influence the SWB dressage horses, except for Oldenburg in the very last few years of the studied period.

The higher percentage of thoroughbred ancestry in some horses may be likely the reason why the DAPC failed to correctly assign those animals to one specific subpopulation. Likewise, the presence of horses with a higher percentage of thoroughbred ancestry may be the reason why the BIC analysis showed a similar likelihood for two and three subpopulations in the data. However, the distribution of EBVs point at two major SWB subpopulations. By adding a very small third subpopulation would, in this study, not contribute much information.

The low average F_{st} between the SJ and NS SWB subpopulations is in line with the relatively recent specialization process towards different disciplines. Therefore, even though genomic divergence was detected within the breed, there is no evidence to claim the presence of two separate breeds.

3.4.6 Genomic Divergence Between SJ and NS SWB Horses

In the past few years, pursuing signatures of selection in the genome has been widely used to understand the genetic mechanism behind phenotypes, and several approaches have been proposed to detect selection footprints [193]. In this study, two representative methods, F_{st} and XP-EHH were employed, to explore potential signature of selection in show jumping and non-show jumping horses. Both the F_{st} and XP-EHH tests are based on population differentiation, and they can be considered complementary in time scale. It takes time to reach fixation, especially for complex traits where the desired phenotype results from mutations in a network of genes, rather than from monogenic point mutations. Compared to allele frequency difference, long-range haplotypes persist for relatively shorter periods of time before being broken down by recombination [179]. Signatures of selection at haplotype level can therefore be interpreted as more recent selection compared to F_{st}. Nevertheless, both F_{st} and XP-EHH are effective when the traits selected in one subpopulation are not selected in the other subpopulation, resembling the case of show jumping traits.

None of the potentially under selection regions found in SJ or NS horses overlapped with the genomic signature of selections in SWB from the comparison to Exmoor ponies. Genomic regions found from the comparison to Exmoor might be of general importance for equestrian sport performance, while genomic regions under selection in only one of the two subpopulations (SJ and NS) of SWB might harbour genes of more specific importance for a sport discipline.

All but one of the nine SNPs with $F_{st} > 0.20$ were located within genes or, alternatively, in their 5' or 3' untranslated regions, indicating a possible functional role for these genes in SWB horses. It is important to note that these SNPs may not be the causal variants themselves, but in LD with potential causal variants. One of these SNPs was instead located in a region where a 1 Mbp long CNV gain was reported in a Quarter horse [194]. In this region, the divergence seen between SJ and NS SWB horses may therefore be due to heterozygosity of an indel rather than a higher homozygosity. The remaining eight divergently selected SNPs are located within five genes: cyclic GMP–dependent protein kinase 1 (*PRKG1*) on ECA1; NME/NM23 family member 7 (*NME7*), and leucin-

rich-repeat-containing 7 (*LRRC7*) on ECA5; Rac Family Small GTPase 1 (*RAC1*), activated kinase 5 (*PAK5*), and phospholipase C beta 1 (*PLCB1*) on ECA22. The three genes *PRKG1*, *NME7*, and *PAK5* all encode kinases that phosphorylate proteins in cell signalling pathways. The Ser/Thr kinase encoded by the *PRKG1* gene has a crucial function in smooth muscle activity, and has been associated with thoracic aortic aneurysm in humans [178]. This gene was also reported in a potential selection signature for ocular size, memory and learning in Thoroughbred horses [195]. The kinase encoded by *NME7* catalyses nucleotides in the cell biosynthesis, and the Ser/Thr kinase encoded by *PAK5* influences the outgrowth of neurons critical in locomotion and cognition [196]. The gene *LRRC7* encodes densin-180, an abundant scaffold protein within the excitatory postsynaptic density (PSD) [197]. The gene *PLCB1* plays an important role in signal transduction from G-protein coupled receptors (*GPCRs*), which have been associated with risk of developing schizophrenia in humans [198, 199]. Overall, most of the genes located within the highest divergent regions detected by F_{st} are involved in cell signalling cascades in the central nervous system, indicating the potential presence of mental and neuronal differences between horses selected for show jumping or dressage.

Overall, more and longer regions were found under potential selection in SJ compared to NS horses, which may be a consequence of the more rapid genetic gain in show jumping performance than in non–show jumping performance traits in SWB. It is noteworthy that, within significant regions, both protein-coding genes and non-coding RNAs (ncRNA) were detected. Non-coding RNAs are involved in gene regulation by epigenetic modifications, and may regulate other genes in their genomic vicinity [200, 201]. Mutations in regulatory elements may be a quicker way for genomes to evolve, compared to coding mutations with high constrains to prevent accumulation of deleterious missense mutations [202].

3.4.7 Selection for High Mobility and Relaxed Locomotion in NS SWB Horses

In NS horses, five genes were overrepresented in the biological processes and pathways analysis; sec61 translocon beta subunit (*SEC61B*); sphingomyelin synthase 1 (*SGMS1*); phosphatase and tensin homologue (*PTEN*); 30-phosphoadenisine-50-phosphosulphate synthase 2 (*PAPSS2*); and transforming growth factor beta receptor 1 (*TGFBR1*). Three of these genes—*SGMS1*, *PTEN*, and *PAPSS2*—are located within the significant region on ECA1:40,592,555–43,510,660. This ECA1 region was found to be under selection in racing Quarter Horses [137], and overlaps with a known QTL for free jumping in Hanoverian horses [203], suggesting that the homozygous haplotype in the NS SWB horses is unfavourable for jumping performance. The selection pressure on this ECA1 region is further supported by the high divergence detected by F_{st} between SJ and NS horses, as mentioned above. Mice haplo-insufficient for *PTEN* show abnormal social behaviour and exaggerated reactions to sensory stimuli [204, 205], and, in humans, *PAPSS2* influences the odds ratio for being an exerciser [206], suggesting the importance of the region for equine sport performance. It was also shown that *PAPSS2* mRNA is down-regulated by transforming growth factor beta (*TGF-beta*) signalling, mediated by its receptor *TGFBR1* [207, 208]. The equine gene *TGFBR1*, located on ECA25:5,277,465–5,984,107, was found to be overrepresented in the activin receptor signalling pathway and TGF-beta signalling pathway. Mutations in *TGFBR1* have been associated with aortic aneurysms, as well as impaired connective tissues leading to joint laxity [209]. The overrepresented gene SEC61 Translocon Beta Subunit (*SEC61B*) is located only 60 kb downstream of *TGFBR1*. The primary function of the SEC61 complex is to translocate proteins across the endoplasmatic reticulum (ER) membrane, and, together with the Discoidin Domain Receptor 1, it regulates collagen IV synthesis, which is upregulated in fibrosis [210, 211].

The gene collagen type XV alpha 1 chain *COL15A1* is likewise located within the same region under selection on ECA25 as *TGFBR1* and *SEC61B*, although not among the overrepresented genes. Mice deficient of *COL15A1* expression show exercise-induced skeletal myopathy and cardiovascular defects [212]. Recently, it was also shown that

COL15A1 gene expression is downregulated in skin fibroblasts derived from patients with the kyphoscoliotic form of Ehlers–Danlos syndrome (kEDS) [213]. This form of kEDS is caused by mutations in the genes FK506–binding protein 14 (*FKBP14*) and/or procollagen lysine, 2-oxoglutarate 5-dioxygenase 1 (*PLOD1*). Ehlers–Danlos syndrome has, in many cases, been associated with hypermobility, joint laxity and aortic aneurysms in humans [214]. The equine version of kEDS, i.e., Warmblood Fragile Foal Syndrome (WFFS), is caused by a recessive lethal missense mutation in *PLOD1*, where affected foals show severe skin fragility, and joint laxity [215]. Similarly, a region holding a gene with a known function in collagen regulation was found to be under selection in SJ horses by the XP-EHH test, although not confirmed by F_{st}. This region (ECA12:11,851,991–11,854,261) harbours the gene encoding for the zinc transporter solute carrier family 39 member 13 (*SLC39A13*). Mutations in this gene have been associated with collagen deficiency in the spondylocheiro dysplastic form of Ehlers–Danlos syndrome [216]. Overall, these findings suggest a strong selection on connective tissue functions in all sport horses, highlighting the importance of high mobility and relaxed locomotion patterns. This is, to some extent, in agreement with results from the comparison between SWB horses and Exmoor ponies, in which other genes involved in muscle contraction and development were shown to be under putative selection in the SWB breed.

3.4.8 Selection for Mentality and Postsynaptic Signalling in SJ SWB Horses

In the SJ horses a total of 26 overrepresented genes were found in the GO term analysis for biological function and pathways. The genes were mainly related to organism development, gene regulation and cellular processes, and 20 of them were confirmed by F_{st} analysis. The largest region under putative selection in SJ horses is located on ECA22:27,935,550-31,789,708 with 24 ncRNAs and 37 protein-coding genes, of which nine are overrepresented protein-coding genes—disc large homolog associated protein 4 (*DLGAP4*), myosin light chain 9 (*MYL9*), SAM and HD domain containing deoxynucleoside triphosphate triphosphohydrolase 1 (*SAMHD1*), retinoblastoma 1

(*RBL1*), SRC Proto–Oncogene/Non–Receptor Tyrosine Kinase (*SRC*), Rho GTPase activating protein 40 (*ARHGAP40*), DNA Topoisomerase 1 (*TOP1*), V-set and transmembrane domain containing 2 like (*VSTM2L*)—and the uncharacterized protein *KIAA1755*. Some of the overrepresented genes, for example, *RBL1* and *SAMHD1*, have general functions in the cell, such as cell differentiation and nucleotide metabolism [217]. Most of the 24 ncRNAs are located in the 3'-end of this region and may be involved in the epigenetic regulation of gene expression of genes located within the region. It was previously shown that such epigenetic regulation could alter the gene expression of *DLGAP4* in cerebellar Purkinje cells, causing early-onset non-progressive cerebellar ataxia, and bipolar disorder in humans [218]. The membrane-associated guanylate kinase, encoded by the gene *DLGAP4*, the Rho GTPase activating protein 40, and the non-receptor tyrosine kinase encoded by the gene SRC, are all involved in the excitatory post-synaptic density (PSD), a large protein complex important in synaptic plasticity. Improper function of PSD has been associated with schizophrenia and bipolar disorder [219–222]. The proteins in PSD organize the post-synaptic cytoskeleton by linking glutamate receptors and signalling proteins to regulate receptor turnover in response to oscillating synaptic activity. This suggests that excitatory synaptic plasticity has an important function in SJ horses, possibly influencing the flexibility and reactive capacity needed when jumping. Genes involved in neurological control and signalling pathways were likewise found in shared ROHs both among SWB horses and Hanoverian sport horses [64].

3.4.9 Selection for Growth and Muscle Function in SJ SWB Horses

The *SRC* gene has been shown to be involved in osteopetrosis due to non-functional osteoclasts [223]. The significant region on ECA8:10,891,925–11,400,093, comprises the genes phosphatidylinositol transfer protein (*PITPNB*) and meningioma 1 (*MN1*), that overlap with a reported QTL for body growth trait in horses [95]. In humans, the two genes *PITPNB* and *MN1* are both involved in ossification processes during development [224]. The two genes, *TOP1* and *VSTM2L*, on ECA22, are both overrepresented in

biological processes as well as cellular pathways, with a known function in skeletal–muscle development. In humans, mutations in *TOP1* have been associated with collagen build-up and growth of connective tissue [225], while *VSTM2L* was associated with carcass weight in Gir cattle, which implies a function in growth and muscle development [50].

The myosin light chain encoded by *MYL9* regulates muscle contraction by modulating ATPase activity [226]. The five genes glycogen synthase kinase 3 beta (*GSK3B*), nuclear receptor subfamily 1 group I member 2 (*NR1I2*), popeye domain containing 2 (*POPDC2*), CD80 molecular (*CD80*), and Rho GTPase activating protein 31 (*ARHGAP31*), were overrepresented in SJ horses, and located within the same region on ECA19:41,459,599–41,959,599. The kinase encoded by *GSK3B* regulates several proteins involved in muscle glycogenolysis and glycolysis. One of these proteins is glycogen synthase (GS), encoded by the gene glycogen synthase 1 (*GYS1*). A gain-of-function mutation in *GYS1* is responsible for the muscle disease polysaccharide storage myopathy (PSSM) in horses [227–229], and double-muscling in myostatin deficient cattle is a result of increased *GSK3B* phosphorylation of GS [228]. Furthermore, loss of *PTEN* function in mice also leads to the inactivation of *GSK3B* [230] that in turn, normally phosphorylates *ARHGAP31*. Dominant gain-of-function mutations in the *ARHGAP31* gene is associated to Adams–Oliver syndrome, where the patients show skin and limb defects [231]. The gene Popeye domain containing 2 (*POPDC2*) encodes a membrane-bound protein, affecting potassium levels in striated muscles. Mice with null mutations in *POPDC2* develop muscle dystrophy and cardiac arrhythmia [232, 233]. This ECA19 region, together with the regions on ECA8, and ECA22, appears to be under putative selection in SJ horses, where genes involved in muscle build-up and function are located.

3.4.10 Selection for CNS Reward System and Motoneuronal Control of Coordination in SJ SWB Horses

The regulator of G protein signalling 17 (*RGS17*) gene is located within the significant region on ECA31:13,778,365–14,321,247 for SJ horses, and overlaps with a putative QTL for altitude adaptation in horses [234]. The *RGS17* modulates signalling of opioid

receptors required in the endogenous reward system. The *RGS17* gene is located about 0.8 Mbp downstream of the opioid receptor mu 1 (*OPRM1*) and may influence transcription of this gene. In humans, *RGS17* has been associated with risk for dependence on addictive substances [235]. This implies a possible function of *RGS17* in the equine endogenous reward system. In the genomic region in-between the *RGS17* and *OPRM1* genes, Doan et al. (2012) and Kader et al. (2016) reported one indel variation upstream of the gene prothymosin alpha (*PTMA*) and another indel covering a long non-coding RNA (lncRNA) [194, 236]. Prothymosin alpha has been reported to influence locomotor activity, memory-learning, and anxiety behaviour in mice, as well as modulating linker histone interaction to promote transcription [237, 238]. It is therefore plausible that the genes *RGS17* and *OPRM1* are co-regulated and influenced by *PTMA* and lncRNA structural modifications.

Among the overrepresented biological processes, the gene neuron navigator 2 (*NAV2*) and developing brain homeobox 1 (*DBX1*) are located within the selected region on ECA7:89,669,652–91,203,503. Both genes are involved in locomotion patterns; *NAV2* was shown to be involved in motor coordination and balance in mice [239], while *DBX1* was shown to coordinate left–right locomotor activity [240]. In fact, *DBX1* is expressed in the same dI6 interneurons as doublesex and mab-3 related transcription factors (*DMRT3*) [241], known to affect locomotion pattern in horses [242]. Those findings suggest that show jumping horses are selected for neuromuscular control and coordination, and they might have a well-developed endogenous reward system that triggers them to seek fences to jump.

Chapter 4

4.1 What to do next to preserve local genetic diversity?

The case study on Bardigiano horses, presented in Chapter 2, shed light on the breed's genetic diversity based on pedigree and molecular data, and it proposed breeding strategies to ensure long-term genetic diversity. The analyses based on pedigree data were performed using 9,469 horses, of which 3,416 are alive. The demographic and genetic parameters were estimated on subpopulations to evaluate potential genetic diversity differences among breeding and nonbreeding horses, based on province of origin and among animals showing different breeding values (EBVs). From the first studied birth year cohort (1975–1983) inbreeding steadily increased, reaching in the last birth year cohort a mean value equal to 0.10. The rate of inbreeding per generation was equal to 1.64% which is above the rate suggested by FAO regulation. In addition, significant differences on both average relatedness and inbreeding among horses with high and low EBVs were shown, highlighting a reduced genetic variability in superior animals. Breeding strategies differed also among areas where Bardigiano horses are commonly bred. The higher inbreeding and relationship among individuals bred in mountain area stressed the need for a more uniform genetic flow among areas of breeding to reduce the risk of genetic drift and inbreeding within local area.

A possible solution to monitor future inbreeding and, the population's genetic diversity, is to implement efficient breeding programs. Optimal contribution selection (OCS) might be an effective tool to increase genetic gain while controlling the rate of inbreeding, thus enabling the maintenance of genetic diversity in the Bardigiano. The introduction of the average relationship (AR) as criteria in breeding programs lends itself as an effective strategy since the selection of breeding animals with low AR can minimize the inbreeding in coming generations.

Genomics can increase substantially our understanding of genetic diversity, as demonstrated for the Bardigiano breed (Chapter 2). From the analysis of 88 genotyped Bardigiano horses, it was possible to further decipher the state of genetic diversity in

the breed. The effective population size based on Linkage Disequilibrium (Ne) was equal to 39 horses, and it showed a decline over time. Based on the ROH analyses, we showed a critical reduction of genetic variability mainly driven by past bottlenecks. The average inbreeding based on runs of homozygosity (ROH) was equal to 0.17 (SD = 0.03). The majority of the ROHs were relatively short (91% were ≤2 Mbp long), highlighting the occurrence of older inbreeding, rather than a more recent one. Eight ROH islands were found in the breed; these islands were shared among more than 70% of the Bardigiano horses analysed in this study. Four of them overlapped with known QTLs associated with conformation traits (e.g., body size and coat colour) and disease susceptibility. These homozygous genomic regions might be the outcome of directional selection in recent years, in line with the conversion of Bardigiano horses from agricultural to riding purposes.

Evaluating inbreeding based on molecular data might be an effective strategy for breeding association to design optimal breeding practices and making decisions to ensure the long-term survival of the breed. Genomics might also help to decipher the genetic background of unique traits in local and traditional breeds, such as adaptability and resilience to harsh conditions, and resistance to endemic diseases. Such traits should be considered a fundamental part of a breed-specific heritage that society should strive to preserve. To this end, signatures of selection help to detect the genetic background of those important traits, as shown in the Bardigiano case study. Therefore, signatures of selection detection might be an effective approach in local and small populations which need to redefine their role in the society. The identification of their unique genomic features might be a key aspect in this regard.

The use of molecular data to implement genomic selection might lead to a successful conversion of those breeds from agricultural to riding purposes. However, successful implementation requires all-encompassing collaboration across studbooks [18]. Therefore, future collaborations among local horse breeds sharing the challenge of converting horses from agricultural to riding activity, are suggested to develop potential multibreed genomic predictions.

Moreover, deleterious mutations in a population can nowadays be more easily prevented by means of genomics. Beside phenotypically diseased animals, mutant-

allele carriers can be identified and removed from the breeding program. For this reason, further research is required to detect lethal alleles in a population and to develop specific tests for carrier identification.

The study on Bardigiano horse based on SNP data presented in this thesis is the first one of this kind conducted in the context of Italian local breeds, to the best of the author's knowledge. Therefore, next steps should broaden the available knowledge in the field of genetic diversity by introducing additional Italian native breeds. This last step will increase our understanding of Italian horse genetic diversity and it will promote new research opportunities on the origins of and connections among different Italian native breeds. Closed populations with low diversity can be potentially reinvigorated by the genetic contribution from historically and genetically related breeds without the risk of losing the breed's identity.

4.2 What to do next to implement genomics in horse breeding?

Chapter 3 has shown that genes associated with behavioural and physical abilities, and fertility appear to be the target of selection in the Swedish Warmblood Breed. The SWB horses reveal putative signatures of selection in genomic regions containing genes primarily involved in nervous system functionality, as well as muscle contraction and development. Such evidence was obtained by comparing results with those calculated for a different breed not selected for sport purpose, the Exmoor ponies. As expected, due to the complex nature of sport horse performance, many of the genes under selection in SWB interact with each other in complex biological networks. In line with this expectation, the SWB case study reveals that selection for sport performance has occurred in numerous genomic regions.

Chapter 3 also showed genetic divergence due to the specialization towards different disciplines in SWB horses. The two main subpopulations of SWB horses were on average moderately differentiated, except in eleven chromosomes showing more differentiation. Both measures of genomic diversity and extended haplotypes showed significant chromosomal regions with signatures of positive selection for either show jumping or non-show jumping horses. In show jumping horses, genes primarily related to the endogenous reward system, excitatory synaptic plasticity, neuromuscular

control, and coordination, seemed to be under selection. On the other hand, genes involved in joint laxity, collagen build-up and muscle function are potentially under selection in non-show jumping horses, suggesting the presence of numerous genes involved in flexibility of movements in performing horses. Furthermore, many of the selected chromosomal regions comprised noncoding elements that are putative regulators of gene expression, implying a further dimension to the variability of the phenotypes selected in the sport horses.

The results of Chapter 3 pointed at many interesting genes to be validated by further functional genomic studies such as resequencing, RNA-sequencing and/or detection of epigenetic DNA modifications. Resequencing would provide a complete profile of the genetic variability and would create new opportunities to study the role of the potentially selected genes in sport horses. While SNP data are beneficial to pinpoint regions of interest, sequence and gene expression data will provide extra information, thus leading to a superior level of comprehension of complex traits in horses.

A further research step might be to combine the above results with genome-wide association studies (GWAS), also by using additional sources of information (e.g. copy number variation, CNV). However, the association of genomic regions with phenotypes is highly dependent on how traits have been recorded. If any bias associated with data collection takes place, e.g. due to inaccuracy in the trait definition (low repeatability and reproducibility), results from the association studies might be compromised. To date, and especially for temperament-related traits, little research has been done to efficiently define those traits. Therefore, a possible next step is the investigation of strategies to optimise the identification of performance-related traits, especially in the field of behavioural traits. The main goal might be to move beyond the traditional scoring system and to improve, via novel strategies (e.g. advanced sensor technologies) the traits' recording system. The expected result is optimizing the association studies by using more objective recording systems for the identification of performance traits especially related to temperament traits.

List of publications

The following publications were produced during this three-year PhD program:

Journal publications

1. Enrico Mancin*, **Michela Ablondi***, Roberto Mantovani, Giuseppe Pigozzi, Alberto Sabbioni, Cristina Sartori. 2020. "Genetic Variability in the Italian Heavy Draught Horse from Pedigree Data and Genomic Information". Animals. Available from: https://doi.org/10.3390/ani10081310. *Shared first authorship
2. Marica Simoni, Rokia Temmar, Davide Augusto Bignamini, Andreas Foskolos, Alberto Sabbioni, **Michela Ablondi**, Afro Quarantelli, Federico Righi. 2020. "Effects of the combination between selected phytochemicals and the carriers silica and Tween 80 on dry matter and neutral detergent fibre digestibility of common feeds". Italian Journal of Animal Science. Available from: https://doi.org/10.1080/1828051X.2020.1787882
3. **Michela Ablondi**, Christos Dadousis, Matteo Vasini, Susanne Eriksson, Sofia Mikko, Alberto Sabbioni. 2020. "Genetic Diversity and Signatures of Selection in a Native Italian Horse Breed Based on SNP Data". Animals. Available from: https://doi.org/10.3390/ani10061005
4. Elena Mariani, Andrea Summer, **Michela Ablondi***, Alberto Sabbioni. 2020. "Genetic Variability and Management in Nero di Parma Swine Breed to Preserve Local Diversity". Animals. Available from: https://doi.org/10.3390/ani10030538. *Corresponding author
5. Alberto Sabbioni, Valentino Beretti, Paola Superchi, **Michela Ablondi**. 2020. "Body weight estimation from body measures in Cornigliese sheep breed". Italian Journal of Animal Science. Available from: https://doi.org/10.1080/1828051X.2019.1689189
6. **Michela Ablondi**, Andrea Summer, Matteo Vasini, Marica Simoni, Alberto Sabbioni. 2020. "Genetic Parameters Estimation in an Italian Horse Native Breed to Support the Conversion from Agricultural Uses to Riding Purposes". Journal of Animal Breeding and Genetics. Available from: https://doi.org/10.1111/jbg.12425

7. **Michela Ablondi**, Susanne Eriksson, Sasha Tetu, Alberto Sabbioni, Åsa Viklund, Sofia Mikko. 2019. "Genomic Divergence in Swedish Warmblood Horses Selected for Equestrian Disciplines". Genes. Available from: https://doi.org/10.3390/genes10120976
8. Marina Solé*, **Michela Ablondi***, Amrei Binzer-Panchal, Brandon D Velie, Nina Hollfelder, Nadine Buys, Bart J Ducro, Liesbeth François, Steven Janssens, Anouk Schurink, Åsa Viklund, Susanne Eriksson, Anders Isaksson, Hanna Kultima, Sofia Mikko, Gabriella Lindgren. 2019. "Inter-and intra-breed genome-wide copy number diversity in a large cohort of European equine breeds". BMC Genomics. Available from: https://doi.org/10.1186/s12864-019-6141-z. *Shared first authorship
9. **Michela Ablondi**, Åsa Viklund, Gabriela Lindgren, Susanne Eriksson, Sofia Mikko. 2019. "Signatures of Selection in the Genome of Swedish Warmblood Horses Selected for Sport Performance". BMC Genomics. Available from: https://doi.org/10.1186/s12864-019-6079-1
10. Alberto Sabbioni, Valentino Beretti, Ernesto Mario Zambini, Paola Superchi, **Michela Ablondi**. 2019. "Allometric Coefficients for Physical-Chemical Parameters of Meat in a Local Sheep Breed". Small Ruminant Research. Available from: https://doi.org/10.1016/j.smallrumres.2019.04.001
11. Paola Superchi, Roberta Saleri, Sven Menčik, Silvia Dander, Valeria Cavalli, Chiara Izzi, **Michela Ablondi**, Alberto Sabbioni. 2019. "Relationships among Maternal Backfat Depth, Plasma Adipokines and the Birthweight of Piglets". Livestock Science. Available from: https://doi.org/10.1016/j.livsci.2019.03.012
12. **Michela Ablondi**, Matteo Vasini, Valentino Beretti, Paola Superchi, Alberto Sabbioni. 2018. "Exploring genetic diversity in an Italian horse native breed to develop strategies for preservation and management". Journal of Animal Breeding and Genetics. Available from: http://doi.wiley.com/10.1111/jbg.12357

Conference publications

1. Michela Ablondi, Paola Superchi, Valentino Beretti, Alberto Sabbioni. Exploring genetic diversity in Nero di Parma pig breed to develop strategies for conservation and management. Abstract presented at the XPIGMED, X International Symposium of Mediterranean pig 2019.

2. Michela Ablondi, Susanne Eriksson, Alberto Sabbioni, Åsa Viklund, Sofia Mikko. Jumping into selective sweeps for sport horse performance. Abstract presented at the ISAG 37th conference 2019.

3. Michela Ablondi, Claudio Cipolat-Gotet, Jan-Thijs Van Kaam, Massimo Malacarne, Alberto Sabbioni, Andrea Summer. Signatures of selection in the genome of Italian Holstein cattle for cheese production. Abstract presented at the ASPA 2019 conference.

4. Michela Ablondi, Valentino Beretti, Marica Simoni, Federico Righi, Paola Superchi, Alberto. Sabbioni Body weight estimation from linear measures in Cornigliese sheep breed. Abstract presented at the ASPA 2019 conference.

5. Michela Ablondi, Åsa Viklund, Carl Rubin, Gabriella Lindgren, Susanne Eriksson, Sofia Mikko. ROH as hint of selection in the genome of a modern sport horse breed. Abstract presented at the EAAP 2018 69th Conventional and traditional livestock production systems.

6. Michela Ablondi, Valentino Beretti, Matteo Vasini, Paola Superchi, Alberto Sabbioni. Genetic parameters in the Bardigiano horse breed population. Abstract presented at the EAAP 2018 69th Conventional and traditional livestock production systems.

7. Michela Ablondi, Matteo Vasini, Valentino Beretti, Alberto Sabbioni. Explore genetic diversity at pedigree level in an Italian native mountain horse breed to develop strategies for breed preservation and management. Abstract presented at the 1st European Symposium on Livestock Farming in Mountain Areas 2018.

References

1. Cothran EG, Podhajsky AW. Evolution of the horse. Encyclopaedia Britannica. 2020. https://www.britannica.com/animal/horse/Evolution-of-the-horse.

2. Librado P, Gamba C, Gaunitz C, Der Sarkissian C, Pruvost M, Albrechtsen A, et al. Ancient genomic changes associated with domestication of the horse. Science (80-). 2017;356:442–5. doi:10.1126/science.aam5298.

3. Forrest S. Age of the horse. Cambridge: Groups West; 2016. doi:10.1017/CBO9781107415324.004.

4. Waran N. The Welfare of horses. In: The Welfare of horses. Kluwer Academi Publishers; 2003. p. 1–30. doi:10.1017/CBO9781107415324.004.

5. Scherf B, Pilling D. The second report on the state of the world's animal genetic resources for food and agriculture. Rome; 2015. http://www.fao.org/3/a-i4787e/index.html.

6. Wallner B, Vogl C, Shukla P, Burgstaller JP, Druml T, Brem G. Identification of Genetic Variation on the Horse Y Chromosome and the Tracing of Male Founder Lineages in Modern Breeds. PLoS One. 2013;8:e60015. doi:10.1371/journal.pone.0060015.

7. Druml T. Functional Traits in Early Horse Breeds of Mongolia, India and China from the Perspective of Animal Breeding. In: Pferde in Asien: Geschichte, Handel und Kultur. Verlag der österreichischen Akademie der Wissenschaften; 2009. p. 9–16. doi:10.2307/j.ctvmd83w6.6.

8. European Horse Network. 2015. http://www.europeanhorsenetwork.eu/horse-industry/economic-impact/. Accessed 1 Sep 2020.

9. Khadka R, Jan Philipsson S, Georg Thaller S. Global Horse Population with respect to Breeds and Risk Status. 2010.

10. Groeneveld LF, Lenstra JA, Eding H, Toro MA, Scherf B, Pilling D, et al. Genetic diversity in farm animals - A review. Anim Genet. 2010;41 SUPPL. 1:6–31. doi:10.1111/j.1365-2052.2010.02038.x.

11. Sherf B. World Watch List for Domestic Animal Diversity, 3rd ed. FAO, Rome. 2000.

12. Allendorf FW, Hohenlohe PA, Luikart G. Genomics and the future of conservation genetics. Nat Rev Genet. 2010;11:697–709.

13. Huisman J, Kruuk LEB, Ellis PA, Clutton-Brock T, Pemberton JM. Inbreeding depression across the lifespan in a wild mammal population. Proc Natl Acad Sci. 2016;113:3585–90. doi:10.1073/pnas.1518046113.

14. Eggen A. The development and application of genomic selection as a new breeding paradigm. Anim Front. 2012;2:10–5. doi:10.2527/af.2011-0027.

15. Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME. Invited review: Genomic selection in dairy cattle: Progress and challenges. J Dairy Sci. 2009;92:433–43. doi:10.3168/jds.2008-1646.

16. Schefers JM, Weigel KA. Genomic selection in dairy cattle: Integration of DNA testing into breeding programs. Anim Front. 2012;2:4–9. doi:10.2527/af.2011-0032.

17. Howard JT, Pryce JE, Baes C, Maltecca C. Invited review: Inbreeding in the genomics era: Inbreeding, inbreeding depression, and management of genomic variability. J Dairy Sci. 2017;100:6009–24. doi:10.3168/jds.2017-12787.

18. Stock KF, Jönsson L, Ricard A, Mark T. Genomic applications in horse breeding. Anim Front. 2016;6:45–52. doi:10.2527/af.2016-0007.

19. Stock K, Reents R. Genomic Selection: Status in Different Species and Challenges for Breeding. Reprod Domest Anim. 2013;48 SUPPL.1:2–10. doi:10.1111/rda.12201.

20. Viklund Å, Thorén Hellsten E, Näsholm A, Strandberg E, Philipsson J. Genetic parameters for traits evaluated at field tests of 3- and 4-year-old Swedish Warmblood horses. animal. 2008;2:1832–41. doi:10.1017/S1751731108003030.

21. Thorén Hellsten E, Viklund Å, Koenen EPCPC, Ricard A, Bruns E, Philipsson J, et al. Review of genetic parameters estimated at stallion and young horse performance tests and their correlations with later results in dressage and show-jumping competition. Livest Prod Sci. 2006;103:1–12. doi:10.1016/j.livsci.2006.01.004.

22. Ducro BJ, Koenen EPC, van Tartwijk JMFM, Bovenhuis H. Genetic relations of movement and free-jumping traits with dressage and show-jumping performance in competition of Dutch Warmblood horses. Livest Sci. 2007;107:227–34. doi:10.1016/j.livsci.2006.09.018.

23. Koenen EPC, van Veldhuizen 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 Prod Sci. 1995;43:85–94.

24. Duensing J, Stock KF, Krieter J. Implementation and Prospects of Linear Profiling in the Warmblood Horse. J Equine Vet Sci. 2014;34:360–8. doi:10.1016/j.jevs.2013.09.002.

25. Viklund Å, Eriksson S. Genetic analyses of linear profiling data on 3-year-old Swedish Warmblood horses. J Anim Breed Genet. 2018;135:62–72. doi:10.1111/jbg.12311.

26. Raudsepp T, Finno CJ, Bellone RR, Petersen JL. Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era. Anim Genet. 2019;50:569–97.

27. Hendricks B Lou. International encyclopedia of horse breeds. University of Oklahoma press; 1995.

28. Libro Genealogico Cavallo Bardigiano. La diversità genetica del cavallo Bardigiano ad oggi. First edit. Parma; 2018. www.cavallobardigiano.it.

29. Di Stasio L, Perrotta G, Blasi M, Lisa C. Genetic characterization of the Bardigiano horse using microsatellite markers. Ital J Anim Sci. 2008;7:243–50.

30. Hill WG, Mackay TFC. D. S. Falconer and introduction to quantitative genetics. Genetics. 2004;167:1529–36.

31. Sairanen J, Nivola K, Katila T, Virtala A-M, Ojala M. Effects of inbreeding and other genetic components on equine fertility. animal. 2009;3:1662–72. doi:10.1017/S1751731109990553.

32. Maignel L, Boichard D, Verrier E. Genetic variability of French dairy breeds estimated from pedigree information. Interbull Bull. 1996;14:49–53.

33. Bussiman FO, Perez BC, Ventura R V., Peixoto MGCD, Curi RA, Balieiro JCC. Pedigree analysis and inbreeding effects over morphological traits in Campolina horse population. animal. 2018;12:2246–55. doi:10.1017/S175173111800023X.

34. Duru S. Pedigree analysis of the Turkish Arab horse population: structure, inbreeding and genetic variability. animal. 2017;11:1449–56. doi:10.1017/S175173111700009X.

35. Giontella A, Pieramati C, Silvestrelli M, Sarti FM. Analysis of founders and performance test effects on an autochthonous horse population through pedigree analysis: structure, genetic variability and inbreeding. animal. 2019;13:15–24. doi:10.1017/S1751731118001180.

36. Hamann H, Distl O. Genetic variability in Hanoverian warmblood horses using

pedigree analysis1. J Anim Sci. 2008;86:1503–13. doi:10.2527/jas.2007-0382.

37. Hasler H, Flury C, Menet S, Haase B, Leeb T, Simianer H, et al. Genetic diversity in an indigenous horse breed - implications for mating strategies and the control of future inbreeding. J Anim Breed Genet. 2011;128:394–406. doi:10.1111/j.1439-0388.2011.00932.x.

38. Kjöllerström HJ, Gama LT, Oom MM. Impact of inbreeding on fitness-related traits in the highly threatened Sorraia horse breed. Livest Sci. 2015;180:84–9. doi:10.1016/j.livsci.2015.08.001.

39. Onogi A, Shirai K, Amano T. Investigation of genetic diversity and inbreeding in a Japanese native horse breed for suggestions on its conservation. Anim Sci J. 2017;88:1902–10. doi:10.1111/asj.12867.

40. Schurink A, Arts DJG, Ducro BJ. Genetic diversity in the Dutch harness horse population using pedigree analysis. Livest Sci. 2012;143:270–7. doi:10.1016/j.livsci.2011.10.005.

41. Vicente AA, Carolino N, Gama LT. Genetic diversity in the Lusitano horse breed assessed by pedigree analysis. Livest Sci. 2012;148:16–25. doi:10.1016/j.livsci.2012.05.002.

42. Vostrá-Vydrová H, Vostrý L, Hofmanová B, Krupa E, Zavadilová L. Pedigree analysis of the endangered Old Kladruber horse population. Livest Sci. 2016;185:17–23. doi:10.1016/j.livsci.2016.01.001.

43. Bokor Á, Jónás D, Ducro B, Nagy I, Bokor J, Szabari M. Pedigree analysis of the Hungarian Thoroughbred population. Livest Sci. 2013;151:1–10.

44. Kim E-S, Cole JB, Huson H, Wiggans GR, Van Tassell CP, Crooker BA, et al. Effect of Artificial Selection on Runs of Homozygosity in U.S. Holstein Cattle. PLoS One. 2013;8:e80813. doi:10.1371/journal.pone.0080813.

45. Purfield DC, Berry DP, McParland S, Bradley DG. Runs of homozygosity and population history in cattle. BMC Genet. 2012;13:70. doi:10.1186/1471-2156-13-70.

46. Muñoz M, Bozzi R, García-Casco J, Núñez Y, Ribani A, Franci O, et al. Genomic diversity, linkage disequilibrium and selection signatures in European local pig breeds assessed with a high density SNP chip. Sci Rep. 2019;9:13546. doi:10.1038/s41598-019-49830-6.

47. Talenti A, Bertolini F, Pagnacco G, Pilla F, Ajmone-Marsan P, Rothschild MF, et al. The Valdostana goat: a genome-wide investigation of the distinctiveness of its selective sweep regions. Mamm Genome. 2017;28:129.

48. Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, et al. Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. PLoS Biol. 2012;10:e1001258. doi:10.1371/journal.pbio.1001258.

49. 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. doi:10.1371/journal.pone.0054997.

50. Maiorano AM, Lourenco DL, Tsuruta S, Toro Ospina AM, Stafuzza NB, Masuda Y, et al. Assessing genetic architecture and signatures of selection of dual purpose Gir cattle populations using genomic information. PLoS One. 2018;13:1–24. doi:10.1371/journal.pone.0200694.

51. Khayatzadeh N, Mészáros G, Utsunomiya YT, Garcia JF, Schnyder U, Gredler B, et al. Locus-specific ancestry to detect recent response to selection in admixed Swiss Fleckvieh cattle. Anim Genet. 2016;47:637–46. doi:10.1111/age.12470.

52. Msalya G, Kim ES, Laisser ELK, Kipanyula MJ, Karimuribo ED, Kusiluka LJM, et al. Determination of genetic structure and signatures of selection in three strains of Tanzania Shorthorn Zebu, Boran and Friesian cattle by genome-wide SNP analyses. PLoS One. 2017;12:1–18.

53. Peripolli E, Munari DP, Silva MVGBGB, Lima ALFF, Irgang R, Baldi F. Runs of homozygosity: current knowledge and applications in livestock. Anim Genet. 2017;48:255–71. doi:10.1111/age.12526.

54. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows into population history and trait architecture. Nat Rev Genet. 2018;19:220–34. doi:10.1038/nrg.2017.109.

55. 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. doi:10.1016/j.ajhg.2008.08.007.

56. Curik I, Ferenčaković M, Sölkner J. Inbreeding and runs of homozygosity: A possible

solution to an old problem. Livest Sci. 2014;166:26–34. doi:10.1016/j.livsci.2014.05.034.

57. Ferenčaković M, Hamzić E, Gredler B, Solberg TR, Klemetsdal G, Curik I, et al. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. J Anim Breed Genet. 2013;130:286–93. doi:10.1111/jbg.12012.

58. Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, et al. Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. Anim Genet. 2015;46:110–21. doi:10.1111/age.12259.

59. Grilz-Seger, Neuditschko, Ricard, Velie, Lindgren, Mesarič, 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. doi:10.3390/genes10070491.

60. 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. doi:10.1371/journal.pone.0215913.

61. 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. doi:10.1186/s12864-019-5564-x.

62. Druml T, Neuditschko M, Grilz-Seger G, Horna M, Ricard A, Mesarič M, et al. Population Networks Associated with Runs of Homozygosity Reveal New Insights into the Breeding History of the Haflinger Horse. J Hered. 2018;109:384–92. doi:10.1093/jhered/esx114.

63. Salek Ardestani S, Aminafshar M, Zandi Baghche Maryam MB, Banabazi MH, Sargolzaei M, Miar Y. Whole-Genome Signatures of Selection in Sport Horses Revealed Selection Footprints Related to Musculoskeletal System Development Processes. Animals. 2019;10:53. doi:10.3390/ani10010053.

64. Metzger J, Karwath M, Tonda R, Beltran S, Águeda L, Gut M, et al. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. BMC Genomics. 2015;16:764. doi:10.1186/s12864-015-1977-3.

65. Moravčíková N, Kasarda R, Kadlečík O, Trakovická A, Halo M, Candrák J. Runs of

homozygosity as footprints of selection in the norik of muran horse genome. Acta Univ Agric Silvic Mendelianae Brun. 2019;67:1165–70.
66. 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.
67. 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.
68. Gutierrez JP, Goyache F. A note on ENDOG: a computer program for monitoring genetic variability of populations using pedigree information. J Anim Sci. 2005;122:172–9.
69. SAS® 9.4 Fourth Edition. SAS Institute Inc. 2015. http://support.sas.com/documentation/cdl/en/lestmtsref/68024/PDF/default/lestmtsref.pdf.
70. Caballero A, Toro MA. Interrelations between effective population size and other pedigree tools for the management of conserved populations. Genet Res. 2000;75:331–43.
71. ISTAT. Principali statistiche geografiche sui comuni. 2018. https://www.istat.it/it/archivio/156224. Accessed 1 Jun 2018.
72. Lacy RC. Analysis of founder representation in pedigrees: Founder equivalents and founder genome equivalents. Zoo Biol. 1989;8:111–23.
73. Boichard D, Maignel L, Verrier É. The value of using probabilities of gene origin to measure genetic variability in a population. Genet Sel Evol. 1997;29:5. doi:10.1186/1297-9686-29-1-5.
74. Meuwissen T, Luo Z. Computing inbreeding coefficients in large populations. Genet Sel Evol. 1992;24:305. doi:10.1186/1297-9686-24-4-305.
75. Ablondi M, Summer A, Vasini M, Simoni M, Sabbioni A. Genetic parameters estimation in an Italian horse native breed to support the conversion from agricultural uses to riding purposes. J Anim Breed Genet. 2020;137:200–10. doi:10.1111/jbg.12425.
76. Kalbfleisch TS, Rice ES, DePriest MS, Walenz BP, Hestand MS, Vermeesch JR, et al. Improved reference genome for the domestic horse increases assembly contiguity and

composition. Commun Biol. 2018;1:1–8.

77. Beeson SK, Schaefer RJ, Mason VC, McCue ME. Robust remapping of equine SNP array coordinates to EquCab3. Anim Genet. 2019;50:114–5. doi:10.1111/age.12745.

78. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MARR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75. doi:10.1086/519795.

79. 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. doi:10.3389/fgene.2015.00109.

80. Sved JA, Feldman MW. Correlation and probability methods for one and two loci. Theor Popul Biol. 1973;4:129–32. doi:10.1016/0040-5809(73)90008-7.

81. Beeson SK, Mickelson JR, McCue ME. Exploration of fine-scale recombination rate variation in the domestic horse. Genome Res. 2019;29:1744–52.

82. Pitt D, Bruford MW, Barbato M, Orozco-terWengel P, Martínez R, Sevane N. Demography and rapid local adaptation shape Creole cattle genome diversity in the tropics. Evol Appl. 2019;12:105–22.

83. Biscarini F, Cozzi P, Gaspa G, Marras G. detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes. 2019. https://cran.r-project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html#references.

84. R Development Core Team. R: A Language and Environment for Statistical Computing. 2011.

85. Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, et al. The Ensembl gene annotation system. Database. 2016;2016:baw093. doi:10.1093/database/baw093.

86. Karolchik D. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 2004;32:493D – 496.

87. Hu Z-L, Park CA, Reecy JM. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. Nucleic Acids Res. 2018;47 November 2018:1–10. doi:10.1093/nar/gky1084.

88. Schurink A, Wolc A, Ducro BJ, Frankena K, Garrick DJ, Dekkers JCM, et al. Genome-wide association study of insect bite hypersensitivity in two horse populations in the

Netherlands. Genet Sel Evol. 2012;44:1. doi:10.1186/1297-9686-44-31.

89. 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. doi:10.1111/jbg.12436.

90. Metzger J, Ohnesorge B, Distl O. Genome-wide linkage and association analysis identifies major gene loci for guttural pouch tympany in Arabian and German warmblood horses. PLoS One. 2012;7.

91. Haase B, Signer-Hasler H, Binns MM, Obexer-Ruff G, Hauswirth R, Bellone RR, et al. Accumulating Mutations in Series of Haplotypes at the KIT and MITF Loci Are Major Determinants of White Markings in Franches-Montagnes Horses. PLoS One. 2013;8:1–10.

92. 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;44:467–71. doi:10.1111/age.12031.

93. 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;8:1–12.

94. 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;7:1–6.

95. Skujina I, Winton CL, Hegarty MJ, McMahon R, Nash DM, Davies Morel MCG, 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. doi:10.1139/gen-2018-0006.

96. Na L, Minjun S, Wen H. Impacts of the Euro sovereign debt crisis on global trade and economic growth : A General Equilibrium Analysis based on GTAP model. In: 16th Annual Conference on Global Economic Analysis, Shanghai, China. 2013.

97. Sabbioni A, Beretti V, Zanon A, Pagani GP, Superchi P, Bonomi A, et al. Caratterizzazione demografica e variabilità genetica nel cavallo Bardigiano attraverso l'analisi di dati genealogici. In: 59th SISVet Congress. 2005. p. 451–2.

98. Olsen HF, Meuwissen T, Klemetsdal G. Optimal contribution selection applied to the

Norwegian and the North-Swedish cold-blooded trotter - a feasibility study. J Anim Breed Genet. 2013;130:170–7. doi:10.1111/j.1439-0388.2012.01005.x.

99. Druml T, Baumung R, Sölkner J. Pedigree analysis in the Austrian Noriker draught horse: genetic diversity and the impact of breeding for coat colour on population structure. J Anim Breed Genet. 2009;126:348–56. doi:10.1111/j.1439-0388.2008.00790.x.

100. Verrier E, Leroy G, Blouin C, Mériaux JC, Rognon X, Hospital F. Estimating the effective size of farm animals populations from Pedigree or molecular data: a case study on two French draught horse breeds. 9th World Congr Genet Appl to Livest Prod Leipzig. 2010; May.

101. Valera M, Molina A, Gutiérrez JP, Gómez J, Goyache F. Pedigree analysis in the Andalusian horse: population structure, genetic variability and influence of the Carthusian strain. Livest Prod Sci. 2005;95:57–66. doi:10.1016/j.livprodsci.2004.12.004.

102. Sørensen MK, Sørensen AC, Baumung R, Borchersen S, Berg P. Optimal genetic contribution selection in Danish Holstein depends on pedigree quality. Livest Sci. 2008;118:212–22. doi:10.1016/j.livsci.2008.01.027.

103. Fabbri MC, de Rezende MPG, Dadousis C, Biffani S, Negrini R, Carneiro PLS, et al. Population structure and genetic diversity of Italian beef breeds as a tool for planning conservation and selection strategies. Animals. 2019;9.

104. Mariani E, Summer A, Ablondi M, Sabbioni A. Genetic Variability and Management in Nero di Parma Swine Breed to Preserve Local Diversity. Animals. 2020;10:538. doi:10.3390/ani10030538.

105. Flury C, Tapio M, Sonstegard T, Drögemüller C, Leeb T, Simianer H, et al. Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. J Anim Breed Genet. 2010;127:339–47. doi:10.1111/j.1439-0388.2010.00862.x.

106. Goyache F, Álvarez I, Fernández I, Pérez-Pardal L, Royo LJ, Lorenzo L. Usefulness of molecular-based methods for estimating effective population size in livestock assessed using data from the endangered black-coated Asturcón pony. J Anim Sci. 2011;89:1251–9.

107. Sadeghi R, Moradi-Shahrbabak M, Ashtiani SRM, Schlamp F, Cosgrove EJ, Antczak

DF. Genetic diversity of Persian arabian horses and their relationship to other native iranian horse breeds. J Hered. 2019;110:173–82.

108. Druml T, Curik I, Baumung R, Aberle K, Distl O, Sölkner J. Individual-based assessment of population structure and admixture in Austrian, Croatian and German draught horses. Heredity (Edinb). 2007;98:114–22.

109. Gutiérrez JP, Cervantes I, Molina A, Valera M, Goyache F. Individual increase in inbreeding allows estimating effective sizes from pedigrees. Genet Sel Evol. 2008;40:359–78. doi:10.1051/gse:2008008.

110. Gutiérrez JP, Cervantes I, Goyache F. Improving the estimation of realized effective population sizes in farm animals. J Anim Breed Genet. 2009;126:327–32. doi:10.1111/j.1439-0388.2009.00810.x.

111. Kamiński S, Hering DM, Jaworski Z, Zabolewicz T, Ruść A. Assessment of genomic inbreeding in Polish Konik horses. Pol J Vet Sci. 2017;20:603–5.

112. Schurink A, Shrestha M, Eriksson S, Bosse M, Bovenhuis H, Back W, et al. The Genomic Makeup of Nine Horse Populations Sampled in the Netherlands. Genes (Basel). 2019;10:480. doi:10.3390/genes10060480.

113. Velie BD, Solé M, Fegraeus KJ, Rosengren MK, Røed KH, Ihler CF, et al. Genomic measures of inbreeding in the Norwegian-Swedish Coldblooded Trotter and their associations with known QTL for reproduction and health traits. Genet Sel Evol. 2019;51:1–10. doi:10.1186/s12711-019-0465-7.

114. Solé M, Gori AS, Faux P, Bertrand A, Farnir F, Gautier M, et al. Age-based partitioning of individual genomic inbreeding levels in Belgian Blue cattle. Genet Sel Evol. 2017;49:1–18. doi:10.1186/s12711-017-0370-x.

115. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic Patterns of Homozygosity in Worldwide Human Populations. Am J Hum Genet. 2012;91:275–92. doi:10.1016/j.ajhg.2012.06.014.

116. Takasuga A. PLAG1 and NCAPG-LCORL in livestock. Anim Sci J. 2016;87:159–67. doi:10.1111/asj.12417.

117. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010;467:832–8. doi:10.1038/nature09410.

118. Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson B V, Zusmanovich P, et al. Many sequence variants affecting diversity of adult human height. Nat Genet. 2008;40:609–15. doi:10.1038/ng.122.

119. Weedon MN, Lango H, Lindgren CM, Wallace C, Evans DM, Mangino M, et al. Genome-wide association analysis identifies 20 loci that influence adult height. Nat Genet. 2008;40:575–83. doi:10.1038/ng.121.

120. Flori L, Fritz S, Jaffrézic F, Boussaha M, Gut I, Heath S, et al. The Genome Response to Artificial Selection: A Case Study in Dairy Cattle. PLoS One. 2009;4:e6595. doi:10.1371/journal.pone.0006595.

121. Eberlein A, Takasuga A, Setoguchi K, Pfuhl R, Flisikowski K, Fries R, et al. Dissection of Genetic Factors Modulating Fetal Growth in Cattle Indicates a Substantial Role of the Non-SMC Condensin I Complex, Subunit G (NCAPG) Gene. Genetics. 2009;183:951–64. doi:10.1534/genetics.109.106476.

122. Loughran G, Huigsloot M, Kiely PA, Smith LM, Floyd S, Ayllon V, et al. Gene expression profiles in cells transformed by overexpression of the IGF-I receptor. Oncogene. 2005;24:6185–93.

123. Peeters LM, Janssens S, Coussé A, Buys N. Insect bite hypersensitivity in Belgian warmblood horses: Prevalence and risk factors. Vlaams Diergeneeskd Tijdschr. 2014;83:240–9.

124. Hauswirth R, Haase B, Blatter M, Brooks SA, Burger D, Drögemüller C, et al. Mutations in MITF and PAX3 cause “splashed white” and other white spotting phenotypes in horses. PLoS Genet. 2012;8.

125. Rieder S, Taourit S, Mariat D, Langlois B, Guérin G. Mutations in the agouti (ASIP), the extension (MCIR), and the brown (TYRP1) loci and their association to coat color phenotypes in horses (Equus caballus). Mamm Genome. 2001;12:450–5.

126. SWB. SWBs avelsplan. 2015. http://swb.org/wp-content/uploads/2016/11/Avelsplan-fîr-SWB.pdf.

127. Graaf K. Den svenska varmblodshästens historia under 200 år. Västervik, AB C O Ekblad & Co Tryckeri. 2004.

128. Viklund, Å., Granberg, L., Eriksson S. Genetic analysis of data from Swedish stallion performance test. In: Proc. 69th Annual Meeting of the EAAP. 2018. p. 660.

129. Viklund Å, Näsholm A, Strandberg E, Philipsson J. Genetic trends for performance of Swedish Warmblood horses. Livest Sci. 2011;141:113–22. doi:10.1016/j.livsci.2011.05.006.

130. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. Nature. 2007;449:913–8. doi:10.1038/nature06250.

131. Bomba L, Nicolazzi EL, Milanesi M, Negrini R, Mancini G, Biscarini F, et al. Relative extended haplotype homozygosity signals across breeds reveal dairy and beef specific signatures of selection. Genet Sel Evol. 2015;47:25. doi:10.1186/s12711-015-0113-9.

132. Randhawa IAS, Khatkar MS, Thomson PC, Raadsma HW. Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. BMC Genet. 2014;15:1–19.

133. Frischknecht M, Flury C, Leeb T, Rieder S, Neuditschko M. Selection signatures in Shetland ponies. Anim Genet. 2016;47:370–2. doi:10.1111/age.12416.

134. Moon S, Lee JW, Shin D, Shin KY, Kim JJ, Choi IY, et al. A Genome-wide scan for selective sweeps in racing horses. Asian-Australasian J Anim Sci. 2015;28:1525–31.

135. Marchiori CM, Pereira GL, Maiorano AM, Rogatto GM, Assoni AD, Augusto II V. Silva J, et al. Linkage disequilibrium and population structure characterization in the cutting and racing lines of Quarter Horses bred in Brazil. Livest Sci. 2019;219:45–51. doi:10.1016/j.livsci.2018.11.013.

136. Petersen JL, Valberg SJ, Mickelson JR, McCue ME. Haplotype diversity in the equine myostatin gene with focus on variants associated with race distance propensity and muscle fiber type proportions. Anim Genet. 2014;45:827–35. doi:10.1111/age.12205.

137. Avila F, Mickelson JR, Schaefer RJ, McCue ME. Genome-Wide Signatures of Selection Reveal Genes Associated With Performance in American Quarter Horse Subpopulations. Front Genet. 2018;9 JUL:249. doi:10.3389/fgene.2018.00249.

138. Lopes MS, MendonÃ§a D, Rojer H, Cabral V, Bettencourt SX, da CÃ¢mara Machado A. Morphological and genetic characterization of an emerging Azorean horse breed: the Terceira Pony. Front Genet. 2015;6. doi:10.3389/fgene.2015.00062.

139. Ovchinnikov I V., Dahms T, Herauf B, McCann B, Juras R, Castaneda C, et al. Genetic diversity and origin of the feral horses in Theodore Roosevelt National Park. PLoS One.

2018;13:1–18.

140. Hendricks BL. International encyclopedia of horse breeds. 2007.

141. 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. doi:10.1371/journal.pone.0152966.

142. Schaefer RJ, Schubert M, Bailey E, Bannasch DL, Barrey E, Bar-Gal GK, et al. Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds. BMC Genomics. 2017;18:565. doi:10.1186/s12864-017-3943-8.

143. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010;38 suppl_2:W214–20. doi:10.1093/nar/gkq537.

144. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B. PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. Genome Res. 2003;13:2129–41. doi:10.1101/gr.772403.

145. Nei M. Molecular evolutionary genetics. Columbia Univ Press. 1987.

146. Qanbari S, Gianola D, Hayes B, Schenkel F, Miller S, Moore S, et al. Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. BMC Genomics. 2011;12:318. doi:10.1186/1471-2164-12-318.

147. Delaneau O, Coulonges C, Zagury J-F. Shape-IT: new rapid and accurate algorithm for haplotype inference. BMC Bioinformatics. 2008;9:540. doi:10.1186/1471-2105-9-540.

148. Gautier M, Vitalis R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics. 2012;28:1176–7. doi:10.1093/bioinformatics/bts115.

149. 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.

150. Kalbfleisch TS, Rice E, DePriest MS, Walenz BP, Hestand MS, Vermeesch JR, et al. EquCab3, an Updated Reference Genome for the Domestic Horse. bioRxiv.

2018;:306928. doi:10.1101/306928.

151. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science (80-). 2009;326:865–7. doi:10.1126/science.1178158.

152. Mardia KV. Some properties of clasical multi-dimesional scaling. Commun Stat - Theory Methods. 1978;7:1233–41. doi:10.1080/03610927808827707.

153. R Foundation for Statistical Computing. R: A language and environment for statistical computing. R Core Team. 2018.

154. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet. 2010;11:94. doi:10.1186/1471-2156-11-94.

155. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24:1403–5. doi:10.1093/bioinformatics/btn129.

156. Akey JM. Interrogating a High-Density SNP Map for Signatures of Natural Selection. Genome Res. 2002;12:1805–14. doi:10.1101/gr.631202.

157. Wright S. Evolution and the genetics of populations. Volume 4: variability within and among natural populations. University of Chicago Press, Chicago, IL, USA.; 1978.

158. Makina SO, Muchadeyi FC, Van Marle-Köster E, Taylor JF, Makgahlela ML, Maiwashe A. Genome-wide scan for selection signatures in six cattle breeds in South Africa. Genet Sel Evol. 2015;47:1–14.

159. Zhao F, McParland S, Kearney F, Du L, Berry DP. Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genet Sel Evol. 2015;47:49. doi:10.1186/s12711-015-0127-3.

160. Gautier M, Naves M. Footprints of selection in the ancestral admixture of a New World Creole cattle breed. Mol Ecol. 2011;20:3128–43. doi:10.1111/j.1365-294X.2011.05163.x.

161. Benjamini Y, Hochberg Y. Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. J R Stat Soc. 1995;Methodolog:289–300. doi:10.2307/2346101.

162. Thomas PD, Kejariwal A, Guo N, Mi H, Campbell MJ, Muruganujan A, et al. Applications for protein sequence-function evolution data: mRNA/protein expression

analysis and coding SNP scoring tools. Nucleic Acids Res. 2006;34 Web Server:W645–50. doi:10.1093/nar/gkl229.

163. Sevane N, Dunner S, Boado A, Cañon J. Polymorphisms in ten candidate genes are associated with conformational and locomotive traits in Spanish Purebred horses. J Appl Genet. 2017;58:355–61. doi:10.1007/s13353-016-0385-y.

164. Rare Breeds Survival Trust. 2014. https://www.rbst.org.uk/Rare-and-Native-Breeds/Equine/Exmoor-Pony.

165. Vitti JJ, Grossman SR, Sabeti PC. Detecting Natural Selection in Genomic Data. Annu Rev Genet. 2013;47:97–120. doi:10.1146/annurev-genet-111212-133526.

166. Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, et al. Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 2010;464:587–91. doi:10.1038/nature08832.

167. Auton A, Bryc K, Boyko AR, Lohmueller KE, Novembre J, Reynolds A, et al. Global distribution of genomic diversity underscores rich complex history of continental human populations. Genome Res. 2009;19:795–803. doi:10.1101/gr.088898.108.

168. Slatkin M. Linkage disequilibrium — understanding the evolutionary past and Mapping the Medical Future. Nat Rev Genet. 2016;9:477–85.

169. Méjat A, Ramond F, Bassel-Duby R, Khochbin S, Olson EN, Schaeffer L. Histone deacetylase 9 couples neuronal activity to muscle chromatin acetylation and gene expression. Nat Neurosci. 2005;8:313–21.

170. Paoletti P, Bellone C, Zhou Q. NMDA receptor subunit diversity: impact on receptor properties, synaptic plasticity and disease. Nat Rev Neurosci. 2013;14:383–400. doi:10.1038/nrn3504.

171. Jäderkvist Fegraeus K, Hirschberg I, Árnason T, Andersson L, Velie BD, Andersson LS, 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.

172. 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. doi:10.1186/s12863-018-0670-3.

173. Lipscombe D, Andrade A, Allen SE. Alternative splicing: Functional diversity among

voltage-gated calcium channels and behavioral consequences. Biochim Biophys Acta - Biomembr. 2013;1828:1522–9. doi:10.1016/j.bbamem.2012.09.018.

174. Åberg ND, Brywe KG, Isgaard J. Aspects of Growth Hormone and Insulin-Like Growth Factor-I Related to Neuroprotection, Regeneration, and Functional Plasticity in the Adult Brain. Sci World J. 2006;6:53–80. doi:10.1100/tsw.2006.22.

175. Damaj L, Lupien-Meilleur A, Lortie A, Riou É, Ospina LH, Gagnon L, et al. CACNA1A haploinsufficiency causes cognitive impairment, autism and epileptic encephalopathy with mild cerebellar symptoms. Eur J Hum Genet. 2015;23:1505–12. doi:10.1038/ejhg.2015.21.

176. Zeng L, Ming C, Li Y, Su LY, Su YH, Otecko NO, et al. Rapid evolution of genes involved in learning and energy metabolism for domestication of the laboratory rat. Mol Biol Evol. 2017;34:3148–53.

177. Thorén Hellsten E, Jorjani H, Philipsson J. Connectedness among five European sport horse populations. Livest Sci. 2008;118:147–56. doi:10.1016/j.livsci.2007.12.028.

178. Zhang W, Han Q, Liu Z, Zheou W, Cao Q, Zhou W. Exome sequencing reveals a de novo PRKG1 mutation in a sporadic patient with aortic dissection. BMC Med Genet. 2018;19:1–5.

179. Chen M, Pan D, Ren H, Fu J, Li J, Su G, et al. Identification of selective sweeps reveals divergent selection between Chinese Holstein and Simmental cattle populations. Genet Sel Evol. 2016;48:76. doi:10.1186/s12711-016-0254-5.

180. Wieczorek L, Majumdar D, Wills TA, Hu L, Winder DG, Webb DJ, et al. Absence of Ca2+-stimulated adenylyl cyclases leads to reduced synaptic plasticity and impaired experience-dependent fear memory. Transl Psychiatry. 2012;2:e126–e126. doi:10.1038/tp.2012.50.

181. Bosse KE, Charlton JL, Susick LL, Newman B, Eagle AL, Mathews TA, et al. Deficits in behavioral sensitization and dopaminergic responses to methamphetamine in adenylyl cyclase 1/8-deficient mice. J Neurochem. 2015;135:1218–31. doi:10.1111/jnc.13235.

182. Wieczorek L, Maas JW, Muglia LM, Vogt SK, Muglia LJ. Temporal and Regional Regulation of Gene Expression by Calcium-Stimulated Adenylyl Cyclase Activity during Fear Memory. PLoS One. 2010;5:e13385. doi:10.1371/journal.pone.0013385.

183. Gong R, Ding C, Hu JJ, Lu Y, Liu F, Mann E, et al. Role for the membrane receptor guanylyl cyclase-C in attention deficiency and hyperactive behavior. Science (80-). 2011;333:1642–6. doi:10.1126/science.1207675.

184. Gharibi B, Ghuman MS, Cama G, Rawlinson SCF, Grigoriadis AE, Hughes FJ. Site-specific differences in osteoblast phenotype, mechanical loading response and estrogen receptor-related gene expression. Mol Cell Endocrinol. 2018;477:140–7. doi:10.1016/j.mce.2018.06.011.

185. Olofsson B. Rho guanine dissociation inhibitors: Pivotal molecules in cellular signalling. Cell Signal. 1999;11:545–54.

186. Wang X, Harimoto K, Liu J, Guo J, Hinshaw S, Chang Z, et al. Spata4 promotes osteoblast differentiation through Erk-activated Runx2 pathway. J Bone Miner Res. 2011;26:1964–73.

187. Fujimoto H, Ogi T, Mimura J, Hikida M, Ohmori H, Fujii-Kuriyama Y. Expression of human and mouse genes encoding polkappa: testis-specific developmental regulation and AhR-dependent inducible transcription. Genes to Cells. 2003;6:943–53.

188. Park SJ, Kim SJ, Rhee Y, Byun JH, Kim SH, Kim MH, et al. Fidgetin-Like 1 Gene Inhibited by Basic Fibroblast Growth Factor Regulates the Proliferation and Differentiation of Osteoblasts. J Bone Miner Res. 2007;22:889–96. doi:10.1359/jbmr.070311.

189. Mori S, Kou I, Sato H, Emi M, Ito H, Hosoi T, et al. Association of genetic variations of genes encoding thrombospondin, type 1, domain-containing 4 and 7A with low bone mineral density in Japanese women with osteoporosis. J Hum Genet. 2008;53:694–7. doi:10.1007/s10038-008-0300-4.

190. Kawane K, Tanaka H, Kitahara Y, Shimaoka S, Nagata S. Cytokine-dependent but acquired immunity-independent arthritis caused by DNA escaped from degradation. Proc Natl Acad Sci. 2010;107:19432–7. doi:10.1073/pnas.1010603107.

191. Viklund Å, Braam Å, Näsholm A, Strandberg E, Philipsson J. Genetic variation in competition traits at different ages and time periods and correlations with traits at field tests of 4-year-old Swedish Warmblood horses. animal. 2010;4:682–91. doi:10.1017/S1751731110000017.

192. Hellsten ET, Näsholm A, Jorjani H, Strandberg E, Philipsson J. Influence of foreign

stallions on the Swedish Warmblood breed and its genetic evaluation. Livest Sci. 2009;121:207–14. doi:10.1016/j.livsci.2008.06.014.

193. Storz JF. Invited review: Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol Ecol. 2005;14:671–88. doi:10.1111/j.1365-294X.2005.02437.x.

194. Doan R, Cohen N, Harrington J, Veazy K, Juras R, Cothran G, et al. Identification of copy number variants in horses. Genome Res. 2012;22:899–907.

195. Lee W, Park K-D, Taye M, Lee C, Kim H, Lee H-K, et al. Analysis of cross-population differentiation between Thoroughbred and Jeju horses. Asian-Australasian J Anim Sci. 2018;31:1110–8. doi:10.5713/ajas.17.0460.

196. Nekrasova T, Jobes ML, Ting JH, Wagner GC, Minden A. Targeted disruption of the Pak5 and Pak6 genes in mice leads to deficits in learning and locomotion. Dev Biol. 2008;322:95–108. doi:10.1016/j.ydbio.2008.07.006.

197. Carlisle HJ, Luong TN, Medina-Marino A, Schenker L, Khorosheva E, Indersmitten T, et al. Deletion of Densin-180 Results in Abnormal Behaviors Associated with Mental Illness and Reduces mGluR5 and DISC1 in the Postsynaptic Density Fraction. J Neurosci. 2011;31:16194–207. doi:10.1523/JNEUROSCI.5877-10.2011.

198. Udawela M, Scarr E, Hannan AJ, Thomas EA, Dean B. Phospholipase C beta 1 expression in the dorsolateral prefrontal cortex from patients with schizophrenia at different stages of illness. Aust N Z J Psychiatry. 2011;45:140–7.

199. Weng YT, Chien T, Kuan II, Chern Y. The TRAX, DISC1, and GSK3 complex in mental disorders and therapeutic interventions 06 Biological Sciences 0604 Genetics 11 Medical and Health Sciences 1103 Clinical Sciences. J Biomed Sci. 2018;25:1–14.

200. Fernandes JCR, Acuña SM, Aoki JI, Floeter-Winter LM, Muxel SM. Long non-coding RNAs in the regulation of gene expression: Physiology and disease. Non-coding RNA. 2019;5.

201. Qureshi IA, Mehler MF. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. Nat Rev Neurosci. 2012;13:528–41. doi:10.1038/nrn3234.

202. Carneiro M, Rubin C-J, Di Palma F, Albert FW, Alfoldi J, Barrio AM, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication.

Science (80-). 2014;345:1074–9. doi:10.1126/science.1253714.
203. Schröder W, Klostermann A, Stock KF, Distl O. A genome-wide association study for quantitative trait loci of show-jumping in Hanoverian warmblood horses. Anim Genet. 2012;43:392–400. doi:10.1111/j.1365-2052.2011.02265.x.
204. Napoli E, Ross-Inta C, Wong S, Hung C, Fujisawa Y, Sakaguchi D, et al. Mitochondrial dysfunction in Pten Haplo-insufficient mice with social deficits and repetitive behavior: Interplay between Pten and p53. PLoS One. 2012;7:1–13.
205. Kwon CH, Luikart BW, Powell CM, Zhou J, Matheny SA, Zhang W, et al. Pten Regulates Neuronal Arborization and Social Interaction in Mice. Neuron. 2006;50:377–88. doi:10.1016/j.neuron.2006.03.023.
206. Rankinen T, Roth SM, Bray MS, Loos R, Pérusse L, Wolfarth B, et al. Authors and Disclosures From Medicine and Science in Sports and Exercise® Advances in Exercise, Fitness, and Performance Genomics. Med Sci Sports Exerc. 2010;42:835–46. doi:10.1249/MSS.0000000000000300.
207. Coricor G, Serra R. TGF-β regulates phosphorylation and stabilization of Sox9 protein in chondrocytes through p38 and Smad dependent mechanisms. Sci Rep. 2016;6 August:1–11. doi:10.1038/srep38616.
208. Chavez RD, Coricor G, Perez J, Seo HS, Serra R. SOX9 protein is stabilized by TGF-β and regulates PAPSS2 mRNA expression in chondrocytes. Osteoarthr Cartil. 2017;25:332–40.
209. Loeys BL, Schwarze U, Holm T, Callewaert BL, Thomas GH, Pannu H, et al. Aneurysm Syndromes Caused by Mutations in the TGF-β Receptor. N Engl J Med. 2006;355:788–98. doi:10.1056/NEJMoa055695.
210. Borza CM, Su Y, Tran T-L, Yu L, Steyns N, Temple KJ, et al. Discoidin domain receptor 1 kinase activity is required for regulating collagen IV synthesis. Matrix Biol. 2017;57–58:258–71. doi:10.1016/j.matbio.2016.11.009.
211. Chiusa M, Hu W, Liao H-J, Su Y, Borza CM, de Caestecker MP, et al. The Extracellular Matrix Receptor Discoidin Domain Receptor 1 Regulates Collagen Transcription by Translocating to the Nucleus. J Am Soc Nephrol. 2019;30:1605–24. doi:10.1681/ASN.2018111160.
212. Eklund L, Piuhola J, Komulainen J, Sormunen R, Ongvarrasopone C, Fässler R, et al.

Lack of type XV collagen causes a skeletal myopathy and cardiovascular defects in mice. Proc Natl Acad Sci U S A. 2001;98:1194–9.

213. Lim, Lindert, Opitz, Hausser, Rohrbach, Giunta. Transcriptome Profiling of Primary Skin Fibroblasts Reveal Distinct Molecular Features Between PLOD1- and FKBP14-Kyphoscoliotic Ehlers–Danlos Syndrome. Genes (Basel). 2019;10:517. doi:10.3390/genes10070517.

214. Castori M, Tinkle B, Levy H, Grahame R, Malfait F, Hakim A. A framework for the classification of joint hypermobility and related conditions. Am J Med Genet Part C Semin Med Genet. 2017;175:148–57.

215. Monthoux C, de Brot S, Jackson M, Bleul U, Walter J. Skin malformations in a neonatal foal tested homozygous positive for Warmblood Fragile Foal Syndrome. BMC Vet Res. 2015;11:12. doi:10.1186/s12917-015-0318-8.

216. Giunta C, Elçioglu NH, Albrecht B, Eich G, Chambaz C, Janecke AR, et al. Spondylocheiro Dysplastic Form of the Ehlers-Danlos Syndrome-An Autosomal-Recessive Entity Caused by Mutations in the Zinc Transporter Gene SLC39A13. Am J Hum Genet. 2008;82:1290–305.

217. Mushtaq M, Gaza HV, Kashuba EV. Role of the RB-Interacting Proteins in Stem Cell Biology. 2016. p. 133–57. doi:10.1016/bs.acr.2016.04.002.

218. Minocherhomji S, Hansen C, Kim HG, Mang Y, Bak M, Guldberg P, et al. Epigenetic remodelling and dysregulation of DLGAP4 is linked with early-onset cerebellar ataxia. Hum Mol Genet. 2014;23:6163–76.

219. Rasmussen AH, Rasmussen HB, Silahtaroglu A. The DLGAP family: Neuronal expression, function and role in brain disorders. Molecular Brain. 2017;10:1–13.

220. Banerjee A, Wang HY, Borgmann-Winter KE, MacDonald ML, Kaprielian H, Stucky A, et al. Src kinase as a mediator of convergent molecular abnormalities leading to NMDAR hypoactivity in schizophrenia. Mol Psychiatry. 2015;20:1091–100. doi:10.1038/mp.2014.115.

221. Schob C, Morellini F, Ohana O, Bakota L, Hrynchak M V., Brandt R, et al. Cognitive impairment and autistic-like behaviour in SAPAP4-deficient mice. Transl Psychiatry. 2019;9.

222. Ellenbroek SIJ, Collard JG. Rho GTPases: functions and association with cancer. Clin

Exp Metastasis. 2007;24:657–72. doi:10.1007/s10585-007-9119-1.

223. Lowe C, Yoneda T, Boycet BF, Chent H, Mundy GR, Sorianot P. Osteopetrosis in Src-deficient mice is due to an autonomous defect of osteoclasts. Proc Natl Acad Sci USA. 1993;90 May:4485–9.

224. Davidson TB, Sanchez-Lara PA, Randolph LM, Krieger MD, Wu SQ, Panigrahy A, et al. Microdeletion del(22)(q12.2) encompassing the facial development-associated gene, MN1 (meningioma 1) in a child with Pierre-Robin sequence (including cleft palate) and neurofibromatosis 2 (NF2): A case report and review of the literature. BMC Med Genet. 2012;13:19. doi:10.1186/1471-2350-13-19.

225. Hu PQ, Fertig N, Medsger TA, Wright TM. Molecular Recognition Patterns of Serum Anti-DNA Topoisomerase I Antibody in Systemic Sclerosis. J Immunol. 2014;173:2834–41.

226. Nevitt C, Tooley JG, Schaner Tooley CE. N-terminal acetylation and methylation differentially affect the function of MYL9. Biochem J. 2018;475:3201–19. doi:10.1042/BCJ20180638.

227. Rylatt DB, Aitken A, Bilham T, Condon GD, Embi N, Cohen P. Glycogen Synthase from Rabbit Skeletal Muscle. Eur J Biochem. 2005;107:529–37. doi:10.1111/j.1432-1033.1980.tb06060.x.

228. Yang S, Li X, Liu X, Ding X, Xin X, Jin C, et al. Parallel comparative proteomics and phosphoproteomics reveal that cattle myostatin regulates phosphorylation of key enzymes in glycogen metabolism and glycolysis pathway. Oncotarget. 2018;9:11352–70. doi:10.18632/oncotarget.24250.

229. Lewis SS, Nicholson AM, Williams ZJ, Valberg SJ. Clinical characteristics and muscle glycogen concentrations in warmblood horses with polysaccharide storage myopathy. Am J Vet Res. 2017;78:1305–12. doi:10.2460/ajvr.78.11.1305.

230. Zhou J, Blundell J, Ogawa S, Kwon CH, Zhang W, Sinton C, et al. Pharmacological inhibition of mTORCl suppresses anatomical, cellular, and behavioral abnormalities in neural-specific PTEN knock-out mice. J Neurosci. 2009;29:1773–83.

231. Hassed S, Li S, Mulvihill J, Aston C, Palmer S. Adams–Oliver syndrome review of the literature: Refining the diagnostic phenotype. Am J Med Genet Part A. 2017;173:790–800.

232. Schindler RFR, Scotton C, Zhang J, Passarelli C, Ortiz-Bonnin B, Simrick S, et al. POPDC1S201F causes muscular dystrophy and arrhythmia by affecting protein trafficking. J Clin Invest. 2015;126:239–53. doi:10.1172/JCI79562.

233. Swan AH, Gruscheski L, Boland LA, Brand T. The Popeye domain containing gene family encoding a family of cAMP-effector proteins with important functions in striated muscle and beyond. J Muscle Res Cell Motil. 2019.

234. Hendrickson SL. A genome wide study of genetic adaptation to high altitude in feral Andean Horses of the páramo. BMC Evol Biol. 2013;13.

235. Shin JH, Adrover MF, Wess J, Alvarez VA. Muscarinic regulation of dopamine and glutamate transmission in the nucleus accumbens. Proc Natl Acad Sci. 2015;112:8124–9.

236. Kader A, Liu X, Dong K, Song S, Pan J, Yang M, et al. Identification of copy number variations in three Chinese horse breeds using 70K single nucleotide polymorphism BeadChip array. Anim Genet. 2016;47:560–9.

237. Ueda H, Sasaki K, Halder SK, Deguchi Y, Takao K, Miyakawa T, et al. Prothymosin alpha-deficiency enhances anxiety-like behaviors and impairs learning/memory functions and neurogenesis. J Neurochem. 2017;141:124–36.

238. George EM, Brown DT. Prothymosin α is a component of a linker histone chaperone. FEBS Lett. 2010;584:2833–6. doi:10.1016/j.febslet.2010.04.065.

239. McNeill EM, Klöckner-Bormann M, Roesler EC, Talton LE, Moechars D, Clagett-Dame M. Nav2 hypomorphic mutant mice are ataxic and exhibit abnormalities in cerebellar development. Dev Biol. 2011;353:331–43.

240. Lanuza GM, Gosgnach S, Pierani A, Jessell TM, Goulding M. Genetic identification of spinal interneurons that coordinate left-right locomotor activity necessary for walking movements. Neuron. 2004;42:375–86.

241. Griener A, Zhang W, Kao H, Haque F, Gosgnach S. Anatomical and electrophysiological characterization of a population of dI6 interneurons in the neonatal mouse spinal cord. Neuroscience. 2017;362:47–59.

242. Andersson LLS, Larhammar M, Memic F, Wootz H, Schwochow D, Rubin C-JJ, et al. Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. Nature. 2012;488:642–6. doi:10.1038/nature11399.

Appendix I

Supplementary Table 1 Genes located within consensus homozygous segments in Exmoor ponies.

Genes table ID	Gene name	ECA
ENSECAG00000000016	FAM149B1	1
ENSECAG00000000353	HPSE2	1
ENSECAG00000001038		1
ENSECAG00000001924		1
ENSECAG00000002062		1
ENSECAG00000002815		1
ENSECAG00000004257	DNAJC9	1
ENSECAG00000007097	CRTAC1	1
ENSECAG00000007985	GOLGA7B	1
ENSECAG00000008070	ZFYVE27	1
ENSECAG00000008577	PPP3CB	1
ENSECAG00000010218	CFAP70	1
ENSECAG00000011253	SECISBP2L	1
ENSECAG00000012316	DTWD1	1
ENSECAG00000013079	USP54	1
ENSECAG00000013743	FAM227B	1
ENSECAG00000015163	SHC4	1
ENSECAG00000016676	SEMA6D	1
ENSECAG00000016729		1
ENSECAG00000017631	GALK2	1
ENSECAG00000017677	HERC2	1
ENSECAG00000017875		1
ENSECAG00000019666	SLC12A1	1
ENSECAG00000026853	CEP152	1
ENSECAG00000026876		1
ENSECAG00000022704	ECD	1
ENSECAG00000024709	COPS2	1
ENSECAG00000004877		11
ENSECAG00000008172		11
ENSECAG00000009723	IGF2BP1	11
ENSECAG00000013064		11
ENSECAG00000013134		11
ENSECAG00000013143		11
ENSECAG00000013354	B4GALNT2	11
ENSECAG00000014174		11
ENSECAG00000015552	GDPD1	11
ENSECAG00000016585	SUPT4H1	11
ENSECAG00000016953	HLF	11
ENSECAG00000017737	RNF43	11
ENSECAG00000018575	YPEL2	11
ENSECAG00000021160	UBE2Z	11
ENSECAG00000022092	SNF8	11
ENSECAG00000027279	RF00012	11
ENSECAG00000026538	RF00586	11
ENSECAG00000001252	PCDHB14	14

ENSECAG00000001934		14
ENSECAG00000002973		14
ENSECAG00000003046		14
ENSECAG00000004740		14
ENSECAG00000005331	ISOC1	14
ENSECAG00000005465	PCDHGA2	14
ENSECAG00000005500	PCDHGA1	14
ENSECAG00000005542	SLC25A2	14
ENSECAG00000005591	PCDHB15	14
ENSECAG00000005655		14
ENSECAG00000006848	SLC27A6	14
ENSECAG00000007418	PCDHAC2	14
ENSECAG00000007513	HSPA9	14
ENSECAG00000007923		14
ENSECAG00000009100	SLCO4C1	14
ENSECAG00000009534	PCDHA12	14
ENSECAG00000012445	CHD1	14
ENSECAG00000012655	CYSTM1	14
ENSECAG00000013041	ETF1	14
ENSECAG00000014338	EGR1	14
ENSECAG00000015036	ANKHD1	14
ENSECAG00000015317	REEP2	14
ENSECAG00000016123	PCDHGC5	14
ENSECAG00000016183	STARD4	14
ENSECAG00000016696	CAMK4	14
ENSECAG00000017001	WDR36	14
ENSECAG00000018062	KDM3B	14
ENSECAG00000018369	RGMB	14
ENSECAG00000022141		14
ENSECAG00000023991	CTNNA1	14
ENSECAG00000026692	RF00015	14
ENSECAG00000025989	RF00090	14
ENSECAG00000025479	RF00154	14
ENSECAG00000027617	RF00154	14
ENSECAG00000000699	SLC6A1	16
ENSECAG00000001022	UCN2	16
ENSECAG00000003695		16
ENSECAG00000006875	IP6K2	16
ENSECAG00000008659	SHISA5	16
ENSECAG00000010761	TREX1	16
ENSECAG00000010784	ATRIP	16
ENSECAG00000011234		16
ENSECAG00000012037	NCKIPSD	16
ENSECAG00000013120	SLC26A6	16
ENSECAG00000014159		16
ENSECAG00000014543	CCDC51	16
ENSECAG00000014767	PLXNB1	16
ENSECAG00000015069	BHLHE40	16
ENSECAG00000016006	CELSR3	16
ENSECAG00000017338	PFKFB4	16
ENSECAG00000020016	TMEM89	16

ENSECAG00000020020	UQCRC1	16
ENSECAG00000020184	VGLL4	16
ENSECAG00000020933	ITPR1	16
ENSECAG00000026276	eca-mir-711	16
ENSECAG00000000996		18
ENSECAG00000001030		18
ENSECAG00000013822	KIF5C	18
ENSECAG00000014080	SCN3A	18
ENSECAG00000015805	MARCH7	18
ENSECAG00000018415	TANC1	18
ENSECAG00000019894	RBMS1	18
ENSECAG00000020320	CD302	18
ENSECAG00000022703	LY5	18
ENSECAG00000024839	MBD5	18
ENSECAG00000026355	RF00026	18
ENSECAG00000008905	MCF2L2	19
ENSECAG00000015809	ZBTB20	19
ENSECAG00000016912	ATP11B	19
ENSECAG00000000907		2
ENSECAG00000003961	BBS12	2
ENSECAG00000007347	MFSD8	2
ENSECAG00000009877	HSPA4L	2
ENSECAG00000010816		2
ENSECAG00000010831	SLC25A31	2
ENSECAG00000011414	DAB1	2
ENSECAG00000011831	INTU	2
ENSECAG00000014312	PRKCZ	2
ENSECAG00000019168	GABRD	2
ENSECAG00000024212	MYSM1	2
ENSECAG00000024492		2
ENSECAG00000024272	RIMS1	20
ENSECAG00000002884		22
ENSECAG00000004657		22
ENSECAG00000004797		22
ENSECAG00000007522	TMX4	22
ENSECAG00000007706	PHACTR3	22
ENSECAG00000010314	HAO1	22
ENSECAG00000018583		22
ENSECAG00000021201	BMP2	22
ENSECAG00000026497	RF00548	22
ENSECAG00000000122	GDA	23
ENSECAG00000006788	PCSK5	23
ENSECAG00000007568		23
ENSECAG00000008123		23
ENSECAG00000009671		23
ENSECAG00000010004		23
ENSECAG00000011532		23
ENSECAG00000015794	ANXA1	23
ENSECAG00000015973	RCL1	23
ENSECAG00000018363	GNAQ	23
ENSECAG00000019624	OSTF1	23

ENSECAG00000020544	GLIS3	23
ENSECAG00000020865	VPS13A	23
ENSECAG00000021484	CEMIP2	23
ENSECAG00000021902	SPATA6L	23
ENSECAG00000023826	ZFAND5	23
ENSECAG00000025900	RF00026	23
ENSECAG00000020433	TENM3	27
ENSECAG00000000599	PCDH7	3
ENSECAG00000002673		3
ENSECAG00000009164	PPARGC1A	3
ENSECAG00000018240	PHKB	3
ENSECAG00000019489		3
ENSECAG00000021168	ABCC11	3
ENSECAG00000022436	LONP2	3
ENSECAG00000024558		3
ENSECAG00000024818		3
ENSECAG00000014636	WDR27	31
ENSECAG00000018858	PHF10	31
ENSECAG00000026735	RF00100	31
ENSECAG00000002370	GPR141	4
ENSECAG00000003867		4
ENSECAG00000004761		4
ENSECAG00000006765	CDK14	4
ENSECAG00000008474		4
ENSECAG00000012115	SEMA3D	4
ENSECAG00000013914		4
ENSECAG00000021252	ITPRID1	4
ENSECAG00000021847	PPP1R17	4
ENSECAG00000022057	BBS9	4
ENSECAG00000026084	RF00026	4
ENSECAG00000000450	TMCO1	5
ENSECAG00000003566		5
ENSECAG00000006120		5
ENSECAG00000012244		5
ENSECAG00000014292	CR1	5
ENSECAG00000022247		5
ENSECAG00000000564		6
ENSECAG00000009337	TSPAN9	6
ENSECAG00000010929	MRPS35	6
ENSECAG00000012527		6
ENSECAG00000012558	A2M	6
ENSECAG00000013446	MANSC4	6
ENSECAG00000013455	KLHL42	6
ENSECAG00000014331	PRMT8	6
ENSECAG00000014708	PTHLH	6
ENSECAG00000014912	CCDC91	6
ENSECAG00000015666	PHC1	6
ENSECAG00000016652	PPFIBP1	6
ENSECAG00000021038		6
ENSECAG00000021235		6
ENSECAG00000023247	GRIN2B	6

ENSECAG00000024571		6
ENSECAG00000027409	RF01169	6
ENSECAG00000000878		7
ENSECAG00000001374		7
ENSECAG00000001466	PLEKHB1	7
ENSECAG00000002421		7
ENSECAG00000002448	OR51A7	7
ENSECAG00000002654	OR51G2	7
ENSECAG00000003751		7
ENSECAG00000003766		7
ENSECAG00000004301		7
ENSECAG00000005126		7
ENSECAG00000006096		7
ENSECAG00000006917	P4HA3	7
ENSECAG00000007348	TTC36	7
ENSECAG00000007379	KCNE3	7
ENSECAG00000007430		7
ENSECAG00000007743	KMT2A	7
ENSECAG00000007949	AKIP1	7
ENSECAG00000008164	CALR	7
ENSECAG00000008973	TMEM25	7
ENSECAG00000009012	IER2A	7
ENSECAG00000009498		7
ENSECAG00000010509		7
ENSECAG00000010608		7
ENSECAG00000010687		7
ENSECAG00000010696		7
ENSECAG00000010781	RNF169	7
ENSECAG00000011964	IL27RA	7
ENSECAG00000012408	SYCE2	7
ENSECAG00000012495	GCDH	7
ENSECAG00000012624	POLD3	7
ENSECAG00000012790		7
ENSECAG00000012804	DENND5A	7
ENSECAG00000013426		7
ENSECAG00000014897	RLN3	7
ENSECAG00000015139	FARSA	7
ENSECAG00000015582		7
ENSECAG00000015965	IPO7	7
ENSECAG00000015967	ASF1B	7
ENSECAG00000017517	LIPT2	7
ENSECAG00000018803	PGM2L1	7
ENSECAG00000018889	ZNF143	7
ENSECAG00000020398	STK33	7
ENSECAG00000020774	WEE1	7
ENSECAG00000020829		7
ENSECAG00000022115	MRPL48	7
ENSECAG00000022426	SWAP70	7
ENSECAG00000022671		7
ENSECAG00000022679		7
ENSECAG00000022685		7

ENSECAG00000022694	RAB6A	7
ENSECAG00000022805		7
ENSECAG00000022810	ST5	7
ENSECAG00000022876		7
ENSECAG00000024207		7
ENSECAG00000024508	DAND5	7
ENSECAG00000024607	RAD23A	7
ENSECAG00000026407	RF00026	7
ENSECAG00000009506		9
ENSECAG00000010100		9
ENSECAG00000010182	RGS22	9
ENSECAG00000010526	STK3	9
ENSECAG00000018644		9
ENSECAG00000019852	VPS13B	9
ENSECAG00000025757		9
ENSECAG00000026298	MIR599	9

Appendix II

Supplementary Table 2 Details on the five overrepresented biological processes and the genes involved in each biological term from the ROH analysis.

PANTHER GO-Slim Biological Process	Equus caballus - Reference list (20497)	Gene list	P-value
actin cytoskeleton reorganization (GO:0031532)	17	*THSD7A*	0.0157
cellular macromolecule catabolic process (GO:0044265)	18	*DNASE2*	0.0165
apoptotic process (GO:0006915)	264	*HDAC9, DNASE2*	0.0224
glycoprotein metabolic process (GO:0009100)	29	*B3GAT1*	0.0260
synaptic signaling (GO:0099536)	314	*CACNA1A, GRIN2B*	0.0308

Appendix III

Supplementary Table 3 SNP markers found to be under potential selection when analysed by the XP-EHH test before FDR control. SNP markers still significant after FDR control are indicated in bold.

MARKER	CHR	POSITION	XP-EHH	LOG10(p-value)
AX-103660314	1	22,450,649	4.63	5.453
AX-104568168	1	22,499,055	4.95	6.145
AX-103761897	1	22,506,483	4.77	5.754
AX-103872914	1	22,507,754	4.6	5.388
AX-104069920	1	22,511,416	4.47	5.125
AX-103175323	**1**	**148,660,078**	**5.02**	**6.306**
AX-103108018	**1**	**148,660,387**	**5.25**	**6.835**
AX-104738367	**1**	**148,662,209**	**4.67**	**5.525**
AX-104145670	**1**	**148,663,404**	**4.86**	**5.936**
AX-104576862	**1**	**148,665,804**	**4.83**	**5.883**
AX-104674825	2	74,631,100	4.38	4.93
AX-104481381	2	75,008,834	4.91	6.045
AX-104298228	**4**	**15,977,755**	**4.92**	**6.074**
AX-103314178	**4**	**15,978,736**	**4.6**	**5.389**
AX-104487403	**4**	**15,990,296**	**4.35**	**4.869**
AX-104845778	**4**	**16,034,364**	**4.39**	**4.951**
AX-104196592	**4**	**16,039,971**	**4.4**	**4.98**
AX-104489435	**4**	**16,067,700**	**4.57**	**5.324**
AX-104102646	**4**	**16,072,377**	**5.07**	**6.402**
AX-103590406	**4**	**16,077,767**	**5.82**	**8.24**
AX-104386645	**4**	**16,078,057**	**5.3**	**6.956**
AX-104532722	**4**	**16,084,453**	**4.97**	**6.179**
AX-104477686	**4**	**16,090,862**	**5.17**	**6.643**
AX-104191303	**4**	**16,091,738**	**4.73**	**5.66**
AX-103103005	**4**	**16,323,434**	**4.44**	**5.057**
AX-104391775	4	16,537,081	4.48	5.135
AX-104086274	4	18,252,703	4.38	4.932
AX-104000509	4	18,275,058	4.5	5.183
AX-104023205	4	18,277,966	4.66	5.52
AX-103302128	4	18,283,486	4.47	5.115
AX-104296343	**4**	**19,913,037**	**4.72**	**5.629**
AX-104197593	**4**	**19,917,780**	**4.94**	**6.111**
AX-104854659	**4**	**19,918,093**	**5.36**	**7.089**
AX-104389264	**4**	**19,921,582**	**5.09**	**6.46**
AX-104565694	**4**	**19,923,593**	**5.19**	**6.694**
AX-104010517	4	20,472,902	4.46	5.102
AX-104620084	4	20,485,064	4.46	5.103
AX-103623304	4	21,801,777	4.4	4.971
AX-103451735	4	21,944,846	4.79	5.783
AX-103690970	4	21,944,978	4.46	5.101
AX-104377051	4	22,032,688	4.47	5.111

AX-104193616	4	22,034,171	4.8	5.806
AX-104486616	4	22,774,104	4.38	4.935
AX-104259910	4	27,286,775	4.51	5.198
AX-104775554	4	50,004,622	4.79	5.796
AX-104572429	4	50,005,726	4.39	4.948
AX-103550556	**6**	**42,096,241**	**4.46**	**5.105**
AX-104457897	**6**	**42,097,131**	**4.37**	**4.921**
AX-103740850	**6**	**42,099,387**	**4.39**	**4.963**
AX-104732197	**6**	**42,300,651**	**4.67**	**5.537**
AX-103902824	**6**	**42,301,314**	**4.81**	**5.84**
AX-103062645	**6**	**42,301,673**	**4.68**	**5.555**
AX-104059326	**6**	**42,329,856**	**4.41**	**4.999**
AX-104537836	6	63,418,556	4.65	5.497
AX-103889758	6	67,577,583	4.71	5.614
AX-103669524	6	67,578,609	4.94	6.125
AX-103830705	6	67,579,009	4.6	5.38
AX-104724675	17	23,660,636	4.53	5.242
AX-102965486	17	23,661,239	4.67	5.524
AX-103624241	26	22,701,987	4.51	5.197
AX-1049762851	26	22,702,755	4.52	5.219

Appendix IV

Supplementary Table 4 Details on the six overrepresented biological processes and the genes involved in each biological term from the overlapping regions from the XP-EHH and Fst tests.

PANTHER GO-Slim Biological Process	Equus caballus - REFLIST (20497)	Genes	P-value
cellular response to stimulus (GO:0051716)	1834	*FIGNL1, ARHGDIB, IGFBP1, GUCY2C, ADCY1, RERG, IGFBP3, ERP27*	0.001
ribonucleotide biosynthetic process (GO:0009260)	41	*GUCY2C,ADCY1*	0.001
signal transduction (GO:0007165)	1563	*ARHGDIB, IGFBP1, GUCY2C, ADCY1, RERG,IGFBP3*	0.008
regulation of cell communication (GO:0010646)	360	*IGFBP1,ADCY1, IGFBP3*	0.008
peptidyl-arginine modification (GO:0018195)	9	*ART4*	0.012
mRNA cis splicing, via spliceosome (GO:0045292)	16	*WBP11*	0.020

Appendix V

Supplementary Table 5 List of genes located within Fst windows around the nine top SNPs with Fst>0.20. Genes overrepresented in a PANTHER analysis are shown in bold.

ECA	Gene start (bp)	Gene end (bp)	Gene stable ID	Gene name	Gene type
1	34416060	34451419	ENSECAG00000036814		protein_coding
1	34525905	34581643	ENSECAG00000028267		protein_coding
1	34598182	34625689	ENSECAG00000037369		protein_coding
1	34698524	34748015	ENSECAG00000039250	CYP2C113	protein_coding
1	34788981	34820402	ENSECAG00000030406		protein_coding
1	34843443	34882160	ENSECAG00000018810	HELLS	protein_coding
1	34859227	34859557	ENSECAG00000026309	RF00100	misc_RNA
1	34891019	34970348	ENSECAG00000021438	**TBC1D12**	protein_coding
1	42837204	43179795	ENSECAG00000014017	PRKG1	protein_coding
1	43205006	43208327	ENSECAG00000017837	DKK1	protein_coding
5	5564149	5603417	ENSECAG00000017771	**ATP1B1**	protein_coding
5	5586963	5813575	ENSECAG00000011690	**NME7**	protein_coding
5	5813930	5836672	ENSECAG00000017383	**BLZF1**	protein_coding
5	5839550	5883472	ENSECAG00000008361	CCDC181	protein_coding
5	5915751	5934495	ENSECAG00000015797	**SLC19A2**	protein_coding
5	5964176	6039725	ENSECAG00000006860	F5	protein_coding
5	88362886	88448537	ENSECAG00000016794	ANKRD13C	protein_coding
5	88456290	88485587	ENSECAG00000017756	SRSF11	protein_coding
5	88498862	88547324	ENSECAG00000019554	LRRC40	protein_coding
5	88569124	89057869	ENSECAG00000006937	**LRRC7**	protein_coding
22	14079530	14083395	ENSECAG00000038331		protein_coding
22	14297747	14475340	ENSECAG00000020220	PAK5	protein_coding
22	14484015	14484254	ENSECAG00000037777		protein_coding
22	14484658	14500221	ENSECAG00000009366	**LAMP5**	protein_coding
22	14505665	14508154	ENSECAG00000033273		lncRNA
22	14530500	14776538	ENSECAG00000012335	**PLCB4**	protein_coding
22	15087058	15753614	ENSECAG00000021224	**PLCB1**	protein_coding
22	15864191	15906960	ENSECAG00000007522	TMX4	protein_coding
22	15935386	15985720	ENSECAG00000010314	HAO1	protein_coding

Appendix VI

Supplementary Table 6 List of genes and genomic elements located with all significant XP-EHH regions in SJ horses. Genes overrepresented in the PANTHER analysis are shown in bold.

ECA	Gene start (bp)	Gene end (bp)	Gene stable ID	Gene name	Gene type
7	89626174	90023454	ENSECAG00000017807	**NAV2**	protein_coding
7	90049571	90053641	ENSECAG00000019994	**DBX1**	protein_coding
7	90062343	90062464	ENSECAG00000025696	eca-mir-1302c-5	miRNA
7	90174293	90178174	ENSECAG00000030698		protein_coding
7	90229868	90245531	ENSECAG00000020077	HTATIP2	protein_coding
7	90248266	90383616	ENSECAG00000021124	PRMT3	protein_coding
7	90298515	90301838	ENSECAG00000029128		protein_coding
7	90433642	90488500	ENSECAG00000022340	SLC6A5	protein_coding
7	90502233	91321050	ENSECAG00000024835	NELL1	protein_coding
8	10862639	10919434	ENSECAG00000010198	PITPNB	protein_coding
8	10963357	11007491	ENSECAG00000019170	MN1	protein_coding
8	11299224	11332027	ENSECAG00000029258		lncRNA
8	11340793	11346622	ENSECAG00000031780		lncRNA
8	11341304	11346622	ENSECAG00000036895		lncRNA
8	11349511	11350844	ENSECAG00000033180		lncRNA
12	11495383	11674749	ENSECAG00000024739	C12H11orf49	protein_coding
12	11608467	11608570	ENSECAG00000025453	RF00026	snRNA
12	11673931	11684150	ENSECAG00000024744	ARFGAP2	protein_coding
12	11685623	11690380	ENSECAG00000025103	PACSIN3	protein_coding
12	11705160	11728840	ENSECAG00000008711	DDB2	protein_coding
12	11729758	11738090	ENSECAG00000013128	ACP2	protein_coding
12	11743566	11749748	ENSECAG00000015726	**NR1H3**	protein_coding
12	11754435	11788181	ENSECAG00000016022	**MADD**	protein_coding
12	11786269	11807257	ENSECAG00000024318	MYBPC3	protein_coding

			ENSECAG00000024457		
12	11809660	11825650		SPI1	protein_coding
12	11847995	11848996	ENSECAG00000032379		lncRNA
			ENSECAG00000024590		
12	11849997	11858269		SLC39A13	protein_coding
			ENSECAG00000024786		
12	11858729	11864419		PSMC3	protein_coding
			ENSECAG00000003289		
12	11875857	11886619		RAPSN	protein_coding
			ENSECAG00000007741		
12	11898228	11949746		CELF1	protein_coding
			ENSECAG00000012816		
12	11975180	11981432		PTPMT1	protein_coding
			ENSECAG00000013258		
12	11981975	11985604		KBTBD4	protein_coding
			ENSECAG00000013454		
12	11984628	11990759		NDUFS3	protein_coding
			ENSECAG00000014163		
12	11992120	11993762		FAM180B	protein_coding
12	11995006	11995980	ENSECAG00000014184		pseudogene
			ENSECAG00000014198		
12	12016876	12032587		MTCH2	protein_coding
			ENSECAG00000018720		
12	12064825	12108787		AGBL2	protein_coding
12	28801394	28812904	ENSECAG00000035459		protein_coding
			ENSECAG00000018075		
12	28822350	28828292		ARL2	protein_coding
			ENSECAG00000018173		
12	28831842	28840339		SNX15	protein_coding
			ENSECAG00000016053		
12	28841141	28843665		SAC3D1	protein_coding
			ENSECAG00000018798		
12	28844216	28854600		NAALADL1	protein_coding
			ENSECAG00000019578		
12	28863461	28879240		CDCA5	protein_coding
			ENSECAG00000019712		
12	28879548	28883204		ZFPL1	protein_coding
			ENSECAG00000020462		
12	28887608	28896985		VPS51	protein_coding
			ENSECAG00000021419		
12	28897104	28901396		TM7SF2	protein_coding
			ENSECAG00000021748		
12	28901818	28903029		ZNHIT2	protein_coding
			ENSECAG00000021956		
12	28905425	28906895		FAU	protein_coding
			ENSECAG00000022196		
12	28907055	28909823		MRPL49	protein_coding
			ENSECAG00000022486		
12	28910958	28917148		SYVN1	protein_coding
12	28961120	28962272	ENSECAG00000031458		protein_coding

			ENSECAG00000022942		
12	28968152	28991540		CAPN1	protein_coding
12	28992996	29016395	ENSECAG00000023574		protein_coding
			ENSECAG00000024194		
12	29041816	29066727		POLA2	protein_coding
			ENSECAG00000006867		
12	29086121	29086765		CDC42EP2	protein_coding
			ENSECAG00000001650		
12	29095703	29109188		DPF2	protein_coding
			ENSECAG00000006896		
12	29112607	29114022		TIGD3	protein_coding
			ENSECAG00000006906		
12	29122705	29127934		SLC25A45	protein_coding
			ENSECAG00000007115		
12	29131484	29152443		FRMD8	protein_coding
			ENSECAG00000031012		
12	29165908	29166029		RF01955	misc_RNA
			ENSECAG00000037092		
12	29166272	29166387		RF01956	misc_RNA
			ENSECAG00000036949		
12	29166603	29166761		RF01957	misc_RNA
			ENSECAG00000028832		
12	29300604	29300724		RF01955	misc_RNA
			ENSECAG00000032012		
12	29300961	29301073		RF01956	misc_RNA
			ENSECAG00000037032		
12	29301287	29301445		RF01957	misc_RNA
			ENSECAG00000034865		
12	29304008	29304122		RF01956	misc_RNA
13	27182897	27385137	ENSECAG00000022181		protein_coding
			ENSECAG00000021501		
13	27403862	27426741		UQCRC2	protein_coding
			ENSECAG00000020496		
13	27427791	27440523		PDZD9	protein_coding
13	27447476	27511140	ENSECAG00000020342		protein_coding
			ENSECAG00000009842		
13	27526717	27580812		VWA3A	protein_coding
13	27539875	27543222	ENSECAG00000040281		protein_coding
			ENSECAG00000008735		
13	27585017	27600546		SDR42E2	protein_coding
			ENSECAG00000022924		
13	27620138	27669038		EEF2K	protein_coding
13	27674709	27677909	ENSECAG00000035892		protein_coding
			ENSECAG00000015801		
13	27687817	27718707		POLR3E	protein_coding
			ENSECAG00000014770		
13	27730933	27753546		CDR2	protein_coding
			ENSECAG00000026484		
13	27768928	27769142		RF00012	snoRNA
13	27781469	27785226	ENSECAG00000032658		lncRNA
13	27810710	27854689	ENSECAG00000017507		protein_coding

19	41389013	41607791	ENSECAG00000024935	GSK3B	protein_coding
19	41598898	41627600	ENSECAG00000008728	NR1I2	protein_coding
19	41642775	41704190	ENSECAG00000016423	MAATS1	protein_coding
19	41712232	41714364	ENSECAG00000037526	COX17	protein_coding
19	41728306	41744201	ENSECAG00000021468	POPDC2	protein_coding
19	41761592	41786165	ENSECAG00000022707	PLA1A	protein_coding
19	41789983	41802421	ENSECAG00000023958	ADPRH	protein_coding
19	41814054	41840948	ENSECAG00000002211	CD80	protein_coding
19	41822782	41822938	ENSECAG00000025246	RF00003	snRNA
19	41842839	41863178	ENSECAG00000005889	TIMMDC1	protein_coding
19	41868556	41888733	ENSECAG00000010757	POGLUT1	protein_coding
19	41892242	41922685	ENSECAG00000020653	TMEM39A	protein_coding
19	41931945	42037957	ENSECAG00000007228	ARHGAP31	protein_coding
22	27886819	27974154	ENSECAG00000011582	DLGAP4	protein_coding
22	27990413	27994240	ENSECAG00000007545	MYL9	protein_coding
22	28029100	28039167	ENSECAG00000038953		protein_coding
22	28058222	28067217	ENSECAG00000010948	RAB5IF	protein_coding
22	28068662	28086618	ENSECAG00000014135	SLA2	protein_coding
22	28094843	28147979	ENSECAG00000016390	NDRG3	protein_coding
22	28169673	28255972	ENSECAG00000019507	SOGA1	protein_coding
22	28265229	28274043	ENSECAG00000008787	TLDC2	protein_coding
22	28277115	28327778	ENSECAG00000008959	SAMHD1	protein_coding
22	28346022	28407057	ENSECAG00000024316	RBL1	protein_coding
22	28420229	28469484	ENSECAG00000011332	MROH8	protein_coding
22	28469505	28524034	ENSECAG00000018381	RPN2	protein_coding
22	28529919	28533665	ENSECAG00000011318	GHRH	protein_coding

22	28562881	28563048	ENSECAG00000011341	MANBAL	protein_coding
22	28633300	28649766	ENSECAG00000012072	SRC	protein_coding
22	28653017	28654180	ENSECAG00000032280		lncRNA
22	28695652	28741584	ENSECAG00000035784		lncRNA
22	28749044	28749307	ENSECAG00000030259	BLCAP	protein_coding
22	28751140	28752958	ENSECAG00000036225		protein_coding
22	28892970	29050687	ENSECAG00000017144	CTNNBL1	protein_coding
22	29076735	29113525	ENSECAG00000037272	VSTM2L	protein_coding
22	29139685	29183745	ENSECAG00000013760	TTI1	protein_coding
22	29183841	29234284	ENSECAG00000018017	RPRD1B	protein_coding
22	29270953	29302727	ENSECAG00000006290	TGM2	protein_coding
22	29342302	29382418	ENSECAG00000017454	KIAA1755	protein_coding
22	29406550	29439276	ENSECAG00000022085	BPI	protein_coding
22	29446602	29470936	ENSECAG00000014553	LBP	protein_coding
22	29554174	29554308	ENSECAG00000026243	RF00056	snoRNA
22	29555965	29556097	ENSECAG00000025565	RF00056	snoRNA
22	29556720	29558949	ENSECAG00000031862		lncRNA
22	29556918	29557032	ENSECAG00000027074	RF00056	snoRNA
22	29557241	29557372	ENSECAG00000028924	RF00056	snoRNA
22	29570357	29575993	ENSECAG00000038773		lncRNA
22	29571351	29571482	ENSECAG00000038811	RF00056	snoRNA
22	29573213	29573345	ENSECAG00000025684	RF00056	snoRNA
22	29574565	29574698	ENSECAG00000026160	RF00056	snoRNA
22	29604080	29714277	ENSECAG00000004465	RALGAPB	protein_coding
22	29638394	29641620	ENSECAG00000031039		protein_coding
22	29663719	29663822	ENSECAG00000027501	RF01210	snoRNA
22	29683620	29687553	ENSECAG00000019945	ADIG	protein_coding
22	29713363	29733545	ENSECAG00000020026	ARHGAP40	protein_coding

22	29795146	29799965	ENSECAG00000022995	SLC32A1	protein_coding
22	29811014	29833704	ENSECAG00000024781	ACTR5	protein_coding
22	29859462	29953734	ENSECAG00000008784	PPP1R16B	protein_coding
22	29960186	29980315	ENSECAG00000013238	FAM83D	protein_coding
22	29987494	30055371	ENSECAG00000013881	DHX35	protein_coding
22	30093906	30095844	ENSECAG00000035615		lncRNA
22	30116487	30118630	ENSECAG00000031493		lncRNA
22	30144262	30146487	ENSECAG00000038194		lncRNA
22	30194028	30232341	ENSECAG00000030288		lncRNA
22	30327270	30327683	ENSECAG00000040636		protein_coding
22	30559120	30575763	ENSECAG00000037456		lncRNA
22	30748291	30748353	ENSECAG00000036790	eca-mir-1255b	miRNA
22	30853621	30871059	ENSECAG00000039049		lncRNA
22	30923375	30932893	ENSECAG00000030067		lncRNA
22	31049823	31087624	ENSECAG00000036579		lncRNA
22	31216808	31218301	ENSECAG00000034048		lncRNA
22	31433933	31434904	ENSECAG00000024705	MAFB	protein_coding
22	31596334	31602136	ENSECAG00000031009		lncRNA
22	31602198	31705814	ENSECAG00000030230		lncRNA
22	31721865	31805610	ENSECAG00000000383	TOP1	protein_coding
31	13821280	13821603	ENSECAG00000035662		protein_coding
31	14019394	14137652	ENSECAG00000038753		lncRNA
31	14289361	14315806	ENSECAG00000022231	RGS17	protein_coding

Appendix VII

Supplementary Table 7 List of genes and genomic elements located with all significant XP-EHH regions in NS horses. Genes overrepresented in the PANTHER analysis are shown in bold.

ECA	Gene start (bp)	Gene end (bp)	Gene stable ID	Gene name	Gene type
1	40592175	40611043	ENSECAG00000040253		lncRNA
1	40645693	40672650	ENSECAG00000033700		lncRNA
1	40721632	40767981	ENSECAG00000033405		lncRNA
1	40857487	40941794	ENSECAG00000017012	**PTEN**	protein_coding
1	40984941	41032681	ENSECAG00000000473	ATAD1	protein_coding
1	41041725	41182864	ENSECAG00000008426	**PAPSS2**	protein_coding
1	41049587	41049712	ENSECAG00000025194	RF00619	snRNA
1	41182850	41267017	ENSECAG00000012951	MINPP1	protein_coding
1	41536973	41656352	ENSECAG00000000125	**SGMS1**	protein_coding
1	41673609	41775939	ENSECAG00000006259		protein_coding
1	41815374	41863836	ENSECAG00000010424	A1CF	protein_coding
1	41969486	41988124	ENSECAG00000034071		lncRNA
1	42072015	42142487	ENSECAG00000034975		protein_coding
1	42647313	42649172	ENSECAG00000004701	CSTF2T	protein_coding
1	42837204	43179795	ENSECAG00000014017	PRKG1	protein_coding
1	43205006	43208327	ENSECAG00000017837	DKK1	protein_coding
1	43294744	43423003	ENSECAG00000036776		lncRNA
23	51238106	51393666	ENSECAG00000008989	AUH	protein_coding
23	51494049	51496836	ENSECAG00000028478		lncRNA
23	51505144	51543324	ENSECAG00000029666		lncRNA
23	51545997	51548741	ENSECAG00000028528		protein_coding
23	51617289	51682884	ENSECAG00000009488	SYK	protein_coding
23	51801949	51803551	ENSECAG00000037086		lncRNA
23	51842783	51843382	ENSECAG00000002969	DIRAS2	protein_coding
23	51844418	51846991	ENSECAG00000029144		protein_coding

25	5209140	5537995	ENSECAG00000009066	GABBR2	protein_coding
25	5562009	5609393	ENSECAG00000016143	ANKS6	protein_coding
25	5631576	5691074	ENSECAG00000016546	GALNT12	protein_coding
25	5739662	5841756	ENSECAG00000021211	COL15A1	protein_coding
25	5898574	5923520	ENSECAG00000004757	TGFBR1	protein_coding
25	5977703	5981706	ENSECAG00000011920	ALG2	protein_coding
25	5982047	5989768	ENSECAG00000026930	SEC61B	protein_coding
25	37398473	37402845	ENSECAG00000037647		protein_coding
25	37400945	37423271	ENSECAG00000000436		protein_coding
25	37472259	37474494	ENSECAG00000036361		lncRNA
25	37502247	37510693	ENSECAG00000032115		lncRNA
25	37519941	37558211	ENSECAG00000016957	OLFM1	protein_coding
25	37673074	37674802	ENSECAG00000034348		lncRNA
25	37691307	37691451	ENSECAG00000033548	eca-mir-9042	miRNA
25	37796640	37810483	ENSECAG00000033001		lncRNA
25	37809693	37810697	ENSECAG00000038736		lncRNA
25	37813294	37816926	ENSECAG00000038957	PPP1R26	protein_coding
25	37823576	37827267	ENSECAG00000029848	C9orf116	protein_coding
25	37827523	37831501	ENSECAG00000032143	MRPS2	protein_coding
25	37862367	37865282	ENSECAG00000032975		protein_coding
25	37872685	37875886	ENSECAG00000029369		protein_coding