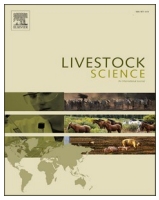
Livestock Science 284 (2024) 105480

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci



Genomic variability of the MHC region: Empirical evidence from five horse breeds

Nina Moravčíková [a], Radovan Kasarda [a,*], Hana Vostra Vydrova [b], Lubos Vostry [b], Barbora Karásková [b], Juraj Candrák [a], Marko Halo [a]

[a] Faculty of Agrobiology and Food Resources, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 94976 Nitra, Slovakia
[b] Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Kamýcká 129, 165 00 Prague, Czech Republic

HIGHLIGHTS

- Forces shaping the genetic structure of populations can affect heterozygosity of usually highly polymorphic genomic regions such as major histocompatibility complex.
- Despite differences in linkage disequilibrium between tested horse breeds, it appears that they share common genetic patterns regarding MHC class I and class II genes.
- HRR islands and balancing selection footprints inside the MHC region indicated that they may be concentrated around class II genes.

ARTICLE INFO

Keywords:
Balancing selection
Genetic diversity
Heterozygosity
Horses
Major histocompatibility complex

ABSTRACT

The purpose of this study was to analyse the level of variability in the autosomal genome, especially in the equine major histocompatibility complex region, in five horse breeds and identify heterozygosity-rich regions and potential footprints of balancing selection. Depending on data quality control, the dataset consisted of 51,168 or 53,874 single nucleotide polymorphism markers, available for 514 individuals (89 Lipizzan, 238 Old Kladruber, 47 Shagya Arabian, 61 Czech Warmblood and 81 Slovak Warmblood horses). Genomic variability within and between breeds was examined based on levels of heterozygosity (observed and expected), genomic inbreeding, Wright's F_{IS} index and linkage disequilibrium. Subsequently, the screening of heterozygosity-rich regions and balancing selection signals derived from Tajima's D positive values was performed. As expected, due to the polymorphic nature of the major histocompatibility complex, the genomic variability level was generally higher when analysing only markers located in this area (mainly around genes belonging to class I). The Slovak and Czech Warmblood horses, as breeds with open herdbook, showed higher average values of heterozygosity indices than Lipizzan, Old Kladruber or Shagya Arabian breeds. Concerning only markers in the major histocompatibility complex region in complete or very high linkage disequilibrium, common patterns were found close to *EQMHCB2, MHCB3* and *EQMHCC1* genes belonging to class I and *DQA1, DRB2, DRB3* and *HLA-DOB* genes from class II. Genome-wide, the number of heterozygosity-rich regions per animal ranged from 345.25 (Old Kladruber) to 603.33 (Czech Warmblood). Across all breeds, 254 heterozygosity-rich regions were detected directly in the major histocompatibility complex region (194 in class I and 60 in class II). Among them, the highest overlap showed regions found in the genomes of historically connected Czech and Slovak Warmblood breeds. The results suggested that the frequency of markers in heterozygosity-rich regions increased in Lipizzan, Old Kladruber and Shagya Arabian breeds in the genomic region of *EQMCE1* gene (class I) and in Czech and Slovak Warmblood horses in *DQB1, DQA2, DQB2, DQA3* and *DRB2* genes (class II). Although the identified heterozygosity-rich regions formed 330 islands across the genomes of tested breeds, these islands were outside the major histocompatibility complex region. On the other hand, four of 425 balancing selection signals detected across breeds were located directly in the major histocompatibility complex region, close to *DRA, DRB1, DQA1, DQB1* and

* Corresponding author at: Institute of Nutrition and Genomics, Faculty of Agrobiology and Food Resources, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2. 94976 Nitra, Slovakia.
E-mail address: radovan.kasarda@uniag.sk (R. Kasarda).

https://doi.org/10.1016/j.livsci.2024.105480
Received 26 April 2023; Received in revised form 2 May 2024; Accepted 6 May 2024
Available online 8 May 2024

DQB2 genes (class II). Shared islands and balancing selection footprints among breeds were found mainly on chromosomes 7 and 11.

1. Introduction

Conservation of genetic diversity is an irreplaceable part of sustainable development. Horses belong to species with a long history and well-organised breeding work. Most famous breeds (e.g. Old Kladruber, Lipizzan or Shagya Arabian) have generations-long pedigrees in their herdbook covering the most important lines and families, allowing them to trace up to founders. However, even if the pedigree information is robust, founder analysis is sometimes not strong enough to characterise the genetic variability of breeds (Cervantes et al., 2009). Bartolomé et al. (2022) pointed out that even if pedigree completeness is essential for reliable genetic evaluations to enhance genetic progress, genomic data will help reveal real contributions in the future.

Previous studies have shown that different types of genetic markers can be successfully used to analyse genomic variability and the uniqueness of the horse population gene pool (Khaudov et al., 2018). Microsatellites were historically popular due to their utility in parentage verification for over two decades (Bowling, 1997). They have been extensively employed to assess the genetic variability and population structure of breeds like Lipizzan, Old Kladruber, Arabian or Slovak Warmblood (Curik et al., 2003a; Dovc et al., 2006; Vostrý et al., 2011; Kasarda et al., 2016; Putnová et al., 2019; Machmoum et al., 2020). Alongside microsatellites, mitochondrial DNA (mtDNA) sequencing and SNP markers genotyping (low and high-density arrays) are also often used to quantify genetic variability. Maternally inherited mtDNA plays an important role in energy metabolism, and variance in mtDNA genes might contribute to differences in performance traits, as already pointed out in humans and racehorses. Moreover, genotyping of mtDNA SNPs allows for tracing the evolution of maternal lineages far before pedigree recordings (Engel et al., 2022). On the other hand, the sequence of the non-recombining part of the Y chromosome has been used in different horse breeds to discover parental lineages and phylogeny of breed development, detect admixture, and create a basis for future conservation programs (Khaudov et al., 2018; Radovic et al., 2022). Out of all the listed marker types, SNP markers in the autosomal genome are currently the primary choice for assessing genomic variation, population structure, admixture, and genotype-phenotype interaction in various livestock species, including horses (Nolte et al., 2019; Grilz-Seger et al., 2019; Schaefer and McCue, 2020; Santos et al., 2021; Vostrý et al., 2021; Szmatoła et al., 2022; Nolte et al., 2022).

The major histocompatibility complex (MHC) refers to a relatively large locus on the mammalian DNA covering a group of closely linked polymorphic genes divided into MHC classes I, II and III (Rothbard et al., 1991). MHC genes encode cell-surface proteins with an essential role in innate and adaptive immune response, i.e. genetic variation in this region influences an individual's response to challenges from various pathogens (Brinkmeyer-Langford et al., 2013). The horse MHC region, so-called equine leukocyte antigen (ELA), is localised on chromosome 20. Early studies indicated 16 to 33 MHC class I genes in the equine genome; however, all of these genes are unlikely to be expressed (Carpenter et al., 2001). Tallmadge et al. (2005, 2010) identified 15 horse MHC class I genes spread in three regions and expressed in different haplotypes. Horecky et al. (2018) developed an additional microsatellite panel of the MHC class I and class III genomic sub-regions in horses. The first long-read sequence assembly of the horse MHC class II region with gene annotation has been provided by Viļuma et al. (2017). Previous studies demonstrated the role of horse MHC genes in the control of immune-mediated diseases like insect bite hypersensitivity (Klumperova et al., 2013), myositis (Duward-Akhurst et al., 2016) or equine sarcoids (Staiger et al., 2016). However, MHC genes could also be responsible for other important phenotypic traits in horses, including MHC-linked female preferences (Burger et al., 2017) and embryo survival during uterine nesting (Jeannerat et al., 2020).

The MHC usually exhibits a high level of variability, making it one of the most polymorphic genetic regions in the mammalian genome (Kulski et al., 2002). This variability is characterised by the presence of numerous alleles across MHC loci (alleles identified in Equids species are summarised in the ELA Database available at https://www.ebi.ac.uk/ipd/mhc/group/ELA/), contributing to the extensive diversity across and within populations. MHC polymorphism is assumed to be shaped by a combination of factors, including genetic drift, mutation, recombination, and natural selection, acting by two mechanisms: overdominant selection (also called heterozygote advantage) and balancing selection. In the case of overdominant selection, the heterozygote advantage may arise from a certain degree of dominance of resistance, which allows heterozygotes to respond to a broader range of pathogens or pathogen strains compared to homozygotes, or from overdominance, where heterozygotes inherently have higher fitness than homozygotes. Balancing selection, on the other hand, encompasses a variety of mechanisms that maintain genetic diversity within populations by actively selecting against the fixation of any single allele. This can occur through various processes such as heterozygote advantage (as in overdominant selection), frequency-dependent selection, and spatial or temporal habitat heterogeneity (Charlesworth, 2006; Radwan et al., 2020).

This study aimed to analyse and compare the level of variability in the autosomal genome, especially in the MHC region, in five horse breeds and test the distribution of heterozygosity-rich regions and potential selection footprints arising in the genome of each breed as a consequence of balancing selection.

2. Material and methods

2.1. Genomic datasets

Genome-wide SNP data of five breeds were tested: Lipizzan (Czech population $N = 24$, Slovak population $N = 65$), Old Kladruber ($N = 238$) and Shagya Arabian ($N = 47$) as representatives of breeds with closed herdbook and Czech Warmblood ($N = 61$) and Slovak Warmblood ($N = 81$) as breeds with open breeding programs. Biological samples (hair roots) were collected in collaboration with Slovak and Czech national stud farms and private owners to cover the gene pool of each breed reliably. After DNA extraction, samples were genotyped by two arrays, GGP Equine 70k and Illumina Equine 80k, in the commercial lab with an average call rate of 98.41 %.

Two different data quality control processes were performed in PLINK v1.9 (Chang et al., 2015), depending on the type of analysis. The first dataset containing only animals ($N = 514$) and autosomal SNP markers with a call rate higher than 90 % and minor allele frequency higher than 1 % ($N = 51{,}168$) was prepared to quantify the level of genomic variability. In the second dataset intended to analyse heterozygosity-rich region distribution and potential footprints of balancing selection in the genome of tested breeds, only the call rate for animals ($N = 514$) and autosomal SNPs was considered (cut-off value 90 %) ($N = 53{,}874$). In the next step, only SNPs in the MHC region (102 or 108, depending on the quality control) were filtered from both datasets. The start and end position of the horse MHC region was characterised by MHC class I and class II genes (Table S1), according to information stored in the Ensembl (Cunningham et al., 2022) and NCBI (reference genome assembly EquCab3.0) databases.

2.2. Genomic variability

Genomic variability on the genome-wide level and in MHC region within and between breeds was derived from the calculation of commonly accepted genetic diversity parameters (observed and expected heterozygosity, minor allele frequency, genomic inbreeding, Wright's fixation index) and variability in linkage disequilibrium (LD).

Observed (H_o) and expected (H_e) heterozygosity, minor allele frequency (MAF) and pairwise LD were calculated using PLINK v1.9 and Wright's fixation index F_{IS} using R package snpR (Hemstrom and Jones, 2023). Obtained values were averaged across SNPs on the genome-wide level and in the MHC region (class I and class II) separately for each breed. H_o, H_e and F_{IS} were then also estimated using short haplotypes to reduce the ascertainment bias of the Equine 70k and 80k chips (Simčič et al., 2015). Short haplotypes ($N = 5912$) were created by splitting the SNP data into non-overlapping blocks of 4 SNPs with an inter-marker distance of less than 50 kb for neighbouring SNPs and the maximum length of each block being less than 150 kb. Genomic inbreeding (F_{HOM}) was estimated by the method-of-moments from the observed and expected homozygous counts for each animal by PLINK v1.9 and averaged per breed.

LD differences in MHC region between breeds were analysed by calculating standardised varLD score separately for each breed combination according to the approach described by Teo et al. (2009) using varLD (Ong and Teo, 2010). The LD values, expressed by r^2, were calculated over sliding windows containing 50 SNPs for all SNP pairs.

2.3. Detection of heterozygosity-rich regions and footprints of balancing selection

The distribution of heterozygosity-rich regions (HRRs) across the genome and in the MHC region was analysed separately for each breed by R package DetectRUNS (Biscarini et al., 2018) using a consecutive SNP-based approach. As proposed by several studies on various livestock species (Marras et al., 2018; Biscarini et al., 2020; Santos et al., 2021; Selli et al., 2021) following criteria for HRRs detection were applied: the minimum number of SNPs in HRR was 10, the minimum length of HRR was 0.25 Mbp, the maximum gap between SNPs in HRR was 1 Mbp, the maximum number of homozygous SNPs in HRR was three and the maximum number of missing SNPs was set to 2. Subsequently, HRR islands were identified based on the frequency at which SNP markers occurred in the HRR in the population gene pool, i.e. the proportion of animals that shared particular SNP in HRR. The cut-off value characterising HRR islands was set to the top 0.99 of the percentile of SNP frequency in HRR. Each HRR island consisted of a minimum of 4 SNPs above the cut-off value.

Footprints of balancing selection across the genome of breeds, particularly in the MHC region, were tested by Tajima's D method. Tajima's D is a statistical measure assessing population diversity and an indicator of balancing selection. It compares the observed nucleotide variation within a population to what is expected under neutral evolution. A positive Tajima's D value suggests balancing selection, where alleles are maintained at intermediate frequencies, resulting in more nucleotide differences between pairs of sequences than expected. In contrast, negative Tajima's D values are associated with positive selection, reducing genetic diversity and nucleotide differences (Tajima, 1989; Simonsen et al., 1995; Danecek et al., 2011). Tajima's D values were calculated in 0.25 Mbp sliding windows using VCFtools (Danecek et al., 2011). Potential footprints of balancing selection were defined by the top 0.99 of the percentile of Tajima's D values ($D > 0$). R package ggplot2 (Wickham, 2016) was used for the graphical visualisation.

2.4. Annotation of QTLs and genes connected to HRR islands and footprints of balancing selection

The Animal QTL Database (Hu et al., 2019) was used to identify quantitative trait loci (QTLs) associated with different phenotypic traits in horses. Only QTLs located directly in genomics regions of HRR islands and footprints of balancing selection were considered.

Protein-coding genes in the area of interest were retrieved from the Ensembl database using the web-based tool Biomart, according to the latest reference genome assembly EquCab3.0. Initially, the biological function of identified genes was examined by a comprehensive literature survey. Subsequently, gene enrichment analyses were performed using the Database for Annotation, Visualisation, and Integrated Discovery (DAVID) v2021 (Huang et al., 2009) to find significantly enriched Gene Ontology (GO) for biological processes and annotation clusters of enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Fisher's exact test was applied to set the threshold value (p-value equal to 0.05).

3. Results

3.1. Genomic variability

The level of genomic variability within analysed breeds was quantified by calculating average values of observed and expected heterozygosity, minor allele frequency, genomic inbreeding derived from the observed and expected count of homozygous genotypes, Wright's fixation index and linkage disequilibrium separately across the autosomal genome and markers localised directly in MHC region (class I and class II). The observed results are presented in Table 1. As expected, due to the polymorphic nature of the MHC, the genomic variability level was generally higher when analysing only SNPs in the MHC region (Figure S1). Representatives of breeds with open herdbooks, the Czech and Slovak Warmblood horses, showed higher heterozygosity and MAF than others regardless of the analysed region (autosomal genome or MHC region). This was also confirmed by an alternative approach for estimating H_o, H_e and F_{IS} by constructing short blocks of SNPs (haplotypes) that serve as multiallelic markers, allowing a better analysis of intrapopulation genetic variability than biallelic SNPs. A comparison of the obtained results for MHC class I and class II regions showed that the degree of intrapopulation variability differentiated between breeds and applied approaches for heterozygosity estimation. Haplotypes-based estimates indicated higher variability in the MHC class I region in all breeds except Old Kladruber. In contrast, the SNP-based method suggested higher variability in MHC class I only in Lipizzan and Old Kladruber.

Screening of F_{HOM} on a genome-wide level showed that, except for Old Kladruber, heterozygous genotypes were slightly predominant in the genome of all tested breeds. However, at the population level, this was confirmed only in the Lipizzan and Shagya Arabian breeds, where the F_{IS} index became negative. Nevertheless, it is not possible to clearly state the predominance of homozygous genotypes over heterozygous ones on the genome-wide level or in the MHC region, as the average F_{IS} and F_{HOM} values for all breeds were close to zero, indicating relatively balanced genotype frequencies in their genomes.

Lipizzan, Old Kladruber, and Shagya Arabian breeds showed higher genome-wide LD levels than Czech and Slovak Warmblood horses. This could signalise the more substantial effect of selection pressure on the genome structure in the case of breeds with closed herdbooks compared to those with open breeding programs. A closer look at the MHC region showed a similar trend, especially for the class I region. In the case of MHC class II, a higher average LD level was found in the Czech and Slovak Warmblood breeds. Fig. 1 shows the pairwise LD values only for SNPs in the MHC region separately for each breed. Comparison between breeds revealed that the highest proportion of SNP pairs in complete LD in the MHC region showed Old Kladruber horse (17.27 %), followed by Shagya Arabian (10.58 %), Lipizzan (10.37 %), Slovak Warmblood (4.72 %) and Czech Warmblood horse (3.35). Concerning only SNP pairs in complete or very high LD ($r^2 > 0.95$), common patterns across breeds were found mainly in connection to *EQMHCB2, MHCB3* and *EQMHCC1*

Table 1
Average values (standard errors in brackets) of tested diversity parameters across the autosomal genome and in the MHC region.

Breed	H_o	$H_{oBlocks}$	H_e	$H_{eBlocks}$	MAF	F_{HOM}	F_{IS}	$F_{ISBlocks}$	LD (r^2)
Genome-wide level									
Lipizzan	0.280	0.599	0.275	0.585	0.203	−0.018	−0.011	−0.017	0.520
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.004)	(0.001)	(0.001)	(0.001)
Old Kladruber	0.274	0.583	0.280	0.595	0.205	0.022 (0.003)	0.020 (0.001)	0.022 (0.001)	0.515
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)				(0.001)
Shagya Arabian	0.298	0.610	0.286	0.584	0.213	−0.043	−0.028	−0.033	0.525
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.005)	(0.001)	(0.001)	(0.001)
Czech Warmblood	0.338	0.664	0.337	0.659	0.251	−0.005	0.003 (0.001)	0.002 (0.001)	0.490
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)			(0.001)
Slovak	0.337	0.662	0.336	0.660	0.250	−0.002	0.005 (0.001)	0.003 (0.001)	0.491
Warmblood	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)			(0.001)
MHC region									
Class I.									
Lipizzan	0.306	0.723	0.308	0.705	0.235	0.004 (0.034)	0.007 (0.017)	−0.020	0.443
	(0.021)	(0.030)	(0.020)	(0.024)	(0.019)			(0.025)	(0.021)
Old Kladruber	0.286	0.608	0.286	0.595	0.212	−0.001	−0.010	−0.023	0.560
	(0.022)	(0.053)	(0.022)	(0.054)	(0.019)	(0.026)	(0.014)	(0.024)	(0.030)
Shagya Arabian	0.266	0.599	0.281	0.601	0.210	0.054 (0.062)	0.052 (0.023)	0.008 (0.033)	0.498
	(0.019)	(0.062)	(0.019)	(0.062)	(0.018)				(0.023)
Czech Warmblood	0.342	0.731	0.346	0.738	0.258	0.012 (0.043)	0.023 (0.017)	0.020 (0.021)	0.426
	(0.018)	(0.033)	(0.016)	(0.023)	(0.016)				(0.018)
Slovak	0.365	0.774	0.353	0.742	0.269	−0.033	−0.030	−0.040	0.443
Warmblood	(0.018)	(0.008)	(0.017)	(0.018)	(0.017)	(0.045)	(0.014)	(0.023)	(0.018)
Class II.									
Lipizzan	0.295	0.568	0.280	0.544	0.194	−0.052	−0.056	−0.043	0.497
	(0.026)	(0.029)	(0.025)	(0.039)	(0.023)	(0.050)	(0.010)	(0.028)	(0.034)
Old Kladruber	0.258	0.745	0.254	0.744	0.198	−0.017	−0.017	0.002 (0.020	0.475
	(0.037)	(0.029)	(0.035)	(0.014)	(0.033)	(0.031)	(0.013)		(0.047)
Shagya Arabian	0.277	0.574	0.289	0.577	0.228	0.038 (0.078)	0.031 (0.023)	0.015 (0.009)	0.459
	(0.032)	(0.096)	(0.034)	(0.097)	(0.033)				(0.036)
Czech Warmblood	0.352	0.661	0.372	0.691	0.293	0.052 (0.048)	0.053 (0.020)	0.055 (0.022)	0.503
	(0.026)	(0.072)	(0.026)	(0.062)	(0.027)				(0.033)
Slovak	0.365	0.658	0.395	0.690	0.311	0.075 (0.050)	0.068 (0.023)	0.053 (0.013)	0.511
Warmblood	(0.022)	(0.059)	(0.023)	(0.057)	(0.024)				(0.028)

H_o – observed heterozygosity ($H_{oBlocks}$ – haplotypes-based estimates), H_e – expected heterozygosity ($H_{eBlocks}$ – haplotypes-based estimates), MAF – minor allele frequency, F_{HOM} – genomic inbreeding derived from the observed and expected homozygous counts, F_{IS} – Wright fixation index ($F_{ISBlocks}$ – haplotypes-based estimates), LD – linkage disequilibrium.

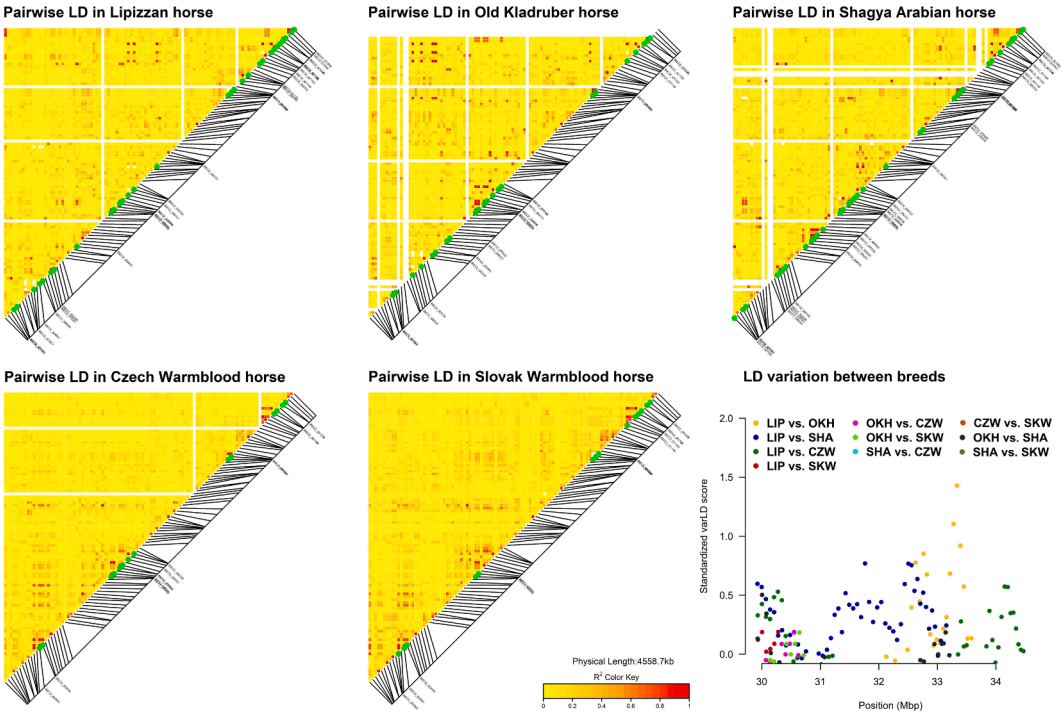


Fig. 1. Pairwise LD values for SNPs in the MHC region (SNPs in complete LD shown in green) and LD variation between breeds.

genes belonging to MHC class I and *DQA1, DRB2, DRB3* and *HLA-DOB* genes from MHC class II. The LD variability in the MHC region was then tested based on the pairwise comparison of LD levels between breeds by calculating standardised varLD scores. As shown in Fig. 1, the obtained varLD score pointed to only low LD differences between breeds, which indicates that they share common LD patterns. The highest signal was detected when comparing LD in Lipizzan and Old Kladruber at position 32.76 Mbp between the EQMHCC1 and DRA genes.

3.2. Detection of heterozygosity-rich regions and footprints of balancing selection

Genome-wide screening of HRRs distribution showed that the number of HRRs per animal ranged from 345.25 (Old Kladruber) to 603.33 (Czech Warmblood) (Table 2). A summary of the identified HRRs in the autosomal genome of tested breeds is given in Table S2. Based on selected criteria for HRR detection, only segments up to 2.5 Mbp were identified with average length from 0.45 Mbp (Lipizzan and Old Kladruber) to 0.77 Mbp (Slovak Warmblood horse). The maximum number of SNPs in HRRs on a genome-wide level ranged from 31 (Lipizzan) to 54 (Shagya Arabian). The highest average number of HRRs per animal and average number of SNPs in HRRs in the MHC region was found in Slovak Warmblood, followed by the Czech Warmblood, Shagya Arabian, Lipizzan and Old Kladruber horse (Table 2). Across all breeds, a total of 254 HRRs were detected in the MHC region (194 in class I and 60 in class II region). As can be seen in Fig. 2B, some of them totally overlapped across different breeds. The highest proportion of common HRRs in the MHC class I region was found in Czech and Slovak Warmblood breeds ($N = 8$). In addition, they shared 4 HRRs with Shagya Arabian, 3 with Lipizzan, and 1 with Old Kladruber horses. 3 HRRs were common for Lipizzan, Old Kladruber and Czech and Slovak Warmblood, 2 for Lipizzan, Old Kladruber and Shagya Arabian, and 1 HRR for Old Kladruber and Shagya Arabian breeds. In the MHC class II region, 5 HRRs overlapped in Czech and Slovak Warmblood breeds, which also shared 2 segments with Lipizzan and 2 with Old Kladruber horses.

The frequency of SNPs in HRRs (%) was also used to identify potential HRR islands in the autosomal genome, especially the MHC region. The cut-off value for their identification was set to 36.4 %, 33.8 %, 38.3 %, 36.7 % and 37.2 % for Lipizzan, Old Kladruber, Shagya Arabian, Czech Warmblood and Slovak Warmblood horses, respectively. Fig. 2 shows the genome-wide distribution of detected HRR islands in metapopulation (C) and per breed (D). As expected, the distribution of HRR islands was non-uniform.

Table 2
Summary of HRRs distribution across genome and in MHC region.

Breed	No. of HRRs per animal	HRR Length (Mbp)		SNPs in HRR	
	$\bar{x}$(SE)	$\bar{x}$(SE)	max	$\bar{x}$(SE)	max
Genome-wide level					
Lipizzan	366.216 (2.826)	0.450 (0.001)	1.747	12.031 (0.014)	31
Old Kladruber	345.248 (1.954)	0.450 (0.001)	1.956	12.009 (0.009)	38
Shagya Arabian	455.127 (4.072)	0.462 (0.001)	2.410	12.302 (0.009)	54
Czech Warmblood	603.333 (5.249)	0.475 (0.001)	2.021	12.567 (0.019)	41
Slovak Warmblood	596.704 (5.440)	0.477 (0.001)	2.404	12.555 (0.013)	38
MHC region					
Lipizzan	1.372 (0.082)	0.507 (0.027)	1.043	11.797 (0.297)	19
Old Kladruber	1.208 (0.065)	0.510 (0.026)	1.332	11.460 (0.202)	19
Shagya Arabian	1.731 (0.162)	0.537 (0.039)	1.168	11.556 (0.274)	19
Czech Warmblood	1.744 (0.149)	0.506 (0.027)	1.256	12.067 (0.273)	21
Slovak Warmblood	2.078 (0.166)	0.525 (0.018)	1.124	12.632 (0.204)	24

HRRs – heterozygosity-rich regions, $\bar{x}$ – average value, SE – standard error, max – maximum value.

As can be seen in Fig. 2A, SNPs located in the MHC region have not reached the required threshold values. In Lipizzan, Old Kladruber and Shagya Arabian horses, the potential HRR island (average SNP frequency in HRRs: Lipizzan 19.32 %, Old Kladruber 11.52 % and Shagya Arabian 23.40 %) located in position from 31.23 to 31.53 Mbp overlapped with *EQMCE1* gene (MHC class I). In the Czech and Slovak Warmblood MHC region, the highest frequency of SNPs in HRRs was found in positions from 31.15 to 31.23 Mbp (in average 21.67 % and 29.22 %) between *EQMHCB2* and *EQMCE1* genes (MHC class I) and from 33.79 to 34.14 Mbp (in average 20.91 % and 28.08 %) overlapping with *DQB1, DQA2, DQB2, DQA3* and *DRB2* genes (MHC class II).

Genome-wide, the number of detected HRR islands above the significance threshold ranged from 65 (Lipizzan and Shagya Arabian) to 74 (Czech Warmblood). Detailed information about identified HRR islands is given in Table S3. The minimum size of the HRR island was 0.01 Mbp (chromosome 2), and the maximum was 4.69 Mbp (chromosome 17). 3 HRR islands located on chromosomes 7 and 11 were shared by all breeds (Table 3). Overall, 23 genes were detected in the HRR island on chromosome 7 and 5 genes in both islands on chromosome 11. Subsequent gene enrichment analysis revealed that detected genes are included in 3 functional Gene Ontology terms and 1 KEGG pathway (Table S5). Inside the first HRR island on chromosome 11, QTL associated with insect bite hypersensitivity was also found.

Potential footprints of balancing selection were identified by Tajima's D statistics when only positive D values above a cut-off value were taken into account (Lipizzan – 3.47, Old Kladruber – 3.87, Shagya Arabian – 3.25, Czech Warmblood 3.43, Slovak Warmblood – 3.59). Table S4 provides a summary of the detected signals in the autosomal genome of analysed breeds. The highest number of signals ($N = 8$) was identified on chromosome 1 in Slovak Warmblood (average D 3.78), chromosome 2 in Shagya Arabian (average D 3.44), chromosome 4 in Czech Warmblood (average D 3.68) and chromosome 8 in Old Kladruber (average D 4.21). Based on the set criteria for footprint detection, only four signals in three breeds (Old Kladruber, Slovak Warmblood and Czech Warmblood) were identified directly in the MHC regions (Fig. 3A). Two of them in Czech Warmblood and Slovak Warmblood horses were connected to *DRA, DRB1* and *DQA1* genes (MHC class II) and another two in Slovak Warmblood and Old Kladruber to *DQB1* and *DQB2* genes (MHC class II). In connection with the HRR island analysis, it seems that the increase in heterozygosity in the regions around particular MHC class II genes in Czech Warmblood and Slovak Warmblood breeds may be a consequence of balancing selection. Fig. 3 also shows the distribution of signals within chromosomes in metapopulation (B) and individual breeds (C). As shown in Table 3, similar to the HRR island analysis, shared signals by all breeds were observed on chromosomes 7 and 11 directly or close to the position of detected HRR islands. This overlap is evident mainly in connection to the CA10 gene on chromosome 11. In addition, another shared signal was detected on chromosome 13 (Table 3). In connection with the signals, 14 protein-coding genes were identified. Five of them were linked to the same KEGG pathway as in the enrichment analysis of genes inside HRR islands. On chromosome 13, QTL for recurrent airway obstruction was localised.

4. Discussion

Initial molecular-genetic studies of heterozygosity in horses utilised mainly microsatellite markers to quantify intra and interpopulation variability levels. Curik et al. (2003b) reported average individual heterozygosity of 0.67 for Lipizzan breeding mares from seven European national stud farms. A similar level of heterozygosity was also found in

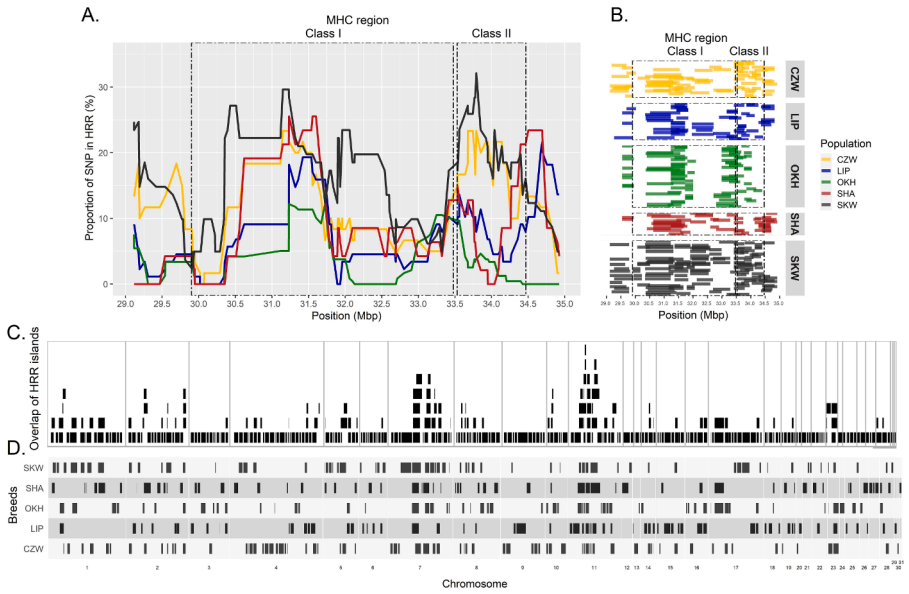


Fig. 2. Graphical visualisation of SNP frequency in HRRs (%) in MHC region (A., the position of MHC region is coloured in black), HRRs distribution in MHC region (B., the position of MHC region is colored in black) and distribution of HRR islands per chromosomes in metapopulation (C.) and breeds (D.). Gaps between consecutive HRR islands within and between chromosomes do not correspond to the real distance in bp (LIP – Lipizzan, OKH – Old Kladruber, SHA – Shagya Arabian, CZW – Czech Warmblood, SKW – Slovak Warmblood).

Lipizzan horses in further studies in the later period (Achman et al., 2004; Dovc et al., 2006; Barcaccia et al., 2013; Kasarda et al., 2016). Even though the Old Kladruber horse has a smaller effective population size and thus a limited gene flow than the Lipizzan breed, previous microsatellite-based studies showed a comparable level of heterozygosity to that in the Lipizzan horses (Vostrý et al., 2011; Janova et al., 2013; Kasarda et al., 2016; Vostrý et al., 2018). However, further studies utilising different types of genetic markers, including high-density SNP arrays, revealed that the estimated level of genetic variability strongly depends on the polymorphic nature and number of markers analysed as well as the applied methodological approach (Janova et al., 2013; Simčič et al., 2015; Cortés et al., 2019), which was also confirmed in this study. Moreover, due to different intensities of selection pressure on particular genomic regions controlling preferred phenotypic traits, such regions may display reduced levels of heterozygosity compared to others (Petersen et al., 2013; Metzger et al., 2015). For example, Janova et al. (2013) found in Old Kladruber horse that even if microsatellite-based heterozygosity estimates pointed to a relatively high amount of heterozygosity, there was a significant decrease if they used SNPs localised close to immunity-related, coat colour and MHC genes. The SNP-based estimates of observed and expected heterozygosity and minor allele frequency across autosomal genome of analysed breeds were consistent with the previous results for Warmblood breeds (Petersen et al., 2013), Arabian horses (Cosgrove et al., 2020), Thoroughbred horses (Kim et al., 2019) and Italian local horse breeds (Criscione et al., 2022). The observed increase in heterozygosity in the MHC region compared to that on a genome-wide level may reflect the fact that the MHC gene complex responsible for many important biological traits, including the immune system, mating preference and pregnancy outcome, is considered one of the most polymorphic in vertebrates (Curik et al., 2003a; Sommer, 2005; Tang et al., 2022). But, as this study suggests, different breeding strategies could affect the level of heterozygosity stored in the equine MHC region as well. The Czech Warmblood and Slovak Warmblood horses, as breeds with intensive gene flow, revealed a higher level of heterozygosity compared to Lipizzan, Old Kladruber or Shagya Arabian horses as representatives of breeds with closed herdbooks. Similar results were presented by Jaworska et al. (2020) for primitive and draft horse breeds.

Higher average genome-wide LD values in Lipizzan, Old Kladruber and Shagya Arabian horses signalise that long-term selection for traits of interest and less intense gene flow increased LD extent compared to Czech and Slovak Warmblood breeds. This argument supports previous studies that have shown that continuous selection can affect not only the phenotypic traits of horses but also their genome structure, including MHC genes (Andersson et al., 2012; Jaworska et al., 2020). It was shown that loci affecting phenotypic traits under strong selection pressure usually display an increase in LD compared to others and are transmitted to offspring in the form of haplotypes (Al-Mamun et al., 2015; Gurgul et al., 2019; Oyelami et al., 2020). However, LD extent can also be affected by other genetic forces, such as genetic drift, gene flow, admixture or population fragmentation (Slatkin, 2008). Genetic drift can interact with selection in surprising ways. Response of tightly linked loci to selection can be weakened because of a small amount of LD produced by genetic drift due to the Hill-Robertson effect (Hill and Robertson, 1966). This effect can be strong mainly if many loci under selection are tightly linked (Slatkin, 2008). Populations with low effective sizes are more susceptible to genetic drift than larger populations. Thus, genetic drift randomly generates LD at different loci at a rate inversely proportional to the effective population size (Waples et al., 2016). On the other hand, gene flow in connection with admixture and recombination could reduce LD extent (Lucek and Willi, 2021; Liang et al., 2022). The gene flow itself as a result of the gametes exchange among subpopulations increases LD in each subpopulation while the frequency of alleles differs among them. If the frequency of alleles among subpopulations is maintained by selection, then the LD will persist to the next generations (Slatkin, 2008). However, admixture events gradually decrease linkage disequilibrium in the admixed population in the subsequent generations as a result of recombination (Liang et al., 2022).

Based on the results of LD extent analysis in the MHC region, it appears that tested breeds shared common patterns that resulted in LD increase in areas connected to both MHC class I (*EQMHCB2, MHCB3, EQMHCC1*) as well as class II genes (*DQA1, DRB2, DRB3* and *HLA-DOB*). The analysis of LD extent in MHC complex was used in previous studies to evaluate the variation of serum IgE levels (Curik et al., 2003a) or identify equine MHC haplotypes (Tseng et al., 2010).

Several studies have been focused on characterising the distribution of HRR segments in the livestock genome (Williams et al., 2016; Ferenčaković et al., 2016; Biscarini et al., 2020; Selli et al., 2021; Santos

Table 3
HRR islands and signals of balancing selection overlapping in all tested breeds.

Chr	Start (Mbp)	End (Mbp)	Size (Mbp)	QTLs	Protein-coding genes
HRR islands					
7	50.010	50.651	0.640		ENSECAG00000059850, ENSECAG00000043020, ENSECAG00000047942, ENSECAG00000050713, ENSECAG00000052645, ENSECAG00000046337, ENSECAG00000029788, ENSECAG00000055839, ENSECAG00000053433, ENSECAG00000044631, ENSECAG00000048367, ENSECAG00000059680, ENSECAG00000050915, ENSECAG00000057812, ENSECAG00000058001, ENSECAG00000043034, ENSECAG00000047363, ENSECAG00000052648, ACP5, CNN1, ELOF1, ECSIT, ZNF653
11	26.971	27.251	0.281	QTL:29,289	ENSECAG00000048039, ENSECAG00000057286, ENSECAG00000053616, CA10
11	27.374	27.820	0.447		CA10, ENSECAG00000052855
Signals of balancing selection					
7	48.750	49.000	0.250		ENSECAG00000048134, ENSECAG00000049820, ENSECAG00000050249, ENSECAG00000050553, ENSECAG00000044810, ENSECAG00000045529, ENSECAG00000055949, ENSECAG00000041598, ENSECAG00000056107, ENSECAG00000041852, ENSECAG00000041001
11	27.500	27.750	0.250		CA10
11	28.250	28.500	0.250		
13	15.500	15.750	0.250	QTL:28,353	ENSECAG00000058502, AUTS2

et al., 2021; Ablondi et al., 2022; Chen et al., 2022). These studies have shown that HRR segments are much rarer and shorter than runs of homozygosity (ROH). In contrast to ROH, HRRs are formed by heterozygote loci clustered in regions with different sizes, frequencies and chromosomal locations in the genome. Comparison between the above studies showed that the number and length of identified HRRs depend on the species analysed, the density of the SNP array, and the method and settings used for HRRs screening. A common feature was that the minimum number of SNP markers in the HRR was much smaller compared to ROH, with a decrease in the minimum number of SNPs leading to a decrease in their number and length rather than an increase (Santos et al., 2021; Selli et al., 2021). Similarly, the minimum number of opposite and missing genotypes differs between the studies. Biscarini et al. (2020) and Selli et al. (2021) reported that the number of opposite genotypes allowed in analysis increases the number of identified HRR segments.

The average number of HRRs per animal across the genome of tested breeds was higher compared to those reported for the Mangalarga Marchador or Quarter Horse horse breeds (Santos et al., 2021, 2023). However, the average length of HRRs was comparable with previous studies. We didn't find HRR longer than 2.5 Mbp (average maximum length across breeds 2.11 Mbp). The HRRs number per animal in the MHC region followed the distribution of HRRs on the genome-wide level. The Slovak Warmblood and Czech Warmblood horses showed a higher average number of HRRs per animal compared to other analysed breeds. Thus, the observed distribution of HRRs across the autosomal genome as well as in the MHC region suggested that the genomic variability of horse breeds with closed herdbooks may be reduced compared to those with open breeding programs probably due to limited gene flow between stud farms, a lower proportion of admixture as well as accumulation of inbreeding and homozygous regions coming from common ancestors across generations. On the other hand, a high proportion of common HRRs in Czech and Slovak Warmbloods could result from their historical relatedness. This is in line with the study of Faria et al. (2019), which suggests that many inversions are maintained polymorphic within populations by balancing selection, which impedes divergence and speciation. Contrary to expectations, we detected HRR islands outside the MHC region. However, the results suggested that the frequency of SNPs in HRRs increased in Lipizzan, Shagya Arabian and Old Kladruber horses mainly in the genomic region of MHC class I gene

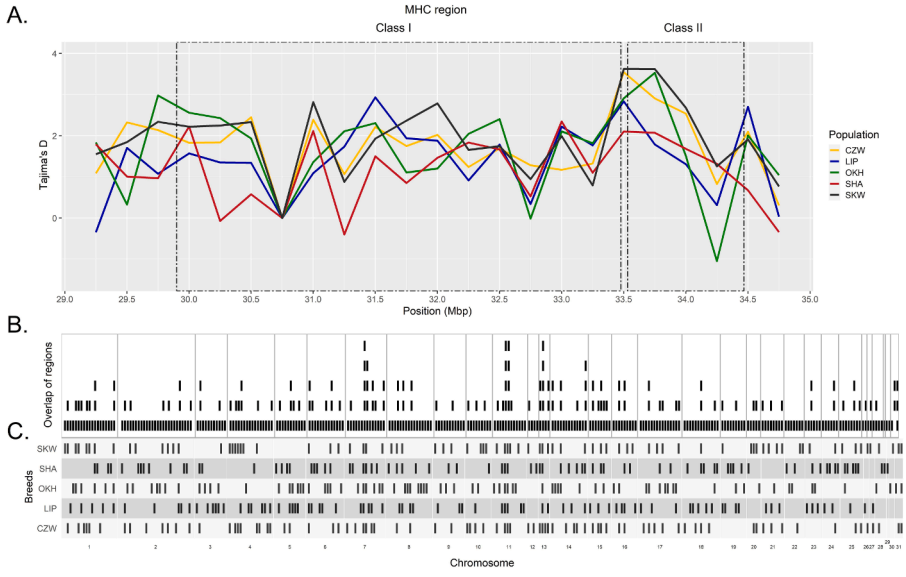


Fig. 3. Graphical visualisation of Tajima's D values in the MHC region (A., the position of MHC region is coloured in black) and distribution of balancing selection signals per chromosomes in metapopulation (B.) and breeds (C.). Gaps between signals within and between chromosomes do not correspond to the real distance in bp (LIP – Lipizzan, OKH – Old Kladruber, SHA – Shagya Arabian, CZW – Czech Warmblood, SKW – Slovak Warmblood).

EQMCE1 and in Czech and Slovak Warmbloods in the region of MHC class II genes *DQB1, DQA2, DQB2, DQA3* and *DRB2*. Further analysis of balancing selection footprints indicated that the increase in heterozygosity around MHC class II genes in Czech and Slovak Warmbloods could be a consequence of balancing selection pressure. MHC class II genes are integral for the immune system, as they encode proteins that present antigens to specialised immune cells, helping to initiate immune responses against pathogens and foreign substances. These genes are mainly expressed on specialised antigen-presenting cells such as dendritic cells, macrophages, and B-cells, as well as on activated T cells (Vasoya et al., 2023). A significant effect of selection on the MHC genomic region was also found in other livestock species, including cattle (Gautier et al., 2009; Nayak et al., 2023), sheep (Kijas et al., 2012) or pigs (Tong et al., 2023).

Observed genome-wide distribution of shared HRR islands and selection signals by all breeds confirmed that excess in heterozygosity may be strongly associated with balancing selection pressure on certain genomic regions. Both HRR islands and selection signals were concentrated in adjacent genomic regions on chromosomes 7 and 11. Overlap between regions was evident mainly on chromosome 11 close to the *CA10* gene and QTL:29,289 associated with insect bite hypersensitivity in Shetland Pony (Schurink et al., 2012). While specific research on the role of the *CA10* gene (carbonic anhydrase) in horses has not been documented yet, studies in other species suggest its role in the central nervous system (Sterky et al., 2017).

5. Conclusion

A comparison of variability stored in the autosomal genome, especially in the MHC region, between five horse breeds suggested that breeds with open breeding programs can generally exhibit higher heterozygosity levels compared to breeds whose gene pool is limited to a relatively low number of sire lines or maternal families. Although tested breeds revealed certain differences in LD extent, it appears that they shared common LD patterns close to the several MHC class I and class II genes. Screening of HRRs confirmed that even if they are present in the MHC region of all breeds, their position, size and number depend on the evaluated breed. The highest proportion of shared HRRs revealed historically connected Czech and Slovak Warmblood breeds. Detailed analysis of the distribution of HRR islands and balancing selection footprints inside the MHC region indicated that they may be concentrated around several class II genes. A further study utilising whole genome sequence data can bring more light on the interaction between genetic variants responsible for one of the most important complexes of genes related to immunity response in vertebrates.

CRediT authorship contribution statement

Nina Moravčíková: Writing – original draft, Visualization, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Radovan Kasarda:** Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Hana Vostra Vydrova:** Writing – review & editing, Validation, Data curation. **Lubos Vostry:** Writing – review & editing, Validation, Supervision, Investigation, Data curation. **Barbora Karásková:** Data curation. **Juraj Candrák:** Writing – review & editing, Validation, Resources. **Marko Halo:** Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgement

We would like to thank National Stud Farm Topolčianky, National Stud at Kladruby nad Labem, private owners and all the people who were involved in the study and contributed to collecting biological material.

Funding

This work was supported by the Slovak Research and Development Agency [project numbers APVV-17-0060 and APVV-20-0161]; and the GACR [project number 24-14325L] and by “S grant of CZU”.

Author Statement

Dear Editor, we are sending a revised version of manuscript entitled "Comparison of genomic variability of chromosome 20 and MHC region in horse populations with closed and open herdbooks". We certify that this paper consists of original, unpublished work which is not under consideration for publication elsewhere. We hope that the revised manuscript meets the high standards of Livestock Science journal and replies to all reviewers comments. Topic of the Manuscript has been changed according to the reviewer suggestion, additional co-author was included and another project has been acknowledged. We are looking forward to receiving a positive response from you regarding the acceptance of our revised manuscript and positive approval from reviewers.

Sincerely,
Radovan Kasarda

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.livsci.2024.105480.

References

Ablondi, M., Eriksson, S., Mikko, S., 2022. Haplotype blocks and heterozygosity rich regions on ECA2 in Swedish Warmblood horses. In: Veerkamp, R.F., de Haas, Y. (Eds.), Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP). Wageningen Academic Publishers, pp. 3159–3162.

Achmann, R., Curik, I., Dovc, P., Kavar, T., Bodo, I., Habe, F., Marti, E., Sölkner, J., Brem, G., 2004. Microsatellite diversity, population subdivision and gene flow in the Lipizzan horse. Anim. Genet. 35, 285–292. https://doi.org/10.1111/j.1365-2052.2004.01157.x.

Al-Mamun, H.A., Clark, S.A., Kwan, P., Gondro, C., 2015. Genome-wide linkage disequilibrium and genetic diversity in five populations of Australian domestic sheep. Genet. Sel. Evol. 47, 90. https://doi.org/10.1186/s12711-015-0169-6.

Andersson, L.S., Swinburne, J.E., Meadows, J.R., Broström, H., Eriksson, S., Fikse, W.F., Frey, R., Sundquist, M., Tseng, C.T., Mikko, S., Lindgren, G., 2012. The same ELA class II risk factors confer equine insect bite hypersensitivity in two distinct populations. Immunogenetics 64, 201–208. https://doi.org/10.1007/s00251-011-0573-1.

Barcaccia, G., Felicetti, M., Galla, G., Capomaccio, S., Cappelli, K., Albertini, E., Buttazzoni, L., Pieramati, C., Silvestrelli, M., Verini Supplizi, A., 2013. Molecular analysis of genetic diversity, population structure and inbreeding level of the Italian Lipizzan horse. Livest. Sci. 151, 124–133. https://doi.org/10.1016/j.livsci.2012.11.022.

Bartolomé, E., Valera, M., Fernández, J., Rodríguez-Ramilo, S.T., 2022. Effects of selection on breed contribution in the caballo de deporte español. Animals 12, 1635. https://doi.org/10.3390/ani12131635.

Biscarini, F., Cozzi, P., Gaspa, G., Marras, G., 2018. detectRUNS: an R package to detect runs of homozygosity and heterozygosity in diploid genomes. https://rdrr.io/cran/detectRUNS/f/vignettes/detectRUNS.vignette.Rmd (accessed 10 April 2023).

Biscarini, F., Mastrangelo, S., Catillo, G., Senczuk, G., Ciampolini, R., 2020. Insights into genetic diversity, runs of homozygosity and heterozygosity-rich regions in maremmana semi-feral cattle using pedigree and genomic data. Animals 10, 2285. https://doi.org/10.3390/ani10122285.

Bowling, A.T., Eggleston-Stott, M.L., Byrns, G., Clark, R.S., Dileanis, S., Wictum, E., 1997. Validation of microsatellite markers for routine horse parentage testing. Anim. Genet. 28, 247–252. https://doi.org/10.1111/j.1365-2052.1997.00123.x.

Brinkmeyer-Langford, C.L., Cai, J.J., Gill, C.A., Skow, L.C., 2013. Microsatellite variation in the equine MHC. Anim. Genet. 44, 267–275. https://doi.org/10.1111/age.12003.

Burger, D., Meuwly, C., Marti, E., Sieme, H., Oberthür, M., Janda, J., Meinecke-Tillmann, S., Wedekind, C., 2017. MHC-correlated preferences in diestrous female

horses (Equus caballus). Theriogenology 89, 318–323. https://doi.org/10.1016/j.theriogenology.2016.09.015.
Carpenter, S., Baker, J.M., Bacon, S.J., Hopman, T., Maher, J., Ellis, S.A., Antczak, D.F., 2001. Molecular and functional characterization of genes encoding horse MHC class I antigens. Immunogenetics 53, 802–809. https://doi.org/10.1007/s00251-001-0384-x.
Cervantes, I., Gutiérrez, J.P., Molina, A., Goyache, F., Valera, M., 2009. Genealogical analyses in open populations: the case of three Arab derived Spanish horse breeds. J. Anim. Breed. Genet. 126, 335–347. https://doi.org/10.1111/j.1439-0388.2008.00797.x.
Chang, C.C., Chow, C.C., Twllier, L.C.A.M., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7. https://doi.org/10.1186/s13742-015-0047-8.
Charlesworth, D., 2006. Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet 2, e64. https://doi.org/10.1371/journal.pgen.0020064.
Chen, Z., Zhang, Z., Wang, Z., Zhang, Z., Wang, Q., Pan, Y., 2022. Heterozygosity and homozygosity regions affect reproductive success and the loss of reproduction: a case study with litter traits in pigs. Comput. Struct. Biotechnol. J. 20, 4060–4071. https://doi.org/10.1016/j.csbj.2022.07.039.
Cortés, O., Eusebi, P., Dunner, S., Sevane, N., Cañón, J., 2019. Comparison of diversity parameters from SNP, microsatellites and pedigree records in the Lidia cattle breed. Livest. Sci. 219, 80–85. https://doi.org/10.1016/j.livsci.2018.11.006.
Cosgrove, E.J., Sadeghi, R., Schlamp, F., Holl, H.M., Moradi-Shahrbabak, M., Miraei-Ashtiani, S.R., Abdalla, S., Shykind, B., Troedsson, M., Stefaniuk-Szmukier, M., Prabhu, A., Bucca, S., Bugno-Poniewierska, M., Wallner, B., Malek, J., Miller, D.C., Clark, A.G., Antczak, D.F., Brooks, S.A., 2020. Genome diversity and the origin of the arabian horse. Sci. Rep. 10, 9702. https://doi.org/10.1038/s41598-020-66232-1.
Criscione, A., Mastrangelo, S., D'Alessandro, E., Tumino, S., Di Gerlando, R., Zumbo, A., Marletta, D., Bordonaro, S., 2022. Genome-wide survey on three local horse populations with a focus on runs of homozygosity pattern. J. Anim. Breed. Genet. 139, 540–555. https://doi.org/10.1111/jbg.12680.
Cunningham, F., et al., 2022. Ensembl 2022. Nucleic. Acids. Res. 50 (D1), D988–D995. https://doi.org/10.1093/nar/gkab1049.
Curik, I., Fraser, D., Eder, C., Achmann, R., Swinburne, J., Crameri, R., Brem, G., Sölkner, J., Marti, E., 2003a. Association between the MHC gene region and variation of serum IgE levels against specific mould allergens in the horse. Genet. Sel. Evol. 35 (Suppl 1), S177–S190. https://doi.org/10.1186/1297-9686-35-S1-S177.
Curik, I., Zechner, P., Sölkner, J., Achmann, R., Bodo, I., Dovc, P., Kavar, T., Marti, E., Brem, G., 2003b. Inbreeding, microsatellite heterozygosity, and morphological traits in Lipizzan horses. J. Hered. 94, 125–132. https://doi.org/10.1093/jhered/esg029.
Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group, 2011. The variant call format and VCFtools. Bioinformatics. 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.
Dovc, P., Kavar, T., Sölkner, H., Achmann, R., 2006. Development of the Lipizzan horse breed. Reprod. Domest. Anim. 41, 280–285. https://doi.org/10.1111/j.1439-0531.2006.00726.x.
Durward-Akhurst, S.A., Finno, C.J., Barnes, N., Shivers, J., Guo, L.T., Shelton, G.D., Valberg, S.J., 2016. Major histocompatibility complex i and ii expression and lymphocytic subtypes in muscle of horses with immune-mediated myositis. J. Vet. Intern. Med. 30, 1313–1321. https://doi.org/10.1111/jvim.14371.
Engel, L., Becker, D., Nissen, T., Russ, I., Thaller, G., Krattenmacher, N., 2022. Mitochondrial DNA variation contributes to the aptitude for dressage and show jumping ability in the holstein horse breed. Animals 12, 704. https://doi.org/10.3390/ani12060704.
Faria, R., Johannesson, K., Butlin, R.K., Westram, A.M., 2019. Evolving Inversions. Trend Ecol. Evol. 34, 239–248. https://doi.org/10.1016/j.tree.2018.12.005.
Ferenčaković, M., Banadinović, M., Mercvajler, M., Khayatzadeh, N., Mészáros, G., Cubric-Curik, V., Curik, I., Sölkner, J., 2016. Mapping of heterozygosity rich regions in Austrian pinzgauer cattle. Acta Agric. Slov. 5, 41–44.
Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., Moazami-Goudarzi, K., Foulley, J.L., 2009. A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. BMC Genom 10, 550. https://doi.org/10.1186/1471-2164-10-550.
Grilz-Seger, G., Neuditschko, M., Ricard, A., Velie, B., Lindgren, G., Mesarič, M., Cotman, M., Horna, M., Dobretsberger, M., Brem, G., Druml, T., 2019. Genome-Wide Homozygosity Patterns and Evidence for Selection in a Set of European and Near Eastern Horse Breeds. Genes (Basel) 10, 491. https://doi.org/10.3390/genes10070491.
Gurgul, A., Jasielczuk, I., Semik-Gurgul, E., Pawlina-Tyszko, K., Stefaniuk-Szmukier, M., Szmatoła, T., Polak, G., Tomczyk-Wrona, I., Bugno-Poniewierska, M., 2019. A genome-wide scan for diversifying selection signatures in selected horse breeds. PLoS ONE 14, e0210751. https://doi.org/10.1371/journal.pone.0210751.
Hemstrom, W., Jones, M., 2023. snpR: user friendly population genomics for SNP data sets with categorical metadata. Mol. Ecol. Resour. 23, 962–973. https://doi.org/10.1111/1755-0998.13721.
Hill, W.G., Robertson, A., 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8, 269–294.
Horecky, C., Horecka, E., Futas, J., Janova, E., Horin, P., Knoll, A., 2018. Microsatellite markers for evaluating the diversity of the natural killer complex and major histocompatibility complex genomic regions in domestic horses. HLA 91, 271–279. https://doi.org/10.1111/tan.13211.
Hu, Z.L., Park, C.A., Reecy, J.M., 2019. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. Nucleic. Acids. Res. 47, D701–D710. https://doi.org/10.1093/nar/gky1084.
Huang, D., Sherman, B., Lempicki, R., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57. https://doi.org/10.1038/nprot.2008.211.
Janova, E., Futas, J., Klumplerova, M., Putnova, L., Vrtkova, I., Vyskocil, M., Frolkova, P., Horin, P., 2013. Genetic diversity and conservation in a small endangered horse population. J. Appl. Genet. 54, 285–292. https://doi.org/10.1007/s13353-013-0151-3.
Jaworska, J., Ropka-Molik, K., Wocławek-Potocka, I., Siemieniuch, M., 2020. Inter- and intrabreed diversity of the major histocompatibility complex (MHC) in primitive and draft horse breeds. PLoS ONE 15, e0228658. https://doi.org/10.1371/journal.pone.0228658.
Jeannerat, E., Marti, E., Thomas, S., Herrera, C., Sieme, H., Wedekind, C., Burger, D., 2020. Embryo survival in the oviduct not significantly influenced by major histocompatibility complex social signaling in the horse. Sci. Rep. 10, 1056. https://doi.org/10.1038/s41598-020-58056-w.
Kasarda, R., Vostrý, L., Moravčíková, N., Vostrá-Vydrová, H., Dovč, P., Kadlečík, O., 2016. Detailed insight into genetic diversity of the Old Kladruber horse substructure in comparison to the Lipizzan breed. Acta Agric. Scand. A Anim. Sci. 66, 67–74. https://doi.org/10.1080/09064702.2016.1249400.
Khaudov, A.D., Duduev, A.S., Kokov, Z.A., Amshokov, K.K., Zhekamukhov, M.K., Zaitsev, A.M., Reissmann, M., 2018. Genetic analysis of maternal and paternal lineages in Kabardian horses by uniparental molecular markers. Open Vet. J. 8, 40–46. https://doi.org/10.4314/ovj.v8i1.7.
Kijas, J.W., Lenstra, J.A., Hayes, B., Boitard, S., Porto Neto, L.R., San Cristobal, M., Servin, B., McCulloch, R., Whan, V., Gietzen, K., Paiva, S., Barendse, W., Ciani, E., Raadsma, H., McEwan, J., Dalrymple, B., International Sheep Genomics Consortium Members, 2012. Genome-Wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol 10, e1001258. https://doi.org/10.1371/journal.pbio.1001258.
Kim, S., Lim, B., Do, K., 2019. Identification of divergently selected regions in the thoroughbred and Jeju Pony by using Equine SNP Chip Array. Intl. J. Agric. Biol. 22, 1079–1086. https://doi.org/10.17957/IJAB/15.1172.
Klumplerova, M., Vychodilova, L., Bobrova, O., Cvanova, M., Futas, J., Janova, E., Vyskocil, M., Vrtkova, I., Putnova, L., Dusek, L., Marti, E., Horin, P., 2013. Major histocompatibility complex and other allergy-related candidate genes associated with insect bite hypersensitivity in Icelandic horses. Mol. Biol. Rep. 40, 3333–3340. https://doi.org/10.1007/s11033-012-2408-z.
Kulski, J.K., Shiina, T., Anzai, T., Kohara, S., Inoko, H., 2002. Comparative Genomic analysis of the mhc: the evolution of class i duplication blocks, diversity and complexity from shark to man. Immunol. Rev. 190, 95–122. https://doi.org/10.1034/j.1600-065X.2002.19008.x.
Liang, M., Shishkin, M., Mikhailova, A., Shchur, V., Nielsen, R., 2022. Estimating the timing of multiple admixture events using 3-locus linkage disequilibrium. PLoS Genet 18, e1010281. https://doi.org/10.1371/journal.pgen.1010281.
Lucek, K., Willi, Y., 2021. Drivers of linkage disequilibrium across a species' geographic range. PLoS Genet 17, e1009477. https://doi.org/10.1371/journal.pgen.1009477.
Machmoum, M., Boujenane, I., Azelhak, R., Badaoui, B., Petit, D., Piro, M., 2020. Genetic diversity and population structure of arabian horse populations using microsatellite markers. J. Equine Vet. Sci. 93, 103200 https://doi.org/10.1016/j.jevs.2020.103200.
Marras, G., Wood, B.J., Makanjuola, B., Malchiodi, F., Peeters, K., van As, P., Baes, C.F., Biscarini, F., Turkeys, H., 2018. Characterization of runs of homozygosity and heterozygosity-rich regions in a commercial turkey (Meleagris gallopavo) population. In: Proceedings of the 11th World Congress of Genetics Applied to Livestock Production, Species - Avian, 1, p. 763.
Metzger, J., Karwath, M., Tonda, R., Beltran, S., Águeda, L., Gut, M., Gut, I.G., Distl, O., 2015. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. BMC Genomics 16, 764.. https://doi.org/10.1186/s12864-015-1977-3.
Nayak, S.S., Panigrahi, M., Kumar, H., Rajawat, D., Sharma, A., Bhushan, B., Dutt, T., 2023. Evidence for selective sweeps in the MHC gene repertoire of various cattle breeds. Anim. Biotechnol. 34, 4167–4173. https://10.1080/10495398.2023.2196317.
Nolte, W., Alkhoder, H., Wobbe, M., Stock, K.F., Kalm, E., Vosgerau, S., Krattenmacher, N., Thaller, G., Tetens, J., Kühn, C., 2022. Replacement of microsatellite markers by imputed medium-density SNP arrays for parentage control in German warmblood horses. J. Appl. Genet. 63, 783–792. https://doi.org/10.1007/s13353-022-00725-9.
Nolte, W., Thaller, G., Kuehn, C., 2019. Selection signatures in four German warmblood horse breeds: tracing breeding history in the modern sport horse. PLoS ONE 14, e0215913. https://doi.org/10.1371/journal.pone.0215913.
Ong, R.T., Teo, Y.Y., 2010. varLD: a program for quantifying variation in linkage disequilibrium patterns between populations. Bioinformatics 26, 1269–1270. https://doi.org/10.1093/bioinformatics/btq125.
Oyelami, F.O., Zhao, Q., Xu, Z., Zhang, Z., Sun, H., Zhang, Z., Ma, P., Wang, Q., Pan, Y., 2020. Haplotype block analysis reveals candidate genes and QTLs for meat quality and disease resistance in chinese jiangquhai pig breed. Front. Genet. 11, 752. https://doi.org/10.3389/fgene.2020.00752.
Petersen, J.L., Mickelson, J.R., Rendahl, A.K., Valberg, S.J., Andersson, L.S., Axelsson, J., Bailey, E., Bannasch, D., Binns, M.M., Borges, A.S., Brama, P., da Câmara Machado, A., Capomaccio, S., Cappelli, K., Cothran, E.G., Distl, O., Fox-Clipsham, L., Graves, K.T., Guérin, G., Haase, B., Hasegawa, T., Hemmann, K., Hill, E.W., Leeb, T., Lindgren, G., Lohi, H., Lopes, M.S., McGivney, B.A., Mikko, S., Orr, N., Penedo, M.C., Piercy, R.J., Raekallio, M., Rieder, S., Røed, K.H., Swinburne, J., Tozaki, T.,

Vaudin, M., Wade, C.M., McCue, M.E., 2013. Genome-wide analysis reveals selection for important traits in domestic horse breeds. PLoS Genet 9, e1003211. https://doi.org/10.1371/journal.pgen.1003211.

Putnová, L., Štohl, R., Vrtková, I., 2019. Using nuclear microsatellite data to trace the gene flow and population structure in Czech horses. Czech J. Anim. Sci. 64, 67–77. https://doi.org/10.17221/2/2018-CJAS.

Radovic, L., Remer, V., Krcal, C., Rigler, D., Brem, G., Rayane, A., Driss, K., Benamar, M., Machmoum, M., Piro, M., Krischke, D., Butler-Wemken, I.V., Wallner, B., 2022. Y Chromosome haplotypes enlighten origin, influence, and breeding history of North African Barb Horses. Animals 12, 2579. https://doi.org/10.3390/ani12192579.

Radwan, J., Babik, W., Kaufman, J., Lenz, T.L., Winternitz, J., 2020. Advances in the evolutionary understanding of MHC polymorphism. Trends Genet 36, 298–311. https://doi.org/10.1016/j.tig.2020.01.008.

Rothbard, J.B., Gefter, M.L., 1991. Interactions between immunogenic peptides and MHC proteins. Annu. Rev. Immunol. 9, 527–565. https://doi.org/10.1146/annurev.iy.09.040191.002523.

Santos, W., Schettini, G., Fonseca, M.G., Pereira, G.L., Chardulo, L.A., Neto, O., Baldassini, W.A., Oliveira, H., Curi, R., 2021. Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the Mangalarga Marchador horse breed. J. Anim. Breed. Genet. 138, 161–173. https://doi.org/10.1111/jbg.12508.

Santos, W.B., Pereira, C.B., Maiorano, A.M., Arce, C.D.S., Baldassini, W.A., Pereira, G.L., Chardulo, L.A.L., Neto, O.R.M., Oliveira, H.N., Curi, R.A., 2023. Genomic inbreeding estimation, runs of homozygosity, and heterozygosity-enriched regions uncover signals of selection in the Quarter Horse racing line. J. Anim. Breed. Genet. 140, 583–595. https://doi.org/10.1111/jbg.12812.

Schaefer, R.J., McCue, M.E., 2020. Equine Genotyping Arrays. Vet. Clin. North. Am. Equine Pract. 36, 183–193. https://doi.org/10.1016/j.cveq.2020.03.001.

Schurink, A., Wolc, A., Ducro, B.J., Frankena, K., Garrick, D.J., Dekkers, J.C.M., van Arendonk, J.A.M., 2012. Genome-wide association study of insect bite hypersensitivity in two horse populations in the Netherlands. Genet. Sel. Evol. 44, 31. https://doi.org/10.1186/1297-9686-44-31.

Selli, A., Ventura, R.V., Fonseca, P.A.S., Buzanskas, M.E., Andrietta, L.T., Balieiro, J.C.C., Brito, L.F., 2021. Detection and Visualization of Heterozygosity-Rich Regions and Runs of Homozygosity in Worldwide Sheep Populations. Animals 11, 2696. https://doi.org/10.3390/ani11092696.

Simčič, M., Smetko, A., Sölkner, J., Seichter, D., Gorjanc, G., Kompan, D., Medugorac, I., 2015. Recovery of native genetic background in admixed populations using haplotypes, phenotypes, and pedigree information–using Cika cattle as a case breed. PLoS ONE 10, e0123253. https://10.1371/journal.pone.0123253.

Simonsen, K.L., Churchill, G.A., Aquadro, C.F., 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141, 413–429. https://10.1093/genetics/141.1.413.

Slatkin, M., 2008. Linkage disequilibrium–understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. 9, 477–485. https://doi.org/10.1038/nrg2361.

Sommer, S., 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front. Zool. 2, 16. https://doi.org/10.1186/1742-9994-2-16.

Staiger, E.A., Tseng, C.T., Miller, D., Cassano, J.M., Nasir, L., Garrick, D., Brooks, S.A., Antczak, D.F., 2016. Host genetic influence on papillomavirus-induced tumors in the horse. Int. J. Cancer. 139, 784–792. https://doi.org/10.1002/ijc.30120.

Sterky, F.H., Trotter, J.H., Lee, S.J., Recktenwald, C.V., Du, X., Zhou, B., Zhou, P., Schwenk, J., Fakler, B., Südhof, T.C., 2017. Carbonic anhydrase-related protein CA10 is an evolutionarily conserved pan-neurexin ligand. Proc. Natl. Acad. Sci. USA. 114, E1253–E1262. https://10.1073/pnas.1621321114.

Szmatoła, T., Gurgul, A., Jasielczuk, I., Oclon, E., Ropka-Molik, K., Stefaniuk-Szmukier, M., Polak, G., Tomczyk-Wrona, I., Bugno-Poniewierska, M., 2022. Assessment and distribution of runs of homozygosity in horse breeds representing different utility types. Animals 12, 3293. https://doi.org/10.3390/ani12233293.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595. https://10.1093/genetics/123.3.585.

Tallmadge, R.L., Campbell, J.A., Miller, D.C., Antczak, D.F., 2010. Analysis of MHC class I genes across horse MHC haplotypes. Immunogenetics 62, 159–172. https://doi.org/10.1007/s00251-009-0420-9.

Tallmadge, R.L., Lear, T.L., Antczak, D.F., 2005. Genomic characterization of MHC class I genes of the horse. Immunogenetics 57, 763–774. https://doi.org/10.1007/s00251-005-0034-9.

Tang, Y., Liu, G., Zhao, S., Li, K., Zhang, D., Liu, S., Hu, D., 2022. Major histocompatibility complex (MHC) diversity of the reintroduction populations of endangered przewalski's horse. Genes (Basel) 13, 928. https://doi.org/10.3390/genes13050928.

Teo, Y.Y., Fry, A.E., Bhattacharya, K., Small, K.S., Kwiatkowski, D.P., Clark, T.G., 2009. Genome-wide comparisons of variation in linkage disequilibrium. Genome Res 19, 1849–1860. https://doi.org/10.1101/gr.092189.109.

Tong, X., Chen, D., Hu, J., Lin, S., Ling, Z., Ai, H., Zhang, Z., Huang, L., 2023. Accurate haplotype construction and detection of selection signatures enabled by high quality pig genome sequences. Nat. Commun. 14, 5126. https://doi.org/10.1038/s41467-023-40434-3.

Tseng, C.T., Miller, D., Cassano, J., Bailey, E., Antczak, D.F., 2010. Identification of equine major histocompatibility complex haplotypes using polymorphic microsatellites. Anim. Genet. 41, 150–153. https://doi.org/10.1111/j.1365-2052.2010.02125.x.

Vasoya, D., Tzelos, T., Benedictus, L., Karagianni, A.E., Pirie, S., Marr, C., Oddsdóttir, C., Fintl, C., Connelley, T., 2023. High-Resolution Genotyping of Expressed Equine MHC Reveals a Highly Complex MHC Structure. Genes (Basel) 14, 1422. https://10.3390/genes14071422.

Viļuma, A., Mikko, S., Hahn, D., Skow, L., Andersson, G., Bergström, T.F., 2017. Genomic structure of the horse major histocompatibility complex class II region resolved using PacBio long-read sequencing technology. Sci. Rep. 7, 45518. https://doi.org/10.1038/srep45518.

Vostrý, L., Kracíková, O., Hofmanová, B., Czerneková, V., Kott, T., Přibyl, J., 2011. Intra-line and inter-line genetic diversity in sire lines of the Old Kladruber horse based on microsatellite analysis of DNA. Czech J. Anim. Sci. 56, 163–175.

Vostry, L., Vostra-Vydrova, H., Citek, J., Gorjanc, G., Curik, I., 2021. Association of inbreeding and regional equine leucocyte antigen homozygosity with the prevalence of insect bite hypersensitivity in Old Kladruber horse. Anim. Genet. 52, 422–430. https://doi.org/10.1111/age.13075.

Vostry, L., Vydrová-Vostrá, H., Hofmanová, B., Moravčíková, N., Veselá, Z., Majzlík, I., 2018. Genetic diversity in five Czech native horse breeds assessed using microsatellite markers. Acta Fytotechn. Zootechn. 21, 190–193. https://doi.org/10.15414/afz.2018.21.04.190-193.

Waples, R.K., Larson, W.A., Waples, R.S., 2016. Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. Heredity (Edinb) 117, 233–240. https://doi.org/10.1038/hdy.2016.60.

Wickham, H., 2016. ggplot2: Elegant Graphics For Data Analysis. Springer-Verlag, New York.

Williams, J.L., Hall, S.J.G., Del Corvo, M., Ballingall, K.T., Colli, L., Ajmone Marsan, P., Biscarini, F., 2016. Inbreeding and purging at the genomic Level: the Chillingham cattle reveal extensive, non-random SNP heterozygosity. Anim. Genet. 47, 19–27. https://doi.org/10.1111/age.12376.