Original Paper

Preliminary analysis of utilization of genomic relationship in mating plan of Old Kladruber horse

Jiří Bauer*, Josef Přibyl, Luboš Vostrý

Institute of Animal Science, Praha–Uhříněves, Czech Republic

The study analyzed 48 Old Kladruber horses genotyped by Illumina Equine SNP70 BeadChip for usefulness of genomic data in determining of mating plan. Totally 12 variants of data filtering and their impact on calculations in dependence of different parameters of GenCall Score, Minor Allele Frequency and assumed average values of loci of ancestors was investigated. For possibility of comparison between genomic and commonly evaluated relationships, pedigree based relationship matrix was constructed and subsequently subtraction of pedigree from genomic matrix was performed. All matrices were thoroughly inspected and most suitable setting of parameters was chosen. Evaluation of genomic relationships can be successfully implemented in more precise method of mating plan design of Old Kladruber horses. Further genotyping and development of method for rescaling of differences between genomic and pedigree relationship matrices´ elements is advised for a purpose of better interpretation of results by breeders.

Keywords: horse, Old Kladruber, genomic relationship, mating plan, SNP

1   Introduction

New methods of analyses developed in last decade enabled to use more detailed genetic information about animals. One of these methods utilizes massive detecting of single nucleotide polymorphisms by biochips and the information is used in single-step and multi-step prediction of genomic BLUP breeding values of cattle. However, detected SNPs can be used in many other ways such as in association studies for traits of interest (e.g., Tiezzi et al., 2015), more accurate relationships among animals (VanRaden 2008), genetic characterization of breed (McKay et al., 2008) and genetic distances among breeds (Gautier et al., 2010). If the breed belongs to population established as genetic resource, preservation of genetic character, variability and sustainability of population are essential concerns and these new methods can help to more accurately choose genetically suitable parents.

In horses, genotyping by biochips for SNP identification is ongoing subsequent implementation. All of aforementioned directions of research are being developed (e.g., Haberland et al., 2012; Boyko et al., 2014; Petersen et al., 2013 or Schubertová et al., 2014). Moreover use of microsatellites for identification of horse breeds characteristics and parentage confirmation is now routinely carried out.

Old Kladruber horse as UNESCO heritage of mankind and Czech official genetic resource is aim of research for improvement of breeding conservation program and diversity of population by microsatellites was investigated by Vostrý et al., 2011. The next step in the process is inclusion of SNP detection in determination of genomic relationships among horses and incorporation of the results simultaneously with improved pedigree analysis method done by Vostrá-Vydrová et al. 2016 into mating plan design. Our study is focused on development of method of construction of genomic relationships from SNPs identified by biochips and their practical implementation in mating plan of Old Kladruber horse.

---

* **Corresponding Author:** Jiří Bauer, Institute of Animal Science, Přátelství 815, Praha – Uhříněves 104 00, Czech Republic.
E-mail:bauer.jiri@vuzv.cz

Faculty of Agrobiology and Food Resources

2   Material and methods

For preliminary analysis, 48 Old Kladruber horses (10 stallions, 38 mares) active in reproduction were genotyped by Illumina Equine SNP70 BeadChip for identification of 65 157 SNPs. Totally 12 variants of data filtering and their impact on calculations in the dependence of GenCall Score (GCS), Minor Allele Frequency (MAF) and different assumed average values of loci of ancestors ($\bar{x}$l) was investigated and variants are described in Table 1. From filtered data, genomic relationship matrix (**G**) was calculated by VanRaden 2008 method and rescaled to average diagonal element value of 1 (Forni et al., 2011).

Table 1   Overview of variants of SNP data filtering included in the study

variant	inclusion of data		average ancestors´ loci value
	GenCall Score	Minor Allele Frequency	
1	0.5 and higher	higher than 5 %	average of genotyped population
2	0.7 and higher	higher than 5 %	average of genotyped population
3	0.5 and higher	no filtering	0 (i.e., homozygote A)
4	0.7 and higher	no filtering	0 (homozygote A)
5	0.5 and higher	higher than 5 %	1 (heterozygote)
6	0.7 and higher	higher than 5 %	1 (heterozygote)
7	0.5 and higher	no filtering	average of genotyped population
8	0.7 and higher	no filtering	average of genotyped population
9	0.5 and higher	higher than 5 %	0 (homozygote A)
10	0.7 and higher	higher than 5 %	0 (homozygote A)
11	0.5 and higher	no filtering	1 (heterozygote)
12	0.7 and higher	no filtering	1 (heterozygote)

homozygote of type A had in calculations value 0, heterozygote 1 and homozygote B corresponded to value 2

For possibility of comparison between genomic and common relationships, pedigree based relationship matrix (**A**) was constructed on the basis of calculation of Vostrá-Vydrová et al. 2016. Shifting of **G** to equal average element value as in **A**, which enables subtraction of **A** from **G** (Vitezica et al. 2011), was further carried out. This adjustment corresponded to similar procedure done in single-step GBLUP method (Misztal et al., 2009). Resulting genomic and pedigree relationship matrices were inspected on the basis of their statistical characteristics, distribution of individual values and correlation coefficients between **A** and **G** for all 12 variants of data filtering.

The most suitable setting for parameters GCS, MAF, $\bar{x}$l was chosen on the basis of comparison among variants by aforementioned inspected criteria and possibility of use in additional calculations, ease of interpretation and usability by breeders were also taken into account. Values of diagonal and nondiagonal elements of matrices were analyzed for variant selection as well.

3   Results and discussion

Several loci were not successfully analyzed by laboratory in genotyped population and consequently average number of identified loci per horse was 63 912. Average GCS was 0.72 with standard deviation 0.186. The distribution of GCS in genotyped horses is shown in Figure 1. High amount of homozygous and close to homozygous loci were detected in genotyped horses (see Table 2). Chips usually comprise of heterozygous loci with high predicative ability cross the breeds. The homozygote condition of these loci could therefore be characteristic for Old Kladruber breed. From this point of view, SNP loci can be partitioned into group which characterizes the breed (about 18 % of loci) and group of loci which characterize the interbreed variability and relationship (remaining proportion). However more genotyped individuals would be necessary to confirm this hypothesis. The number of SNP in the analysis after filtering for different GCS and MAF conditions in 12 variants of calculation and in the case of filtering MAF lower than 2 % are described in Table 3 and basic characteristics of genomic relationship matrices in different variants are in Table 4.

Faculty of Agrobiology and Food Resources

In calculation of genomic relationship matrix, different settings of average ancestors´ loci change definitions of relationships among animals, most noticeably in comparison of diagonal elements of **G** among variants. In $\bar{x}l = 0$ cluster, loci where most animals are homozygous are counted to relationship value and therefore minimal element of **G** matrix is high; in the opposite case when $\bar{x}l$ equals mean of genotyped population, individual relationships are expressed in relation to the mean and mutual relation among individuals can be much lower (negative) than average or much higher as in case of diagonal elements with value considerably higher than 1.
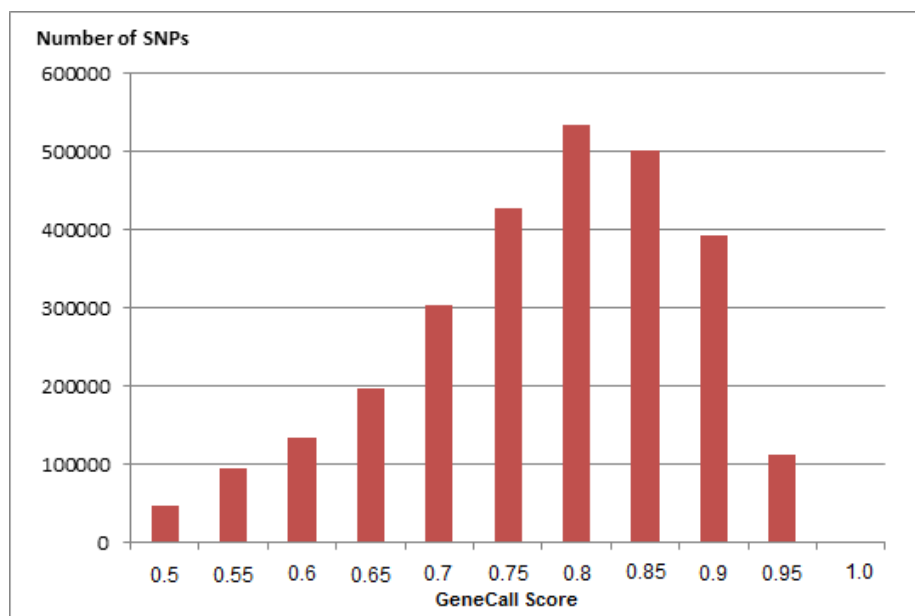
GCS=0.5 and higher; with GCS filtering of 0.7 and higher the distribution was very similar

Figure 1   Distribution of GenCall Score of SNP allele detection in genotyped population

Table 2   Detected almost homozygous and homozygous SNP loci in genotyped horses

Ratio of homozygous alleles in individual SNP loci (%)	Share of loci to total number (%)
99–100	12.72
95–99	4.95
90–95	2.83
70–90	26.8
less than 70	52.71

Note: GSC higher than 0.5

Table 3   Average number of SNPs for individual horse after GCS and MAF filtering

Inclusion of GCS	MAF minimum	average no. of SNPs	min.	max.
0.5 and higher	N/A	57199	56 349	57 502
0.7 and higher	N/A	44422	43 259	44 832
0.5 and higher	2%	45027	44 367	45 135
0.7 and higher	2%	35147	34 405	35 241
0.5 and higher	5%	39879	39 532	39 953
0.7 and higher	5%	34588	30 485	31 247

Table 4 Basic characteristics of genomic and pedigree relationship matrices in 12 variants of calculations (after rescaling by Forni et al., 2011 method)

GCS	MAF minimum	correction for	average value of matrix elements	st. dev.	min.	max.
0.5 and higher	5%	mean	0.020	0.223	-0.191	1.787
0.7 and higher	5%	mean	0.020	0.223	-0.191	1.770
0.5 and higher	N/A	0	0.870	0.036	0.762	1.018
0.7 and higher	N/A	0	0.864	0.039	0.747	1.018
0.5 and higher	5%	1	0.490	0.125	0.248	1.059
0.7 and higher	5%	1	0.486	0.126	0.242	1.060
0.5 and higher	N/A	mean	0.023	0.223	-0.182	2.123
0.7 and higher	N/A	mean	0.023	0.224	-0.183	2.112
0.5 and higher	5%	0	0.081	0.047	0.717	1.021
0.7 and higher	5%	0	0.809	0.047	0.711	1.025
0.5 and higher	N/A	1	0.660	0.093	0.397	1.043
0.7 and higher	N/A	1	0.649	0.096	0.376	1.045
pedigree relationship matrix			0.342	0.183	0.135	1.189

Detected differences among values of genomic matrices of MAF 5% and nonMAF variants and among their subtractions by **A** were very small. Correlations among those variants were from 0.97 to 1.0. Variants differing in GSC filtering were almost identical with correlations 0.999 to each other. Correction for ancestor population average loci consisting of homozygotes or heterozygotes did not resulted in distinctly different genomic matrices even if various settings of MAF were considered (correlations among variants were always higher than 0.96). On the other hand if the average value of genotyped population loci was chosen for the correction, correlations to values resulting from 1 and 0 settings were from 0.8 to 0.93. Therefore the variants were divided into two large clusters according to the use of correction value (i.e., $\bar{x}$l 01 cluster and $\bar{x}$l genotyped cluster). Average correlation between **A** and **G** for all variants was 0.87, for **A** and $\bar{x}$l genotyped cluster **G** 0.94 and for **A** and $\bar{x}$l 01 cluster **G** 0.84 (0.89 in case of MAF 5 % and 0.78 for variants without MAF filtering). Individual correlations of **G** variants to **A** ranged from 0.77 to 0.95. Example of distribution of values after subtraction of genomic from pedigree relationship matrix is presented in Figure 2 and 3 for two aforementioned clusters.

While subtraction of nondiagonal elements **G** – **A** shows similar results in both clusters (2b, 3b Figures), subtracted diagonal elements were quite different (2a, 3a). The main cause was one animal that was according to genotyping distinctly less related to other horses in population. In $\bar{x}$l genotyped cluster this animal was much more related to itself than to rest of population and for this reason the diagonal element (i.e., relationship to himself expressed relatively to genotyped population mean) was very high (1.77 opposite to 0.88 to 1.06 of other horses after rescaling). In $\bar{x}$l =0 cluster this situation was not apparent as all relationships were expressed to homozygous average ancestor loci and in that case differences of all animals of current population to ancestors´ state were similar. However in nondigonal **G** elements of $\bar{x}$l =0 cluster, difference in relationships of this particular horse and population was explicit as all **G** relationships in population were 0.76 to 0.92 opposite to 0.71-0.73 of this horse. This distinction was not detectable in $\bar{x}$l genotyped cluster.

Negative values in diagonal elements in Fig. 2a are caused by shifting of G matrix to average value of **A** when average value of **G** matrix in this variant was very high and therefore subtraction value distinctly lowered values of diagonal elements (by 0.48).

For purposes of breeders we are inclined to prefer $\bar{x}$l 01 cluster variants. The main reason for this preference is the ease of interpretation of genomic matrix elements by breeders, as the elements are expressing relationships on the basis of common alleles in loci on scale 0 to 100%. Also values among genotyped individuals in this variant will not be affected by further genotyping of more individuals, which can change classification of homozygotness or heterozygotness of individual loci. However even more practical would be values of **G** – **A** subtraction matrix which would notify breeders about distinct differences between pedigree and genomic based relationship. This output is though hindered by difficulties of interpretation of rescaled, shifted and subtracted values in meaningful way.

The suggestion for utilization of $\bar{x}$I 01 cluster and clustering of results can distinctly change by more genotyped Old Kladruber horses, especially because of relatively low number of genotypes now available. Repeated evaluation of parameters for calculation after further genotyping is therefore recommended.
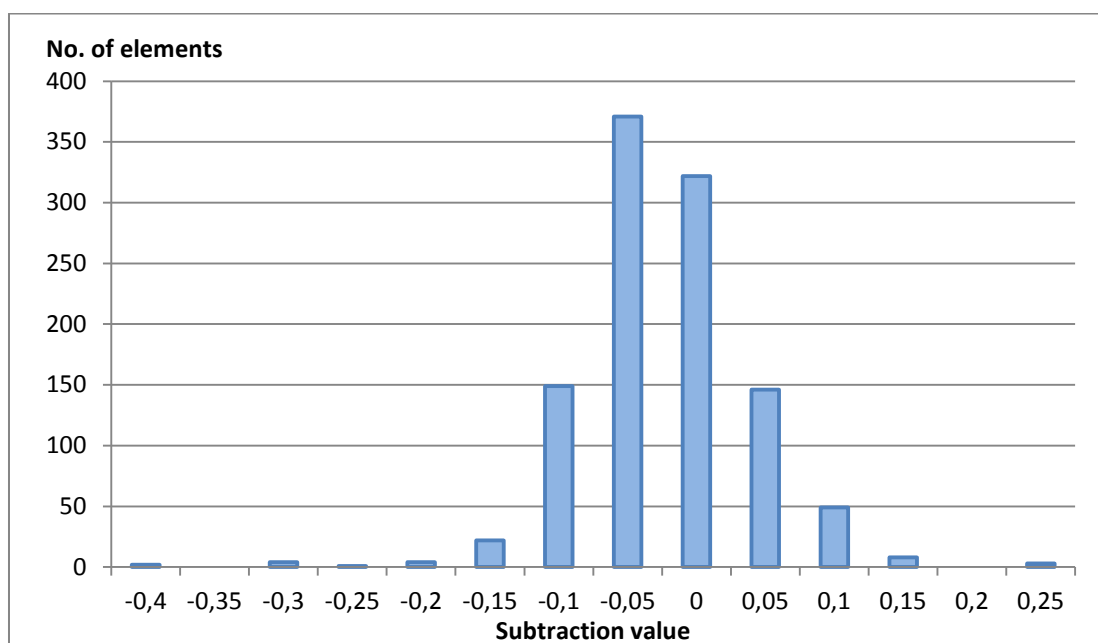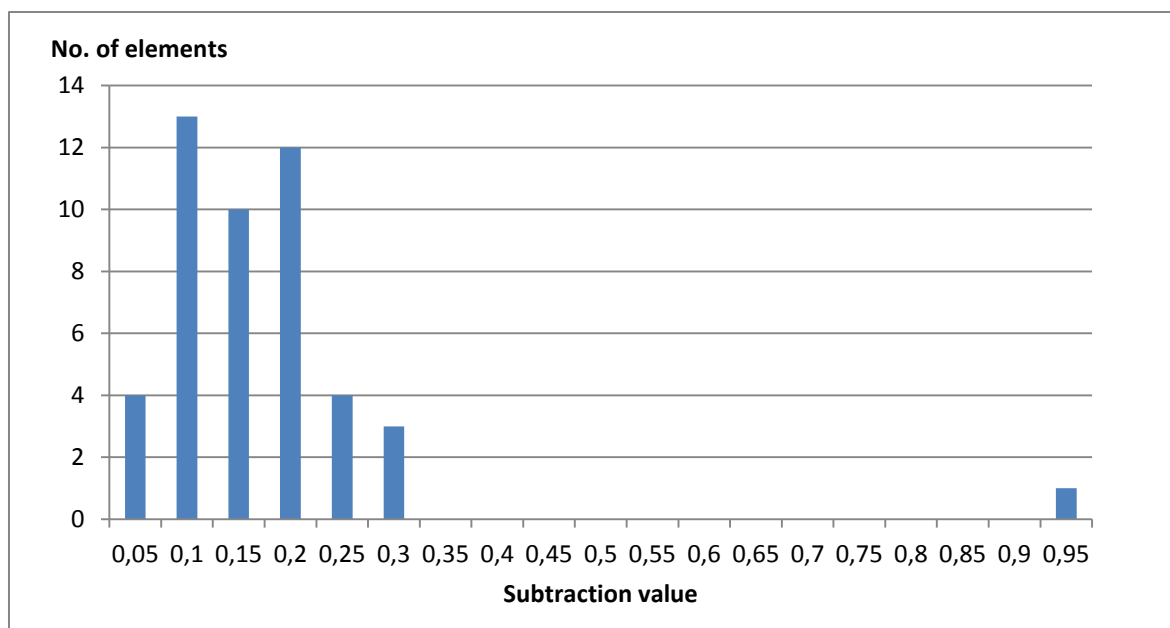

Figure 2a, b   Distribution of values after subtraction of **A** from **G** (filtering for GCS higher than 0.7, MAF 5%, $\bar{x}$I average of genotyped) – diagonal (a) and nondiagonal elements (b)
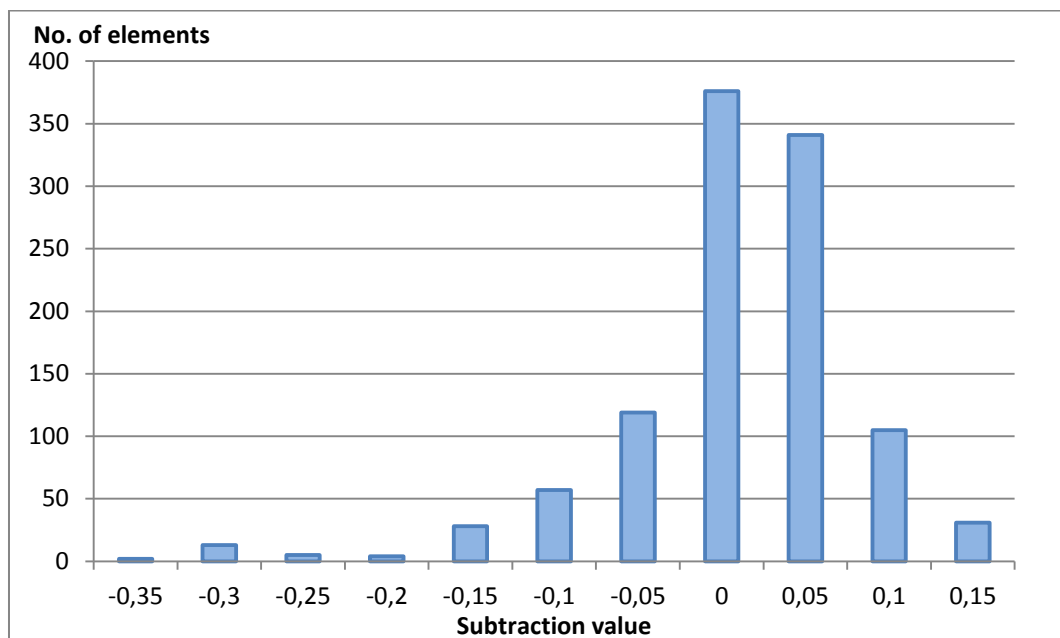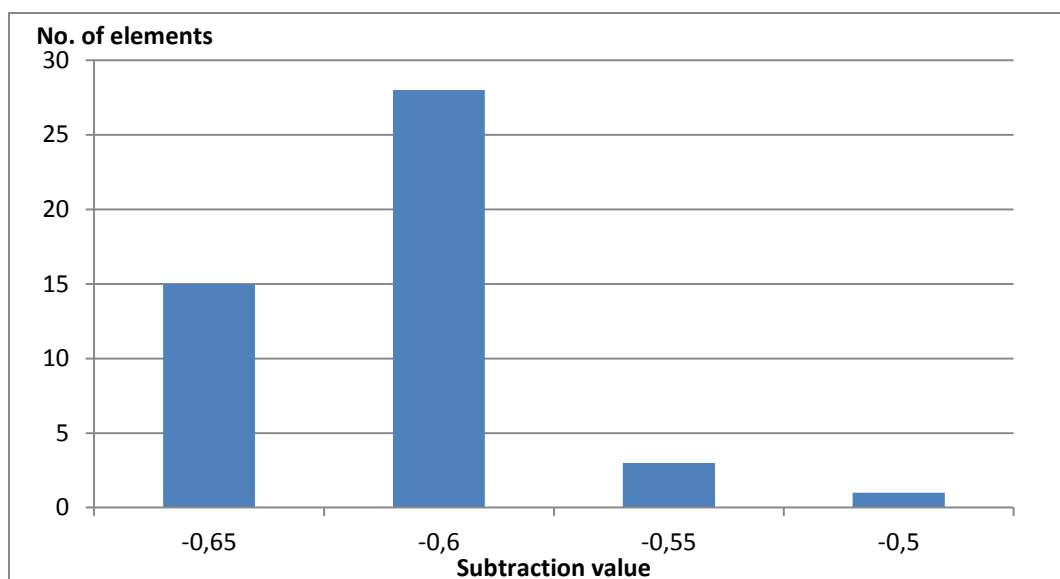
Figure 3a,b   Distribution of values after subtraction of **A** from **G** (filtering for GCS higher than 0.7, MAF 5%, $\bar{x}$l = homozygote A) – diagonal (a) and nondiagonal elements (b)

4   Conclusions

Genotyping by biochips and calculation of genomic matrix can be successfully implemented into instruments for more precise design of mating plan of Old Kladruber horses. Further genotyping and development of method for rescaling of differences between genomic and pedigree relationship matrices´ elements for better interpretation of results by breeders is advised.

Acknowledgments

References

BOYKO, A. R. et al. (2014) Genomic analysis establishes correlation between growth and laryngeal neuropathy in Thoroughbreds. *BMC genomics*, vol. 15, no. 1, pp. 259. doi: http://dx.doi.org/10.1186/1471-2164-15-259

FORNI S., AGUILAR I. and MISZTAL I. (2011) Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genetics Selection Evolution*, vol. 43, no. 1. doi: http://dx.doi.org/10.1186/1297-9686-43-1

GAUTIER, M., LALOË D. and MOAZAMI-GOUDARZI K. (2010) Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS One*, vol. 5, no. 9, e13038. doi: http://dx.doi.org/10.1371/annotation/ee4e4e98-80fc-4724-a96b-891cf652b9e1

HABERLAND, A. M. et al. (2012) Integration of genomic information into sport horse breeding programs for optimization of accuracy of selection. *Animal*, vol. 6, no.9, pp. 1369–1376. doi: http://dx.doi.org/10.1017/S1751731112000626

MISZTAL, I., LEGARRA A. and AGUILAR I. (2009) Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science*, vol. 92, no. 9, pp.4648–4655. doi: http://dx.doi.org/10.3168/jds.2009-2064

MCKAY, S. D. et al. (2008) An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. *BMC Genetics*, vol. 9, no.1, pp. 1. doi: http://dx.doi.org/10.1186/1471-2156-9-37

PETERSEN, J. L. et al. (2013) Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One*, vol. 8, no. 1, e54997. doi: http://dx.doi.org/10.1371/journal.pone.0054997

SCHUBERTOVÁ, Z., PAVLÍK O. and CANDRÁK J. (2014) The influence of the genomic data on the showjumping horses performance evaluation. *Journal of Central European Agriculture*, vol. 15, no. 2, pp. 33-40. doi: http://dx.doi.org/10.5513/JCEA01/15.2.1447

TIEZZI, F. et al. (2015) A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. *PloS One*, vol. 10, no. 2, e0114919. doi: http://dx.doi.org/10.1371/journal.pone.0114919

VANRADEN P.M. (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science*, vol. 91, no. 11, pp. 4414–4423. doi: http://dx.doi.org/10.3168/jds.2008-1514

VITEZICA et al. (2011) Bias in genomic predictions for populations under selection. *Genetics Research*, vol. 93, no. 5, pp. 357–366. doi: http://dx.doi.org/10.1017/S001667231100022X

VOSTRÝ L. et al. (2011) Intra-line and inter-line genetic diversity in sire lines of the Old Kladruber horse based on microsatellite analysis of DNA. *Czech Journal of Animal Science*, vol. 56, no. 4, pp. 163–175.

VOSTRÁ-VYDROVÁ H. et al. (2016) Pedigree analysis of the endangered Old Kladruber horse population. *Livestock Science*, vol. 185, pp. 17–23. http://dx.doi.org/10.1016/j.livsci.2016.01.001