

RUNS OF HOMOZYGOSITY AS FOOTPRINTS OF SELECTION IN THE NORIK OF MURAN HORSE GENOME

Nina Moravčíková¹, Radovan Kasarda¹, Ondrej Kadlečík¹, Anna Trakovická¹, Marko Halo², Juraj Candrák¹

¹ Department of Animal Genetics and Breeding Biology, Faculty of Agrobiology and Food Resources, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 76 Nitra, Slovakia

² Department of Animal Husbandry, Faculty of Agrobiology and Food Resources, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 76 Nitra, Slovakia

To link to this article: <https://doi.org/10.11118/actaun201967051165>

Received: 12. 8. 2019, Accepted: 24. 9. 2019

To cite this article: MORAVČÍKOVÁ NINA, KASARDA RADOVAN, KADLEČÍK ONDREJ, TRAKOVICKÁ ANNA, HALO MARKO, CANDRÁK JURAJ. 2019. Runs of Homozygosity as Footprints of Selection in the Norik of Muran Horse Genome. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 67(5): 1165–1170.

Abstract

The aim of this study was to analyse the genome-wide distribution of runs of homozygosity (ROH) segments in the genome of Norik of Muran horse and to identify the regions under strong selection pressure. Overall, 25 animals genotyped by the GGP Equine70k chip were included in the study. After SNP pruning, 54479 SNPs (75.72%) covering 2.25 Gb of the autosomal genome were retained for scan of ROH segments distribution. The ROHs were present in the genome of all animals and covered in average 13.17% (295.29 Mb) of autosomal genome expressed by the SNP loci. The highest number of ROHs was identified on autosome 1 (404), while the lowest proportion of autosome residing in ROH showed ECA31 (38). The footprints of selection, characterized by SNPs with extreme frequency in ROHs across specific genomic regions, were defined by the top 0.01 percentile of signals. Overall, nine genomic regions located on seven autosomes (3, 6, 9, 11, 15, 23) were identified. The strongest signal of selection showed three autosomes ECA3, ECA9 and ECA11. The protein-coding genes located within these regions suggested that the identified footprints of selection are most likely consequences of intensive breeding for traits of interest during the grading-up process of the Norik of Muran horse.

Keywords: autozygosity, horse, genomic data, local population, selection sweeps

INTRODUCTION

Modern horse breeds represents heterogeneous populations selected for specific appearance and performance traits. In general, genomic regions under strong selection pressure due to specific breeding for traits of interest display low genetic variability resulting in increase of proportion of continuous homozygous segments that are common in individuals of target population or breed. These long stretches of consecutive homozygous genotypes located in the genome are known as runs of homozygosity (ROH) (Metzger *et al.*, 2015). It was shown that ROH segments

give insight into a populations genetic events, demographic evolution of a population over time and genetic relatedness among individuals. The frequency, size and distribution of ROH segments in the genome depends on various factors, including recombination, linkage disequilibrium, population structure, mutation rate as well as natural and artificial selection (Peripolli *et al.*, 2017). The analysis of ROH segments distribution in the genome can be used to estimate the level of genomic inbreeding in population or to find genomic regions under strong selection pressure that are mostly described as selection footprints. Assuming that the frequency

of certain alleles increases due to positive selection the selection footprints can be valuable resource for mapping of causative mutations. Until now, the distribution of footprints of selection across genome have been reported for various species, including cattle (Kasarda *et al.*, 2015; Kukučková *et al.*, 2017; Gurgul *et al.*, 2019), goat (Brito *et al.*, 2017), sheep (Kim *et al.*, 2016), dog (Cagan and Blass, 2016), pig (Gurgul *et al.*, 2018) and horse (Nolte *et al.*, 2019).

The Norik of Muran is a unique draught horse breed named for part of the Gemer region in central southern Slovakia. The Norik of Muran belongs to the traditional horse breeds bred in Slovakia mainly due to its good character, constitution, excellent manoeuvrability, high cultural and historical value. It is used for various purposes, including sport riding, draught power as well as hippotherapy (Porter *et al.*, 2016; Halo *et al.*, 2018).

The history of the Norik of Muran breed in Slovakia started in the 1950s, when a stud farm in Velka Luka near Muran was founded. One of the reasons for establishing this stud farm was the need to produce horses for forestry and the military. During the time, the breeding goal of the Norik of Muran horse has changed several times, depending on the demands on the exterior and performance traits for work in forestry and agriculture. At the beginning, the Norik of Muran was developed by crossing Hutsul breed and local heavy horses with Norik and Belgian Draught horses (Fjord, Hafling). At the end of the 1960s, the coldblood mares from Bohemia and Moravia were imported in the Norik of Muran stud farms. In the second half of the 70th the stallions and mares of Silesian Noriker and Hungarian Coldblood horse have been implemented in the Norik of Muran population. Since the early 1980s, only purebred Norik of Muran stallions are used in the breeding. The selection strategy applied in breeding of this breed resulted in development of multi-purpose and hard-working leisure horse with an excellent character (Halo *et al.*, 2006). With respect to Noric of Muran horse only several pedigree-based studies describing the state of diversity have been published (Pjontek *et al.*, 2012; Halo *et al.*, 2018).

The objective of this study was to analyse the distribution of ROH segments in the genome of Norik of Muran horse and to describe the genomic regions under strong selection pressure.

MATERIALS AND METHODS

In this study overall 25 Norik of Muran horses genotyped by GGP Equine70k (71947 SNPs) chip were included. The SNP pruning was performed using PLINK v1.9 (Chang *et al.*, 2015) to exclude i) SNPs located on sex chromosomes or markers with unknown chromosomal position, ii) all of autosomal SNPs and animals with call rate lower than 90% and iii) SNPs with minor allele frequency lower than 1% in population.

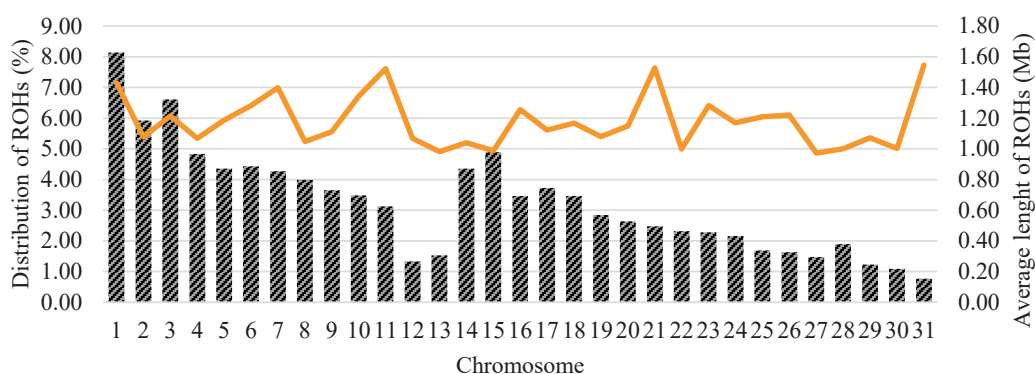
The distribution of ROH segments in the genome of Norik of Muran horse was tested by using PLINK v1.9 (Chang *et al.*, 2015) and R package detectRUNS (Biscarini *et al.*, 2018). The ROH segments were defined according to Nolte *et al.* (2019) as follows: minimum number of SNPs in a run 15, minimum length of run 500 kb, maximum distance between consecutive SNPs in a window 1000 kb, lower density limit 1SNP per 100kb and no missing or heterozygous SNPs in a run. The detection of selection footprints was based on the assumption that the identified runs of homozygosity across the genome of Norik of Muran horse are a results of selective breeding for traits of interest defined in its breeding objectives. The footprints of selection, characterized by SNPs with extreme frequency in ROHs across specific genomic regions, were determined by the calculation of runs incidence per each SNP marker. The genome-wide occurrence of SNPs in ROH was expressed as the frequency (%) of overlapping ROH shared among animals. Genomic regions under selection were defined by the top 0.01 percentile of signals. For functional analysis, genomic regions covering footprints of selection were scanned for annotated genes in the equine reference assembly EquCab2.0 by web based tool Biomart from Ensembl database (<https://www.ensembl.org/>).

RESULTS AND DISCUSSION

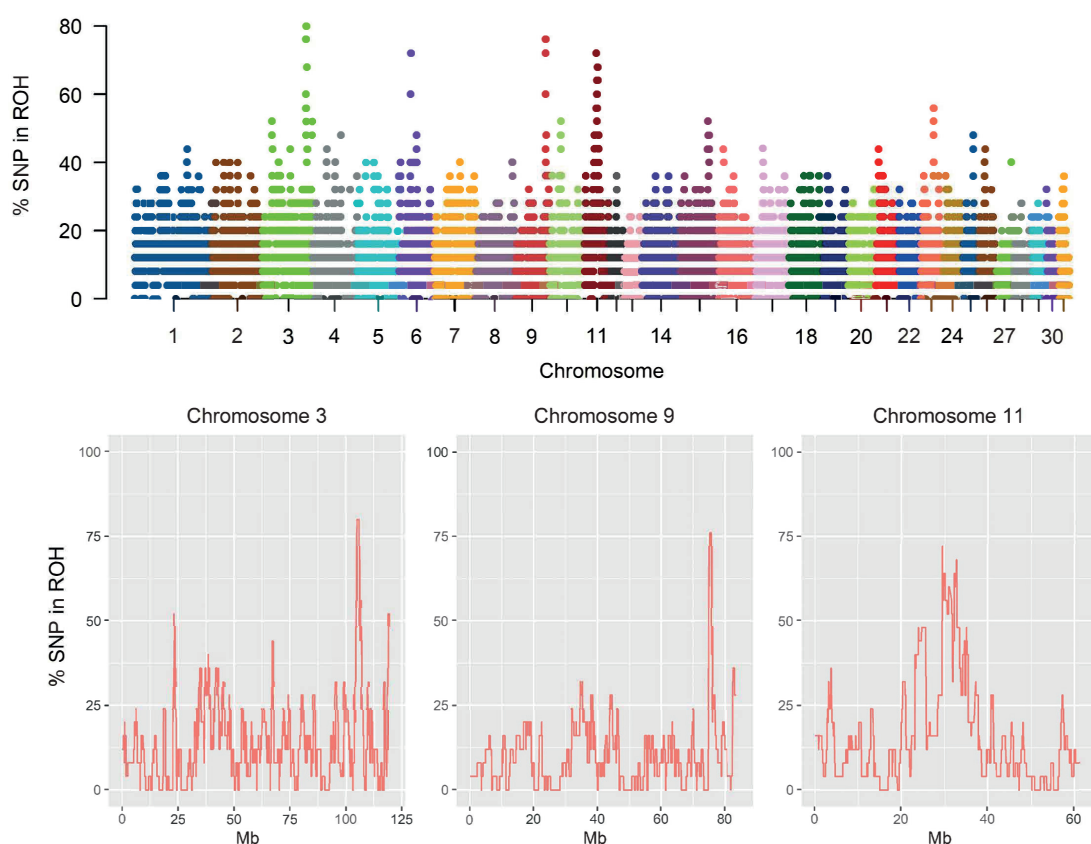
After quality control of genotyping data, 54479 SNPs (75.72%) covering 2.25 Gb of the autosomal genome were retained for scan of ROH segments distribution. The average distance between adjacent SNP loci in the final dataset was 41.35 ± 39.39 kb. All of genotyped individuals in analysis showed genotyping rate higher than 90% and the total genotyping rate across samples and SNPs at level 98.23% was in accordance to previous studies (Kamiński *et al.*, 2017; Nolte *et al.*, 2019).

The scan of ROHs distribution in the Norik of Muran horse genome showed that the ROH segments were present in the genome of all animals under consideration. The ROHs detected in this study covered 13.17% (295.29 Mb) of the autosomal genome expressed by the SNP loci. The total genome length covered by ROHs is comparable with study of Grilz-Seger *et al.* (2018) and Druml *et al.* (2018) that showed across various breeds the overall genome-wide ROH coverage at level 305.1 Mb. But, the study of Grilz-Seger *et al.* (2019a) revealed by use of 700k genotyping array that the genome ROH coverage in Noriker horse depending on the colour variety resulting from different mating strategies.

The distribution and length of ROH segments varied across autosomes (Fig. 1). The highest number of ROHs was identified on autosome 1 (404), while the lowest proportion of autosome residing in ROH showed ECA31 (38). The subsequent analysis of selection footprints in the genome of Norik



1: Distribution and average length of ROHs detected across autosomal genome



2: Genome-wide occurrence of SNPs in ROH across autosomal genome of Norik of Muran horse

of Muran horse resulting from the assumption that the extreme ROH frequencies are most likely consequences of intensive breeding for traits of interest during the grading-up process of the breed. The footprints of selection were defined based on the SNPs with extreme frequency in ROHs across specific genomic regions. Only top 0.1 percentile of signals were considered as reliable footprints of selection. According to this, nine genomic regions located within seven autosomes (3, 6, 9, 11, 15, 23) were identified (Fig. 2).

The shortest genomic region was detected on ECA3 (0.03 Mb), whereas the longest region was

located on ECA11 (2.25 Mb) (Tab. I). The strongest signal of selection showed three autosomes ECA3, ECA9 and ECA11 (Fig. 2). Similar results reported for Austrian Noriker horse Grilz-Seger *et al.* (2019a) that identified overall seven ROH islands shared by more than 50% of the entire analysed sample. The high frequency of ROH islands on ECA11 was also found in genome of Lipizan horse (Grilz-Seger *et al.*, 2019b).

Various protein-coding genes are located within the identified genomic regions under selection pressure (Tab. I). Within the regions on ECA3 the CALB2, KCNIP4, SLIT2 and MIR218-1 genes were

I: Genomic region under strong selection pressure in Norik of Muran horse

Region	Chr	Start position (Mb)	End position (Mb)	Region size (Mb)	Protein-coding genes
1	3	23.13	23.16	0.04	MIR9079, CALB2
2	3	104.58	106.52	1.93	KCNIP4, PACRGL, SLIT2, MIR218-1
3	6	28.77	30.02	1.26	CACNA1C, CECR2, ADA2, HDHD5, TMEM121B, IL17RA, CACNA1C, DCP1B, CACNA2D4, LRTM2, ADIPOR2
4	9	74.90	75.62	0.72	EFR3A, OC90, HHLA1, KCNQ3, LRRC6, TMEM71, PHF20L1
5	10	27.40	28.10	0.70	EQUCABV1R-POS919, EQUCABV1R-PS920, EQUCABV1R-PS921, ZNF671, C10H19orf18, ZNF606
6	11	29.56	31.81	2.25	ANKFN1, TOM1L1, STXBP4, HLF, MMD, TMEM100, NOG, DGKE, C11H17orf67, TRIM25, COIL, SCPEP1, AKAP1
7	11	32.29	32.83	0.54	MSI2, CCDC182, MRPS23, CUEDC1, VEZF1, SRSF1, DYNLL2, EPX, MKS1, LPO
8	15	66.22	67.11	0.89	EDH3, CAPN14, GALNT14, CAPN13, LCLAT1
9	23	33.93	34.15	0.23	NFIB

found. Blum *et al.* (2018) considered the CALB2 gene as the most sensitive and specific marker for the diagnosis of malignant mesothelioma in humans. The KCNIP4 gene, which play a role in the calcium ion binding, potassium and voltage-gated ion channel activity, was significantly associated with yearling weight in Canchim beef cattle (Xia *et al.*, 2017). The SLIK2 gene is a neural axon guidance and chemorepellent protein that stimulates motility in a variety of cell types. In mammals, the SLIT family has a role in the proper development of the central nervous system, and organs including the lung, kidney, and mammary gland (Saunders *et al.*, 2016). Cerro-Herreros *et al.* (2018) confirmed that the MIR218-1 is one of the factors which downregulates MBNL proteins in this cell line, responsible for the neuromuscular disease myotonic dystrophy type 1.

Overall seven protein-coding genes were found in a specific region on ECA9. One of this genes is the OC90 gene. In mammals and birds, the Otoconin 90 gene is a highly glycosylated protein that together with OTOL1, modulates calcite crystal morphology and growth kinetics (Wang *et al.*, 1998). Among others genes, e.g. the TMEM71 was associated with increasing grades of glioma in humans (Wang *et al.*, 2019).

Within the regions on ECA11 sequences of 23 genes were found. From these, the TMEM100 gene

was in mice associated with embryonic lethality due to impaired differentiation of arterial endothelium and defects of vascular morphogenesis (Somekawa *et al.*, 2012). The NOG gene has been suggested to be critical for normal bone and joint development. The human noggin protein encoded by NOG gene binds to bone morphogenetic protein (BMP) of the transforming growth factor- β superfamily and prevents its binding to the cognate receptor (Takano *et al.*, 2016). The TRIM25 gene encodes the E3 ubiquitin ligase enzyme that is involved in various cellular processes, including regulation of the innate immune response against viruses (Martín-Vicente *et al.*, 2017).

On the ECA15 directly in the region showing signal of selection, the CAPN13 and CAPN14 genes are located. Both of these genes are members of calpains family that are involved in a variety of regulatory processes, including cytoskeletal dynamics, cell-cycle progression, signal transduction, gene expression, and apoptosis. Both CAPN13 and CAPN14 are derived from a single classical calpain present in the jawed vertebrate ancestor. In domesticated animals the CAPN14 orthologues share 76% to 88% identity with the human protein; homology of human CAPN14 to other mammals and birds, reptiles, and frogs is 61% to 77% and 44% to 61%, respectively (Litosh *et al.*, 2017).

CONCLUSION

The analysis of selection footprints revealed totally nine genomic regions significantly affected by positive selection. The shortest genomic region was detected on ECA3, whereas the longest region was located on ECA11. The ROH segments shared by more than 70% of the entire analysed sample was found on ECA3, ECA9 and ECA11. Based on the subsequent functional analysis overall 59 protein-coding genes were identified. The biological role of those genes suggested that the identified footprints of selection are most likely consequences of intensive breeding for traits of interest during the grading-up process of the Norik of Muran horse. Assuming that the frequency of certain alleles

increases due to positive selection, the SNP markers located in regions strongly affected by selection can be valuable resource for future mapping of causative mutations important for adaptive and performance traits in horses.

Acknowledgements

This study was supported by the Slovak Research and Development Agency (APVV-14-0054 and APVV-17-0060) and VEGA (1/0742/17).

REFERENCES

- BISCARINI, F., COZZI, P., GASPA, G. et al. 2018. *detectRUNS: Detect runs of homozygosity and runs of heterozygosity in diploid genomes*. Available at: <http://orca.cf.ac.uk/108906/> [Accessed: 2019, August 15].
- BRITO, L. F., KIJAS, J. W., VENTURA, R. V. et al. 2017. Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genomics*, 18(1): 229.
- CAGAN, A. and BLASS, T. 2016. Identification of genomic variants putatively targeted by selection during dog domestication. *BMC Evol. Biol.*, 16: 10.
- CERRO-HERREROS, E., SABATER-ARCIS, M., FERNANDEZ-COSTA, J. M. et al. 2018. miR-23b and miR-218 silencing increase Muscleblind-like expression and alleviate myotonic dystrophy phenotypes in mammalian models. *Nature Communications*, 9: 2482.
- DRUML, T., NEUDITSCHKO, M., GRILZ-SEGER, G. et al. 2018. Population Networks Associated with Runs of Homozygosity Reveal New Insights into the Breeding History of the Haflinger Horse. *J Hered.*, 109(4): 384–392.
- GRILZ-SEGER, G., DRUML, T., NEUDITSCHKO, M. et al. 2019a. Analysis of ROH patterns in the Noriker horse breed reveals signatures of selection for coat color and body size. *Anim. Genet.*, 50(4): 334–346.
- GRILZ-SEGER, G., DRUML, T., NEUDITSCHKO, M. et al. 2019b. High-resolution population structure and runs of homozygosity reveal the genetic architecture of complex traits in the Lipizzan horse. *BMC Genomics*, 20: 174.
- GRILZ-SEGER, G., MESARIČ, M., COTMAN, M. et al. 2018. Runs of Homozygosity and Population History of Three Horse Breeds With Small Population Size. *J. Equine. Vet. Sci.*, 71: 27–34.
- GURGUL, A., JASIELCZUK, I., ROPKA-MOLIK, K. et al. 2018. A genome-wide detection of selection signatures in conserved and commercial pig breeds maintained in Poland. *BMC Genet.*, 19: 95.
- GURGUL, A., JASIELCZUK, I., SEMIK-GURGUL, E. et al. 2019. Diversifying selection signatures among divergently selected subpopulations of Polish Red cattle. *J. Appl. Genet.*, 60(1): 87–95.
- HALO, M., GRÁČZ, F., ŠMELKO, V. et al. 2006. *The herd book of Norik of Muran status*. [Online]. Available at: http://www.horses.sk/joomla/index.php?option=com_content&view=article&id=52:tatut-plemennej-knihy-norik-muransky&catid=41:norik-muransky&Itemid=18 [Accessed: 2019, May 15].
- HALO, M., MLYNEKOVÁ, E., HORNÁ, M. et al. 2018. Evaluation of Genetic Variability of the Breed Norik of Muran according to Pedigree Information. *Czech J. Anim. Sci.*, 63(5): 195–200.
- CHANG, C. C., CHOW C. C., TELLIER, L. C. A. M. et al. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, 4(1): 7.
- KAMIŃSKI, S., HERING, D. M., JAWORSKI, Z. et al. 2017. Assessment of genomic inbreeding in Polish Konik horses. *Pol. J. Vet. Sci.*, 20(3): 603–605.
- KASARDA, R., MORAVČÍKOVÁ, N., TRAKOVICKÁ, A. et al. 2015. Genome-wide selection signatures in Pinzgau cattle. *Potravinárstvo*, 9(1): 268–274.
- KIM, E. S., ELBELTAGY, A. R., ABOUL-NAGA, A. M. et al. 2016. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*, 116(3): 255–264.
- KUKUČKOVÁ, V., MORAVČÍKOVÁ, N., FERENČAKOVIČ, M. et al. 2017. Genomic characterization of Pinzgau cattle: genetic conservation and breeding perspectives. *Conserv. Genet.*, 18(4): 893–910.
- LITOSH, V. A., ROCHMAN, M., RYMER, J. K., et al. 2017. Calpain-14 and its association with eosinophilic esophagitis. *J. Allergy Clin. Immunol.*, 139(6): 1762–1771.
- MARTÍN-VICENTE, M., MEDRANO, L. M., RESINO, S. et al. 2017. TRIM25 in the Regulation of the Antiviral Innate Immunity. *Front. Immunol.*, 8: 1187.
- METZGER, J., KARWATH, M., TONDA, R., et al. 2015. Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC Genomics*, 16: 764.
- NOLTE, W., THALLER, G. and KUEHN, C. 2019. Selection signatures in four German warmblood horse breeds: Tracing breeding history in the modern sport horse. *PLoS One*, 14(4): e0215913.

- PERIPOLLI, E., MUNARI, D. P., SILVA, M. V. G. B. et al. 2017. Runs of homozygosity: current knowledge and applications in livestock. *Anim. Genet.*, 48(3): 255–271.
- PJONTEK, J., KADLEČÍK, O., KASARDA, R. et al. 2012. Pedigree analysis in four Slovak endangered horse breeds. *Czech J. Anim. Sci.*, 57(2): 54–64.
- PORTER, V., ALDERSON, L., HALL, S. J. G. et al. 2016. *Mason's World Encyclopedia of Livestock Breeds and Breeding, 2 Volume Pack*. Oxfordshire: CABI.
- SAUNDERS, J., WISIDAGAMA, D. R., MORFORD, T. et al. 2016. Maximal Expression of the Evolutionarily Conserved Slit2 Gene Promoter Requires Sp1. *Cell. Mol. Neurobiol.*, 36(6): 955–964.
- SOMEKAWA, S., IMAGAWA, K., HAYASHI, H. et al. 2012. Tmem100, an ALK1 receptor signaling-dependent gene essential for arterial endothelium differentiation and vascular morphogenesis. *PNAS*, 109(30): 12064–12069.
- TAKANO, K., OGASAWARA, N., MATSUNAGA, T. et al. 2016. A novel nonsense mutation in the NOG gene causes familial NOG-related symphalangism spectrum disorder. *Human Genome Variation*, 3: 16023.
- WANG, K. Y., HUANG, R. Y., TONG, X. Z. et al. 2019. Molecular and clinical characterization of TMEM71 expression at the transcriptional level in glioma. *CNS Neurosci. Ther.*, 25(9): 965–975.
- WANG, Y., KOWALSKI, P. E., THALMANN, I. et al. 1998. Otoconin-90, the mammalian otoconial matrix protein, contains two domains of homology to secretory phospholipase A2. *Proc. Natl. Acad. Sci. USA*, 95(26): 15345–15350.
- XIA, J., FAN, H., CHANG, T. et al. 2017. Searching for new loci and candidate genes for economically important traits through gene-based association analysis of Simmental cattle. *Sci. Rep.*, 7: 42048.

Contact information

Nina Moravčíková: nina.moravcikova1@gmail.com
Radovan Kasarda: radovan.kasarda@uniag.sk
Ondrej Kadlečík: ondrej.kadlecik@uniag.sk
Anna Trakovická: anna.trakovicka@uniag.sk
Marko Halo: marko.halo@uniag.sk
Juraj Candrák: juraj.candrak@uniag.sk