Selection sweeps through variation in linkage disequilibrium in horses

N. Moravčíková*, R. Kasarda, M. Halo, J. Candrák

Slovak University of Agriculture in Nitra
Faculty of Agrobiology and Food Resources
Department of Animal Genetics and Breeding Biology
Department of Animal Husbandry

Introduction

- **Selection sweeps**
  - ✓ genetic changes that shaped the genome due to natural as well as artificial selection

  - ✓ insight into the genetic adaptation to specific environments and history of selection for specific traits of interest

  - ✓ domestication-related loci, immunity-related loci, loci associated with desired phenotypic traits...

- Why did we choose **Norik of Muran** and **Lipizzan horses**?

fppt.com

Objectives

Detection of selection sweeps in the genomes of Norik of Muran and Lipizzan horses bred in Slovakia

Identification of genomic regions subjected to selection histories of those local populations

Functional analysis through identification of annotated genes located directly in genomic regions covering signals of positive selection
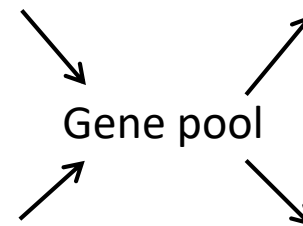
Material & Methods

Analysed population (overall 71 animals)

Norik of Muran → 25

Gene pool

Lipizzan → 36

Genotyping of animals:

GGP Equine70k chip (71,947 SNPs)

fppt.com

Material & Methods

Quality control of genotyping data

PLINK 1.9

- autosomal SNPs → 68,214 SNPs

- quality control criteria:
    missing genotypes per sample max. 10 %
    min. SNPs call rate 90 %

- final database → **62,730 SNPs** covering 2.25 Gbp of the autosomal genome

fppt.com

Material & Methods

Analysis of selection signatures

Integrated Haplotype Score (iHS) statistic

Variation in linkage disequilibrium (LD)

Material & Methods

Integrated Haplotype Score (iHS) statistic

- haplotype reconstruction according to Scheet and Stephens (2006) → fastPHASE

- iHS score calculation
  - ➢ according to Voight et al. (2006) for each SNPs → R package *rehh* (Gautier and Vitalis, 2012)

  - ➢ iHS score averaged into non-overlapping 500 kb segments across the genome
- genomic region under selection → 0.1 percentile of signals

Scheet, P., Stephens, M. 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. American Journal of Human Genetics, 78, 629-644.
Voight, B. F., Kudaravalli, S., Wen, X., Pritchard, J. K. 2006. A map of recent positive selection in the human genome. PLoS Biology, 4, e72.
Gautier, M., Vitalis, R. 2012. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*, vol. 28, p. 1176-1177.

Material & Methods

Variation in linkage disequilibrium (LD)

- VarLD (Ong and Teo, 2010)

  ➢ differences in genome-wide LD patterns between populations over sliding windows of 50 SNPs

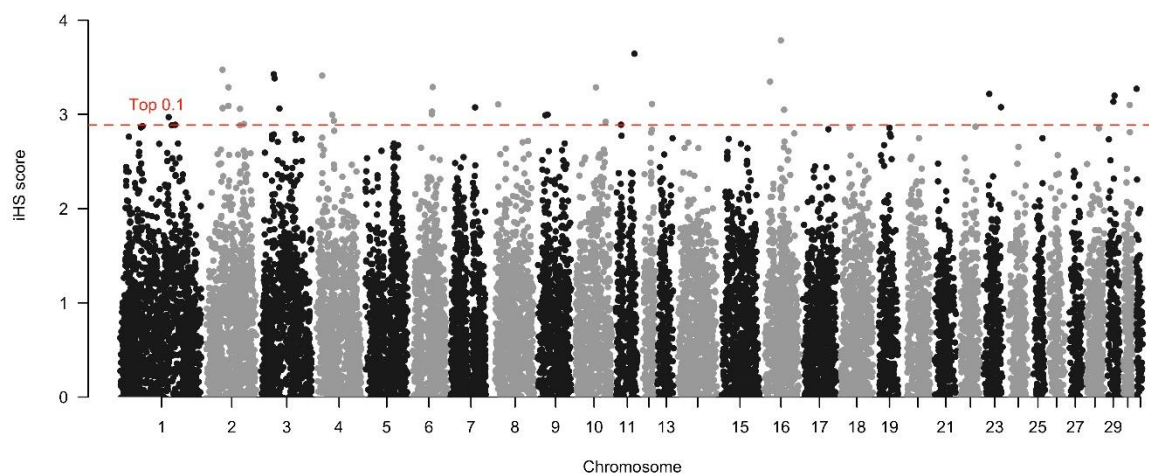 ➢ genomic region under selection → 0.1 percentile of signals

Ong, R. T., Teo, Y. Y. 2010. varLD: a program for quantifying variation in linkage disequilibrium patterns between populations. Bioinformatics, 26 (9), 1269-1270.

Results

iHS statistic - Norik of Muran horse

➢ 24 regions showing signal of selection

(16 autosomes)

    largest region ECA3

➢ 28 protein-coding genes

Genome-wide distribution of iHS score in Norik of Muran breed
(upper abline is showing the genome-wide significance threshold)

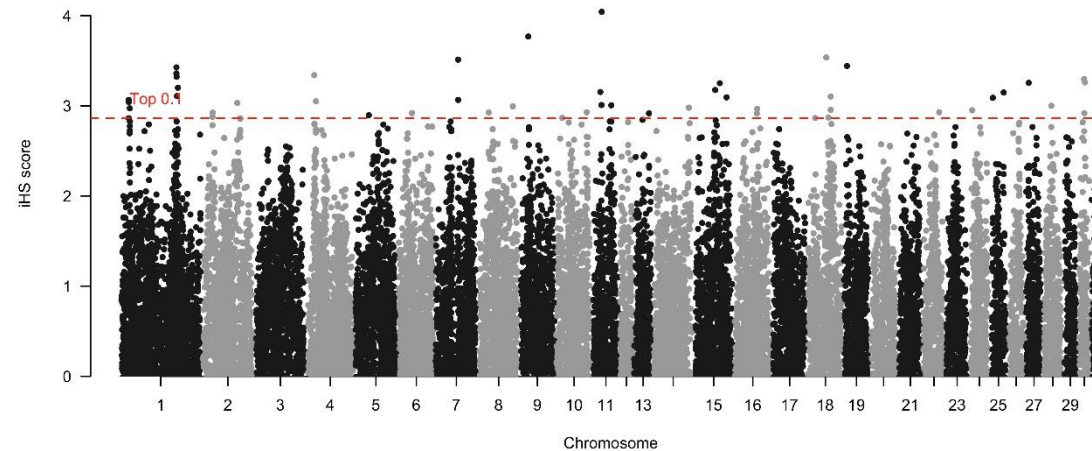| ECA	Start position (bp)	End position (bp)	Protein-coding genes
1	51598295	51668319	ARID5B
1	108470884	108564455	NSMCE3, FAM189A1
1	115410633	115643433	FAH, ZFAND6, BCL2A1
2	40497351	40585624	MTOR, ANGPTL7
2	53077765	53302391	ADAM28
2	78357465	78545882	TDO2
3	32982818	33085736	GSE1
3	43616288	43671949	-
4	17504389	17627063	TNS3
4	39476805	39598486	SLC25A13
6	49134640	49160929	-
7	58359803	58363847	-
8	10266286	10311520	TTC28
9	19271883	19376228	ADHFE1
10	45949629	46074234	LOC111775320
10	67343978	67390785	SLC35F1
11	46184501	46213848	SRR, TSR1, SGSM2
12	23226819	23300790	MRPL16,STX3
16	19839423	19862920	-
16	43395541	43500192	ZNF197, ZKSCAN7, ZNF660, ZNF 852
23	15921848	16031193	-
29	17087244	17185444	MALRD1
30	19203699	19237606	-
31	3942988	4062482	PDE10A

Results

iHS statistic - Lipizzan

➢ 26 regions showing signal of selection

(19 autosomes)

   largest region ECA11

➢ 41 protein-coding genes

Genome-wide distribution of iHS score in Lipizzan breed
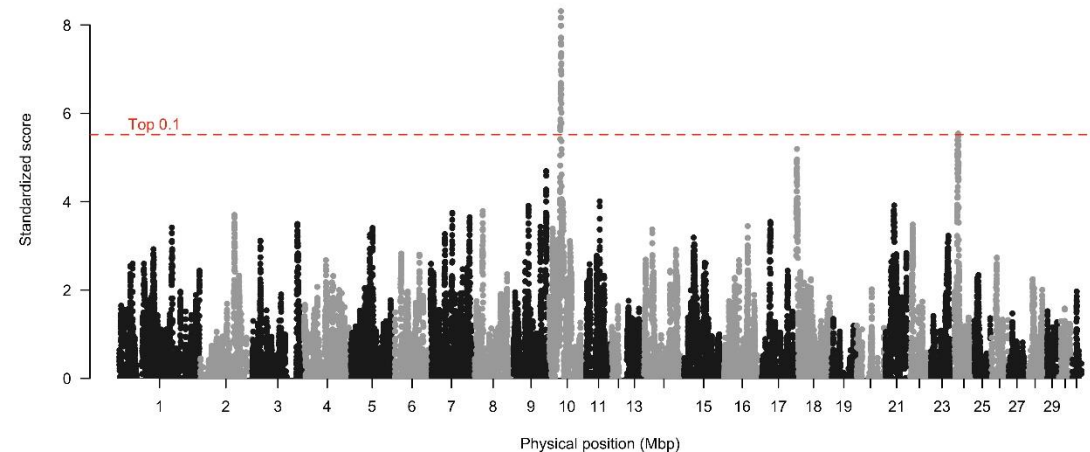(upper abline is showing the genome-wide significance threshold)

ECA	Start position (bp)	End position (bp)	Protein-coding genes
1	19994949	19999539	-
1	128762946	128838395	OAZ2, RBMS2, ZNF609
2	25818018	25893580	-
2	82025402	82035073	GATB
4	17502281	17555905	TSN3
5	33553127	33580262	VANGL2
6	32497593	32553297	CRACR2A
7	53032339	53146886	LOC106783417
8	24639666	24686529	RHOF, SETD1B, HPD
8	79554051	79564639	ZNF532
9	20807398	20883170	LOC106780965
11	20362916	20517990	WNK4, VPS25, RAMP2, EZH1, CCR10, TUBG2, RETREG3
11	20905582	21055397	CNP, TTC25, KLHL10, KLHL11, ACLY
14	83002277	83064180	EDIL3
15	49349209	49430214	-
15	59962832	59972358	SOS1
15	75360373	75456847	TDRD15
16	53663824	53781295	OSBPL10
18	43281898	43317552	LOC100061409
19	8243335	8246579	-
24	6975588	6979483	CCDC175
25	7211035	7273342	-
25	31814627	31886991	ENG, AK1, ST6GALNAC6, ST6GALNAC4
27	7772707	7862277	BRF2, ADGRA2, PLPBP
28	19451358	19455260	EEA1
30	14448031	14552810	TGFB2, RRP15

Results

Variation in linkage disequilibrium (LD)
Norik of Muran vs. Lipizzan

ECA10: 26,232,615 - 28,980,724 bp

ECA24: 10,013,252 – 10,551,084 bp

Differences in genome-wide LD patterns between populations
(upper abline is showing the genome-wide significance threshold)

fppt.com

Results

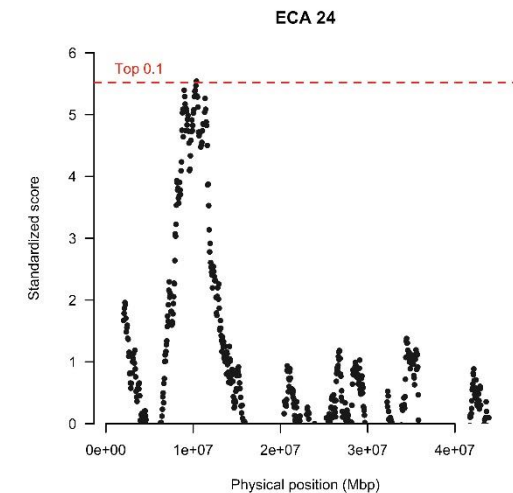## Variation in linkage disequilibrium (LD) Norik of Muran vs. Lipizzan
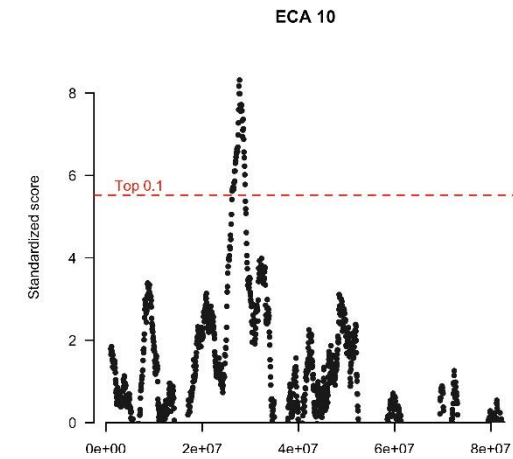
ECA10
→ 122 protein-coding genes

PEG3 – regulates skeletal muscle growth
AURKC – fertility
genes encoded zink finger proteins
…

ECA24
→ 8 protein-coding genes

RHOJ, GPHB5, PPP2R5E, WDR89, …

fppt.com

Conclusion

- various protein-coding genes identified within detected selection sweeps

- signals were strong mainly in genomic regions containing genes responsible for immunity related traits (e. g. toll like receptors), growth and muscle mass, body weight and reproduction

Conclusion

- applied approach → alternative strategy to identify the genomic regions and genetic markers related to traits with biological importance

- identification of genomic regions affected by positive and/or natural selection improves our understanding of the connection between changes in genome structure as result of interaction with local environment in which animal live

Thank you for your attention!

Ing. Nina Moravčíková, PhD.
Slovak University of Agriculture in Nitra
Department of Animal Genetics and Breeding Biology
Tr. A. Hlinku 2, 949 76 Nitra, Slovakia
nina.moravcikova1@mail.com