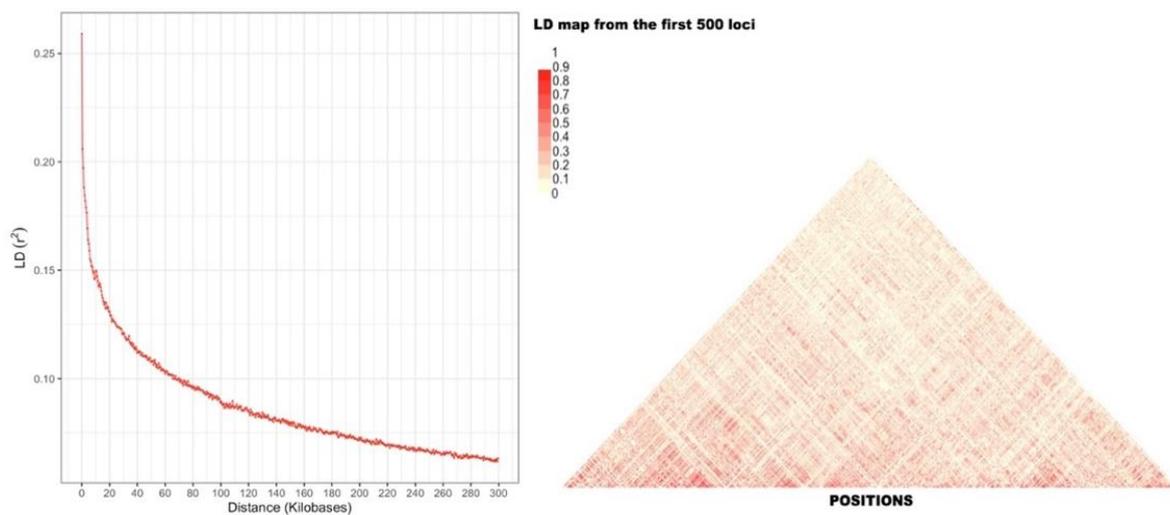


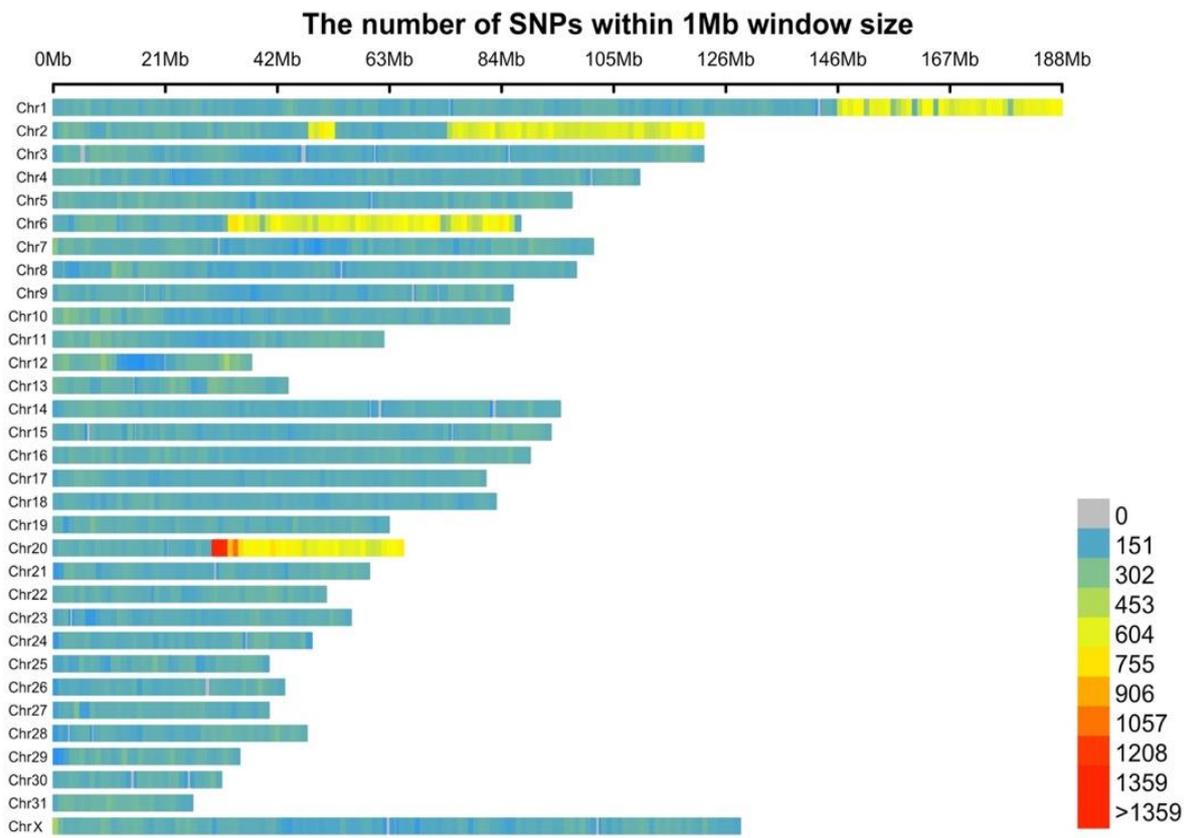
# Genome-wide scans for signatures of selection in Mangalarga Marchador horses using high-throughput SNP genotyping

Wellington B. Santos<sup>1\*</sup>, Gustavo P. Schettini<sup>1</sup>, Amanda M. Maiorano<sup>1</sup>; Fernando O. Bussiman<sup>2</sup>, Júlio C. C. Balieiro<sup>3</sup>, Guilherme C. Ferraz<sup>1</sup>, Guilherme L. Pereira<sup>4</sup>, Welder Angelo Baldassini<sup>4</sup>; Otávio R. M. Neto<sup>4</sup>, Henrique N. Oliveira<sup>1</sup> & Rogério A. Curi<sup>4</sup>

## Supplementary material: Additional file 1



**Fig. S1** Genome-wide linkage disequilibrium (LD) decay plot for 192 Mangalarga Marchador based on 347,935 SNP markers

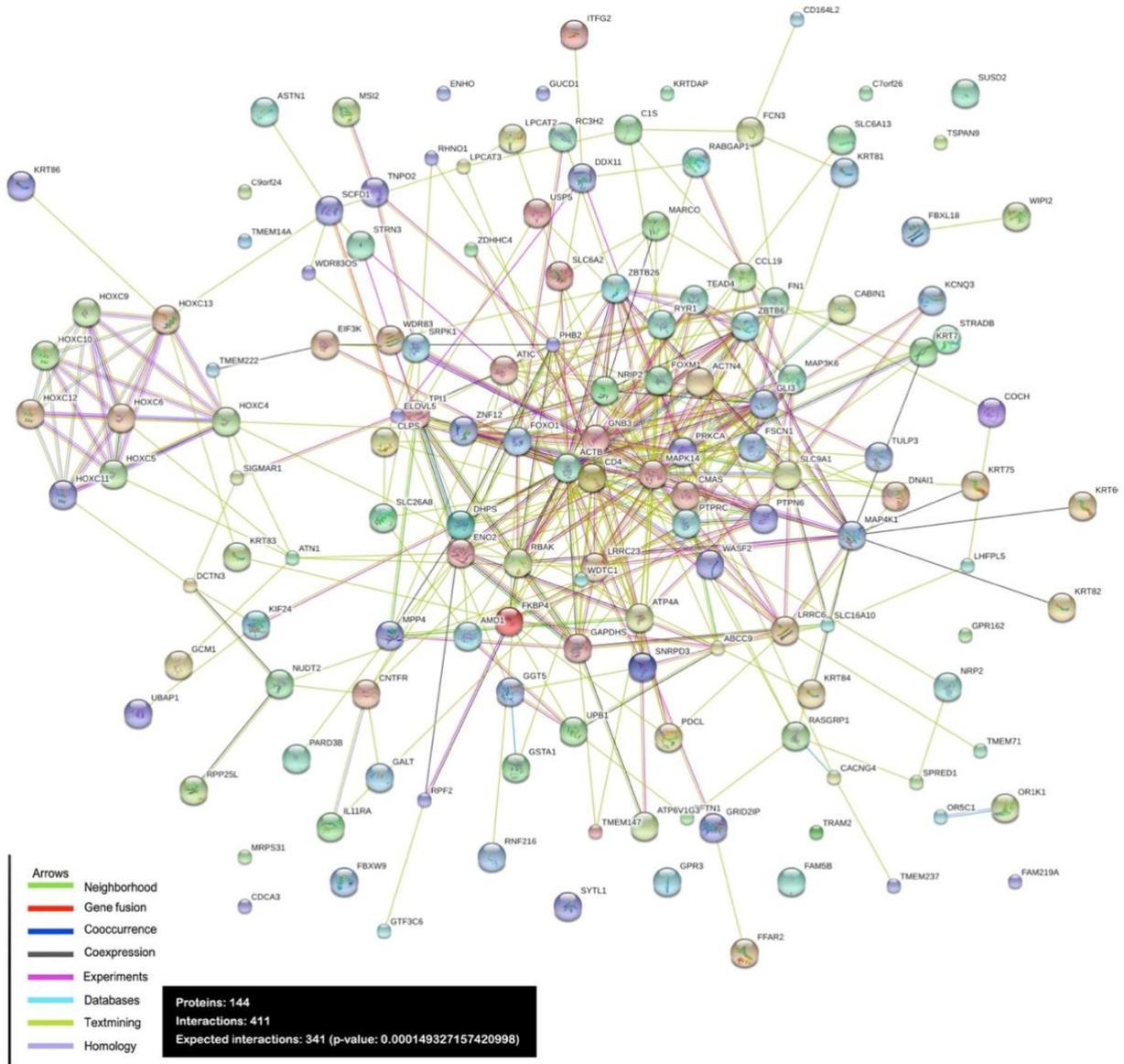


**Fig. S2** Final density of 545,219 SNP in the Mangalarga Marchador horse genome after Axiom™ Analysis Suite pruning

## Gene network analysis

Gene networks can shed light on the complex behavior of horse genes related to gait patterns, diseases, performance, and physiology. This analysis makes it possible to study functional interactions among proteins encoded by a list of genes. We used this analysis to add meaning to the list of genes related to signatures of selection and to investigate whether these candidate genes have some relationship with the *DMRT3* gene.

The networks were constructed considering candidate genes resulting from the Tajima's D, iHS, and ROH methods, and also by the biological pathways known in the equine sector in a wider sense. Each gene annotated in the previous stage of gene enrichment was used, except for iHS, which, due to the high number of signals, only contemplated the pruned genes according to the three criteria established in the iHS results section. The interactions were calculated by GeneMANIA [1] and STRINGdb R package [2]. Direct (physical) and indirect (functional) associations between genes were included in the interactions [3]. We merged the 169 candidate genes (TD=27, iHS=104, and ROH=38 genes) to conduct the analysis (known and predicted protein-protein interactions). The STRING identifiers could not map 16 genes, and nine genes were common between the methods, remaining 144 genes for gene network analysis. Interconnections were observed between most of the annotated genes (Additional file 1: Fig. S3).



**Fig. S3** Interaction networks of candidate genes identified from signatures of selection. Different colored arrows indicate the types of evidence used in predicting the associations.

The image shows the associations are separated into two clusters of genes. Basically, the HOX family genes were involved in the first cluster. The second cluster included the majority of the annotated genes which activate many pathways. On the black box in the plot, a p-value was inserted that represents the probability of the

expectation of an equal or greater number of interactions by chance. This p-value is practically null, meaning that the interactions were not caused by chance. An additional network analysis included the DMRT3 gene in the database, but did not identify clusters around this gene, and was only finding a few associations of low co-expression.

### **Supplementary References**

1. Franz M, Rodriguez H, Lopes C, Zuberi K, Montojo J, Bader GD, Morris Q. GeneMANIA update 2018. *Nucleic Acids Res.* 2018;46:W60–W64.

2. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 2013;41:808–15.

3. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic acids research.* 2015;43:D447–D452.