2023

Whole Genome Sequencing for Identification of Putative Causative Alleles for Hemochromatosis in Horses

Anna M. Rauber-Ramos

Whole Genome Sequencing for Identification of Putative Causative Alleles for Hemochromatosis in Horses

Anna M. Rauber-Ramos

Summary

Hemochromatosis is liver disorder in which excess iron is stored in the liver, leading to iron overload and extensive liver damage that can prove fatal. Based on human precedence, a genetic cause for idiopathic hemochromatosis in horses is suspected. Two horses with hemochromatosis were whole genome sequenced, with variants in candidate genes prioritized in analyses of potential pathogenicity. No clear putative variants were found to be present in both horses, suggesting an etiology more complex than the initially suspected Mendelian inheritance. A heterozygous nonsense mutation in *STEAP3 metalloreductase* (*STEAP3*) was identified as a potential contributor to hemochromatosis in one of the horses.

Introduction

Hemochromatosis in horses is a disorder caused by the deposition of hemosiderin in the functional hepatocytes of the liver, resulting in tissue damage and dysfunction that can ultimately lead to liver failure and death (Smith, 2009). Clinical symptoms include weight loss, lethargy, and anorexia, and no functional treatment has yet been identified (Foreman, 2022). In humans, the most common causes of hemochromatosis are genetic (Di Bisceglie, n.d.), which, along with cases of hemochromatosis in horses not being associated with excessive dietary iron intake (Pearson, et al., 1994), suggests a potential genetic etiology in horses as well. Hemochromatosis is rare in horses and may be underdiagnosed, resulting in a limited pool of diagnosed horses that can be used to study and thus making identification of causative alleles difficult. Whole genome sequencing has proven to be a powerful tool for identifying putative causative alleles, even for rare alleles with as few as one diseased individual available for study (Michot, et al., 2015). Identification of rare disease-causing alleles by using next-generation genome sequencing and a species reference genome has been used in human research for thirteen years (Boycott, et al., 2013), and knowledge of the genetics of disease etiology in humans has been used to help identify candidate genes and causative alleles in livestock species (Peters, et al., 2015).

The equine genome has been sequenced and high-throughput genotyping arrays, containing information on variation in single nucleotide polymorphisms (SNPs), have been developed of increasingly high density, facilitating genotyping and research, with the MNEc670k containing over 670,000 SNPs (Schaefer & McCue, 2020). A large-scale catalog of genetic variation in horses has been developed, based upon 534 individuals and providing a significantly improved understanding of background genetic diversity in the horse that vastly increases the ability to identify disease-causing alleles (Durward-Akhurst, et al., 2021). These types of genome wide association studies, using older, less dense high-throughput genotyping arrays, have already been used to identify causative alleles for other genetic diseases in horses, such as dwarfism (Orr, et al., 2010) and Lavender Foal Syndrome (Brooks, et al., 2010).

Material and Methods

Horses and Genomes

Two individual horses with a diagnosis of hemochromatosis were paired-end whole genome sequenced using Illumina technology to a coverage depth of ~20x. A large-scale catalog of horse genetic variants based on 534 horses from 46 breeds was used as a comparison (Durward-Akhurst, et al., 2021).

Mapping and Variant Calling

Fastqc 0.11.8 and Trimmomatic 0.38 were used for standard fastq quality control and trimming. The Burrows-Wheeler Aligner (BWA) (Li & Durbin, 2009) was used to align reads to the EquCab 3.0 reference horse assembly (Kalbfleish, et al., 2018), with modification for joint variant calling by GATK haplotype caller (McKenna, et al., 2010) and BCFtools mpileup (Li, et al., 2009). GATK version 4.1.0.0 was also used for indel realignment and base-quality score

recalibration (DePristo, et al., 2011), using a modified version of Genome Analysis Toolkit best practices (Van der Auwera, et al., 2018). Picard tools (https://broadinstitute.github.io/picard/) version 2.18.27 was used to detect and remove PCR duplicates. From these, a combined VCF was produced, as well as individual VCFs for each horse.

Prioritization of Candidate Genes

Genes known to cause hemochromatosis were found through Online Mendelian Inheritance in Animals (Sydney School of Veterinary Science, n.d.) (Lenffer, et al., 2006) and Online Mendelian Inheritance in Man (McKusick-Nathans Institute of Genetic Medicine, n.d.) (Amberger, et al., 2015). Horse genes with ontologies related to iron metabolism and transport were collected from AmiGO 2 (Carbon & Mungall, 2018).

Variant Annotation and Analysis

Python was used to filter for variants that had allele frequencies of 0.2 or less and were within 100 base pairs of each candidate gene. The allele frequency was set high to account for likely underdiagnosis of hemochromatosis in horses. Variants below the allele frequency threshold in all other genes were also analyzed. Consurf (Yariv, et al., 2023) was used to assess the conservation of residues across species, and Mutpred2 (Pejaver, et al., 2020), WS-SNPs&GO (Capriotti, et al., 2013), and PredictSNP (Bendl, et al., 2014) were used to estimate pathogenicity. For one of the horses, the PROVEAN web server (Choi & Chan, 2015) was also used to analyze variants. The server was retired before it could be used to analyze the variants of the other horse. Splice region variants and synonymous variants were not analyzed due to a lack of readily available analysis tools.

Results

Below the allele frequency threshold, no variants in the candidate genes were found in the homozygous state in both horses. Two gene variants were present in both horses in a heterozygous state. None of the algorithms identified either of these variants as potentially deleterious, and they were excluded as causes.

A nonsense variant in *STEAP3 metalloreductase* (*STEAP3*) was identified as heterozygous in one horse with a MAF of 5.3%. The variant's PROVEAN score was -1055, with scores below -2.5 being considered deleterious. *STEAP3* is an essential component of iron recycling of erythrocytes. Senescent erythrocytes are phagocytosed into endosomes by macrophages. Fe^{3+} is reduced to Fe^{2+} by STEAP3, allowing for transport across the endosome membrane (Sendamarai, et al., 2008). A similar mutation in humans causes hypochromic microcytic anemia with iron overload (OMIM: 615234) and is inherited as a dominant trait with partial penetrance (Grandchamp, et al., 2011). A knockout study in mice demonstrated increased liver iron levels with non-functional *STEAP3* (Zhang, et al., 2012). In the case described by Grandchamp, et al., the level of mRNA expression of the alleles varied, accounting for the incomplete penetrance. The associated anemia and inability to remove iron from the macrophages aligns with the normal serum iron levels and iron accumulation in macrophages found in some cases of idiopathic

hemochromatosis in horses (Pearson, et al., 1994). Hematology results were not available for the two genotyped horses in this study.

Among all gene variants below the allele frequency threshold, three were found to be homozygous in both horses. None were associated with liver or iron metabolism function nor identified as potentially deleterious by a notable number of algorithms. No gene variants that were homozygous in one horse and heterozygous in the other or heterozygous in both horses were found that were likely to cause iron metabolism disorders.

Discussion

No obvious Mendelian cause for hemochromatosis in horses was found, suggesting a more complex genetic etiology. Heterogenous causes are possible as a *STEAP3* nonsense mutation in one of the two horses analyzed in this study is promising as a potential cause. Hemochromatosis may also be polygenic and associated with gene-gene and gene-environment interactions. Future genome sequencing of additional horses and more research into equine iron metabolism will be needed to further elucidate genetic factors related to the development of hemochromatosis in horses.

Limitations include the paucity of research into equine iron metabolism resulting in the use of human iron metabolism information for the determination and prioritization of candidate genes. Several types of genetic variants were also not analyzed due to lack of access to analysis tools, although no rare variants were homozygous in both horses.

Acknowledgements

I would like to thank Dr. Christine Larson, Dr. Michael Gonda, Dr. Sian Durward-Akhurst, Dr. Molly McCue, and Jillian Marlowe for their invaluable aid in this project.

References

Amberger, J. S. et al., 2015. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Research,* 43(D1), pp. D789-D798.

Bendl, J. et al., 2014. PredictSNP: Robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Computational Biology,* 10(1), p. e1003440.

Boycott, K. M., Vanstone, M. R., Bulman, D. E. & MacKensie, A. E., 2013. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nature Reviews: Genetics,* 14(10), pp. 681-691.

Brooks, S. A. et al., 2010. Whole-Genome SNP Association in the Horse: Identification of a Deletion in Myosin Va Responsible for Lavender Foal Syndrome. *PLoS Genetics,* 6(4).

Capriotti, E. et al., 2013. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomis,* 14(Suppl 3), p. S6.

Carbon, S. & Mungall, C., 2018. *Gene Ontology Data Archive.* s.l.:Zenodo.

Choi, Y. & Chan, A. P., 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics,* 31(16), pp. 2745-2747.

DePristo, M. A. et al., 2011. 10.1093/bioinformatics/btp324. *Nature Genetics,* 43(5), pp. 491-498.

Di Bisceglie, A. M., n.d. *Hemochromatosis.* [Online]
Available at: https://www.niddk/nih.gov/health-information/live-disease/hemochromatosis
[Accessed 14 May` 2022].

Durward-Akhurst, S. A. et al., 2021. Genetic variation and the distribution of variant types in the horse. *Frontiers in Genetics,* Volume 12, p. 758366.

Foreman, J. H., 2022. *Miscellaneous Hepatic Disorders in Large Animals.* [Online]
Available at: https://www.merckvetmanual.com/digestive-system/hepatic-disease-in-large-animals/miscellaneous-hepatic-disorders-in-large-animals
[Accessed 22 January 2023].

Grandchamp, B. et al., 2011. A novel type of congenital hypochromic anemia associated with a nonsense mutation in the STEAP3/TSAP6 gene. *Blood,* 118(25), pp. 6660-6666.

Kalbfleish, T. S. et al., 2018. EquCab3, an updated reference genome for the domestic horse. *bioRxiv.*

Lenffer, J. et al., 2006. OMIA (Online Mendelian Inheritance in Animals): an enhanced platform and integration into the Entrez search interface at NCBI. *Nucelic Acids Research,* 34(suppl_1), pp. D599-D601.

Li, H. & Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25(14), pp. 1754-1760.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25(16), pp. 2078-2079.

McKenna, A. et al., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research,* 20(9), pp. 1297-1303.

McKusick-Nathans Institute of Genetic Medicine, J. H. U. (. M., n.d. [Online]
Available at: https://omim.org
[Accessed 2 February 2023].

Michot, P. et al., 2015. Whole-genome sequencing identifies a homozygous deletion encompassing exons 17 to 23 of the integrin beta 4 gene in a Charolais calf with junctional epidermolysis bullosa. *Genetics Selection Evolution,* 47(1), p. 37.

Orr, N. et al., 2010. Genome-wide SNP association–based localization of a dwarfismgene in Friesian dwarf horses. *Animal Genetics,* Volume 41, pp. 2-7.

Pearson, E. G., Hedstrom, O. R. & Poppenga, R. H., 1994. Hepatic cirrhosis and hemochromatosis in three horses. *Journal of American Veterinary Medical Association,* 204(7), pp. 1053-1056.

Pejaver, V. et al., 2020. Inferring the molecular and phenotypic impact of amino acid varaints with MutPred2. *Nature Communications,* Volume 11, p. 5918.

Peters, M. et al., 2015. DNA-based diagnosis of rare diseases in veterinary medicine: a 4.4 kb deletion of ITGB4 is associated with epidermolysis bullosa in Charolais cattle. *BMC Veterinary Research,* 11(48).

Schaefer, R. J. & McCue, M. E., 2020. Equine genotyping arrays. *The Veterinary Clinics of North America: Equine Practice,* 36(2), pp. 183-193.

Sendamarai, A. K., Ohgami, R. S., Fleming, M. D. & Lawrence, C. M., 2008. Structure of the membrane proximal oxidoreductase domain of human Steap3, the dominant ferrireductase of the erythroid transferrin cycle. *Proceedings of the National Academy of Sciences of the United States of America,* 105(21), pp. 7410-7415.

Smith, B. P., 2009. *Large Animal Internal Medicine.* 4th ed. St. Louis, Missouri: Mosby, Inc., an affiliate of Elsevier Inc..

Sydney School of Veterinary Science, n.d. [Online]
Available at: https://omia.org/
[Accessed 12 February 2023].

Van der Auwera, G. A. et al., 2018. From FastQ data to high-confidence variant calls: The Genome Anaylsis Tookit Best Practices Pipeline. *Current Protocols in Bioinformatics,* pp. 11.10.1-11.10.33.

Yariv, B. et al., 2023. Using evolutionary data to make sense of macromolecules with a 'face-lifted' ConSurf. *Tools for Protein Science.*

Zhang, F. et al., 2012. Metalloreductase Steap3 coordinates the regulation of iron homeostasis and inflammatory responses. *Haematologica,* 97(12), pp. 1826-1835.