Genetic evidence from zooarchaeological samples suggests Iron Age horses not specialized in size or speed

Boopalakrishnan S. Arul
Division of Biological Sciences, University of California San Diego, San Diego, CA
ANAR166

Introduction

The many breeds of horses, with different coat colors, sizes, and other conserved properties, are variations within the species created by this event. Today, horse breeds fall into three main categories and several intermediate ones. Ponies are horses which have a mature height at the withers (the portion of the spine that protrudes above the shoulder blades) less than 148 cm, and are used by children or trainee riders. Draft, cold-blooded, or brachymorphic horses are tall, heavy-boned, and covered in short, thick muscles heavy in slow-twitch fibers for slow contraction. They can pull heavy loads such as carriages. Light, warm-blooded, or dolichomorphic horses have longer spines and long, thin muscles with fast-twitch fibers, and are specialized for sport and racing (Dall'Olio et al. 2010).

This paper will investigate what early civilizations sought to do with their selective power over horses, and specifically which physical conformation they considered best for their purposes. Did they favor tall, strong draft horses, or leaner horses bred for speed?

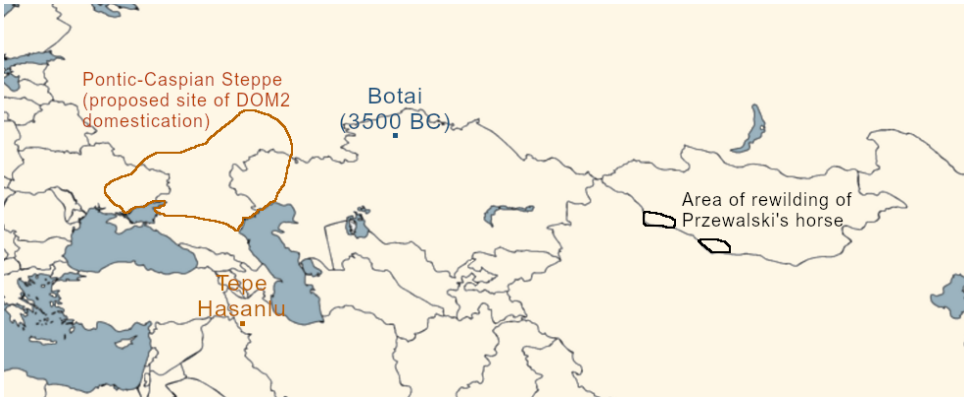

Figure 1: Map of sites frequently mentioned in this paper.

The Threshold of Domestication

The first domestication of horses is thought to have occurred in the Eneolithic Botai culture (named for its primary archeological site, a pithouse village on the Ishim River straddling the Kazakh-Russian border) dating to 3500 BC. Here, a large assemblage of nearly exclusively

horse remains was found, along with a material culture of horse-bone tools and decorated objets d'art. Outram et al. used the proportions of deuterium in two varieties of fatty acids found on ceramics in order to identify their origins, and concluded that some of them were from horse milk (Outram et al. 2009, Fig. 3). Some of these early findings have since been challenged. Genomic analysis has proven the Botai horses to ancestors of *Equus przewalskii*, the wild Przewalski's horse, and not the domestic *E. ferus caballus* (Gaunitz et al. 2018). A more radical challenge to the theory of a horse domestication at Botai rejects the identification of the ceramics as milk containers, as well as the identification of a strip of exposed enamel on the front of a premolar as wear-and-tear from a metal bit (Taylor and Barron-Ortiz 2021). The pro-domestication rebuttal has focused on enclosures or enclosure-like structures found near the ancient village, but has also sought to reaffirm that the fatty acids discovered in 2009 were sourced from horse milk (Outram et al. 2021).

Why is milk so pivotal to the argument for domestication? If humans had access to a mare's milk, then the mare likely gave birth in captivity. It also must have been docile enough to permit a human to milk it. The distinguishing feature of the Botai archaeological site, which sets it apart from sites of particularly successful horse-hunting cultures, is the possible presence of horse milk proteins on its ceramics, which is why this finding is so strongly defended by those who hold that horse domestication occurred at Botai in the 5000s BC (Outram et al. 2009; 2023).

The second case of domestication, occurring on the Pontic steppe north of the Caspian Sea, is more relevant to the origins of modern domestic horses. The evidence for a successful domestication here ranges from the discovery of unique horse milk proteins in the dental calculi of two people from the Krivyanski 9 archaeological site on the Don River, identified with the Yamnaya culture of the early 3000s BC (Niskanen 2023), to the uncovering of oldest chariots in the world (Librado et al. 2021; Niskanen 2023).

The descendants of this domestication share an ancestral component referred to as DOM2 in recent studies. This component spread over the entirety of Eurasia and northern Africa at the turn of the second millennium BC. Native wild populations contributed some ancestry to modern domestic horses, but otherwise they have almost totally disappeared under sustained genetic introgression (Librado et al. 2021). Of the two wild horse groups that remain, one is the aforementioned Przewalski's horse, brought back from the brink of extinction and reintroduced into the wild over the last few decades but still very endangered (Taylor and Barron-Ortiz 2021). The other is the tarpan, a mixture of wild horses native to Europe and horses related to DOM2 (Librado et al. 2021).

Published Data for a New Question

Librado et al. sequenced 264 ancient horse genomes. This paper will focus on thirteen of those genomes, four of which were previously sequenced by two other studies (Gaunitz et al. 2018; Fages et al. 2019). The first eight genomes are from the Iron Age layers of Tepe Hasanlu, a site of continuous settlement for thousands of years until it was abandoned in the last millennium, and are between 2863 and 2613 years old.

Tepe Hasanlu was a horse breeding center in the early Iron Age, as attested in Assyrian records (Danti and Cifarelli 2013, p. 7). Nestled in the Solduz valley of the Zagros Mountains, southwest of Lake Urmia, the hinterland of the settlement consisted of arable and well-watered land. In the surrounding uplands, barley and other fodder were grown for the sheep, goats, and horses in the summer. The site also promised a diverse selection of horses, as it was the site of a great battle. Around 800 BC, the period corresponding with layer IVb of the site, the city was sacked and abandoned. The horse Hasanlu2405 was found in a burned building containing horse harnesses and skeletons (Gaunitz et al. 2018). It may have been a horse trained for work or war, which died in its stable. Hasanlu1140, V31E, 2327, and 3398 lived during the subsequent recovery and repopulation of the city under foreign occupation (Fages et al. 2019).

The eight horses' sequence alignment files were downloaded from the EBI database, where Librado et al. made their data available. Each file contains reads of all the DNA molecules harvested from a given sample and their most likely positions within the genome. These files were compared at nine positions to check for single nucleotide polymorphisms (SNP), alternative sequences differing from the reference by a single base at a single position. Horses as diploid organisms have two copies of each gene, which might be the same or different alleles, or variant sequences of the same gene. The specific pair is called a genotype. These nine positions were chosen based on their proximity to genes identified as key contributors to horse phenotypes such as height, weight, and performance in races (Han et al. 2022; Makvandi-Nejad et al. 2012; Tozaki et al. 2011). The Makvandi-Nejad et al. and Tozaki et al. variant positions were aligned with an older reference genome (EquCab2.0, completed September 2007) than the one used for all other data in this paper (EquCab3.0, completed January 2018). The older coordinates were located in the new reference genome with UCSC Genome Broswer's LiftOver tool.

A

```
samtools mpileup --verbosity 4 -r chr22:1176610-1364857 -f ../Equus_caballus.EquCab3.0.dna.toplevel.fa
 Hasanlu1140_CGG-1-019998_Ira_m663.Bowtie2.EquCab3.pmds1.r.t.m.s.q25.bam Hasanlu2327_CGG-1-019995_Ira_
m768.Bowtie2.EquCab3.pmds1.r.t.m.s.q25.bam Hasanlu2405_CGG-1-019992_Ira_m868.Bowtie2.EquCab3.pmds1.r.t
.m.s.q25.bam Hasanlu3394_CGG-1-019997_Ira_m790.Bowtie2.EquCab3.pmds1.r.t.m.s.q25.bam Hasanlu3398_CGG-1
-019986_Ira_m768.Bowtie2.EquCab3.pmds1.r.t.m.s.q25.bam Hasanlu3461_CGG-1-020003_Ira_m913.Bowtie2.EquCa
b3.pmds1.r.t.m.s.q25.bam Hasanlu368_CGG-1-019994_Ira_m878.Bowtie2.EquCab3.pmds1.r.t.m.s.q25.bam Hasanl
uV31E_CGG-1-021461_Ira_m768.Bowtie2.EquCab3.pmds1.r.t.m.s.q25.bam > pileups_speed/SYNDIG1.mpileup
```

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9	Column 10	Column 11	Column 12	Column 13	Column 14	Column 15
chr5	52312504	C	0	*	*	1	,	J	0	*	*	3	...	JJJ
chr22	23460297	T	1	.	J	1	.	J	1	A	J	0	*	*

B

```
java -jar VarScan.jar pileup2snp greps_hasanlu.mpileup --min-coverage 2 --min-reads 2 --min-avg-qual 1
5 --min-var-freq 0.01 --p-value 0.99
```

Sample1	Sample2	Sample3	Sample4	Sample5	Sample6	Sample7
./.:.:1	./.:.:1	./.:.:1	./.:.:0	./.:.:0	1/1:0:3:3:0:3:100%:9.8E-1:0:40:0:0:1:2	1/1:0:2:2:0:2:100%:9.8E-1:0:41:0:0:1:1
./.:.:0	./.:.:1	./.:.:2	./.:.:3	./.:.:0	./.:.:1	./.:.:0

Figure 2: a) Example input and output of the samtools mpileup utility. For each variant position, a pileup file that collates all the reads from all the samples was generated. In this example, the range is set by the start and end positions of the SYNDIG1 gene (chromosome 22, 1,176,610 - 1,264,857). b) Example input and output of the VarScan pileup2snp utility which assesses the most likely genotype (pair of alleles) at each variant position.

However, the elimination of sources of error also drastically reduced the total amount of DNA in each of the eight samples. Librado et al. treated their DNA samples with enzymes that cleave the molecule on either side of zones affected by cytosine deamination. The affected zones, however, are lost.

Not much ancient DNA survived the centuries to begin with, so this resulted in low coverage. Each of the eight original sequence alignment files is expected to contain duplicates and overlaps. If fifty reads overlapped a position in the genome (achieving 50x coverage at that position), it would be expected that equal proportions of those fifty reads come from each of the two homologous chromosomes containing that position. When the coverage is low, and this position is only represented in one read or even none at all, the genotype cannot be confidently identified. Getting around this requires imputation, or filling in the missing information with the most likely genotype, as determined from a panel of organisms from the same species whose genomes have already been sequenced with high coverage (Sousa da Mota et al. 2023).

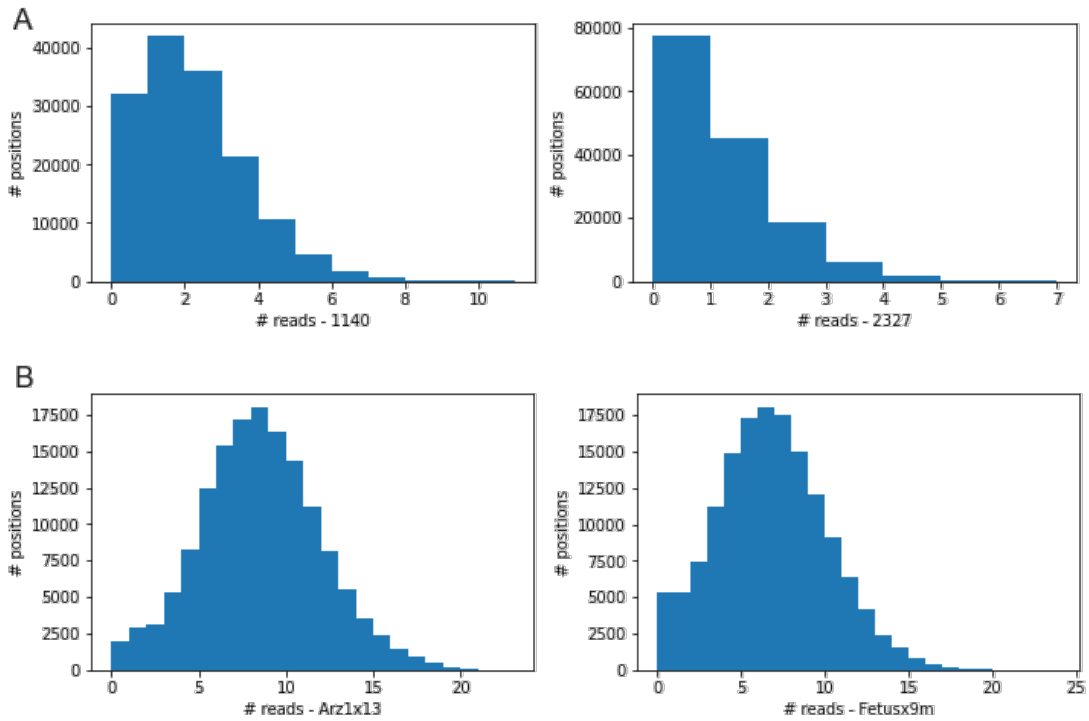

Figure 3: Histogram of coverage across the SLC16A1 gene between two horses each from the Hasanlu group and the high coverage group. In the latter group's samples, fewer positions have low or no coverage. There are therefore more positions where the genotype can be determined.

This paper instead uses a second set of Iron Age horses from the same dataset that have higher coverage, to illustrate how genotyping would work with more ideal data. The point of this inclusion is to show that although difficulties were faced in genotyping the Hasanlu horses, this difficulty is not from a qualitative difference in the data, only a quantitative one. The second set of horses simply have more data to offer. The paper could have been oriented to focus only on this second set, but what information can be gleaned from the Hasanlu horses still paints a useful picture.

The Results of Genotyping

Sample Name	Site	Country	Age (Years BP)	Mean Coverage
Hasanlu1140	Tepe Hasanlu	Iran	2613	1.80
Hasanlu2327	Tepe Hasanlu	Iran	2718	0.72
Hasanlu2405	Tepe Hasanlu	Iran	2818	2.63
Hasanlu3394	Tepe Hasanlu	Iran	2740	2.77

Hasanlu3398	Tepe Hasanlu	Iran	2718	1.08
Hasanlu3461	Tepe Hasanlu	Iran	2863	5.56
Hasanlu368	Tepe Hasanlu	Iran	2828	1.76
HasanluV31E	Tepe Hasanlu	Iran	2718	1.99
Arz1x13	Arzhan-I	Russia	2804	8.18
Fetusx9m	Els Vilars	Spain	2425	6.37
GVA9035	Tatsyn Ereg	Mongolia	2974	5.77
Kuymx2	Kuyum	Russia	2530	7.07
UE10287	La Monédière	France	2567	8.68

Table 1: The thirteen samples used in this paper. The first eight are from Tepe Hasanlu. The last five are also from the early Iron Age (first millennium BC), and have the highest mean coverage out of all the samples in that category. Information adapted from Librado et al. 2021, Supplementary Table 1.

Gene	Associated With	Full name, Role	Variant Position	Ref. Allele
SLC16A1	speed	solute carrier family 16 member 1; catalyzes movement of lactate and pyruvate across plasma membrane, anaerobic fermentation (Han et al. 2022)	chr5: 52312504	C
MYLK2	speed	myosin light chain kinase 2; muscle contraction, synaptic transmission, skeletal muscle cell differentiation, muscle filament sliding (Han et al. 2022)	chr22: 23460297	T
NTM	speed	neurotrimin; neural growth and synapse formation, learning and memory; positively selected during horse domestication (Han et al. 2022)	chr7: 41381993	G
G6PC2	speed	glucose-6-phosphatase catalytic subunit 2; major component of glycolysis (Han et al. 2022)	chr18: 48638568	G
SYNDIG1	speed	synapse differentiation inducing 1; regulates development of excitatory synapses (Han et al. 2022)	chr22: 1289141	C
LCORL	strength	ligand dependent nuclear receptor corepressor-like; transcription factor associated with human height (Makvandi-Nejad et al. 2012)	chr3: 107374136	T
ZFAT	strength	zinc finger and AT hook domain containing; associated with height in multiple human population	chr9: 77656570	T

		(Makvandi-Nejad et al. 2012)		
LASP1	strength	LIM and SH3 protein 1; mediates cell migration and survival (Makvandi-Nejad et al. 2012)	chr11: 23334511	A
MSTN	strength	myostatin, GDF-8; member of transforming growth factor β family, thought to inhibit growth and number of muscle fibers (Tozaki et al. 2011)	chr18: 66654989	A

Table 2: The nine variant positions examined for SNPs, which gene they are associated with (either within, upstream, or downstream), the role and purpose of that gene, and whether it is counted as a “strength” or “speed” gene for purposes of this paper.

	SLC16A1	MYLK2	NTM	G6PC2	SYNDIG1	LCORL	ZFAT	LASP1	MSTN
H. 1140						TT			
H. 2327									
H. 2405			GG	AA	CC	TT	CC		GG
H. 3394	CC		GG				CC	GG	AA
H. 3398						TT		GG	
H. 3461	CC	AA		GG	CC	TT		GG	AA
H. 368	CC	AA			CC				GG
H. V31E		TT	AA		CC	TT			
Arz1x13	CC	TT	GA	AA		TT	CC	GG	GG
Fetusx9m	CG	TT		GG	CT	TT	CC	GG	AA
GVA9035	CC	TT		GG	CC	TT	CC	GG	AA
Kuymx2	CC	TT	GG	GA	CC	TT	CC		AG
UE10287	CG	TT			CC	TT	TC	GG	AA

Table 3: Likely genotypes at each variant position listed in Table 2, for each of the samples listed in Table 1. A blank cell means the genotype could not be determined, because of insufficient (less than 3 for the high coverage samples, less than 2 for the Hasanlu samples) reads.

These genotypes, to the extent that they can even be determined, are not a complete picture of variation in complex phenotypic traits like speed, body size, or strength. The genes in Tables 2 and 3 are not solely responsible for any trait. Nor is any gene sufficient to compel a particular phenotype: genes are selectively activated and regulated, affecting the number of RNA transcripts produced from them, and these are in turn regulated— delayed, marked for

destruction— on the way to being translated into proteins, which share responsibility for complex traits. Advances in ancient RNA recovery and sequencing promise information on how different tissues, and different cells within those tissues, expressed different genes to varying levels (Fromm et al. 2020). It is even possible that ancient RNA might uncover extinct genes, RNA transcripts of DNA genomic sequences that don't code for proteins in modern animals.

Genomic data still has its uses. Highly conserved sequences can serve as markers of phylogenetic distance between species, or between members of a single species. Additionally, all the regulation extending outward from the gene cannot change the protein that the gene codes for— if both copies of a gene have a mutation that creates an early stop codon, ending transcription early, then the original version of that protein cannot be produced, and the result will in most cases be a nonfunctional protein. This allows SNPs to serve as markers associated with particular traits or disorders.

The association of a particular genetic mutation with a particular phenotype has to be evaluated statistically. For example, four variant positions in horses can explain 83.5% of the variation in size (Makvandi-Nejad et al. 2012). Three of these are the variant positions upstream of LCORL, downstream of ZFAT, and in LASP1 listed in Table 2. All twelve of the samples in Table 3 which have genotypes determined for at least one of these positions display alleles associated with small size (Makvandi-Nejad et al. 2012).

Meanwhile, on the variant position in gene MSTN, four horses (Hasanlu2405, Hasanlu368, and Arz1x13) display the genotype GG, which was associated with a low mean ratio of body weight to withers height in Thoroughbreds (Tozaki et al. 2011). Five horses had genotype AA and one, Kuyxm2, had genotype AG. AA was associated with greater weight-to-height ratios in September, but in every other month of the six-month period where Tozaki et al. conducted their study the AA genotype horses trailed the AG genotype horses in weight-to-height ratio, the GG horses trailing both by a larger margin. Since none of these horses are likely to have been particularly tall, going off the variant positions described above, it's likely that horses with the GG genotype at this variant position put on less muscle, and consequently weighed less.

The variant position in SLC16A1 is noteworthy because a G in that position creates an early stop codon, usually something that makes a protein nonfunctional. However, this novel variant was identified in Arabian horses, and was significantly associated with racing breeds (Han et al. 2022). This allele was seen in Fetusx9m and UE10287. The A allele for the variant position in MYLK2 has a frequency of 0.65 in Arabian horses, meaning that 65% of the alleles within that population (each horse has 2 alleles) are the A allele (Han et al. 2022). Its frequency

in other racing breeds ranges from 0.45 to 0.28. Hasanlu3461 and Hasanlu368 are homozygous for this allele. HasanluV31E and Arz1x13 have the racing-associated allele (A) for NTM. Hasanlu2405, Arz1x13, and Kuymx2 have the racing-associated allele (A) for G6PC2. Fetusx9m has the racing-associated allele (T) for SYNDIG1.

Overall, Fetusx9m and Arz1x13 have racing-associated alleles at two loci each, while the other horses have them at one locus or none. For both of these horses, one locus has to do with sugar catabolism (SLC16A1, G6PC2), while the other has to do with nervous system growth and differentiation (NTM, SYNDIG1). These two horses are the most likely to be faster than their peers.

Discussion

Large-scale phylogenetic analyses have assigned all modern and ancient domestic horses into a single cluster, distinguished by an ancestral component named DOM2 (Gaunitz et al. 2018; Fages et al. 2019; Librado et al. 2021). A more recent paper put forward a timeline of DOM2's development and spread. Western Eurasia's horses were, in the middle Holocene, partitioned into several distinct habitats with minimal migration or gene flow between them, a finding anticipated by earlier climate simulations (Librado et al. 2021; Leonardi et al. 2018). The DOM2 ancestral component first became the majority of ancestry in the Volga-Don group of horses (Librado et al. 2021, Fig. 1 e,f) in the 5000s BC, and starting from 2200 BC at the earliest it rapidly dispersed through Europe and Asia, overrunning previous barriers to horse migration. Other horse groups, such as the native horses of western Europe, have left only traces on the ancestry of modern horses (Librado et al. 2021).

The two trademark mutations of the DOM2 horses don't concern size or speed. One mutation is upstream of the GSDMC gene, and may affect a transcription factor binding site. In other mammals this gene is silenced by the inclusion of extraneous elements, because when expressed it is a marker for lumbar spinal stenosis, a chronically painful hardening of the lumbar vertebrae. Instead, DOM2 horses have a mutation that may encourage the expression of this gene. The second mutation is closest to ZFPM1, involved in the serotonergic neurons of the dorsal raphe, and likely leading to greater mood regulation and decreased aggression. The qualities that drove this horse's success were sturdiness and docility (Librado et al. 2021).

One thousand years later, in the first millennium BC, were any other core traits added onto domestic horses? The picture created by the data analyzed in this study suggests that horses were not yet specialized through breeding to isolate particular batches of alleles. This paper's sample size was only 13 horses— however, the median of the mean coverages across

all positions within the 135 DOM2 samples in the Librado et al. 2021 dataset was only 1.8. The difficulties in genotyping caused by low coverage in one sample would not be affected or offset by adding even more low-coverage samples.

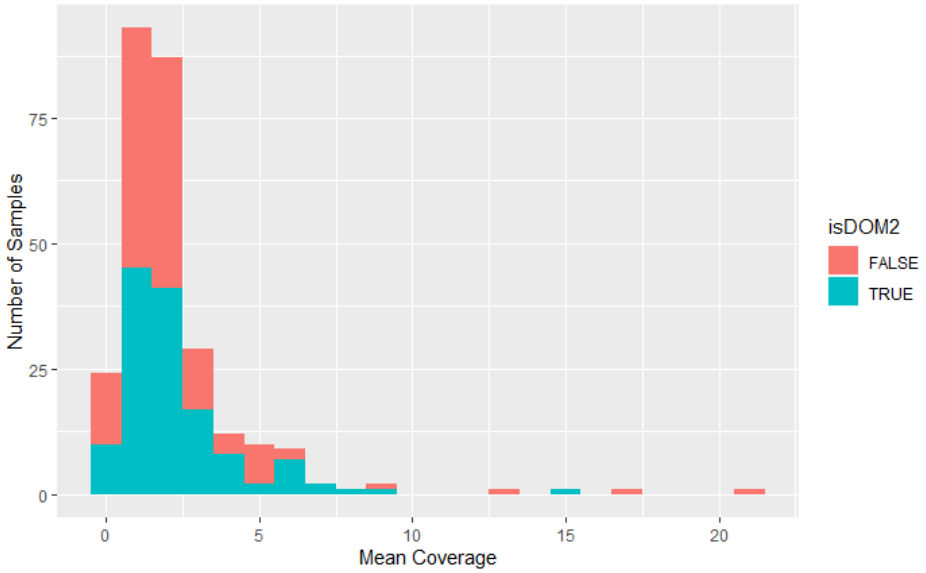

Figure 4: Histogram of mean coverages for the 135 DOM2 and 138 non-DOM2 horses in Librado et al. 2021.

This paper's methods also run the risk of taking the part for the whole. Although each of the nine variant positions in Table 2 were chosen because they explained the variance in a trait or were associated with an increase or decrease in a trait, a variant allele should not be assumed to cause a particular effect all on its own. Although discovering the causes of conditions within the genome is part of the promise of genomics, a mutation can't immediately be cajoled into doing much more than testifying to its own existence. This is the premise of phylogeny: mutations in highly conserved sequences occur so rarely that closely related species will share that sequence down to the letter, even if their genomes differ elsewhere. This allows clusters of relative closeness or distance to be defined among a heterogenous group of species. In fact, in a previous paper Orlando and Librado use this principle to define mutations as "likely to be deleterious": their assumption is that a conserved sequence contributes to fitness by remaining conserved, meaning that mutations in such a sequence probably set the organism back somehow (Orlando and Librado 2019). The result of such a mutation may be guessed at by looking at the affected sequence's functional annotation, but the study does not definitively assign a function, never mind an effect size, to the variants it screens for.

A future paper might begin by separating the population into two groups assumed to differ on as few traits as possible, and then screen for variants with significantly different frequencies between the groups. For example, one could take horses living in colder and hotter climates, and look for variants linked to heat stress tolerance. This could explain how a species domesticated in Siberia and the Kazakh-Russian borderland would have subgroups at home in such sweltering locations as the Karakum, the Syrian Desert, and the American Southwest.

Conclusion

This paper invesftigated whether ancient horse breeders created breeds of domesticated horses specialized for qualities such as size, strength, or speed. Thirteen horses from the first millennium BC were screened for variants at nine positions in the genome. From genotyping on SNP positions in or close to genes associated with relevant traits, it was found that none of the horses had alleles suggesting a large size or excess muscle growth, but that two horses did hold alleles suggesting increased speed. This paper concludes that strong specialization for one use or another was not a goal for animal breeders in this period.

References

Dall'Olio, Stefania, Luca Fontanesi, Leonardo Nanni Costa, Marco Tassinari, Laura Minieri, and Adalberto Falaschini. "Analysis of Horse Myostatin Gene and Identification of Single Nucleotide Polymorphisms in Breeds of Different Morphological Types." *Journal of Biomedicine and Biotechnology* 2010 (2010): 1–11. https://doi.org/10.1155/2010/542945.

Outram, Alan K. et al. The Earliest Horse Harnessing and Milking. Science 323, 1332-1335 (2009). DOI:10.1126/science.1168594

Gaunitz, Charleen et al. Ancient genomes revisit the ancestry of domestic and Przewalski's horses. Science 360, 111-114 (2018). DOI:10.1126/science.aao3297

Taylor, W.T.T., Barrón-Ortiz, C.I. Rethinking the evidence for early horse domestication at Botai. Sci Rep 11, 7440 (2021). https://doi.org/10.1038/s41598-021-86832-9

Outram, A., Bendrey, R., Evershed, R. et al. Rebuttal of Taylor and Barrón-Ortiz 2021 Rethinking the evidence for early horse domestication at Botai. Preprint (2021). [https://doi.org/10.5281/zenodo.5142604]

Librado, P., Khan, N., Fages, A. et al. The origins and spread of domestic horses from the Western Eurasian steppes. Nature 598, 634–640 (2021). https://doi.org/10.1038/s41586-021-04018-9

Leonardi, Michela et al. “Late Quaternary horses in Eurasia in the face of climate and vegetation change”. Sci. Adv.4 (2018). 10.1126/sciadv.aar5589

Niskanen, Markku. “The prehistoric origins of the domestic horse and horseback riding”, Bulletins et mémoires de la Société d’Anthropologie de Paris [Online] 35, 1 (2023) https://doi.org/10.4000/bmsap.11881

Fages, Antoine, Kristian Hanghøj, Naveed Khan, Charleen Gaunitz, Andaine Seguin-Orlando, Michela Leonardi, Christian McCrory Constantz, et al. “Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series.” Cell 177, no. 6 (2019). https://doi.org/10.1016/j.cell.2019.03.049.

Danti, Michael D., and Megan Cifarelli. Hasanlu V: The Late Bronze and Iron I Periods. University of Pennsylvania Press, 2013. http://www.jstor.org/stable/j.ctt4cggt4.

Sousa da Mota, B., Rubinacci, S., Cruz Dávalos, D.I. et al. Imputation of ancient human genomes. Nat Commun 14, 3660 (2023). https://doi.org/10.1038/s41467-023-39202-0

Makvandi-Nejad S, Hoffman GE, Allen JJ, Chu E, Gu E, Chandler AM, et al. Four Loci Explain 83% of Size Variation in the Horse. PLoS ONE 7(7) (2012). https://doi.org/10.1371/journal.pone.0039929

Tozaki T, Sato F, Hill EW, Miyake T, Endo Y, Kakoi H, Gawahara H, Hirota K, Nakano Y, Nambo Y, Kurosawa M. Sequence variants at the myostatin gene locus influence the body composition of Thoroughbred horses. J Vet Med Sci. 2011 Dec;73(12):1617-24. https://doi.org/10.1292/jvms.11-0295.

Han, H., McGivney, B.A., Allen, L. et al. Common protein-coding variants influence the racing phenotype in galloping racehorse breeds. Commun Biol 5, 1320 (2022). https://doi.org/10.1038/s42003-022-04206-x

Orlando L, Librado P. Origin and Evolution of Deleterious Mutations in Horses. Genes (Basel). 2019 Aug 28;10(9):649. https://doi.org/10.3390/genes10090649.