Title

Short Tandem Repeat Analysis of Genetic Diversity Metrics in American Standardbreds and an Investigation on the Cause of the Rabicano Coat Color Phenotype Utilizing Short and Long Read Sequencing Methods.

Author

Esdaile, Elizabeth Susan

Publication Date

2021

Peer reviewed|Thesis/dissertation

Short Tandem Repeat Analysis of Genetic Diversity Metrics in American Standardbreds and an Investigation on the Cause of the Rabicano Coat Color Phenotype Utilizing Short and Long Read Sequencing Methods.

By

ELIZABETH ESDAILE

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Animal Biology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Rebecca Bellone, Chair

Felipe Avila

C. Titus Brown

Committee in Charge

2021

Table of Contents

Acknowledgments

I would like to express my deepest appreciation to my advisors, Dr. Rebecca Bellone and Dr. Felipe Avila, for all of their support, mentorship, patience, and encouragement throughout my degree. I am deeply indebted to Rayann Eaves for the invaluable work she did analyzing the short-read data and I am extremely grateful to Jocelyn Tanaka for all of her previous work on Rabicano. I would like to extend a sincere thank you Dr. C. Titus Brown for his support and help with this project and for providing the long-read data that helped make the Rabicano project possible. We will continue to analyze these data for some time to come! I also wish to thank Daniela Soto for her guidance in analyzing the long-read data and her beautifully written instructions. I must also thank T.C. Lane and Bruce Brinkerhoff with the United States Trotting Association for their support in the Standardbred research, and Bureau Veritas Laboratory for providing samples and STR data which made that work possible. I must also thank the Bellone Laboratory and the UC Davis Veterinary Genetics Laboratory for all of their support and advice throughout the last two years. I would like to thank Nicole Kingsley for preparing the sample for long read sequencing, for showing me the ropes, and for her support. And I cannot forget to thank Julia Malvick for her wisdom and technical support in ordering and preparing primers for genotyping.

Finally, I would like to express the deepest gratitude to my friends and family for their support and interest in this work, and I cannot wait to share it with you.

Abstract

The governing body for American Standardbreds, the United States Trotting Association (USTA), closed the studbook in 1973, halting new geneflow into the population. In 2009, the USTA capped the studbook of new sires at 140 mares bred per year to maintain genetic diversity within the breed. To investigate the state of genetic diversity within the American Standardbred during this shift in breeding practices, genotypes for sixteen STRs from American Standardbred horses foaled from 2010 to 2015 and their sires and dams (n=50,621) were investigated. Allelic richness (A_r), expected heterozygosity (H_E), observed heterozygosity (H_O), inbreeding coefficient (F_{IS}), and fixation index (F_{ST}) were calculated. These analyses found that trotting and pacing sires were less genetically diverse than dams (H_E $p_{bonferroni}$=.030 and <.001, respectively) and their offspring (A_r $p_{bonferroni}$ =.03 and <.001, respectively), and only pacing offspring were significantly less diverse than their dams (H_E $p_{bonferroni}$ =.002). Inbreeding coefficients for trotters (F_{IS}=-0.014) and pacers (F_{IS}=-0.012) suggest that breeding practices have maintained diversity within the breed. Moderate levels of genetic differentiation (.066<F_{ST}<0.105) were found between pacing and trotting groups, suggesting little interbreeding between the two groups based on gait. No statistically significant differences were identified in any of the metrics evaluated across the 6-year period, however, evaluating diversity prior to the studbook cap and across a longer time period will help to best define the impact of the studbook cap on the population. These data provide a basis with which to continue to monitor trends in inbreeding and develop strategies to maintain diversity for generations to come.

In addition to genetic diversity, horse breeders are concerned with selecting for traits that are of economic importance and prevent detrimental conditions. One such trait of economic importance is coat color, which can impact the value of a horse. One popular equine coat color is rabicano, a phenotype with an unknown genetic cause that can be found in many breeds of horses, including Quarter Horses and Arabians. Rabicano can be recognized by white ticking in the flank,

sometimes expanding forward up the barrel, and white banding on the tailhead. Using a candidate gene approach, Illumina short read sequencing data from four rabicano horses, 17 non-rabicano controls, and one indeterminate phenotype were utilized to identify and prioritize coding variants from 659 pigmentation-related genes for further investigation. Long read sequencing data from one of the rabicano horses were also evaluated to identify and investigate structural variants in the same set of candidate genes as the potential genetic cause of this phenotype. Six SNPs in *ANKRD27*, *CEP290*, *CRB1*, *FMN1*, *KIF13A*, and *OCA2* were investigated in a larger sample set comprised of 61 rabicano cases and 36 non-rabicano controls. While not perfectly concordant, the variant in *CEP290* (ENSECAT00000052715.2:c.538A>G) was the most concordant with phenotype, with a p-value of 2.24×10^{-07}. This suggests that either none of these variants are the cause of rabicano across breeds or a more complex mode of inheritance best explains this phenotype. Given that the variant in *CEP290* was most concordant with the rabicano phenotype, is located proximally to *KITLG,* and the phenotypic similarities of *KITLG* variants in other species, putative regulatory and structural variants flanking *KITLG* were also investigated. This analysis identified a potential 1.7Mb haplotype on ECA28 surrounding *KITLG* associated with the rabicano phenotype. The most concordant SNP was ECA28:rs397240012 (NC_009171.3:g.15967332G>A) ($p=7.36 \times 10^{-09}$). This SNP is located within a H3K4me1 annotated peak (indicative of a gene enhancer) in skin and is 160kb from the start of transcription for *KITLG*. Additional analyses are required to refine the haplotype across breeds and identify the causal variant of rabicano in horses.

Abbreviations

ABI	Applied BioSystems
AI	Artificial Insemination
Al	Serum albumin locus
ANKRD27	Ankyrin Repeat Domain 27
AR	Arabian Horse
A_r	Allelic richness
ATOH1	Atonal BHLH Transcription Factor 1
ATP2B1	ATPase Plasma Membrane Ca2+ Transporting 1
bp	Base pair
BLAT	BLAST-like alignment tool
BTBD11	BTB Domain Containing 11
BWA	Burrows-Wheeler Aligner
CP	Caspian Pony
CEP290	Centrosomal Protein 290
CNV	Copy number variant
CPM	Chip prep module
CRB1	Crumbs Cell Polarity Complex Component 1
CSNB	Congenital Stationary Night Blindness
DCT	Dopachrome tautomerase
DMRT3	Doublesex And Mab-3 Related Transcription Factor 3
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DUSP6	Dual Specificity Phosphatase 6
E	Extension
ECA	Equine chromosome
FAANG	Functional Annotation of Animal Genomes
FAM	Fluorescein amidites
F_{IS}	Inbreeding coefficient
FMN1	Formin 1
F_{ST}	Fixation index
GALNT4	Polypeptide N-Acetylgalactosaminyltransferase 4
GO	Gene ontology
GWAS	Genome-wide association study
H_E	Expected heterozygosity
H_O	Observed heterozygosity
HPLC	High performance liquid chromatography

HGVS	Human Genome Variation Society
HWE	Hardy-Weinberg Equilibrium
HYPP	Hyperkalemic periodic paralysis
IDT	Integrated DNA Technologies
IGV	Intgrative Genomics Viewer
INDEL	Insertion-deletion mutation
KIF13A	Kinesin family member 13A
kb	Kilobases
KIT	KIT proto-oncogene, receptor tyrosine kinase
KITLG	KIT Ligand
KWPN	Royal Warmblood Studbook of the Netherlands (“Dutch Warmblood”)
LD	Linkage disequilibrium
LEF1	Lymphoid enhancer binding factor 1
LINE	Long interspersed nuclear element
lm	Linear model
Lp	Leopard Pattern
LTR	Long terminal repeat
MA	MassARRAY
MC1R	Melanocortin 1 receptor
MCHR1	Melanin concentrating hormone receptor 1
MEI1	Meiotic double-stranded break formation protein 1
MGAT4C	Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase C
MGF	Mast cell growth factor (now known as KITLG)
MGI	Mouse Genome Informatics
MITF	Melanocyte inducing transcription factor
MO	Morgan Horse
MSA	Microsatellite Analyzer
NCBI	National Center for Biotechnology Information
OCA2	Oculocutaneous albinism II
PATN1	Appaloosa Pattern-1
PCR	Polymerase chain reaction
PCRP	Pre-amplification PCR primers
POC1B	POC1 centriolar protein B
POMC	Proopiomelanocortin
PPFIA2	Protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 2
QH	Quarter Horse
RER	Recurrent rhabdomyolysis

RHOH	Ras Homolog Family Member H
Rn	Roan
ROH	Runs of homozygosity
SAP	Shrimp alkaline phosphatase
SIFT	Sorting intolerant from tolerant
SINE	Short interspersed nuclear element
SNAP	Screening for Non-Acceptable Polymorphisms
STX17	Syntaxin 17
sl	Steel locus in mice
SLC24A5	Solute carrier family 24 member 5
SNP	Single nucleotide polymorphism
STR	Short tandem repeat
SV	Structural variant
TB	Thoroughbred Horse
TMTC3	Transmembrane O-mannosyltransferase targeting cadherins 3
TNRC6B	Trinucleotide repeat containing adaptor 6B
TRPM1	Transient receptor potential cation channel subfamily M member 1
TRF	Tandem repeat finder
TYR	Tyrosinase
TYRP1	Tyrosinase related protein 1
UCSC	University of California, Santa Cruiz
UEP_SEQ	Extension Primer
USH2A	Usherin
USTA	United States Trotting Association
VEP	Variant Effect Predictor
VGL	University of California, Davis Veterinary Genetics Laboratory
WB	Warmblood Horse
WE	Welsh Pony
WT	Wild type
XPNPEP3	X-prolyl aminopeptidase 3

List of Figures

List of Tables

Appendices

Chapter 1: A History and Review of Genetic Diversity in American Standardbreds and Roan-like Phenotypes in Mammals.

1 American Standardbreds

1.1 History of the American Standardbred

The American Standardbred is a breed of horse developed for racing in the 19th century. They differ from the more common breeds of racehorses, as American Standardbreds were developed for harness racing at a trot or a pace while pulling a one-person cart called a sulky. Standardbreds descend primarily from road horses and racing breeds extant in the early- to mid-19th century, including Thoroughbreds, Trotters, Norfolk Trotters, Morgans, Barbs, Canadian Pacers, and other driving horses imported from Europe. Notable founding sires include the famous Thoroughbred, Messenger (foaled in 1780), and the famous Trotter, Hambletonian 10 (foaled in 1849, commonly referred to as Rysdyk's Hambletonian).

The first attempt at a Standardbred studbook was John H. Wallace's American Stud-book, Volume 1, published in 1867. Although this book focused primarily on Thoroughbreds, it contained a supplement for all horses from the earliest recorded trotting races through 1866 who had trotted a mile in a public race under a standard time, plus all of their known ancestors (Wallace 1867). Appreciation from the Trotting horse breeders and opposition from the Thoroughbred breeders caused Wallace to shift his studbook efforts solely to trotting horses. In 1871, Wallace published what is now widely regarded as the first Standardbred studbook, Wallace's American Trotting Register (Wallace 1871). The formation of the first national registry followed shortly thereafter with the 1876 establishment of the National Association of Trotting Horse Breeders. In 1879, admission standards into the registry were established, recognizing approved horses as "a standard trotting-bred animal" (McCarr 1971; Wallace 1885). These admission requirements are the foundation of the breed we now know as the Standardbred (McCarr 1971; Wallace 1897).

Selection for horses that did not "break gait", or canter, during a race resulted in the American Standardbred population becoming fixed for the "gait keeper" variant (Promerová *et al.* 2014). The "gait keeper" variant is a nonsense mutation in the last exon of the doublesex and mab-3 related transcription factor 3 gene (*DMRT3*) that is expected to produce a truncated protein (rs1150690013, NC_009166.3::g.22391254C>A, ENSECAT00000024805.3:c.1391C>A, *ENSECAP00000020624.3:p.(Ser464*)*)(Andersson *et al.* 2012; Promerová *et al.* 2014). This variant impacts coordination and locomotion and allows for alternative gaits. (Andersson *et al.* 2012). *DMRT3*-null mice have been found to manifest significantly increased stride length, swing time, stance time, and hind leg propulsion as well as difficulties running at higher speeds (Andersson *et al.* 2012). Most horses with the "gait keeper" mutation exhibit a parallel phenotype to mice, gaining the ability to perform alternate gaits such as the pace, amble, tölt/rack, running walk, foxtrot, or *marcha batida* (Staiger *et al.* 2017); all diagonal or lateral couplet gaits found in certain horse breeds. Trotting Standardbred horses are the most common exception to this genotype-phenotype correlation, as although they are fixed for this "gait keeper" mutation, they are unable to pace. Instead, they demonstrate the standard gaits, walk, trot, and canter, but manifest a faster trot with greater stride lengths and difficulty cantering, traits ideal for individuals racing at the trot. In the last century, selection based on gait resulted in two subpopulations of American Standardbreds: trotters and pacers. In the present-day American Standardbred population, it is thought to be rare to encounter interbreeding between trotting and pacing horses, resulting in further differentiation between the two subpopulations.

Governance of the breed underwent many changes in the late 19th and early 20th century, including Wallace being removed from updates to his stud-book and the foundation of a host of competing breed organizations (McCarr 1972). In the 1930s, an effort to unite the breed resulted in the foundation of the United States Trotting Association (USTA) in 1939 (McCarr 1972). From then on, horses had to be registered with the USTA to be eligible to race (McCarr 1972). Although an effort was made to close the studbook during the formation of this new registry, the American

Standardbred studbook was not officially closed until 1973 (King 1992; McCarr 1972). This closure limits registration of Standardbreds to offspring of horses already registered with the USTA or global partner associations. Closing a studbook inherently limits the breeding pool and prevents additional genetic diversity from entering the population. Moreover, selection pressures can further decrease genetic diversity within a population, resulting in inbreeding depression (rare deleterious alleles increasing in frequency and negatively affecting the health and fertility of the breed) (Hartl 2020).

1.2 Breeding Population

In an effort to prevent loss of genetic diversity within the breed, the USTA implemented a cap on the number of mares a stallion could breed each year. The book size of trotting sires who debuted in 2009 or later was capped at to 140 mares per year, while new pacing stallions had their 2009 book capped to 160 mares, their 2010 book capped at 150 mares, and their 2011 book and beyond capped at 140 mares. This studbook cap was determined based on genetic diversity work by Dr. Gus Cothran, who completed an unpublished study for the USTA in the mid-2000s (Nevills 2019).

This studbook cap was implemented to increase the number of breeding individuals in the population. A limited breeding pool increases both the genetic load of a population (a measure of deleterious mutations present in an individual or population) and decreases overall diversity. This can result in inbreeding and increases the likelihood of an individual inheriting a detrimental recessive condition Examples of such detrimental conditions include atrial fibrillation, tarsal osteochondrosis, and recurrent exertional rhabdomyolysis (RER), three moderately heritable diseases that can impact performance in American Standardbreds (Kraus *et al.* 2017, 2018; Lykkjen *et al.* 2014; McCoy *et al.* 2016, 2019; Norton *et al.* 2016). Although associated genetic risk factors for osteochondrosis and RER have been identified in American Standardbreds (McCoy *et*

al. 2019; Valberg *et al.* 2019), these are believed to be complex diseases with no commercially available genetic tests to aid in breeding decisions.

In addition to reducing genetic diversity and potentially propagating genetic disorders through influential sire lines, high levels of inbreeding have been correlated with reproductive failure in some breeds. One example of this can be found in Friesian horses, a breed which underwent a significant bottleneck at the start of the 20th century with only three stallions registered in 1917, followed by a second bottleneck of 1,000 registered horses and 500 registered matings in the late 1960s (Sevinga, Vrijenhoek, *et al.* 2004). High levels of inbreeding were found to be correlated with an increase in the rate of retained placenta in Friesian horses, a disease that affects 53.9%-65.2% of Friesian horse pregnancies (Sevinga, Barkema, *et al.* 2004; Sevinga, Vrijenhoek, *et al.* 2004). Previous studies in pacing American Standardbreds suggest that increased rates of inbreeding result in lower conception and foaling rates. However, this association was found to impact, at most, 2% of the variation in conception or foaling rates (Cothran *et al.* 1986; MacCluer *et al.* 1983).

Additionally, a correlation between increased inbreeding and decreased reproductive success has been noted in other species, such as dairy cattle (González-Recio *et al.* 2007; Gutiérrez-Reinoso *et al.* 2020). Cumulatively, these findings suggest that the correlation between inbreeding and reproductive success should be monitored in Standardbreds to prevent problems in future generations. Nevertheless, the rate of inbreeding is not always correlated with reproductive failure, as the age of the mare, breeding method, and other environmental factors have also been shown to have a greater impact on reproductive success (Cothran *et al.* 1986; Müller-Unterberg *et al.* 2017; Todd *et al.* 2020).

It should be noted that breeding decisions impact the number of offspring a stallion can sire each year. This is important because increasing the live-foaling rate could increase the prevalence of a sire's contribution to the gene pool and can directly impact diversity and increase the frequency of disease variants in a population. As one example, in Finnish Standardbreds, the

foaling rate decreased from 75.1% in 1991 to 65.9% in 2005 (Katila *et al.* 2010). This decrease corresponds to an aging broodmare population and a shift in popular artificial insemination (AI) techniques, from on-site AI to the less successful transported and frozen semen methods, both of which have been shown to decrease fertility rates (Katila *et al.* 2010; Müller-Unterberg *et al.* 2017; Todd *et al.* 2020). These data suggest that in addition to a studbook cap, other breeding management strategies can positively impact genetic diversity in a population. For example, one consequence of the studbook cap may be a shift in breeding management to younger, more fertile mares, or advancement in reproductive technology to ensure each breeding results in a foal. Evaluating genetic diversity is essential to monitor trends within the breed, develop genomic resources to assist in maintaining diversity and develop tools to monitor and unravel genetic mechanisms of heritable traits including genetic diseases and performance phenotypes.

1.3 Genetic Diversity of American Standardbreds

In the 1980s, inbreeding of American Standardbreds was analyzed using short tandem repeats (STRs) and deep pedigree analysis of more than 14 generations. These pedigree analyses found trotters to be more inbred than pacers. In pacers, a decrease in pedigree inbreeding values was associated with increased expected heterozygosity from STR data ($p<.001$). These results suggest that pacers are a more diverse and less inbred population than trotters (Cothran *et al.* 1986).

Additionally, genetic diversity in the American Standardbred was investigated using ten blood markers of 4,404 trotters foaled between 1973 and 1989 and 12,271 pacers foaled between 1971 and 1989 (King 1992). King found that in both trotters and pacers, the mean percent of observed total heterozygosity was statistically significantly decreasing over time. This was seen more so in trotters than pacers. This may be due to a decrease in the size of the breeding population owing to the closure of the studbook, as in 1987, 2% of the sire population (60 horses) were responsible for 26% of the foal crop (King 1992).

In 2013, Petersen *et al.* published a report of single nucleotide polymorphism (SNP)-based diversity metrics in many horse breeds (Petersen *et al.* 2013). Significant excess homozygosity (F_{IS}) was identified in a small sample set of American Standardbreds (seven trotters and eight pacers) which was not identified in Norwegian Standardbreds (n=25), a breed consisting of only trotters and no pacers. This suggests that excess homozygosity in American Standardbreds could be due to substructure based on gait or increased inbreeding in the individuals examined (Petersen *et al.* 2013). Additional work needs to be done in the American Standardbred to further tests these hypotheses.

Metrics to evaluate genetic diversity using STRs include allelic richness, observed heterozygosity, expected heterozygosity, inbreeding coefficient, and fixation index. Allelic richness (A_r) is a measure of the number of alleles in a population and can reflect the amount of genetic differentiation within a population. Observed heterozygosity (H_O) is the amount of genetic variation in a population, measuring the proportion of heterozygotes. Expected heterozygosity (H_E) is an estimation of the amount of heterozygosity in a population following Hardy-Weinberg equilibrium (HWE). Observed heterozygosity that is larger than expected heterozygosity is desired, as it suggests that the population is genetically diverse. Observed heterozygosity that is lower than expected heterozygosity suggests loss of genetic diversity (inbreeding) within a population and may be cause for concern. Inbreeding coefficient (F_{IS}) is an estimation of the amount of inbreeding in an individual, with values close to or at zero being most desirable as these indicate low levels of relatedness between its parents. Higher values may be found in populations with a small breeding population where inbreeding is common. Fixation index (F_{ST}) is a measure of differentiation between populations due to genetic substructure. Higher fixation index values indicate greater differentiation between the two populations, with 5%-15% indicating moderate differentiation, 15%-25% indicates great differentiation, and over 25% indicates very great differentiation (Sewall Wright 1978). Values under 5% suggest little differentiation or that the two populations are genetically similar. All of these metrics are easily accessible measures for many

equine breed associations as they can be calculated with STR data collected during parentage verification.

Knowledge of genetic diversity within a breed provides essential information for the breed association, breeding management decisions, and provides a basis to monitor trends over time. These data can also provide tools and resources with which to ask questions about heritable traits in the breed and can lead to genetic tests enabling additional breeding management tools within and across breeds. Traits of highest economic interest among breeders include disease, performance, and coat color. While many of the American Standardbreds are bay in coat color (shade of red body with black lower legs, mane, and tail) some do have white patterning in their coat of unknown genetic mechanisms. A popular white spotting pattern across multiple breeds is the rabicano pattern, a form of roaning, or white hairs intermixed in the coat. Discovering the genetic cause(s) of this trait would assist several breeds in mate selection and allow for the determination of whether these alleles are present and at what frequency in the Standardbred and other breeds.

2 Roan and roan-like Phenotypes in Mammals

Roaning is a term used by pigmentation biologists to describe the phenotype of white hairs mixed with pigmented hair in the coat of mammals. This is a broad term as multiple slight phenotypic differences can occur within and between species, each caused by separate genetic mechanisms. However, in most cases, roaning can be considered as a dominant, epistatic trait that is found in several different mammals. Broadly, roan is characterized by a coat consisting of both unpigmented and pigmented hairs mixed like salt and pepper, spanning the entirety of the body or specific to some region(s) of the body.

2.1 Roan and Roan-like Phenotypes in Horses

In horses, classic roan is found predominantly in Quarter Horses but has been identified in a wealth of other breeds. The classic roan phenotype consists of intermixed white hairs with

pigmented hairs across the horse's body with fully pigmented hairs on the head, legs below the knee, mane, and tail. This is commonly called “classic roan” to distinguish it from similar roan-like phenotypes found in the horse that have distinguishable characteristics, such as varnish roan in Appaloosas and related breeds, as well as the rabicano phenotype described below. In horses with classic roan, the roan ticking rarely continues down the legs past the hock and knees and onto the head. This phenotype may be due to other white variants amplifying the expression of roan. Roan is epistatic, acting on top of the base coat resulting in different “colors” of roan. For example, chestnut horses with classic roan are referred to as a red roan and those with a black base coat are referred to as blue roan. Interestingly, and contrary to other coat colors, injuries on roan horses yield solid colored scars with no white ticking, called “corn spots”, as opposed to other coat colors which produce white, hypopigmented scars (D. Phillip Sponenberg & Bellone 2017; “The Roan Quarter Horse Color” 2018). The molecular mechanism of this abnormality has not yet been identified.

Gray is occasionally confused for roan, as although gray is a progressive phenotype, there may be a stage where a gray horse resembles a roan horse. Gray horses are born a solid color and gradually lose hair pigmentation over several years, frequently resulting in a near all-white phenotype in older horses. The most significant difference between a roan and a gray, if no history of the horse is known, is that gray horses have a homogenous coat color phenotype across the entire body, including the head and legs, whereas classic roan horses typically have fully pigmented head and legs (with the exception of other white markings unrelated to roan).

Gray is the result of a 4.6kb duplication in intron 6 of *syntaxin 17 (STX17)*(Rosengren Pielberg *et al.* 2008; Sundström *et al.* 2012). Two *melanocyte inducing transcription factor (MITF)* binding sites exist in this duplication, resulting in a marked upregulation of *MITF* (Sundström *et al.* 2012). *MITF* plays an important role in the regulation of melanocyte development (Levy *et al.* 2006). Mutations in *MITF* are known to cause deafness and absence of pigmentation due to its role in regulation of three enzymes that play an important role in the

conversion of tyrosine to melanin in melanocytes: tyrosinase (TYR), tyrosinase related protein 1 (TYRP1), dopachrome tautomerase (DCT) (Levy *et al.* 2006; Spritz & Hearing 2013). It is thought that the initial increase in melanocytes due to upregulation of *MITF* gradually depletes the cells available to produce pigment, resulting in progressive loss of pigmentation (D. Phillip Sponenberg & Bellone 2017). Gray horses are also predisposed to vitiligo and melanoma, with black horses most affected (MacKay 2019). 51% of gray Quarter Horses over 15 years of age have been found to be affected by melanoma (MacKay 2019). This number jumps to 78% in gray Lipizzaners over the age of 15, potentially due to the high prevalence of the black base coat color within the breed (MacKay 2019). Additionally, horses homozygous for the 4.6kb *STX17* deletion turn grey faster than heterozygotes, and are more likely to develop melanoma and vitiligo (D. Phillip Sponenberg & Bellone 2017). Also, four variants (one insertion in the promotor, on frameshift mutation, and two large deletions) have been identified in *MITF* that explain splashed white (SW) in some horses: SW1, SW3, SW5, and SW6 (Magdesian *et al.* 2020). The splashed white phenotype is characterized by large white regions on the head, legs, belly and lower neck, blue eyes in many cases, and sometimes deafness (D. Phillip Sponenberg & Bellone 2017). Furthermore, two other splashed white variants have been identified in *paired box 3 (PAX3)*, called SW2 and SW4.

Additional roan-like phenotypes in horses include frost, sabino, varnish roan, and rabicano. Sabino is an incompletely dominant phenotype resulting in regions of complete hypopigmentation. Typical heterozygous sabino horses have a large white face marking, partial to full white legs, and white on the belly, wrapping upwards around the barrel. Additionally, the white on the barrel may disperse into white ticking continuing upwards. Minimal heterozygotes may have white markings only on the lower legs with an inconspicuous face marking thought to be due to an unknown modifier. Homozygous sabino horses have extensive white markings, with a hair coat at least 90% white. The sabino phenotype is the result of a SNP 13 base pairs (bp) upstream of exon 17 of the *proto-oncogene c-KIT (KIT)*, SNPAX-103727726

(ENSECAT00000014185.3:c.2686-13A>T), resulting in altered splicing and a loss of exon 17 (Brooks & Bailey 2005; Druml *et al.* 2018). This variant has been named "Sabino 1" as it was the first variant identified connected to this phenotype and other phenotypes are thought to occur. The sabino-1 variant is thought to impact the normal function of *KIT*, resulting in a lack of pigment (Brooks & Bailey 2005; Druml *et al.* 2018). *KIT* encodes a transmembrane receptor tyrosine kinase. This receptor plays a major role in melanogenesis, melanocyte proliferation, migration, physiology, and survival (Alexeev & Yoon 2006) (D. Phillip Sponenberg & Bellone 2017; Spritz & Hearing 2013). Mutations in *KIT* are responsible for a host of white spotting phenotypes in horses and other mammals (Artesi *et al.* 2020; Brooks & Bailey 2005; David *et al.* 2014; Picardo & Cardinali 2011; D. Phillip Sponenberg & Bellone 2017; Sun *et al.* 2020; Wong *et al.* 2013)

Varnish roan is a progressive roan coat color primarily found in Appaloosa horses. The progressive depigmentation avoids the bony surfaces of the horse, including parts of the face and prominent leg joints, leaving hair in these areas fully pigmented (Bellone *et al.* 2013). Varnish roan is the result of leopard complex spotting (*Lp*), an incompletely dominant variant that is permissive to allow multiple white patterning in the coat (Bellone *et al.* 2010). The causal variant is a 1,378bp insertion of a long terminal repeat (LTR) of an endogenous retrovirus into intron one of the *transient receptor potential cation channel subfamily M member 1 (TRPM1)* gene (Bellone *et al.* 2013). Additionally, horses homozygous for *Lp* are afflicted by a disorder of the retina, congenital stationary night blindness (CSNB). As the name implies, affected individuals have difficulty seeing in low light conditions. *TRPM1* plays a role in pigmentation by suppressing melanoma metastasis, regulating calcium homeostasis, and decreasing the amount of tyrosine in melanosomes (Setaluri & Jayanthy 2013). Additionally, *TRPM1* plays a critical role in depolarizing ON-bipolar cells of photoreceptors in response to light. Loss of function of ON-bipolar cells due to *TRPM1* mutations is a cause of CSNB in humans; however, no loss of pigment has been identified in these cases (D. Phillip Sponenberg & Bellone 2017; van Genderen *et al.* 2009).

An additional white ticking phenotype is Frost, a milder version of varnish roan characterized by light ticking along the top of the neck, back, and hindquarters, as if the horse bore a light layer of snow (D. Phillip Sponenberg & Bellone 2017). The modifier that causes frost has not yet been identified.

Rabicano, a roan-like phenotype, results in a solid-colored horse with white ticking isolated to the flanks and horizontal white markings at the top of the tail, known as "coontail". Expression of rabicano varies from minimal ticking in the flanks to extensive white ticking up the barrel of the horse and what is known as "barring", denser bands of white along the ribs, and a coontail with multiple white bands around the tail head. Although rare, it is possible for extreme cases of rabicano to be confused with the sabino phenotype.

Proper coat color phenotyping is important for horse identification and registration purposes. Standardization of coat color terminology within breeds, along with genetic testing has helped to correctly phenotype horses. However, previous errors due to overlapping phenotypes among roaned horses make it difficult to study the inheritance based on pedigrees of some of these roan and roan-like patterns. As such, the overlap in phenotype of rabicano with other roan-like patterns often results in the rabicano phenotype improperly classified. Like roan and other ticking phenotypes across species, rabicano is believed to be a dominant trait based on mode of inheritance in some families of horses (D. Phillip Sponenberg & Bellone 2017)., Identifying the genetic cause of rabicano and other white ticking phenotypes could aid in proper coat color classification across breeds.

2.2 Genetic Investigations of Roan Phenotypes in Horses

The genetic cause of classic roan was first reported to be linked to the extension locus (E, now known as the melanocortin 1 receptor gene, *MC1R*) and the serum albumin locus (*Al*) in Belgian horses. Fifty-seven breedings of a Belgian Horse stallion heterozygous for extension (*E/e*) and roan (*Rn/rn*) to eight chestnut mares (*e/e rn/rn*) resulted in a recombination rate of 0.035 ±

0.024 between the extension and roan genotypes (D P Sponenberg *et al.* 1984). Having narrowed down the location of the *Rn* locus to the region around *MC1R*, these findings opened the doors to further research. In 1999, Marklund *et al.* identified a strong association between the classic roan phenotype and *KIT*, and although associated variants were found, no causative mutation was identified (Marklund *et al.* 1999).

A haplotype within and downstream of *KIT* was identified in Noriker horses, Quarter Horses, Slovenian Coldbloods, Murgese, and a Belgian Draught horse. This haplotype was not found in classic roan Shetland Ponies, German Sport Horses, or a Trakehner (Grilz-Seger *et al.* 2020). The lack of the same haplotype in all classic roan horses may suggest multiple causal mutations for the classic roan phenotype, or simply a loss of linkage between the haplotype and the causal mutation in some breeds.

For decades, classic roan was believed by many to be homozygous lethal due to the ratio of roan: non-roan offspring (Marklund *et al.* 1999; D P Sponenberg *et al.* 1984). This question of homozygous lethality was investigated in Icelandic horses, where an analysis of 67 roan x roan matings established a ratio of 4.6:1 of roan:non-roan foals (Voß *et al.* 2020). These findings were significantly different from the 2:1 roan:non-roan ratio that was expected if roan was indeed homozygous lethal, suggesting that homozygous roan horses are indeed viable (Voß *et al.* 2020). Additional *KIT* variants associated with roan were identified in Icelandic horses and although the causative mutation remains elusive, it is probable that classic roan in horses is caused by a yet unidentified variant(s) impacting the function of *KIT* (Voß *et al.* 2020).

Although the genetic basis of classic roan is not fully elucidated, an associated haplotype was identified by the University of California, Davis Veterinary Genetics Laboratory and is offered as a test for horse owners and breeders (UC Davis Veterinary Genetics Laboratory n.d.). Data from this commercial test may one day assist in the discovery of the genetic cause of roan.

2.3 Genetic Investigation of Roan in Cattle

The roan phenotype has also been identified in cattle, exhibiting white hairs across the body and head, a white belly, and solid colored legs. This roan phenotype varies from light roan (many white hairs and few pigmented hairs) to dark roan (many pigmented hairs and few white hairs) depending on modifiers inherited from the parents (Barrington & Pearson 1906). A causative mutation was identified following almost a century of work on the inheritance of roan in cattle. It was first investigated in Shorthorn cattle, starting with a cross of white Shorthorns and black Galloways, two breeds and colors that breed nearly true (Barrington & Pearson 1906). This resulted in an F1 generation of "blue-grey" cattle, suggesting a mode of simple mendelian inheritance (Barrington & Pearson 1906). Pedigree records and interviews with breeders of blue-grey cattle supported this hypothesis, leading to the conclusion that roan is a codominant trait in Shorthorn and Belgian Blue cattle (Barrington & Pearson 1906; Laughlin 1911; Wilson 1908). In 1995, microsatellites were used to link the roan locus in Belgian Blue cattle to the same chromosome segment that contains the steel (*sl)* locus in mice, a locus encompassing the mast cell growth factor gene (*MGF*), also known as *KIT* Ligand (*KITLG*) (Charlier *et al.* 1996). The causative mutation, a non-conservative missense variant in the seventh exon of *KITLG* (AC_000162.1:g.18339001C>A, ENSBTAT00000023349.5:c. c.572C>A, NM_174375.2:p.(Ala191Asn),was identified in 1999 by Seitz, *et al* (Jansen *et al.* 2013; Seitz *et al.* 1999). This codominant variant results in heterozygous individuals with a roan phenotype and homozygous alternate individuals with an all-white phenotype (Seitz *et al.* 1999).

KITLG works by binding to the *KIT* receptor to initiate the *KIT* signaling pathway, causing a cascade of events critical to melanogenesis (Alexeev & Yoon 2006). Disruption of *KITLG* inhibits its ability to bind to the KIT receptor which affects skin and hair pigmentation, melanocyte proliferation, melanin synthesis and distribution, and activation of *MITF* (D'Mello *et al.* 2016; Picardo & Cardinali 2011; Spritz & Hearing 2013). The missense variant associated with roan in cattle replaces a hydrophobic residue with an acidic one. This change is expected to alter the

hydrophobicity and the α-helix structure of this domain of the protein. It is possible that this mutation results in a structural change to the ligand, impairing ligand binding and disrupting the *KIT* signaling pathway (Seitz *et al.* 1999).

2.4 Genetic Investigation of Roan in Goats

The roan coat color has also been identified in goats, exhibiting the classic mix of white and pigmented hairs throughout the body with phenotypic variation ranging from almost white to almost completely pigmented, similar to that of cattle (Talenti *et al.* 2018). Unlike horses, roan goats commonly have the roan pattern on their head and face but they retain the solid colored legs and some retain a solid colored band of hair across the neck and spine. Recently, an analysis comparing 35 roan Pakistani goats of the Barri and Beetal Muki Cheni breeds to 740 non-roan Pakistani and Italian goats of 41 different breeds deduced that the causative mutation of roan in goats likely impacts *KITLG* (Talenti *et al.* 2018). This analysis utilized 47,406 autosomal SNPs to identify runs of homozygosity (ROH) associated with the roan phenotype in goats. A 1.7Mb ROH was identified on chromosome 5, with an H score of 0.71, the highest H value identified. Haplotype analysis identified a region overlapping the ROH of interest and *KITLG*, suggesting that this region is under strong selection in roan goats (Talenti *et al.* 2018). Although additional analyses are necessary to confirm the association of roan in goats to *KITLG* and identify the causal mutation, these data provide compelling evidence that *KITLG* may be responsible for the roan coat color in goats.

2.5 Genetic Investigation of Roan in Dogs

A dominant roan phenotype has also been identified in dogs, namely in Australian Cattle Dogs and Wirehaired Pointing Griffons. Notably, the definition of roan in dogs differs from that of roan in horses, cows, pigs, and mink. Roan in dogs is epistatic to existing solid white markings, such as piebald, causing the individual to present a homogenous mix of pigmented hairs within an otherwise unpigmented region (D P Sponenberg & Rothschild 2001). As such, presence of the

roan genotype cannot be identified in solid colored dogs. Notably, “ticking” is an additional coat color in dogs. Ticking is similar to roan, in that pigmented hairs are found in existing white regions, however, ticking differs in that the solid colored hairs group into small spots within the white region (D P Sponenberg & Rothschild 2001). Finally, flecking is a recessive modifier that removes white hairs from tick marks, resulting in the Dalmatian phenotype (D P Sponenberg & Rothschild 2001). Without this modifier, ticking spots are more muddled and roan-like (D P Sponenberg & Rothschild 2001). Dogs can be both roan and ticked and the pigmented hairs present in roan, ticking, and flecking are the same color as the existing base coat color (Kawakami *et al.* 2021).

An 11.4kb tandem duplication (NC_006620.3:g.11131835_11143237dup) in an intronic region of the usherin gene (*USH2A*) was recently identified as the likely causative mutation of roan in dogs (Kawakami *et al.* 2021). This region contains three clusters of highly conserved sequences and overlaps with transcription binding sites and DNAse I hypersensitive sites. *USH2A* encodes for the usherin protein and is commonly associated with progressive hearing and vision loss, including retinitis pigmentosa, in humans (Kawakami *et al.* 2021; Liu *et al.* 2007). Mutations in *USH2A* can impact stereocilia in the ear and connecting cilium in photoreceptors expressed primarily in the neural retina, resulting in deafness and progressive vision loss (Liu *et al.* 2007). As no association between *USH2A* and hair or skin pigmentation has been reported, Kawakami *et al.* hypothesize that the 11.4kb tandem duplication in *USH2A* in dogs may result in ectopic expression in skin melanocytes or creates a novel isoform due to alternative splicing (Kawakami *et al.* 2021). The roll of *USH2A* in human deafness may also explain the high prevalence of congenital deafness in breeds of dogs with heavy roaning and Dalmatians, but this hypothesis has yet to be investigated (Kawakami *et al.* 2021).

Additionally, the 11.4kb duplication was not associated with the ticking phenotype, suggesting that this is not the only variant that causes roan-like phenotypes in dogs. This 11.4kb tandem duplication was also found in Dalmatians, all of whom had the same duplication or a

similar haplotype. It has been suggested that the Dalmatian (non-flecked) phenotype is associated with an intronic SNP (NC_006585.3:g.72316930G>A) in the Ras Homolog Family Member H (*RHOH*) gene (Kawakami *et al.* 2021). An analysis of 262 Dalmatians found 98% to be homozygous for this variant. No wild type Dalmatians were identified (Kawakami *et al.* 2021). This variant was also found in other breeds with Dalmatian-like spots, as well as four dogs without Dalmatian-like spots (Kawakami *et al.* 2021).

Although not associated with a roan phenotype, a recently discovered 6kb copy number variant (CNV), NC_006597.3*:g.29821450_29832950[2_?],* located 152kb upstream of *KITLG* was identified in dogs (Weich *et al.* 2020). This CNV is responsible for differences in pheomelanin and eumelanin pigment intensity in hair (Weich *et al.* 2020). Again, supporting the role of *KITLG* in melanin distribution and suggesting that the presence of this copy number variant affects pigmentation intensity.

2.6 Genetic Investigation of Roan in Pigs

Roaning has also been found in pigs, namely in an Italian local grey pig population and other mixed breed pigs. Using the word "grey" to refer to roan pigs appears to be a result of terminology in different languages, as in French they are called grey, Fontanesi *et al.* refers to them as grey-roan, and in English they are called roan (Fontanesi *et al.* 2010; Lauvergne & Canope 1979). Roan pigs can be identified by their mixture of white and colored hairs, as one would expect. Two different mutations in *KIT* have been associated with roaning in pigs. The first, a dominant 4bp deletion in intron 18 (NC_010450.4:g.41488583_41488586del) was identified in a local Italian grey pig population (Fontanesi *et al.* 2010). All 41 grey pigs investigate carried at least one copy of this deletion and all solid colored pigs were homozygous for the reference allele. Additionally, 123 pigs from three all-white breeds (Italian Large White, Italian Landrace, and Belgian Landrace) were found to be heterozygous for this deletion except for a one pig that was homozygous reference. Also, almost all of the white pigs had two previously identified dominant

white variants, a ~450kb duplication encompassing the entirety of *KIT* and a splice site mutation causing exon 17 to be skipped (Fontanesi *et al.* 2010). The remaining two white pigs had the duplication but not the splice site mutation (Fontanesi *et al.* 2010). All grey pigs had the homozygous reference genotype for these variants (Fontanesi *et al.* 2010).

Shortly after this 4bp deletion was identified, a separate *KIT* variant was discovered in a family of roan Landrace and Korean Native pigs (Lim *et al.* 2011). A U_{26} repeat with a G interruption in intron 5 that causes exon 5 to be skipped was identified as the most likely causative variant. The predicted impact of skipping exon five is a loss of 169 base pairs, shifting the reading frame and resulted in a truncated trans-membrane and intracellular domains in the resulting KIT protein. It was noted that the roan individuals with this variant also had the previously discovered 4bp deletion in intron 18 due to a long-range linkage disequilibrium (LD) block (Lim *et al.* 2011). More research is needed to determine which of these variants is causal for roan in pigs.

2.7 Genetic Investigation of Roan in Mink

The roan phenotype has also been identified in American Mink. Although wild mink are solid colored (non-roan), three distinct roan phenotypes have been identified in the captive population. The "cross" phenotype is a codominant trait similar to the roan phenotype seen in Shorthorn and Belgian Blue cattle. These individuals vary from solid black (no roan variant) to a black and white roan-like pattern (heterozygous) to a 95% white (thought to be homozygous for a roan variant). Meanwhile, the "stardust" phenotype is similar to that of classic roan in horses. Stardust has a dominant pattern of inheritance, with the color dependent on the base color of the individual. A third phenotype, "cinnamon," is a recessive trait of a grayish-brown and white roan with a white undercoat. No causative mutations for any of these roan patterns have been identified to date. To investigate the genetic cause of these three coat colors, six candidate genes selected due to phenotypic similarities in other species and implicated effect on pigmentation (*ATOH1, KIT, KITLG, MITF, POMC, SLC24A5) (Anistoroaei et al. 2012).*Using a microsatellite-based

approached, two microsatellite markers were developed for each candidate gene, ten of which were polymorphic in the families of interest (Anistoroaei *et al.* 2012). Co-segregation was expected between each phenotype and a marker(s), but was not observed (Anistoroaei *et al.* 2012). As such, *KIT*, *ATOH1*, and *POMC* were excluded as causal genes for all three phenotypes *(Anistoroaei et al. 2012)*. Additionally, *KITLG* was excluded from the stardust phenotype, and *MITF* and *SLC24A5* were excluded from the cinnamon phenotype *(Anistoroaei et al. 2012)*. All three (*KITLG, MITF, and SLC24A5)* were excluded as candidates for the cross phenotype (Anistoroaei *et al.* 2012). Finally, exclusion of *MITF* and *SLC24A5* for stardust was not possible due to insufficient proof to be excluded, and *KITLG* was unable to be excluded from cinnamon due to inconclusive markers (Anistoroaei *et al.* 2012). The role of regulatory elements on roan-like and other pigmentation phenotypes, suggests that a regulatory region could play a role in the genetic basis of rabicano, either due to a structural variant (SV) (such as in dogs) or a SNP (such as in humans). Identifying the genetic cause of rabicano could provide valuable information to aid further investigations into roan in goats and mink.

2.8 Additional *KITLG* Hair Pigmentation Variant in Humans

In humans, a SNP (rs12821256) associated with blonde hair was identified 350kb upstream of *KITLG* (Sulem *et al.* 2007). Although one explanation for this finding was that the SNP is associated with a haplotype in linkage disequilibrium with the causal mutation, additional research confirmed that this SNP impacts a highly conserved nucleotide which alters a lymphoid enhancer binding factor 1 (*LEF1*) transcription factor (Guenther *et al.* 2014; Sulem *et al.* 2007).Thus this variant impacts enhancer activity and alters hair pigmentation (Guenther *et al.* 2014).

These functional data from humans suggest that some of the difficulty in identifying the cause of roan and roan-like phenotypes in other mammals may be due to the causal mutation being located in a regulatory region, distant from the transcriptional start site of the most obvious

candidate genes *KIT* or *KITLG*. Better characterization of regulatory regions, such as the work being done by the Functional Annotation of Animal Genomes (FAANG) project may aid in identifying these variants.

2.9 Conclusion

To summarize, a trend can be seen between the phenotype of roan and association with *KIT* and *KITLG*, most notably in cows, goats, pigs, and classic roan horses. Similarity between these phenotypes and rabicano suggest that *KIT* and *KITLG* may be valuable functional candidate genes in the search for the genetic cause of rabicano in horses.

Chapter 2: Analysis of Genetic Diversity in the American Standardbred Horse Utilizing Short Tandem Repeats

An Excerpt from Esdaile, E., Avila, F., Bellone, RR. (2021) ***Analysis of Genetic Diversity in the American Standardbred Horse Utilizing Short Tandem Repeats and Single Nucleotide Polymorphisms.*** **Manuscript submitted for publication.**

1 Introduction

The American Standardbred was developed during the 19th century as a harness racing horse breed that competes at one of two different gaits: the trot (a symmetrical two-beat, diagonally opposed gait) or the pace (a symmetrical two-beat, lateral gait). After the formation of several Standardbred associations in the early 20th century, the United States Trotting Association (USTA) was established in 1938 as the unifying Standardbred registry in the United States (McCarr 1972). The establishment of this new registry included a requirement that all horses must be registered with the USTA to be eligible to race (McCarr 1972). The USTA closed their studbook in 1973, limiting registration to offspring of already registered American Standardbreds (Adelman 1981; Cothran *et al.* 1986).

Closing a studbook limits the introduction of new alleles into the gene pool and thus can impact genetic diversity and the health of animals. Furthermore, commonly utilized breeding strategies to improve performance can also impact levels of genetic diversity, namely line-breeding and inbreeding. Such strategies can have adverse effects on the health of the breed due to inbreeding depression (Adams *et al.* 2016; Yue *et al.* 2015). Inbreeding depression caused by selective pressures was found to increase the genetic load of modern horse breeds (Orlando & Librado 2019).

In the 1980s, inbreeding was analyzed in American Standardbreds using pedigree analysis of 14 generations. These studies support differences in inbreeding between trotters and pacers, with trotters showing higher levels of inbreeding (Cothran *et al.* 1984, 1986; MacCluer *et al.* 1983). Additionally, in analyzing short tandem repeats (STRs), increased expected heterozygosity was associated with decreased inbreeding values calculated from pedigree analysis in pacers ($p<.001$).

A statistically significant association between higher expected heterozygosity and higher foaling and conception rates was also noted in pacers ($p<.05$) (Cothran *et al.* 1986; MacCluer *et al.* 1983). To further evaluate genetic diversity within the breed, an investigation was conducted from 1971 to 1989 using ten blood markers from 4,404 trotters and 12,271 pacers. These data determined the mean observed total heterozygosity in both trotters and pacers to be statistically significantly decreasing over time, more so in trotters than pacers (King 1992). The author hypothesized that such a decrease might be attributed to the closed breeding structure and the significant impact a single sire can have on a breed. These results further supported past research that found that pacers were less inbred than trotters (Cothran *et al.* 1984, 1986; MacCluer *et al.* 1983).

Due to the reported decrease in heterozygosity, the closure of the studbook, and a concern for potential health risks due to inbreeding depression, the USTA imposed a studbook limit of 140 mares covered per year for all trotting stallions debuting in 2009. Similarly, a studbook limit of 160 mares covered per year was imposed on pacing sires who debuted in 2009, decreasing to 150 for those who debuted in 2010 and 140 for those who debuted in 2011 or later. It is unknown how studbook restrictions have impacted diversity or if they have mitigated further loss of heterozygosity. To date, the only single nucleotide polymorphism (SNP)-based study to investigate diversity in Standardbreds utilized data from the 50K SNP array to compare diversity indices of 15 American Standardbreds (seven trotters and eight pacers), 25 Norwegian Standardbreds (trotters), in addition to 35 other horse breeds. Unlike Norwegian Standardbreds, American Standardbreds were found to have significant excess homozygosity (F_{IS}), which could be attributed to increased inbreeding in the cohort examined or substructure based on gait (Petersen *et al.* 2013).

Given the previous studies from the 1980s, which highlighted differences in inbreeding of trotting and pacing populations when analyzing STRs, excess homozygosity observed in a small sample of American Standardbreds using SNP data, and the concern for the impact of book size on diversity indices, we aim to evaluate genetic diversity in the American Standardbred using

STRs after the book size limit was imposed. We also aim to evaluate an extensive data set comprising 50,621 American Standardbreds to further investigate whether pacers are more genetically diverse than trotters (MacCluer *et al.* 1983). Finally, we hypothesize that dams are maintaining genetic diversity in American Standardbreds.

2 Methods & Materials

2.1 Data Collection

To establish baseline values of genetic diversity indices to monitor trends over time and the potential effects of the 2009 studbook cap, we evaluated the breeding stock and first six foal crops after the rule was imposed (2010-2015). All foals ("offspring") born between 2010 and 2015, as well as their sires ("sires"), and dams ("dams"), were genotyped by Bureau Veritas Laboratories as part of the routine parentage verification process for registration with the USTA and data was provided for use in this study.

2.2 Data Analysis

Records from 191 pacing dams, 77 trotting dams, 244 pacing offspring, and 40 trotting offspring were removed from the analysis due to duplicate samples and incomplete records, leaving 50,621 horses in our analyses. Horses were grouped according to gait (trotters or pacers) and designated subpopulations (sires, dams, and offspring) (Table 1). To compare the impact of book size within and between gait types, the sires and offspring were divided into three groups based on both the total number of offspring for the years under investigation and sample sizes within each group. Groupings were divided based on approximately 1/3rd of the total offspring and categorized as "high-book", "mid-book", and "low-book", accordingly. The highest producing sires, those that cumulatively sired 40% of trotters and 32.9% of pacers over the six years were classified as "high-book sires". Sires that cumulatively sired 27. 6% of trotters and 33.9% of pacers were classified as "mid-book sires", and the remaining, lowest producing sires were classified as

"low-book sires" (Table 1). Offspring were classified by sire book size in addition to year of birth (Table 1).

Table 1:Groupings of American Standardbreds foaled from 2010-2015 and their sires and dams utilized to investigate genetic diversity within the breed. The percentage is the size of the subgroup as compared to the respective group. Trotters: high-book sires ranged from an average of 70.7 to 107 offspring per year (=78.3), mid-book sires ranged from an average of 22 to 66.8 offspring per year (=36.9), low-book sires ranged from an average of 0.3 to 21.2 offspring per year (=5.9). Pacers: high-book sires ranged from an average of 82.3 to 135.7 offspring per year (=101.6), mid-book sires ranged from an average of 34.5 to 72.5 offspring per year (=52.3), low-book sires ranged from an average of 0.7 to 33.5 offspring per year (=9.3).

Population Group *Subgroup*	**Number of Trotters** *(percent of group type)*	**Number of Pacers** *(percent of group type)*
All	**22,731**	**27,890**
Dams	7,314	9,260
Sires	171	140
High Book Sires	13 *(7.6%)*	10 *(7.1%)*
Mid Book Sires	19 *(11.1%)*	20 *(14.3%)*
Low Book Sires	139 *(81.3%)*	110 *(78.6%)*
Offspring	15,246	18,490
Offspring of High Book Sires	6,105 *(40%)*	6,078 *(32.9%)*
Offspring of Mid Book Sires	4,206 *(27.6%)*	6,259 *(33.9%)*
Offspring of Low Book Sires	4,935 *(32.4%)*	6,153 *(33.3%)*
2010 Offspring	2,728 *(17.9%)*	3,284 *(17.8%)*
2011 Offspring	2,883 *(18.9%)*	3,497 *(18.9%)*
2012 Offspring	2,814 *(18.5%)*	3,337 *(18.0%)*
2013 Offspring	2,544 *(16.7%)*	2,956 *(16.0%)*
2014 Offspring	2,287 *(15.0%)*	2,886 *(15.6%)*
2015 Offspring	1,990 *(13.1%)*	2,530 *(13.7%)*

Sixteen STR loci commonly used for routine parentage testing were analyzed: AHT4, AHT5, ASB17, ASB2, ASB23, CA425, HMS1, HMS2, HMS3, HMS6, HMS7, HTG10, HTG4, HTG7,

LEX33, and VHL20. STR data was formatted for analysis with Microsatellite Analyser 4.05 (MSA) (Dieringer & Schlötterer 2003). For each marker, MSA was used to calculate allelic richness, expected heterozygosity, observed heterozygosity, inbreeding coefficient, and pairwise fixation index (F_{ST}) using 20,000 permutations (Weir & Cockerham 1984). Mean allelic richness (A_r), mean expected heterozygosity (H_E), mean observed heterozygosity (H_O), and mean inbreeding coefficient (F_{IS}) were then calculated by averaging the values of each locus. T-test comparisons of H_E, H_O, F_{IS}, and A_r, between and within populations, as well as linear models (lm) of offspring foaled from 2010-2015, were calculated using R version 4.0.3 (R Core Team 2020). A Bonferroni corrected significance level of <.05 was used to determine the significance of all t-tests ($p_{Bonferroni} = 1 - (1 - x)^n$).

3 Results

3.1 Allelic Richness (A_r)

In both trotting and pacing subpopulations, high-book sires and mid-book sires had significantly lower A_r than low-book sires (trotters A_r = 4.4 and 5.0; t-test: $p_{Bonferroni}$ = 1.6x10-5 and 9.3x10-3; pacers A_r = 4.0 and 4.7; t-test: $p_{Bonferroni}$ < 0.001 and 0.017) (Table 2). Sires as a whole had lower A_r when compared to dams and offspring (trotters A_r = 5.9, 8.1, 8.2; t-test: $p_{Bonferroni}$ = 4.8x10-3 and 0.034; pacers A_r = 6.0, 8.4, 8.9; t-test: $p_{Bonferroni}$ = 2.4x10-3 and 6.9x10-6). Even when parsed by book size, offspring had higher A_r than their respective sire groups (A_r - high-book sires and their offspring; trotters = 4.4 and 7.6, pacers = 4.0 and 7.9, mid-book sires and their offspring; trotters = 5.0 and 7.6, pacers = 4.7 and 7.8, low-book sires and their offspring; trotters = 5.9 and 7.4, pacers = 5.9 and 7.9; all $p_{Bonferroni}$ < 0.02). A_r of offspring did not differ when parsed by stallion book size or offspring year of birth (t-tests: $p_{Bonferroni}$ > 0.05). A_r did not differ when comparing respective groups of trotters to pacers, including the comparison of all trotters as compared to all pacers (t-test: p > 0.05)(Table 2).

Table 2: Measures of diversity in American Standardbreds calculated using 16 STRs. Presented are allelic richness (Ar), expected and observed heterozygosity (H_E and H_O), inbreeding coefficient (F_{IS}), and pairwise fixation index (F_{ST}) values of trotting and pacing Standardbreds foaled from 2010-2015 and their sires and dams.

Population Group *Subgroup*	**Trotters**				**Pacers**				**Pairwise F_{ST} of Trotters and Pacers**
	A_r	**H_E**	**H_O**	**F_{IS}**	**A_r**	**H_E**	**H_O**	**F_{IS}**	
All	8.63	0.665	‡ 0.675	-0.014	9.13	0.669	‡ 0.677	-0.012	0.076
Dams	ABC 8.06	AB 0.674	‡, A 0.684	AB -0.015	ABC 8.38	ABCDEFGHIJK 0.680	‡, ABCDEFGHI 0.692	-0.017	0.066
Sires	CLMNOPQRSTU 5.94	B 0.654	0.678	-0.041	CLMNOPQRSTU 6.00	A 0.657	0.676	-0.029	0.092
High Book Sires	BDEFG 4.44	0.638	0.680	-0.095	BDEFG 4.00	0.654	0.718	-0.114	0.105
Mid Book Sires	AHIJK 5.00	0.673	‡ 0.741	A -0.116	AHIJK 4.69	0.632	0.646	-0.041	0.076
Low Book Sires	ABDEFGHIJK 5.88	AC 0.653	0.669	-0.029	ABDEFGHIJK 5.94	CL 0.661	0.678	-0.026	0.093
Offspring	GKL 8.19	0.660	‡ 0.670	-0.015	GKL 8.88	B 0.663	‡, A 0.670	-0.012	0.082
Offspring of High Book Sires	DIO 7.56	0.645	‡ 0.668*	BCD -0.036	DIO 7.94	D 0.655	‡ 0.670	A -0.023	0.099
Offspring of Mid Book Sires	EHN 7.63	0.665	0.672	D -0.011	EHN 7.75	E 0.652	‡, B 0.661	-0.015	0.080
Offspring of Low Book Sires	FJM 7.44	C 0.667	A 0.671	AC -0.006	FJM 7.94	AL 0.672	‡, C 0.679	A -0.011	0.077
2010 Offspring	P 7.50	0.660	0.665	-0.006	P 7.69	F 0.665	‡, D 0.675	-0.016	0.080

Population Group *Subgroup*	**Trotters**				**Pacers**				**Pairwise F_{ST} of Trotters and Pacers**
	A_r	**H_E**	**H_O**	**F_{IS}**	**A_r**	**H_E**	**H_O**	**F_{IS}**	
2011 Offspring	Q 7.31	0.660	0.664	-0.005	Q 7.38	G 0.660	‡, E 0.669	-0.014	0.082
2012 Offspring	R 7.13	0.663	0.667	-0.005	R 7.88	H 0.661	F 0.666	-0.009	0.083
2013 Offspring	S 7.25	0.661	0.665	-0.004	S 7.50	I 0.664	G 0.670	-0.010	0.081
2014 Offspring	T 7.19	0.658	0.664	-0.008	T 7.38	J 0.663	H 0.669	-0.011	0.083
2015 Offspring	U 7.19	0.657	0.664	-0.010	U 7.38	K 0.662	‡, I 0.670	-0.012	0.084

Unique letters in the upper right corner of each cell denote all statistically significant pairwise comparisons within a column based on a t-test and $p_{Bonferroni} < .05$. ‡ = signifies when H_O was statistically significantly different from H_E.

3.2 Expected Heterozygosity (H_E)

Trotting and pacing dams had higher average H_E as compared to sires (H_E trotters = 0.67 and 0.65; H_E pacers = 0.68 and 0.66; t-test: $p_{Bonferroni}$ = 0.03 and <0.001, respectively). Pacing dams had statistically significantly higher H_E than that of offspring (H_E = 0.68 and 0.66, respectively; t-test: $p_{Bonferroni}$ = 0.002) (Table 2). Offspring of low-book sires had significantly higher H_E than their sire group (H_E trotters = 0.67 and 0.65; H_E pacers = 0.67 and 0.66; t-test: trotters, $p_{Bonferroni}$ = 0.046, pacers $p_{Bonferroni}$ = 0.024). No significant difference in H_E was found when comparing trotters and pacers, both as a whole and by group. (t-test: $p_{Bonferroni}$ > 0.05)(Table).

3.3 Observed Heterozygosity (H_O)

H_O of pacing dams was significantly higher than that of offspring as a whole and offspring of mid- and low-book sires (H_O = 0.69, 0.67, 0.66, and 0.68; t-test: $p_{Bonferroni}$ < 0.001, 3.8x10-3, 0.012, respectively). In trotters, only H_O of dams was significantly higher than that of offspring of low-book sires (H_O = 0.68 and 0.67, respectively; t-test: $p_{Bonferroni}$ = 3.2x10-3). No other comparisons of H_O were significantly different among groups (Table 2).

3.4 H_O and H_E

H_O was statistically significantly higher than H_E in both trotting and pacing horses as a whole, dams, total offspring, and offspring of high-book sires. It was also statistically significantly higher in pacing offspring of mid and low-book sires, pacing offspring foaled in 2010, 2011, and 2015, and trotting mid-book sires (Table 2, t-test: $p_{Bonferroni}$ > 0.05).

3.5 Inbreeding Coefficients (F_{IS})

Estimated inbreeding coefficients (F_{IS}) below zero were observed for all groups (Table 2). In trotters, offspring of high-book sires had lower F_{IS} than offspring of mid-book sires, offspring of low-book sires, and dams ($p_{Bonferroni}$ = 3.3x10-5, 3.7x10-4, and 4.5x10-3 respectively) (Table 2). In

trotters, dams had higher F_{IS} than mid-book sires ($p_{Bonferroni}$ = 0.045). In pacers, only offspring of high-book sires had significantly lower F_{IS} than offspring of low-book sires ($p_{Bonferroni}$ = 0.03). No significant differences in F_{IS} were observed when comparing respective trotters and pacers as a whole or by respective groups (t-test: $p > 0.05$).

3.6 Fixation Index (F_{ST})

Moderate genetic differentiation ($0.05 < F_{ST} < 0.15$)(Wright, 1978) was found when comparing respective groups of trotters to pacers using pairwise F_{ST} (Table 1). Dams had the lowest pairwise F_{ST} between respective groups (F_{ST} = 0.066). High-book sires had the highest pairwise F_{ST} between respective groups (F_{ST} = 0.11), with their offspring not far behind (F_{ST} = 0.099). Within the trotting and pacing subpopulations, pairwise F_{ST} was below 0.05 in all intergroup comparisons, indicating little to no genetic differentiation within gait subgroups.

3.7 Longitudinal Analysis

Pacing dams had statistically significantly higher H_E and H_O than their respective offspring when grouped by year ($p_{Bonferroni} < 0.029$), while pacing and trotting sires had lower A_r than offspring grouped by year ($p_{Bonferroni} < 0.031$) (Supplementary Table 2). When offspring were parsed by year, A_r, H_E, and H_O followed a downward trend across the six years investigated, but no statistically significant differences were detected (Figure 1). Although still below zero, F_{IS} of trotting offspring experienced an overall decrease across the six years, but increased in pacing offspring. Across this same period, pairwise F_{ST} between offspring of trotters and pacers increased, indicating continued differentiation between the groups. There were no statistically significant changes in any of the metrics examined across the six years investigated ($p > 0.05$) (Figure 1*)*.

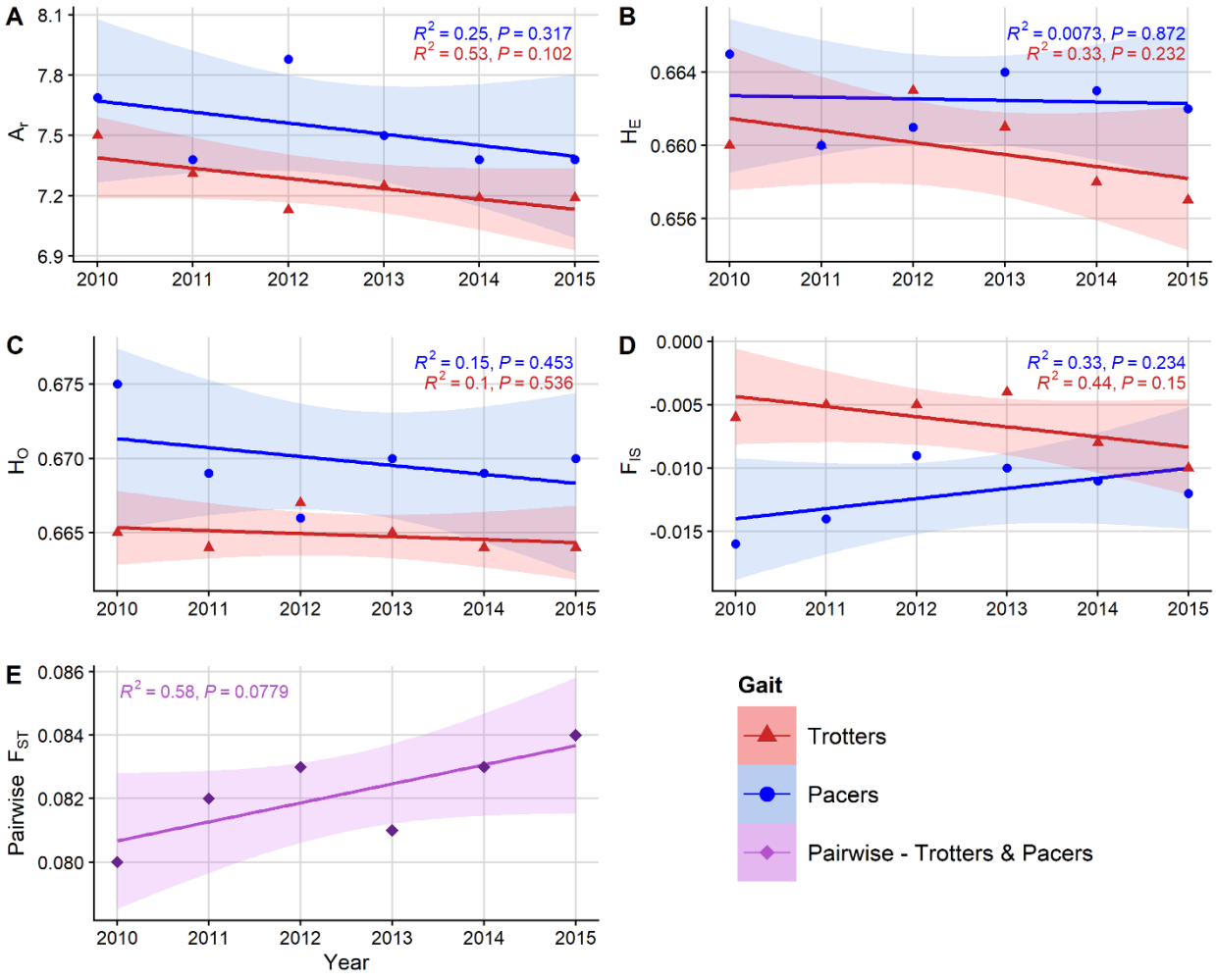

Figure 1: Trends in diversity measures of American Standardbred horses foaled from 2010 to 2015. A_r (A), H_E (B), and H_O (C) trended downward, suggesting ongoing loss of diversity within the breed. F_{IS} (D) differed between groups, increasing in pacers and decreasing in trotters. Pairwise F_{ST} (E) trended upwards, suggesting ongoing differentiation between the two subpopulations.

4 Discussion

In this study, genotyping records of 16 STRs in 50,621 American Standardbreds were used to investigate genetic diversity metrics in the breed across six years from 2010 to 2015. Additionally, STRs were utilized to analyze the contribution of dams to within-breed diversity as well as diversity between pacers and trotters. The application of these markers is widespread in population genetic studies due to their robustness to discern genetic differences between

individuals using a limited number of highly polymorphic loci and low-quality input DNA. Furthermore, because STR genotypes are readily available for every registered Standardbred, this approach allows for an unbiased investigation of diversity in the population as a whole in a reproducible and cost-effective way. In contrast, high-density SNP genotyping is rapidly evolving as an alternative strategy to investigate more global genetic diversity, albeit with increased costs.

Analyses utilizing STRs suggested a lack of allelic richness (A_r) in Standardbred sires as compared to offspring and dams, indicating lower diversity in the sire population. When parsed by percentage contribution of offspring, high-book and mid-book sires demonstrated lower A_r than low-book sires. While it is possible that this difference can be attributed to a smaller number of horses in the high-book sire (n=13 trotters and 10 pacers) and mid-book sire (n=19 trotters and 20 pacers) groups, these may represent true differences in diversity among the groups, and this metric should continue to be monitored to make comparisons across generations.

H_E of dams was significantly higher than sires and pacing offspring, while offspring as a whole were not significantly different from sires. However, offspring of low-book sires had significantly greater H_E as compared to their own sire group. A statistically significant difference in H_E between the offspring of high- and mid-book sires and their respective sire groups was not detected. This may be due to the short time period examined or small group size, which only includes 23 and 39 high and mid-book sires, respectively. The relatively small number of markers evaluated by the STR analysis could have also hindered the ability to identify small differences in H_E between sire offspring groups.

Significantly higher H_O as compared to H_E estimates were detected using STRs, suggesting that genetic diversity is being maintained in American Standardbreds over time. Further, high levels of diversity observed by higher values of H_O as compared to H_E are supported by negative inbreeding coefficients, with both metrics indicating that in the years studied, the breeding practices utilized may be maintaining diversity in the population.

Although they did not reach significance, the opposing trends of F_{IS} in trotters and pacers should be further investigated by a longer longitudinal analysis using a denser set of markers. The trend of decreasing values of A_r, H_O, and H_E in trotting and pacing offspring from 2010 to 2015 support the need for continued monitoring in future generations and comparison to previous generations. As although the trend was not statistically significant, that may be due to the short time period investigated and it may reach significance over ten or twenty years. Although the trend was not statistically significant, it is likely a reflection of the interval evaluated, less than one generation, therefore continuing to monitor this trend will be important to understand how book size has impacted diversity.

In support of this, from 2010 to 2015, a decreasing number of sires contributed to an increasing percentage of the annual foal crop. In comparison to King's report stating that 60 sires (2%) were responsible for 26% of the 1987 foal crop (King 1992), at most 15 sires (5% of sires) were responsible for the same 26% of annual foal crop each year, from 2010-2015 (13,15,15,14,12, and 10 sires, respectively). Meanwhile, the top 60 sires were responsible for 66% of the offspring from 2010-2015. This increase in the percent of offspring produced by top sires is concerning because a small sire pool leads to an increase in inbreeding and homozygosity within the breed, which in turn can increase the frequency of recessive genetic diseases having a devastating economic impact (Adams *et al.* 2016). It is also possible that a statistically significant difference was not observed in genetic metrics evaluated over the 6-year time frame because the studbook cap of 140 mares per year stabilized any genetic loss that may have occurred in generations prior. Future studies comparing these data to both previous and future generations will help to gain a better understanding of the impact of inbreeding and book size limits within this closed population.

The higher genetic diversity found in dams compared to other groups (as measured by A_r, H_O, and H_E) may indicate that dams are indeed maintaining the genetic diversity of offspring and counteracting the decreasing number of sires contributing to a larger percentage of the foal crop.

If the broodmare population experiences a significant loss of diversity due to shifts in breeding practices or the continued use of a small number of sires, the diversity of the American Standardbred could be threatened.

When comparing subpopulations by gait, the STR analysis found no significant differences in the amount of inbreeding between the trotting and pacing American Standardbreds. This differs from the results of previous pedigree studies, which did support differences in inbreeding of trotters and pacers (Cothran *et al.* 1984, 1986; MacCluer *et al.* 1983). Nonetheless, the moderate F_{ST} values we found indicate that pacing and trotting groups are indeed differentiating from each other at the genetic level, and more so each year. This shows that although the two subpopulations have similar levels of within-population genetic diversity, they are moderately genetically different from each other. This differentiation was expected as interbreeding between trotting and pacing lines is believed to be scarce.

These data constitute a snapshot of diversity trends in American Standardbreds over a 6-year period and illustrate the value in utilizing available STR data to assist in monitoring population-level trends in diversity. Future SNP-based studies using larger populations of trotting and pacing Standardbreds are required to explore genome-wide diversity in the breed and should help monitor trends over time. Thus, these data provide a basis of breed-wide genetic diversity in American Standardbreds and serve as a reference point for future studies.

Chapter 3: The Search for the Variant Causing Rabicano in the Horse

1 Introduction

Rabicano is a roan-like white spotting pattern in horses consisting of a mixture of white and pigmented hairs in the flanks and white hairs sometimes referred to as "bars" along the top of the tail ("coontail"). Rabicano can be found in American Quarter Horses (QHs) and Arabians and has also been reported in Thoroughbreds and Warmbloods ("The Roan Quarter Horse Color" 2018; Wahler 2011). Rabicano is thought to be a dominant trait, like roan, based on inheritance pattern reported in some families of horses (D. Phillip Sponenberg & Bellone 2017).

Phenotypic variation in the amount of roaning in the flanks is present across breeds, with some horses exhibiting a small amount of ticking (white hairs) in the flanks and tailhead to individuals with ticking extending forward onto the horse's barrel, "barring" (white stripes down the ribs), and extensive coontail present in extreme cases (Figure 3) (D. Phillip Sponenberg & Bellone 2017). Minimal expression of this white pattern has also been found, with so little ticking present in the flanks or at the base of the tail head that the presence of rabicano cannot be easily determined. White markings are generally more extensive in chestnut horses, lessening in bays, followed by minimal white markings on black horses and it is likely that the same phenomenon is

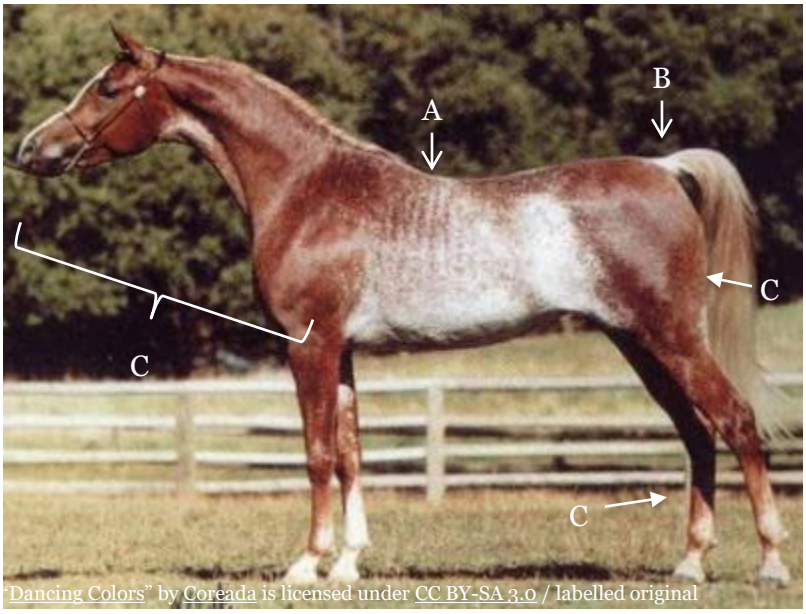

Figure 2: An extreme chestnut rabicano Arabian horse. Note the "barring" at the front of the rib cage (A), the white at the top of the tail (B), and the solid colored head, neck, shoulders, hindquarters, and legs (C). The light tail is likely a result of the flaxen variant (undiscovered). This phenotype could be mistaken by few for sabino or varnish roan.

true for rabicano, with chestnut horses typically displaying more producing prominent white ticking (D. Phillip Sponenberg & Bellone 2017).

The variety of phenotypic expression makes classification of rabicano difficult based on appearance alone. Furthermore, rabicano is similar and often confused with other white spotting patterns, such as sabino, varnish roan, and classic roan (D. Phillip Sponenberg & Bellone 2017). Rabicano is also difficult to identify in horses with dilute coat colors, such as palomino and cremello, as the white hairs are not always obvious against the lighter pigmented hairs of dilute horses. Phenotyping for rabicano is further complicated by horses who have ticking in the flanks but no coontail, or vice versa, and those who have multiple white spotting patterns. For example, it is possible for a horse to have both the classic roan and rabicano phenotypes. Additionally, complications for accurate phenotyping for rabicano include white hairs due to injury and age-related roaning. Given its similarity in phenotype to other roan-like patterns and the fact that some horses have more than one white spotting pattern, rabicano is frequently misidentified. Proper coat color classification is important for animal identification and registration purposes. In addition, rabicano is a desirable phenotype in many breeds. Therefore, proper rabicano phenotyping for both identification and mate selection purposes would be greatly aided by the identification of the causal genetic variant and a subsequent DNA test.

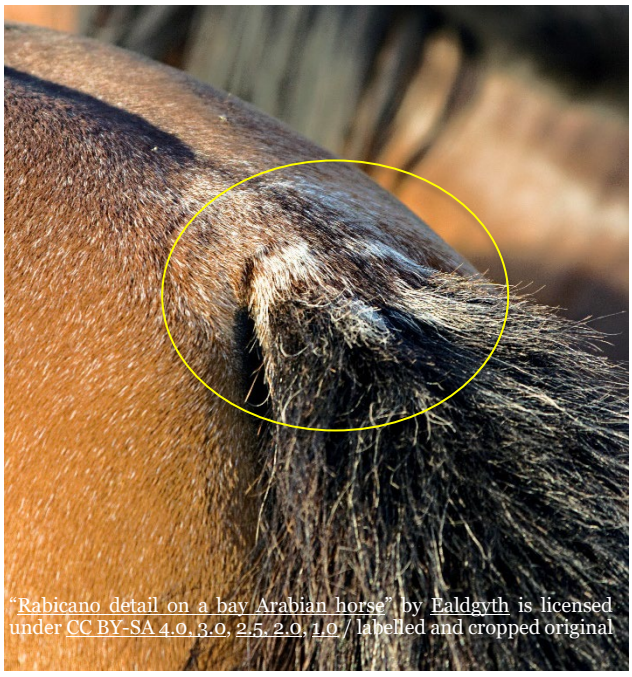

Figure 3: Coontail phenotype of a rabicano horse. The white "bars" (circled) at the tailhead of this day horse is referred to as a "coontail", a trait that is part of the rabicano coat color phenotype in horses.

It is possible that the phenotypic variation in rabicano due to epistatic or polygenic modes of inheritance. A similar trait is white spotting in Appaloosas and related breeds. The leopard complex variant (*Lp),* NC_009144.2g.108297929_108297930ins[1378], permits leopard complex patterns varying from minimal leopard complex markings to an almost completely white coat (Bellone *et al.* 2013; Raudsepp *et al.* 2019). An additional variant, *PATN1*, NC_009146.2:g.23658447T>G, in combination with *Lp* results in a pattern with high amounts of white (at least 60% at birth) (Holl *et al.* 2016; D. Phillip Sponenberg & Bellone 2017). It is possible that rabicano involves similar modifying loci that result in a minimal rabicano patterning, with a second variant increasing the amount of white. However, this has yet to be investigated.

Very little is known about the genetics of roan and roan-like phenotypes in the horse. Classic roan is a coat color phenotype consisting of white hairs intermixed with solid colored hairs throughout the coat that typically leaves the head and lower legs fully pigmented. A haplotype associated with the KIT proto-oncogene receptor tyrosine kinase gene (*KIT*) has been identified in Quarter Horses, Noriker horses, Icelandic horses, Slovenian Coldbloods, Murgese, and a Belgian Draught horse but was not associated with roan Shetland Ponies, German Sport Horses, or a Trakehner (Grilz-Seger *et al.* 2020). The causal mutation for roan in horses has yet to be identified.

Figure 4. A classic bay roan Quarter Horse. Note the white ticking throughout the body and neck of the horse, but the solid white head, legs, mane, and tail. Roan horses are unique in that injuries lose the white ticking as they heal, resulting in solid colored marks. The solid brown lines on the horse's hindquarters (A) are the result of this phenomenon.

Another roan-like phenotype in the horse is the sabino white spotting pattern. Sabino is characterized by extensive white markings on the face and legs as well as white spotting in the belly. In some cases, the white belly

markings extend to include notable white ticking upwards onto the barrel, but usually does not occur across the entire body. While several different genetic variants are expected to cause a similar phenotype, only one of these have been identified, Sabino 1. Sabino 1 is caused by a SNP in intron 16 of *KIT,* SNP AX-103727726 (ENSECAT00000014185.3:c.2686-13A>T), 13 base pairs (bp) upstream of exon 17 resulting in exon skipping during splicing (Brooks & Bailey 2005; Druml *et al.* 2018). This variant is inherited in an incompletely dominant fashion with heterozygous individuals having irregular white markings on the legs, abdomen, and face, with minimal to extensive roaning on the body while homozygous sabino horses are at least 90% white. (Brooks & Bailey 2005; Druml *et al.* 2018).

Roaning throughout the coat has also been identified in several other species, including horses, cattle, goats, pigs, dogs, and mink (Anistoroaei *et al.* 2012; Fontanesi *et al.* 2010; Kawakami *et al.* 2021; Seitz *et al.* 1999; D P Sponenberg *et al.* 1984; Talenti *et al.* 2018). Several of these have a known genetic cause of roan. For example, variants near *KIT* are associated with roan in multiple breeds of pigs, with a repeated element in an intron likely causing exon 5 to get skipped (Lim *et al.* 2011). The causative mutation for roan in cattle is a non-conservative missense mutation in KIT Ligand (*KITLG*), ENSBTAT00000074582.1:p.(Ala191Asn), resulting in an incomplete dominance phenotype. Cattle without the variant have a fully pigmented coat, those who are heterozygous have the roan phenotype, and those who are homozygous for the variant are all-white animals (Seitz *et al.* 1999). *KITLG* is also suspected to play a role in roan in goats but no causal mutation has been identified (Talenti *et al.* 2018). Although not associated with roan, a 6kb copy number variant upstream of *KITLG,* NC_006597.3*:g.29821450_29832950[2_?]* affects pheomelanin and eumelanin pigment intensity in dogs (Weich *et al.* 2020). Additionally, a SNP in a regulatory region 350kb upstream of *KITLG, rs12821256 ,* affects human hair pigmentation and is associated with blonde hair (Guenther *et al.* 2014). Finally, an 11kb tandem duplication in an intronic region of usherin

(*USH2A*), NC_006620.3:g.11131835–11143237dup is likely responsible for a dominant roan phenotype in dogs and appears to be fixed in dalmatians (Kawakami *et al.* 2021).

Given the numerous associations of roan and roan-like phenotypes to both *KIT* and *KITLG* across species, these are strong functional candidate genes for rabicano in the horse. However, a total of 327 genes have been identified to contribute to pigmentation across mammals and therefore it is also likely that rabicano is causes by one of these other genes (Baxter *et al.* 2019). Utilizing a candidate gene approach including this comprehensive list of pigmentation-related genes in mammals, here we aim to identify the genetic cause of rabicano in horses.

2 Methods & Materials

2.1 Sample Collection, Phenotyping, and Grouping

Hair and/or blood samples, photographic records, and pedigrees were collected from 244 horses for potential enrollment in this study. DNA from mane or tail hair follicles were extracted with a crude hair lysis protocol, based off the protocol described by Locke *et al. (Hack et al. 2021; Locke et al. 2002)*. For each horse, five to seven hair bulbs were collected and transferred into a 0.5ml tube with 5µl of 2mg/ml Proteinase K, 10µl of PCR buffer, 10µl of 25mM $MgCl_2$, 0.5µl of Tween 20, and 79.5µl of H_2O. Samples were incubated at 60°C for 45 minutes immediately followed by 95°C for 45 minutes. Samples were then stored at 4°C until use. DNA was extracted from blood samples with the Puregene® Blood Core Kit C and DNA Purification from Whole Blood or Bone Marrow protocol from Qiagen according to the manufacturer's protocol (Germantown, Maryland).

Horses were genotyped by the University of California, Davis Veterinary Genetics Laboratory (VGL) for the commercial full coat color panel (agouti, cream, champagne, dun, grey, leopard complex, appaloosa pattern-1, pearl, splashed white 1-4, extension, silver, lethal white overo, sabino 1, tobiano, W5, W10, W20, and W22) and classic roan. This classic roan test

genotypes an associated haplotype discovered by the VGL (UC Davis Veterinary Genetics Laboratory n.d.).

Given the difficulty in phenotyping rabicano (see Introduction), four phenotypers independently rated all 245 horses for the rabicano coat color using the following criteria: flank ticking, coontail, and no extensive roaning. Each trait could be marked as present, absent, or indeterminate. When available, images of each horse showed the full right and left sides, close-ups of the right and left flanks, and an image of the tailhead. Horses with flank ticking and a coontail but no extensive roaning were classified as cases (rabicano), while horses with none of the traits were considered controls. Horses with indeterminate flank ticking, indeterminate coontail, or flank ticking but no coontail and vice versa were considered an indeterminate phenotype and were excluded from the study with one exception; see family of interest description below. Horses who had extensive roaning or were roan by genotype and those with one or more known white spotting variants (except for W20) were also excluded from the study. Consensus among at least three of the phenotypers was necessary to include a horse as a case or control for further analysis.

Horses were categorized by breed and Quarter Horses were additionally classified as "family of interest" or "non-family of interest". The family of interest contained 44 offspring of a popular rabicano sire and the sire himself (Table 3). This half-sibling family of interest was initially included in this study because the sire's production record suggested he may have been homozygous for a dominant variant causing the rabicano phenotype. However, 29 of his offspring phenotyped as indeterminate as they either had inconclusive ticking in the flank or tailhead or at least three of the four phenotypers (blinded to pedigree information) did not agree that they showed enough white ticking to be included as a case. Another 4 offspring were excluded due to extensive roaning. However, one indeterminate offspring remained in the study to prioritize variants if this stallion was in fact homozygous. Six offspring were phenotyped as controls by at least three raters and were included in the analysis.

Table 3: Number of horses phenotyped as rabicano cases and controls. Quarter Horses were additionally grouped into those from the half-sibling family and those who were not descendants of the sire of interest.

Breed	Breed	Rabicano	Control	Total
Quarter Horses	*Family of Interest*	*39*	*6*	*45*
	Non-Family of Interest	*17*	*22*	*39*
	Group Total	56	29	85
Non-Quarter Horses	Arabian	1	3	4
	Caspian Pony	1	0	1
	KWPN	1	0	1
	Warmblood	1	0	1
	Welsh Pony	1	0	1
	Morgan	0	2	2
	Thoroughbred	0	2	2
	Group Total	5	7	12
	Total	**61**	**36**	**97**

2.2 Short Read Data and Analysis

Four horses from the Quarter Horse family of interest were sequenced using the Illumina Novaseq 6000 S4 platform at the University of California, San Francisco, Genomics Cores Family, with 150bp paired end reads and an insert size of 350bp. Average depth of coverage was 28x. Samples included the sire of the half-sibling family (rabicano), two of his rabicano offspring, and one offspring with an indeterminate phenotype.

The HTStream pipeline was used to process and filter the sequencing data for quality (Streett 2017). These data were then aligned to EquCab 3 using Burrows-Wheeler Aligner (BWA) (Kalbfleisch *et al.* 2018; Li & Durbin 2009). Sequencing data from one rabicano Quarter Horse with shared ancestry with the family of interest, four generations back, was provided by a collaborating laboratory for use in the analyses (unpublished). Whole genome sequencing data from six Friesians, ten Haflingers, and one Tennessee Walking Horse were used as non-rabicano controls as Friesians and Haflingers have not been reported to have the rabicano white pattern

and the Tennessee Walking Horse was utilized in another study and phenotyped based on photographic records as a control (Bellone *et al.* 2017; Hack *et al.* 2021; Hisey *et al.* 2020).

2.3 Long Read Data and Analysis

The sire of the Quarter Horse family of interest was sequenced by the DNA Technologies & Expression Analysis Core at the University of California, Davis Genome Center using Oxford Nanopore long-read sequencing with the LSK109 library prep kit and flowcell version R9.4.1 (PRO002). Base-calling was completed by the same group using ont-guppy-for-minknow 3.0.3. Two directories of fastq.tar files were created based on Phren scores, with <Q7 considered "failed" reads and >Q7 considered "pass" reads. All reads >Q7 were used in the structural variant (SV) identification analysis. Additional quality assessment was performed by NanoPlot v 1.38.0 (De Coster *et al.* 2018). Minimap2 version 2.17-r941 was used to map the reads to the most up-to-date reference genome, EquCab3.0 (Kalbfleisch *et al.* 2018; Li 2018). Quality of long-read mapping was evaluated with samtools, which was then used to convert and index the output to a bam file. SVs were called with sniffles version 1.0.11, and SV distribution was calculated with SURVIVOR version 1.0.3 (Jeffares *et al.* 2017; Sedlazeck *et al.* 2018). Scripts for the long-read analysis can be found in this GitHub repository: https://github.com/EEsdaile/Thesis.

2.4 Variant Identification and Prioritization

Variant identification utilized a list of 659 candidate genes (Appendix A). Candidate genes were identified with a literature search, including 323 from Baxter *et al.* 2018 reported to be involved in human and mouse pigmentation, as well as other genes implicated in mammalian pigmentation using a search of pigmentation related gene ontology (GO) terms, and a search of The Mouse Genome Informatics (MGI) website (Ashburner *et al.* 2000; Baxter *et al.* 2019; Blake *et al.* 2017; Law & Shaw 2018; The Gene Ontology Consortium 2019). Candidate gene coordinates were retrieved from Ensembl and the National Center for Biotechnology Information (NCBI) and converted to the reference genome, EquCab3, as necessary (Kalbfleisch *et al.* 2018; NCBI

Resource Coordinators 2016; Zerbino *et al.* 2018). Freebayes v. 1.3.1-17, bcftools v. 1.10.2, and samtools were used to prioritize coding variants within the 659 candidate genes involved with pigmentation (Appendix A)(Danecek *et al.* 2021; Garrison & Marth 2012). Additionally, 1Mb both up and downstream of each gene of interest were investigated in an effort to capture potential regulatory variants.

Variants were prioritized for further evaluation according to a dominant hypothesis, i.e. controls (n=17) were homozygous reference but cases (n= 5 or 6) were homozygous alternate or heterozygous. Variant identification was performed with and without the distantly related quarter horse to allow for the investigation of phenocopies. Concerning the one indeterminate horse, analyses were performed with this horse included as a case (i.e. cases: n=5, controls: n=17), repeated with the horse included as a control (i.e. cases: n=4, controls: n=18), and again repeated with this horse excluded (i.e. cases: n=4, controls: n=17). All identified variants were further considered as to not miss any variants that could lead to incomplete penetrance or minimal expression of rabicano phenotype observed in this horse. Variants were further annotated by SNPeff v. 4.3 (Cingolani *et al.* 2012) and only high and moderately impactful variants were prioritized.

During the initial investigation, without data from the rabicano horse outside of the half-sibling family, only regions within gene boundaries on the candidate gene list were investigated. Twelve moderately impactful variants were identified when the indeterminate horse was classified as a case and five were found when classified as a control. These variants were selected for further examination based on a.) the role of the impacted gene in melanogenesis as identified by a literature search and b.) the SIFT score of the variant as calculated by PredictSNP (Ng & Henikoff 2003). This produced three variants that were selected for further investigation. ECA1:114,505,701C>A and ECA20:17,241,786C>T were identified when the indeterminate individual was classified as a case and ECA10:4,049,610C>T was identified when classified as a control. When data from the more distantly related case was added to the analysis, three

additional variants were identified when the indeterminate individual was classified as a case (ECA30:26,482,977C>T, ECA1:155,328,511G>T, and ECA1:114,505,701C>A), and two were identified when that individuals was classified as a control (ECA10:4,049,610C>T and ECA28:15,431,302C>T).

Additional putative regulatory variants involving *KITLG* and 1 Mb flanking this gene and known histone marks in the skin were prioritized for further investigation using Microsoft Excel (Kingsley *et al.* 2021).

SVs and insertions and deletions (INDELs) in the long-read data were identified with Variant Effect Predictor (VEP) v. 104.1 and then visualized in Integrative Genomics Viewer (IGV) v. 2.8.2 utilizing a candidate gene approach (Appendix A)(McLaren *et al.* 2016; Robinson *et al.* 2011; Thorvaldsdóttir *et al.* 2013). Additionally, SVs and SNPs 1Mb up and downstream of, and within, *KITLG* were identified in IGV and with samtools and freebayes (respectively). In IGV, the long and short read genome sequencing data from the sire was investigated along with short read data from additional rabicanos and controls.

We visually inspected all SV of interest in the University of California, Santa Cruiz (UCSC) Genome Browser using EquCab3 as the reference genome and classified overlap with genomic features of interest, including short and long interspersed nuclear elements (SINE and LINEs, respectively), long terminal repeats (LTRs), and simple tandem repeats (Benson 1999; Kent *et al.* 2002; Smit *et al.* n.d.).

2.5 Primers

Primers for SVs were designed with Primer3 v. 0.4.0 and MassARRAY (MA) primers were designed using the MassARRAY Typer Assay Design v. 4.0.0.2 (Agena Bioscience, San Diego, CA) (Koressaar & Remm 007; Kõressaar et al. 2018; Untergasser et al. 2012). Primers were manufactured by Integrated DNA Technologies (IDT, Coralville, IA). Primers for INDELs were designed to flank the variant of interest, except for the deletion spanning ECA28:13,938,253-

13,941,507 which required an internal primer to genotype the variant correctly (Table 4). Forward primer 1 and reverse primer 1 were placed outside of the deletion at ECA28:13,938,253-13,941,507to detect the deletion, while forward primer 2 was placed inside of the region to detect the reference allele in conjunction with reverse primer 1 (Table 4). The short-read data indicated a 76bp SINE in the middle of this deletion so primers were designed to determine the presence or absence this SINE (Figure 5). To do so, one primer, reverse primer 2,was placed within the SINE and was paired with forward primer 1 (outside the deletion on the 5' end) to detect the retention of the SINE in individuals with the deletion. Forward primer 3 was placed within the deleted region and paired with reverse primer 2 to detect the presence of the SINE in individuals with the reference sequence. Additionally, if an individual retained the SINE within the deletion, like individual C in Figure 5 appears to have, the PCR product would be 76bp longer than expected if the individual had the true, 3,254bp,deletion.

The INDEL primers had a fluorescent tag, FAM or Yakima Yellow, on either the forward or reverse primer so that size variation could be determined using ABI3730 genetic analyzer (Applied Biosystems, Thermo Fisher, Waltham, MA) (Table 4). Primers were rehydrated with Applied Biosystems' 1X PCR Buffer II to manufacturer specifications for a final concentration of 100uM. Rehydrated primers were analyzed on the Eppendorf BioPhotometer (Hamburg, Germany) and A260 values were recorded. Concentration was calculated as follows: concentration (μM)=(A260 value x 1,000)/(molar extinction coefficient x dilution factor), (dilution factor = 100). Primer mixes were diluted with the same Applied Biosystems' 1X PCR Buffer II to achieve the correct concentration in each reaction (Table 4).

2.6 Genotyping & Analysis

Cases (n=61) and controls (n=36) were genotyped for SVs using PCR reactions (Appendix B) and analyzed on the ABI 3730. SNP variants were genotyped with the MassARRAY System with CPM 96. Genotypes for SVs and SNPs were identified using STRand v. 2.4.131 (Toonen & Hughes 2001)

Table 4: Variants, primers, and primer concentrations of prioritized variants genotyped on the ABI 3730 and MassARRAY (MA). Forward primers 1 and 3 and reverse primer 2 for ECA28:13,938,253-13,941,507 were used to detect the presence or absence of a SINE within the deleted region.

Coordinates	Reference Allele	Alternate Allele	Platform	Primer Concentration (μM)	Primer Type	Sequence
ECA28:10,019,065 -10,019,144	-	79bp Deletion	ABI 3730	0.10	Forward Primer	AGTCGCTTTACCCGTCAAAA
			ABI 3730	0.09	Reverse Primer	/56-FAM/TCTGCAAAGATGAATGTTGGTC
ECA28:13,938,253 -13,941,507	-	3254bp Deletion	ABI 3730	0.10	Forward Primer 1	TGCAGAAACCTGTTGCATTT
			ABI 3730	0.09	Forward Primer 2	TGAACCTGACTGGAAGCAAA
			ABI 3730	0.09	Forward Primer 3	GGTAGTTGCCAGAGGGAAAG
			ABI 3730	0.10	Reverse Primer 1	/56-FAM/AGAGAACCCACCTCATGCTC
			ABI 3730	0.09	Reverse Primer 2	/5YakYel/GGGCAAATCTTCCTCAGTAAA
ECA28:15,067,299	A	G	MA	14	UEP_SEQ	cTTCCTTAATAAATGTCCTTTTTT
			MA	1	2nd-PCRP	ACGTTGGATGTGTGAAGCCACGTTTATTCC
			MA	1	1st-PCRP	ACGTTGGATGTTGCATTTCACACCTGCTTC
ECA28:15,431,302	T	C	MA	14	UEP_SEQ	TCTATTTGTTTCTGATAGTCAATAA
			MA	1	2nd-PCRP	ACGTTGGATGGTCACTGTCTTCTCCTCTTC
			MA	1	1st-PCRP	ACGTTGGATGCTTCAGAATGAACAGCTTCG
ECA28:15,442,561	C	T	MA	7	UEP_SEQ	ACTGTTTGGACCACTG
			MA	1	2nd-PCRP	ACGTTGGATGTGACATTTTCGACTGTTTGG
			MA	1	1st-PCRP	ACGTTGGATGCCAAAAGGAAAACTTATCATC
ECA28:15,446,190	C	T	MA	11.6	UEP_SEQ	AACACCATTAATGACAGCC
			MA	1	2nd-PCRP	ACGTTGGATGCCCTGTCAGGCCTCTTTTAT
			MA	1	1st-PCRP	ACGTTGGATGGACCTTCGGAAGGTGATTAC
ECA28:15,485,999	A	-	MA	14	UEP_SEQ	cTCTTATATAGCTCCTTTTTTTTTT
			MA	1	2nd-PCRP	ACGTTGGATGAGAGTAGGCAGTCTTTGTTC
			MA	1	1st-PCRP	ACGTTGGATGCTGATCTCTATAGTAAATGG

Coordinates	Reference Allele	Alternate Allele	Platform	Primer Concentration (μM)	Primer Type	Sequence
ECA28:15,503,208	-	TTT	ABI 3730	0.2	Forward Primer	/56-FAM/AGGGAAACCGGATGAGAAAA
			ABI 3730	0.2	Reverse Primer	AGATGGGCATGGATGTTAGC
ECA28:15,504,948	G	T	MA	9.3	UEP_SEQ	GACAAGGCAAAGGCTCA
			MA	1	2nd-PCRP	ACGTTGGATGTCTTGCATGACTGACAAGGC
			MA	1	1st-PCRP	ACGTTGGATGAGAGAGCAAAGCTAGGACAG
ECA28:15,518,406-15,518,409	GAGA	-	ABI 3730	0.2	Forward Primer	/5YakYel/CAGGCAGGATAGAATCAGCA
			ABI 3730	0.2	Reverse Primer	TCCTTCTCCAGGTTTTGGAA
ECA28:15,967,332	G	A	MA	11.6	UEP_SEQ	taTGTAATTACCTGCCAGAG
			MA	1	2nd-PCRP	ACGTTGGATGGACTAAGATCAAACAATCTG
			MA	1	1st-PCRP	ACGTTGGATGATGCAACCTAATGCTGCCTC
ECA28:16,465,850	A	T	MA		UEP_SEQ	GCCCCCCTCCTCACC
			MA	1	2nd-PCRP	ACGTTGGATGAACACTTTCTCCGAGACCTG
			MA	1	1st-PCRP	ACGTTGGATGGGTTACTAGGGCAACCTTTG
ECA28:16,616,952	T	A	MA	11.6	UEP_SEQ	CTGATCTATACTGATACCTCA
			MA	1	2nd-PCRP	ACGTTGGATGCTTGAGCGATTGGATTAACTG
			MA	1	1st-PCRP	ACGTTGGATGAACGAGTTCACAGAGAGGAC
ECA28:16,764,837	G	A	MA	7	UEP_SEQ	aTTGTTCCCACCACACT
			MA	1	2nd-PCRP	ACGTTGGATGCTGCTGCCTTCATTTCTCTG
			MA	1	1st-PCRP	ACGTTGGATGGAAAATGAGAACAAGCAGCC
ECA28:31,573,538-31,573,779	-	241bp Deletion	ABI 3730	0.12	Forward Primer	GTGTCAGGCTGTCCTGCAT
			ABI 3730	0.10	Reverse Primer	/56-FAM/AGGGATCTTCTCCCACATCC
ECA28:38,000,030	-	238bp Insertion	ABI 3730	0.11	Forward Primer	TTCTGCGGAGTGCCCTATAA
			ABI 3730	0.10	Reverse Primer	/56-FAM/GCTAGGTCTCGGGCCTTATT
ECA28:38,605,308	-	234bp Insertion	ABI 3730	0.10	Forward Primer	GAAGAGGTCCCACAGGTCAA
			ABI 3730	0.10	Reverse Primer	/5YakYel/CACTTCCCCTCATCTATTTGG
ECA28:39,293,149	G	A	MA	9.3	UEP_SEQ	TGGATGATCTCGCCTCT
			MA	1	2nd-PCRP	ACGTTGGATGAAGACAGCAGCTATGAGGAG

Coordinates	Reference Allele	Alternate Allele	Platform	Primer Concentration (µM)	Primer Type	Sequence
			MA	1	1st-PCRP	ACGTTGGATGTCTGAGTATGTGGAACAGGG
ECA28:39,296,303	C	T	MA	9.3	UEP_SEQ	ATCTTCCTCTCACAAGTT
			MA	1	2nd-PCRP	ACGTTGGATGCGGAGTTCATTATCTTCCTC
			MA	1	1st-PCRP	ACGTTGGATGCATAGCCTTCAGCTAGGGAG
ECA28:39,300,931	G	A	MA	14	UEP_SEQ	TCAGTTTACACAGATAGTATCTA
			MA	1	2nd-PCRP	ACGTTGGATGCCCGTGTCAGGAAAAATCAG
			MA	1	1st-PCRP	ACGTTGGATGAGAATGCCACACAGTGCTTG
ECA1:114,505,701	C	A	MA	7	UEP_SEQ	CGGGTCTAAAGACCTCT
			MA	1	2nd-PCRP	ACGTTGGATGTCATCCAGTTCTTCAGCTCC
			MA	1	1st-PCRP	ACGTTGGATGCTTGTGAGGAATTCCTCAGC
ECA1:155,328,511	G	T	MA	14	UEP_SEQ	TCCTTCCCCATCTGC
			MA	1	2nd-PCRP	ACGTTGGATGAAAGAGGAAGGGAGTTGCTG
			MA	1	1st-PCRP	ACGTTGGATGACAAGAAGAAAAGGGCGTGG
ECA10:4,049,610	C	T	MA	7	UEP_SEQ	AGTTTAGCGCACACG
			MA	1	2nd-PCRP	ACGTTGGATGCACATTACTGCTCCGAGATG
			MA	1	1st-PCRP	ACGTTGGATGGATCCAGAACAGGCTGAAAG
ECA20:17,241,786	C	T	MA	7	UEP_SEQ	AGTCATCCTCCACCACG
			MA	1	2nd-PCRP	ACGTTGGATGTCCCACTTTCACTGGAGTTC
			MA	1	1st-PCRP	ACGTTGGATGGGAAGTGATGCGTGTTACAG
ECA30:26,482,977	C	T	MA	14	UEP_SEQ	ACCAAAATCTGGAATTCTTC
			MA	1	2nd-PCRP	ACGTTGGATGCTTCAAAGGCTGTGTCCAAG
			MA	1	1st-PCRP	ACGTTGGATGGATTCCGGGATGAATTGCTC

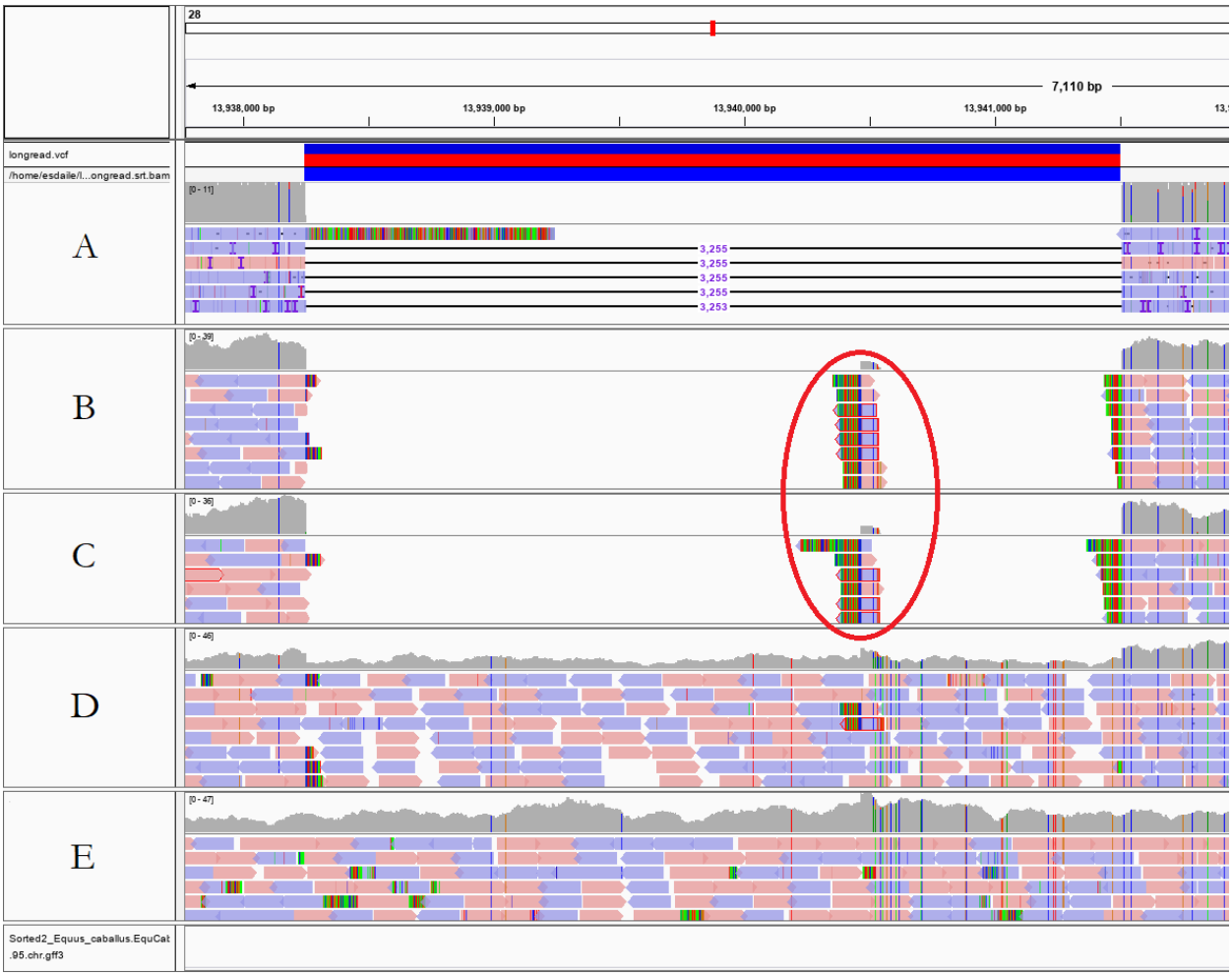

Figure 5: View of 3,254bp deletion at ECA28:13,938,253-13,941,507 in the Integrative Genomics Viewer. Circled is the region from ECA28:13,940,468-13,940,544 that appeared as an "island" in the short-read deletion. This appears to be a portion of a SINE that mapped incorrectly, corroborated by the absence of this sequence in the long-read data. Individuals are as follows: (A) sire of family of interest – long read data, (B) same sire of family of interest, short read data, (C) offspring in family of interest, short read data, (D & E) Friesian and Haflinger controls, respectively, short read data.

and the MassARRAY Typer Analyser v. 5.0.2, respectively. Contingency tables and Fisher exact p-values for each variant were calculated with the Microsoft Excel add-on Real Statistics Resource Pack (Release 7.6)(Zaiontz 2021), assuming a dominant mode of inheritance. Using the same software and mode of inheritance, chi-squared was calculated for each variant with n>5 for each genotype of each group. To investigate potential associations between variants and the rabicano

phenotype, conditional inference trees and variable importance plots were calculated in Rstudio v1.3.1093 (R v4.0.3)with the packages randomForest and partykit (Hothorn & Zeileis 2015; Hothorn *et al.* 2006; Liaw & Wiener 2002; Microsoft Corporation 2018; Team 2020). Potential haplotypes associated with rabicano were identified within 1.6Mb surrounding *KITLG*. Haplotypes were visualized in Microsoft Excel to identify the concordance of rabicano with *KITLG* (Microsoft Corporation 2018).

3 Results

3.1 Phenotyping Demographics

Using strict phenotyping criteria, 61 rabicano cases and 36 controls were utilized in this study. Of the cases, 56 were Quarter Horses, 39 of which were members of the half-sibling family of interest, and the remaining 5 were non-Quarter Horses (one Arabian, one Dutch Warmblood, one unspecified warmblood, one Welsh Pony, and one Caspian Pony) (Table 3). Of the 36 controls, 29 were Quarter Horses, 6 were members of the family of interest, and 7 were non-Quarter horses (three Arabians, two Morgans, and two Thoroughbreds) (Table 3).

3.2 Long Read Data Quality Assessment

Quality assessment of the long-read data produced an N50 read length of 28.2kb, a mean quality score of Q8.8, and 29.3 gigabases (Gb) total (Table 5 5). Quality assessment of reads >Q7 retained 26.6Gb, with an N50 read length of 28,721bp and mean quality score of Q10.7 (Table 5). Seventy-two percent of reads were >Q7 and were included in the analysis (Table 6). The four longest reads were <Q7 and could not be included in the analysis (Table 7). The highest quality score attained was Q17.9, with a read length of 290bp (Table 7). Average genome-wide coverage of 10x was attained.

Table 5: Summary of the long-read data of the sire of the Quarter Horse family of interest.

General Summary	All	>Q7
Mean read length:	14,098.6bp	17,895.7 bp
Mean read quality:	Q8.8	Q10.7
Median read length:	8,883bp	14,502 bp
Median read quality:	Q10	Q11
Number of reads:	2,075,239	1,488,970
Read length N50:	28,169bp	28,721 bp
Total Bases:	**29,258,011,814bp**	**26,646,134,789 bp**

Table 6: Number of long reads above quality cutoffs. Number of reads and total megabases (Mb) above each Phred score (Q).

Minimum Phred Score	Number of Reads (percentage of total reads)	Mb
>Q5	1646101(79.3%)	28,412.1
>Q7	1488970(71.7%)	26,646.1
>Q10	1035928 (49.9%)	19,295.4
>Q12	342579(16.5%)	6,456.0
>Q15	90 (0%)	0.5

Table 7: Highest quality scores (>Q7) and read lengths from the long-read data from the sire of the Quarter Horse family of interest.

Rank	5 Highest Mean Quality Scores and Read Lengths	Mean Quality Scores of 5 Longest Reads (All)	Mean Quality Scores of 5 Longest Reads (>Q7)
1	17.9 (290 bp)	3.1 (712,732)	7.5 (296,275 bp)
2	16.6 (173 bp)	3.1 (587,727)	8.9 (243,961 bp)
3	16.5 (2860 bp)	3.5 (322,095)	11 (202,928 bp)
4	16.2 (298 bp)	3.1 (311,266)	10.8 (200,682 bp)
5	15.9 (1839 bp)	7.5 (296,275 bp)	8.9 (188,398 bp)

3.3 Identification of Variants

A total of seventeen SNPs, seven INDELs, and one SV variant of interest were identified and pursued through further investigation (Table 9). Of the six initial variants initially investigated (Table 8), three variants, *ECA10:4,049,610C>T, ECA28:15,431,302T>C*, and *ECA20:17,241,786C>T*, had Pearson's X^2 and Fisher's Exact Test p-values <0.05 when evaluating genotype data from the 61 cases and 36 controls (Table 10). A conditional inference tree of these six variants of interest and variable inference analysis suggested that ECA28:15,431,302T>C (CEP290:p.Ile173Val) best predicts rabicano phenotype (Figure 6 and Figure 7). Together, these data and the proximity of ECA28:15,431,302 to *KITLG* suggest that variants impacting *KITLG* should be further investigated as the cause for rabicano.

Table 8: Six variants identified in initial analysis, with 4-5 rabicano cases and 17 controls.

Loci	Predict SNP	PhD-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP
ECA1:114,505,701C>A (*OCA2*:p.(Cys93Phe))	D (0.50)	N (0.68)	D (0.74)	U (-)	N (0.67)	D (0.72)
ECA1:155,328,511 G>T (*FMN1*:p.(Pro189Gln))	N (0.73)	N (0.78)	N (0.66)	D (0.47)	N (0.77)	N (0.50)
ECA10:4,049,610C>T (*ANKRD27*:p.(Ala627Thr))	N (0.82)	N (0.78)	U (-)	U (-)	N (0.75)	N (0.66)
ECA20:17,241,786C>T (*KIF13A* p.(Arg848His))	N (0.65)	D (0.57)	N (0.66)	D (0.39)	N (0.64)	N (0.55)
ECA28:15,431,302C>T (*CEP290*:p.Ile173Val)	N (0.82)	N (0.83)	N (0.66)	N (0.63)	N (0.79)	N (0.61)
ECA30:26,482,977C>T (*CRB1*:p.(Pro755Ser))	N (0.82)	N (0.71)	U (-)	U (-)	N (0.68)	N (0.50)

D = Deleterious
N = Neutral
U = Unknown

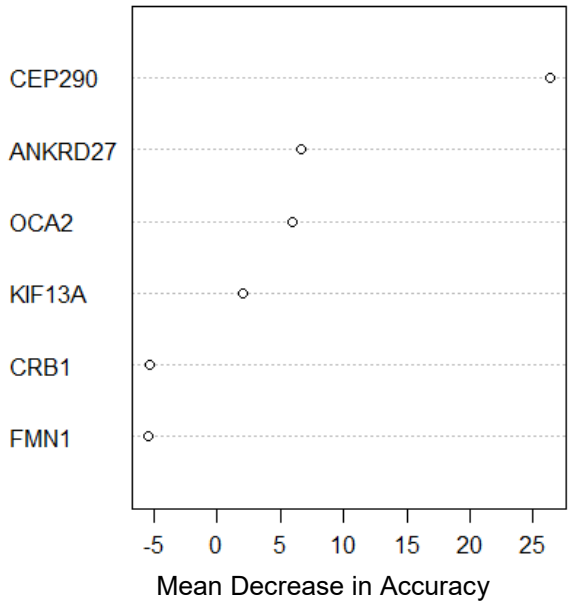

Figure 6: randomForest variable inference plot showing the importance of variants associated with rabicano. Variants shown are from an analysis of ECA1:114,505,701 (*OCA2*), ECA1:155,328,511 (*FMN1*), ECA10:4,049,610 (*ANKRD27*), ECA20:17,241,786 (*KIF13A*), ECA28:15,431,302 (*CEP290*), and ECA30:26,482,977 (*CRB1*). ECA28:15,431,302 (*CEP290*) was determined to be most predictive of rabicano phenotype. The mean decrease in accuracy suggests that removing ECA28:15,431,302 (*CEP290*) from the model would result in the largest loss in accuracy of phenotypic classification.

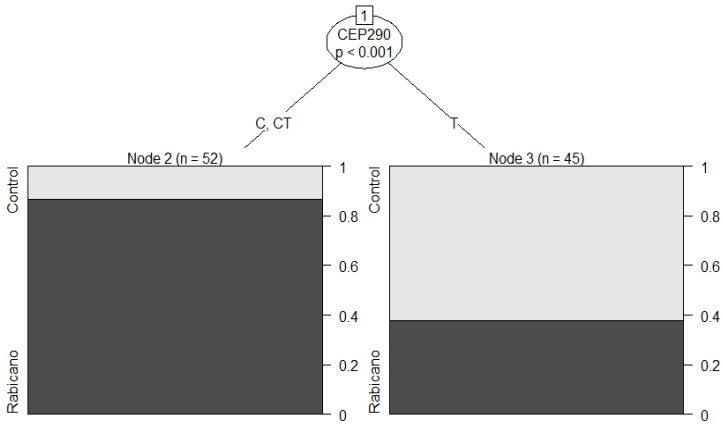

Figure 7: Conditional inference tree from an analysis of ECA1:114,505,701 (OCA2), ECA1:155,328,511 (*FMN1*), ECA10:4,049,610 (*ANKRD27*), ECA20:17,241,786 (*KIF13A*), ECA28:15,431,302 (*CEP290*), and ECA30:26,482,977 (CRB1). ECA28:15,431,302 (*CEP290*) was determined to be the most significantly associated variant with additional inferences not increasing the accuracy in predicting the phenotype.

3.4 Analysis of Potential Regulatory Variants Involving *KITLG*

An analysis of SNPs in the short-read data within 1Mb of *KITLG* (ECA28:14,726,503-16,807,871), excluding the indeterminate individual and including the rabicano outside of the family of interest, identified 166 variants. Of these 166 variants, 7 SNPs were selected for further analysis due to their location within known regulatory histone marks in the skin: ECA28:15,442,561C>T, ECA28:15,446,190C>T, ECA28:15,504,948G>T, ECA28:15,967,33G>A, ECA28:16,465,850A>T, ECA28:16,616,952T>A, ECA28:16,764,837G>A (Kingsley *et al.* 2021) (Kingsley *et al.* 2019). Only two variants near *KITLG* were identified when filtering variants in the short-read data with the indeterminate horse classified as a case and including the rabicano outside the family of interest: ECA28:15,485,999del and ECA28:15,067,299A>G. These two variants also were included in further analysis.

A 3bp insertion (ECA28:15,503,208TTT) and 4bp deletion (ECA28:15,518,406-15,518,409) located between *KITLG* and *CEP290* were identified in IGV and included in the analysis. Another six INDELs of interest were identified by viewing SVs near *KITLG* that were identified in the long-read data by sniffles in IGV to genotype of cases and controls. This identified one variant of interest near *KITLG* (ECA28:15,485,999del) and the five variants 2-22Mb up and downstream from *KITLG*: ECA28:10,019,065-10,019,144del(79), ECA28:13,938,253-13,941,507del(3254), ECA28:31573538-31573779ins(241), ECA28:38,000,030 (238), and ECA28:38,605,308 (234).

Analysis of the short-read data in IGV suggests that the largest INDEL investigated, the 3,254bp deletion at ECA28:13,938,253-13,941,507, retained a 76bp sequence from ECA28:13,940,468-13,940,544 (Figure 5). This region was identified by RepeatMasker as a SINE and was present elsewhere in the genome when searched in the BLAST-like alignment took (BLAT)(Kent 2002). This SINE was not identified in the long-read data; the sequenced horse is homozygous for the deletion (Figure 5). Additionally, this sequence was not detected when genotyping horses homozygous for the deletion, with forward primer 1 paired with reverse primer

Table 9: Genetic variants investigated for concordance with the rabicano phenotype. Includes Human Genome Variation Society (HGVS)-nomenclature for the genomic location of each variant and coding and protein information when possible. Also included is the nearest gene and annotated skin histone mark when applicable.

Abbreviated Coordinates (chr:bp)	HGVS-nomenclature			Reference SNP ID	Nearest Gene/Affected Skin Histone Mark
	Genomic	Coding	Protein		
ECA28:10,019,065-10,019,144	NC_009171.3:g.10019065_10019144del(79)	-	-	-	Part of a simple Tandem Repeat in *PPFIA2* (intron 1) – 60bp of which is part of a SINE
ECA28:13,938,253-13,941,507	NC_009171.3:g.13938253_13941507del(3254)	-	-	-	*MGAT4C* (intron 4) comprises 1 LTR, a SINE insertion, and a LINE insertion with two of its own SINE insertions.
ECA28:15,067,299	NC_009171.3:g.15067299A>G	-	-	rs1151576566	LOC111771013/ENSECAG00000043517.1 (intergenic)
ECA28:15,431,302	NC_009171.3:g.15431302T>C	ENSECAT00000052715.2:c.538A>G	ENSECAP00000018792.3:p.Ile173Val	rs1148517979	*CEP290* (exon 8)
ECA28:15,442,561	NC_009171.3:g.15442561C>T	-	-	rs1144478014	*TMTC3* (intron 1) - H3K4me3/H3K27ac
ECA28:15,446,190	NC_009171.3:g.15446190C>T	-	-	rs396641903	*TMTC3* (intron 1) – within H3K27ac
ECA28:15,485,999	NC_009171.3:g.15485999del			-	*TMTC3* (intron 10)
ECA28:15,503,208	NC_009171.3:g.15503208_15503208insTTT	-	-	-	*TMTC3* (intergenic) - 40bp from H3K4me1
ECA28:15,504,948	NC_009171.3:g.15504948G>T	-	-	rs1146185275	*TMTC3* (intergenic) – within H3K4me1
ECA28:15,518,406-15,518,409	NC_009171.3:g.15518406_15518409del	-	-	-	*TMTC3* (intergenic)
ECA28:15,967,332	NC_009171.3:g.15967332G>A	-	-	rs397240012	ENSECAG00000033697.2 (intron 1) - within H3K4me1
ECA28:16,465,850	NC_009171.3:g.16465850A>T	-	-	rs69379733	*DUSP6* (intron 8) – within H3K4me1, H3K4me3 & H3K27ac

Abbreviated Coordinates (chr:bp)	HGVS-nomenclature			Reference SNP ID	Nearest Gene/Affected Skin Histone Mark
	Genomic	Coding	Protein		
ECA28:16,616,952	NC_009171.3:g.16616952T>A	-	-	rs1136201810	*POC1B* (intronic)/*GALNT4* (exon 1) -within H3K4me3 & H3K27ac
ECA28:16,764,837	NC_009171.3:g.16764837G>A	-	-	rs395184158	*ATP2B1* (intron 1) – within H3K4me1
ECA28:31,573,538-31,573,779	NC_009171.3:g.31573538_31573779del(241)	-	-	-	*BTBD11* (intron 16) – 226bp of the variant is part of a SINE
ECA28:38,000,030	NC_009171.3:g.38000030_38000031ins(238)	-	-	-	*TNRC6B* (intron 3) – SINE Insertion
ECA28:38,605,308	NC_009171.3:g.38605308_38605309ins(234)	-	-	-	*XPNPEP3* (intron 7) – SINE insertion into a LINE
ECA28:39,293,149	NC_009171.3:g.39293149G>A	ENSECAT00000001925.2:c.2458G>A	ENSECAP00000001393.2:p.(Ala820Thr)	rs1138390911	*MEI1* (exon 20)
ECA28:39,296,303	NC_009171.3:g.39296303C>T	-	-	rs1137730908	*MEI1* (intron 20)
ECA28:39,300,931	NC_009171.3:g.39300931G>A	-	-	rs1146714219	*MEI1* (intron 22)
ECA1:114,505,701	NC_009144.3:g.114505701C>A	ENSECAT00000062592.2:c.278G>T	ENSECAP00000027833.2:p.(Cys93Phe)	rs1136953351	*OCA2* (exon 4)
ECA1:155,328,511	NC_009144.3:g.155328511G>T	ENSECAT00000045675.2:c.566C>A	ENSECAP00000009955.2:p.(Pro189Gln)	rs1142586788	*FMN1* (exon 4)
ECA10:4,049,610	NC_009153.3:g.4049610C>T	ENSECAT00000021221.2:c.1879G>A	ENSECAP00000017468.2:p.(Ala627Thr)	rs1140238710	*ANKRD27* (exon 18)
ECA20:17,241,786	NC_009163.3:g.17241786C>T	ENSECAT00000004103.2:c.2543G>A	ENSECAP00000002833.2:p.(Arg848His)	rs396511301	*KIF13A* (exon 21)
ECA30:26,482,977	NC_009173.3:g.26482977C>T	ENSECAT00000061612.2:c.2389C>T	ENSECAP00000008702.2:p.(Pro755Ser)	rs1147290580	*CRB1* (exon 7)

Table 10: Fishers Exact Test and chi-squared p-values for all prioritized variants of interest using a dominant model. The variant most concordant with rabicano when investigating all horses was at ECA28:15,967,332. The most concordant variant in Quarter Horses is at ECA28:13,938,253-13,941,507, 2Mb downstream of *KITLG*. Light green indicates p<0.05 and medium green indicates p<0.01, and dark green indicates p<0.001.

	All		Quarter Horses		Quarter Horse, Family of Interest		Quarter Horses, Non-Family of Interest		Non-Quarter Horse	
Variant Position	**Fishers**	**χ2 p-value**	**Fishers**	**χ2 p-value**	**Fishers**	**χ2 p-value**	**Fishers**	**χ2 p-value**	**Fishers**	**χ2 p-value**
ECA28:10,019,065-10,019,144del	6.3E-03	4.0E-03	0.017	7.9E-03	1.000	-	0.318	-	0.47	-
ECA28:13,938,253-13,941,507del	7.8E-09	5.0E-09	**4.5E-09**	-	0.015	-	1.1E-03	-	1.0	-
ECA28:15,067,299A>G	6.5E-07	5.3E-07	3.5E-06	1.9E-06	0.025	-	3.6E-03	-	0.52	-
ECA28:15,431,302 T>C	2.2E-07	2.2E-07	5.1E-06	2.9E-06	0.035	-	4.0E-03	2.4E-03	0.15	-
ECA28:15,442,561C>T	2.2E-05	1.3E-05	2.2E-04	1.3E-04	0.035	-	0.049	0.029	0.18	-
ECA28:15,446,190C>T	5.4E-06	3.0E-06	6.0E-05	4.1E-05	0.035	-	0.025	0.015	0.15	-
ECA28:15,485,999del	5.4E-06	3.0E-06	6.0E-05	4.1E-05	0.035	-	0.025	0.015	0.15	-
ECA28:15,503,208del	1.6E-05	1.0E-05	3.0E-04	1.4E-04	0.035	-	0.054	0.032	0.15	-
ECA28:15,504,948G>T	5.4E-06	3.0E-06	6.0E-05	4.1E-05	0.035	-	0.025	0.015	0.15	-
ECA28:15,518,406-15,518,409del	5.4E-06	3.0E-06	6.0E-05	4.1E-05	0.035	-	0.025	0.015	0.15	-
ECA28:15,967,332G>A	**7.4E-09**	**4.5E-09**	9.4E-07	4.6E-07	0.035	-	**3.3E-04**	-	**0.010**	-
ECA28:16,465,850A>T	8.1E-08	8.9E-08	1.7E-06	1.2E-06	0.025	-	4.0E-03	2.4E-03	0.15	-
ECA28:16,616,952T>A	8.1E-08	8.9E-08	1.7E-06	1.2E-06	0.025	-	4.0E-03	2.4E-03	0.15	-
ECA28:16,764,837G>A	1.6E-05	1.0E-05	3.0E-04	1.4E-04	0.035	-	0.054	0.032	0.15	-
ECA28:31,573,538-31,573,779del	1.0	-	1.0	-	1.0	-	1.000	-	0.42	-
ECA28:38,000,030ins(2388)	7.8E-06	6.7E-06	1.5E-04	8.8E-05	0.448	-	0.282	-	0.15	-
ECA28:38,605,308ins(234)	0.019	0.011	0.035	0.022	0.356	-	1.0	1.0	1.0	-
ECA28:39,293,149G>A	0.078	0.054	0.22	0.22	1.0	-	0.75	0.64	0.52	-
ECA28:39,296,303C>T	0.078	0.054	0.22	0.22	1.0	-	0.75	0.64	0.52	-

ECA28:39,300,931G>A	0.078	0.054	0.22	0.22	1.0	-	0.75	0.64	0.52	-
ECA1:114,505,701 C>A	0.078	0.054	0.16	0.11	0.13	-	0.21	0.18	0.58	-
ECA1:155,328,511 G>T()	0.20	0.19	0.64	0.59	0.65	-	0.75	0.7	0.24	-
ECA10:4,049,610 C>T	3.58E-3	3.4E-03	0.016	0.012	0.67	-	1.000	-	0.42	-
ECA20:17,241,786 C>T	6.8E-04	4.7E-04	9.3E-04	5.5E-04	1.0	-	0.26	-	1.0	-
ECA30:26,482,977 C>T	0.56	-	0.76	-	1.0	-	0.24	-	1.0	-

2 producing no PCR product (Table 4, Appendix C). The PCR product identifying the 3,254bp deletion confirmed this finding, as it was the length expected. This SINE was detected in all horses with at least one copy of the reference genotype and absent in all horses homozygous for the deletion (n=97) (Appendix C).

In addition to the known three base pair insertion, a second alternate allele was identified in STRand at ECA28:15,503,194 . This 1bp insertion was deemed WT2, as it segregated with the reference alleles of the two variants flanking that locus: ECA28:15,485,999 and ECA28:15,504,948 (Appendix D: Table D1.1. All 13 horses heterozygous for the wild type and wild type 2 alleles (WT/WT2) were homozygous reference for all other variants in the haplotype, save for the last variant have the haplotype, ECA28:16,764,837 (Table 11). The two horses who were homozygous for the WT2 insertion atECA28:15,503,194 failed to genotype for theECA28:15,442,561 SNP (Table 11). Both horses were controls, one Morgan and one QH not in the family of interest (Table 11). All horses homozygous for the expected insertion and all but one horse heterozygous for the expected insertion and WT2 allele atECA28:15,503,194 had the T allele at ECA28:15,442,561 (Table 11). The remaining horse that was heterozygous for the expected insertion and WT2 allele at ECA28:15,503,194 failed to genotype at ECA28:15,446,190. These findings suggest that there may be a null allele at ECA28:15,446,190 that is commonly inherited with the alternate insertion (WT2) at ECA28:15,503,194 (Table 11).

In regards to the region within 1Mb upstream/downstream of *KITLG*, when assessed in all animals using a dominant model, all 12 variants from ECA28:15,067,299 to ECA28:16,764,837 were found to be significantly associated with the rabicano phenotype ($p=<1.6x10^{-5}$) with two additional variants at ECA28:38,000,030 and ECA28:38,605,308 also significantly associated ($p=7.9x10^{-5}$ and 0.019)(Table 10). Two more INDELs, at ECA28:10,019,065-10,019,144 and ECA28:13,938,253-13,941,507 were found to be statistically significantly associated with rabicano when analyzing all animals (p=0.0063 and $7.8x10^{-9}$) and Quarter Horses (p=0.017 and $4.5x10^{-9}$), but the deletion at ECA28:10,019,065-10,019,144 did not reach significance when examining the

Table 11: Potential haplotypes identified within 1Mb upstream and downstream of *KITLG*. 31/39 of rabicano horses in the half-sibling family of interest family, and 45/61 total rabicano horses had at least one copy of the alternate haplotype from ECA28-15,431,302 to ECA28:16,616,952. An additional four rabicano horses had the reference haplotype across that same region but at least one alternate allele at the variant of interest, ECA28:15,967,332A>G, accounting for 80.3% of the rabicano horses. Similarly, 80.5% of controls were homozygous for the reference allele. Although ECA28:16,616,952T>A appears to be concordant with several potential haplotypes, its neighbor, ECA28:16,764,837G>A, appears to be outside the likely haplotype boundary.

n rabicanos/n controls														
All	QH	Family of Interest	QH, non-family of interest	Non-QH	ECA28: 15,067,299	ECA28: 15,431,302	ECA28: 15,442,561	ECA28: 15,446,190	ECA28: 15,485,999	ECA28: 15,503,208	ECA28: 15,504,948	ECA28: 15,518,406-15,518,409	ECA28: 15,967,332	ECA28: 16,465,850	ECA28: 16,616,952	ECA28: 16,764,837
7/0	7/0	6/0	1/0	0/0	G	C	T	T	Del	Ins	G	Del	**A**	T	A	A
2/0	1/0	0/0	1/0	1^{AR}/0	A	T	C	C	WT	WT	T	WT	**A**	A	T	G
1/0	1/0	0/0	1/0	0/0	G	C	T	T	Del	Ins	G	Del	**A**	T	A	G
1/0	0/0	0/0	0/0	1^{KW}/0	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	**A**	T	A	A
21/4	21/4	15/2	6/2	0/0	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	**AG**	TA	TA	AG
10/1	9/1	7/0	2/1	1^{WE}/0	GA	CT	T	CT	Del/WT	WT2/Ins	GT	Del/WT	**AG**	TA	TA	AG
3/0	3/0	3/0	0/0	0/0	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	**AG**	TA	TA	A
1/0	0/0	0/0	0/0	1^{WB}/0	A	T	C	C	WT	WT	T	WT	**AG**	A	T	G
1/0	1/0	0/0	1/0	0/0	GA	T	C	C	WT	WT	T	WT	**AG**	A	T	G
2/1	2/1	1/0	1/1	0/0	G	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	**AG**	TA	TA	AG
0/1	0/1	0/0	0/1	0/0	GA	CT	T	CT	Del/WT	WT2/Ins	GT	Del/WT	**AG**	TA	TA	A
1/0	1/0	1/0	0/0	0/0	A	T	C	C	WT	WT	T	WT	**G**	TA	TA	G
1/0	1/0	0/0	1/0	0/0	A	T	C	C	WT	WT/WT2	T	WT	**G**	A	T	AG
6/6	6/6	4/1	2/5	0/0	A	T	C	C	WT	WT/WT2	T	WT	**G**	A	T	G
1/1	1/0	1/0	0/1	0/0	GA	T	C	C	WT	WT	T	WT	**G**	A	T	G
0/1	0/1	0/0	0/1	0/0	A	T	CT	CT	Del/WT	Ins/WT	GT	Del/WT	**G**	A	T	AG
0/1	0/1	0/0	0/1	0/0	A	T	T	CT	Del/WT	WT2/Ins	GT	Del/WT	**G**	A	T	G

0/1	0/1	0/0	0/1	0/0	A	T	C	C	WT	WT	T	WT	**G**	A	T	AG
0/2	0/1	0/0	0/1	0/1MO	A	T	/	C	WT	WT2	T	WT	**G**	A	T	G
0/1	0/1	0/0	0/1	0/0	A	T	/	C	WT	WT2/Ins	T	WT	**G**	A	T	AG
0/1	0/0	0/0	0/0	0/1^{AR}	G	T	C	C	WT	WT	T	WT	**G**	A	T	G
3/15	2/10	2/3	0/7	1^{CP}/5*	A	T	C	C	WT	WT	T	WT	**G**	A	T	G

Key:

AR	Arabian	Green = Homozygous reference allele		
KWPN	KWPN	Purple = Heterozygous allele		
WE	Welsh Pony	Blue = Homozygous alternate allele		
WB	Warmblood	/ = suspected null allele		
MO	Morgan	Orange = Additional alternate allele		
CP	Caspian Pony	**Bold = Most concordant SNP**		
*	2 Arabians, 2 Thoroughbreds, 1 Morgan	Del = Deletion	Ins = Insertion	WT= reference allele (wild type)

Quarter Horse family of interest. The only variant statistically significantly associated with rabicano in non-Quarter Horses was ECA28:15,967,332, but this could reflect the small sample size (p=0.010). However, this variant is also the most concordant with phenotype when examining all individuals ($7.4x10^{-9}$). These variants around *KITLG*, from ECA28:15,067,299 to ECA28:16,764,837, suggest that a rabicano specific haplotype is present at this locus (Table 11). The most concordant SNP from the region is ECA28:15,967,332A>G, which is located within an H3K4me1 associated enhancer in skin and is located within intron 1 of the novel gene, ENSECAG00000033697.2.

4 Discussion and Conclusion

Investigating whole-genome sequencing data identified six SNPs in candidate pigmentation genes for further investigation. Three of these variants were not associated with rabicano phenotyping when evaluating them across breeds. One SNP (ECA28: 28:15,431,302T>C) was strongly but not perfectly concordant with phenotype and conditional inference analysis suggested it was highly predictive of rabicano phenotype. Together these data and ECA28: 28:15,431,302T>C proximity to *KITLG* suggest that variants impacting *KITLG* could be the cause of rabicano. To evaluate this further, we utilized both long and short-read data to identify additional variants for investigation. In evaluating 20 variants on ECA28, we identified a potential 1.6Mb haplotype associated with rabicano that centered around *KITLG*. While we did not identify any coding variants in *KITLG*, based on our analysis, the associated haplotype suggests that the causative variant could be in a regulatory region that affects the expression of *KITLG* in

melanocytes (Allen *et al.* 2014). Additionally, two variants over 20Mb away from *KITLG* were significantly associated with rabicano in all horses as well as in the Quarter Horse group: ECA28:38,000,030 ($p=7.8x10^{-6}$) and ECA28:38,605,308 ($p=0.019$). Both of these variants are within 500kb of melanin concentrating hormone receptor 1 (*MCHR1)*. Although *MCHR1* was not

initially in our candidate gene list, it has been identified in melanocytes and melanoma cells, and the presence of antibodies for *MCHR1* are associated with vitiligo in humans (Hoogduijn *et al.* 2002; Madelaine *et al.* 2020; Unal *et al.* 2021). It is also possible that an SV not detected in our long-read data but involving *KITLG* and even *MCHR1* could be the cause of rabicano. Significantly associated variants 800kb from *KITLG* may suggest breakpoints of an inversion or other complex SV and must be thoroughly investigated using long-read data from additional rabicano and non-rabicano horses.

The most concordant SNP from this region, ECA28:15,967,332G>A, is located in an H3K4me1 annotated histone mark in skin. H3K4me1 is most frequently associated with gene enhancers. Further, this SNP is located approximately 160kb from the start of transcription for *KITLG*, therefore it may be responsible for enhancing *KITLG* expression. More work is needed to evaluate this SNP and others as potential causal regulatory mutations. However, given that phenotypes of 49/61 rabicano horses included in this study could be explained by this SNP, but 12 were homozygous reference for this SNP (8 rabicano Quarter Horses in the family of interest, three not in the family of interest, and 1 Caspian Pony), other variants may be at play (Table 11). Thus, further refinement of this haplotype testing additional variants from this region as well as investigating the dams from the half-sibling family is necessary. Additionally, seven controls have at least one alternate allele for this variant (2 Quarter Horses in the family of interest and five outside of the half-sibling family). These horses may represent cases of incomplete penetrance or variable expression of the trait. Therefore, more detailed phenotyping is necessary to investigate if there is a phenotypic difference between these subsets of horses to further explore whether or not this variant is associated with white ticking. Also, this variant could be tagging the causal variant but has undergone recombination, and given that 1.6Mb is a large region for a haplotype, more detailed investigation and haplotype analysis is needed. While 8 recombinants in a single half-sibling family seem unlikely given the size of this hapolotype, this region maybe a recombination hot spot. Thus further work is needed to investigate this.

A crucial limitation to the short-read sequencing data analysis was a lack of closely related Quarter Horse controls, which may help to refine the *KITLG* haplotype. It is also possible that the underlying genetic cause of rabicano varies by breed. To mitigate this, an attempt was made to use control horses for each breed represented. This was possible with Quarter Horses and Arabians, but not yet for the Warmbloods, Morgan, Welsh Pony, or Caspian Pony. Additionally, although rabicano has not been reported in Haflinger horses or closely related breeds, the Sabino1 variant has been. This variant has an atypical presentation in Haflingers, producing little to no white markings (Druml *et al.* 2018). This could be concerning when using Haflingers as controls, as they may have the rabicano phenotype and modifiers suppressing white expression since white patterning is selected against in this breed.

To date, this association of rabicano with variants flanking KITLG is the most promising lead in identifying the cause of the rabicano coat color in horses. Nevertheless, more analyses are needed to determine if this region is truly associated by collecting, phenotyping, and genotyping additional horses across breeds. Additional analyses to find the cause of rabicano may include an across-breed genome-wide association study (GWAS) to replicate association to ECA28 and further interrogate variants flanking KITLG.

Identifying the causative variant for rabicano would enable further investigation into similar phenotypes, including other forms of sabino and frost. Plus, identifying the causal variant for rabicano could assist in mate selection and individual animal identification, particularly in the Quarter Horse, Arabian, and Warmblood breeds where this is a desirable phenotype. We hope that one day a genetic test can be offered to horse breeders and owners alike.

Bibliography

Adams, H. A., Sonstegard, T. S., VanRaden, P. M., Null, D. J., Van Tassell, C. P., Larkin, D. M., & Lewin, H. A. (2016). Identification of a nonsense mutation in APAF1 that is likely causal for a decrease in reproductive efficiency in Holstein dairy cattle. *Journal of Dairy Science*, 99/8: 6693–6701. DOI: 10.3168/jds.2015-10517

Adelman, M. L. (1981). The First Modern Sport in America: Harness Racing in New York City, 1825-1870. *Journal of Sport History*, 8/1: 5–32.

Alexeev, V., & Yoon, K. (2006). Distinctive role of the cKit receptor tyrosine kinase signaling in mammalian melanocytes. *The Journal of Investigative Dermatology*, 126/5: 1102–1110. DOI: 10.1038/sj.jid.5700125

Allen, J. R. L. R., de Guzman Strong, C., & Johnson, S. (2014). Highlighting blondes: a tissue-specific KITLG enhancer shows us how. *Pigment Cell & Melanoma Research*, 27/6: 1007–1008. DOI: 10.1111/pcmr.12311

Andersson, L. S., Larhammar, M., Memic, F., Wootz, H., Schwochow, D., Rubin, C.-J., Patra, K., *et al.* (2012). Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. *Nature*, 488/7413: 642–646. DOI: 10.1038/nature11399

Anistoroaei, R., Markakis, M. N., Vissenberg, K., & Christensen, K. (2012). Exclusion of candidate genes for coat colour phenotypes of the American mink (Neovison vison). *Animal Genetics*, 43/6: 813–816. DOI: 10.1111/j.1365-2052.2012.02339.x

Artesi, M., Tamma, N., Deckers, M., Karim, L., Coppieters, W., Van den Broeke, A., Georges, M., *et al.* (2020). Colour-sidedness in Gloucester cattle is associated with a complex structural variant impacting regulatory elements downstream of KIT. *Animal Genetics*, 51/3: 461–465. DOI: 10.1111/age.12932

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., *et al.* (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25/1: 25–29. DOI: 10.1038/75556

Barrington, A., & Pearson, K. (1906). On the inheritance of coat-colour in cattle: Part i. Shorthorn crosses and pure shorthorn crosses and pure shorthorns. *Biometrika*.

Baxter, L. L., Watkins-Chow, D. E., Pavan, W. J., & Loftus, S. K. (2019). A curated gene list for expanding the horizons of pigmentation biology. *Pigment Cell & Melanoma Research*, 32/3: 348–358. DOI: 10.1111/pcmr.12743

Bellone, R. R., Forsyth, G., Leeb, T., Archer, S., Sigurdsson, S., Imsland, F., Mauceli, E., *et al.* (2010). Fine-mapping and mutation analysis of TRPM1: a candidate gene for leopard complex (LP) spotting and congenital stationary night blindness in horses. *Briefings in functional genomics*, 9/3: 193–207. DOI: 10.1093/bfgp/elq002

Bellone, R. R., Holl, H., Setaluri, V., Devi, S., Maddodi, N., Archer, S., Sandmeyer, L., *et al.* (2013). Evidence for a retroviral insertion in TRPM1 as the cause of congenital stationary night blindness and leopard complex spotting in the horse. *Plos One*, 8/10: e78280. DOI: 10.1371/journal.pone.0078280

Bellone, R. R., Liu, J., Petersen, J. L., Mack, M., Singer-Berk, M., Drögemüller, C., Malvick, J., *et al.* (2017). A missense mutation in damage-specific DNA binding protein 2 is a genetic risk factor for limbal squamous cell carcinoma in horses. *International Journal of Cancer*, 141/2: 342–353. DOI: 10.1002/ijc.30744

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, 27/2: 573–580. DOI: 10.1093/nar/27.2.573

Blake, J. A., Eppig, J. T., Kadin, J. A., Richardson, J. E., Smith, C. L., Bult, C. J., & the Mouse Genome Database Group. (2017). Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. *Nucleic Acids Research*, 45/D1: D723–D729. DOI: 10.1093/nar/gkw1040

Brooks, S. A., & Bailey, E. (2005). Exon skipping in the KIT gene causes a Sabino spotting pattern in horses. *Mammalian Genome*, 16/11: 893–902. DOI: 10.1007/s00335-005-2472-y

Charlier, C., Denys, B., Belanche, J. I., Coppieters, W., Grobet, L., Mni, M., Womack, J., *et al.* (1996). Microsatellite mapping of the bovine roan locus: a major determinant of White Heifer disease. *Mammalian Genome*, 7/2: 138–142. DOI: 10.1007/s003359900034

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., *et al.* (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6/2: 80–92. DOI: 10.4161/fly.19695

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*, 34/15: 2666–2669. DOI: 10.1093/bioinformatics/bty149

Cothran, E. G., MacCluer, J. W., Weitkamp, L. R., & Guttormsen, S. A. (1986). Genetic variability, inbreeding, and reproductive performance in standardbred horses. *Zoo Biology*, 5/2: 191–201. DOI: 10.1002/zoo.1430050213

Cothran, E. G., MacCluer, J. W., Weitkamp, L. R., Pfennig, D. W., & Boyce, A. J. (1984). Inbreeding and reproductive performance in standardbred horses. *The Journal of Heredity*, 75/3: 220–224. DOI: 10.1093/oxfordjournals.jhered.a109916

D'Mello, S. A. N., Finlay, G. J., Baguley, B. C., & Askarian-Amiri, M. E. (2016). Signaling pathways in melanogenesis. *International Journal of Molecular Sciences*, 17/7. DOI: 10.3390/ijms17071144

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., *et al.* (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10/2. DOI: 10.1093/gigascience/giab008

David, V. A., Menotti-Raymond, M., Wallace, A. C., Roelke, M., Kehler, J., Leighty, R., Eizirik, E., *et al.* (2014). Endogenous retrovirus insertion in the KIT oncogene determines white and white spotting in domestic cats. *G3 (Bethesda, Md.)*, 4/10: 1881–1891. DOI: 10.1534/g3.114.013425

Dieringer, D., & Schlötterer, C. (2003). microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular ecology notes*, 3/1: 167–169. DOI: 10.1046/j.1471-8286.2003.00351.x

Druml, T., Grilz-Seger, G., Neuditschko, M., Horna, M., Ricard, A., Pausch, H., & Brem, G. (2018). Novel insights into Sabino1 and splashed white coat color patterns in horses. *Animal Genetics*, 49/3: 249–253. DOI: 10.1111/age.12657

Fontanesi, L., D'Alessandro, E., Scotti, E., Liotta, L., Crovetti, A., Chiofalo, V., & Russo, V. (2010). Genetic heterogeneity and selection signature at the KIT gene in pigs showing different coat colours and patterns. *Animal Genetics*, 41/5: 478–492. DOI: 10.1111/j.1365-2052.2010.02054.x

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv*.

González-Recio, O., López de Maturana, E., & Gutiérrez, J. P. (2007). Inbreeding depression on female fertility and calving ease in Spanish dairy cattle. *Journal of Dairy Science*, 90/12: 5744–5752. DOI: 10.3168/jds.2007-0203

Grilz-Seger, G., Reiter, S., Neuditschko, M., Wallner, B., Rieder, S., Leeb, T., Jagannathan, V., *et al.* (2020). A Genome-Wide Association Analysis in Noriker Horses Identifies a SNP Associated With Roan Coat Color. *Journal of equine veterinary science*, 88: 102950. DOI: 10.1016/j.jevs.2020.102950

Guenther, C. A., Tasic, B., Luo, L., Bedell, M. A., & Kingsley, D. M. (2014). A molecular basis for classic blond hair color in Europeans. *Nature Genetics*, 46/7: 748–752. DOI: 10.1038/ng.2991

Gutiérrez-Reinoso, M. A., Aponte, P. M., Cabezas, J., Rodriguez-Alvarez, L., & Garcia-Herreros, M. (2020). Genomic Evaluation of Primiparous High-Producing Dairy Cows: Inbreeding Effects on Genotypic and Phenotypic Production-Reproductive Traits. *Animals : an open access journal from MDPI*, 10/9. DOI: 10.3390/ani10091704

Hack, Y. L., Crabtree, E. E., Avila, F., Sutton, R. B., Grahn, R., Oh, A., Gilger, B., *et al.* (2021). Whole-genome sequencing identifies missense mutation in GRM6 as the likely cause of congenital stationary night blindness in a Tennessee Walking Horse. *Equine Veterinary Journal*, 53/2: 316–323. DOI: 10.1111/evj.13318

Hartl, D. L. (2020). Inbreeding and population structure. *A primer of population genetics and genomics*, pp. 47–74. Oxford University Press. DOI: 10.1093/oso/9780198862291.003.0003

Hisey, E. A., Hermans, H., Lounsberry, Z. T., Avila, F., Grahn, R. A., Knickelbein, K. E., Duward-Akhurst, S. A., *et al.* (2020). Whole genome sequencing identified a 16 kilobase deletion on ECA13 associated with distichiasis in Friesian horses. *BMC Genomics*, 21/1: 848. DOI: 10.1186/s12864-020-07265-8

Holl, H. M., Brooks, S. A., Archer, S., Brown, K., Malvick, J., Penedo, M. C. T., & Bellone, R. R. (2016). Variant in the RFWD3 gene associated with PATN1, a modifier of leopard complex spotting. *Animal Genetics*, 47/1: 91–101. DOI: 10.1111/age.12375

Hoogduijn, M. J., Ancans, J., Suzuki, I., Estdale, S., & Thody, A. J. (2002). Melanin-concentrating hormone and its receptor are expressed and functional in human skin. *Biochemical and Biophysical Research Communications*, 296/3: 698–701. DOI: 10.1016/s0006-291x(02)00932-4

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, 15/3: 651–674. DOI: 10.1198/106186006X133933

Hothorn, T., & Zeileis, A. (2015). partykit: A Modular Toolkit for Recursive Partytioning in R. *Journal of Machine Learning Research*.

Jansen, S., Aigner, B., Pausch, H., Wysocki, M., Eck, S., Benet-Pagès, A., Graf, E., *et al.* (2013). Assessment of the genomic variation in a cattle population by re-sequencing of key animals at low to medium coverage. *BMC Genomics*, 14: 446. DOI: 10.1186/1471-2164-14-446

Jeffares, D. C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F., *et al.* (2017). Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nature Communications*, 8: 14061. DOI: 10.1038/ncomms14061

Kalbfleisch, T. S., Rice, E. S., DePriest, M. S., Walenz, B. P., Hestand, M. S., Vermeesch, J. R., O Connell, B. L., *et al.* (2018). Improved reference genome for the domestic horse increases assembly contiguity and composition. *Communications Biology*, 1: 197. DOI: 10.1038/s42003-018-0199-z

Katila, T., Reilas, T., Nivola, K., Peltonen, T., & Virtala, A.-M. (2010). A 15-year survey of reproductive efficiency of Standardbred and Finnhorse trotters in Finland--descriptive results. *Acta veterinaria Scandinavica*, 52: 40. DOI: 10.1186/1751-0147-52-40

Kawakami, T., Jensen, M. K., Slavney, A., Deane, P. E., Milano, A., Raghavan, V., Ford, B., *et al.* (2021). R-locus for roaned coat is associated with a tandem duplication in an intronic region of USH2A in dogs and also contributes to Dalmatian spotting. *Plos One*, 16/3: e0248233. DOI: 10.1371/journal.pone.0248233

Kent, W. J. (2002). BLAT--the BLAST-like alignment tool. *Genome Research*, 12/4: 656–664. DOI: 10.1101/gr.229202

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Research*, 12/6: 996–1006. DOI: 10.1101/gr.229102

King, J. A. (1992). *Changes in Heterozygosity Through Time in American Standardbred and American Saddlebred Horses (1960-1990)* (Master thesis). University of North Texas.
Kingsley, N. B., Hamilton, N. A., Lindgren, G., Orlando, L., Bailey, E., Brooks, S., McCue, M., *et al.* (2021). Adopt-a-Tissue" Initiative Advances Efforts to Identify Tissue-Specific Histone Marks in the Mare. *Frontiers in genetics*, 12: 649959. DOI: 10.3389/fgene.2021.649959
Kingsley, N. B., Kern, C., Creppe, C., Hales, E. N., Zhou, H., Kalbfleisch, T. S., MacLeod, J. N., *et al.* (2019). Functionally Annotating Regulatory Elements in the Equine Genome Using Histone Mark ChIP-Seq. *Genes*, 11/1. DOI: 10.3390/genes11010003
Kraus, M., Physick-Sheard, P. W., Brito, L. F., Sargolzaei, M., & Schenkel, F. S. (2018). Marginal ancestral contributions to atrial fibrillation in the Standardbred racehorse: Comparison of cases and controls. *Plos One*, 13/5: e0197137. DOI: 10.1371/journal.pone.0197137
Kraus, M., Physick-Sheard, P. W., Brito, L. F., & Schenkel, F. S. (2017). Estimates of heritability of atrial fibrillation in the Standardbred racehorse. *Equine Veterinary Journal*, 49/6: 718–722. DOI: 10.1111/evj.12687
Laughlin, H. H. (1911). The Inheritance of Color in Shorthorn Cattle A Study in Somatic Blends Accompanying Gametic Segregation and Intra-Zygotic Inhibition and Reaction. *The American Naturalist*, 45/540: 705–742. DOI: 10.1086/279250
Lauvergne, J., & Canope, I. (1979). Étude de quelques variants colorés du porc Créole de la Guadeloupe. *Annales de genetique et de selection animale*, 11/4: 381–390. DOI: 10.1186/1297-9686-11-4-381
Law, M., & Shaw, D. R. (2018). Mouse Genome Informatics (MGI) Is the International Resource for Information on the Laboratory Mouse. *Methods in Molecular Biology*, 1757: 141–161. DOI: 10.1007/978-1-4939-7737-6_7
Levy, C., Khaled, M., & Fisher, D. E. (2006). MITF: master regulator of melanocyte development and melanoma oncogene. *Trends in Molecular Medicine*, 12/9: 406–414. DOI: 10.1016/j.molmed.2006.07.008
Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34/18: 3094–3100. DOI: 10.1093/bioinformatics/bty191
Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25/14: 1754–1760. DOI: 10.1093/bioinformatics/btp324
Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2/3: 18–22.
Lim, H. T., Zhong, T., Cho, I. C., Seo, B. Y., Kim, J. H., Lee, S. S., Ko, M. S., *et al.* (2011). Novel alternative splicing by exon skipping in KIT associated with whole-body roan in an intercrossed population of Landrace and Korean Native pigs. *Animal Genetics*, 42/4: 451–455. DOI: 10.1111/j.1365-2052.2011.02225.x
Liu, X., Bulgakov, O. V., Darrow, K. N., Pawlyk, B., Adamian, M., Liberman, M. C., & Li, T. (2007). Usherin is required for maintenance of retinal photoreceptors and normal development of cochlear hair cells. *Proceedings of the National Academy of Sciences of the United States of America*, 104/11: 4413–4418. DOI: 10.1073/pnas.0610950104
Locke, M. M., Penedo, M. C. T., Bricker, S. J., Millon, L. V., & Murray, J. D. (2002). Linkage of the grey coat colour locus to microsatellites on horse chromosome 25. *Animal Genetics*, 33/5: 329–337. DOI: 10.1046/j.1365-2052.2002.00885.x
Lykkjen, S., Olsen, H. F., Dolvik, N. I., Grøndahl, A. M., Røed, K. H., & Klemetsdal, G. (2014). Heritability estimates of tarsocrural osteochondrosis and palmar/plantar first phalanx osteochondral fragments in Standardbred trotters. *Equine Veterinary Journal*, 46/1: 32–37. DOI: 10.1111/evj.12058
MacCluer, J. W., Boyce, A. J., Dyke, B., Weitkamp, L. R., Pfenning, D. W., & Parsons, C. J. (1983). Inbreeding and pedigree structure in Standardbred horses. *Journal of Heredity*, 74/6: 394–399. DOI: 10.1093/oxfordjournals.jhered.a109824

MacKay, R. J. (2019). Treatment options for melanoma of gray horses. *The Veterinary clinics of North America. Equine practice*, 35/2: 311–325. DOI: 10.1016/j.cveq.2019.04.003
Madelaine, R., Ngo, K. J., Skariah, G., & Mourrain, P. (2020). Genetic deciphering of the antagonistic activities of the melanin-concentrating hormone and melanocortin pathways in skin pigmentation. *PLoS Genetics*, 16/12: e1009244. DOI: 10.1371/journal.pgen.1009244
Magdesian, K. G., Tanaka, J., & Bellone, R. R. (2020). A de novo MITF deletion explains a novel splashed white phenotype in an american paint horse. *The Journal of Heredity*, 111/3: 287–293. DOI: 10.1093/jhered/esaa009
Marklund, S., Moller, M., Sandberg, K., & Andersson, L. (1999). Close association between sequence polymorphism in the KIT gene and the roan coat color in horses. *Mammalian Genome*, 10/3: 283–288. DOI: 10.1007/s003359900987
McCarr, K. (1971). THE REGISTER -- One hundred Years old! *Hoof Beats*, 28, .
——. (1972). A century of standardbred registration; The second harness racing registrar. *Hoof Beats*.
McCoy, A. M., Beeson, S. K., Splan, R. K., Lykkjen, S., Ralston, S. L., Mickelson, J. R., & McCue, M. E. (2016). Identification and validation of risk loci for osteochondrosis in standardbreds. *BMC Genomics*, 17: 41. DOI: 10.1186/s12864-016-2385-z
McCoy, A. M., Norton, E. M., Kemper, A. M., Beeson, S. K., Mickelson, J. R., & McCue, M. E. (2019). SNP-based heritability and genetic architecture of tarsal osteochondrosis in North American Standardbred horses. *Animal Genetics*, 50/1: 78–81. DOI: 10.1111/age.12738
McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., *et al.* (2016). The ensembl variant effect predictor. *Genome Biology*, 17/1: 122. DOI: 10.1186/s13059-016-0974-4
Microsoft Corporation. (2018). *Microsoft Excel*. Computer software, Microsoft Corporation.
Müller-Unterberg, M., Wallmann, S., & Distl, O. (2017). Effects of inbreeding and other systematic effects on fertility of Black Forest Draught horses in Germany. *Acta veterinaria Scandinavica*, 59/1: 70. DOI: 10.1186/s13028-017-0338-4
NCBI Resource Coordinators. (2016). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 44/D1: D7–19. DOI: 10.1093/nar/gkv1290
Nevills, J. (2019). Ten Years Later: How The Stud Book Cap Has Affected Harness Racing. *Paulick Report*.
Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, 31/13: 3812–3814. DOI: 10.1093/nar/gkg509
Norton, E. M., Mickelson, J. R., Binns, M. M., Blott, S. C., Caputo, P., Isgren, C. M., McCoy, A. M., *et al.* (2016). Heritability of recurrent exertional rhabdomyolysis in standardbred and thoroughbred racehorses derived from SNP genotyping data. *The Journal of Heredity*, 107/6: 537–543. DOI: 10.1093/jhered/esw042
Orlando, L., & Librado, P. (2019). Origin and evolution of deleterious mutations in horses. *Genes*, 10/9. DOI: 10.3390/genes10090649
Petersen, J. L., Mickelson, J. R., Cothran, E. G., Andersson, L. S., Axelsson, J., Bailey, E., Bannasch, D., *et al.* (2013). Genetic diversity in the modern horse illustrated from genome-wide SNP data. *Plos One*, 8/1: e54997. DOI: 10.1371/journal.pone.0054997
Picardo, M., & Cardinali, G. (2011). The genetic determination of skin pigmentation: KITLG and the KITLG/c-Kit pathway as key players in the onset of human familial pigmentary diseases. *The Journal of Investigative Dermatology*, 131/6: 1182–1185. DOI: 10.1038/jid.2011.67
Promerová, M., Andersson, L. S., Juras, R., Penedo, M. C. T., Reissmann, M., Tozaki, T., Bellone, R., *et al.* (2014). Worldwide frequency distribution of the “Gait keeper” mutation in the DMRT3 gene. *Animal Genetics*, 45/2: 274–282. DOI: 10.1111/age.12120

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Computer software, R Foundation for Statistical Computing.
Raudsepp, T., Finno, C. J., Bellone, R. R., & Petersen, J. L. (2019). Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era. *Animal Genetics*, 50/6: 569–597. DOI: 10.1111/age.12857
Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29/1: 24–26. DOI: 10.1038/nbt.1754
Rosengren Pielberg, G., Golovko, A., Sundström, E., Curik, I., Lennartsson, J., Seltenhammer, M. H., Druml, T., *et al.* (2008). A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nature Genetics*, 40/8: 1004–1009. DOI: 10.1038/ng.185
Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., & Schatz, M. C. (2018). Accurate detection of complex structural variations using single-molecule sequencing. *Nature Methods*, 15/6: 461–468. DOI: 10.1038/s41592-018-0001-7
Seitz, J. J., Schmutz, S. M., Thue, T. D., & Buchanan, F. C. (1999). A missense mutation in the bovine MGF gene is associated with the roan phenotype in Belgian Blue and Shorthorn cattle. *Mammalian Genome*, 10/7: 710–712. DOI: 10.1007/s003359901076
Setaluri, V., & Jayanthy, A. (2013). Coat color mutations, animals. *Brenner's encyclopedia of genetics*, pp. 58–60. Elsevier. DOI: 10.1016/B978-0-12-374984-0.00275-8
Sevinga, M., Barkema, H. W., Stryhn, H., & Hesselink, J. W. (2004). Retained placenta in Friesian mares: incidence, and potential risk factors with special emphasis on gestational length. *Theriogenology*, 61/5: 851–859. DOI: 10.1016/s0093-691x(03)00260-7
Sevinga, M., Vrijenhoek, T., Hesselinks, J. W., Barkema, H. W., & Groen, A. F. (2004). Effect of inbreeding on the incidence of retained placenta in Friesian horses. *Journal of Animal Science*, 82/4: 982–986. DOI: 10.2527/2004.824982x
Sewall Wright. (1978). *Evolution and the Genetics of Populations, Volume 4: Variability Within and Among Natural Populations*. University of Chicago Press.
Smit, A. F. A., Hubley, R., & Green, P. (n.d.). *RepeatMasker Open-3.0*. Computer software, http://www.repeatmasker.org/. 1996-2010.
Sponenberg, D P, Harper, H. T., & Harper, A. L. (1984). Direct evidence for linkage of roan and extension loci in Belgian horses. *The Journal of Heredity*, 75/5: 413–414. DOI: 10.1093/oxfordjournals.jhered.a109968
Sponenberg, D P, & Rothschild, M. F. (2001). Genetics of coat colour and hair texture. Ruvinsky A. & Sampson J. (eds) *The genetics of the dog*, pp. 61–85. CABI: Wallingford. DOI: 10.1079/9780851995205.0061
Sponenberg, D. Phillip, & Bellone, R. (2017). *Equine Color Genetics*, 4th ed., p. 352. Hoboken, NJ, USA: Wiley-Blackwell.
Spritz, R. A., & Hearing, V. J. (2013). Chapter 145 - Abnormalities of Pigmentation. Rimoin D., Pyeritz R., & Korf B. (eds) *Emery and Rimoin's Principles and Practice of Medical Genetics*, 6th ed. Elsevier Ltd.
Staiger, E. A., Almén, M. S., Promerová, M., Brooks, S., Cothran, E. G., Imsland, F., Jäderkvist Fegraeus, K., *et al.* (2017). The evolutionary history of the DMRT3 "Gait keeper" haplotype. *Animal Genetics*, 48/5: 551–559. DOI: 10.1111/age.12580
Streett, D. (2017). *HTStream: A Toolkit for High Throughput Sequencing Analysis | Theses and Dissertations Collection* (Master thesis). University of Idaho. Retrieved January 1, 2019, from <https://www.lib.uidaho.edu/digital/etd/items/streett_idaho_0089n_11229.html>
Sulem, P., Gudbjartsson, D. F., Stacey, S. N., Helgason, A., Rafnar, T., Magnusson, K. P., Manolescu, A., *et al.* (2007). Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nature Genetics*, 39/12: 1443–1452. DOI: 10.1038/ng.2007.13

Sun, G., Liang, X., Qin, K., Qin, Y., Shi, X., Cong, P., Mo, D., *et al.* (2020). Functional analysis of KIT gene structural mutations causing the porcine dominant white phenotype using genome edited mouse models. *Frontiers in genetics*, 11: 138. DOI: 10.3389/fgene.2020.00138

Sundström, E., Komisarczuk, A. Z., Jiang, L., Golovko, A., Navratilova, P., Rinkwitz, S., Becker, T. S., *et al.* (2012). Identification of a melanocyte-specific, microphthalmia-associated transcription factor-dependent regulatory element in the intronic duplication causing hair greying and melanoma in horses. *Pigment Cell & Melanoma Research*, 25/1: 28–36. DOI: 10.1111/j.1755-148X.2011.00902.x

Talenti, A., Bertolini, F., Williams, J., Moaeen-Ud-Din, M., Frattini, S., Coizet, B., Pagnacco, G., *et al.* (2018). Genomic analysis suggests KITLG is responsible for a roan pattern in two pakistani goat breeds. *The Journal of Heredity*, 109/3: 315–319. DOI: 10.1093/jhered/esx093

Team, Rs. (2020). RStudio: Integrated Development Environment for R.

The Roan Quarter Horse Color. (2018). *American Quarter Horse Association*. Retrieved July 24, 2021, from <https://www.aqha.com/-/horse-breeding-the-roan-quarter-horse-color>

The Gene Ontology Consortium. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, 47/D1: D330–D338. DOI: 10.1093/nar/gky1055

Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14/2: 178–192. DOI: 10.1093/bib/bbs017

Todd, E. T., Hamilton, N. A., Velie, B. D., & Thomson, P. C. (2020). The effects of inbreeding on covering success, gestation length and foal sex ratio in Australian thoroughbred horses. *BMC Genetics*, 21/1: 41. DOI: 10.1186/s12863-020-00847-1

Toonen, R. J., & Hughes, S. (2001). Increased throughput for fragment analysis on an ABI PRISM 377 automated sequencer using a membrane comb and STRand software. *Biotechniques*, 31/6: 1320–1324.

UC Davis Veterinary Genetics Laboratory. (n.d.). Roan Zygosity Test. *UC Davis Veterinary Medicine, Veterinary Genetics Laboratory*. Retrieved September 4, 2021, from <https://vgl.ucdavis.edu/test/roan>

Unal, A., Ozkol, H. U., Bayram, Y., & Akdeniz, N. (2021). Comparison of tyrosinase antibody, tyrosinase-related protein-1 and -2 antibodies, melanin-concentrating hormone receptor antibody levels with autologous serum skin test and autologous plasma skin test results in patients with vitiligo. *Postepy dermatologii i alergologii*, 38/3: 473–479. DOI: 10.5114/ada.2020.93272

Valberg, S. J., Soave, K., Williams, Z. J., Perumbakkam, S., Schott, M., Finno, C. J., Petersen, J. L., *et al.* (2019). Coding sequences of sarcoplasmic reticulum calcium ATPase regulatory peptides and expression of calcium regulatory genes in recurrent exertional rhabdomyolysis. *Journal of Veterinary Internal Medicine*, 33/2: 933–941. DOI: 10.1111/jvim.15425

van Genderen, M. M., Bijveld, M. M. C., Claassen, Y. B., Florijn, R. J., Pearring, J. N., Meire, F. M., McCall, M. A., *et al.* (2009). Mutations in TRPM1 are a common cause of complete congenital stationary night blindness. *American Journal of Human Genetics*, 85/5: 730–736. DOI: 10.1016/j.ajhg.2009.10.012

Voß, K., Tetens, J., Thaller, G., & Becker, D. (2020). Coat Color Roan Shows Association with KIT Variants and No Evidence of Lethality in Icelandic Horses. *Genes*, 11/6. DOI: 10.3390/genes11060680

Wahler, B. (2011). *Arabian Coat Color Patterns' '* , pp. 2–3. Arabian Horse Association.

Wallace, J. H. (1867). *Wallace's American Stud-Book*, Vol. 1. New York: W A Townsend & Adams.

——. (1871). *Wallace's American Trotting Register*, Vol. 1. New York: Geo. E. Woodward, Orange Judd & Co.

——. (1885). *Wallace's American Trotting Register …*, Vol. V. New York: Office of Wallace's Monthly.
——. (1897). *The Horse of America in His Derivation, History and Development*. New York: J H Wallace.
Weich, K., Affolter, V., York, D., Rebhun, R., Grahn, R., Kallenberg, A., & Bannasch, D. (2020). Pigment Intensity in Dogs is Associated with a Copy Number Variant Upstream of KITLG. *Genes*, 11/1. DOI: 10.3390/genes11010075
Weir, B. S., & Cockerham, C. C. (1984). Estimating f-statistics for the analysis of population structure. *Evolution*, 38/6: 1358–1370. DOI: 10.1111/j.1558-5646.1984.tb05657.x
Wilson, J. (1908). Mendelian Characters among Shorthorns. *Nature*, 77/2005: 509–509. DOI: 10.1038/077509a0
Wong, A. K., Ruhe, A. L., Robertson, K. R., Loew, E. R., Williams, D. C., & Neff, M. W. (2013). A de novo mutation in KIT causes white spotting in a subpopulation of German Shepherd dogs. *Animal Genetics*, 44/3: 305–310. DOI: 10.1111/age.12006
Yue, X.-P., Dechow, C., & Liu, W.-S. (2015). A limited number of Y chromosome lineages is present in North American Holsteins. *Journal of Dairy Science*, 98/4: 2738–2745. DOI: 10.3168/jds.2014-8601
Zaiontz, C. (2021). *Real Statistics Resource Pack*. Computer software, www.real-statistics.com.
Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., Billis, K., *et al.* (2018). Ensembl 2018. *Nucleic Acids Research*, 46/D1: D754–D761. DOI: 10.1093/nar/gkx1098

Appendix A

Table A1. Pigmentation Candidate Gene List

ABCA4	*ATOX1*	*CASKIN1*	*COL17A1*	*DNM1L*	*ESCO2*	*GNAI3*	*HTR2C*
ABCB6	*ATP13A2*	*CASP3*	*COL18A1*	*DOCK7*	*ETS1*	*GNAQ*	*IDS*
ABI2	*ATP6AP2*	*CASR*	*COL1A1*	*DPH1*	*F2RL1*	*GNAS*	*IDUA*
ACD	*ATP7A*	*CBL*	*COP1*	*DPH6*	*FANCA*	*GNAT1*	*IFIH1*
ACE2	*ATP7B*	*CBS*	*COPA*	*DRAM2*	*FANCC*	*GNPAT*	*IGFBP3*
ACVR2A	*ATR*	*CCR2*	*CORIN*	*DRD2*	*FANCD2*	*GPC3*	*IHH*
ADAM10	*ATRN*	*CD44*	*COTL1*	*DSG4*	*FANCE*	*GPNMB*	*IKBKB*
ADAM17	*BACE2*	*CD46*	*CPD*	*DST*	*FANCI*	*GPR143*	*IKBKG*
ADAMTS20	*BARX2*	*CD63*	*CRAMP1*	*DSTYK*	*FAS*	*GPR161*	*ILK*
ADAMTS9	*BAX*	*CDC123*	*CRB1*	*DTNBP1*	*FAT1*	*GRK3*	*IMMP2L*
ADAMTSL4	*BBS2*	*CDC42*	*CRB2*	*DUSP7*	*FBN1*	*GRM1*	*IRF4*
ADAR	*BBS4*	*CDC42EP5*	*CRX*	*ECE1*	*FGFR2*	*GRM6*	*IRS2*
ADCY6	*BBS5*	*CDH3*	*CRYBA1*	*EDA*	*FGFR3*	*GRN*	*ITGB1*
AEBP2	*BBS7*	*CDK2*	*CRYBG1*	*EDAR*	*FIG4*	*GSTA4*	*JMJD6*
AFG3L2	*BCL2*	*CDK20*	*CSNK1A1*	*EDARADD*	*FLNA*	*GTF2IRD1*	*KANSL1*
ALDH2	*BCL2L11*	*CDK4*	*CSNK2A1*	*EDDM3B*	*FMN1*	*H2AFY*	*KAT14*
ALG1	*BEND3*	*CDK5*	*CTC1*	*EDN1*	*FMR1*	*H2AFY2*	*KDM8*
ALMS1	*BEST1*	*CDK6*	*CTLA4*	*EDN3*	*FOXC1*	*HDAC1*	*KIF13A*
ANK1	*BLM*	*CDK7*	*CTNNA1*	*EDNRB*	*FOXC2*	*HDAC2*	*KIT*
ANKRD27	*BLOC1S1*	*CDKN1A*	*CTNNB1*	*EED*	*FOXD3*	*HECTD1*	*KITLG*
AP1G1	*BLOC1S2*	*CDKN1B*	*CTNS*	*EFEMP1*	*FOXN1*	*HELLS*	*KRAS*
AP1M1	*BLOC1S3*	*CDKN2A*	*CXCL10*	*EGFR*	*FREM2*	*HEPH*	*KRT1*
AP3B1	*BLOC1S4*	*CDX1*	*CXCL17*	*EIF3C*	*FRYL*	*HERC2*	*KRT14*
AP3D1	*BLOC1S5*	*CEP290*	*CYB561*	*ELOVL3*	*FSCN1*	*HES1*	*KRT17*
APC	*BLOC1S6*	*CERS1*	*CYP11A1*	*ELOVL4*	*FUCA1*	*HGF*	*KRT2*
APOB	*BMP7*	*CFH*	*CYP19A1*	*EMX2*	*FUZ*	*HIF1A*	*KRT27*
APOE	*BMPR1A*	*CHEK1*	*DCC*	*EN1*	*FZD1*	*HPRT1*	*KRT4*
ARCN1	*BMPR1B*	*CHIC2*	*DCP2*	*ENDOG*	*FZD4*	*HPS1*	*KRT5*
ARHGAP35	*BMPR2*	*CHM*	*DCT*	*ENPP1*	*FZD7*	*HPS3*	*KRT75*
ARL6	*BMS1*	*CIDEA*	*DCTN1*	*EPG5*	*GAS1*	*HPS4*	*KRT76*
ARPC1B	*BNC2*	*CISD2*	*DCTN2*	*ERBB3*	*GATA3*	*HPS5*	*KRT9*
ASCL1	*BRAF*	*CITED1*	*DDB1*	*ERCC2*	*GBA*	*HPS6*	*KXD1*
ASIP	*BRCA1*	*CLCN2*	*DDB2*	*ERCC3*	*GGT1*	*HR*	*L1CAM*
ATE1	*BRCA2*	*CLCN7*	*DDX3X*	*ERCC4*	*GGT5*	*HRAS*	*LAMA4*
ATF1	*BRIP1*	*CLDN6*	*DDX59*	*ERCC5*	*GLI3*	*HS2ST1*	*LCA5*
ATG7	*BTD*	*CLN8*	*DEFB103A*	*ERCC6*	*GLMP*	*HSF4*	*LDLR*
ATM	*CARMIL2*	*CNGB1*	*DKC1*	*ERP44*	*GNA11*	*HSP90B1*	*LEF1*

LEP	*MFSD2A*	*NSDHL*	*POFUT1*	*RAB7B*	*SDC2*	*SNX5*	*TICAM1*
LGI1	*MFSD8*	*NSMCE2*	*POGLUT1*	*RAB9A*	*SEMA3C*	*SOD2*	*TIMP3*
LIPH	*MGRN1*	*NTRK1*	*POLA1*	*RAB9B*	*SEMA4A*	*SOX10*	*TINF2*
LMNA	*MITF*	*NUAK1*	*POLE4*	*RABGGTA*	*SEMA4C*	*SOX18*	*TLR4*
LMO7	*MKKS*	*NUAK2*	*POLG*	*RAC1*	*SENP7*	*SOX2*	*TMEM138*
LMX1A	*MKLN1*	*NXF1*	*POLH*	*RACK1*	*SERPINC1*	*SOX5*	*TMEM163*
LOC100053137	*MLANA*	*OAT*	*POMC*	*RAD50*	*SERPINF1*	*SOX9*	*TMEM173*
LRMDA	*MLH1*	*OCA2*	*POU1F1*	*RAD9A*	*SETD5*	*SPAG9*	*TMEM30B*
LRP2	*MLPH*	*OPHN1*	*POU3F2*	*RAF1*	*SFRP4*	*SPARC*	*TMEM79*
LRP5	*MPV17*	*OSTM1*	*PPARGC1A*	*RAG1*	*SFXN1*	*SPNS2*	*TP53*
LRRK2	*MPZL3*	*OTC*	*PPP1R32*	*RALY*	*SGPL1*	*SPRED1*	*TP63*
LVRN	*MRAP*	*OTX2*	*PPP5C*	*RAPH1*	*SHOC2*	*SPTA1*	*TPCN2*
LYST	*MREG*	*PAH*	*PRKAR1A*	*RB1*	*SHROOM2*	*SRC*	*TRAF4*
MAB21L1	*MSH2*	*PALB2*	*PRKAR2B*	*RBP1*	*SHROOM3*	*SRSF4*	*TRAF6*
MAGOH	*MSH6*	*PARD3*	*PRKCI*	*RBPJ*	*SIK2*	*ST3GAL5*	*TRAPPC6A*
MAP2K1	*MSX2*	*PARN*	*PRKCQ*	*RCC2*	*SIK3*	*STAR*	*TRP63*
MAPK1	*MTA1*	*PAX1*	*PRKCZ*	*RDH8*	*SIRT7*	*STK11*	*TRPM1*
MAPK3	*MTAP*	*PAX2*	*PRKDC*	*RECQL4*	*SKP1*	*STRA6*	*TRPM7*
MBTPS1	*MYC*	*PAX3*	*PROM1*	*RERE*	*SLC17A5*	*STX17*	*TRPS1*
MBTPS2	*MYL2*	*PAX4*	*PRPF3*	*REST*	*SLC24A4*	*SUFU*	*TSC1*
MC1R	*MYO10*	*PAX6*	*PRPF31*	*RET*	*SLC24A5*	*SUZ12*	*TSC2*
MC2R	*MYO5A*	*PCARE*	*PRPF8*	*RFWD3*	*SLC26A4*	*SZT2*	*TUB*
MC5R	*MYO7A*	*PCBD1*	*PRPH2*	*RHBDF2*	*SLC29A3*	*TACO1*	*TULP1*
MCHR1	*MYSM1*	*PCNT*	*PSENEN*	*RHO*	*SLC30A4*	*TAF4*	*TWIST2*
MCM2	*NADK2*	*PDE3B*	*PSMB7*	*RIT1*	*SLC31A1*	*TAL1*	*TYR*
MCM4	*NAGLU*	*PDE4D*	*PTCH1*	*RLN3*	*SLC35D3*	*TBC1D32*	*TYRP1*
MCOLN3	*NBN*	*PDE6B*	*PTEN*	*RLPB1*	*SLC36A1*	*TBX10*	*UNC119*
MCPH1	*NCOA6*	*PDGFB*	*PTPN11*	*RPE65*	*SLC45A2*	*TBX15*	*UQCRFS1*
MCRIP1	*NDOR1*	*PDGFC*	*PTPN6*	*RPGR*	*SLC6A19*	*TBX19*	*USF2*
MDM1	*NDP*	*PDPK1*	*PTS*	*RPL24*	*SLC6A8*	*TBX2*	*USH2A*
MDM2	*NF1*	*PEPD*	*PXDN*	*RPL27A*	*SLC7A11*	*TBX3*	*USP13*
MDM4	*NHLRC1*	*PER2*	*PYGO1*	*RPL38*	*SMARCA4*	*TDO2*	*USP39*
MED1	*NMNAT1*	*PFAS*	*RAB10*	*RPS19*	*SMARCA5*	*TERF1*	*USP9X*
MEF2C	*NNT*	*PHACTR4*	*RAB11A*	*RPS20*	*SMARCAL1*	*TERF2*	*UVSSA*
MEMO1	*NOTCH1*	*PICALM*	*RAB11B*	*RPS6*	*SMARCC1*	*TERF2IP*	*VAC14*
MEN1	*NOTCH2*	*PIKFYVE*	*RAB17*	*RPS7*	*SMC3*	*TERT*	*VANGL1*
MERTK	*NPHP4*	*PITX2*	*RAB1A*	*RS1*	*SMCHD1*	*TET1*	*VHL*
MET	*NR0B1*	*PITX3*	*RAB27A*	*RUVBL2*	*SMO*	*TFAP2A*	*VLDLR*
METTL16	*NR2E1*	*PKNOX1*	*RAB27B*	*RXRA*	*SMOC1*	*TFE3*	*VPS33A*
METTL7B	*NR2F1*	*PLXNB2*	*RAB29*	*S1PR2*	*SNAI2*	*TFEB*	*VPS33B*
MFN2	*NR2F2*	*PMCH*	*RAB32*	*SAMD9*	*SNAPIN*	*TFEC*	*VSX2*
MFRP	*NRAS*	*PMEL*	*RAB38*	*SASH1*	*SNX1*	*TGFBR2*	*WASHC5*
MFSD12	*NRL*	*PMS2*	*RAB7A*	*SBNO1*	*SNX13*	*TH*	*WDR12*

WFDC5	*ZDHHC21*	*ZNF804B*
YWHAE	*ZEB1*	*ZWIM5*
YWHAZ	*ZEB2*	*ZZEF1*
YY1	*ZIC2*	
ZBTB17	*ZMPSTE24*	

Appendix B
Thermocycler and reaction protocols

Table B1: Thermocycler protocol for INDELs. Asterisk denotes time that was lengthened to 45 seconds when genotyping the INDELs and ECA28:38,000,030 and ECA28:38,605,308. Volumes of reagents are for genotyping one sample.

Temperature (°C)	Time	Cycles
95	5 minutes	1
85	5 minutes	
95	1 minute	5
60	30 seconds	
72	30 seconds*	
95	45 seconds	28
60	30 seconds	
72	30 seconds*	
72	30 minutes	1
10	∞	

Table B2: Reaction protocol for INDELs

Reagent	Volume (µl)
DNA	2.0
Primer Mix	5.0
Water	4.53
10X Buffer	1.67
25mM $MgCl_2$	1.67
8mM dNTPs	1.67
DMSO	0.33
5µ/ul GoTaq G2 Flexi DNA Polymerase	0.13
Chill-Out™ Liquid Wax	15
Total Reaction Size	**32**

Table B3: Thermocycler protocol for MassARRAY PCR.

Temperature (°C)	Time	Cycles
94	2 minutes	1
94	30 seconds	45
56	30 seconds	
72	1 minute	
72	5 minutes	1
10	∞	

Table B4: Reaction protocol for MassARRAY PCR.

Reagents	Volume (µl)
HPLC Water	2.3
10X PCR Buffer with 20mM MgCl2	0.50
25mM MgCl2	0.40
25mM dNTP Mix	0.10
1uM Primer Mix	0.50
5µ/ul PCR enzyme	0.20
Total Reagent Vol	**4**

Table B5: Thermocycler protocol for MassARRAY shrimp alkaline phosphatase (SAP) cleanup.

Temperature (°C)	Time	Cycles
37	40 minutes	1
85	5 minutes	
10	∞	

Table B6: Reaction protocol for MassARRAY shrimp alkaline phosphatase (SAP) cleanup.

Reagents	Volume (µl)
HPLC Water	1.53
10X SAP Buffer	0.17
SAP Enzyme (1.7µ/ul)	0.30
Total Reagent Vol	**2**

Table B7: Thermocycler protocol for MassARRAY extension.

Temperature (°C)	Time	Cycles
94	30 seconds	1
94	5 seconds	40
52	5 seconds	
80	3 seconds	
52	5 seconds	
80	3 seconds	
52	5 seconds	
80	3 seconds	
52	5 seconds	
80	3 seconds	
52	5 seconds	
80	3 seconds	
72	3 minutes	1
10	∞	

Table B8: Reaction protocol for MassARRAY extension.

Reagents	Volume (µl)
HPLC Water	0.619
iPLEX® Buffer	0.20
iPLEX® Termination Mix	0.20
Extend Primer Mix	0.94
iPLEX® Enzyme	0.041
Total Volume	**2**

Appendix C

STRand images of the 3,254bp deletion at ECA28:13,938,253-13,941,507.

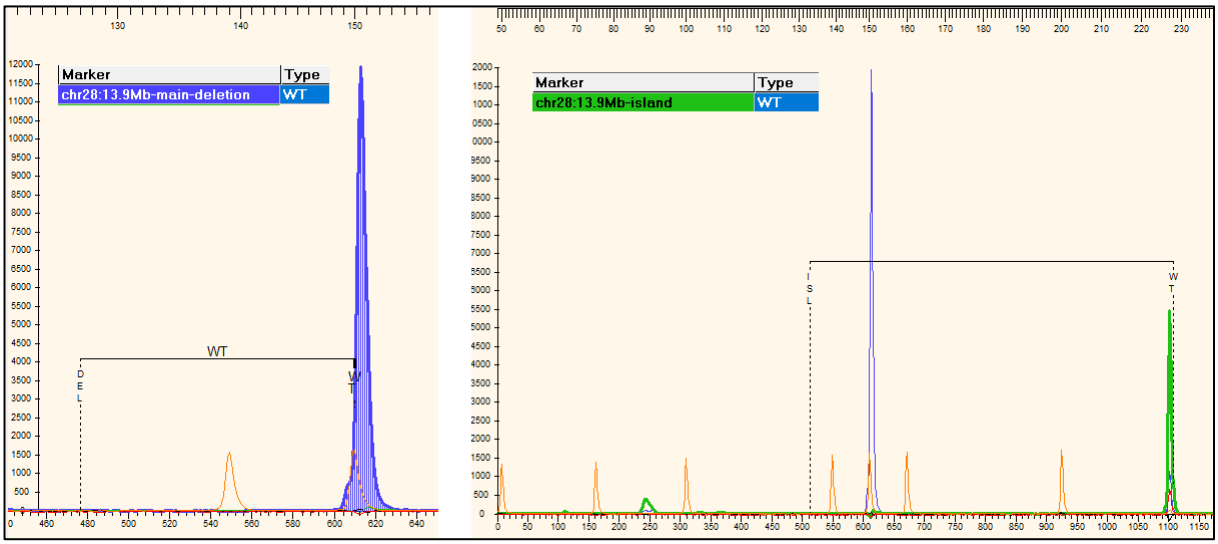

Figure C1: Representation of a wild type individual for the 3,254bp deletion at ECA28:13,938,253-13,941,507 as viewed in STRand. The left pane shows the presence of the wild type allele, as captured with a primer within the deletion and one outside. The right pane shows the presence of the "island" at ECA28:13,940,468-13,940,544, as captured by a primer on that region and a primer within the deletion.

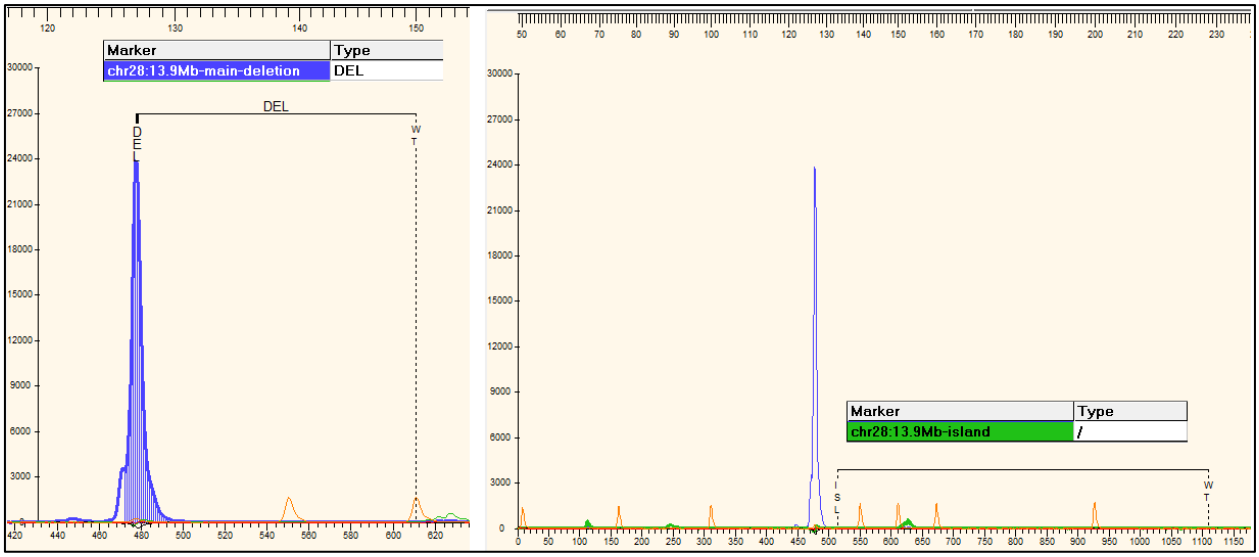

Figure C2: Representation of a homozygous alternate individual for the 3,254bp deletion atECA28:13,938,253-13,941,507 as viewed in STRand. The left pane shows the presence of the deletion as captured by a primer on each end of the deletion. The right pane shows the absence of the "island" spanning fromECA28:13,940,468-13,940,544. Presence of the island in a homozygous alternate individual would have been captured by a reverse primer within the "island" and a forward primer outside the deletion. The product size of the 3,254bp deletion was as expected, further suggesting a lack of the interior region.

Appendix D

Genotypes of all cases and controls.

Table D1.1: Phenotypes, breeds, membership in family of interest, and genotypes of all cases and controls. IDs were assigned arbitrarily so individuals can be found in both Table D1.1 and Table D1.2.

ID	Phenotype	Breed	Part of Family of Interest	ECA28:10,019,065-10,019,144	ECA28:13,938,253-13,941,507	ECA28:15,067,299	ECA28:15,431,302	ECA28:15,442,561	ECA28:15,446,190	ECA28:15,485,999	ECA28:15,503,208	ECA28:15,504,948	ECA28:15,518,406-15,518,409	ECA28:15,967,332	ECA28:16,465,850
C1	Control	AR	No	Del/WT	WT	G	T	C	C	WT	WT	T	WT	G	A
C2	Control	AR	No	Del/WT	Del/WT	A	T	C	C	WT	WT	T	WT	G	A
C3	Control	AR	No	Del/WT	WT	A	T	C	C	WT	WT	T	WT	G	A
C4	Control	MOR	No	Del/WT	Del/WT	A	T	C	C	WT	WT	T	WT	G	A
C5	Control	MOR	No	WT	Del	A	T	/	C	WT	WT2	T	WT	G	A
C6	Control	QH	No	WT	WT	A	T	C	C	WT	WT	T	WT	G	A
C7	Control	QH	No	WT	WT	A	T	C	C	WT	WT/WT2	T	WT	G	A
C8	Control	QH	No	Del/WT	Del/WT	A	T	C	C	WT	WT/WT2	T	WT	G	A
C9	Control	QH	No	WT	WT	A	T	C	C	WT	WT/WT2	T	WT	G	A
C10	Control	QH	No	Del/WT	Del/WT	G	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
C11	Control	QH	No	WT	WT	A	T	/	C	WT	WT2	T	WT	G	A
C12	Control	QH	No	Del/WT	Del/WT	GA	CT	T	CT	Del/WT	WT2/Ins	GT	Del/WT	AG	TA
C13	Control	QH	No	Del/WT	WT	A	T	C	C	WT	WT	T	WT	G	A
C14	Control	QH	No	Del/WT	WT	GA	T	C	C	WT	WT	T	WT	G	A
C15	Control	QH	No	Del/WT	WT	A	T	C	C	WT	WT	T	WT	G	A
C16	Control	QH	No	Del/WT	WT	A	T	C	C	WT	WT	T	WT	G	A
C17	Control	QH	No	Del/WT	Del/WT	GA	CT	T	CT	Del/WT	WT2/Ins	GT	Del/WT	AG	TA
C18	Control	QH	No	Del/WT	WT	A	T	C	C	WT	WT	T	WT	G	A
C19	Control	QH	No	Del	WT	A	T	T	CT	Del/WT	WT2/Ins	GT	Del/WT	G	A
C20	Control	QH	No	Del	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
C21	Control	QH	No	WT	WT	A	T	/	C	WT	WT2/Ins	T	WT	G	A
C22	Control	QH	No	WT	WT	A	T	C	C	WT	WT	T	WT	G	A

C23	Control	QH	No	Del/WT	WT	A	T	C	C	WT	WT	T	WT	G	A
C24	Control	QH	No	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
C25	Control	QH	No	WT	WT	A	T	C	C	WT	WT/WT2	T	WT	G	A
C26	Control	QH	No	WT	WT	A	T	C	C	WT	WT/WT2	T	WT	G	A
C27	Control	QH	No	WT	WT	A	T	CT	CT	Del/WT	Ins/WT	GT	Del/WT	G	A
C28	Control	QH	Unknown	WT	WT	A	T	C	C	WT	WT	T	WT	G	A
C29	Control	QH	Yes	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
C30	Control	QH	Yes	Del/WT	Del/WT	A	T	C	C	WT	WT	T	WT	G	A
C31	Control	QH	Yes	Del/WT	WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
C32	Control	QH	Yes	Del/WT	Del/WT	A	T	C	C	WT	WT	T	WT	G	A
C33	Control	QH	Yes	Del/WT	WT	A	T	C	C	WT	WT	T	WT	G	A
C34	Control	QH	Yes	Del/WT	Del/WT	A	T	C	C	WT	WT/WT2	T	WT	G	A
C35	Control	TB	No	Del	WT	A	T	C	C	WT	WT	T	WT	G	A
C36	Control	TB	No	WT	WT	A	T	C	C	WT	WT	T	WT	G	A
R1	Rabicano	AR	No	Del	WT	A	T	C	C	WT	WT	T	WT	A	A
R2	Rabicano	CP	No	Del/WT	WT	A	T	C	C	WT	WT	T	WT	G	A
R3	Rabicano	KWPN	No	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	A	T
R4	Rabicano	QH	No	WT	Del/WT	G	C	T	T	Del	Ins	G	Del	A	T
R5	Rabicano	QH	No	Del	Del	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R6	Rabicano	QH	No	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R7	Rabicano	QH	No	WT	Del/WT	GA	T	C	C	WT	WT	T	WT	AG	A
R8	Rabicano	QH	No	WT	Del/WT	A	T	C	C	WT	WT/WT2	T	WT	G	A
R9	Rabicano	QH	No	Del	Del/WT	GA	CT	T	CT	Del/WT	WT2/Ins	GT	Del/WT	AG	TA
R10	Rabicano	QH	No	Del/WT	WT	A	T	C	C	WT	WT/WT2	T	WT	G	A
R11	Rabicano	QH	No	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R12	Rabicano	QH	No	Del/WT	Del/WT	GA	CT	T	CT	Del/WT	WT2/Ins	GT	Del/WT	AG	TA
R13	Rabicano	QH	No	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R14	Rabicano	QH	No	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R15	Rabicano	QH	No	Del/WT	Del/WT	G	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R16	Rabicano	QH	No	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R17	Rabicano	QH	No	Del	WT	A	T	C	C	WT	WT	T	WT	A	A
R18	Rabicano	QH	No	Del	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R19	Rabicano	QH	No	Del	Del	G	C	T	T	Del	Ins	G	Del	A	T
R20	Rabicano	QH	No	WT	WT	A	T	C	C	WT	WT/WT2	T	WT	G	A
R21	Rabicano	QH	Yes	Del/WT	Del	G	C	T	T	Del	Ins	G	Del	A	T
R22	Rabicano	QH	Yes	Del	Del	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA

R23	Rabicano	QH	Yes	Del	Del	A	T	C	C	WT	WT/WT2	T	WT	G	A
R24	Rabicano	QH	Yes	Del	Del	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R25	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R26	Rabicano	QH	Yes	Del/WT	Del/WT	A	T	C	C	WT	WT	T	WT	G	A
R27	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	T	CT	Del/WT	WT2/Ins	GT	Del/WT	AG	TA
R28	Rabicano	QH	Yes	Del	Del	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R29	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	T	CT	Del/WT	WT2/Ins	GT	Del/WT	AG	TA
R30	Rabicano	QH	Yes	Del	Del	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R31	Rabicano	QH	Yes	Del	Del	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R32	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R33	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R34	Rabicano	QH	Yes	Del/WT	Del/WT	A	T	C	C	WT	WT/WT2	T	WT	G	A
R35	Rabicano	QH	Yes	Del/WT	Del/WT	A	T	C	C	WT	WT/WT2	T	WT	G	A
R36	Rabicano	QH	Yes	Del/WT	Del/WT	A	T	C	C	WT	WT/WT2	T	WT	G	A
R37	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R38	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	T	CT	Del/WT	WT2/Ins	GT	Del/WT	AG	TA
R39	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R40	Rabicano	QH	Yes	Del/WT	Del/WT	GA	T	C	C	WT	WT	T	WT	G	A
R41	Rabicano	QH	Yes	Del/WT	Del	G	C	T	T	Del	Ins	G	Del	A	T
R42	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	T	CT	Del/WT	WT2/Ins	GT	Del/WT	AG	TA
R43	Rabicano	QH	Yes	Del	Del	G	C	T	T	Del	Ins	G	Del	A	T
R44	Rabicano	QH	Yes	Del/WT	Del/WT	G	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R45	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R46	Rabicano	QH	Yes	Del	Del	G	C	T	T	Del	Ins	G	Del	A	T
R47	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	T	CT	Del/WT	WT2/Ins	GT	Del/WT	AG	TA
R48	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	T	CT	Del/WT	WT2/Ins	GT	Del/WT	AG	TA
R49	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	T	CT	Del/WT	WT2/Ins	GT	Del/WT	AG	TA
R50	Rabicano	QH	Yes	WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R51	Rabicano	QH	Yes	WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R52	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R53	Rabicano	QH	Yes	Del/WT	Del/WT	A	T	C	C	WT	WT	T	WT	G	A
R54	Rabicano	QH	Yes	Del/WT	Del/WT	A	T	C	C	WT	WT	T	WT	G	TA
R55	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R56	Rabicano	QH	Yes	Del/WT	Del	G	C	T	T	Del	Ins	G	Del	A	T
R57	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA

R58	Rabicano	QH	Yes	Del/WT	Del/WT	G	C	T	T	Del	Ins	G	Del	A	T
R59	Rabicano	QH	Yes	Del/WT	Del/WT	GA	CT	CT	CT	Del/WT	Ins/WT	GT	Del/WT	AG	TA
R60	Rabicano	WB	No	Del/WT	Del/WT	A	T	C	C	WT	WT	T	WT	AG	A
R61	Rabicano	WEL	No	Del/WT	Del/WT	GA	CT	T	CT	Del/WT	WT2/Ins	GT	Del/WT	AG	TA

Table D1.2: Continuation of phenotypes, breeds, membership in family of interest, and genotypes of all cases and controls. IDs were assigned arbitrarily so individuals can be found in both Table D1.1 and Table D1.2.

ID	Phenotype	Breed	Part of Family of Interest	ECA28:16,616,952	ECA28:16,764,837	ECA28:31,573,538-31,573,779	ECA28:38,000,030	ECA28:38,605,308	ECA28:39,293,149	ECA28:39,296,303	ECA28:39,300,931	ECA1:114,505,701	ECA1:114,505,7012	ECA10:4,049,610	ECA20:17,241,786	ECA30:26,482,977
C1	Control	AR	No	T	G	Del/Del2	WT	Ins	G	C	G	CA	G	C	C	C
C2	Control	AR	No	T	G	Del	WT	WT	G	C	G	C	G	C	C	C
C3	Control	AR	No	T	G	Del/Del2	WT	WT	G	C	G	C	G	C	C	CT
C4	Control	MOR	No	T	G	Del	WT	WT/Ins	G	C	G	CA	GT	C	C	C
C5	Control	MOR	No	T	G	Del	WT	WT	AG	CT	GA	C	G	C	CT	C
C6	Control	QH	No	T	G	Del/Del2	WT	Ins	G	C	G	C	G	C	C	T
C7	Control	QH	No	T	G	Del/Del2	WT/Ins	Ins	AG	CT	GA	C	GT	C	T	CT
C8	Control	QH	No	T	G	Del/Del2	WT	WT/Ins	AG	CT	GA	CA	T	C	C	CT
C9	Control	QH	No	T	G	Del	WT	WT/Ins	AG	CT	GA	C	GT	CT	C	C
C10	Control	QH	No	TA	AG	Del	WT	WT	AG	CT	GA	C	G	C	C	C
C11	Control	QH	No	T	G	Del	WT/Ins	WT/Ins	A	T	A	CA	GT	C	C	C
C12	Control	QH	No	TA	AG	Del	WT/Ins	WT/Ins	A	T	A	CA	G	C	C	C
C13	Control	QH	No	T	G	Del	WT	WT	G	C	G	A	GT	C	CT	CT
C14	Control	QH	No	T	G	Del	WT	WT	G	C	G	C	G	C	C	C
C15	Control	QH	No	T	G	Del/Del2	WT	WT/Ins	G	C	G	CA	G	C	C	C
C16	Control	QH	No	T	G	Del	WT	WT	G	C	G	A	GT	C	C	CT
C17	Control	QH	No	TA	A	Del	WT	WT/Ins	AG	CT	GA	CA	GT	C	C	C
C18	Control	QH	No	T	G	Del/WT	WT	WT	G	C	G	C	G	C	C	CT

C19	Control	QH	No	T	G	Del2	WT	WT	G	C	G	CA	GT	C	C	CT
C20	Control	QH	No	TA	AG	Del/WT	WT	WT	AG	CT	GA	CA	G	C	C	T
C21	Control	QH	No	T	AG	Del	WT	WT	AG	CT	GA	CA	GT	C	C	CT
C22	Control	QH	No	T	AG	Del	WT	WT/Ins	G	C	G	CA	GT	C	CT	CT
C23	Control	QH	No	T	G	Del	WT	Ins	G	C	G	C	GT	C	C	C
C24	Control	QH	No	TA	AG	Del	WT/Ins	WT/Ins	A	T	A	CA	GT	C	C	CT
C25	Control	QH	No	T	G	Del/Del2	WT	WT	AG	CT	GA	A	G	C	C	CT
C26	Control	QH	No	T	G	Del	WT	WT	AG	CT	GA	CA	GT	C	C	T
C27	Control	QH	No	T	AG	Del	WT	WT	G	C	G	CA	G	C	C	CT
C28	Control	QH	Unknown	T	G	Del	WT	WT/Ins	G	C	G	C	GT	CT	CT	C
C29	Control	QH	Yes	TA	AG	Del/WT	WT/Ins	WT/Ins	A	T	A	A	GT	C	CT	T
C30	Control	QH	Yes	T	G	Del	WT/Ins	WT/Ins	A	T	A	A	GT	C	CT	CT
C31	Control	QH	Yes	TA	AG	Del	WT/Ins	Ins	AG	CT	GA	CA	G	C	CT	CT
C32	Control	QH	Yes	T	G	Del	WT/Ins	Ins	AG	CT	GA	C	GT	CT	CT	C
C33	Control	QH	Yes	T	G	Del	WT	WT	A	T	A	CA	GT	T	CT	CT
C34	Control	QH	Yes	T	G	Del	WT/Ins	WT/Ins	AG	CT	GA	CA	G	CT	T	CT
C35	Control	TB	No	T	G	Del	WT	WT	G	C	G	CA	GT	C	C	CT
C36	Control	TB	No	T	G	Del	WT	Ins	G	C	G	A	G	C	C	C
R1	Rabicano	AR	No	T	G	Del	WT/Ins	Ins	G	C	G	CA	GT	CT	C	C
R2	Rabicano	CP	No	T	G	Del	WT	WT	G	C	G	CA	G	C	C	C
R3	Rabicano	KWPN	No	A	A	Del/Del2	WT	WT	AG	CT	GA	C	GT	C	C	CT
R4	Rabicano	QH	No	A	G	Del	WT/Ins	WT/Ins	AG	CT	GA	C	GT	C	C	C
R5	Rabicano	QH	No	TA	AG	Del	Ins	Ins	AG	CT	GA	A	GT	CT	T	CT
R6	Rabicano	QH	No	TA	AG	Del	WT	WT	G	C	G	C	GT	C	C	C
R7	Rabicano	QH	No	T	G	Del	WT	WT	G	C	G	CA	GT	C	C	CT
R8	Rabicano	QH	No	T	G	Del/WT	WT	WT	G	C	G	CA	G	C	C	C
R9	Rabicano	QH	No	TA	AG	Del	WT/Ins	WT/Ins	A	T	A	A	GT	C	C	CT
R10	Rabicano	QH	No	T	AG	Del	WT	WT/Ins	G	C	G	CA	GT	C	C	C
R11	Rabicano	QH	No	TA	AG	Del	WT/Ins	WT/Ins	A	T	A	C	G	C	C	CT
R12	Rabicano	QH	No	TA	AG	Del	WT/Ins	WT/Ins	A	T	A	CA	GT	C	C	CT
R13	Rabicano	QH	No	TA	AG	Del	WT	WT	A	T	A	C	G	C	C	CT
R14	Rabicano	QH	No	TA	AG	Del	WT/Ins	WT/Ins	G	C	G	C	GT	C	CT	C
R15	Rabicano	QH	No	TA	AG	Del/Del2	WT	WT	AG	CT	GA	C	G	C	C	C
R16	Rabicano	QH	No	TA	A	Del	WT	/	G	C	G	C	G	C	CT	C
R17	Rabicano	QH	No	T	G	Del/Del2	WT	WT	G	C	G	C	G	C	C	CT
R18	Rabicano	QH	No	TA	AG	Del2	WT	WT/Ins	AG	CT	GA	C	GT	C	CT	CT

R19	Rabicano	QH	No	A	A	Del	WT	WT	G	C	G	CA	G	C	C	C
R20	Rabicano	QH	No	T	G	Del/Del2	WT	WT	G	C	G	CA	G	C	CT	C
R21	Rabicano	QH	Yes	A	A	Del	WT/Ins	WT/Ins	G	C	G	A	T	C	CT	C
R22	Rabicano	QH	Yes	TA	AG	Del	WT/Ins	Ins	AG	CT	GA	CA	GT	C	CT	T
R23	Rabicano	QH	Yes	T	G	Del	WT/Ins	WT/Ins	A	T	A	CA	GT	CT	CT	T
R24	Rabicano	QH	Yes	TA	AG	Del	WT/Ins	WT/Ins	G	C	G	CA	G	C	CT	CT
R25	Rabicano	QH	Yes	TA	AG	Del	Ins	Ins	A	T	A	A	G	C	CT	T
R26	Rabicano	QH	Yes	T	G	Del	Ins	Ins	AG	CT	GA	CA	GT	CT	CT	C
R27	Rabicano	QH	Yes	TA	AG	Del	WT/Ins	WT/Ins	A	T	A	A	T	C	CT	C
R28	Rabicano	QH	Yes	TA	AG	Del	Ins	Ins	A	T	A	CA	GT	CT	CT	CT
R29	Rabicano	QH	Yes	TA	AG	Del	WT/Ins	Ins	AG	CT	GA	A	GT	T	CT	CT
R30	Rabicano	QH	Yes	TA	AG	Del	Ins	Ins	A	T	A	CA	GT	CT	CT	CT
R31	Rabicano	QH	Yes	TA	AG	Del	WT/Ins	Ins	G	C	G	A	GT	CT	CT	CT
R32	Rabicano	QH	Yes	TA	AG	Del	Ins	Ins	AG	CT	GA	A	G	CT	CT	CT
R33	Rabicano	QH	Yes	TA	AG	Del	Ins	Ins	A	T	A	A	GT	C	CT	CT
R34	Rabicano	QH	Yes	T	G	Del	WT/Ins	WT/Ins	A	T	A	CA	G	C	CT	CT
R35	Rabicano	QH	Yes	T	G	Del	WT/Ins	WT/Ins	AG	CT	GA	A	GT	CT	T	C
R36	Rabicano	QH	Yes	T	G	Del/WT	Ins	Ins	AG	CT	GA	A	GT	CT	CT	T
R37	Rabicano	QH	Yes	TA	AG	Del/Del2	WT/Ins	Ins	AG	CT	GA	CA	T	C	CT	T
R38	Rabicano	QH	Yes	TA	AG	Del	WT/Ins	WT/Ins	AG	CT	GA	A	T	CT	CT	CT
R39	Rabicano	QH	Yes	TA	AG	Del	WT/Ins	Ins	AG	CT	GA	CA	GT	C	T	CT
R40	Rabicano	QH	Yes	T	G	Del	WT/Ins	WT/Ins	AG	CT	GA	CA	G	CT	CT	CT
R41	Rabicano	QH	Yes	A	A	Del	WT/Ins	Ins	AG	CT	GA	CA	GT	C	T	CT
R42	Rabicano	QH	Yes	TA	AG	Del	WT/Ins	WT/Ins	AG	CT	GA	A	G	C	CT	T
R43	Rabicano	QH	Yes	A	A	Del	Ins	Ins	A	T	A	CA	T	CT	C	T
R44	Rabicano	QH	Yes	TA	AG	Del	WT/Ins	WT/Ins	AG	CT	GA	CA	GT	CT	CT	CT
R45	Rabicano	QH	Yes	TA	AG	Del	Ins	Ins	AG	CT	GA	CA	G	CT	CT	CT
R46	Rabicano	QH	Yes	A	A	Del	WT/Ins	Ins	AG	CT	GA	CA	GT	C	CT	CT
R47	Rabicano	QH	Yes	TA	AG	Del	WT	WT	A	T	A	A	T	C	CT	CT
R48	Rabicano	QH	Yes	TA	AG	Del/Del2	WT/Ins	Ins	G	C	G	CA	G	CT	CT	T
R49	Rabicano	QH	Yes	TA	AG	Del/Del2	WT	Ins	G	C	G	CA	GT	C	CT	CT
R50	Rabicano	QH	Yes	TA	AG	Del/WT	WT/Ins	WT/Ins	AG	CT	GA	CA	GT	CT	C	C
R51	Rabicano	QH	Yes	TA	A	Del/WT	WT/Ins	WT/Ins	AG	CT	GA	A	GT	CT	C	CT
R52	Rabicano	QH	Yes	TA	A	Del	Ins	Ins	AG	CT	GA	A	GT	C	T	CT
R53	Rabicano	QH	Yes	T	G	Del	WT/Ins	WT/Ins	AG	CT	GA	CA	GT	CT	CT	CT
R54	Rabicano	QH	Yes	TA	G	Del/Del2	WT/Ins	Ins	AG	CT	GA	CA	T	CT	CT	T

R55	Rabicano	QH	Yes	TA	AG	Del	Ins	Ins	AG	CT	GA	CA	GT	CT	CT	T
R56	Rabicano	QH	Yes	A	A	Del	WT	WT	AG	CT	GA	CA	GT	CT	CT	C
R57	Rabicano	QH	Yes	TA	AG	Del	WT/Ins	WT/Ins	AG	CT	GA	CA	G	CT	CT	C
R58	Rabicano	QH	Yes	A	A	Del	WT/Ins	Ins	AG	CT	GA	CA	G	CT	CT	C
R59	Rabicano	QH	Yes	TA	AG	Del	Ins	Ins	A	T	A	CA	GT	CT	CT	C
R60	Rabicano	WB	No	T	G	Del	WT	Ins	AG	CT	GA	A	GT	C	C	CT
R61	Rabicano	WEL	No	TA	AG	Del2	WT/Ins	WT/Ins	G	C	G	CA	GT	C	C	C

Footnote:

Breed Abbreviation	**Breed Name**
AR	Arabian
CP	Caspian Pony
KWPN	Dutch Warmblood
QH	Quarter Horse
WB	Warmblood
WEL	Welsh Pony
MOR	Morgan
TB	Thoroughbred

Appendix E

Contingency tables using a dominant model of inheritance for variants of interest.

Table E1: Dominant model contingency tables of the ECA28:10,019,065-10,019,144del

		Del & Del/WT	WT	Total
All	Rabicano	**55**	**6**	61
	Control	**24**	**12**	36
	Total	79	18	97
Quarter Horses	Rabicano	**50**	**6**	56
	Control	**19**	**10**	29
	Total	69	16	85
Quarter Horses - Family of Interest	Rabicano	**37**	**2**	39
	Control	**6**	**0**	6
	Total	*43*	*2*	*45*
Quarter Horses - Non-Family of Interest	Rabicano	**13**	**4**	17
	Control	**13**	**9**	22
	Total	26	13	39
Non-Quarter Horses	Rabicano	**5**	**0**	5
	Control	**5**	**2**	7
	Total	10	2	12

Table E2: Dominant model contingency tables of the ECA28:13,938,253-13,941,507del

		Del & Del/WT	WT	Total
All	Rabicano	**56**	**5**	61
	Control	**13**	**23**	36
	Total	69	28	97
Quarter Horses	Rabicano	**53**	**3**	56
	Control	**10**	**19**	29
	Total	63	22	85
Quarter Horses - Family of Interest	Rabicano	**39**	**0**	0
	Control	***4***	***2***	*2*
	Total	43	2	45
Quarter Horses - Non-Family of Interest	Rabicano	**14**	**3**	17
	Control	**6**	**16**	22
	Total	20	19	39
Non-Quarter Horses	Rabicano	**3**	**2**	5
	Control	**3**	**4**	7
	Total	6	6	12

Table E3: Dominant model contingency tables of the ECA28:15,067,299A>G

		GA & G	A	Total
All	Rabicano	**47**	**14**	61
	Control	**9**	**27**	36
	Total	56	41	97
Quarter Horses	Rabicano	**45**	**11**	56
	Control	**8**	**21**	29
	Total	53	32	85
Quarter Horses - Family of Interest	Rabicano	**32**	**7**	39
	Control	**2**	**4**	6
	Total	34	11	45
Quarter Horses - Non-Family of Interest	Rabicano	**13**	**4**	17
	Control	**6**	**16**	22
	Total	19	20	39
Non-Quarter Horses	Rabicano	**2**	**3**	5
	Control	**1**	**6**	7
	Total	3	9	12

Table E4: Dominant model contingency tables of the ECA28:15,431,302T>C (*CEP290*:p.Ile173Val)

		T & TA	A	Total
All	Rabicano	**45**	**16**	61
	Control	**7**	**29**	36
	Total	52	45	97
Quarter Horses	Rabicano	**43**	**13**	56
	Control	**7**	**22**	29
	Total	50	35	85
Quarter Horses - Family of Interest	Rabicano	**31**	**8**	39
	Control	**2**	**4**	6
	Total	33	12	45
Quarter Horses - Non-Family of Interest	Rabicano	**12**	**5**	17
	Control	**5**	**17**	22
	Total	17	22	39
Non-Quarter Horses	Rabicano	**2**	**3**	5
	Control	**0**	**7**	7
	Total	2	10	12

Table E5: Dominant model contingency tables of the ECA28:15,442,561C>T

		T & CT	C	Total
All	Rabicano	**45**	**16**	61
	Control	**9**	**24**	33
	Total	54	40	94
Quarter Horses	Rabicano	**43**	**13**	56
	Control	**9**	**18**	27
	Total	52	31	83
Quarter Horses -	Rabicano	**31**	**8**	39

Table E6: Dominant model contingency tables of the ECA28:15,446,190C>T

		T & CT	C	Total
All	Rabicano	**45**	**16**	61
	Control	**9**	**27**	36
	Total	54	43	97
Quarter Horses	Rabicano	**43**	**13**	56
	Control	**9**	**20**	29
	Total	52	33	85
Quarter Horses -	Rabicano	**31**	**8**	39

Family of Interest	Control	**2**	**4**	7
	Total	33	12	45

Quarter Horses - Non-Family of Interest		T & CT	C	Total
	Rabicano	**12**	**5**	17
	Control	**7**	**13**	20
	Total	19	18	37

Non-Quarter Horses		T & CT	C	Total
	Rabicano	**2**	**3**	5
	Control	**0**	**6**	6
	Total	2	9	11

Family of Interest	Control	**2**	**4**	7
	Total	33	12	45

Quarter Horses - Non-Family of Interest		T & CT	C	Total
	Rabicano	**12**	**5**	17
	Control	**7**	**15**	22
	Total	19	20	39

Non-Quarter Horses		T & CT	C	Total
	Rabicano	**2**	**3**	5
	Control	**0**	**7**	7
	Total	2	10	12

Table E7: Dominant model contingency tables of the ECA28:15,485,999del

All		Del & Del/A	A	Total
	Rabicano	**45**	**16**	61
	Control	**9**	**27**	36
	Total	54	43	97

Quarter Horses		Del & Del/A	A	Total
	Rabicano	**43**	**13**	56
	Control	**9**	**20**	29
	Total	52	33	85

Quarter Horses - Family of Interest		Del & Del/A	A	Total
	Rabicano	**31**	**8**	39
	Control	**2**	**4**	7
	Total	33	12	45

Table E8: Dominant model contingency tables of the ECA28:15,503,208insTTT

All		WT/Ins & Ins/Ins & Ins/WT2	WT & WT2 & WT/WT2	Total
	Rabicano	**45**	**16**	61
	Control	**10**	**26**	36
	Total	55	42	97

Quarter Horses		WT/Ins & Ins/Ins & Ins/WT2	WT & WT2 & WT/WT2	Total
	Rabicano	**43**	**13**	56
	Control	**10**	**19**	29
	Total	53	32	85

Quarter Horses - Family of Interest		WT/Ins & Ins/Ins & Ins/WT2	WT & WT2 & WT/WT2	Total
	Rabicano	**31**	**8**	39
	Control	**2**	**4**	6
	Total	33	12	45

Quarter Horses - Non-Family of Interest		Del & Del/A	A	Total
	Rabicano	**12**	**5**	17
	Control	**7**	**15**	22
	Total	19	20	39

Quarter Horses - Non-Family of Interest		WT/Ins & Ins/Ins & Ins/WT2	WT & WT2 & WT/WT2	Total
	Rabicano	**12**	**5**	17
	Control	**8**	**14**	22
	Total	20	19	39

Non-Quarter Horses		Del & Del/A	A	Total
	Rabicano	**2**	**3**	5
	Control	**0**	**7**	7
	Total	2	10	12

Non-Quarter Horses		WT/Ins & Ins/Ins & Ins/WT2	WT & WT2 & WT/WT2	Total
	Rabicano	**2**	**3**	5
	Control	**0**	**7**	7
	Total	2	10	12

Table E9: Dominant model contingency tables of the ECA28:15,504,948G>T

All		G & GT	T	Total
	Rabicano	**45**	**16**	61
	Control	**9**	**27**	36
	Total	54	43	97

Quarter Horses		G & GT	T	Total
	Rabicano	**43**	**13**	56
	Control	**9**	**20**	29
	Total	52	33	85

Quarter Horses - Family of Interest		G & GT	T	Total
	Rabicano	**31**	**8**	39
	Control	**2**	**4**	6
	Total	33	12	45

Quarter Horses - Non-Family of Interest		G & GT	T	Total
	Rabicano	**12**	**5**	17
	Control	**7**	**15**	22
	Total	19	20	39

Table E10: Dominant model contingency tables of the ECA28:15,518,406-15,518,409del

All		Del & Del/WT	WT	Total
	Rabicano	**45**	**16**	61
	Control	**9**	**27**	36
	Total	54	43	97

Quarter Horses		Del & Del/WT	WT	Total
	Rabicano	**43**	**13**	56
	Control	**9**	**20**	29
	Total	52	33	85

Quarter Horses - Family of Interest		Del & Del/WT	WT	Total
	Rabicano	**31**	**8**	39
	Control	**2**	**4**	6
	Total	33	12	45

Quarter Horses - Non-Family of Interest		Del & Del/WT	WT	Total
	Rabicano	**12**	**5**	17
	Control	**7**	**15**	22
	Total	19	20	39

		G & GT	T	Total
Non-Quarter Horses	Rabicano	**2**	**3**	5
	Control	**0**	**7**	7
	Total	2	10	12

		Del & Del/WT	WT	Total
Non-Quarter Horses	Rabicano	**2**	**3**	5
	Control	**0**	**7**	7
	Total	2	10	12

Table E11: Dominant model contingency tables of the ECA28:15,967,332G>A

		A & AG	G	Total
All	Rabicano	**49**	**12**	61
	Control	**7**	**29**	36
	Total	56	41	97

		A & AG	G	Total
Quarter Horses	Rabicano	**45**	**11**	56
	Control	**7**	**22**	29
	Total	52	33	85

		A & AG	G	Total
Quarter Horses - Family of Interest	Rabicano	**31**	**8**	39
	Control	**2**	**4**	6
	Total	33	12	45

		A & AG	G	Total
Quarter Horses - Non-Family of Interest	Rabicano	**14**	**3**	17
	Control	**5**	**17**	22
	Total	19	20	39

		A & AG	G	Total
Non-Quarter Horses	Rabicano	**4**	**1**	5
	Control	**0**	**7**	7
	Total	4	8	12

Table E12: Dominant model contingency tables of the ECA28:16,465,850A>T.

		T & TA	A	Total
All	Rabicano	**46**	**15**	61
	Control	**7**	**29**	36
	Total	53	44	97

		T & TA	A	Total
Quarter Horses	Rabicano	**44**	**12**	56
	Control	**7**	**22**	29
	Total	51	34	85

		T & TA	A	Total
Quarter Horses - Family of Interest	Rabicano	**32**	**7**	39
	Control	**2**	**4**	6
	Total	34	11	45

		T & TA	A	Total
Quarter Horses - Non-Family of Interest	Rabicano	**12**	**5**	17
	Control	**5**	**17**	22
	Total	17	22	39

		T & TA	A	Total
Non-Quarter Horses	Rabicano	**2**	**3**	5
	Control	**0**	**7**	7
	Total	2	10	12

Table E13: Dominant model contingency tables of the ECA28:16,616,952T>A

		A & AT	T	Total
All	Rabicano	**46**	**15**	61
	Control	**7**	**29**	36
	Total	53	44	97
Quarter Horses	Rabicano	**44**	**12**	56
	Control	**7**	**22**	29
	Total	51	34	85
Quarter Horses - Family of Interest	Rabicano	**32**	**7**	39
	Control	**2**	**4**	6
	Total	34	11	45
Quarter Horses - Non-Family of Interest	Rabicano	**12**	**5**	17
	Control	**5**	**17**	22
	Total	17	22	39
Non-Quarter Horses	Rabicano	**2**	**3**	5
	Control	**0**	**7**	7
	Total	2	10	12

Table E14: Dominant model contingency tables of the ECA28:16,764,837G>A

		A & AG	G	Total
All	Rabicano	**45**	**16**	61
	Control	**10**	**26**	36
	Total	55	42	97
Quarter Horses	Rabicano	**43**	**13**	56
	Control	**10**	**19**	29
	Total	53	32	85
Quarter Horses - Family of Interest	Rabicano	**31**	**8**	39
	Control	**2**	**4**	6
	Total	33	12	45
Quarter Horses - Non-Family of Interest	Rabicano	**12**	**5**	17
	Control	**8**	**14**	22
	Total	20	19	39
Non-Quarter Horses	Rabicano	**2**	**3**	5
	Control	**0**	**7**	7
	Total	2	10	12

Table E15 Dominant model contingency tables of the ECA28:31,573,538-31,573,77del.

		Del & Del/WT& Del/Del2	WT & Del2	Total
All	Rabicano	**59**	**2**	61
	Control	**35**	**1**	36
	Total	94	3	97
Quarter Horses	Rabicano	**55**	**1**	56

Table E16: Dominant model contingency tables of the ECA28:38,000,030ins(238)

		Ind & Ins/WT	WT	Total
All	Rabicano	**44**	**17**	61
	Control	**9**	**27**	36
	Total	53	44	97
Quarter Horses	Rabicano	**42**	**14**	56

	Control	**28**	**1**	29
	Total	83	2	85

	Control	**9**	**20**	29
	Total	51	34	85

Quarter Horses - Family of Interest		Del & Del/WT& Del/Del2	WT & Del2	Total
	Rabicano	**39**	**0**	39
	Control	**6**	**0**	6
	Total	45	0	45

Quarter Horses - Family of Interest		Ind & Ins/WT	WT	Total
	Rabicano	**36**	**3**	39
	Control	**5**	**1**	6
	Total	41	4	45

Quarter Horses - Non-Family of Interest		Del & Del/WT& Del/Del2	WT & Del2	Total
	Rabicano	**16**	**1**	17
	Control	**21**	**1**	22
	Total	37	2	39

Quarter Horses - Non-Family of Interest		Ind & Ins/WT	WT	Total
	Rabicano	**6**	**11**	17
	Control	**4**	**18**	22
	Total	10	29	39

Non-Quarter Horses		Del & Del/WT& Del/Del2	WT & Del2	Total
	Rabicano	**4**	**1**	5
	Control	**7**	**0**	7
	Total	11	1	12

Non-Quarter Horses		Ind & Ins/WT	WT	Total
	Rabicano	**2**	**3**	5
	Control	**0**	**7**	7
	Total	2	10	12

Table E17: Dominant model contingency tables of the ECA28:38,605,308(234)

		Ins & Ins/WT	WT	Total
All	Rabicano	**48**	**12**	60
	Control	**20**	**16**	36
	Total	68	28	96
Quarter Horses	Rabicano	**45**	**10**	55
	Control	**17**	**12**	29
	Total	62	22	84
Quarter Horses - Family of Interest	Rabicano	**37**	**2**	39
	Control	**5**	**1**	6
	Total	42	3	45
Quarter Horses - Non-Family of Interest	Rabicano	**8**	**8**	16
	Control	**11**	**11**	22
	Total	19	19	38
Non-Quarter Horses	Rabicano	**3**	**2**	5
	Control	**3**	**4**	7
	Total	6	6	12

Table E18: Dominant model contingency tables of the ECA28:39,293,149G>A

		A & AG	G	Total
All	Rabicano	**44**	**17**	61
	Control	**19**	**17**	36
	Total	63	34	97
Quarter Horses	Rabicano	**42**	**14**	56
	Control	**18**	**11**	29
	Total	60	25	85
Quarter Horses - Family of Interest	Rabicano	**34**	**5**	39
	Control	**6**	**0**	6
	Total	40	5	45
Quarter Horses - Non-Family of Interest	Rabicano	**8**	**9**	17
	Control	**12**	**10**	22
	Total	20	19	39
Non-Quarter Horses	Rabicano	**2**	**3**	5
	Control	**1**	**6**	7
	Total	3	9	12

Table E19: Dominant model contingency tables of the ECA28:39,296,303C>T

		T & TC	C	Total
All	Rabicano	**44**	**17**	61
	Control	**19**	**17**	36
	Total	63	34	97
Quarter Horses	Rabicano	**42**	**14**	56
	Control	**18**	**11**	29
	Total	60	25	85
Quarter Horses - Family of Interest	Rabicano	**34**	**5**	39
	Control	**6**	**0**	6
	Total	40	5	45
Quarter Horses - Non-Family of Interest	Rabicano	**8**	**9**	17
	Control	**12**	**10**	22
	Total	20	19	39
Non-Quarter Horses	Rabicano	**2**	**3**	5
	Control	**1**	**6**	7
	Total	3	9	12

Table E20: Dominant model contingency tables of the ECA28:39,300,931G>A

		A & AG	G	Total
All	Rabicano	**44**	**17**	61
	Control	**19**	**17**	36
	Total	63	34	97
Quarter Horses	Rabicano	**42**	**14**	56
	Control	**18**	**11**	29
	Total	60	25	85
Quarter Horses - Family of Interest	Rabicano	**34**	**5**	39
	Control	**6**	**0**	6
	Total	40	5	45
Quarter Horses - Non-Family of Interest	Rabicano	**8**	**9**	17
	Control	**12**	**10**	22
	Total	20	19	39
Non-Quarter Horses	Rabicano	**2**	**3**	5
	Control	**1**	**6**	7
	Total	3	9	12

Table E21: Dominant model contingency tables of the chr1:114,505,701C>A (*OCA2*:p.Cys93Phe)

		A & CA	C	Total
All	Rabicano	**51**	**10**	61
	Control	**24**	**12**	36
	Total	75	22	97
Quarter Horses	Rabicano	**47**	**9**	56
	Control	**20**	**9**	29
	Total	67	18	85
Quarter Horses -	Rabicano	**39**	**0**	39

Table E22: Dominant model contingency tables of the chr1:155,328,511G>T (*FMN1*:p.Pro189Gln)

		T & GT	G	Total
All	Rabicano	**42**	**19**	61
	Control	**20**	**16**	36
	Total	62	35	97
Quarter Horses	Rabicano	**38**	**18**	56
	Control	**18**	**11**	29
	Total	56	29	85
Quarter Horses -	Rabicano	**29**	**10**	39

Family of Interest	Control	**5**	**1**	6
	Total	44	1	45

Family of Interest	Control	**4**	**2**	6
	Total	33	12	45

		A & CA	C	Total
Quarter Horses - Non-Family of Interest	Rabicano	**8**	**9**	17
	Control	**15**	**7**	22
	Total	23	16	39

		T & GT	G	Total
Quarter Horses - Non-Family of Interest	Rabicano	**9**	**8**	17
	Control	**13**	**9**	22
	Total	22	17	39

		A & CA	C	Total
Non-Quarter Horses	Rabicano	**4**	**1**	5
	Control	**4**	**3**	7
	Total	8	4	12

		T & GT	G	Total
Non-Quarter Horses	Rabicano	**4**	**1**	5
	Control	**2**	**5**	7
	Total	6	6	12

Table E23: Dominant model contingency tables of the chr10:4,049,610C>T (*ANKRD27*:p.Ala627Thr)

		T & TC	C	Total
All	Rabicano	**26**	**35**	61
	Control	**5**	**31**	36
	Total	31	66	97

		T & TC	C	Total
Quarter Horses	Rabicano	**25**	**31**	56
	Control	**5**	**24**	29
	Total	30	55	85

		T & TC	C	Total
Quarter Horses - Family of Interest	Rabicano	**24**	**15**	39
	Control	**3**	**3**	6
	Total	27	18	45

		T & TC	C	Total
Quarter Horses - Non-Family of Interest	Rabicano	**1**	**16**	17
	Control	**1**	**21**	22
	Total	2	37	39

Table E24: Dominant model contingency tables of the chr20:17,241,786C>T *KIF13A*:p.Arg848His)

		T & CT	C	Total
All	Rabicano	**41**	**20**	61
	Control	**11**	**25**	36
	Total	52	45	97

		T & CT	C	Total
Quarter Horses	Rabicano	**41**	**15**	56
	Control	**10**	**19**	29
	Total	51	34	85

		T & CT	C	Total
Quarter Horses - Family of Interest	Rabicano	**36**	**3**	39
	Control	**6**	**0**	6
	Total	42	3	45

		T & CT	C	Total
Quarter Horses - Non-Family of Interest	Rabicano	**5**	**12**	17
	Control	**3**	**19**	22
	Total	8	31	39

Non-Quarter Horses		T & TC	C	Total
	Rabicano	**1**	**4**	5
	Control	**0**	7	7
	Total	1	11	12

Non-Quarter Horses		T & CT	C	Total
	Rabicano	**0**	**5**	5
	Control	**1**	**6**	7
	Total	1	11	12

Table E25: Dominant model contingency tables of the chr30:26,482,977C>T (*CRB1*:p.Pro755Ser)

All		C & CT	T	Total
	Rabicano	**51**	**10**	61
	Control	**32**	**4**	36
	Total	83	14	97

ChrECA30:26,482,977 SNP				
Quarter Horses		C & CT	T	Total
	Rabicano	**46**	**10**	56
	Control	**25**	**4**	29
	Total	71	14	85

ChrECA30:26,482,977 SNP				
Quarter Horses - Family of Interest		C & CT	T	Total
	Rabicano	**29**	**10**	39
	Control	**5**	**1**	6
	Total	34	11	45

ChrECA30:26,482,977 SNP				
Quarter Horses - Non-Family of Interest		C & CT	T	Total
	Rabicano	**17**	**0**	17
	Control	**19**	**3**	22
	Total	36	3	39

ChrECA30:26,482,977 SNP				
Non-Quarter Horses		C & CT	T	Total
	Rabicano	**5**	**0**	5
	Control	**7**	**0**	7
	Total	12	0	12