2014

USE OF GENOMIC TOOLS TO DISCOVER THE CAUSE OF CHAMPAGNE DILUTION COAT COLOR IN HORSES AND TO MAP THE GENETIC CAUSE OF EXTREME LORDOSIS IN AMERICAN SADDLEBRED HORSES

Deborah G. Cook
University of Kentucky, in2paints@gmail.com

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Deborah G. Cook, Student

Dr. Ernest Bailey, Major Professor

Dr. Dan Howe, Director of Graduate Studies

USE OF GENOMIC TOOLS TO DISCOVER THE CAUSE OF CHAMPAGNE DILUTION COAT COLOR IN HORSES AND TO MAP THE GENETIC CAUSE OF EXTREME LORDOSIS IN AMERICAN SADDLEBRED HORSES

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Agriculture, Food and Environment
At the University of Kentucky

By
Deborah G. Cook
Lexington, Kentucky

Director: Dr. Ernest Bailey, Department of Veterinary Science
Lexington, Kentucky 2014

ABSTRACT OF DISSERTATION

USE OF GENOMIC TOOLS TO DISCOVER THE CAUSE OF
CHAMPAGNE DILUTION COAT COLOR IN HORSES AND TO MAP THE
GENETIC CAUSE OF EXTREME LORDOSIS
IN AMERICAN SADDLEBRED HORSES

Champagne dilution of coat color in horses is caused by dominant gene action. Three sire families were identified as segregating for this trait. Genome wide linkage analysis using 104 microsatellite DNA markers was used to map the gene to ECA14 (LOD > 11.0). Four genes, namely *SPARC*, *SLC36A1*, *SLC36A2* and *SLC36A3*, were selected from the region implicated by linkage and their exons sequenced to identify genetic variation. DNA sequences were compared for two homozygotes for Champagne dilution, two heterozygotes and two horses without dilution. A single base change in exon 2 of *SLC36A1* was found unique to horses exhibiting Champagne dilution. *SLC36A1* encodes a eukaryote specific proton dependent small amino acid transporter also known as LYAAT-1 in rats. This change in base 188 of the cDNA from a *C* to a *G* (Genbank REFSEQ: Non-champagne EU432176 and Champagne EU432177), is predicted to cause an amino acid change from threonine to arginine in the first transmembrane region of the protein. This may disrupt placement of the transmembrane portion of the protein in the membrane interrupting function of the ion channel. No exceptions to the association of this mutation with the *Champagne dilution* (*CH)* allele were identified based on testing 182 additional horses representing 15 breeds, suggesting that this mutation may be the causative for the dilution phenotype. While this gene is expected to function as a proton/amino acid symporter based on its sequence and gene family, this observation for the horse is the first evidence for a phenotypic effect of mutation of the gene. Identifying this variant also gives breeders a new tool for selecting breeding stock with or without this dilution genotype.

Extreme lordosis is a condition in which the dorsal to ventral curvature of the back is accentuated. The condition is considered a conformation fault by many Saddlebred breeders and is commonly referred to as swayback, softback or lowback. Previous studies suggested the existence of a hereditary component in American Saddlebred horses. A whole genome association study was performed utilizing the Illumina SNP50 beadchip containing over 59,000 SNPs. A 3 Mb region on ECA20 was found associated with extreme lordosis. The distribution of markers indicated that the trait had a simple Mendelian recessive mode of inheritance. Subsequently, the association was confirmed and the region was narrowed to just over 500 kb by a higher density SNP assay using 47 SNPs selected from the equine SNP database. Identification of this associated region will allow future studies a better opportunity to locate the mutation responsible for the lordotic phenotype in young American Saddlebred Horses.

KEYWORDS: horse, champagne, lordosis, swayback, mapping

Deborah G. Cook
Student's Signature

June 17, 2014
Date

USE OF GENOMIC TOOLS TO DISCOVER THE CAUSE OF CHAMPAGNE DILUTION COAT COLOR IN HORSES AND TO MAP THE GENETIC CAUSE OF EXTREME LORDOSIS IN AMERICAN SADDLEBRED HORSES

By

Deborah G. Cook

Dr. Ernest Bailey
Director of Thesis

Dr. Daniel K. Howe
Director of Graduate Studies

June 17, 2014
Date

In loving memory of my father

Albert “Shorty” Cook Jr.
December 3, 1948 – July 9, 1983

I was his shadow.
His love made me feel so safe.
His death made me strong.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my friends Samantha Brooks and Stephanie Reedy. These two women kept pulling me back from the edge of quitting and reminded me time after time that I can do this. Both have offered me safe haven for focused writing and positive support that made me keep believing in myself.

I would also like to thank my lab mate John Eberth and my friends in horse genetics Stephen Colemen, Mike Meineltowski, and Rebecca Bellone. The good times we shared both in the lab and at conferences helped keep life in research fun.

I thank my committee for their patience and understanding during my time here and for their ideas and insights about my work.

I want to thank all the horse owners and breeders who happily provided samples for my research.

I also want to thank all the employees and support staff at Gluck Equine Research Center and at University of Kentucky. The many helpful people at Gluck have made me feel more welcome than family.

A special thanks to my graduate adviser Ernie Bailey. He never gave up on me and continued to tell me I could do this every time I doubted myself.

A special thank you to Dan Howe as well, for he and Ernie have helped to move mountains for me to finish, even after I thought all was lost.

I would like to thank my Mom and all my friends and family who have been there for me during this journey. I cannot possible mention everyone who has been there for me over the years, but know that I appreciate everyone that has been there for me while I worked for this degree

Lastly I would like to thank my loving husband, Daniel Ward and my son, Vincent Cook. They are the lights of my life and I love them more than words can express.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: Gene Mapping and Current Applications in Horses

Introduction

Genetic tools have been valuable for medical and agricultural research. Applications of genetics have led to understanding of basic phenotypic variations, diseases, pathogens and the creation of therapeutic treatments. With the advent of improved molecular genetic techniques and the completion of whole genome sequences for many livestock species, including the horse, scientists have uncovered genes related to diseases, color patterns and performance traits. This dissertation describes two approaches to gene discovery used to investigate a coat color gene (Champagne dilution) and a morphological trait (extreme lordosis). The work is a product of modern technology but has deep roots in the history of genetics.

Section I: Nature of Natural and Artificial Selection

Natural Selection

In the middle of the 19th century, Charles Darwin, an English naturalist, noted that organisms within a species possess individual differences which are less pronounced in populations living in confined regions, but exhibit more pronounced variations between populations of different regions, especially with differing environmental pressures (Darwin, 1859). Inherited variation and evolution had long been recognized before Charles Darwin published his book On the Origin of Species, but he recognized natural selection as the subtle, but powerful, force that joined these two observations. Darwin did not know how genetic variation originated but he recognized that when it occurred, resulting differences enabled some individuals in a population to be more efficient in obtaining food and reproducing. Those individuals would leave more offspring leading to an increase in that genetic variant in future generations. He proposed that changes in the environment and changes in genetic variation could lead to evolution of new species. Variables in nature were the driving force of selection. One of his examples is the difference among beak functions and sizes possessed by finches populating different isolated islands of the Galapagos archipelago. He deduced differences such as these were under selective pressure from the environment and must have conferred reproductive advantage in order to get passed on from one generation to the next generation (natural selection).

Artificial Selection

One of the major arguments that Darwin advanced about the power of natural selection was derived from observations of artificial selection (Darwin, 1859). Artificial selection exemplified by plant cultivation and animal domestication in which people select for plants and animals that have desirable traits. Over time, artificial selection has led to production of wheat with higher yields, faster growth in livestock and greater power and speed among horses. Farmers select the most phenotypically desirable individuals for reproduction of the next generation, leading to varieties distinct from those found in nature, but are superbly adapted to agricultural use. This continued desire to reproduce desirable traits and avoid undesirable ones for efficiency and profitability is a driving force behind continued advances in molecular technologies and research.

Selective Pressure

Selective pressure can be the result of natural selection in the wild, such as the female mocking bird's selection of a mate depending upon his song or ecological selection caused by a trait making an individual more fit for reproduction in a particular environment. This pressure can also be the result of artificial selection as described above. Disease causing genes can be under selective pressure within certain populations. Genetic variants can be subjected to negative and positive selection at the same time. If the selection pressures are equivalent, then the genes are said to be under balanced selection and the variant will be maintained in the population.

The genetics of thalassemia, a microcytic anemia, is a good example of how selective pressure can result in fixation of an otherwise abnormal gene in a population. The genetic variant α-thalassemia has been found in a uniformly high frequency in Tharu people of Southern Nepal. Most Tharu people are homozygous for the α-/α-genotype, with an overall *α-thal* gene (α-) frequency of .8 in the population. It is suggested that holoendemic malaria resulted in preferential survival of people with α-*thal* genetic variant, enabling this population to survive for centuries despite the high occurrence of malaria (Modiano *et al.,* 1991). Higher frequencies of the genes for sickle cell anemia, Glucose-6-phosphate dehydrogenase deficiency and Hemoglobin C are also associated with selective pressure from malaria in areas of Africa and around the Mediterranean region (Allison 1954; Ruwende and Hill, 1998 and Agarwal *et al.,* 2000).

Section II: Discrete Nature of Genes

Gregor Mendel, an Augustinian friar and Austrian scientist, established that the inheritance of particular traits in peas followed set patterns. He studied traits that were easy to characterize and only occurred in one of two forms; such as; yellow or green seeds, tall or short stems, white or purple flowers. He proposed that traits are passed on from parent to progeny as descrete units (later called genes). He characterized these genes as being dominant or recessive; dominant genes were always expressed in at least one parent while recessive genes would not be expressed unless two copies were present; one inherited from each parent. His work led to two major conclusions. First, one of the two genes in a parent is passed to its offspring (Law of Segregation). Second, genes for separate traits are passed on to offspring independent of one another (Law of Independent Assortment) (Mendel, 1866). These basic principles of genes and heredity were reestablished independently by three scientists around 1900 (deVries, 1890; Correns, 1900; von Tschermak-Seysenegg, 1900-1901) who are collectively credited with rediscovery of Mendel's Laws of Independent Assortment and Segregation.

As an interesting side note, during this same period, the presence of chromosomes in the cell nucleus was noted by cytologists including Walther Flemming, Wilhelm von Waldeyer-Hartz and Edouard Van Beneden (Flemming, 1878; Waldeyer-Hartz, 1888). After the turn of the century, the relationship between chromosomes and heredity was first suggested by Theodor Boveri and expanded upon by Walter Sutton (Sutton, 1902). Morgan (1910) proved that chromosomes are the source of hereditary material.

Mendel's basic laws of heredity are the foundation of current studies to identify the genes responsible for simple and complex traits. During the 1900s many Mendelian genes were uncovered, one of the most famous being the ABO blood group system. In 1900, Karl Landsteiner identified four blood types A, B, AB and O in humans; however the specific genetic basis underlying the ABO system was not described until 24 years later (Bernstein 1924)

Modes of Inheritance

The mode of inheritance is the manner in which a gene and its corresponding phenotype are passed down from one generation to another. Modal inheritance patterns can be categorized as follows; autosomal dominant, autosomal recessive, X-linked dominant, sex-linked, complex and mitochondrial. (Autosomes are the chromosomes that occur in duplicate pairs in individuals, one coming from each parent; Sex chromosomes are the X and the Y chromosome that determine the gender of the individual.) Autosomal dominant genes are always phenotypically expressed when present, whether occurring as a singleton or in duplicate. If the gene has a different effect when present in two copies, then it is referred to as an incomplete dominant gene. Cream dilution in horses has an incomplete dominant mode of inheritance as shown by the dilution of brown pigment when present in one copy and dilution of all pigment when present in two copies (Adalsteinsson, 1974). Autosomal recessive conditions are only manifested in individuals bearing two copies of the variant allele, or, to put it another way, in the absence of a dominant gene. Chestnut coat color is due to a recessive gene resulting in a loss of function variation that prevents production of black pigment (Marklund *et al.,* 1996).

Sex linked inheritance is associated with genes on the X or Y chromosome in mammals. Females have two copies of the X chromosome while males have only one X and one Y chromosome. Males possess a single copy of genes on the X chromosome and will express whatever gene is present. As a consequence, genes which are recessive and rarely detected in females are readily expressed in males. Males are said to be hemi-zygous for the X chromosome since they have but one copy. X-linked recessive traits are manifested in females where both X chromosomes bear the variation and in males carrying just the single X chromosome with the causative variation. Hemophilia is a hemorrhagic disease inherited as a sex linked recessive trait. Queen Victoria of England was a carrier of hemophilia. It was long speculated which form of hemophilia was carried by the royal family and no living carriers exist. Discovery of tombs containing Czar Nicolas II's entire family provided the necessary genetic material, namely from Czar Alexis who had the disease. This led to the discovery that the royal hemophilia was hemophilia B caused by a mutation of the *F9* gene (Lannoy and Hermans, 2010)

Some phenotypically similar traits have heterogeneous origins, with different genetic variations within the same gene responsible for a common phenotype, such as is the case with Dominant White in horses. In some horse breeds all white or partial white horses arise from the matings of solid colored parents. In many of these cases the white color is inherited as a monogenic autosomal dominant trait that spontaneously arose in one of the parent's germ line cells (Haase *et al.,* 2007, Haase *et al.,* 2009, Haase *et al.,* 2010). Duchenne's muscular dystrophy (Emery, 1984) is the result of many different individual variations, some of which occurred spontaneously in germ line cells. Genetic variations in different genes can also cause a single phenotype. Osteogeneisis imperfecta (OI) is a connective tissue disorder with different genes (namely collagen gens and genes coding for proteins that form complexes with them) responsible for different types of OI.

Complex inheritance is the result of multiple genes, their interactions with other genes and/or interactions with the environment, all playing a role in the risk of disease or trait development. Athletic performance is influenced by training in addition to hereditary characters influencing strength, endurance, speed and intellect. For example: a 287 bp in/del identified in the angiotensin converting enzyme, ACE gene, was found to be associated with serum levels of the product enzyme (Rigat *et al.,* 1990). Frequency of the insertion allele was also found to be associated with endurance 1996 study of Australian rowers (Gayagay *et al.,* 1998). In a study of British mountaineers, the ACE insertion allele was associated with the ability to climb greater than 8000 feet without oxygen with no deletion homozygotes present in this elite group. The deletion genotype has been found to be associated with sprinters and short distance swimmers (Montgomery *et al.,* 1998). Observed associations such as this only account for part of the heritability of these traits. There are other yet unknown genes involved in these abilities or phenotypes.

Mitochondria are organelles found in the cytoplasm of cells; mitochondrial DNA is a closed circular molecule 16,569 bp long, encodes 13 genes and is only inherited from the mother (Anderson *et al.,* 1981, Giles *et al.,* 1980). Many of the inherited mitochondrial disorders are due to mutations in the approximately 1500 nuclear genes that are targeted to the mitochondria and only half of which have been identified (Calvo *et al.,* 2006).

Linear Organization of Genes and Recombination

The first linkage between genes was discovered, but not understood, in studies of dihybrid crosses in pea plants similar to Mendel's work, by William Bateson, Edith Rebecca Saunders and Reginald Punnett (Bateson *et al.,* 1905). This linkage between traits was later elucidated by Thomas Hunt Morgan in his studies of *Drosophala melanogaster* (fruit flies) at Columbia University. Morgan ascertained the presence of the discrete hereditary sex determining unit known as the X chromosome in breeding experiments with a white eyed fly (Morgan, 1910). He observed co-segregation of white eyes and gender corresponded with the presence of chromosomes. Morgan suspected crossing over/recombination of chromosomes when he noticed that the "linked" traits of small wing and white eye would sometimes separate (Morgan, 1911). Morgan's student, Alfred Sturtevant, was able to calculate the linear genetic relationship between the trait loci and piece together the first map of genetic linkage by calculating recombination frequency/crossing over events between linked loci (Sturtevant, 1913). This research demonstrated that genes were discrete hereditary units which were joined in a linear arrangement in relation to one other like beads on a string (Sturtevant and Beadle, 1939). Some chromosomes are very large and it is not possible to detect linkage between genes residing on opposite ends of a chromosome. Therefore, the presence of testable markers that give good coverage of a chromosome is important.

Section III: Nature of Genetic Variation

Every cell of mammals, except germ line cells, contains the full complement of DNA, called the genome. Its code is comprised of different combinations of only four bases; adenine, thymine, guanine, and cytosine (A, T, G, C). Phenotypic differences in organisms are the result of physical and structural changes to this code. There are also many physical and structural differences of the DNA that will not be associated with phenotypic differences. These variations present in the genome are discussed below.

SNPs

Single nucleotide polymorphisms, SNPs, are a single base change in the sequence of DNA. In 2001 a human SNP map was compiled using 1.42 million SNPs. This included all the SNPs that were publically available from multiple sources in November of 2000. This indicated about one SNP per 1.9 kbp. An estimated 60,000 of these SNPs were believed to fall within coding regions (Sachidanandam *et al.*, 2001). By 2008, a sequencing project of an individual person identified over 3 million SNPs. This indicates the presence of approximately 1 SNP per 1000 bases. More than 10,000 of the SNPs cause amino-acid substitutions in known coding sequences (Wheeler *et al.*, 2008). The number of SNPs present in the genome can vary among other species, as the grapevine (Vitis vinifera L.) has SNPs occurring one in every 47 bp in non-coding regions and one in 69 bp in coding regions (Lijavetzky *et al.*, 2007)

Insertions and Deletions: In/Dels

Insertions and deletions (IN/DELS) can involve a single base or several thousand bases and have the potential to cause disease or phenotypic change if they occur in a location that alters a gene's transcript or expression. Messenger RNA transcripts are read by the translation machinery in sets of three base pairs at a time, called codons. The insertion or deletion of a single base, or any group of bases not in a multiple of three, results in a "frameshift". A frameshift results in a different codon reading frame from the point forward from a mutation resulting in an altered product or introducing a premature stop codon.

Duchenne's (DMD) and Becker's (BMD) Muscular Dystrophy are examples of genetically heterogenous X-linked recessive lethal diseases caused by various deletions, duplications and point mutations in the dystrophin gene. The dystrophin gene is a large gene coding for a long protein product that is part of the motor network in muscle cells. Disease severity depends on whether the in/del creates a frame shift or remains in frame. The frame-shift variants cause a more severe phenotype, because the altered reading frame introduces premature termination codons, whereas in frame variants are still functional (Prior and Bridgeman 2005). The dystrophin gene is located on the X chromosome in a highly morphogenic region, also known as a mutational "hot spot". Deletions or duplications account for about two-thirds of the cases of DMD and BMD with the remainder of the cases accounted for by smaller deletions and point mutations as reviewed by Prior and Bridgeman (2005).

Tandem Repeats

Tandem repeats are categorized into two groups; micro- and mini- satellites. Microsatellites can range from 1 to 4 base pairs; single nucleotide, dinucleotide trinucleotide and tetranucleotide. Repeats of units 6 to 64 bases long are minisatellites and can span from 100 bp to 10 kb (Jeffries *et al.,* 1985). Satellites are units from 5 to 171 bp and repeats can span from 100 kb to several Mb. A mega satellite is a repeat composed of units several kilobases long and can span several hundred kilobases. The number of repeats in succession can be polymorphic in the population due to polymerase slippage at DNA replication or mismatched recombination during meiosis.

Microsatellites in the form of simple tandem repeats with the dinucleotide sequence [i.e., $(dT\text{-}dG)_n$], are interspersed throughout the human genome with approximately 50,000 copies present. They can be highly polymorphic and their variability makes them suitable for gene mapping and parentage testing (Litt and Luty 1989, Weber and May 1989). They are usually absent from exons of genes because a dinucleotide expansion would disrupt the reading frame of the exon in which they resided, resulting in a frame shift. Polymorphism of a dinucleotide repeat in the *STAT 6* gene is associated with allergic diseases (Tamura *et al.,* 2001). Trinucleotide repeats are the most common form responsible for repeat expansion diseases and disorders.

Huntington's disease, a neuronal degenerative disease in humans, is a genetic disorder inherited in an autosomal dominant fashion and is caused by an exonic trinucleotide $(CAG)_n$ expansion. In the normal population this repeat is highly polymorphic ranging in size from 11 to 34 repeats. Huntington's cases possess 48 or more repeats of this trinucleotide. Expansions of the repeat occur from one generation to the next and exhibits "anticipation", which is an earlier age of onset with each affected generation due to increased repeat length. Lack of recombination of the affected chromosome and transmission from the father is typical (HDCRG 1993). Myotonic dystrophy 1, DM1, is caused by a trinucleotide CTG expansion in 3' untranslated region of the *dystrophia myotonica-protein kinase* gene, *DMPK*; whereas, an intronic tetranucleotide CCTG expansion in intron 1 of *zinc finger protein 9, ZNF9* is responsible for dystrophia myotonica 2, DM2 (Liquori *et al.*, 2001).

Minisatellites can be highly polymorphic. A human DNA "fingerprint" was purified and cloned in 1986. This 6.3 kilobase long minisatellite contained multiple copies of a 37 bp repeat unit. It was found to have 77 different sized alleles containing 14 to 524 repeats among 79 individuals (Wong *et al.*, 1986). The human *Aggrecan* (*ACAN*) gene contains several repeat domains which are recognition sites for the binding of other molecules. Its keratin sulfate attachment domain has 11 repeats of a hexameric amino acid sequence. The chondroitin sulfate binding region, CS1, has a highly conserved repeat region that translates to the same 19 amino acids repeated 19 times (Doege *et al.*, 1997). Larger repeats, which are so big they cytogenetically alter the chromosomes, have been identified. A repeat spanning over 1 Mb of chromosome 15 has been identified. The repeat region contains unprocessed partial duplicates of the GABRA5 gene and unprocessed duplicates the immunoglobulin heavy chain (IgH) D segment gene and the neurofibromatosis type 1 (NF1) gene (Ritchie *et al.*, 1998).

Transposable Elements

The first transposable elements were discovered in maize (McClintock, 1950). Transposable elements can be separated into two categories, Class I retrotransposons and Class II transposons. Class II elements are DNA transposons which are movable elements that are excised from the genome then inserted at a new location. This family of transposable elements includes; *Drosophila* P elements, Helitrons and Maverick (Richard *et al.,*. 2008). The Class I elements relocate in a copy and paste method utilizing RNA as the copy which is then reverse transcribed and pasted back into the genome in another location. These elements include long terminal repeats (LTRS), long interspersed nuclear elements (LINES), short interspersed nuclear elements (SINES), Dictyostelium intermediate repeat sequence (DIRS) elements and Penelope-like elements (PLE). In the horse, ERE1 and ERE2 are two members of the SINE family of transposable elements (Gallagher *et al.,* 1999). There are five members of the LINE family of transposable elements represented in the horse (Adelson *et al.,* 2010).

Alu elements of the SINE category are roughly 280 nucleotides long with no introns and are the most common transposable elements across the human genome with nearly 1 million copies present (Schmid, 1998). *Alu* elements are commonly found associated with some microsatellites and may play a role in microsatellite expansion or transposition within the genome (Arcot *et al.,* 1995). The insertion of *Alu* elements into or near genes can disrupt gene function. Genetic mutations in genes due to insertion of *Alu* elements account for approximately 0.1% of human diseases or disorders including, but not limited to; neurofibromatosis, hemophilia, breast cancer, Apert syndrome, cholinesterase deficiency and complement deficiency (Deininger & Batzer, 1999). An *Alu* insertion into exon 5 of the *Factor IX* gene was found to be the cause of hemophilia B in one patient. This insertion interrupted the reading frame of the gene and introduced a stop codon within the inserted sequence (Vidaud *et al.,* 1992).

Copy Number Variations (CNV)

When insertions or deletions involve an entire gene, and possibly its promoter region, copy number variations arise. An increased or decreased number of the same gene creates a gene dosage effect. For example, a duplication of the *CC chemokine 3-like 1* (*CCL3L1*) gene has been found to decrease host susceptibility to infection by HIV. The product of this gene is a ligand for the main co-receptor for HIV, the *CC chemokine receptor 5* (*CCR5*). Individuals possessing more copies than the population average of two to four copies (depending upon the geographic region/population) for the *CCL3L1* gene have decreased susceptibility to HIV infection. Individuals with the lowest copy numbers for these populations had between 69 and 97% higher chance of getting HIV (Gonzalez *et al.*, 2005). Copy number variations have been found associated with autism (Sanders *et al.*, 2011). A duplication found at a site known for microdeletions, which cause Williams–Beuren's Syndrome, has been found to have the opposite effect of the microdeletions, which are known for expressive language, but poor drawing capabilities. The patient with the duplication had severe language difficulties, especially with expressive language, but had increased drawing capabilities for expressing thoughts (Somerville *et al.*, 2005).

Pseudogenes

Pseudogenes are related to functional genes, but are defective. Pseudogenes fall into two categories; genes that retain the regulatory and intervening sequences and the more abundant variety that lacks introns and regulatory sequences. The latter category, known as processed pseudogenes, are the result of reintegration mature mRNA into the genome; a process also known as retrotransposition. Processed pseudogenes lack introns and other regulatory sequences, have poly A tails and are flanked by direct repeats of 7-17 base pairs (Vanin, 1985). Pseudogenes can complicate gene discovery with their similarities to functional genes. Beta-tubulin is a member of a multi gene family. Four sequences are represented in the human genome. One sequence is the expressed gene which yields 1.8 kb and 2.6 kb mRNAs resulting from two different polyadenylation sites. The other three members of this family are processed pseudogenes each flanked by different short direct repeats; two derived from the 1.8 kb mRNA and the other derived

from the 2.6 kb mRNA (Gwo-Shu Lee *et al.,* 1983). Not all retroposons result in pseudogenes. The human phosphoglycerate kinase family includes a functional intronless PK2 gene and the pseudogene psi hPgk-1 (McCarrey, 1990).

Redundancy can be created in a genome when one or several genes or even the entire genome is duplicated in its entirety. Segmental duplication can happen during unequal "crossing over" resulting in homologous recombination between similar sequences. When complete genes are included in the duplicated sequence, the gene duplication events can lead to growth of gene paralogues. Resulting gene families contribute to copy number variation and increased gene dosage effect. When mutations collect in one or more of a gene's duplicates, they may be silenced into a pseudogene or become a unique gene with a new distinct or specialized function. Approximately 40% of the coding regions in the olfactory receptor cluster on human chromosome 17 are taken up by pseudogenes (Sharon *et al.,* 1999).

Inversions

Inversions are DNA segments that have been flipped compared to their relative position in the genome. In heterozygotes they have been noted to suppress recombination of the region in which they reside. An inversion will cause phenotypic effects if it disrupts a gene or regulatory region that controls expression of a gene or genes.

Many cases of Hemophilia A were found to be caused by an inversion. Hemophilia A is an X-linked disorder caused by various mutations in the factor VIII gene (Lakich *et al.,* 1993). Approximately 43% of individuals with this disease possess an inversion involving this gene due to chromosomal homologous recombination beginning mainly in male germ cells (Antonarakis *et al.,* 1995). In horses, a 43Mb inversion near the KIT gene has been found responsible for the Tobiano coat pattern (Brooks *et al.,* 2007).

Translocations

Mismatching during meiosis can lead to translocations where one segment of a chromosome is removed from its normal location and moved to a different position on that chromosome or to another chromosome. Large translocations are usually

microscopic in size and categorized as cytogenetic defects. Smaller translocations have to be characterized on a molecular level. Autosomal translocations, characterized cytogenetically, have been associated with repeated early embryonic loss in horses (Lear *et al.,* 2008). Translocations of genetic material from the X or Y chromosomes can have developmental effects on sex determination (Villagómez *et al.,* 2011). Examples of human diseases caused by translocations include acute myeloid and acute lymphoblastic leukemia. In these diseases the Myeloid/Lymphoid (MLL) gene has been implicated because it spans the breakpoint of an 11q23 translocation (Thirman *et al.,* 1993).

Methylation/Imprinting

The methylation status of any genomic region can determine whether that region is packaged tightly into heterochromatin and transcriptionally silenced or whether it is more loosely packed into euchromatin exposing it to molecular factors for its active transcription. The methylation status of some regions or genes can be maternally or paternally predetermined at conception, a phenomena known as imprinting. The hybrid offspring of horses and donkeys are known as hinnies (product of female donkey male horse) and mules (product of female horse and male donkey). Their physical and temperamental differences are suspected to be the result of preferentially expressed maternal/paternal genes (Allen *et al.,* 1993). Hinnies tend to be smaller in stature like the donkey dam and have longer manes and shorter ears like the horse sire. Mules tend to be taller like the horse dam with longer ears and sparse mane more like the donkey sire. Both tend to have temperaments more similar to the sire as well. In a recent study comparing, horses, donkeys, mules and hinnies, it was found that 15 known imprinted genes, play a major role in placental development. These 15 equine imprinted genes matched imprinting direction as previously seen in human, mouse, hinny and mule. The five maternally expressed genes included *insulin-like growth factor 2 receptor* (*IGF2R*) and *nucleosome assembly protein 1-like 4* (*NAP1L4*). The ten paternally expressed genes included *Insulin-like growth factor 2* (*IGF2*) and *paternally expressed gene 3* (*PEG3*). An additional 78 paternally biased candidate imprinted genes were also identified (Wang *et al*., 2013). The different prenatal environments resulting from these differentially expressed genes may explain a few of these differences.

There are some genes that are dosage dependent and one active copy is all that is required for development and life. Two active copies may result in a disease state. To protect against this, one copy, the mother's or the father's (depending upon the gene), is silenced by methylation of cytosines in GC rich regions within that genomic area. These genes are inherited from the mother or the father in a silenced or active state known as imprinting. The *insulin like growth factor 2* (*IGF-2*) gene is affected by tissue specific maternal imprinting in mice. Both alleles are transcriptionally active in the choroid plexus and leptomeninges, but only the paternal *IGF 2* allele is active in all other embryotic tissues (DeChiara *et al.,* 1991). Genes may also be silenced or activated by increased or decreased methylation due to environmental cues or point mutations that alter the GC content of a region. The changes to DNA methylation or histone acetylation as a result of exposure to carcinogens is implicated in some cancers by the mechanism of increased methylation of DNA repair genes or deacetylation of histones allowing reactivation of developmental genes, both of which may possibly also be a catalyst for increased mutation rate of surrounding DNA or destabilization of nearby microsatellites. Excessive methylation throughout a 5' "CG island" region of the pi-class glutathione S-transferase gene (GSTP1) is associated with human prostate cancer (Lee *et al.,* 1997).

Heterogeneity of Phenotypes

Not all phenotypes are the result of single discrete Mendelian variation. There are some diseases with well defined phenotypes that are caused by different variations in different families or populations, but within the same gene. Alternately, there are diseases with similar, yet difficult to distinguish, phenotypes which can be the result of variations in different genes that play roles in the same pathway or process. Severe combined immunodeficiency (SCID) is a group of autosomal recessive diseases in people that is genetically and phenotypically heterogeneous. Two of the first phenocopies of SCID characterized in the 1970s were purine nucleoside phosphorolase deficiency and adenosine deaminase deficiency both of which are inherited in an autosomal recessive manner (Giblett *et al.,* 1975, Giblett *et al.,* 1972 and WHO Scientific Group, 1995). X-linked forms of SCID are due to mutations in the γ chain of IL-2 receptor (Schwaber and Rosen, 1990). The more rare causes of SCID are the result of MHC class II deficiency,

reticular dysgenesis, CD3γ or CD3ε deficiency or CD8 deficiency (Reith *et al.,* 1988, Ownby *et al.,* 1976,). There is heterogeneity of variants within the genes responsible for each type of SCID. SCID is a condition in horses and mice, but equine and murine SCID are a result of a deficiency in *DNA-PK*; a gene not implicated in any cases of human SCID(Perryman and Torbek 1980, Pla and Mahouy 1991).

Cystic Fibrosis (CF) is an autosomal recessive disease that is the result of many different individual mutations in the *CFTR* gene. Several of the variations responsible for CF and differences between their manifestations of the disease are covered in a review by Lommatzsch and Aris (2009). The *CFTR* locus was mapped by positional cloning to human chromosome 7q in 1989 and was the first genetic disease mapped in this fashion (OMIM 1998). This disease has a broad range of manifestations depending on the mutation(s) present in the individual. Generalized atrophic benign epidermolysis bullosa (GABEB) is another autosomal recessive disease responsible for sub-epidermal blistering, alopecia, dystrophic nails and dysplastic teeth which is caused by different independent mutations in the same gene. This disease is due to reduced or eliminated expression of the collagen17 gene, *COL17A1* also known as *BPAG2*, resulting from nonsense mutations, which are genetic variations that introduce premature stop codons (PTCs) into a transcript. A few of the mutations responsible for GABEB are SNPs altering an amino acid residue directly to a PTC (Darling *et al.*, 1997). Another causative variation is in the form of a double base deletion which creates a frame shift in the reading frame of the resulting RNA creating a PTC downstream of the deletion (McGrath *et al.,* 1996). These variations each cause disease when an individual is homozygous for any one, but individuals can also be affected when they are compound heterozygotes (Darling *et al.,* 1997). Fragile-X Syndrome, hemophilia, Duchenne's and Becker's Muscular Dystrophy and red-green color blindness are other examples of X-linked trait with heterogeneity of mutations within a single gene.

Complex Genetic Modes of Inheritance

Many common traits and diseases are not the result of simple Mendelian inheritance of a single variant. These are defined as quantitative traits that result from the cumulative action of multiple genes that interact with each other in a biological pathway or process. Causative genes identified in the study of complex traits are known as quantitative trait loci (QTLs). In humans, height clearly has a genetic component but is the product of many genes, deduced from early genetic studies over 95 years ago (Galton, 1886, Fisher, 1918). Heritability of height is approximately 0.8, meaning it has a very strong genetic component, but roughly 20% of height is determined by environmental factors (Visscher *et al.,* 2010, Allen *et al.,* 2010). Hypertension (high blood pressure) is a common disorder that is the result of complex polygenic inheritance (Rapp, 1983). Hypertension can lead to stroke, heart attack or other forms of cardiovascular disease. Cardiovascular disease is the leading cause of death in the world (WHO, 2011). Initial investigations to locate QTLs for hypertension were done crossing hypertensive rat strains with normotensive strains then crossing the F1 progeny to produce offspring to be tested for segregating marker alleles (Rapp, 1983; Rapp, 1987 and Rapp, 1991).

Milk yield, fat percentage, fat yield, protein percentage and protein yield are polygenic traits in dairy cattle. These traits are the consequence of the interaction of multiple genes and are partially affected by environmental factors. The first statistically significant QTLs for milk production in Holstein dairy cattle were mapped to five chromosomes implementing interval mapping and “progeny testing” using microsatellites (Georges *et al.,* 1995).

Osteochondrosis dissecans (OCD) is a condition in horses where the bone under the articular cartilage fails to completely form resulting in tears and fissures in the cartilage surface above and can affect any bone/joint surface. It can affect the cartilage of any joint, but clinical significance is attributed to joints that affect soundness. Heritablility for Osteochondrosis of the hock joint (OC) and palmar/plantar osteochondral fragments (POF) was investigated in a study that included 24 Swedish Standardbred trotter sire families. This study concluded a heritability of 0.34 for OC and 0.23 for POF

(Philipsson *et al.*, 1993). It is thought that several genes play a role in location and severity, as well as its age of onset. There appears to be an increased occurrence of this in young horses with rapid growth. This condition is considered to be a complex trait and it is believed diet, injury, exercise and genetics all contribute to its development (Stromberg, 1979). Microsatellite based WGAS conducted over the last decade have identified 14 quantitatitive trait loci that play a role in OCD, but only a few match locations between the different breeds of horse (Distl, 2013).

Section IV: Genetic Markers and Maps

The rediscovery of Mendel's genetic principles led to investigations for Mendelian genetic variation in diverse organisms. For example, this led to investigations of blood groups in animals such as; cow (Ferguson, 1941), pig (Andresen and Baker, 1964) and horse (Stormont and Suzuki, 1964). Biochemical and isozyme alleles were found to be valuable and potentially polymorphic markers to observe for hereditary variation between individuals. Isoenzyme markers also became useful biochemical markers for certain malignancies (Bostick *et al.*, 1978).

During the 1970s, scientists were sequencing the genes they identified. Some of the first sequencing was accomplished through chemical modification of the DNA and cleavage of the modified product at Harvard University (Maxam and Gilbert, 1977). The development of the chain-terminator sequencing method designed by Frederick Sanger at Cambridge University was a groundbreaking step in molecular genetics (Sanger and Coulson, 1975 and Sanger *et al.*, 1977). Higher sequencing throughput increased the discovery rate of polymorphisms in the genome such as SNPs and microsatellites. Restriction fragment length polymorphisms, RFLPs, are the result of polymorphisms in the genome that introduce or eliminate a cleavage recognition site for restriction enzymes. Restriction enzymes are a special class of bacterial endonucleases that recognize specific sequences of nucleotides in double stranded DNA and cause cleavage of both strands in a specific place of the recognition site. Their original function was to degrade bacteriophage DNA to restrict its ability to proliferate in the bacteria. The first specific cleavage site was found for the restriction endonuclease Hae I (Kelly and Smith, 1970). Sites for these enzymes were some of the first markers used for mapping purposes (Botstein *et al.*, 1980). An initial linkage map of the human genome was developed in 1989 with 403 polymorphic loci, including 393 RFLPs linked to approximately 95% of the genome (Donis-Keller *et al.*, 1987). Sickle Cell Anemia, an amino acid variant in hemoglobin resulting from a single base change, was the first instance of a disease phenotype found to be linked to a polymorphism at a restriction site (Kan and Dozy, 1978). The gene for Huntington's disease was mapped to chromosome 4 and the region where it resided was narrowed considerably using RFLPs (Shoulson and Chase, 1975, Gusella *et al.*, 1983, Gilliam *et al.*, 1987). Restriction sites contain only two alleles, so

RFLPs aren't always sufficiently informative for mapping endeavors. The more polymorphic nature of microsatellites made them a new valuable tool for gene mapping and parentage testing (Litt and Luty, 1989, Weber and May, 1989). A second generation linkage map was constructed using family segregation analysis, with 813 markers on 22 autosomes and the X-chromosome spanning regions covering about 90% of the estimated length of the human genome (Weissenbach *et al*., 1992). Over the next two years, the human marker map would be expanded to 2,066 markers (Gyapay *et al*., 1994)

In 1990, the NIH gene-mapping project, also known as the human genome project, was initiated. What began in 1911 with the assignment of a single gene on the X-chromosome became the project of the century in 1990, a 15 year plan with the end goal of sequencing the entire 3 billion base pairs of the human genome. This project, which seemed farfetched at its inception, was completed 2 years ahead of schedule in 2003. A press conference on April 14th of that year and special issues of *Nature: Double Helix at 50* (April 24, 2003) and *Science Building on the DNA Revolution* (April 11, 2003) marked this historic landmark in genetic research.

The human genome of approximately 3 billion base pairs was sequenced and subsequently assembled in 2001 to find that it actually only harbors 20 to 25 thousand genes (Venter *et al.,* 2001, International Human Genome Sequencing Consortium 2004). To explain human complexity from so few gene products, it is believed that about 70% of human genes produce multiple mRNAs from alternate splicing of exons or parts of exons (Black, 2003). A single gene can produce multiple mRNA and/or protein products. Each gene product can affect multiple genes upstream and downstream from it or even affect its own transcription with positive or negative feedback. A common phenotype can arise from multiple variations in a single gene or from variations in a variety of genes within a network. Different variations in a single gene can play a role in many phenotypically different traits or diseases or be the cause of a single disease or phenotype. Even environmental factors can affect gene expression or even result in the introduction of variations which lead to disease phenotypes.

Genome Sequences and SNP Discovery

The Human Genome Project ushered in an era of genomics with the development of methods and tools to cheaply and rapidly sequence DNA and investigate gene expression. With the advent of the genomics era, the complete genomes of more than 180 organisms have been sequenced (http://www.genomenewsnetwork.org/resources/sequenced_genomes/genome_guide_p1.shtml). The genomes of pathogens have been sequenced to aid in the development of vaccines and therapies; Neisseria meningitidis, Heliobacter pylori, Mycobacterium tuburculosis and Pseudomonas aeruginosa (Pizza *et al.,* 2000; Tomb *et al.,* 1997; Cole *et al.,* 1998 and Stover *et al.,* 2000) to name a few. The genomes of model organisms; *Drosophilia melanogaster* (Adams *et al.,* 2000; Myers *et al.,* 2000; Rubin *et al.,* 2000), mouse (Asif *et al.,* 2002) and Norway rat (Gibbs *et al.,* 2004) were sequenced to inform us about genetics, The genomes of many vertebrates, including cattle (Zimin *et al.,* 2009), chickens (Hillier *et al.,* 2004)), dogs (Lindblad-Toh *et al.,* 2005)) and horses (Wade *et al.,* 2009), were sequenced to inform us about the genome organization of vertebrates relative to that of human and mice.

This surge of DNA sequencing led to discovery and use of SNPs which are, at best, only diallelic and often may not be polymorphic in families or even in certain populations. However, these are so numerous that finding a polymorphic SNP within a region is relatively simple since they occur roughly 1 per 1000 bases. Genome-wide studies identify millions of SNPs from which assays of tens of thousands would provide coverage of 20 or more SNPs per centiMorgan. For example; the low density SNP array currently used for the domestic horse includes over 54,000 polymorphic SNPs with approximately 54kb between SNPs (McCue *et al.,* 2012). SNPS can and are used for linkage studies, but the large numbers available make them suitable for genome wide association studies (GWAS), which have power for the mapping and discovery of genes responsible in complex traits and diseases (International Hapmap Consortium, 2007).

Nature of Genome Wide Association Studies (GWAS)

A genome wide association study is a way to investigate genetic associations at a population level without families. This is necessary if the trait is rare, samples are difficult to acquire or if the mode of inheritance is unclear. The discovery of more than a million SNPs and the ability to design and implement a SNP chip to assay thousands of SNPS spanning the entire genome at the same time has provided a way to perform powerful population level studies. A genome-wide association study (GWAS) is the investigation of thousands of common genetic variants in different individuals in search of any variant associated with the trait of interest. It is based on the concept of linkage disequilibrium, LD, shown in Figure 1.1. Variations in the genome arise over time. All individuals possessing the variation also share a common set of markers, some of which are close enough to the variation that they were passed down together from generation to generation. In a GWAS, informative markers in LD with the variant or gene of interest will show statistical significance.

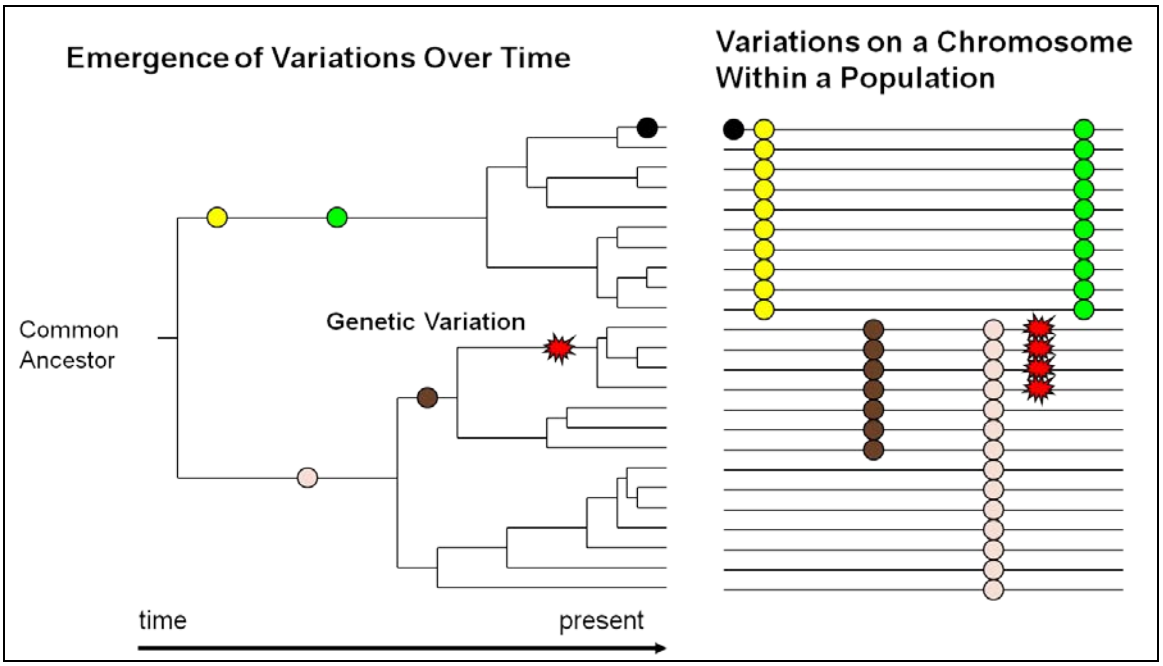


Figure 1.1: Linkage Disequilibrium Explained.
This figure shows the emergence of variations over time. The red starburst is a genetic variation of interest. This is an example of linkage disequilibrium (LD) between a specific variant and nearby markers/variants. The variation of interest is closer in time of emergence with the brown marker, but physically closer and in LD with the white marker. This physical proximity reduces the likelihood these two points will be separated by recombination during meiosis.

The completion of whole genome sequences and discovery of millions of SNPs has made it possible to investigate genetic associations at a population level and without involvement of families. This is important in the following cases:

- The trait is rare and a limited number of affected individuals are available for study,
- The trait is one which is undesirable and breeders will not select or save affected offspring, (Indeed, in extreme cases they may be reluctant to identify affected offspring as coming from their breeding stock.)
- The mode of inheritance is unclear, due to genetic complexity or due to incomplete penetrance.

In such cases, GWASs have proven effective by successfully identifying risk loci for disorders such as rheumatoid arthritis, Crohn's disease, bipolar disorder, hypertension, diabetes and many others (Stranger *et al.*, 2011). GWAS is based on a basic population genetics concept, the presence of linkage disequilibrium between genes

and markers. Animals which belong to the same breed may share common ancestors within 10, 20 and 30 generations spanning 100-300 years. Genes closely linked in the ancestor are unlikely to have been separated by genetic recombination events. As a result, there will be a statistical association between alleles of closely linked loci. In practice, human populations demonstrate linkage disequilibrium between SNPs in haplotype blocks that are from 8.8 to 25.2kb (Hinds *et al.,* 2005). For horses, dogs, and cattle, the LD ranges from 1-8 million bases (McCue *et al.,* 2010; Gray *et al.,* 2009 and Khatkar *et al.,* 2008). Therefore, when a genetic trait exists within a population and animals with and without that trait are compared, using tens of thousands to hundreds of thousands of informative SNPs, statistical associations may be detected.

Marker Maps for Linkage Studies, Especially in the Horse

With the advent of molecular genetics, several types of genetic markers were used in early linkage studies. The two earliest markers used were restriction fragment length polymorphisms (RFLPs) and simple tandem repeats (STRs) or microsatellites. RFLPs, were used to develop genetic maps in hybrid plant lines (Tanksley *et al.,* 1989) and for domestic animals (Fries and Ruddle, 1989). Microsatellites soon proved even more effective than RFLPs since they were more widely distributed throughout the genome than STRs and often had multiple loci ensuring that more loci would be informative in families (Litt and Luty, 1989; Weber and May, 1989). Linkage maps were developed for many of the organisms being studied. By 1992, a second-generation human linkage map was constructed of 814 polymorphic markers including microsatellites and RFLPs which covered about 90% of the estimated length of the human genome (Weissenbach *et al*., 1992).

Early studies of genetic markers in horses involved investigations of genetic systems useful for parentage testing. Three autosomal linkage groups were discovered prior to the initiation of the Horse Gene Mapping Workshop in 1995. The first linkage group included the K blood group locus and the 6-phosphogluconate dehydrogenase locus (Sandberg, 1974). The second linkage group consisted of four loci; serum albumin locus (Al), tobiano (TO), vitamin D binding protein locus (Gc) and the serum esterase locus (Es) (Trommershausen-Smith, 1978; Sanberg and Juneja, 1978; Weitkamp and

Allen, 1979). The third linkage group included the A blood group and lymphocyte alloantigens (Bailey *et al.*, 1979). The coat colors chestnut and roan were found to be linked to loci Al and Es in the second linkage group (Andersson and Sandberg, 1982). Identification of these linkage groups was the beginning of the budding equine marker map. These discoveries were less an effort to make a comprehensive map and more a byproduct of genetic investigations.

In the 1990s, microsatellite DNA markers were identified and found to be suitable for parentage testing (Litt and Luty, 1989; Weber and May, 1989). Over the next decade, sequencing of equine genes and genomic regions began yielding more and more microsatellites to be used in a future equine marker map and well suited for parentage testing

In 1995, a group comprised of 70 scientists from 20 countries gathered to collaborate on mapping and sequencing the horse genome. Publication of the first sizeable equine linkage map occurred in 1998 and was comprised of 100 markers in 25 linkage groups encompassing 18 equine autosomes (Lindgren *et al*., 1998). The next year, a map was published containing 182 microsatellites and 58 random amplified polymorphic DNA (RAPD) markers sorted into 33 linked groups. Twenty two of the groups were assigned to known chromosomes with the remaining 11 groups just provisionally placed upon verification by other methods (Shiue *et al*., 1999). The first linkage maps for horses were reported with 140, 161 and 353 microsatellite markers, respectively, by combining of resources from the International equine gene mapping workshop culminating in a comprehensive map of 766 markers (Lindgren *et al.,* 1998; Guerin *et al.,* 1999; Swinburne *et al.,* 2000 and Penedo *et al.,* 2005).

While linkage maps were effective at identifying genetic distances between polymorphic genetic markers, they did not provide as much information about genes. Gene sequences are highly conserved among species and comparative genetic studies using cross-species fluorescence in situ hybridization (Zoo FISH) indicated strong conservation of genome structure. Radiation hybrid maps were used to visualize the relationship of particular genes and markers across species. A comparative horse map comprised of 4103 markers made up of microsatellites, known genes, and radiation hybrids of markers from other species was compiled in 2008 (Raudsepp *et al.,* 2008). An

entire genome reference sequence was still the final goal. A complete assembled genome sequence would be the most complete marker map compiled, with SNPs, microsatellites and known genes from comparative maps all easily accessible. The whole genome of the horse was completed in 2006 with the first online reference assembly available in January 2007. The final assembly was available in September of that same year. The manuscript detailing sequencing and assembly of the equine genome was published in 2009 (Wade *et al.*, 2009). This publication included characterization of approximately 1 million single nucleotide polymorphisms (SNPs), the next tool in genetic discovery. The discovery of large numbers of SNPs spurred the development and implementation of SNP chips for equine gene discovery and mapping.

Nature of Linkage Studies

Linkage studies require assembly of families including sires, dams and offspring and the use of genetic markers that may be linked to the gene under investigation. Linkage means that the gene and genetic markers are present on the same chromosome and sufficiently close that genetic recombination will not cause their co-segregation to be random (50%). When the gene marker and the gene are very close, for example, 5 to 10 cM apart, then small numbers of offspring (20 or less) may be effective in determining a statistical proof of linkage. If the genetic markers are more distant from the gene, for example 20-30 cM apart, then a large number of offspring will need to be tested to distinguish the linkage relationship from random segregation. Informative families are often difficult to assemble; consequently, it is beneficial to have one or more genetic markers that are close to the gene of interest. This is a challenge since we do not know in advance where the gene and gene markers lie relative to one another. The solution is to have a large number of genetic markers to assure that at least one will be within 5 to 10 cM of the gene of interest. Botstein *et al.* (1980) suggested that a human linkage map with 150 evenly distributed genetic markers would be sufficient to ensure that a gene would be within 10 cM of a marker based on the observation of 3000 cM for the human gene map. At the time, it was not possible to select genetic markers which would be evenly distributed and the solution was to map a greater number of genetic markers so that, by chance, the gene in question would be sufficiently close to a linkage marker.

Family Linkage Studies

Family linkage studies are applied to search for markers in specific genetic regions which co-segregate with the gene of interest from parent to offspring. This kind of study requires assembly of families which include sires, dams and offspring. Genetic markers are used which may be linked to the investigated gene. Microsatellites, also known as short tandem repeats (STRs), are a valuable resource in this type of study.

Genetic linkage is based on the concept that genes and markers located near each other on a chromosome are inherited together. The closer genes are to one another the less likely they are to be separated during meiosis. The base pairs which compose the

genetic markers, loci and genes are all connected to one another in a linear fashion on each chromosome, like beads on a string. This linear connection can be broken and rearranged during meiosis. At the time of meiosis, the paired chromatids (one half of each chromosome received from the parents) are susceptible to recombination between them due to crossing over events.

A linkage map consists of genes and genetic markers shown in relation to one another in terms of recombination frequency are calculated as a distance, cM. A map of known markers is an invaluable tool used to discover genes responsible for specific phenotypes or diseases. Linkage mapping is accomplished with a family study where the recombination frequencies are calculated with a logarithm (base 10) of odds (LOD) score method (Morton, 1955). The LOD score, represented as Z, compares the probability of obtaining the data if the two loci are linked to the probability of obtaining the data if the two loci are not linked. A LOD score of +3 shows that the odds are 1000 to 1 that the linkage observed did not occur by chance. LOD is calculated $LOD = \log_{10} (1- \Theta)^{NR} \times \Theta^{R} / 0.5^{(NR+R)}$, where Θ = number of recombinants(R) / number of recombinants (R) + number of non recombinants (NR).

For a dominant, co-dominant or partially dominant trait, mapping is straightforward. A verified heterozygous sire possesses a trait of interest. Assuming that markers and genes in proximity to each other will be linked, a search of known markers for one that is linked with the phenotype will give an approximate location for the genetic region responsible for the trait. The sire will transmit the gene causing the trait of interest approximately 50% of the time as shown in Figure 1.2, showing a Punnett square using Cream inheritance as the example.

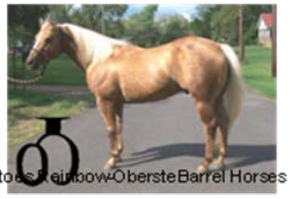
	CR	cr
CR	25% Cremello CR/CR 	25% Palomino cr/CR 
cr	25% Palomino CR/cr 	25% Chestnut cr/cr 

Figure 1.2: Incomplete Dominant Punnett Square Example: Cream Dilution. Palomino sire's genotype *CR/cr*. He will pass on *CR* 50% of the time. This Punnett square uses a palomino dam as well with genotype Cr/cr. This mating would produce Cremello 25% of the time, palomino 50% and sorrel undiluted 25%. This example is given on the chestnut base color.

A family linkage analysis is performed with data derived from a sire heterozygous for the well-defined dominant trait to be studied and 10 to 20 of his offspring. Ten is an adequate number for initial screening sample sets, but 20 or more informative offspring may be necessary to achieve statistical significance. Significance is calculated using the LOD score method to calculate the probability that two loci are linked. The LOD score is calculated as $\log_{10}$ of the Likelihood Ratio (LR): $LR = \Theta^{R} (1-\Theta^{NR})/0.5^{N}$, where the LR is the likelihood of association divided by likelihood of no association due to free recombination. Therefore the LOD value or $Z = \log_{10} \Theta^{R} (1-\Theta^{NR})/0.5^{N}$, where Θ = recombination rate R/ (NR+R), R = number of recombinants and NR = number of nonrecombinants.

Linkage Analysis: Dominant Traits

For example: A heterozygous sire tested with alleles A and B for a particular microsatellite marker. This marker would not be considered informative if the sire were homozygous for a single allele. After testing ten of his offspring, five with sire's trait of interest and five without it: four of the five offspring who share the sire's trait had allele B; the remaining offspring who shared the trait had allele A. Among the five offspring not sharing the trait, all five had allele A. This data is shown below in the bottom half of figure 2.6. In this case, all 10 offspring were informative, meaning there was no ambiguity about which genotype they received from the sire. Offspring are considered uninformative if they have the same two markers as the sire, because it is unknown which marker they inherited from him. If the genotype of the dam is known as well, then the origin of the offspring genotype may be clarified. Occasionally, even with the dam's genotype, some offspring still may not be informative. The one offspring with the sire's phenotype bearing allele A is considered recombinant. This means from 10 offspring there is 1 recombinant and 9 nonrecombinants, therefore, R = 1, NR = 9. Plugging them into the formula above will give this; $LR = (0.1^{1}(1-0.1)^{9}/ 0.5^{10)} = (0.1*(0.387)/0.000977) = (0.0387/0.000977) = 39.62$. Therefore $LOD = \log_{10}(39.62) = 1.6$. What does this mean? Association results are considered to be significant if the Z value is equal to or greater than 3. A LOD value of 3 means there is significant evidence for linkage; this means the odds are 1000:1 in favor of linkage between loci. The LOD value for these 10 offspring was only 1.6. LOD score values are additive. This means LOD scores for different sire families can be added if each individual sire has an inadequate number of offspring to attain significance when LOD values of the different families are combined. So, if a second family is studied and found to have a LOD score of 1.8, then the combined LOD score will be 3.4 and considered statistical evidence for linkage.

Linkage Analysis: Recessive Traits

For a recessive trait, a heterozygote is a carrier (not affected). Carriers are identified by the presence of affected offspring they have produced. They are expected to transmit the gene of interest 50% of the time, but only affected offspring will be informative. Unaffected offspring are un-informative because it is impossible to determine the status of carrier/non-carriers for a yet undiscovered variation. Two or more affected offspring per carrier are necessary for effective linkage analysis. According to Hardy-Weinberg equilibrium (HWE), population frequency of the affected allele can be calculated using the formula $p^2 + 2pq + q^2 = 1$, knowing also that $p + q = 1$, where p = dominant wild type allele and q = recessive variant allele. Frequencies of the different genotypes are represented by p2, 2pq and q^2. In this population the frequency of lordotic horses was 0.04, and if caused by a recessive gene, the frequency of this phenotype is represented as q^2. Consequently, the gene frequency for affected allele (q) would be 0.2. By subtraction, the normal allele (p) would have a frequency of 0.8. According to the HWE equation, unaffected carriers would be present in a frequency of 2pq or 0.32. An example of population frequencies of genotypes from crosses assuming HWE is given in table 1.1.

Table 1.1: Recessive Inheritance Example

Sires / Dams	p = 0.8	q = 0.2
P = 0.8	pp = 0.64	pq = 0.16
q = 0.2	pq = 0.16	qq = 0.04

Population frequency of genotypes occurring in crosses assuming HWE. Frequencies are represented by p2, 2pq and q^2. If the frequency of a recessive trait is 0.04 then a population frequency of the recessive (q) allele can be calculated to be 0.2 using the formula $p^2 + 2pq + q^2 = 1$.

Microsatellites have been successfully used for family studies in horses. For example, genomic position of the *silver* (*Z*) locus was found when significant linkage was shown between the candidate gene *PMEL17* and microsatellite marker *TKY284*. This analysis was accomplished in a family study containing a single sire family that included: 1 heterozygous Silver stallion, 34 of his offspring and their 29 dams which were not silver (Brunberg *et al*., 2006. The Appaloosa gene *leopard complex* (*LP*) was assigned to a region of chromosome 1 in the vicinity of two positional candidate genes for *LP* (*pink eyed dilution* (*p*) and *transient receptor potential cat-ion channel subfamily M. member 1*(*TRPM1*) using two paternal half sib families containing a total of 47 offspring (Terry *et al*., 2004). Several possible causative SNPs (6) were identified after target sequencing of a 300kb region containing *TRPM1* (Bellone *et al*., 2010). The genetic locus for *Sabino 1* (*SB1*), which causes a white spotting pattern, was determined after a successful family study involving 3 sires and 27 offspring and 7 dams (Brooks and Bailey, 2005). The genomic location where the *Cream* (*C*) gene resides was identified using a family consisting of one stallion, 7 mares and 21 offspring (Locke *et al*., 2002). The *C* locus was identified after sequencing the positional candidate gene *MATP* (Mariat *et al*., 2002). Linkage between microsatellite *CORO18* and the *Grey* locus was identified with 7 half sib Quarter Horse families including 276 offspring and 128 dams (Locke *et al*., 2002). The mutant locus was later identified as containing a 4.6kb duplication in intron 6 of the *STX17* gene (Rosengren-Pielberg *et al*., 2008).

When to Use Linkage Analysis and What it Might Miss Relative to GWAS

Family linkage studies are applied when sufficient family members are available for investigation into simple Mendelian traits. Linkage testing requires minimal marker coverage, about one per 10 to 20 cM and minimal sample numbers of 20-50 compared to the hundreds to thousands of samples and markers utilized in GWAS. However, a linkage study can still fail to produce results if markers near the region of interest are not informative. This could either be due to parental homozygosity of necessary marker(s) or unresolved offspring genotypes. Even when a study successfully identifies a linked marker locus, the locus may not show evidence of association in a population of unrelated individuals. This could be the result locus heterogeneity between family groups. For example, family linkage mapping identified the genetic region for *Sabino 1*(*Sb1*) white spotting pattern in Tennessee Walking Horses (Brooks and Bailey, 2005).

When to Use GWAS and What it Might Miss Relative to Linkage

When enough family members cannot be obtained or when a trait has complex etiology, GWAS is recommended, as was the case with lavender foal syndrome (LFS). LFS is a rare recessive lethal trait found in Egyptian Arabian Horses. This rarity precluded the collection of enough family members to perform a linkage study. Therefore, GWAS was performed which successfully identified markers in linkage disequilibrium with the trait from 6 affected samples and 30 unaffected relatives (Brooks *et al.,* 2010).

Complex traits are difficult to study by linkage as they are more often attributed to many genes with minor effect rather than a single gene(s) with major contribution. Minor effects are more readily identified in GWAS case/control population studies that provide more power in associations of limited effect on phenotype. Complex traits can be the result of incomplete penetrance due to; gene-gene interactions, environmental factors or gene-environment interactions. There are still disadvantages involved in case-control study designs compared to family linkage studies. They offer no internal check for genotyping quality. Results for controls may show a departure from Hardy-Weinberg Equilibrium (HWE), which could be caused by genotyping error, selective breeding/mating practices, negative selection, population stratification or pure chance.

The desired reason for departure from HWE would be the genotyped SNP actually plays a role in the disease or susceptibility or is in linkage disequilibrium with the causal variation. Other factors that negatively affect outcome include poor genotyping quality with missing genotypes and population stratification. The issue of population stratification can be avoided by using a family-based study, as was done in the lordosis study discussed in chapter 3.

The search for the cause of Osteochondrosis dissecans (OCD) has been approached from a linkage study standpoint and whole genome association studies with mixed results. Initially, a family linkage study was performed on Hanovarian Warmbloods with 14 paternal half sib families in a search for quantitative trait loci (QTL) using 172 initial microsatellites and 88 additional ones to refine QTLs. QTLs were found on ECA 2, 4, 5 and 16 and chromosome wide significance for QTLs was noted on ECA 2, 3, 4, 5, 15, 16,, 19 and 21 (Dierks *et al.*, 2007). GWAS involving South German Cold Blood horses found association ECA 18 (Wittwerc *et al.*, 2009). GWAS/linkage analysis was done in Norwegian Standardbred Trotters with 162 horses, 80 cases, 82 controls, 22 half sib paternal groups. Two analyses were performed; mixed model analysis and basic association test. Both identified QTLs on the same chromosomes; ECA 5, 10, 27, 28 (Lykkjen *et al.*, 2010). GWAS in French Trotters concluded associations with GM (severity) ECA13, HM (Hock location) ECA3, 13, 14 and other (other locations) ECA 13, 15 (Teyssedres *et al.*, 2012). GWAS investigation of osteochondrosis dissecans in Thoroughbred horses found association on ECA3 (Corbin *et al.*, 2012). These results have some similarities but vary more than they agree. In each study, the testing was done on different breeds with different measures and different affected joints considered. The variation in results might be attributed to the genetic heterogeneity of the condition as well as differing management practices or jobs each horse performed. Overall, multiple loci contributing to OCD have been successfully identified by one or more of the listed studies. A more concise set of parameters regarding the joint affected, severity of lesions and age of onset may be beneficial for more consistent results between studies and breeds.

Section V: Applications of Genetic Tools

The foregoing describes the nature of genetic variation and some examples of resulting phenotypes. When phenotypic variation is apparent within a population, we have two major tools for investigation: linkage studies using families and genome wide association studies using individuals selected from a population merely based on phenotype. The discovery of microsatellite DNA markers spanning the horse genome during the last 20 years has made possible genome wide linkage studies when families are available. The development of the horse genome sequence and the availability of tools to assay tens of thousands of SNPs have made possible genome wide studies in populations. The studies discussed in this dissertation describe the investigation of two different types of phenotypic variation in horses that can be investigated using these tools.

Champagne Dilution

The base coat colors of horses are red and black, encoded by the *MC1R* locus (Marklund *et al.,* 1996). The *Champagne dilution* gene (*CH*) has been found in several breeds in the United States, notably the Tennessee Walking Horse, and its effect is to dilute both the eumelanin (black) and the pheomelanin (red) pigments. The gene is sometimes confused with the effects of the *Cream dilution* gene (*Cr*), which only dilutes eumelanin. Breeders interested in breeding for this trait need an effective test to distinguish the different loci. Since the trait is considered attractive, breeders select for these animals and several families were available for study. This project is the subject of Chapter 2.

Genetics Form and Functional Structure: Extreme Lordosis

Some breeders of American Saddlebred horses are concerned about the occurrence of a condition called swayback that was thought to be hereditary. While some breeders select against the condition, others do not consider it a serious defect and tolerate such horses in their breeding programs. However, no one advertises the condition or actively selects for it. This made it difficult to obtain large and complete families segregating for this trait. Therefore, to investigate the heredity of this condition, a genome wide association study was conducted as described in Chapter 3.

CHAPTER TWO: Champagne Dilution in Horses

Summary

Champagne coat color in horses is controlled by a single, autosomal-dominant gene (*CH*). The phenotype produced by this gene is valued by many horse breeders, but can be difficult to distinguish from the effect produced by the Cream coat color dilution gene (*CR*). Three sires and their families segregating for *CH* were tested by genome scanning with microsatellite markers. The *CH* gene was mapped within a 6 cM region on horse chromosome 14 (LOD = 11.74 for $\theta = 0.00$). Four candidate genes were identified within the region, namely *SPARC* [*Secreted protein, acidic, cysteine-rich (osteonectin)*], *SLC36A1* (*Solute Carrier 36 family A1*), *SLC36A2 (Solute Carrier 36 family A2),* and *SLC36A3 (Solute Carrier 36 family A3)*. *SLC36A3* was not expressed in skin tissue and therefore not considered further. The other three genes were sequenced in homozygotes for *CH* and homozygotes for the absence of the dilution allele (*ch*). *SLC36A1* had a nucleotide substitution in exon 2 for horses with the champagne phenotype, which resulted in a transition from a threonine amino acid to an arginine amino acid (T63R). The association of the single nucleotide polymorphism (SNP) with the champagne dilution phenotype was complete, as determined by the presence of the nucleotide variant among all 85 horses with the champagne dilution phenotype and its absence among all 97 horses without the champagne phenotype. This is the first description of a phenotype associated with the *SLC36A1* gene.

[Most of the work described in this chapter was published 2008 PloS Genetics 4(9): e1000195. doi:10.1371/journal.pgen.1000195]

Introduction

Many horse breeders value animals with specific genetic variation in coat color variations. Several genes are known which diminish the intensity of the coloration and are phenotypically described as "dilutions". Two of these are a result of the *Cremello* and *Silver* loci. The molecular basis for *Cremello* is the result of a single base change in exon 2 of *MATP* gene *(Membrane Associated Transporter Protein),* also referred to as *SLC45A2 (Solute Carrier family 45 member A2*), on ECA21 (Mariat *et al.,* 2002 and Locke *et al.,* 2001). This change results in the replacement of a polar acidic aspartate with a polar neutral asparagine in a putative transmembrane region of the protein coded for by this gene (Locke *et al.,* 2001 and Brunberg *et al.,* 2006). The *Cremello* gene has an incomplete dominant mode of expression. In its heterozygous form it dilutes only pheomelanin (red pigment) whereas homozygosity for *CR* results in extreme dilution of both pheomelanin and eumelanin (black pigment) (Adalsteinsson, 1974).

The *Silver* dilution is the result of a missense mutation of *PMEL17* (Premelanosomal Protein) on ECA6. The base change causes replacement of a cytosolic polar neutral arginine with non-polar neutral cysteine in *PMEL17* (Brunberg *et al.,* 2006). In contrast to *CR,* the *Silver* locus is fully dominant and affects only eumelanin causing little to no visible change in the amount of pheomelanin. The change in eumelanin is most apparent in the mane and tail where the black base color is diluted to white and gray (Bowling, 2000).

The coat color produced by the *CH* locus is similar to that of *CR* in that both can cause dilution phenotypes affecting pheomelanin and eumelanin. However, *CH* differs from *CR* in that; 1) It dilutes both pheomelanin and eumelanin in its heterozygous form and 2) Heterozygotes and homozygotes for *CH* are phenotypically difficult to distinguish. The homozygote may differ by having less mottling or a slightly lighter hair color than the heterozygote. Figure 2.1 displays images of horses with the three base coat colors chestnut, bay and black and the effect of the *CH* locus on these colors.

Figure 2.1: Effect of Champagne Gene action on Base Coat Colors.
A) Chestnut – Horse only produces red pigment. B) Chestnut diluted by *Champagne*. C) Bay – Black pigment is held to the points (eg. Mane, tale, and legs) allowing red pigment produced on the body to show. D) Bay diluted by *Champagne*
E) Black – Red and black pigment produced, red masked by black
F) Black diluted by *Champagne*

Figure 2.2 shows *CH* foals are born with blue eyes, which change color to amber, green, or light brown and pink "pumpkin" skin which acquires a darker, mottled complexion around the eyes, muzzle, and genitalia as the animal matures (Sponenberg 2003). In contrast, foals with one copy of *CR* also have pink skin at birth but their skin is slightly darker and becomes black/near black with age. Indeed, the existence of a gene responsible for a separate dilution was not readily apparent until the discovery that some horses exhibit a dilution phenotype, without *CR*. The champagne phenotype is found among Tennessee Walking Horses, Quarter Horses and several other breeds from the United States. Here we describe family studies that led to mapping the gene and subsequent investigations of candidate genes leading to the identification of a genetic variant that appears to be responsible for the Champagne dilution phenotype.

The purpose of this study was to uncover the molecular basis for the champagne hair color dilution phenotype in horses. Here, we report a DNA base substitution in the second exon of the horse gene *SLC36A1 (Solute Carrier family 36 member A1)* that changes an amino acid in the transmembrane domain of the protein from threonine to arginine. The phenotypic effect of this base change is a diminution of hair and skin color intensity for both red and black pigment in horses and the resulting dilution has become known as champagne. This is the first genetic variant reported for *SLC36A1* and the first evidence for its effect on eye, skin, and hair pigmentation. So far, no other phenotypic effects have been attributed to this gene. This discovery of the base substitution provides a molecular test for horse breeders to test their animals for the *Champagne* gene (*CH*).

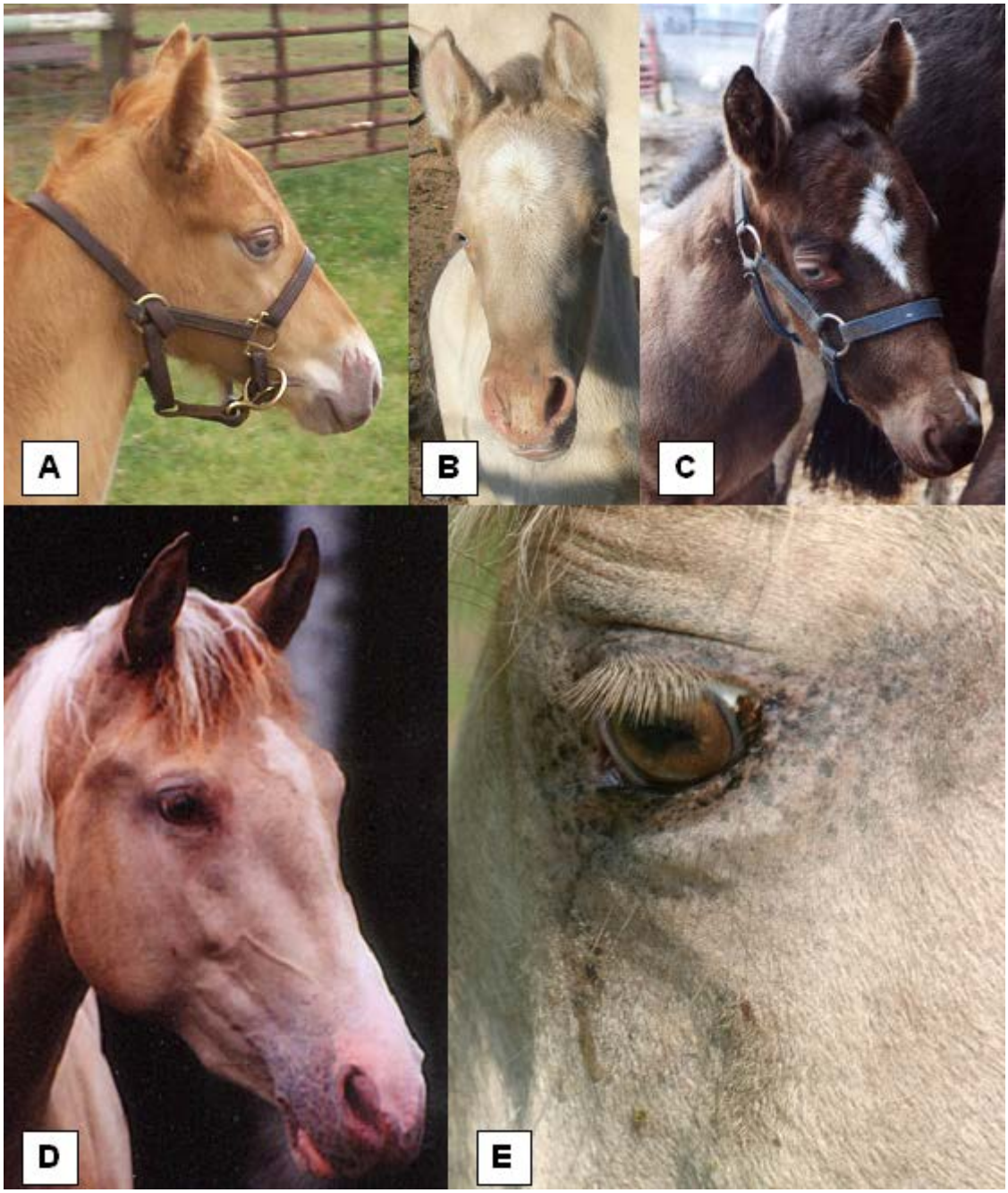


Figure 2.2: Champagne Eye and Skin Traits
A, B and C) Eye and skin color of foals
D and E) Eye color and Skin mottling of adult horse

Materials and Methods

Horses

Three half-sibling families, designated 1, 2 and 3, were used for mapping studies. Family 1 consisted of a Tennessee Walking Horse (TWH) stallion, known heterozygous at the *Champagne locus* (*CH/ch*), and his 17 offspring out of non-dilute mares (ch/ch). Family 2 consisted of an American Paint Horse stallion (*CH/ch*) and his 11 offspring out of non-dilute (*ch/ch*) mares. Family 3 consisted of a TWH stallion (*CH/ch*), 25 offspring and their 10 non-dilute dams (*ch/ch*) and 1 dilute (buckskin) dam (*ch/ch, CR/cr*). Pedigrees of the three sire families are provided in Figure 2.3.

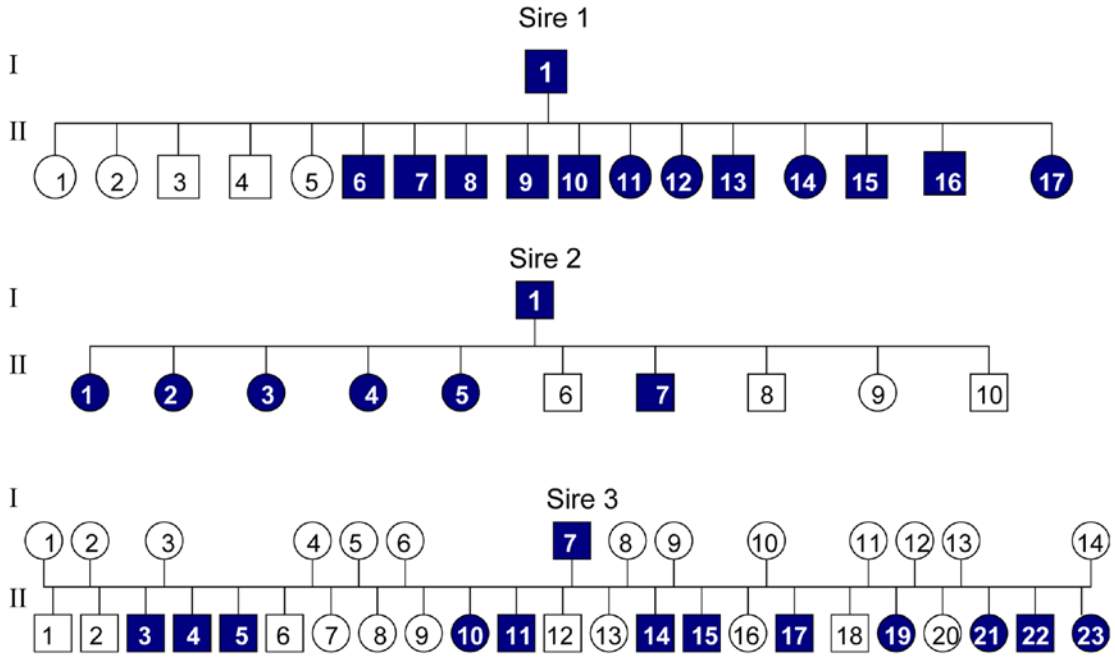


Figure 2.3: Sire and Offspring Pedigrees
Males are represented with squares and females with circles. Individuals exhibiting the Champagne Dilution phenotype are represented by solid colored symbols while non-Champagne individuals are represented with uncolored symbols. There were no dams included for sire families 1 or 2, as sire family 3 was the only family with dams available. In sire family 3, the sire is number 7 in line I which includes him and the dams. Dam's offspring are ones falling between her and the sire or between her and the dam beside her proximal to the sire. For example: Offspring number II1 in sire family 3 is the product of dam I1 and sire who is I7. Offspring II2 and II3 in sire family 3 are from dam I2 and sire who is I7.

To investigate the distribution of the gene among dilute and non-dilute horses of different horse breeds, 97 unrelated (within a single generation) non-champagne horses were chosen from stocks previously collected and archived at the MH Gluck Equine Research Center. These horses were from the following breeds: TWH (20), Thoroughbreds (TB, 35), American Paint Horses (APHA, 32), Pintos (5), American Saddlebreds (ASB, 2), one American Quarter Horse (AQHA), one pony, and one American Miniature (AMH) Horse.

Hair and blood samples from horses with the champagne dilution phenotype were submitted by owners along with pedigree information and photographs showing the champagne color and characteristics of each horse. Samples were collected from the following breeds (85 total): American Miniature Horse (9), American Cream Draft (1), American Quarter Horse (27), American Paint Horse (13, in addition to the family), American Saddlebred (2), Appaloosa (1), ASB/Friesian cross (1), Arabian crossed with APHA or AQHA horses (3), Missouri Foxtrotter(4), Mule (2), Pony (1), Spanish Mustang 1), Spotted Saddle Horse (1), Tennessee Walking Horse (19), in addition to the families).

Color Determination

To be characterized as possessing the champagne phenotype, horses met at least two of the three following criteria: 1) mottled skin around eyes, muzzle and/or genitalia, 2) amber, green, or light brown eyes, or 3) blue eyes and pink skin at birth (Sponenberg 2003). This was accomplished by viewing photo evidence of these traits or by personal inspection. Due to potential confusion between phenotypes of cream dilution and champagne dilution, all DNA samples from horses with the dilute phenotype were tested for the *CR* allele and data from those testing positive were not included in the population data.

DNA Extraction

DNA from blood samples was extracted using Puregene whole blood extraction kit (Gentra Systems Inc., Minneapolis, MN) according to its published protocol. Hair samples submitted by owners were processed using 5-7 hair bulbs according to the method described by Locke *et al.* (2002). The hair bulbs were placed in 100 µl lysis solution of 1X FastStart Taq Polymerase PCR buffer (Roche), 2.5 mM $MgCl_2$ (Roche), 0.5% Tween 20 (JT Baker, Phillipsburg, NJ) and 0.01 mg proteinase K (Sigma-Aldrich, St Louis, MO) and incubated at 60°C for 45 minutes, followed by 95°C for 45 min to deactivate the proteinase K.

Microsatellite Genome Scan

The genome scan was done in polymerase chain reaction (PCR) multiplexes of 3 to 6 microsatellites per reaction. The 102 microsatellite markers used are listed in Table 2-1. Primers for these microsatellites were made available in connection with the USDA-NRSP8 project (Guérin *et al.,* 1999). Two additional microsatellites were used; *TKY329* [18] was selected based on its map location between two microsatellites used for genome scanning (*UM010* and *VHL209*) and *COOK007* was developed in connection with this study based on DNA sequence information from the horse genome sequence V2 as viewed in the UCSC genome browser (Kent *et al.,* 2002) in order to investigate linkage within the identified interval. Primers for *COOK007* were designed using Primer 3 software accessed online (Forward, 5'- 6FAM-CATTCCAAACACCAACAACC - 3'), (Reverse, 5' – GGACATTCCAGCAATACAGAG – 3') (Rozen and Skaletsky, 1998).

Microsatellite PCR reactions were done in multiplexes, with anywhere from 3 to six microsatellites included per reaction. They were grouped according to annealing temperature and color differences of the dye probes; VIC (green), 6FAM (blue), NED (yellow). Multiple markers using the same color probe could be used if the product sizes differed by enough base pairs. PCR product from each assay is filtered to remove excess fluorescent labeled primers and dried down to be rehydrated with formamide for analysis in the ABI 310 sequencer. For example: microsatellite *LEX020,* from table 2.2, has seven known alleles. The fluorescently labeled product from PCR will fall in the approximate size range from 198 to 222 base pairs long. Preferentially, the sire will test for two

differently sized alleles in or near this size range in order for the marker to be informative. Each offspring will produce results for two alleles as well. If the offspring has one or both alleles matching only one allele the sire possesses, then that offspring is deemed informative. The results for that individual are uninformative if it has two alleles which match both of the sire's alleles.

Amplification for fragment analysis was done in 10 µl PCR reactions using 1X PCR buffer with 2.0mM $MgCl_2$, 200 µM of each dNTP, 1 µl genomic DNA from hair lysate, 0.1U FastStart Taq DNA polymerase (Perkin Elmer) and the individual required molarity of each primer from the fluorescently labeled microsatellite parentage panel primer stocks at the MH Gluck Equine Research Center. Samples were run on a PTC-200 thermocycler (MJ research, Inc., Boston, MA) at a previously determined optimum annealing temperature for each multiplex. Capillary electrophoresis of product was run on an ABI 310 genetic analyzer (Applied Biosystems). Results were then analyzed using the current version of STRand microsatellite analysis software (http://www.vgl.ucdavis.edu/informatics/STRand/).

The initial multiplex scans were conducted on a subset of samples from Family 3 which included sire 3, five non-champagne offspring, and five champagne offspring. When the microsatellite allele contribution from the sire was not informative, (e.g. the sire and offspring had the same genotype), dams from family 3 were typed to determine the precise contribution from the sire. When the inheritance of microsatellite markers in family 3 appeared to be correlated with the inheritance of the *CH* allele, then the complete families 1, 2 and 3 were typed and the data analyzed for linkage by LOD score analysis (Morton 1956). Final genotypes and LOD scores for marker *COOK007* are given in figure 2.4.

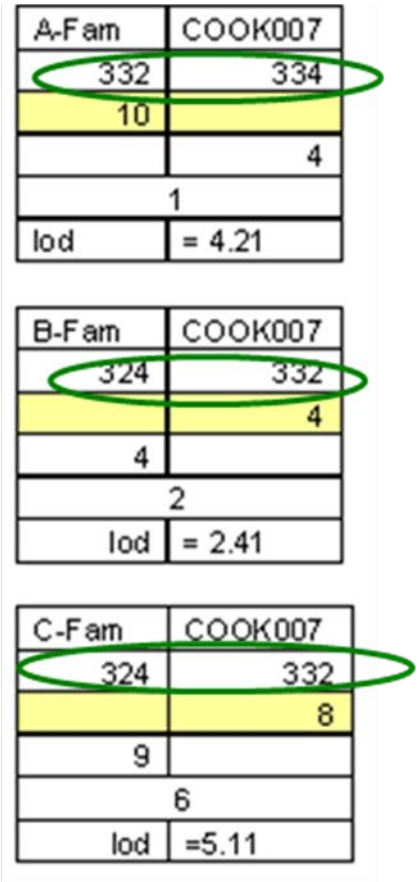


Figure 2.4: Allele Results and LOD Scores for Microsatellite COOK007
Among the three sire families. Sires A, B and C correspond to sires 1, 2 and 3 in that order from previous pedigree image. The numbers 324 and 332 are the length of each allele represented by the microsatellite amplified by the COOK007 primer pair. This microsatellite is on Equine Chromosome 14 at ~25.6 Mb. The data suggests this microsatellite is in linkage disequilibrium with the Champagne allele. The first line of each box identifies the sire family and the microsatellite primer ID. The second line is the pair of alleles the heterozygous champagne sire possesses for that microsatellite. The third line represents the champagne offspring and the allele they received from the sire. The fourth line is non-champagne offspring and the allele they received from the sire. The fifth line contains the offspring (champagne and non-dilute) whose genotype was ambiguous and sires contribution could not be determined. The sixth and last line is the manually calculated LOD score for each family's data.

Table 2.1: Microsatellites Used for Genotype Screening

Marker	Chromosome	Alleles	Size	Dye	Annealing Temp	Optimal concentration (uM)
AHT21	01	8	199-215	VIC	58	0.09
COR100	01	7	212-230	VIC	56	0.15
HMS15	01	11	207-245	6FAM	56	0.6
VIAS-H34	01	7	144-160	VIC	58	0.7
ASB8	01	8	138-164	VIC	58	0.08
LEX020	01	7	198-222	6FAM	58	0.12
ASB41	01	6	156-168	NED	58	0.3
NVHEQ100	01	7	197-217	6FAM	58	0.1
ASB18	02	12	196-213	6FAM	58	0.15
A-14	02	10	220-248	VIC	58	0.8
COR065	02	9	280-292	NED	58	0.3
ASB17	02	17	93-125	6FAM	58	0.1
UM007	02	18	122-176	VIC	58	0.05
COR033	03	9	222-254	VIC	58	0.2
ASB23	03	7	187-213	VIC	58	0.1
UCDEQ437	03	8	167-193	NED	58	0.1
SGCV23	04	8	221-233	VIC	56	0.6
ASB22	04	9	155-177	NED	58	0.1
COR089	04	10	282-304	NED	58	0.1
LEX004	05	6	282-300	NED	58	0.15
LEX069	05	7	248-262	NED	56	0.6
LEX034	05	6	252-262	NED	58	0.05
TKY28	06	7	280-364	NED	58	0.15
COR088	06	7	283-297	NED	58	0.4
COR070	06	12	279-307	NED	58	0.2
NVHEQ82	06	6	133-147	VIC	58	0.01
VIAS-H7	07	12	116-146	6FAM	58	0.3
COR004	07	6	297-319	NED	58	0.5
COR003	08	8	195-215	6FAM	58	0.15
COR056	08	10	194-220	6FAM	58	0.2
LEX023	08	12	233-257	VIC	58	0.15
ASB14	08	9	118-136	6FAM	58	0.05
HTG8	09	7	185-197	6FAM	56	0.6
ASB4	09	6	128-140	VIC	58	0.7
COR008	09	12	251-277	NED	58	0.12
UM037	09	7	108-124	6FAM	58	0.2
ASB6	10	8	185-212	6FAM	58	0.6
COR048	10	10	178-186	NED	58	0.3
ASB9	10	9	67-113	6FAM	58	0.3
NVHEQ18	10	15	119-161	VIC	58	0.06
COR020	10	7	162-176	NED	58	0.2

Table 2.1: cont…

Marker	Chromosome	Alleles	Size	Dye	Annealing Temp	Optimal concentration (uM)
SGCV24	11	10	125-141	VIC	56	0.5
SGCV13	11	5	169-179	NED	58	0.7
LEX068	11	7	162-174	NED	58	0.12
SGCV8	12	8	126-143	VIC	58	0.2
SGCV10	12	6	179-187	NED	56	0.6
AHT17	12	11	123-147	6FAM	58	0.1
COR058	12	12	218-244	VIC	58	0.1
ASB37	13	6	132-146	VIC	58	0.04
COR069	13	8	273-287	NED	58	0.13
VHL047	13	4	134-150	VIC	58	0.05
COR002	14	5	235-243	VIC	58	0.05
UM010	14	7	112-126	6FAM	58	0.05
VHL209	14	6	91-105	6FAM	58	0.2
AHT16	15	7	130-153	VIC	58	0.2
COR014	15	12	149-164	VIC	58	0.04
B-8	15	8	88-110	6FAM	56	0.1
COR075	15	9	202-220	6FAM	58	0.07
HMS20	16	8	116-140	6FAM	58	0.25
L15.2	16	9	147-165	VIC	58	0.07
LEX056	16	7	218-234	VIC	58	0.08
I-18	16	9	93-119	6FAM	58	0.05
COR007	17	9	163-177	NED	58	0.04
LEX055	17	7	216-232	VIC	58	0.15
NVHEQ79	17	7	175-197	6FAM	58	0.08
TKY19	18	9	147-173	NED	56	0.04
COR096	18	8	315-329	NED	58	0.2
LEX054	18	10	170-190	NED	58	0.08
COR092	19	6	191-203	6FAM	58	0.1
LEX036	19	8	148-170	NED	58	0.05
LEX073	19	11	249-277	NED	58	0.15
LEX052	20	7	208-214	VIC	58	0.05
LEX071	20	7	192-211	6FAM	58	0.12
HMS42	20	0	132-140	VIC	58	0.07
UM011	20	11	167-187	NED	58	0.1
COR073	21	8	187-205	6FAM	58	0.1
LEX037	21	4	196-202	6FAM	56	0.07
SGCV16	21	5	154-194	NED	58	0.12
HMS47	22	7	203-215	6FAM	58	0.17
HTG21	22	7	131-143	VIC	58	0.08
COR016	22	7	184-203	6FAM	58	0.12
COR055	23	9	240-270	NED	58	0.1

Table 2.1: cont…

Marker	Chromosome	Alleles	Size	Dye	Annealing Temp	Optimal concentration (uM)
LEX074	24	10	155-175	NED	58	0.15
COR024	24	6	214-226	VIC	58	0.05
COR061	24	11	197-227	6FAM	58	0.1
COR018	25	7	251-283	NED	58	0.1
NVHEQ70	26	7	192-208	6FAM	58	0.08
COR071	26	8	188-210	6FAM	58	0.05
A-17	26	8	102-118	6FAM	58	0.15
COR017	27	12	241-267	VIC	58	0.15
COR031	27	7	210-224	VIC	58	0.08
COR040	27	8	282-300	NED	58	0.25
UCDEQ425	28	8	236-250	VIC	58	0.1
COR027	29	7	231-255	VIC	56	0.12
COR082	29	7	199-233	VIC	58	0.4
L12.2	29	10	136-156	VIC	58	0.05
LEX025	30	7	152-168	VIC	58	0.04
LEX075	30	8	144-164	6FAM	58	0.04
COR038	31	4	210-214	VIC	58	0.1
AHT33	31	8	151-167	VIC	58	0.15
UM038	X	7	120-144	6FAM	58	0.2
LEX022	X	7	110-124	6FAM	58	0.12

The microsatellites on equine chromosome 14 are highlighted in green. The first column is the microsatellite name based upon lab location which identified it. The second column is the chromosome on which the microsatellite is located. The third column shows the number of alleles possible. The fourth column gives the size spectrum for the alleles, given in base pairs. The fifth column is the fluorescent dye label for that primer set. The sixth column is the optimal annealing temperature for PCR of that microsatellite. The last column is the optimal primer concentration for successful PCR.

Sequencing

PCR template for sequencing was amplified in 20 µl PCR reactions using 1X PCR buffer with 2.0 mM $MgCl_2$, 200 µM of each dNTP, 1 µl genomic DNA from hair lysate, 0.2U FastStart Taq DNA polymerase (Perkin Elmer) and 50 nM of each primer. Exon 2 of gene *SLC36A1* was sequenced with the following primers: Forward (5'-CAG AGC CTA AGC CCA GTG TC-3') and Reverse (5'-GGA GGA CTG TGT GGA AAT GG-3') at an annealing temperature of 57°C. Additional primers used to sequence the other exons are listed in Table 2.2. Template product was quantified on a 1% agarose gel, then amplified with BigDye Terminator v1.1 cycle sequencing kit according to manufacturer's instructions (Applied Biosystems, Foster City, CA), cleaned using Centri-Sep columns (Princeton Separations Inc., Adelphia, NJ), and run on and ABI 310 genetic analyzer (Applied Biosystems). Six samples were initially sequenced: 2 suspected homozygous champagnes (based on production of all champagne dilution offspring when bred to at least 10 non-dilute dams), 2 heterozygotes, and 2 non-dilute horses. The results were analyzed and compared by alignment using Vector NTI Advance 10.3 software package (Invitrogen Corporation, Carlsbad, California).

Table 2.2: Primers Used for Exon Screening

Part 1 - Primer sequences for the sequencing of horse *SPARC*		
Exon #	Forward Primer (5'-3')	Reverse primer (5'-3')
1	TCAGGATTGAAAGGGAACTG	ATCTCCACCCAAAAGTCTGC
2	CGAGTTTGGCATCTTCTTGC	AGGTCTGGGAATCAATCAGG
3	CACGCCTTTTCAGCTTGAC	AATTCATGTTCCCCAGAGTCC
4	GCTCTCCTAAATGCTGATTGC	CATCTCCACCTTCTTTTCGTG
5	GGAGCACAAABAGTGGGTTC	GTGGAAGAGATTTGCCCAAG
6	AGCCACTGTGCCAAGGTC	GAGGGGAAGTCTGTGTCCTG
7	CTTTTGTGCATGAGATGCTG	GGAGCAGTGAAGGCTGCTG
8	ATTGCACATCGTCTCCTTCC	GAGTGAGTGCTAATGCTTGAGG
Primer sequences for the sequencing genomic exons of horse *SLC36A1*		
Exon #	Forward Primer (5'-3')	Reverse primer (5'-3')
1	AAGCGCCTGTCTGTCTCTTC	CTCTTCCTCAGCACCAGCTT
2	CAGAGCCTAAGCCCAGTGTC	GGAGGACTGTGTGGAAATGG
3	ATCCCAGGAGCCTCTGTTCT	GAAGCGTTAAGCCAACAGGA
4	CAGCTGCTCAGCATCACAG	CAAAGCAGGAACAGCCCTTA
5	CTCCTGCCATTCCAGTCTTG	AAGGCTCAGTGTGTGAACGA
6	AACCCAGCTCAGACAGTTGG	CCAGAGACCTTTGGCAATGT
7	TTTGTGCATGCTCCAACATT	GTCAAGTCCTCGTGCAAGGT
8	CTTATCGCAGGAGGCAGAAC	GTCCCACATGGGTAACAAGG
9	TATCGTGGAGCTGGTTGTGT	TTCTCTGCACCATCTGGACA
10	CTTGCAGTGAGAGACAGGTTATTC	TGGGGAACATACAGCGGT
Primer sequences for the sequencing genomic exons of horse *SLC36A2*		
Exon #	Forward Primer (5'-3')	Reverse primer (5'-3')
1	CAGAGCCTGCTAAGGCACAC	TTTTCCTCAGCTGCACAATG
2	GACTGTGAGAAGGCCAGGAG	GCTTGTGGAGGCCACTCTAA
3	AAAAGGGCTTGGAAACCAGT	AGCAGAACCTCGCCTTAGGT
4	CTGCCACAGTGTTCTTTCCA	CACCCTGTTGCTGGAGGTAT
5	GGGACAGAAATGGAAACGAC	CCTTGAGGACAGGTCCAAGA
6	CACCCTAACTCGCTGAGACC	ATAAAAGGCTCTGCCCACTG
7	TCCTCTGGCTCTTTGGTTGT	AGGCCAGACGTTGCTTTCT
8	CCTTTCACCCATCAATGGAC	TAGCCTTGAGTCCCCATCAC
9	CTGCTCTGACTCCCTCTTGG	CTGGTTCCTGACCATCCTTC
10	GTCTAAGCCTGGGATGATGC	TCGTAGCTGGTGAATGCTTG

Reverse Transcription (RT-PCR)

RT-PCR was performed in 25 µl reactions using a Titan One Tube RT-PCR Kit (Roche) according to enclosed protocol. The primers used are listed in section 3 of Table S4. RNA from different tissues of non-dilute horses was used to acquire partial cDNAs containing the first two exons for *SLC36A1*, first three exons *SLC36A2* and first 4 exons of *SLC36A3*. The cDNA acquired was sequenced and the resulting sequences were verified for their respective genes with a BLAT search using the equine assembly v2 in ENSEMBL (http://www.ensembl.org/Equus_caballus/index.html) genome browser. RT-PCR was also performed utilizing RNA extracted from skin, kidney and testes of non-dilute animals currently in lab stocks. *SLC36A1* cDNA was produced from the skin and blood using 50 ng RNA per reaction. *SLC36A2* cDNA was produced from testes using 1 mRNA per RT-PCR reaction then following up with a nested PCR for shorter product. *SLC36A2* cDNA was produced from skin using 50 ng mRNA per RT-PCR reaction. Nested PCR was not necessary. *SLC36A3* cDNA was produced from testes using 1 ug mRNA per reaction. 9 µl of initial reaction was visualized on a 2% agarose gel to check for visible bands of product. When product was not initially detected an additional 20 µl PCR was performed in reactions as outlined above using 5 µl of RT product in the place of hair lysate per reaction. Detected product was then sequenced with the protocol listed above. Sequences were then used in a BLAST search using equine genome assembly 2 on ENSEMBL genome browser to verify the correct cDNA was amplified.

Custom TaqMan Probe Assay

A Custom TaqMan® SNP Genotyping Assay (Applied Biosystems) was designed for c.188C/G SNP in filebuilder 3.1 software (Applied Biosystems) to test the population distribution of the *SLC36A1* alleles. A similar assay was also designed to test for the cream SNP. These assays were run on a 7500HT Fast Real Time-PCR System (Applied Biosystems). All dilute horses tested for *SLC36A1* variants were concurrently tested for *SLC45A* variants. Horses testing positive for *CR* alleles were not used in the dataset to avoid any confusion over the origin of their dilution phenotype.

Results

Linkage Analyses

Table 2.3 summarizes the evidence for linkage of the *CH* gene to a region of ECA14. The linkage phase for each family was apparent based on the number of informative offspring in each family. Recombination rates (θ) were calculated based on the recombination rate for all families. Four microsatellites showed significant linkage to the *CH* locus; *VHL209* (LOD = 6.03 for θ = 0.14), *TKY329* (LOD = 3.64 for θ = 0.10), *UM010* (LOD =5.41 for θ = 0.04) and *COOK007* (LOD = 11.74 for θ = 0.00) Most notably, no recombinants were detected among 39 informative offspring between the *CH* and *COOK007* microsatellite locus.

Table 2.3: Linkage Analysis between the Champagne Dilution gene and Microsatellite Markers *UM010*, *COOK007*, *TKY329* and *VHL209*.

			alleles		Sire contribution				Statistics	
Family	*(CH)*	microsatellite	*a/b*	N[1]	a+	a-	b+	b-	LOD score	Θ[2]
3	*(+/-)*	*UM010*	*124/108*	23	14	0	0	9	5.41	
								Σ=	5.41	0.04
1	*(+/-)*	*COOK007*	*332/334*	14	10	0	0	4	4.21	
2	*(+/-)*	*COOK007*	*332/334*	8	4	0	0	4	2.41	
3	*(+/-)*	*COOK007*	*332/324*	17	8	0	0	9	5.12	
								Σ=	11.74	0
1	*(+/-)*	*TKY329*	*117/139*	15	10	2	0	3	1.92	
2	*(+/-)*	*TKY329*	*111/137*	9	5	1	0	3	1.34	
3	*(+/-)*	*TKY329*	*117/139*	18	7	0	1	10	3.64	
								Σ=	6.9	0.1
1	*(+/-)*	*VHL209*	*95/93*	13	4	1	1	7	1.49	
2	*(+/-)*	*VHL209*	*91/93*	12	4	2	1	5	0.46	
3	*(+/-)*	*VHL209*	*95/93*	24	10	1	1	12	4.08	
								Σ=	6.03	0.14

The microsatellites are given in the third column. The first column shows the sire families with data for each microsatellite. The second column gives the sire status which is heterozygous *CH* with [+] designating the champagne allele and [–] designating the non-champagne allele he carries. The fourth column gives the base pair size for the sire's alleles for that microsatellite which have been designated allele [a] and allele [b] for the data in columns 6-9. Column 5 is the number of informative offspring for that sire on that microsatellite. Column 6 is the number of champagne offspring possessing the [a] allele and column 7 is the number of non-champagne offspring possessing the [a] allele. Columns 8 and 9 show the number of champagne offspring and non-champagne respectively possessing the [b] allele. The tenth column is the manually calculated LOD score for each sire family for that microsatellite. The sum of LOD scores for all families for each microsatellite is given below the LODs as Σ=.

1. N = the number of informative meiosis
2. Θ = recombination frequency between that microsatelite and the *champagne gene* for all families, combined

Figure 2.5 identifies the haplotypes for offspring showing recombination between the genetic markers and the *CH* locus. Table 2.4 exhibits the microsatellite haplotype obtained for all individuals tested. The *CH* locus maps to an interval between *UM010* and *TKY329* with microsatellites. No recombinants were detected among 39 informative offspring between the *CH* and *COOK007* locus.

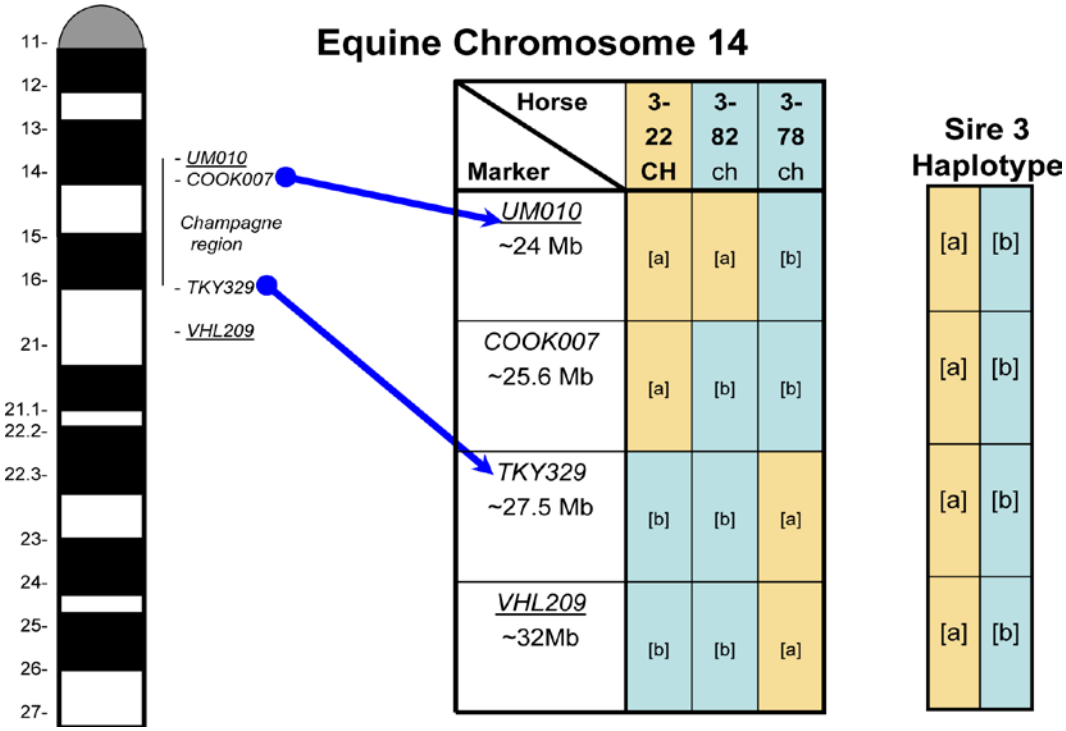


Figure 2.5: Recombinant Haplotypes
Linear relationship from top to bottom between the microsatellites, phenotype, and genotype of recombinant offspring is shown here, with yellow denoting phenotype when coloring block of individual and hypothesized association coloring allele blocks. The [a] alleles represent the speculated paternal champagne genotype and [b] alleles represent the paternal non-champagne genotype. The marker column gives the approximate genomic location for each marker on equine chromosome 14.

Table 2.4: Haplotype Data for 3 Families From Figure 2.3 Pedigree Chart

SIRE 1												
Sire 1	Color	Sex	UM010			COOK007			TKY329			VHl209	
I-1	CH	Sire	124	124		332	334		117	139		93	95
II-1	ch	female	--	--		334	338		117	127		87	93
II-2	ch	female	108	124		324	334		139	139		93	95
II-3	ch	male	116	124		332	334		117	127		85	95
II-4	ch	male	--	--		334	338		127	139		87	93
II-5	ch	female	122	124		334	338		137	139		85	93
II-6	CH	male	122	124		--	--		--	--		93	95
II-7	CH	male	110	124		332	342		117	123		93	95
II-8	CH	male	--	--		332	332		117	137		95	95
II-9	CH	male	122	124		332	338		117	137		85	95
II-10	CH	male	122	124		332	338		117	137		85	95
II-11	CH	female	116	124		--	--		--	--		93	95
II-12	CH	female	--	--		332	338		117	135		93	95
II-13	CH	male	--	--		332	338		117	137		87	95
II-14	CH	female	--	--		332	332		117	137		87	95
II-15	CH	male	--	--		324	332		117	135		87	95
II-16	CH	male	--	--		332	338		117	127		87	93
II-17	CH	female	--	--		332	338		117	127		87	95

SIRE 2												
Sire 2	Color	Sex	UM010			COOK007			TKY329			VHl209	
I-1	CH	Sire	--	--		332	334		111	137		91	93
II-1	CH	female	--	--		332	338		111	139		87	93
II-2	CH	female	--	--		332	332		111	139		91	101
II-3	CH	female	--	--		332	332		111	139		91	93
II-4	CH	female	--	--		--	--		111	139		87	91
II-5	CH	female	--	--		332	334		117	137		87	91
II-6	ch	male	--	--		334	338		111	137		93	95
II-7	CH	male	--	--		330	332		111	139		87	91
II-8	ch	male	--	--		334	338		137	139		93	95
II-9	ch	female	--	--		324	334		137	139		87	93
II-10	ch	male	--	--		332	334		111	127		87	91

The allele designations are the base pair lengths of the amplified microsatellite alleles. Sire family 3 was the only one with dams included. Each heterozygous champagne sire is listed first for each family and is in purple. For family of sire 3 the dams are listed below the sire in pink with their respective offspring immediately below them. Champagne offspring are in tan and non-champagne are in light blue. The genotypes for each offspring derived from the sire are colored in tan and blue depending on their inferred association with the champagne allele from sire.

Table 2.4: cont…

SIRE 3												
Family 3	Color	Sex	UM010			COOK007			TKY329			VHl209	
I-7	CH	Sire	108	124		324	332		117	139		93	95
I-1	ch	Dam	108	124		324	338		127	139		85	93
II-1	ch	male	108	108		324	324		139	139		93	93
I-2	ch	Dam	108	124		324	338		127	139		85	85
II-2	ch	male	108	124		324	338		127	139		85	93
II-3	CH	male	124	124		332	338		117	127		85	95
I-3	ch	Dam	122	124		332	338		127	127		85	93
II-4	CH	male	124	124		332	332		117	127		85	95
II-5	CH	male	124	124		332	338		117	127		85	95
II-6	ch	male	108	122		324	332		127	139		93	93
I-4	ch	Dam	108	108		324	324		139	139		85	93
II-7	ch	female	108	108		324	324		139	139		85	93
I-5	ch	Dam	116	124		338	338		127	127		85	93
II-8	ch	female	108	116		324	338		127	139		93	93
I-6	ch	Dam	118	122		324	332		137	139		85	85
II-9	ch	female	108	122		324	332		137	139		85	93
II-10	CH	female	118	124		324	332		117	139		85	95
II-11	CH	male	118	124		324	332		117	139		85	95
II-12	ch	male	108	118		324	324		139	139		85	93
I-8	ch	Dam	--	--		--	--		--	--		--	--
II-13	ch	female	108	108		324	324		139	139		93	93
I-9	ch	Dam	120	124		334	338		133	137		85	85
II-14	CH	male	108	124		332	338		133	139		85	93
I-10	ch	Dam	108	114		324	342		139	139		93	93
II-15	CH	male	108	124		324	332		117	139		93	95
II-16	ch	female	108	114		324	342		117	139		93	95
I-11	ch	Dam	124	124		332	338		137	139		85	85
II-17	CH	male	124	124		332	332		117	139		85	95
II-18	ch	male	124	124		324	332		137	139		85	93
I-12	ch	Dam	108	116		324	332		133	139		93	95
II-19	CH	female	108	124		324	332		117	139		93	95
I-13	ch	Dam	108	114		324	342		139	139		93	93
II-20	ch	female	108	114		324	342		139	139		93	93
I-14	ch	Dam	122	122		324	332		117	127		85	95
II-21	CH	female	122	124		324	332		117	117		95	95
II-22	CH	male	122	124		332	332		117	127		85	95
II-23	CH	female	122	124		332	332		117	127		85	95

Color legend: Purple = Sire, Pink = Dam, Tan = champagne phenotype (tan data = sire's CH associated allele), Red = ambiguous call

Candidate Genes

Candidate genes were selected on the basis of proximity to the marker *COOK007* and genes previously characterized as influential in the production or migration of any pigment cells. *SPARC* [*Secreted protein, acidic, cysteine-rich (osteonectin*)] was located closest at approximately 90kb downstream from *COOK007* and implicated in migration of retinal pigment epithelial cells in mice (Sheridan *et al.*, 2002). All 9 exons of *SPARC* were sequenced. Multiple single nucleotide polymorphisms (SNPs) were found but all caused synonymous mutations and will not be discussed further in this report.

SLC36A1, *A2* and *A2* were also proximal to this marker. *SLC36A* family members are solute carriers and other solute carrier families have previously been found to play a role in coat color. *SLC36A1* was the second most proximal gene to *COOK007* and is located less than 250kb downstream from *COOK007. SLC36A2 and SLC36A3* are coded for on the plus strand of DNA and are approximately 350 k and 380 k downstream, respectively from COOK007. A2 and A3 have been found to be expressed in a limited range of tissues in humans and mice (Bermingham and Pennington, 2004).

RT-PCR

RT-PCR (reverse transcription-polymerase chain reaction) was used to determine if *SLC36A1*, *SLC36A2* or *SLC36A3* were expressed in skin. *SLC36A1* and *SLC36A2* were expressed in skin and their genomic exons were sequenced. *SLC36A3* is normally only found expressed in testes of other organisms and was not detected in equine skin. Therefore, it was not investigated for the presence of SNPs. Results for RT-PCR of these three genes are shown in Figure 2.6 and primers used are shown in Table 2.6.

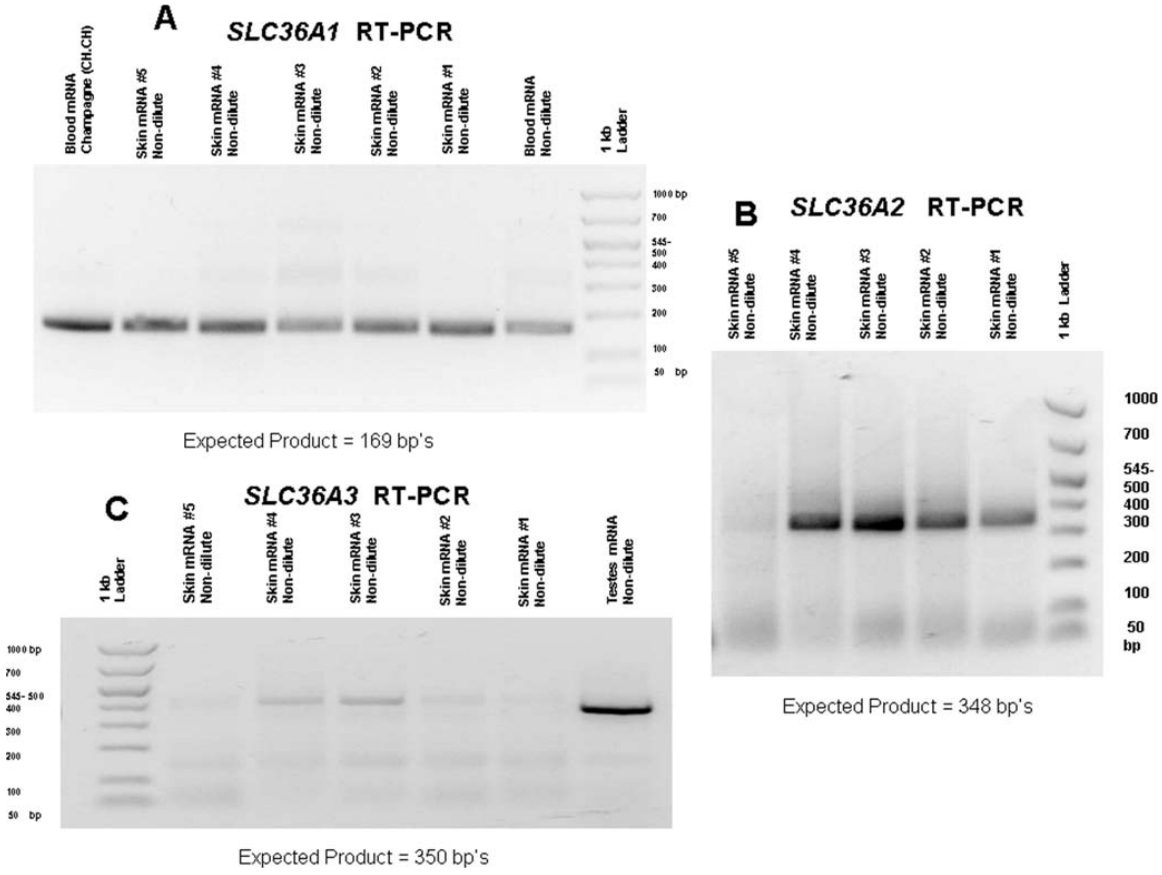


Figure 2.6: RT-PCR Product Results for *SLC36A1*, *A2* and *A3*.
A) RT-PCR results for *SLC36A1*, showing mRNA is present in blood and skin of both dilute horses and lab stock non-dilutes. B) RT-PCR results for *SLC36A2* verifying its expression in skin of lab stock non-dilute horses. C) RT-PCR results for *SLC36A3* verifying its expression only in testes with no mRNA detectible in skin of lab stock non-dilute horses. (Faint bands observed above 400 bp on gel C were sequenced and did not show homology to SLC36A3.)

Table 2.5: Primer Sequences for Reverse Transcription and Sequencing of Partial cDNAs from mRNA

Gene	Forward Primer (5'-3')	Reverse primer (5'-3')	Product (bp)	Exons Included
SLC36A1	ACCAGCGGTTTGGGGAAA	ACGGCCACGATGCCCATCA	169	1-3
SLC36A2 (outer long)	AAGCGTCATGCCCGTGACAAAGAGTGCG	GTGGCAGAAGCGCTGGGCACACCTG	348	1-3
SLC36A2 (outer short)	GCCCGTGACAAAGAGTGCG	AAGCGCTGGGCACACCTG	334	1-3
SLC36A2 (nested)	ACCTCAAACTGGACCTCAGG	ATGCCTGCGTTCCTCACA	208	2-3
SLC36A3	TTGGAAGGGACTACAACAGTGAG	GGTTTCGAGGCTGTACATCATG	350	1-4

Sequencing

All 9 exons of *SPARC* were sequenced. Three SNPs were found in exons but none showed associations with the champagne phenotype. *SLC36A2* was sequenced with discovery of 9 SNPs in exons. None of the SNPs showed associations with CH. These SNPs and all other variations detected during the course of this work are described in Table 2.6.

SLC36A1 was sequenced. Only one SNP was found; a missense mutation involving a single nucleotide change from a C to a G at base 76 of exon 2 (c.188C>G) (Figure 5). These *SLC36A1* alleles were designated *c.188[C/G]*, where c.188 designates the base pair location of the SNP from the first base of *SLC36A1* cDNA, exon 1 (Figure 2.7). Sequencing traces for the partial coding sequence of SLC36A1 exon 2 with part of the flanking intronic regions for one non-champagne horse and one champagne horse were deposited in GenBank with the following accession numbers respectively: EU432176 and EU432177. This single base change at c.188 was predicted to cause a transition from a threonine to arginine at amino acid 63 of the protein (T63R).

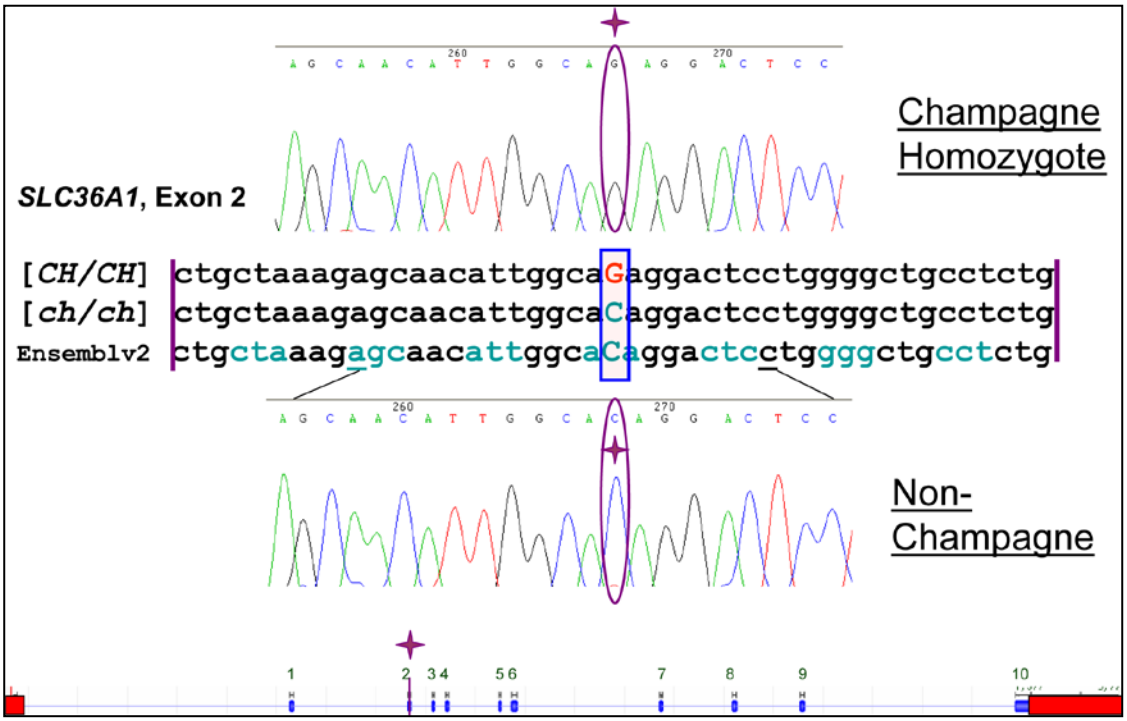


Figure 2.7: Genomic Alignment between Homozygous Champagne, Non-dilute, and Horse Genome Assembly
Allignment is flanked by cut out electropherogram reading of sequencing data for one homozygous champange horse and one homozygous non-champagne horse. Reading frame is marked by alternating colors of codons, which code for specific amino acids which will make up the resulting protein. The bottom of this figure is a diagram of *SLC36A1* highlighting the identified SNP in exon 2 with the sequence and gene layout as given by Ensembl genome browser equine assembly v2. Blue blocks of gene layout are exons and red boxes are the 5′ and 3′ UTRs. The purple star denotes the location where the variant associated with champagne was identified.

Table 2.6: Sequence Variations Detected

Gene	Location	Base Position within location	Variation	Type	5' flank	3'flank	residues
SPARC	In 1-2	1353	A>G		ATG	GAAT CTCC AG	
		1365	C>T		GAATC TCCAG	GAGC TGCT CTCC	
SPARC	Ex 2	60	C>T	syn	GGCTG AGGTG TC	GAGG TACG TGGG	Cys>Cys
SPARC	In 2-3	157	G>A		TGCGT ATTCC CA	T	
		561	C>T		AAGCC TCCCC	CAGT CTTC ACCC	
SPARC	In 3-4	87	C>T		TTCTG TGAGG TT	CCTC CCAA GGGA	
		1397	G>A		AAGG ATTGG TCT	GATT GGGT TGGG	
SPARC	Ex 4	69	T>C	mis	CAGTC CCATG TG	GTGT GCCA GGAC	Cys>Arg
SPARC	In 4-5	1827	C>T		AAAG GTGCG AAA	CCTG TGCTT TGC	

Table 2.6: cont…

Gene	Location	Base Position within location	Variation	Type	5' flank	3'flank	residues
SPARC	In 5-6	19	G>C		CTTTTCCTTGGGC	CTTCTGGGCTTC	
		49	C>A		CGAGGAGAGACC	CTGAACTGACGC	
		78	A>G		GCTGGCTGTGCC	CCCCC	
		2181	A>G		ACCTTGGGCGGG	AGAGTCAGGGTC	
SPARC	In 6-7	645	T>C		AGTCTCCCTGGA	GCTTTGTCACAG	
SPARC	Ex 7	51	A>G	syn	TGGAGACCACCC	GTGGAGCTGCTG	Ser>Gly
SPARC	In 7-8	95	T>G		GGTAGGCTCACT	GATATGGTCAGA	
SLC36A1	Ex 2	76	C>G	mis	GCAACATTGGCA	AGGACTCCTGGG	Thr >Arg
SLC36A2	5'UTR	- 89 from Start	C>	del	GCTGCTGAGA	GGGGTTTGCACT	
SLC36A2	In 2-3	9	G>C		CTGGTAAGAAGG	GCTGCCGCAGGT	
SLC36A2	Ex 6	45	C>T	syn	TTACAA	GAGACAGTGATT	Tyr>Tyr
SLC36A2	In 6-7	1339	T>C		ATTCAGGTCTAC	GTCAGAATACTC	
		1308	T>A		TAAAGGAGCTCA	ACATGCTTCGAG	

Table 2.6: cont…

Gene	Location	Base Position within location	Variation	Type	5' flank	3'flank	residues
SLC36A 2	Ex 9	94	C>T	syn	CCTCT ACGTC TT	GGCA TCCT GTGC	Trp>Trp
		155	G>A	syn	ATCGT CCCCT TC	CTGT CTCC CGCG C	Ser>Ser
SLC36A 2	In 9-10	23	T>G		TCAGC CAGAT GG	GAAG GATG GT	
		33	T>G		GAAG GATGG	?	
SLC36A 2	Ex 10	50	T>G	mis	CCTCT CCCTG GT	GGCT CCAT GAGC	Trp>Gly
		64	G>C	mis	GCTCC ATGAG CA	CAGC GCCC TGGC	Cys>Ser
		67	G>C	mis	CCATG AGCAC CA	CGCC CTGG CCCT	Cys>Ser
		227	A>G	syn	TCTGA TCCAG CC	ACAG ACCA TCTC	Ser>Ser
		265	C>T	mis	CCACC ATTTT CA	TCAG TGAG AATG	Ser>Phe
		269	G>C	mis	GATTT TCACT CA	TGAG AATG GTGC	Cys>Ser

Types of variants found within exons classified as; mis = missense, syn = synonomous. The base position given within location refers to exons/introns.

Protein Alignment

Figure 2.8 shows the alignment of the protein sequence encoded by exons 1 and 2 of *SLC36A1* for seven mammalian species with sequence information from Genbank (horse, cattle, chimpanzee, human, dog, rat and mouse). The alignment was performed using AllignX function of Vector NTI Advance 10 (Invitrogen Corp, Carlsbad, California) and demonstrates that this region is highly conserved among all species. At position 63, the amino acid sequence is completely conserved among these species, with the exception of horses possessing the champagne phenotype. This replacement of threonine with arginine occurs within a putative transmembrane domain of the protein (Boll *et al.*, 2003).

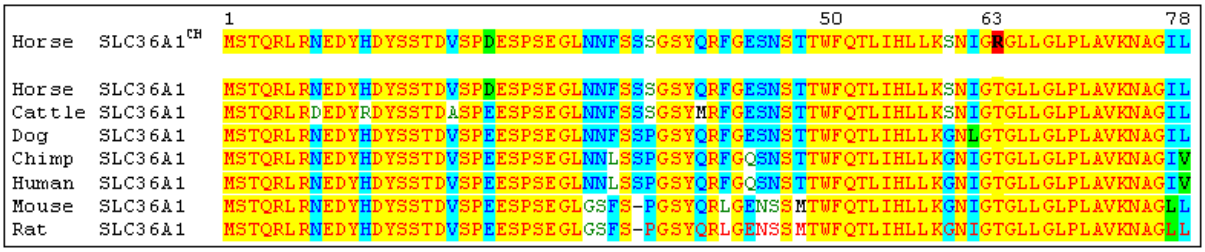


Figure 2.8: Seven Species Protein Sequence Alignment for *SLC36A1* Exons 1 and 2. The R in red is the amino acid replacement associated with the Champagne phenotype.

Population Data

The distribution of *c.188G* allele among different horse breeds and among horses with and without the champagne phenotype was investigated. Table 2.7 is a compilation of the population data collected via the genotyping assay. All dilute horses (85) not possessing the CR gene, tested positive for the *c.188G* allele with genotypes *c.188C/G* or *c.188G/G*. No horses in the non-dilute control group (97) possessed the *c.188G* allele. The horses used for the population study were selected for coat color and not by random selection; therefore measures of Hardy-Weinberg equilibrium are not applicable and were not calculated.

Table 2.7: Results for TaqMan Genotyping of c.188 Locus

	Champagne (CH/CH or CH/ch)		Non-Dilute *(ch/ch, cr/cr)*	
Horse Breeds	*G/G*	*G/C*	*C/C*	*Total*
American Cream Draft	0	1	0	1
American Miniature Horse	0	9	1	10
American Quarter Horse	1	26	1	28
American Paint Horse	0	13	32	45
American Saddlebred	0	2	2	4
Appaloosa	0	1	0	1
Kentucky Mountain	0	1	0	1
Part Arabian	0	3	0	3
Pinto	0	0	5	5
Pony	0	1	1	2
Missouri Foxtrotter	0	4	0	4
Mule	0	2	0	2
Spanish Mustang	0	1	0	1
Spotted Saddle Horse	0	1	0	1
Tennessee Walking Horse	3	16	20	39
Thoroughbred	0	0	35	35
Total	4	81	97	182

First column contains the horse breeds tested. Second column is for individuals homozygous for champagne locus. Third column contains champagne heterozygotes . Fourth column has nondilutes negative for champagne alleles. G is the actual champagne DNA allele and C is the non-dilute wild type allele.

Discussion

Family studies showed linkage of the gene responsible for the champagne dilution phenotype within a 6cM region on ECA14 (Penedo *et al.*, 2005) (Table 1). Based on the Equine Genome Assembly V2 as viewed in the ENSEMBL genome browser (http://www.ensembl.org/Equus_caballus/index.html), this region spans approximately 2.86 Mbp (Kent *et al.*, 2002). Within this region, four candidate genes were investigated; one was selected based on known effects on melanocytes (eg. *SPARC*), the other three were chosen for their similarity to other solute carrier genes previously shown to influence pigmentation (eg, *SLC36A1*, *A2*, and *A3*). While SNPs were found within the exons of SPARC, none were associated with *CH*. Of the other 3 candidate genes, only *SLC36A1* and *SLC36A2* were found to be expressed in skin cells. Therefore, the exons of those two genes were sequenced. A missense mutation in the second exon of *SLC36A1* showed complete association with the champagne phenotype across several breeds. While SNPs were found for *SLC36A2*, none showed associations at the population level for the champagne dilution phenotype.

This data is the first demonstration for a role of *SLC36A1* in pigmentation. Orthologous genes in other species are known to affect pigmentation. For example, the gene responsible for the cream dilution phenotypes in horses, *SLC45A2 (MATP)*, belongs to a similar solute carrier family. In humans, variants in *SLC45A2* have been associated with skin color variation (Graf *et al.*, 2005) and a similar missense mutation (p.Ala111Thr) in *SLC24A5* (a member of potassium-dependent sodium-calcium exchanger family) is implicated in dilute skin colors caused from decreased melanin content among people of European ancestry (Lamason *et al.*, 2005). The same gene, *SLC24A5* is responsible for the *Golden (gol)* dilution as mentioned in the review of mouse pigment research by Hoekstra (Hoekstra, 2006).

We propose that the missense mutation in exon 2 of *SLC36A1* is responsible for the champagne dilution phenotype. While this study provides evidence that this is the mutation responsible for the champagne phenotype, the proof is of a statistical nature and a non-coding causative mutation cannot be ruled out at this point. *SLC36A1*, previously referred to by the name *PAT1* (proton/amino acid transporter 1) in human and mouse (Chen *et al.*, 2003), is a proton-coupled small amino acid transporter located and most

active in the brush border membranes of intestinal epithelial cells. This protein, also known as *lysosomal amino acid transporter 1, LYAAT1*, was first identified in rat brains (Sagne *et al.*, 2001). SLC36A1 is a symporter that facilitates the coupled transport of H+ and select amino acids from both intracellular and extracellular proteolysis. It functions in the membrane of lysozomes of brain neurons and the apical membrane of intestinal epithelial cells. This function is coupled to the V-type H^+ ATPase in lysosomes and with Na^+/H^+ exchanger in the intestinal epithelial cells, which both provide the gradiant which drives transport of amino acids to the cytosol (Boll *et al.*, 2003). *LYAAT1/SLC36A1* is localized in the membrane of lysozomes and the cell membrane of post-synaptic junctions. In lysozomes it allows outward transport of protons and amino acids from the lysozome to the cytosol (Wreden *et al.*, 2003). During purification and separation of early-stage melanosomes *LAMP1/SLC36A1* is found in high concentrations in the fraction containing stage II melanosomes (Kushimoto *et al.*, 2001). Perhaps *SLC36A1* plays a role in transitions from lysozome-like precursor to melanosome. Since organellular pH affects tyrosine processing and sorting (Watabe *et al.*, 2004), an amino acid substitution in this protein may affect cytosolic pH by inhibiting the proper flow of protons from lysozomes and, therefore, the ability to process tyrosine properly. There must be an increase in pH, before tyrosinase can be activated. The cytosolic pH gradient must be maintained for proper sorting and delivery of other proteins required for melanosome development (Watabe *et al.*, 2008). Thus, the pH gradient of the cell may be altered by this mutation.

It is hypothesized that disruption of the protein structure by replacement of a non-polar transmembrane residue with a polar residue results in loss of function for this protein. This is shown here using a figure depicting the topology of the *SLC36A1* gene originally published by Boll *et al* (2003). We modified the figure with horse amino acids overlaying the first two exons with the suspected *CH* variation shown in red (Figure 2.9). This image shows the transmembrane location of variation and its vicinity to the conserved histidine residues which are necessary components of the substrate recognition site for this transporter. The polar residue would negatively affect the structure of this transmembrane region and the regions of the protein flanking it, potentially disrupting the substrate recognition site.

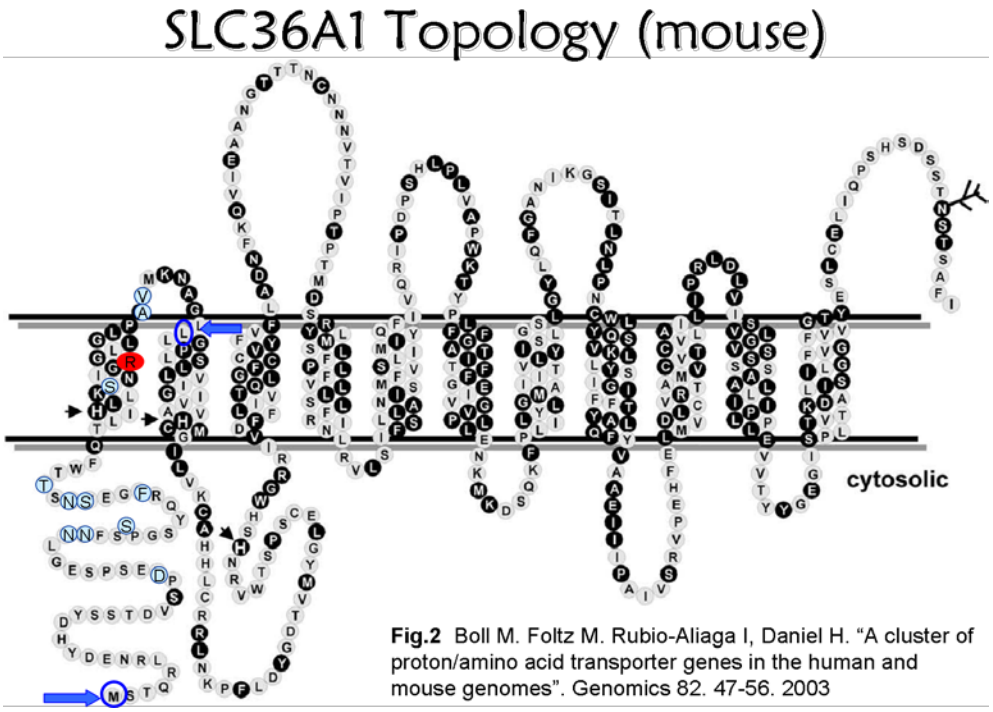


Figure 2.9: First Two Exons of SLC36A1 of the Horse Overlaid on Mouse Topology The residue differences between horse and mouse are overlaid in blue and *CH* variation T – R is highlighted in red. The beginning and end of the 2 exon segment are denoted by blue arrows.

Further investigation into this gene/protein product is necessary to elucidate its role in pigmentation. The transport properties could be compared between wild type and *SLC36A1* and the *CH* variant using the two-electrode voltage clamp technique in a flux study similar to the one by Boll *et al.* to compare the amino acid uptake rates of *SLC36A1*and *SLC36A2* (Boll *et al.*, 2002). Alternatively, cell cultures containing wild type and *CH* variant melanosomes could be compared for processes related to melanin synthesis and any differences in the ability to transfer melanosomes to keratinocytes. In another approach, copies of *SLC36A1 CH* variant could be immunofluorescently labeled and transfected into other cell lines to identify any change in cellular localization.

This variant, discovered in association with a coat dilution in the horse, is the first reported for the *SLC36A1* gene. The phenotype resulting from this mutation, a reduction of pigmentation in the eyes, skin and hair, illustrates previously unknown functions of the of *SLC36A1* protein product. Furthermore, now that a molecular test for champagne dilution is established, the genotyping assay can be used in concert with available tests for cream dilution and silver dilution to clarify the genetic basis of a horse's dilution phenotype. This will give breeders a new tool to use in developing their breeding programs whether they desire to breed for these dilutions or to select against them.

CHAPTER THREE: Genetics of Extreme Lordosis in American Saddlebred Horses

Summary

Extreme lordosis, also called swayback, lowback or softback, can occur as a congenital trait or as a degenerative trait associated with aging. In this study, the hereditary aspect of congenital extreme lordosis was investigated using whole genome association studies of 20 affected and 20 unaffected American Saddlebred (ASB) Horses for 48,165 (SNPs). A statistically significant association was identified on ECA20 (corrected P = 0.017) for SNP BIEC2-532523. Of the 20 affected horses, 17 were homozygous for this SNP when compared to seven homozygotes among the unaffected horses, suggesting a major gene in this region with a recessive mode of inheritance. The result was confirmed by testing an additional 13 affected horses and 166 unaffected horses using 35 SNPs in this region of ECA20 (corrected P = 0.036). Combined results for 33 affected horses and 287 non-affected horses allowed identification of a region of homozygosity defined by four SNPs in the region. Based on the haplotype defined by these SNPs, 80% of the 33 affected horses were homozygous, 21% heterozygous and 9% did not possess the haplotype. Among the non-affected horses, 15% were homozygous, 47% heterozygous and 38% did not possess the haplotype. The differences between the two groups were highly significant (P < 0.00001). The region defined by this haplotype includes 53 known and predicted genes. Exons from three candidate genes, *TRERF1*, *RUNX2* and *CNPY3* were sequenced but distinguishing SNPs were not found. The mutation responsible for extreme lordosis may lie in other genes or in regulatory regions outside exons. This information can be used by breeders to reduce the occurrence of extreme lordosis among their livestock. This condition may serve as a model for investigation of congenital skeletal deformities in other species.

[Most of the work described in this chapter was published in 2010 Animal Genetics 41(Suppl. 2), 64–71. doi:10.1111/j.1365-2052.2010.02108.x]

Spinal Developmental Variations

During vertebral development different spinal variations from the norm can occur. The three major congenital forms of spinal malformation are extreme lordosis/hyperlordosis, kyphosis and scoliosis. The most commonly manifested spinal deformities studied in humans are congenital scoliosis, adolescent idiopathic scoliosis and Scheuermann's kyphosis. Adolescent idiopathic scoliosis (AIS) is characterized by a lateral spinal curvature of at least 10 degrees that occurs between age 10 and the point of skeletal maturity (Ward *et al.,* 2010). Historically, finding the genetic cause for such deformities has been challenging. In humans, cases of known hereditary spinal malformations are rare (Winter *et al.,* 1983). In an extended family study of AIS, the higher risk within the families was attributed to polygenic inheritance (Ward *et al.*, 2010). Even with the polygenic etiology for AIS, one genetic risk variant near LBX1 on human chromosome 10 has been identified by whole genome association study (WGAS) (Takasha *et al.,* 2011).Scheuermann's kyphosis is a result of uneven growth of the thoracic vertebrae resulting in wedge shaped vertebrae and increased kyphotic curve of the spine. No particular genetic cause is yet known, but it is believed to be polygenic (Fotiadis *et al.*, 2008).

Lordosis

Lordosis, specifically the dorsal to ventral concave curvature of the thoracic and lumbar region of the spine, is normal and healthy in most mammals. However, extreme lordosis is associated with pathology in horses (Rooney & Pickett 1967; Rooney, 1969). This condition is also known in horses as swayback, lowback or softback. Rooney & Robertson (1996) noted that variable degrees of lordosis seem to be common in certain lines of American Saddlebred horses. Figure 3.1 shows the characteristic conformation for a horse with lordosis (Figure 3.1 top) and a normal horse (Fig. 3.1middle). A familial aspect was suspected but not established in previous work. A study of extreme lordosis among Saddlebred horses led Gallagher *et al.* (2003) to devise a method to measure the extent of lordosis and to characterize the variation found among horses in this breed. Based on this study, a threshold for considering horses to have the condition was defined,

and 5% of 294 horses were considered affected. Studies of affected families suggested, but did not prove, a recessive mode of inheritance.

In our experience, breeders are of mixed opinions regarding this trait. Extremely lordotic horses have not been routinely identified as experiencing pain, and some horses with this trait have performed well. Most breeders regard the condition as a conformation defect and avoid breeding stock with this condition. Identifying the genetic determinant(s) for this condition would provide breeders with the opportunity to better understand the genetics of the trait and use that information in their selection programs.

Traditionally, family studies have been most useful to identify the genetic locations for hereditary traits in horses such as the tobiano white spotting pattern, equine combined immunodeficiency disease and cream coat color dilution (Trommershausen-Smith 1978; Bailey *et al.,* 1997; Locke *et al.,* 2001). However, while horses with extreme lordosis are not uncommon, breeders avoid mating animals they believe likely to produce affected horses. Consequently, family studies are difficult. However, with the advent of the horse genome sequence and the availability of dense arrays of single-nucleotide polymorphisms (SNPs) for whole genome association (WGA) studies, population studies can be used to investigate the genetics of traits in horses (Wade *et al.,* 2009). The purpose of this study was to determine whether a hereditary component contributes to the lordotic trait in American Saddlebred (ASB) horses and, if so, to identify the location of genes responsible for this trait.

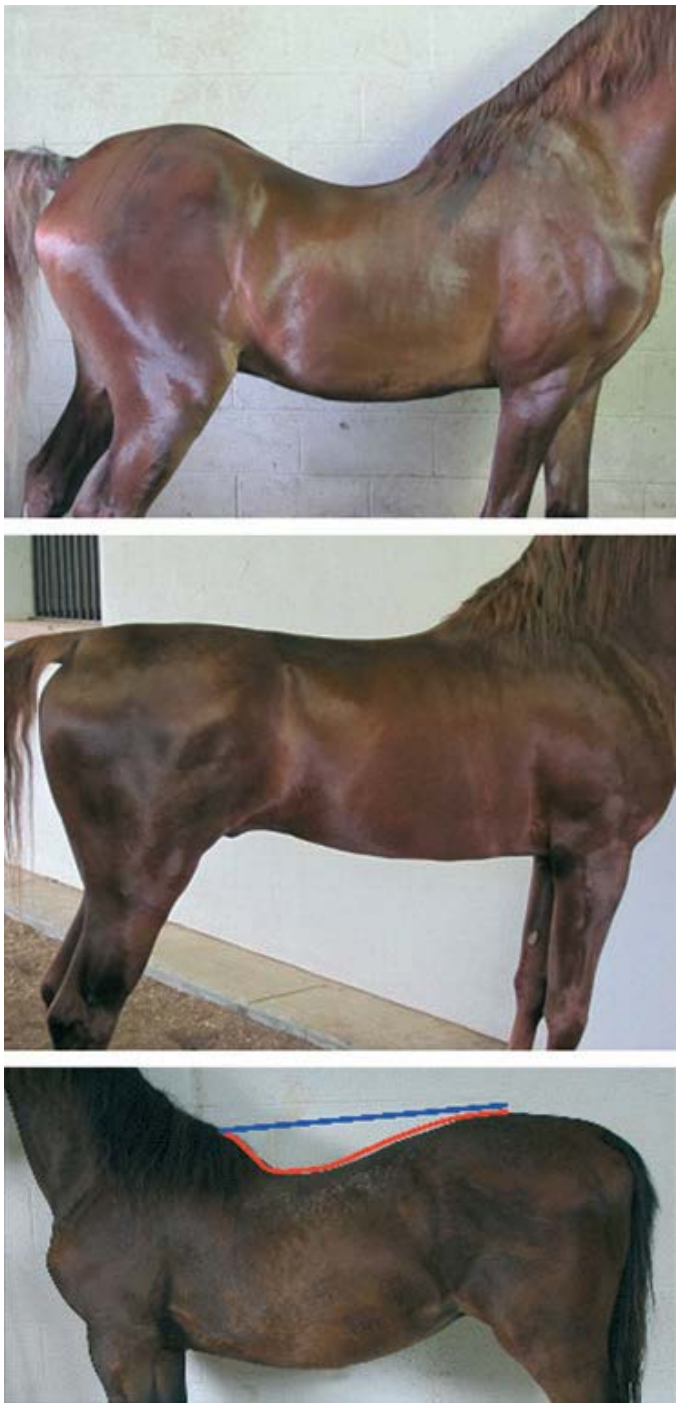


Figure 3.1: Characteristic Conformation Differences Between the Presence and Absence of Extreme Lordosis.
Horses on top and bottom exhibit extreme lordosis. Horse in the middle is consider nonlordotic and has a level topline.

Materials and Methods

Phenotypic Assessment of Lordosis

Phenotypic assessment of lordosis was based on the measurement of back contour (MBC) (Gallagher *et al.*, 2003). Two points are chosen on the horse's back, one at the top of the withers (the highest point of the dorsal spineous process in the region of thoracic vertebrae T2-T3) and one on the point of the rump (the highest point on top of the horse's hips). The shortest distance between those points was measured in centimeters and designated "A". Next, the distance along the contour of the back was measured in centimeters and designated "B". MBC was calculated as the difference between "A" and "B". Figure 3.1c and 3.2 illustrate the points of measurement on a horse exhibiting extreme lordosis.

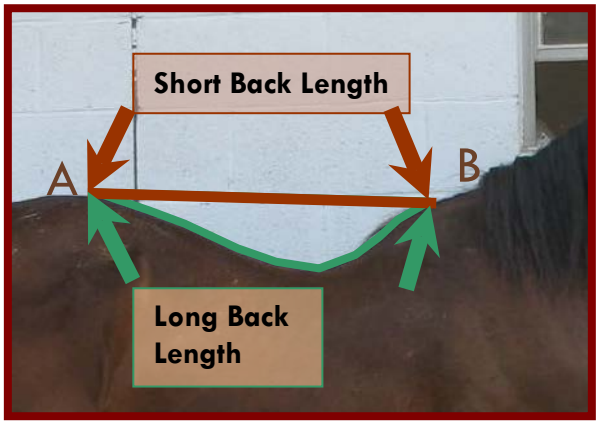


Figure 3.2: Calculating MBC.
Long Back Length – Short Back Length = MBC

Horses Measured for MBC

Measurements and tissue (hair or blood) samples for ASB horses used in this study came from a sample set of 749 American Saddlebred horses collected from private and commercial farms in and around Kentucky. The average age was 7.1 years and ranged from 1 month to 29 years old. All Saddlebred horses used in the Illumina and Sequenom assays were selected from among this group.

Horses for Illumina Assay

For the genome scan using the Illumina Equine SNP50 Chip, 40 ASB horses were selected, based on their lordosis phenotype. Twenty horses were selected based on having an MBC > 8.0 centimeters and 20 were selected based on having MBC < 5.0 cm. The control group included 19 half-siblings for horses in the affected group to reduce the chance of population substructure producing spurious associations.

Horses for Sequenom Assay

A total of 426 horses were selected for a 35 SNPs located in the genomic region suggested by the WGA study**.** Three horses were tested in duplicate to control for the quality of the testing. Twilight, the thoroughbred horse used for the whole equine genome sequencing, was also tested for quality control of known SNPs located in the genomic region suggested by the WGA study. Among the ASB horses, 33 (including the 20 original cases) had values for MBC of 7.0 or greater, 287 had MBC <7.0 cm and 106 were parents, siblings or offspring of lordotic horses. The relatives were tested to assist with haplotype determination.

DNA Extraction

DNA was extracted from blood or hair follicles for testing. DNA from blood samples was extracted using Puregene whole blood extraction kit (Gentra Systems Inc., Minneapolis, MN) according to its published protocol. Hair samples were processed using 20-30 hair bulbs according to a personally optimized Gentra protocol with a Gentra DNA purification kit. The hair bulbs were placed in 200 µl Gentra Cell Lysis solution and 0.01 mg proteinase K (Sigma-Aldrich, St Louis, MO) and incubated at 55°C over night then DNA was purified by following remaining steps from published Gentra protocol.

Illumina Equine SNP50 Genotyping

Initial SNP genotyping of 40 samples in the case/control group was performed utilizing the Illumina Equine 50 SNP chip for a WGA study. DNA was provided to the core facility at the Mayo Clinic in Rochester, MN for genotyping.

Sequenom

Ten SNPs from the WGA study and 39 additional SNPs from the EquCab2.0 SNP database (http://www.broadinstitute.org/ftp/distribution/horse_snp_release/v2/) were selected for testing using the MassArray iPLEX Gold assay on the Sequenom platform. Genotyping was performed at Proactive Genomics, LLC; Clemmons, North Carolina. The 49 SNPS selected were identified as being between positions of the SNPs flanking the region of interest at base pair 41530793 and 44585118 respectively on ECA 20. Following testing, 43 of the 49 SNPs provided quality genotyping data. Of these, 35 SNPS, with minor allele frequency > 0.01, were used for final association and haplotype analysis. Furthermore, 96% of the submitted samples had fewer than 0.05 genotypes missing and were included in analysis.

Genotyping Analysis

Genotyping data analysis was performed using PLINK v1.06 (Purcell *et. al.* 2007). Association analysis by case/control chi-square was performed on the Illumina data. To minimize error, because of the multiplicity of SNPs tested, an MPERM analysis with 10,000 permutations was done and a statistic, EMP 2 (referred to here as corrected P) verified possible associations. Once association was identified, different sized haplotypes were tested by chi-square analysis to identify the size and location of the highest associated haplotype in the region. Selected haplotypes were phased for all horses in this study to clarify familial patterns of transmission. Chi-square association analysis was also performed on the Sequenom data to verify association of the SNPs included from llumina assay and to identify any new associations. The 35 SNP haplotypes were then phased to verify familial patterns and to identify possible recombination locations. Haplotype analysis was also performed in Haploview (Barrett *et al.*, 2005) for validation of PLINK analysis. Haplotypes were identified using the HAP and PHASE options of PLINK. Haplotype frequencies were determined by direct counting.

Sequencing

To investigate candidate genes, exon sequences of two affected horses were compared to those of two unaffected horses. The two affected horses were selected as unrelated horses with MBC >7.0 and homozygous for the lordosis-associated ECA20 haplotype. Unaffected horses were selected as unrelated horses with MBC <7.0 and not possessing the ECA20 haplotype associated with extreme lordosis. When SNPS were identified and confirmed on these four horses, additional control and case horses were tested to determine whether there was an association with the lordotic trait. PCR template for sequencing was amplified in 20 µl PCRs using 1· PCR buffer with 2.0 mM MgCl2, 200 lM of each dNTP, 1 µl genomic DNA from hair lysate, 0.2 U FastStart Taq DNA polymerase (Perkin Elmer, Waltham, Mass.) and 50 nM of each primer. Template product was quantified on a 1% agarose gel, then amplified with BigDye Terminator v1.1 cycle sequencing kit according to manufacturer's instructions (Applied Biosystems), cleaned using Centri-Sep columns (Princeton Separations Inc.), and run on an ABI 310 genetic analyzer (Applied Biosystems). Primers were designed in Primer 3 (Steve and Skaletsky 1998) using 8 intronic sequences and seven exonic sequences of *TRERF1* (transcriptional regulating factor for *CYP11A1*), 15 exonic sequences for *RUNX2* (transcription factor associate with osteoblast differentiation) and six exonic sequences for *CNPY3* (regulates cell surface expression of Toll Receptor 4) (See Table S1).

Results

Distribution of MBC

The distribution of MBC values among the 749 horses in this study was similar to that found by Gallagher *et al*. (2003). MBC for the 749 ASB horses ranged from 0 cm to 17 cm; shown in figure 3.3. Most MBC values appear to fall within a normal population distribution with a mean of 3.6 ± 1.9 cm, median of 4 cm and mode of 4 cm. An MBC of 7 cm or greater was selected to classify horses as affected based on this value being approximately two standard deviations from the mean.

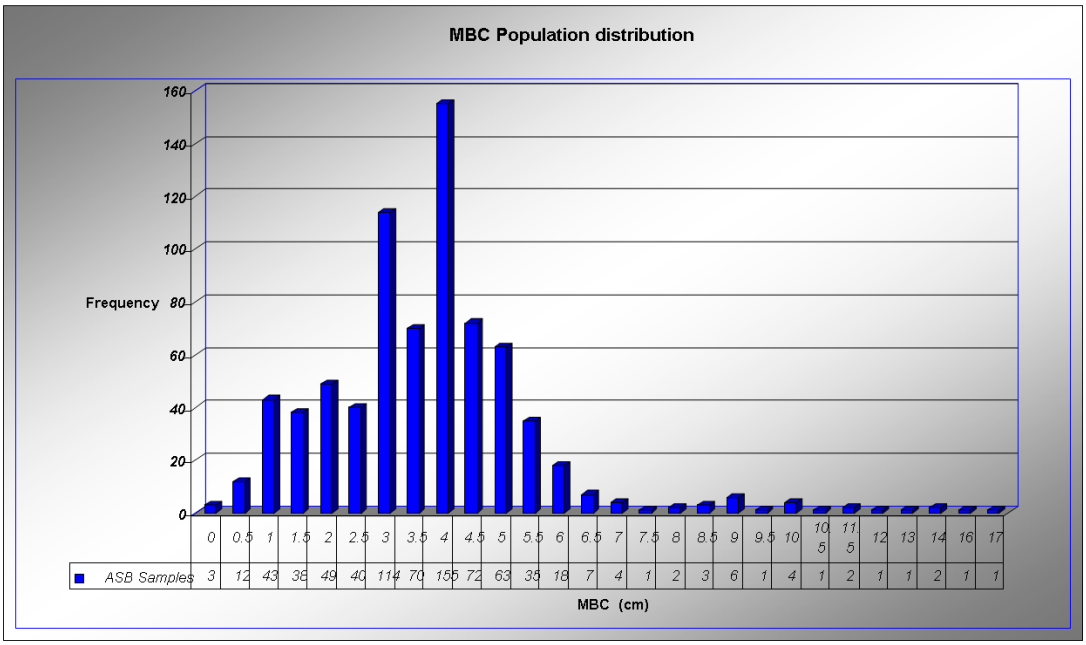


Figure 3.3: Population Distribution of MBC Among ASB Samples
X-axis = MBC measure in centimeters. Y-axis = number of horses.

Whole Genome Scan with Illumina Equine SNP50 Beadchip

The Illumina Equine SNP50 chip (Illumina) was effective in typing DNA from the initial 40 horses. The 40 individuals had an average call rate of 0.96. After filtering for minimum minor allele frequency 0.01 and genotyping of 0.90 per SNP; 48,165 SNPs were retained for data analysis.

Table 3.1 shows the top 10 statistically significant results from a 2 X 2 chi-square analysis comparing the distribution of SNPs in the DNA of affected and non-affected horses. The most significant association was found on ECA20 for SNP BIEC2_532523 (P= 6.69E-06) (Figure3.4). Of the ten most significant P values, 5 occurred for SNPs on ECA20 in the 531 kb region between positions 42,062,440 and 41,530,973. The multiplicity of comparisons will result in the spurious discovery of high chi-square values; therefore, to control for the multiplicity of comparisons, PLINK was used to conduct a Monte Carlo simulation with 10,000 permutations to calculate a corrected P value (EMP2) (Figure 3.5). Only the association with SNP BIEC2-532523 on ECA20 remained significant (EMP2 p-value 0.017).

Table 3.1: Results from Illumina Equine SNP50 assay.

CHR	BP	SNP-ID	CHISQ	P	EMP2
20	41530793	BIEC2-532523	20.28	6.69E-06	0.01699
X	51065036	BIEC2-1124071	15.66	7.60E-05	0.2554
20	42062440	BIEC2-532578	15.53	8.12E-05	0.2904
7	7444796	BIEC2-978005	13.73	0.000211	0.6762
X	51061227	BIEC2-1124068	13.47	0.000243	0.7091
7	18839866	BIEC2-984003	13.33	0.000261	0.7326
28	35119684	BIEC2-739844	13.33	0.000261	0.7326
20	41576546	BIEC2-532530	13.07	0.000301	0.7931
20	41604741	BIEC2-532534	13.07	0.000301	0.7931
20	41625809	BIEC2-532535	13.07	0.000301	0.7931

Case/Control Analysis: 20 lordotic and 20 normal back Saddlebred horses. Chromosome location (CHR), base position on chromosome (BP), SNP identifier (SNP-ID), 2X2 chi-square value (CHISQ), P value (P), and P value from Monte Carlo correction for number of comparison (EMP2) are shown.

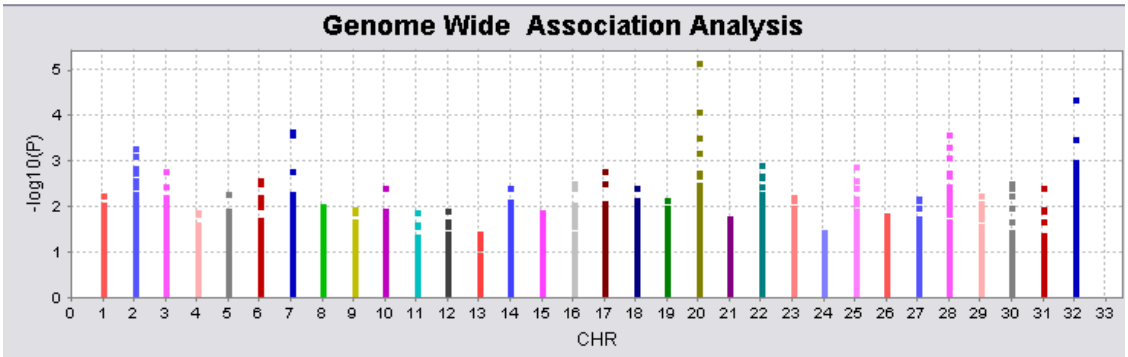


Figure 3.4: The –log P-values from Association Analyses Plotted by Haploview. Chromosomes are differentiated by color. Illumina genotype results from 20 cases vs 20 controls displayed in haploview. Possible statistical significance indicated on equine chromosome 20 with –log10 value greater than 5 which is equivalent to P < .

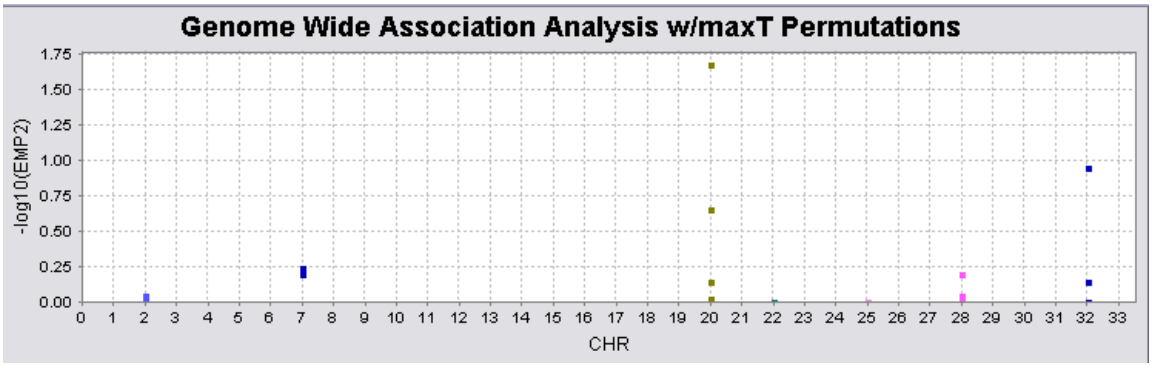


Figure 3.5: Max-T Permutation Results
X-axis = chromosomes, Y-axis = transformed EMP2 value. Ilumina 20 case vs 20 control data. Possible statistical significance is shown on equine chromosome 20.

Among the affected horses with MBC 7cm or greater, 17 out of 20 were homozygous for the T allele of the [BIEC2-532523 T/C] SNP at base pair location 41530793 on ECa 20. Among the 20 controls, only 7 were homozygous for this allele. This SNP fell in a larger region of homozygosity spanning approximately 3 Mb from base position 41,604,741 to 44,512,270, which was identified with the homozygosity function in PLINK scanning sliding windows of 30 SNPS, moving one SNP at a time and allowing one heterozygote per window for each segment to be considered homozygous. A homozygosity plot for chromosome 20 is shown in figure 3.6.

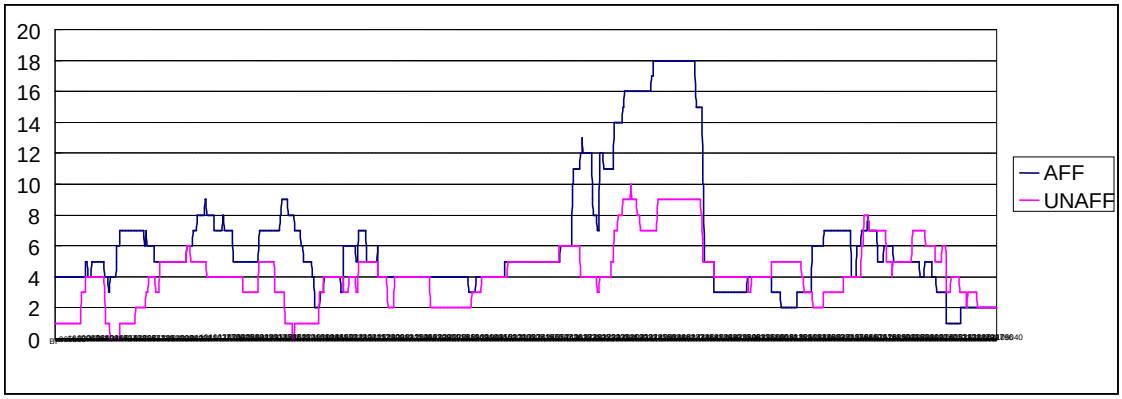


Figure 3.6: Homozygosity Plot Using the SNP Genotypes of 20 Cases vs 20 Controls on Equine Chromosome 20.
The Y-axis is the number of horses. The X-axis is base pair location on equine chromosome 20. The dark blue line is number of affected homozygous for that genetic region. The pink line represents the number of unaffected homozygous for that region. This homozygous region spans approximately 3 Mb from base position 41,604,741 to 44,512,270.

Illumina Haploview Analysis

Haploview SNP analysis verified the results gained in PLINK. The most significant SNPS were still BIEC2-532523 (p-value = 6.6E-6) and BIEC2-532578 (p-value = 8.1E-5). Figure 3.7 shows the linear relationship of the SNPS on ECA20 for the region of association. The chi-square values for each SNP are plotted by the red line above the chromosome. Darker blocks below the chromosome are indicators of high linkage disequilibrium between the two SNPs connected by that particular block.

Haploview analysis using the four-gamete rule was applied: for each marker pair, the population frequencies of the four possible two-marker haplotypes are computed. Recombination is considered to have taken place if all haplotypes were observed with the minimum frequency of 0.01. Figure 3.8 exhibits the results for the region of ECA20 showing significant association patterns using four gamete analyses in Haploview. This region spans from BIEC2-532511 to 532579. The statistical output from the analysis is given back in table 3.2. Blocks 176 and 179 represent the haplotypes bearing the most statistically significant association with extreme lordosis.

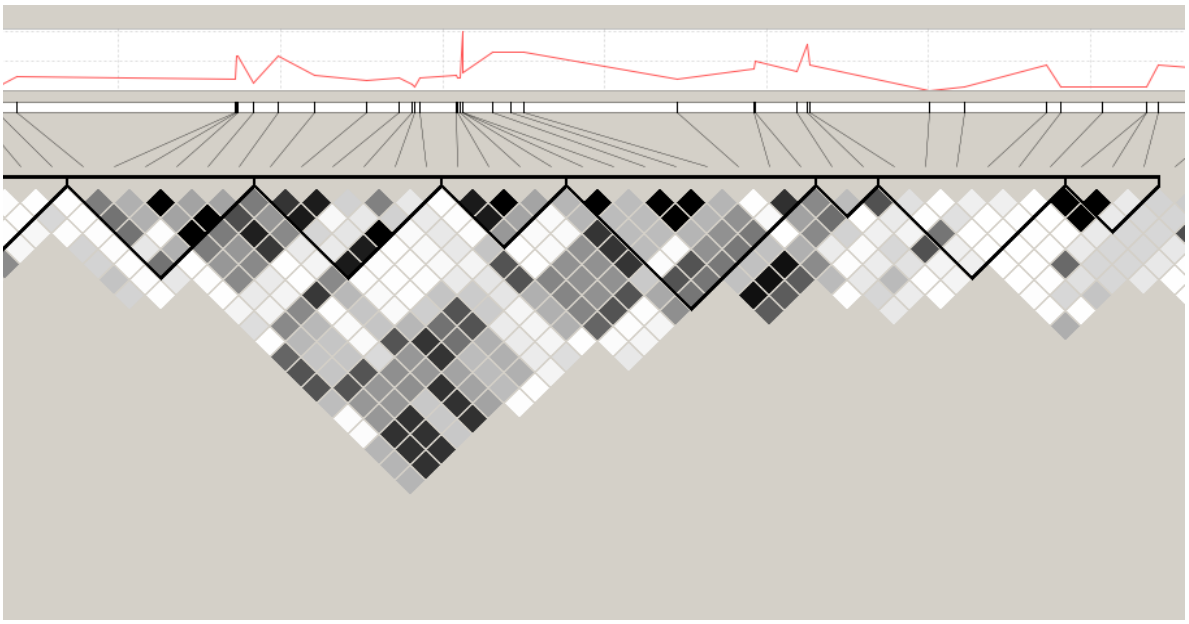


Figure 3.7: Haploview diagram. The Two highest CHISQ peaks in red designate the two most highly associated SNPs; BIEC2-532523 and BIEC2-532578.

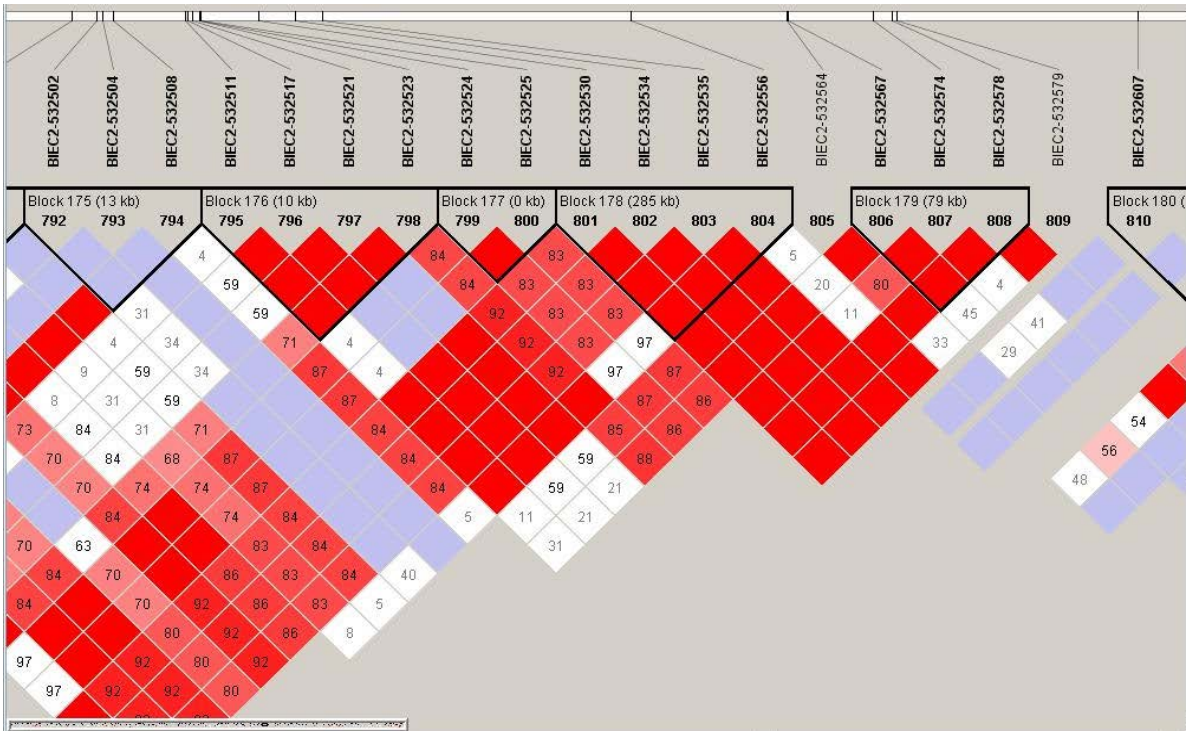


Figure 3.8: ECA20 Region of Association from Haploview Four Gamete Analyses. The SNP IDs are across the top of the diagram. The dark red blocks indicate LD between alleles. Blue blocks indicate likely recombination hot spots.

Table 3.2: Haploview Results - Four Gamete Analysis.

Haplotype	Freq.	Case, Control Ratios	Chi Square	p value
⊞ Block 168				
⊞ Block 169				
⊞ Block 170				
⊞ Block 171				
⊞ Block 172				
⊞ Block 173				
⊟ Block 174				
CAGGAAA	0.762	0.925, 0.600	11.665	6.0E-4
AGGAAAA	0.100	0.025, 0.175	5.0	0.0253
AGAAGCC	0.100	0.050, 0.150	2.222	0.136
AGGAGAA	0.013	0.000, 0.025	1.013	0.3143
AGGAGAC	0.013	0.000, 0.025	1.013	0.3143
AGGAGCC	0.013	0.000, 0.025	1.013	0.3143
⊞ Block 175				
⊟ Block 176				
AAGA	0.734	0.946, 0.521	18.492	1.706E-5
AAGG	0.129	0.004, 0.254	11.145	8.0E-4
GGAG	0.125	0.050, 0.200	4.114	0.0425
GAGG	0.013	0.000, 0.025	1.013	0.3143
⊞ Block 177				
⊟ Block 178				
AGAG	0.750	0.925, 0.575	13.067	3.0E-4
GAGG	0.163	0.050, 0.275	7.44	0.0064
GAGA	0.088	0.025, 0.150	3.914	0.0479
⊟ Block 179				
AGA	0.762	0.950, 0.575	15.531	8.1179E-5
AGG	0.088	0.025, 0.150	3.914	0.0479
GGG	0.075	0.025, 0.125	2.883	0.0895
GAG	0.075	0.000, 0.150	6.486	0.0109
⊞ Block 180				

Data reflects what is shown in figure 3.9 for the region of ECA20 associated with lordosis verifying statistical significance seen with other analysis programs.

Sequenom

To verify the statistical associations found with the Illumina assay, SNPs from the candidate region on ECA20 were tested, including 10 SNPS from the Illumina Equine SNP50 chip and 25 additional SNPs from the EquCab2.0 SNP database. Association chi-squared analyses were performed separately for the additional 13 affected and 181 unaffected horses not previously tested in the WGA study. The association analysis for just these new samples is shown in table 3.3 (CHISQ1 and P1). Based on these 13 affected horses, the association with BIEC2-532523 remained statistically significant with a P-value of 0.036. (Since this was a comparison with new samples, dictated by the original Illumina assay, no statistical correction is necessary to correct for multiplicity of testing as done for the previous experiment.) Of the 35 SNPs, 7 showed statistically significant associations ($P < 0.05$). When all 33 affected horses and 287 controls were compared, 21 of the 35 selected SNPs showed statistical significance in their distributions between the two groups based on their relative frequencies.

Table 3.3: Results from Sequenom assay Using 35 SNPs from the Candidate Region.

SNP-ID	BP	CHISQ	P	CHISQ2	P2
BIEC2_532523*	41530793	4.416	0.0356	23.72	1.12E-06
BIEC2_532534*	41604741	11.19	0.00082	43.72	3.79E-11
BIEC2_559789	41772507	0.7726	0.3794	5.543	0.01855
BIEC2_559792	41823842	0.2935	0.588	9.102	0.002553
BIEC2_532578*	42062440	5.605	0.01791	42.03	8.99E-11
BIEC2_559853	42305090	0.8888	0.3458	1.363	0.2431
BIEC2_559859	42395887	1.484	0.2231	1.669	0.1964
BIEC2_559873	42482150	2.066	0.1506	7.259	0.007055
BIEC2_532658*	42603867	8.571	0.003415	29.79	4.80E-08
BIEC2_559931	42679616	0.5499	0.4584	4.836	0.02787
BIEC2_559976	42757527	0.4859	0.4858	2.909	0.08809
BIEC2_560005	42806550	1.517	0.2181	3.629	0.05679
BIEC2_560049	42906719	3.046	0.08091	0.002347	0.9614
BIEC2_532826	42983718	1.666	0.1969	13.07	0.0003
BIEC2_560102	43022942	0.8888	0.3458	1.828	0.1763
BIEC2_560144	43069858	0.8056	0.3694	1.711	0.1908
BIEC2_560178	43195488	2.306	0.1288	10.9	0.000963
BIEC2_560197	43241900	1.747	0.1862	3.929	0.04746
BIEC2_560209	43319368	9.93	0.001626	20.93	4.76E-06
BIEC2_560255	43433802	1.213	0.2707	9.956	0.001603
BIEC2_533020	43520722	6.459	0.01104	24.51	7.38E-07
BIEC2_560328	43548897	1.173	0.2788	10.58	0.001144
BIEC2_560360	43605858	1.11	0.2921	9.664	0.001879
BIEC2_560386	43690935	2.254	0.1333	13.48	0.000241
BIEC2_560457	43914236	1.056	0.304	2.418	0.1199
BIEC2_560475	43971596	0.9724	0.3241	2.181	0.1397
BIEC2_560489	44034799	2.722	0.09899	11.67	0.000634
BIEC2_533265	44090171	7.979	0.004733	26.6	2.50E-07
BIEC2_560585	44148267	1.407	0.2356	0.964	0.3262

Table 3.3: cont…

SNP-ID	BP	CHISQ	P	CHISQ2	P2
BIEC2_560618	44188313	0.09612	0.7565	1.771	0.1832
BIEC2_533376	44241203	3.691	0.05472	19.71	9.03E-06
BIEC2_560705	44337671	1.659	0.1977	1.456	0.2275
BIEC2_560742	44403289	1.312	0.2521	0.8528	0.3558
BIEC2_560796	44484324	1.056	0.3042	8.726	0.0031
BIEC2_533588	44585118	0.5591	0.4546	1.017	0.3132

SNP identifier (SNP-ID), base position on the chromosome (BP) 2X2 chi-square value comparing the 13 new lordotic horses to 194 controls (CHISQ1), associated P value (P1), chi-square value for 33 lordotic horses and 296 controls (CHISQ2) and associated P values (P2) are shown. SNPs that showed statistically significant association with P1 or P2 are in bold type. SNPs used for the definition of haplotypes for this region are denoted with "*".

Haplotype Analysis

Haplotypes from this region (from bp location 41,530,793 to 44,585,118) were compared among affected and unaffected horses. A minimum haplotype that included the maximum number of affected horses and the lowest number of non-affected horses was identified using the four SNPs BIEC2-532523, 532534, 532578 and 532658 and spanned 1,073,074 bp. Intervening SNPs did not affect haplotype assignment. These four SNPs allowed identification of 13 haplotypes. The haplotypes and their frequencies among affected and unaffected horses are shown in table 3.4. Haplotypes were identified using the –hap-phase option of PLINK. Haplotype frequencies were determined by direct counting. Only 4 of the haplotypes had frequencies above 0.05 and had a cumulative frequency of 0.95 among unaffected horses and 0.93 among affected horses. The most common haplotype, TGTG, was the one associated with lordosis. Haploytpe TGTG had a frequency of 0.80 among affected Saddlebred horses and a frequency 0f 0.39 among non-affected horses. The population distribution for the TGTG haplogype segregated by MBC value can be seen in figure 3.10. Zygosity for this haplotype is shown for affected and non-affected horses in table 3.5. Among the affected horses, 23 (70%) were homozygous for haplotype TGTG while on 15% of the non-affected horses were homozygous. Among the 33 lordotic horses, seven (21%) were heterozygous for haplotype TGTG and three (9%) did not possess haplotype TGTG. Among non-affected, 48% were heterozygous and 35% did not possess haplotype TGTG. Statistical comparison of the combined data set to the subset of horses with lordosis was highly significant (P <0.00001).

Table 3.4: Haplotypes defined by BIEC-532523, 532534, 532578 and 532658 with frequencies

Haplotype	Affected (N=33)	Non-Affected(N=287)
TGTG	0.80	0.388
CACT	0.03	0.232
CACG	0.06	0.167
TACT	0.04	0.164
TGCG	0.05	0.019
TGTT	0.00	0.012
CATG	0.02	0.003
TGCT	0.00	0.003
CGCT	0.00	0.003
CATT	0.00	0.002
TACG	0.00	0.002
TATG	0.00	0.002
CGTG	0.00	0.002

Haplotype frequencies listed among lordotic and non-affected American Saddlebred horses.

Table 3.5: Haplotype by Affected and Unaffected Status

Classification	Number of Horses	TGTG Homozygotes	TGTG Heterozygotes	No TGTG
Lordotic	33	23 (70%)	7 (21%)	3 (9%)
Non-affected	287	44 (15%)	135 (47%)	108 (38%)
Combined	320	67 (21%)	142 (44%)	111 (35%)

Distribution of haplotypes between affected and unaffected individuals

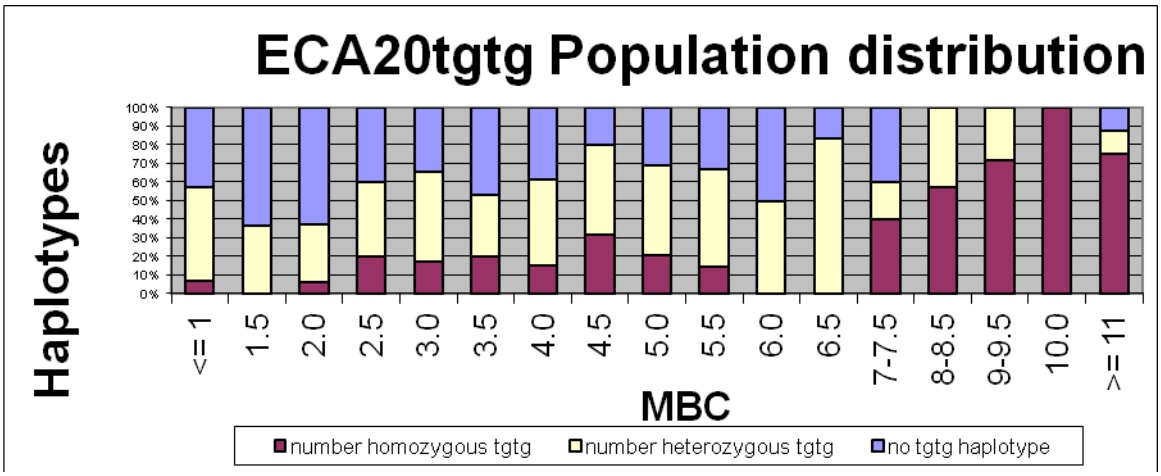


Figure 3.9: TGTG Haplotype Population Distribution Sorted by MBC.
Results are from all animals genotyped. The Y–axis represents the percentage of horses with each pair of haplotypes. The X – axis represents the MBC values.

Haplotype and MBC Related to Age

Extreme hereditary lordosis was the subject of this study, with a focus on early-onset lordosis which is manifested before the horse reaches maturity. To visualize the age of affected horses in the study group as related to haplotype and MBC, figure 3.10 exhibits the individuals grouped by haplotype plotted in a graph with MBC values on the y-axis and age on the x-axis. Interestingly, all of the affected horses under 10 years of age with MBC greater than 7.5cm were homozygous for the associated haplotype.

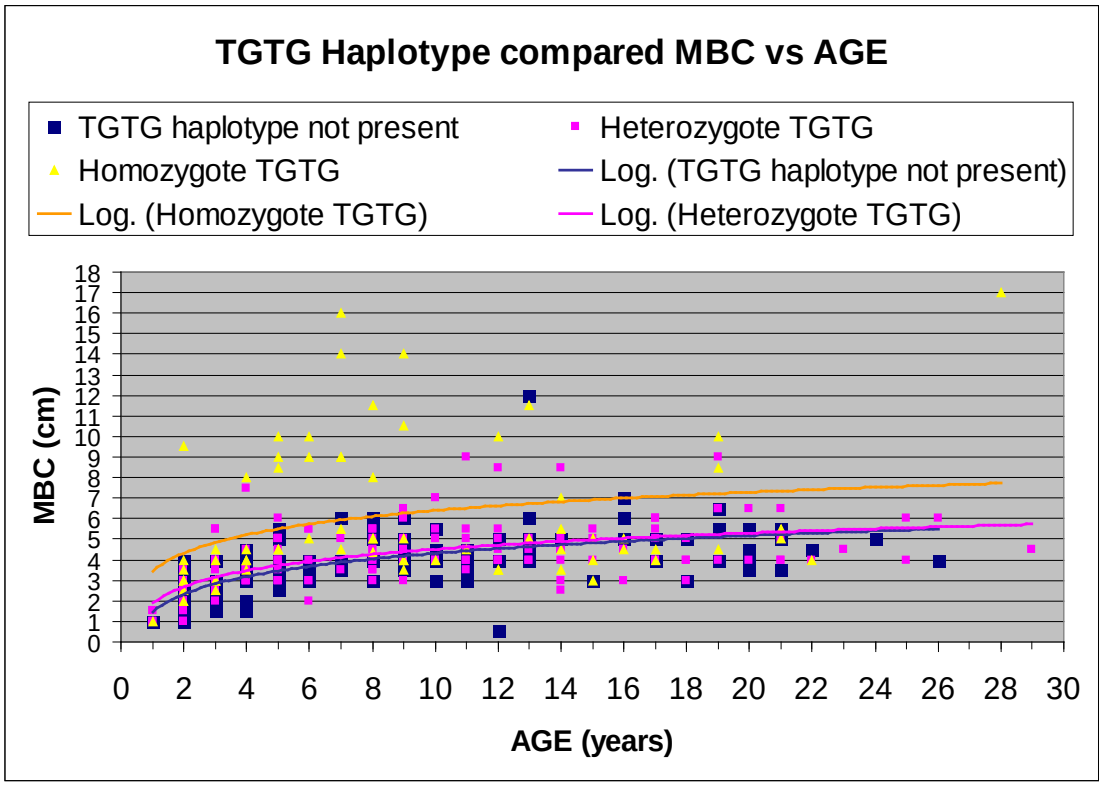


Figure 3.10: Population Distribution MBC vs. AGE.
X-axis = age in years. Y-axis = MBC in centimeters. Interestingly, most of the more extreme MBC values were found in horses less than 10 years of age.

Sequenom Haploview Analysis

Solid Spine Analysis: Haploview analysis of Sequenom data verified associations calculated by PLINK. This method of analysis, searches for a "spine" of strong LD from one marker to the next along each leg of the triangles in the LD chart. Figure 3.11 gives the graphic image of results using the solid spine analysis option. Table 3.6 shows the actual haplotypes, ratios and p-values for each block. Blocks one and two containing a total of 5 SNPs were most significant with p-values < 9E-11.

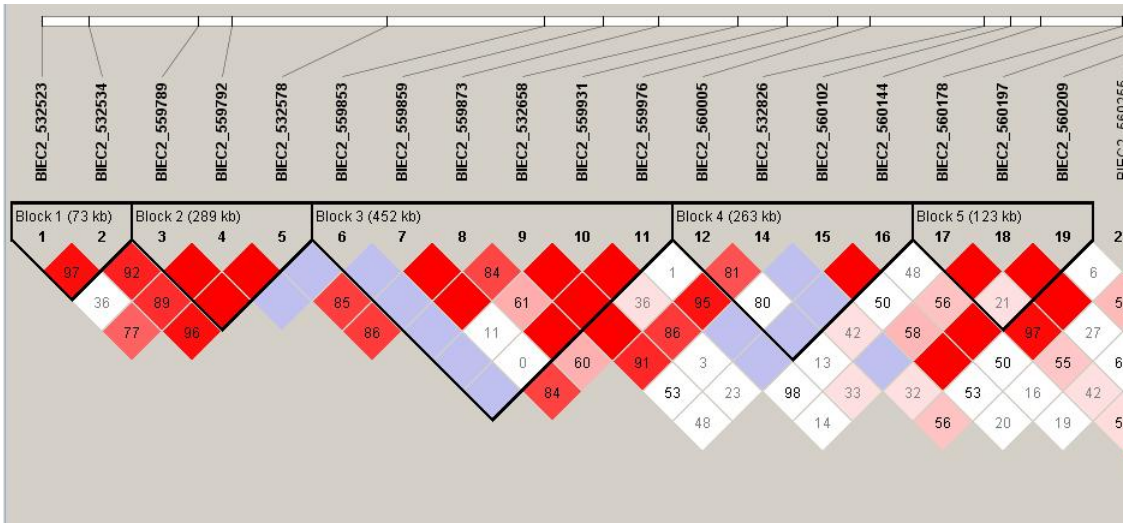


Figure 3.11: Haploview Solid Spine Analysis.
SNP region with LD shown by solid red blocks and recombination spots by blue.

Table 3.6: Solid Spine Analysis Results from Figure 3.11.

Haplotype	Freq.	Case, Control Ratios	Chi Square	p value
Haplotype Associations				
⊟ Block 1				
TG	0.458	56.0 : 10.0, 245.5 : 346.5	44.925	2.0471E-11
CA	0.379	7.0 : 59.0, 242.3 : 349.7	23.269	1.4087E-6
TA	0.158	3.0 : 63.0, 100.7 : 491.3	6.901	0.0086
⊟ Block 2				
AGT	0.442	54.0 : 12.0, 237.0 : 355.0	42.03	8.9878E-11
AAC	0.262	7.0 : 59.0, 165.3 : 426.7	9.206	0.0024
AGC	0.154	2.0 : 64.0, 99.3 : 492.7	8.611	0.0033
GGC	0.142	3.0 : 63.0, 90.4 : 501.6	5.613	0.0178
⊟ Block 3				
TTGGGC	0.582	60.0 : 6.0, 322.8 : 269.2	32.29	1.3276E-8
TTGTGC	0.205	2.0 : 64.0, 132.7 : 459.3	13.709	2.0E-4
TTCTAT	0.064	1.0 : 65.0, 41.0 : 551.0	2.909	0.0881
TACTGC	0.037	1.0 : 65.0, 23.4 : 568.6	0.987	0.3204
TTGTAC	0.036	1.0 : 65.0, 22.7 : 569.3	0.918	0.338
TTCTGC	0.033	0.0 : 66.0, 21.9 : 570.1	2.524	0.1121
CTGGGC	0.016	0.0 : 66.0, 10.8 : 581.2	1.219	0.2696
TTCGGC	0.014	1.0 : 65.0, 8.2 : 583.8	0.0080	0.9276
TACTAC	0.010	0.0 : 66.0, 6.6 : 585.4	0.744	0.3884
⊟ Block 4				
CGAC	0.684	59.9 : 6.1, 390.3 : 201.7	16.99	3.7574E-5
CAAC	0.197	4.1 : 61.9, 125.7 : 466.3	8.502	0.0035
TAAC	0.083	1.9 : 64.1, 52.8 : 539.2	2.802	0.0942
CGGT	0.023	0.0 : 66.0, 15.0 : 577.0	1.711	0.1908
TGAC	0.011	0.1 : 65.9, 7.2 : 584.8	0.664	0.415

Haplotype blocks with frequencies, ratios and p-values.

Four gamete analysis: Alternate 4-gamete analysis of Sequenom data set gave results similar to solid spine analyses. Figure 3.12 shows the haplotype image. Table 3.7 gives frequencies, ratios and p-values. Blocks one and three were most significant. Block one containing 2 SNPS had a p-value of 2.05E-11; whereas block 3 with 6 SNPs had a p-value of 3.38E-11.

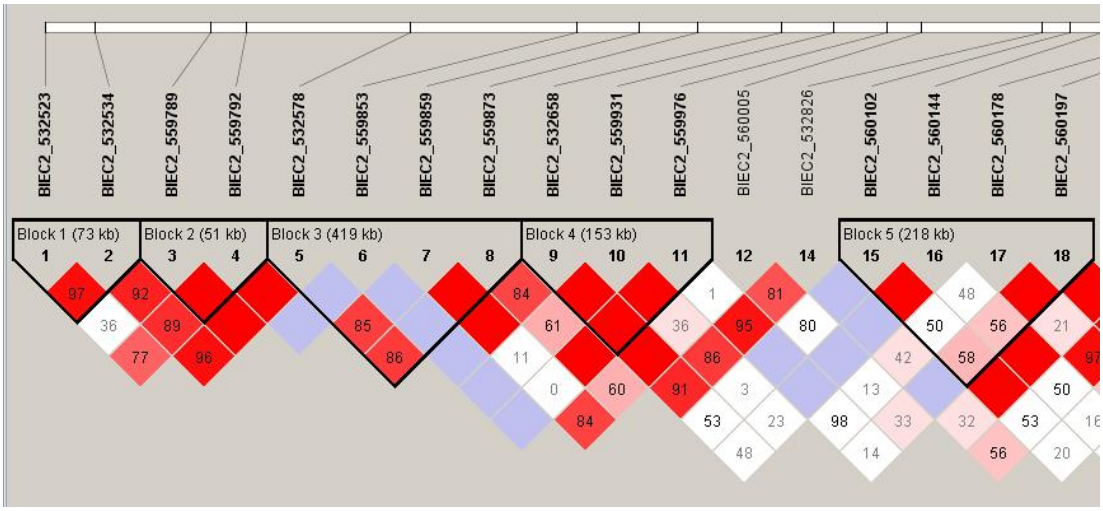


Figure 3.12: Four Gamete Haploview Analysis.
Red blocks indicate LD.

Table 3.7: Statistics from Four Gamete Analysis in Haploview.

Haplotype	Freq.	Case, Control Ratios	Chi Square	p value
Haplotype Associations				
Block 1				
TG	0.458	56.0 : 10.0, 245.5 : 346.5	44.925	2.0471E-11
CA	0.379	7.0 : 59.0, 242.3 : 349.7	23.269	1.4087E-6
TA	0.158	3.0 : 63.0, 100.7 : 491.3	6.901	0.0086
Block 2				
AG	0.598	56.0 : 10.0, 337.2 : 254.8	19.196	1.1793E-5
AA	0.261	7.0 : 59.0, 164.6 : 427.4	9.111	0.0025
GG	0.142	3.0 : 63.0, 90.1 : 501.9	5.576	0.0182
Block 3				
TTTG	0.433	53.9 : 12.1, 231.0 : 361.0	43.946	3.3752E-11
CTTG	0.388	9.1 : 56.9, 246.0 : 346.0	19.254	1.1442E-5
CTTC	0.108	2.0 : 64.0, 69.4 : 522.6	4.713	0.0299
CTAC	0.043	0.9 : 65.1, 27.6 : 564.4	1.515	0.2183
CCTG	0.018	0.0 : 66.0, 12.0 : 580.0	1.363	0.2431
Block 4				
GGC	0.614	61.0 : 5.0, 343.0 : 249.0	29.795	4.8032E-8
TGC	0.275	3.0 : 63.0, 178.0 : 414.0	19.397	1.0619E-5
TAT	0.064	1.0 : 65.0, 41.0 : 551.0	2.909	0.0881
TAC	0.047	1.0 : 65.0, 30.0 : 562.0	1.669	0.1964
Block 5				

Candidate Genes

The SNPs defined an overall region of homozygosity for lordotic horses from 41.5 Mb to 44.5Mb on ECA20. The annotated horse genome at ENSEMBL genome browser (Hubbard *et al.*, 2009) showed that this region contains 53 known and predicted genes. Three genes were selected as possible candidates based on predicted or known function in other species; sequence comparisons for TRERF1 (15 exons), RUNX2 (seven exons, eight introns) and CNPY3 (six exons) between normal and affected horses did not identify SNPs associated with lordosis. Data on SNPs found and their occurrence among the case and control horses are shown in Table S2.

Figure 3.13 shows the region of association on ECA20 as accessed on ENSEMBL genome browser with the location of the statistically significant SNPs labeled. The genes highlighted in blue (*CNPY3*, *TRERF1* and *RUNX2*) were selected for exon screening based on location to associated SNPs and possible roles they play in organismal development. Variations found thus far are in linkage disequilibrium with flanking SNP from the Sequenom assay.

Only one gene, *RUNX2*, in this region had been implicated in skeletal defects based on information from OMIM. From assessment of other likely gene functions,

candidate gene *TRERF1* was identified based on its function as a transcription factor. Another candidate, *RUNX2* has been found to be a scaffold for factors involved in skeletal gene expression (Stein *et al.*, 2004) and it plays a role in osteoblast and chondrocyte differentiation and migration (Fujita *et al.*, (2004). The gene *CNPY3* was selected for exon screening because it contains a trinucleotide repeat in exon 2 and trinucleotide repeat expansions have been identified as playing a role in disease phenotypes in other species. *TRERF1* has 16 exons that were sequenced and compared for 2 normal and 2 lordotic horses. Likewise, *RUNX2* has 9 exons and *CNPY3* has 6 exons that were sequenced and compared. No SNPs were found associated with lordosis in any of the 3 genes screened, compiled results in Table 3.8. Primers used for sequencing of these genes are compiled in table 3.9.

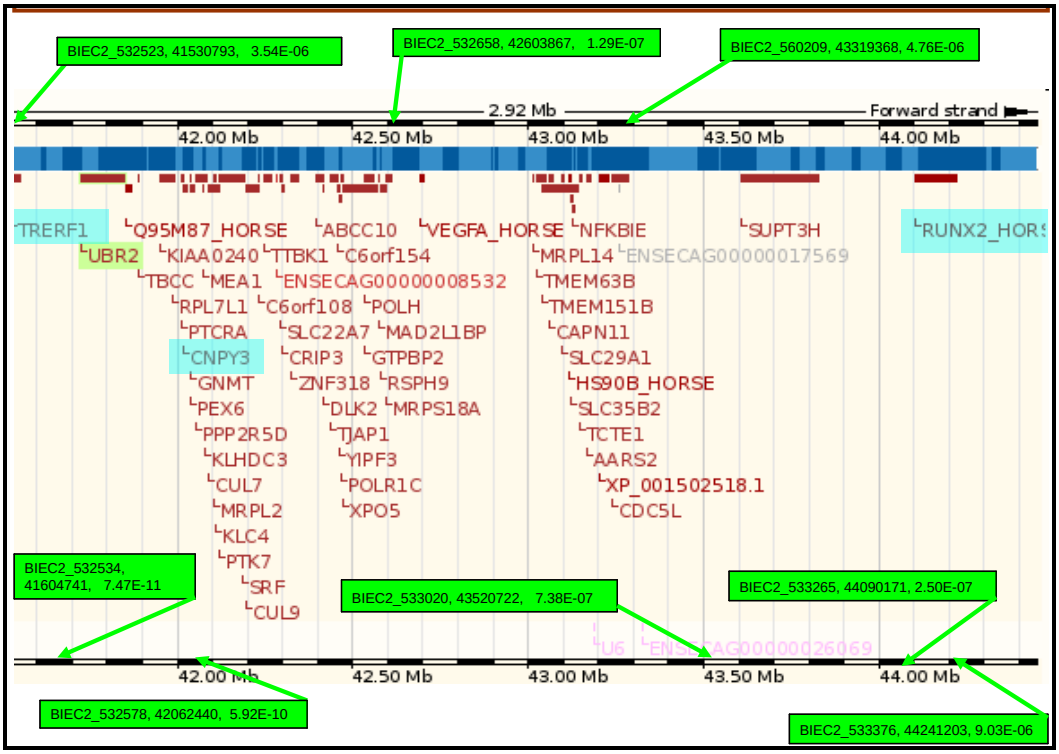


Figure 3.13: Region of Association on ECA 20 as Shown on ENSEMBL Genome Browser.
Genes highlighted in blue had exons sequenced to screen for additional associated variations. Green boxes and arrows identify approximate locations of associated SNPS from table 3.1.

Table 3.8: Results From Sequencing Candidate Genes; Allele Frequencies

Gene	Location	Mutation(q) on EquCab2.0	pp:pq:qq	pp:pq:qq	Conclusion
RUNX2	Intron 1	g.44042908_44042909delAT	11:02:00	11:04:00	No assoc.
RUNX2	Intron 5	g.44188955delT	11:6:0	25:5:0	No assoc.
TRERF1	Exon3	g.41474935_41474936insCAG	30:16:05	5:01:00	No assoc.
CNPY2		None found			No evidence

Allele frequencies for variations found in candidate genes *RUNX2*, *TRERF1* and *CNPY2* in lordotic and normal horses to investigate association of this gene with lordosis. “p” represents the number of horses found with the reference type. “q” represents the number of horses found with the mutation.

Table 3.9: Sequencing Primers for Candidate Genes

Gene	ID	Forward Primer	Reverse Primer	BP begin	BP end
RUNX2	e 1	5’-cacttcgctaacttgtggctg -3’	5’-cgggtgatctcgagaaagag-3’	44102849	44103416
RUNX2	i 1-2	5’-tgttactcaaaggaaagacaccag-3’	5’-cctttccacgttgataaatatgtg-3’	44042741	44043285
RUNX2	i 1-2	5’-agaggagtccaagtcaattagagc-3’	5’-ttaccagaacacacggaatgtaac-3’	44104971	44105504
RUNX2	e 2	5’-attcttggtttttaagctttgctg-3’	5’-aaacactcaaattcatctggacac-3’	44111870	44112245
RUNX2	i 2-3	5’-tgacacccttagttgtctaacgag-3’	5’-attagaagtcaccaatgcttctcc-3’	44114984	44115524
RUNX2	e 3	5’-tctcatttagaataaggggtcctg-3’	5’-ggctaccttatctgggatacattc-3’	44117689	44118023
RUNX2	i 3-4	5’-ttgaatgaatgaaacgtaagcaag-3’	5’-gcttctttgtacagagctttcctc-3’	44142759	44143292
RUNX2	e 4	5’-agatgatgcttatgaagcagtttg-3’	5’-gtgcaagtgaaacctatgagtctg-3’	44169835	44170215
RUNX2	i 4-5	5’-agatcaagtgtggaaagttcattg-3’	5’-tcttccctatcaagaagaatggtc-3’	44178502	44179044
RUNX2	e 5	5’-tctagaaagctttgtgctatgcag-3’	5’-atgagagtgggtttccagttaaag-3’	44188873	44189238
RUNX2	i 5-6	5’-aagtagacccacagagacacacac-3’	5’-tttgcctaaaggaggtagtcattc-3’	44190086	44190643
RUNX2	i 5-6	5’-tgtgctactccctctttacctgag-3’	5’-gggctttacatacattaaccatcc-3’	44201754	44202304
RUNX2	e 6	5’-gtttgtcatcttgaaagtgtttgc-3’	5’-aaaatgggacagtaccaatcagac-3’	44218256	44218605
RUNX2	i 6-7	5’-aaagcaaagatacgagatccaaag-3’	5’-gaaattctcctgagggtcataaac-3’	44218899	44219446
RUNX2	e 7	5’-tgggaagctaaagttttcttcttt-3’	5’-atatattgatacacttgggacgtg-3’	44220029	44220663
CNPY3	e 1	5’-gttgccataggtcgctgaggacaccatcc-3’	5’-gctaactcctaaccaccctctgggcttc-3’	42019587	42020424
CNPY3	e 2	5’-gaaatgacgtgagtaggatgagg-3’	5’-tgaagctctaaggaacaaagcag-3’	42023229	42023634
CNPY3	e 3	5’-aaaaattccttcccatagagctg-3’	5’-atcttccaatgacacccatatca-3’	42024053	42024453
CNPY3	e 4	5’-aaggtgaggtcaagtgtgaagag-3’	5’-tgattctcctgaggaaactgaag-3’	42025686	42026087
CNPY3	e 5	5’-cagtttcctcaggagaatcacag-3’	5’-ttgtcgcaggttattctgtacctg-3’	42026068	42026498

Table 3.9: cont. **Gene**	**ID**	**Forward Primer**	**Reverse Primer**	**BP begin**	**BP end**
CNPY3	e 6	5’-ttaagggtctctcacttcctcac-3’	5’-aggagagaaacgctcagcag-3’	42026614	42027070
TRERF1	e 1	5’-acgtgatctgttctcatgttgaag-3’	5’-ttctccagaggtgtctaattctctc-3’	41554583	41554104
TRERF1	e 2	5’-cagtgacatgagagtgggtgag-3’	5’-gtgcactttctaagtctacctatgc-3’	41501808	41501430
TRERF1	e 3a	5’-aaatactttcctgtcctttgcatc-3’	5’-cacgtagctggacatggtttc-3’	41476107	41475520
TRERF1	e 3b	5’-ttccctcaagatactcgagacg-3’	5’-catttcttgcatcgacatacg-3’	41475646	41474951
TRERF1	e 3c	5’-cagtattacccacagcagcaac-3’	5’-ccctgtagcacactaagacacc-3’	41475013	41474324
TRERF1	e 4	5’-gaagtctcaggtgaggaagagc-3’	5’-taggcacgtggtaatcctgtc-3’	41473091	41472641
TRERF1	e5	5’-tctgaagggtattggagtgatg-3’	5’-aatggaagtgggactgtttgac-3’	41472108	41471624
TRERF1	e 6	5’-acctttctggaaaagcaaagtg-3’	5’-aagagctgtcctcatctccaag-3’	41470844	41470388
TRERF1	e 7	5’-aaagcccaggtatctgaatcg-3’	5’-taagtatttccagcaccagcac-3’	41467098	41466515
TRERF1	e 8	5’-agctccttggagactagtggtg-3’	5’-gtggttaagttggtgtgctctg-3’	41466120	41465719
TRERF1	e9&10	5’-tcttaggatcacgctgtgagg-3’	5’-aaacctgggaaaacagtttcag-3’	41464789	41464197
TRERF1	e 11	5’-agaggaggccccactatctc-3’	5’-aagcagcaattaggacagaagc-3’	41463180	41462699
TRERF1	e 12	5’-ttcagggctgctaaccattatac-3’	5’-ccgatgtcagataccatccttc-3’	41454845	41454445
TRERF1	e 13	5’-ggagtgagaactgaatggatcg-3’	5’-ctctccatagcaggaagtcagg-3’	41452870	41452410
TRERF1	e 14	5’-atgcagaaagaaagcagtgcag-3’	5’-acggagtgaaagaaggaaaacc-3’	41446860	41446362
TRERF1	e 15	5’-tgccctgtcgttagactgtttc-3’	5’-agaccccaaggatggtatgag-3’	41444066	41443606
TRERF1	e16a	5’-caaaccagatgtcagaccacac-3’	5’-aatgtagcccagatgaaacacc-3’	41440537	41439849
TRERF1	e16b	5’-gtggtcgttttcttgcagtttag-3’	5’-acctttttcaaatgacgtgtgtg-3’	41439995	41439529

ID column, e = exon and i = intron.

Discussion

The distribution of MBC measurements in this study confirmed the results from the earlier study that described a normal distribution of the MBC phenotype, with 5% falling two standard deviations above the mean (Gallagher *et al.,* 2003). In that previous study, the mean MBC was 4.05 cm, while the mean found in this study was smaller, 3.6 cm. Differences in age of horses may account for this discrepancy, because the previous study showed a positive correlation for age and MBC; the mean age in the first study was 7.8 years, whereas the mean age in this study was 7.1 years.

The WGA study demonstrated the presence of a recessive gene responsible for the lordotic trait in horses. This was suggested in the initial WGA study with 40 horses and the Illumina SNP chip (corrected $P = 0.017$). The association was confirmed in a subsequent study with a second set of affected horses using SNPs from this targeted area ($P = 0.036$). This haplotype spanned 1 073 074 bases and harbored 53 known and predicted genes.

The TGTG haplotype was the most common and suggests that the haplotype occurred in the breed prior to the mutation causing extreme lordosis. If a gene present in this haplotype was indeed mutated to cause lordosis, then only knowing the specific mutation would allow us to distinguish between these haplotypes. Of course, mutations which subsequently occurred within the lordosis-causing haplotype might allow us to use tests for other markers to identify haplotypes completely associated with extreme lordosis.

The high frequency of this haplotype and this phenotype among ASB horses might be the consequence of selection by breeders. If a single copy of the gene produced a desirable phenotypic effect, such as improved gait, selection for that trait may negate selection against lordosis and result in a net increase in the frequency of the gene in the breed. Comparisons of gene frequencies for the TGTG haplotype among horses with different performance phenotypes will be needed to answer this question.

While the high homozygosity of this haplotype among the lordotic horses strongly suggests the presence of a recessive gene for the trait, not all lordotic horses were homozygous for the region. We found five different haplotypes among the lordotic horses, and 30% were not homozygous for the haplotype associated with the recessive

lordotic condition. The lordotic phenotype may have multiple possible causes, of which the hereditary recessive condition is only one. While 70% of the lordotic horses were homozygous for this haplotype, we observed seven (21%) lordotic horses that were heterozygous and three (9%) that did not have the haplotype at all (Table 4). There may be multiple causes of extreme lordosis among Saddlebred horses, and the recessive gene suggested by this study may be only one of them. Other genes, accidents affecting skeletal integrity or even management practices may lead to extreme lordosis in the absence of the recessive gene implicated by this study. Nevertheless, considering the high prevalence of this haplotype among affected horses, this hereditary recessive condition is probably the most common cause of lordosis among Saddlebred horses.

Increased homozygosity of this region and frequency of the haplotype in the population raised a new question: Is this haplotype frequency a result of transmission distortion (TD) and not just breeder selectivity? TD, also known as segregation distortion, is a preferentially transmitted haplotype that shows a separation from Mendel's law of independent segregation for alleles. A haplotype exhibiting TD in mice, known as the *t* haplotype, was first recognized in 1927 (Dobrovolskaia-Zavadskaia, 1927). In mice this haplotype results from a chromosomal inversion including the mouse *major histocompability complex* (*MHC*) on mouse chromosome 17 (Shin *et al.,* 1982, Silver 1982). The transmission ratio for some of these *TD* alleles can exceed 95% in mice (Braden, 1971). TD has been investigated in an extended American Standardbred family of horses involving the gene *A10* in the MHC. Significant TD was found only within that family and it was theorized *A10* is not the cause of the TD but is linked to the gene responsible in that family (Bailey, 1986). In humans, statistically significant evidence for TD was noted in male parents of European ancestry for chromosome 6p in the region containing the genes SUPT3H and *RUNX2* (Santos *et al.,* 2009). The synteny this region shares with the region of interest in Saddlebreds may warrant a new look at the haplotype associated with extreme lordosis to rule out transmission distortion as a potential cause for its high frequency.

As noted above, the lordosis region contains 53 known or predicted genes. Three of these genes; *RUNX2*, *TRERF1,* and *CNPY3* were selected for exonic sequencing in the hope of identifying a causative variation. From assessment of other likely gene functions,

the candidate gene *TRERF1* was identified based on its function as a transcription factor. *CNPY3*, a trinucleotide repeat-containing gene, was also considered, because repeat expansion has been shown to play a role in various diseases. Spinocerebellar ataxia type 1 is caused by such a repeat expansion within *ataxin 1* gene on human 6p (Kameya *et al.*, 1994). However, exon sequencing of the three candidate genes did not identify SNPs or other genetic variations associated with the trait.

Though no associated variations were identified during the course of this project, *RUNX2* could still be considered a viable candidate. Only exonic sequence based on Ensembl was analyzed. *RUNX2* has two isoforms that have different N-termini and are controlled by two different promoters (Komori and Kishimoto, 1998). Further sequencing is necessary to ensure both isoforms of the gene were completely covered and of the known promoter regions were adequately represented. *RUNX2* is a transcription factor with many functions. It is a scaffold for factors involved in skeletal gene expression (Stein *et al.*, 2004) and has a role in osteoblast and chondrocyte differentiation and migration (Fujita *et al.*, 2004). It is highly expressed in immature osteoblasts, but expression must be down regulated for osteoblasts to mature properly (Maruyama *et al.*, 2007).

Another gene from this region which could be a plausible candidate is *CUL7*, a gene implicated in skeletal development and distally closer than *RUNX2* to the most highly associated SNPs in this study. Short stature syndrome of the Yakut people is attributed to a variation in *CUL7*, a homozygous T insertion at position 4582 in exon 25. Two physiological manifestations of this variant in the Yakut are hyper lordosis of the lumbar region and hypoplasia of the 13th rib (Maksimova *et al.*, 2007). Other *CUL7* variations are responsible for 3M syndrome. There is genetic heterogeneity in 3M syndrome after a large-scale mutation search. The distinguishing features of 3M are growth retardation, long slender tubular bones, and tall vertebral bodies (Huber *et al.*, 2009).

Rooney & Robertson (1996) distinguished between senile lordosis and congenital lordosis in horses. Senile lordosis was a consequence of aging. Congenital lordosis occurred as a consequence of hypoplasia of the articular facets of thoracic vertebrae and followed birth as a result of weight bearing (Rooney & Pickett, 1967). We believe that

the condition we have been investigating in ASB horses is the congenital form because most of the affected horses were under the age of 10. However, this should be confirmed by sequential measurement of MBC in horses of different ages, concentrating especially on young horses, to determine the progression of lordosis. This could be accomplished with a cohort study, following horses from weaning to 2 years, measuring MBC every 30 days. To facilitate future investigations, a better physiological characterization needs to be developed; with photos, measurements and radiographs of the spinal facets, noting the vertebrae involved. This would help better understand the degree of hypoplasia involved with congenital lordosis.

Discovery of the genetic basis for extreme lordosis is a goal that remains elusive. In connection with this project, exons of several candidate genes were sequenced. However, the cause of the trait could be the result of changes in gene expression which would not be uncovered by exon sequencing. As we learn more about the genome of animals, we realize that even the introns and the DNA between genes can play a role in gene regulation. The region of interest might be reduced by further studies using additional genetic markers from this region, including more SNPs, microsatellites or other genetic polymorphisms. Deep resequencing of the genes within the target area is a valid next step toward identifying rare variants associated with lordosis. Even when an associated variant is located, functional studies may be necessary to elucidate the role the variant plays, especially a variant not within a gene coding region or known activator/inhibitor/enhancer site. This was the case when two of the SNPs associated with increased risk for coronary artery disease (CAD) and myocardial infarction (MI) to 9p21.3, a barren, gene free region (Samani *et al., 2007*; Schunkert *et al., 2008, 2011*). It turned out the CAD risk alleles were located in a STAT1 binding site identified during a study that identified enhancers in this genomic region (Harismendy *et al.,* 2011).

Another approach to understanding this condition may be to investigate differences in gene expression between affected and unaffected horses. Discovery of a gene which shows differential expression would help to focus this work. However, the choice of tissues and the age at which horses are tested may be factors which confound such an approach. This could be accomplished with as few as two full sibling weanlings, one exhibiting the lordotic phenotype and one normal. One would sample tissue from the

developing thoracic spinal facets from the two. This tissue could then be used in gene expression analysis. Data could be compared among genes known to interact within the same developmental pathway for bone development.

More insight into the variation responsible for early-onset extreme lordosis in horses may be beneficial for studies of human juvenile kyphosis and juvenile idiopathic scoliosis (IS). As with the horse, these two congenital conditions exhibit an early age of onset. Familial IS only accounts for 10% of all cases in humans, while 90% appear to be sporadic with unknown or environmental aetiological factors (Cheng *et al.*, 2007). Through familial linkage analysis, candidate regions for IS susceptibility have been identified on human chromosomes 6p, 10q and 18q (Wise *et al.*, 2000). More recently, regions on 6, 9, 16 and 17 were identified through genome-wide screening (Miller *et al.*, 2005). It is of particular interest to note that the segment on HSA6 implicated in the Wise paper is syntenic with the region on ECA20 that was found in this study to be associated with extreme lordosis.

CHAPTER FOUR: Discussion of the Champagne and Lordosis Genetic Studies

Introduction

Chapters 2 and 3 describe studies designed to better understand the genetics of two traits in horses, namely champagne dilution and extreme lordosis. Champagne dilution was well known to be the consequence of a Mendelian dominant gene. The task described here was to identify the chromosome region harboring the gene, locate the gene responsible and identify the likely mutation. In contrast, extreme lordosis of American Saddlebred horses was thought to have a hereditary basis but this was not well established. The results of this study confirmed that the trait did have a strong hereditary component, localized a major gene responsible for the trait to chromosome 20, suggested the gene acted with a recessive mode of inheritance, but did not lead to identification of the gene responsible.

These studies were conducted in a conventional fashion, using the best genomic tools available at the time, and the results were classic. Linkage was found in family studies for champagne dilution and association was found with the GWAS study for extreme lordosis. In both cases the underlying hypothesis being tested was that a single major gene was responsible for the trait being studied. However, had the genetics of these traits been different we would have obtained different results. The question arises, how might the results appear under different genetic models and how might they have been interpreted?

Impact of Proper Phenotyping

Champagne Dilution

The *Champagne dilution* gene dilutes both black and red hair pigment. The effect of the *Champagne dilution* gene on red pigment is similar to the effects of the *Cream dilution* gene on red pigment. Chestnuts with the *Champagne* gene are termed Gold champagne but appear very similar to chestnuts with the *Cream dilution* (*CR*) gene (palominos). Had we not known of the difference and pooled data from the two types of horses, the results of our study would not have been very clear. Linkage mapping is based on combined results from testing families. Linkage identified between phenotype and a marker with a calculated LOD score of 3 or greater is considered to be statistically significant. Linkage of a genetic marker with *Champagne* may be identified in one family but not other families. If the gene segregating with the dilution phenotype had actually been *Cream* and not *Champagne* in one of the families, then the data for that family would be in conflict with the data from other families. For example: What if three families, segregating for a hair color dilution gene were genotyped and two of the families had the *CH* gene while the third did not? In this case, the LOD the score for family-1 could be 3.4, family-2 could be 2.7 and family-3 could be -2.6. The score for family-1 is statistically significant since it is greater than 3.0. The score for family-2 is not statistically significant by itself but can be combined with that of family-1 to give a total LOD score of 6.1. , The score from family-3 raises questions because it is less than -2.0 and therefore evidence against linkage. If we add -2.6 to the other two scores, we still have evidence for linkage to those markers because the score is +3.5. However examination of the data from each family would alert us that the genetic basis for dilution in Family-3 lies elsewhere. The ability to detect genetic heterogeneity is one of the strengths of linkage mapping.

It is possible for a horse to possess one or two copies of both *CH* and *CR* simultaneously, because they reside on different chromosomes. A horse possessing both dilutions is called an Ivory Cream and would appear as pale gold. Genotyping a family with a sire possessing both *CH* and *CR* would have proven confounding and inconclusive. Some dilute offspring would segregate with the sire's *CR* allele and some would segregate with the *CH*. With the phenotype considered the same, statistical significance would not be easily attained.

Fortuitous selection of a single true Champagne family would have permitted discovery of linkage and we could have continued the study and identified the gene and mutation as described in Chapter 2. However, when we began to use the test for the gene on the general population we would have found many horses exhibiting a dilution phenotype without having the *Champagne* gene. Every horse with the *Champagne* gene would exhibit dilution so we would know that the results were correct, except there was another genetic basis for dilution.

Examples of Similar Phenotypes Caused by Different Genes

Splashed White

Horses with the splashed white phenotype have blue eyes, and smooth, clearly defined white markings of the legs underbelly and face. They have the appearance of being dipped feet first in white paint. Three different variants in two different genes have been found to be responsible for *Splashed White (SW)* to date: *SW1*, *SW2* and *SW3*. When the first variant responsible was identified, it did not account for all individuals exhibiting the splashed white phenotype. Upon further investigation, two other genetic variants were found responsible for the remaining cases of splashed white. *SW1* is caused by an 11 SNP insertion in *microphthalmia-associated transcription factor* (*MITF*) *promoter 1*; *SW2* is caused by a SNP resulting in an amino acid change in *paired box domain gene family 3* (*PAX3*); and *SW3* is the result of a 5 base pair deletion in exon 5 of *MITF* causing a frame shift and premature stop codon resulting in a truncated protein. These three loci do not account for all the splashed white phenotypes, leaving more to be discovered (Hauswirth et al 2012). .

Sabino (White Spotting Pattern)

Sabino is a white spotting pattern in horses with white patches on the legs belly and face and some white hairs mixed in with the base colored hairs in the midsection of the body. The variant responsible for Sabino in Tennessee Walking Horses and other light horse breeds was identified as a single intronic base change that resulted in skipping of exon 17 in the *KIT* gene. This particular variant was designated *Sabino1 (Sb1*), because it did not explain all the Sabino phenotypes, such as the one possessed by Clydesdale horses (Brooks *et al*., 2005). Presumably a different genetic variant is responsible for sabino color patterns in Clydesdale horses.

Junctional Epidermolysis Bullosa (JEB)

JEB is an autosomal recessive trait that affects the integrity of the skin and mucosa and causes neonatal blistering and skin fragility. Candidate genes for the condition were *LAMA3*, *LAMB3* and *LAMC2* genes which make up the three glycoprotein subunits of Laminin5, a basement membrane protein required for normal function of the dermal-epidermal junction. A defect in any one of these genes could have resulted in JEB, affecting Belgian horses, other draft breeds and American Saddlebred Horses (Frame et al 1988; Kohn et al. 1989; Lieto *et al*., 2002; Milenokovic *et al*., 2003). The variant responsible for JEB in Belgian and other draft horse was a single base insertion that introduced a premature stop codon in the *LAMC2* gene which is located on chromosome 5 (Spirito *et al*., 2002). However, this mutation did not explain JEB for American Saddlebred horses. Subsequent studies revealed that 6589 bp deletion spanning exons 24-27 of the LAMA3 gene on chromosome 8 was the cause of JEB in American Saddlebred Horses. A random sample set from all cases of JEB, irrespective of breed might not provide for an informative GWAS. This situation highlights the value of conducting genetic studies along breed lines.

Impact of Different Mutations at the Same Locus (Genetic Heterogeneity)

Champagne Dilution

Champagne dilution is the result of one variation within a single gene that accounts for all cases of Champagne across all the breeds tested. Some genetic investigations do not end so concisely. Some lead to the discovery that a single variant within that gene does not explain all cases. Different variations in the same gene may be responsible for a single phenotype between different families, ethnicities, geographic populations and/or breeds. This has been noted in a 1989 review of human research of Cystic Fibrosis (CF). About 70% of the cases of CF are caused by one particular mutation and the remainder of the cases caused by almost 230 other variable mutations within the same gene (Kerem *et al*., 1989). Early characterization of the gene responsible for Duchenne's muscular dystrophy found multiple deletion events to be responsible for about half of the cases, with other cases caused by other types of genetic variation within the same gene (Koenig *et al*., 1987).

KIT Heterogeneity

Heterogeneity within a single gene is seen in horses as well. The *KIT* gene has been implicated in many white spotting patterns in mammals. *Dominant White* (*DW*) in horses has been found to be the result of several different spontaneous variations which arise in the *KIT* gene. A single mutation was found responsible for dominant white horses found in the Franches-Montagnes breed, while another variation was found to cause *DW* in a family of Arabian horses. For Camarillo White horses yet another mutation in *KIT* was found to be responsible (Haase *et al.* 2007). Since then even more variants have been found to be responsible for other cases of Dominant White (Haase et al. 2009, Haase et al. 2010). These studies were based on testing individual families. If multiple causative variants were found in that one gene locus for different families or breeds of horse during the course of sequencing the exons of *SLC36A1* for *Champagne*, then it could have been concluded the basis of *Champagne* was heterogeneous.

Aggrecan Heterogeneity

Similar results came from investigation of the *Aggrecan* gene when looking for the cause of dwarfism in miniature horses. A GWAS revealed a point of statistical significance on chromosome 1 showing a region where *Aggrecan (ACAN*) resided as responsible for dwarfism. Based on sequencing *ACAN* for dwarf horses it was discovered that there were at least 4 different recessive variants within the *ACAN* gene responsible for dwarfism (Eberth, 2013). The mutations destroy or alter the function of the *ACAN* gene and any combination of these 4 mutations (in other words, absence of the normal allele) result in a dwarf offspring or fetal/embryonic abortion.

GWAS as a Tool to Find Champagne

If families had been unavailable we could have used the SNP chip and we may have obtained similar results. As long as the sample dilute horses were correctly phenotyped as Champagne, the results would be clear and significant. Several hundred SNPs from equine chromosome 14 are represented on the SNP chip. Some of those would have shown statistically significant linkage to Champagne upon analysis. We would have then proceeded to select candidate genes for sequencing and found the mutation responsible. However, if we had grouped all the Cream and Champagne horses as dilute, and looked for association, the study would have been less conclusive. The number of horses tested would have been small and the genetic associations for *CR* and *CH* might not have been strong enough to achieve statistical significance for either genetic location. Considering the strong evidence that the dilution exhibited a clear Mendelian mode of inheritance, this would have been difficult to explain. The explanation would have to be: 1) The SNPs under study were not sufficiently close to the causative gene for detection. If the gene fell among those which were not mapped and annotated, relegated to chromosome UN, then this might occur. 2) There was genetic complexity confounding the simple model we were testing. Testing a larger number of horses might have provided statistical significance to genotyping analysis results for *CH* and *CR*, but without correct phenotyping the results would be difficult to decipher. Fortunately, for our actual study we could distinguish between *CR* and *CH* and our results were clear, significant and conclusive.

Impact of Unknown Mode of Inheritance

Extreme Lordosis

Extreme lordosis is often associated with old horses whose backs sag due to weakness, overloading or overuse. However there is another form of extreme lordosis found in many breeds in which the condition appears at birth or within the first two years (Rooney & Pickett 1967; Rooney, 1969). Early onset extreme lordosis, the focus of our study, occurs about 1% of the time for most breeds, but in the American Saddlebred an occurrence rate of nearly 4% was noted, giving rise to the hypothesis that extreme lordosis has a hereditary component in this breed (Gallagher *et al*., 2003). At the time we undertook the investigation its mode of inheritance was not known. It could have been dominant with incomplete penetrance, recessive or polygenic with a complex inheritance pattern involving multiple genes of relatively small contribution across the genome.

The lordosis project was different from the search for *Champagne*: the mode of inheritance was unknown and adequate family samples for a linkage study could not be obtained. Many breeders considered it a defect and avoid matings which produce the trait, thus making them less likely to create families for linkage study. The new equine SNP chip had just become available when this project began and sufficient population samples were available for a case vs. control genome wide association study (GWAS). The data collected identified statistically significant association with a genetic location on equine chromosome 20 and indicated a recessive mode of inheritance (Cook *et al*., 2010) If this significant association had not been found, we would have looked into it as having potentially polygenic inheritance. Dominance would not have been considered unless there had been statistical association with a single SNP genotype with little or no homozygosity among the samples. Complex inheritance is responsible for many measureable traits studied in livestock such as milk production in dairy cattle and growth rates in beef cattle. A good review of dairy cattle QTL studies that collated the data across several mapping projects identified two consensus regions on bovine chromosome 6 that affect milk yield (Raadsma, 2004). In beef cattle, QTLs have been mapped in association for growth rate and other meat traits (Kim et al., 2003Nkrumah et al., 2007; Sherman et al., 2008; Gutiérrez et al., 2009; Snelling et al., 2010), QTL investigations of Osteochondrosis

dissecans (OCD) in horses have been undertaken with GWAS using mainly microsatellites. To date, fourteen quantitatitive trait loci have been found to play a role in OCD, but only a few match locations between the different breeds of horse (Distl, 2013). If no single genetic location had been found responsible for lordosis, multiple genetic locations with possible additive effects would have been included and potential heritability would have been calculated for each gene or region.

Complex Mode of Inheritance: Size Variation in the Horse

The many modern breeds of horse encompass a large amount of size variation, from the tiniest of miniature horses at less than three feet tall to the largest of draft breeds that can top out over 6 feet tall at the withers. In a search for genetic components involved in size variation, a GWAS was completed with 48 horses across 16 breeds. All horses had 33 measurements taken of the head, neck, trunk and limbs. Four loci were identified on chromosomes 3, 6, 9 and 11 that together explain 83% of the size variance between these 48 horses (Makvandi-Nejad, 2012).

Impact of Family Availability

If families had been available for use with the SNP chip we may have been able to identify the trait through linkage studies.. Additional family members could have proven helpful in defining borders of the associated homozygous region by exposing possible breaks between the associated haplotype and the location where the gene resides. The American Saddlebreds overall, show a strong incidence close interbreeding, where highly favored sires are found only 4 and 5 generations back on both the sire and dam side of the pedigree for many horses. The lack of genetic diversity in the region would possibly only prove to be confounding in more closely related individuals with the still limited number of markers used.

Current Scientific Tools and Advances

Lavender Foal Syndrome

Lavender foal syndrome is a lethal color dilution in Egyptian Arabian horses with an autosomal recessive mode of inheritance. Lavender's phenotype had strong parallels with two other mutations previously seen in mice one in *Ras-associated protein* (*RAB27A*) and the other in *myosin Va* (*MYO5A*). The genetic location for *Lavender foal* was then identified using the Equine SNP50 Illumina tool to identify a region of homozygosity flanking the *MYO5A* gene (Brooks *et al* 2010). Since no such genetic signature was found for *RAB27A*, this was eliminated as a candidate gene.

Leopard Complex

Leopard complex (LP) is characterized as different white spotting patterns in horses some involve symmetrical patches of white centered over the hips or complete white body and may or may not have pigmented oval spots in the white patterned area (Sponenberg, 1990). *LP* was first mapped to ECA1 with microsatellites in close proximity to *Transient Receptor Potential Cation Channel Subfamily M, Member 1* (*TRPM1*) (Terry 2004). Significant down regulation of TRPM1 was detected in the skin and retina of *LP* homozygotes using qRT-PCR analysis (Bellone *et al*., 2008). Subsequent fine mapping and mutation analysis identified one non-coding SNP that was highly associated with *LP*, but it did not show complete LD with *LP* (Bellone *et al*., 2010). Recently a 1378 bp retroviral insertion was found in intron1 of *TRPM1* that causes a premature poly-adenylation sight. This insertion was identified through novel RNA reads from intron 1 of *TRPM1* found in LP/LP horse skin and retina that had no expression of normal *TRPM1* (Bellone *et al*., 2013).

Lordosis

Following the discovery of association of early onset extreme lordosis with a region on ECA1, several candidate genes (*CNPY3*, *TRERF1* and *RUNX2*) were identified and sequenced. Sequencing was limited to exons with some partial sequencing of introns and the 5' and 3' untranslated regions. In these cases the hypotheses was that the variant responsible would be a SNP within an exon which changes the translated product of that gene in some way. But the variant responsible for lordosis could be any one of the genetic variations discussed in chapter 1 of this work which qualitatively influence expression. The variant may not even be found within the affected gene, but may fall within a regulatory site such as a promoter binding site, inhibitor binding site or splice recognition site. Since many genes have interactions with other genes down and around many different pathways and processes, several genes and their products could be affected. Lordosis may be caused by inadequate development of vertebral facets and. The lordosis gene may influence traits which are being selected for by breeders, such as a higher head carriage, more elevated gate or other specific aspect not yet noted which increase the incidence of lordosis in this breed. Perhaps the genes involved are important for skeletal or cartilage growth and maturation, since the phenotype develops during the young horses growth and maturation. Growth and maturation are under strict control during development giving rise to the idea that it could be caused by a variant involving a regulatory element. The large size of the homozygous region and density of genes increases the difficulty in locating the variation responsible. Some of the horses exhibiting extreme lordosis only possessed one copy of the haplotype of interest or did not possess the haplotype at all. The gene responsible may flank the homozygous region and separated from the haplotype during a rare recombination event.

Whole genome sequencing could provide additional info concerning the developmental pathway and genes involved. Since the first complete equine sequence was assembled seven years ago, newer sequencing technologies have been developed. Whole genome sequencing is now more affordable and could be used to find the gene responsible for lordosis. One approach to discovering the mutation responsible for the trait would be the following. A small sample selection of four individuals could be sequenced and compared: one affected homozygote for the implicated haplotype, one unaffected

homozygote, one affected without the haplotype and one unaffected without the haplotype. Closely related individuals could be most informative, but unrelated horses would suffice. The resulting data might lead to identification of the variant responsible for extreme lordosis and could highlight variant's responsible for others pleotropic traits that are selected for by breeders.

REFERENCES

Adalsteinsson, S. (1974). Inheritance of the palomino color in Icelandic horses. *Journal of heredity*, *65*(1), 15-20.

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., ... & Brokstein, P. (2000). The genome sequence of Drosophila melanogaster. *Science*, *287*(5461), 2185-2195.

Adelson, D. L., Raison, J. M., Garber, M., & Edgar, R. C. (2010). Interspersed repeats in the horse (Equus caballus); spatial correlations highlight conserved chromosomal domains. *Animal genetics*, *41*(s2), 91-99. doi:10.1111/j.1365-2052.2010.02115.x.

Agarwal, A., Guindo, A., Cissoko, Y., Taylor, J. G., Coulibaly, D., Koné, A., Kayentao K., Djimde A., Plowe, C. V., Doumbo O., Wellems T. E. & Diallo, D. (2000). Hemoglobin C associated with protection from severe malaria in the Dogon of Mali, a West African population with a low prevalence of hemoglobin S. *Blood*, 96(7), 2358-2363.

Allen, W. R., Skidmore, J. A., Stewart, F., & Antczak, D. F. (1993). Effects of fetal genotype and uterine environment on placental development in equids.*Journal of reproduction and fertility*, *98*(1), 55-60.

Allison, A. C. (1954). Protection afforded by sickle-cell trait against subtertian malarial infection. *British medical journal*, *1*(4857), 290.

Andersson, L., & Sandberg, K. (1982). A linkage group composed of three coat color genes and three serum protein loci in horses. *Journal of Heredity*, *73*(2), 91-94.

Andresen, E and Baker, L. N. (1964) The C Blood Group System in Pigs and the Detection and Estimation of Linkage Between the C and J Systems. *Genetics*. 49(3), 379-86. PMID:14135401

Arcot, S. S., Wang, Z, Weber, J. L., Deininger, P. L., & Batzer, M.A. (1995) Alu repeats: a source for the genesis of primate microsatellites. *Genomics, 29*(1), 136-144.

Chinwalla, A. T., Cook, L. L., Delehaunty, K. D., Fewell, G. A., Fulton, L. A., Fulton, R. S., ... & Mauceli, E. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature, 420*(6915), 520-562. doi:10.1038/nature01262

Bailey, E., Reid, R. C., Lear, T. L., Skow, L. C., Mathiason, K., & McGuire, T. C. (1997). Linkage of the gene for equine combined immunodeficiency disease to microsatellite markers HTG8 and HTG4; synteny and FISH mapping to ECA9. *Animal genetics, 28*(4), 268-273.

Bailey, E., Stormont, C., Suzuki, Y., Trommershausen Smith A. (1979) Linkage of loci controlling alloantigens on red blood cells and lymphocytes in the horse. *Science*, 204(4399), 1317–1319.

Barrett, J. C., Fry, B., Maller, J. D. M. J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*(2), 263-265. [PubMed ID: 15297300]

Bateson, W., *et al.* (1905) Experimental studies in the physiology of heredity. Reports to the Evolution Committee of the Royal Society. 2, 1–55, 80–99 (1905)

Bauman, J. G. J., Wiegant, J., Borst, P., & Van Duijn, P. (1980). A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochrome-labelled RNA. *Experimental cell research*, *128*(2), 485-490.

Bellone, R. R., Brooks, S. A., Sandmeyer, L., Murphy, B. A., Forsyth, G., Archer, S., ... & Grahn, B. (2008). Differential gene expression of TRPM1, the potential cause of congenital stationary night blindness and coat spotting patterns (LP) in the Appaloosa horse (Equus caballus). *Genetics*, *179*(4), 1861-1870.

Bellone, R., Forsyth, G., Leeb, T., Arche,r S., Sigurdsson, S., Imsland, F., Mauceli, E., Engensteiner, M., Bailey, E., Sandmeyer, L., Grahn, B., Lindblad-Toh, K., Wad,e C. (2010) Fine-mapping and mutation analysis of TRPM1: a candidate gene for leopard complex (LP) spotting and congenital stationary night blindness in horses. *Brief in Functional Genomics and Proteomics*, 9(3), 193-207. doi:101093/bfgp/elq002

Bellone, R. R., Holl, H., Setaluri, V., Devi, S., Maddodi, N., Archer, S., ... & Brooks, S. A. (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.

Bermingham Jr, J. R., & Pennington, J. (2004). Organization and expression of the SLC36 cluster of amino acid transporter genes. *Mammalian genome*, *15*(2), 114-125. doi: 10.1007/s00335-003-2319-3.

Bernstein, F. (1924) Ergebnisse einer biostatistischen zusammenfassenden Betrachtung uber die erblichen Blutstructuren des Menschen. Klin. Wochenschr. 3: 1495-1567.

Binns, M. M., Holmes, N. G., Holliman, A., Scott A. M. (1995) The identification of polymorphic microsatellite loci in the horse and their use in thoroughbred parentage testing. *British Veterinary Journal*. 151(1), 9–15.

Boll, M., Foltz, M., Rubio-Aliaga, I., & Daniel, H. (2003). A cluster of proton/amino acid transporter genes in the human and mouse genomes☆.*Genomics*, *82*(1), 47-56.

Bostick, W. D., Dinsmore, S. R., Mrochek, J. E., Waalkes, T. P. (1978) Separation and analysis of arylsulfatase isoenzymes in body fluids of man. *Clininical Chemistry*. 24(8):1305-16. PMID:28185

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms.*American journal of human genetics*, *32*(3), 314. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1686077/

Bowling A. (2000) Genetics of Color variation. In the Genetics of the Horse Edited by: Bowling A. T. Ruvinsky A. New York: CABI Publishing; 62.

Breen, M., Downs, P., Irvin, Z., & Bell, K. (1994). Six equine dinucleotide repeats: microsatellites MPZ002, 3, 4, 5, 6 and 7. *Animal genetics*, *25*(2), 124-124.

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.

Brooks, S. A., Gabreski, N., Miller, D., Brisbin, A., Brown, H. E., Streeter, C., Mezey, J., Cook, D., and Antczak, D. F. (2010) Whole-genome SNP association in the horse: identification of a deletion in myosin Va responsible for Lavender Foal Syndrome. *PLoS genetics*, *6*(4), e1000909.

Brunberg, E., Andersson, L., Cothran, G., Sandberg, K., Mikko, S., & Lindgren, G. (2006). A missense mutation in PMEL17 is associated with the Silver coat color in the horse. *BMC genetics*, *7*(1), 46. doi: 10.1186/1471-2156-7-46.

Calvo, S., Jain, M., Xie, X., Sheth, S. A., Chang, B., Goldberger, O. A., ... & Mootha, V. K. (2006). Systematic identification of human mitochondrial disease genes through integrative genomics. *Nature genetics*, *38*(5), 576-582.

Caspersson, T., Zech, L., & Johansson, C. (1970). Differential binding of alkylating fluorochromes in human chromosomes. *Experimental cell research*,*60*(3), 315-319.

Castle, W. E. (1954). Coat color inheritance in horses and in other mammals.*Genetics*, *39*(1), 35.

Chen, Z., Fei, Y. J., Anderson, C. M., Wake, K. A., Miyauchi, S., Huang, W., ... & Ganapathy, V. (2003). Structure, function and immunolocalization of a proton-coupled amino acid transporter (hPAT1) in the human intestinal cell line Caco-2. *The Journal of physiology*, *546*(2), 349-361.

Cheng, J. C., Tang, N. L., Yeung, H. Y., & Miller, N. (2007). Genetic association of complex traits: using idiopathic scoliosis as an example. *Clinical orthopaedics and related research*, *462*, 38-44. doi: 10.1097/BLO.0b013e318d09dcc.

Cole, S, Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., ... & Barrell, B.G. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 393(6685), 537-544.

Darwin, Charles (1859). On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life (1st ed.). London: John Murray.

DeChiara, T. M., Robertson, E. J., & Efstratiadis, A. (1991) Parental imprinting of the mouse insulin-like growth factor II gene. *Cell*, 64(4), 849-859.

Doege, K. J., Coulter, S. N, Meek, L. M, Maslen, K., & Wood, J. G. (1997) A Human-specific Polymorphism in the Coding Region of the Aggrecan Gene Variable Number of Tandem Repeats Produce a Range of Core Protein Sizes in the General Population. *Journal of Biological Chemistry*, 272(21), 13974-13979.

Donis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., Stephens, K., ... & Abrahamson, J. (1987) A genetic linkage map of the human genome. Cell, 51(2), 319-337.

Datta, S. P., & Stone, W. H. (1963). Tests for Linkage of the Genes controlling Cattle Transferrins and Blood Groups.

Distl, O. (2013). The genetics of equine osteochondrosis. *The Veterinary Journal*, *197*(1), 13-18.

Dutrillaux, B., & Lejeune, J. (1971). [A new technic of analysis of the human karyotype]. *Comptes rendus hebdomadaires des seances de l'Academie des sciences. Serie D: Sciences naturelles*, *272*(20), 2638-2640.

Eberth, John E. (2013) Chondrodysplasia-Like Dwarfism in the Miniature Horse. Theses and Dissertations--Veterinary Science.Paper 11. http://uknowledge.uky.edu/gluck_etds/11

Emery, A. E. (1984). Genetic heterogeneity in Duchenne muscular dystrophy.*Journal of medical genetics*, *21*(1), 76-77.

Ferguson, L. (1941) Heritable Antigens in the Erythrocytes of Cattle. *Journal of Immunology* 40, 213-242.

Flemming, W. (1878) Zur Kenntniss der Zelle und ihrer Theilungs-Erscheinungen. In: Schriften des Naturwissenschaftlichen Vereins für Schleswig-Holstein 3, 23–27.

Flicek, P., Aken, B. L., Ballester, B., Beal, K., Bragin, E., Brent, S., ... & Durbin, R. (2010). Ensembl's 10th year. *Nucleic acids research*, *38*(suppl 1), D557-D562.

Fotiadis, E., Kenanidis, E., Samoladas, E., Christodoulou, A., Akritopoulos, P., & Akritopoulou, K. (2008). Scheuermann's disease: focus on weight and height role. *European Spine Journal*, *17*(5), 673-678. doi:10.1007/s00586-008-0641-x

Fries, R., & Ruddle, F. H. (1986). Gene mapping in domestic animals. In*Biotechnology for Solving Agricultural Problems* (pp. 19-37). Springer Netherlands.

Fujita, T., Azuma, Y., Fukuyama, R., Hattori, Y., Yoshida, C., Koida, M., Ogita, K. & Komori, T. (2004). Runx2 induces osteoblast and chondrocyte differentiation and enhances their migration by coupling with PI3K-Akt signaling. *The Journal of cell biology*, *166*(1), 85-95. PubMed ID : 15226309

Gall, J. G., & Pardue, M. L. (1969). Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proceedings of the National Academy of Sciences*, *63*(2), 378-383.

Gallagher, P. C., Morrison, S., Bernoco, D., Bailey, E. (2003) Measurement of Back Curvature in American Saddlebred Horses: Structural and genetic basis for early-onset lordosis. *Journal of Equine Veterinary Science*. 23(2), 71-76.

Gallagher, P. C., Lear, T. L., Coogle, L. D., Bailey, E. (1999) Two SINE families associated with equine microsatellite loci. *Mammalian Genome*. 10(2):140-4.

Gayagay, G., Yu, B., Hambly, B., Boston, T., Hahn, A., Celermajer, D. S., & Trent, R. J. (1998). Elite endurance athletes and the ACE I allele–the role of genes in athletic performance. *Human genetics*, *103*(1), 48-50.

Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A. T., ... & Zhao, X. (1995). Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics*, *139*(2), 907-920.

Gibbs, R. A., Weinstock, G. M., Metzker, M. L., Muzny, D. M., Sodergren, E. J., Scherer, S., ... & Green, E. D. (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*, *428*(6982), 493-521. doi:10.1038/nature02426

Giblett, E., Ammann, A., Sandman, R., Wara, D., & Diamond, L. (1975). Nucleoside-phosphorylase deficiency in a child with severely defective T-cell immunity and normal B-cell immunity. *The Lancet*, *305*(7914), 1010-1013.

Giblett, E., Anderson, J., Cohen, F., Pollara, B., & Meuwissen, H. (1972). Adenosine-deaminase deficiency in two patients with severely impaired cellular immunity. *The Lancet*, *300*(7786), 1067-1069.

Giles, R. E., Blanc, H., Cann, H. M., & Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences*, *77*(11), 6715-6719.

Gilliam, T. C., Tanzi, R. E., Haines, J. L., Bonner, T. I., Faryniarz, A. G., Hobbs, W. J., MacDonald, M. E., Cheng, S. V., Folstein, S. E., Conneally, P. M., Wexler, N. S., Gusella, J. F. (1987) Localization of the Huntington's Disease Gene to a Small Segment of Chromosome 4 Flanked by D4S10 and the Telomere. *Cell*. 50(4), 565-571.

Goodrich, H. B. (1935). The development of hereditary color patterns in fish.*American Naturalist*, 267-277.

Graf, J., Hodgson, R., & Van Daal, A. (2005). Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation.*Human mutation*, *25*(3), 278-284.

Gray, M. M., Granka, J. M., Bustamante, C. D., Sutter, N. B., Boyko, A. R., Zhu, L., Ostrander, E.A. & Wayne, R. K. (2009) Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics*, 181(4), 1493-1505. doi: 10.1534/genetics.108.098830

Guérin, G., Bailey, E., Bernoco, D., Anderson, I., Antczak, D. F., Bell, K., Binns, M. M., Bowling, A. T., Brandon, R., Cholewinski, G., Cothran, E. G., Ellegren, H., Förster, M., Godard, S., Horin, P., Ketchum, M., Lindgren, G., McPartlan, H., Mériaux, J –C., Mickelson, J. R., Millon, L. V., Murray, J., Neau, A., Røed, K., Sandberg, K., Shiue, Y-L., Skow, L. C., Stott, M., Swinburne, J., Valberg, S. J., Van Haeringen, H., Van Haeringen, W. A., Ziegle, J. (1999) Report of the International Equine Gene Mapping Workshop: male linkage map. *Animal Genetics* 30(5), 341–354. doi:10.1046/j.1365-2052.1999.00510.x

Gusella, J. F., Wexler, N. S., Conneally, M. P., Naylor, S. L., Anderson, M. A., Tanzi, R. E., Watkins, P.C., Ottina, K., Wallace, M. R,, Sakaguchi, A. Y., Young, A. B., Shoulson, I., Bonnilla, E., and Martin, J. B. (1983) A Polymorphic DNA marker genetically linked to Huntington's disease. *Nature*, 306(5940),234-238.

Gutiérrez-Gil, B., Williams, J. L., Homer, D., Burton, D., Haley, C. S., & Wiener, P. (2009). Search for quantitative trait loci affecting growth and carcass traits in a cross population of beef and dairy cattle. *Journal of animal science*,*87*(1), 24-36.

Gwo-Shu Lee, M., Lewis, S. A., Wilde, C. D., & Cowan, N. J. (1983). Evolutionary history of a multigene family: an expressed human β-tubulin gene and three processed pseudogenes. *Cell*, *33*(2), 477-487.

Gyapay, G., Morissette, J., Vignal, A., Di, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M. & Weissenbach, J. (1994) The 1993−94 Généthon human genetic linkage map. *Nature Genetics* 7, 246 - 339

Haase, B., Brooks, S. A., Schlumbaum, A., Azor, P. J., Bailey, E., Alaeddine, F., ... & Leeb, T. (2007). Allelic heterogeneity at the equine KIT locus in dominant white (W) horses. *PLoS genetics*, *3*(11), e195.

Haase, B., Brooks, S. A., Tozaki, T., Burger, D., Poncet, P. A., Rieder, S., ... & Leeb, T. (2009). Seven novel KIT mutations in horses with white coat colour phenotypes. *Animal genetics*, *40*(5), 623-629.

Haase, B., Obexer-Ruff, G., Dolf, G., Rieder, S., Burger, D., Poncet, P. A., ... & Leeb, T. (2010). Haematological parameters are normal in dominant white Franches–Montagnes horses carrying a KIT mutation. *The Veterinary Journal*, *184*(3), 315-317.

Harismendy, O., Notani, D., Song, X., Rahim, N. G., Heintzman, N., Ren, B., Fu, X. D., Topol, E. J., Rosenfeld, M. G., Frazer, K. A. (2011) 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature*. 470(7333), 264-268.

Hauswirth, R., Haase, B., Blatter, M., Brooks, S. A., Burger, D., Drögemüller, C., ... & Leeb, T. (2012) Mutations in MITF and PAX3 cause "splashed white" and other white spotting phenotypes in horses. *PLoS genetics*, *8*(4).

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., ... & Aerts, J. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, *432*(7018), 695-716.

Hinds, D. A., Stuve, L. L., Nilsen, G. B., Halperin, E., Eskin, E., Ballinger, D. G., Frazer K. A. & Cox, D. R. (2005) Whole-genome patterns of common DNA variation in three human populations. *Science*, 307(5712), 1072-1079. doi:10.1126/science.1105436.

Huber, C., Delezoide, A. L., Guimiot, F., Baumann, C., Malan, V., Le Merrer, M., ... & Cormier-Daire, V. (2009). A large-scale mutation search reveals genetic heterogeneity in 3M syndrome. *European Journal of Human Genetics*,*17*(3), 395-400.

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., ... & Shen, Y. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, *449*(7164), 851-861.

International Human Genome Sequencing Consortium (2001) "Initial sequencing and analysis of the human genome" (PDF). *Nature*, 409(6822): 860–921. doi:10.1038/35057062. PMID 11237011

Jeffreys, A. J., Wilson, V., & Thein, S. L. (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature*, 314(6006), 67-73.

John, H. A., Birnstiel, M. L., & Jones, K. W. (1969). RNA-DNA hybrids at the cytological level.

Jones, A., Montgomery, HE., & Woods, DR. (2002) Human performance: a role for the ACE genotype?. *Exercise and Sport Sciences Reviews*, 30(4), 184-190.

Kameya, T., Abe, K., Aoki, M., Sahara, M., Tobita, M., Konno, H., & Itoyama, Y. (1995). Analysis of spinocerebellar ataxia type 1 (SCA1)-related CAG trinucleotide expansion in Japan. *Neurology*, *45*(8), 1587-1594.

Kan, Y. W. and Dozy, A. M. (1978) Polymorphism of DNA sequence adjacent to human beta-globin structural gene: relationship to sickle mutation. *Proceedings of the National Academy of Sciences*, 75(11), 5631-5635. PMID: 281713

Kearney, J. A. (2012) Advances in epilepsy genetics and genomics. *Epilepsy Currents*. 12(4),143-6.Review. PMID:22936886

Kelly Jr, T. J., & Smith, H. O. (1970) A restriction enzyme from Hemophilus influenzae: II. Base sequence of the recognition site. *Journal of molecular biology*, 51(2), 393-409.

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.

Kerem, B. S., Rommens, J., Buchanan, J. A., Markiewicz, D., Cox T. K., Chakravarti, A., ... & Tsui, L. C. (1989) Identification of the cystic fibrosis gene: genetic analysis. *Science*, *245*(4922), 1073-1080.

Khatkar, M. S., Nicholas, F. W., Collins, A. R., Zenger, K. R., Cavanagh, J. A., Barris, W., Schnabel, R. D., … & Raadsma, H. W. (2008) Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC genomics*, 9(1), 187.

Kim, J. J., Farnir, F., Savell, J., & Taylor, J. F. (2003). Detection of quantitative trait loci for growth and beef carcass fatness traits in a cross between Bos taurus (Angus) and Bos indicus (Brahman) cattle. *Journal of Animal Science*,*81*(8), 1933-1942.

Koenig, M., Hoffman, E. P., Bertelson, C. J., Monaco, A. P., Feener, C., & Kunkel, L. M. (1987) Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell*, *50*(3), 509-517.

Kushimoto, T., Basrur, V., Balencia, J., Mtsunaga, J., Bieira, W. D., Ferrans, V. J., Muller, J., Appella, E., Hearing, V. J. (2001) A model for melanosomes biogenesis based on the purification and analysis of early melanosomes. *Proceedings of the National Academy of Sciences*, 98(19), 10698-10703.

Kwak, K., Zhou, X., Solomon, V., Baracos, V., Davis, J., Bannon, A., Boyle, W., Lacey, D., Han, H. Q. (2004) Regulation of Protein Catabolism by Muscle-Specific and cytokine-Inducible Ubiquitin Ligase E3α-II during Cancer Cachexia. *Cancer Research*. 6(22), 8193-8198.

Lamason, R. L., Mohideen, M. A. P., Mest, J. R., Wong, A. C., Norton, H. L., Aros, M. C., ... & Cheng, K. C. (2005). SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science*, *310*(5755), 1782-1786.

Landsteiner, K. (1900) Zur Kenntnis der antifermentativen, lytischen und agglutinierenden *Wirkungen des Blutserums und der Lymphe. Zentralblatt Bakteriologie*, 27, 357–362.

Lannoy, N., & Hermans, C. (2010). The ‘royal disease’–haemophilia A or B? A haematological mystery is finally solved. *haemophilia*, *16*(6), 843-847. doi: 10.1111/j.1365-2516.2010.02327.x

Lear, T. L., Lundquist, J., Zent, W. W., Fishback Jr., W. D., Clark, A. (2008) Three autosomal chromosome translocations associated with repeated early embryonic loss (REEL) in the domestic horse (*Equus caballus*). *Cytogenetic Genome Research* 120(1-2),117–122 doi:10.1159/000118749.

Lee, W. H., Isaacs, W. B., Bova, G. S., & Nelson, W. G. (1997) CG island methylation changes near the GSTP1 gene in prostatic carcinoma cells detected using the polymerase chain reaction: a new prostate cancer biomarker. *Cancer Epidemiology Biomarkers & Prevention*, 6(6), 443-450.

Lijavetzky, D., Cabezas, J. A., Ibáñez, A., Rodríguez, V., & Martínez-Zapater, J. M. (2007). High throughput SNP discovery and genotyping in grapevine (Vitis vinifera L.) by combining a re-sequencing approach and SNPlex technology.*BMC genomics*, *8*(1), 424.

Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., ... & Aldredge, T. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, *438*(7069), 803-819.

Lindgren, G., Sandberg, K., Persson, H., Marklund, S., Breen, M., Sandgren, B., … & Ellegren, H. (1998) A primary male autosomal linkage map of the horse genome. *Genome Research*. 8(9), 951-966.

Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L., ... & Ranum, L. P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science*, *293*(5531), 864-867.

Litt, M., & Luty, J. A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene.*American journal of human genetics*, *44*(3), 397.

Locke, M. M., Ruth, L. S., Millon, L. V., Penedo, M. C. T., Murray, J. D., & Bowling, A. T. (2001). The cream dilution gene, responsible for the palomino and buckskin coat colours, maps to horse chromosome 21. *Animal genetics*,*32*(6), 340-343.

Ludwig, A., Pruvost, M., Reissmann, M., Benecke, N., Brockmann, G. A., Castaños,P., Cieslak, M., Lippold, S., Llorente, L., Malaspinas, A-S., Slatkin, M. and Hofreiter, M. (2009) Coat Color Variation at the Beginning of Horse Domestication. *Science.* 324(5926), 485. doi:10.1126/science.1172750.

Maksimova, N., Hara, K., Miyashia, A., Nikolaeva, I., Shiga, A., Nogovicina, A., Sukhomyasova, A., Argunov, V., Shvedova, A., Ikeuchi, T., Nishizawa, M., Kuwano, R., Onodera, O. (2007) Clinical, molecular and histopathological features of short stature syndrome with novel CUL7 mutation in Yakuts: new population isolate in Asia .*Journal of Medical Genet*ics, 44(12), 772-778.

Makvandi-Nejad, S., Hoffman, G. E., Allen, J. J., Chu, E., Gu, E., Chandler, A. M., ... & Sutter, N. B. (2012) Four Loci explain 83% of size variation in the horse. *PLoS One*, *7*(7), e39929.

Mariat, D., Taourit, S., Guerin, G. (2002) A Mutation in the MATP gene causes the cream coat colour in the horse. *Genetics Selection Evolution*. 35(1), 119-133.

Marklund, S., Ellegren, H., Eriksson, S., Sandberg, K. and Andersson, L. (1994) Parentage testing and linkage analysis in the horse using a set of highly polymorphic microsatellites. *Animal Genetics*. 25(S1),19–23.

Maruyama, Z., Yoshida, C. A., Furuichi, T., Amizuka, N., Ito, M., Fukuyama, R., Miyazaki, T., Kitaura, H., Nakamura, K., Fujita, T., Kanatani, N., Moriishi, T., Yamana, K., Liu, W.,

Kawaguchi, H., Nakamura, K. and Komori, T. (2007) Runx2 determines bone maturity and turnover rate in postnatal bone development and is involved in bone loss in estrogen deficiency. *Developmental Dyn*amics, 236(7), 1876–1890. doi: 10.1002/dvdy.21187

Maxam, A. M., Gilbert W. (1977) A new method for sequencing DNA. *Proceedings of National. Academy of Science*. 74(2), 560–564.

McCarrey, J. R. (1990) Molecular evolution of the human Pgk-2 retroposon. *Nucleic Acids Research*, 18(4), 949–955.

McClintock, B. (1950) The origin and behavior of mutable loci in maize. *Proceedings of National Academy of Science*,36(6), 344–355. doi:10.1073/pnas.36.6.344.

Mendel, J.G. (1866) Versuche über Pflanzenhybriden Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr, 1865 Abhandlungen:3–47. For the English translation, see: Druery, C.T and William Bateson (1901). Experiments in Plant Hybridization. *Journal of the Royal Horticultural Society* 26: 1–32. http://www.esp.org/foundations/genetics/classical/gm-65.pdf. Retrieved 2009-10-09.

Miller N.H., Justice CM, Marosy B, Doheny KF, Pugh E, Zhang J, Dietz HC 3rd, Wilson AF. (2005) Identification of Candidate Regions for Familial Idiopathic Scoliosis. *Spine*. 30(10), 1181-7.

Modiano, G., Morpurgo, G., Terrenato, L., Novelletto, A., Di Rienzo, A., Colombo, B., Purpura, M., Mariani, M., Santachiara-Benerecetti, S., Brega, A., Dixit, K. A.,Shrestha, S. L., Lania, A., Wanachiwanawin W. & Luzzatto, L. (1991) Protection against malaria morbidity: near-fixation of the α-thalassemia gene in a Nepalese population. *American journal of human genetics*, 48(2), 390-397. PMCID: PMC1683029

Morgan, T. H. (1910) Sex-limited inheritance in Drosophila. *Science*. 32(812), 120–122.

Morgan, T. H. (1911). Random segregation versus coupling in Mendelian inheritance. *Science*, *34*(873), 384-384.

Morton, N. E. (1955). Sequential tests for the detection of linkage. *American journal of human genetics*, *7*(3), 277.PMC 1716611. PMID 13258560

Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA, *et al.*,. (2000) A whole-genome assembly of Drosophila. *Science* 287: 2196-2204.

Nkrumah, J. D., Sherman, E. L., Li, C., Marques, E., Crews, D. H., Bartusiak, R., ... & Moore, S. S. (2007). Primary genome scan to identify putative quantitative trait loci for feedlot growth rate, feed intake, and feed efficiency of beef cattle. *Journal of Animal Science*, *85*(12), 3170-3181.

Ownby, D. R., Pizzo, S., Blackmon, L., Gall, S. A., & Buckley, R. H. (1976) Severe combined immunodeficiency with leukopenia (reticular dysgenesis) in siblings: immunologic and histopathologic findings. *The Journal of Pediatrics*, 89(3), 382-387.

Penedo, M. C. T., Millon, L. V., Bernoco, D., Bailey, E., Binns, M., Cholewinski, G., ... & Tozaki, T. (2005). International Equine Gene Mapping Workshop Report: a comprehensive linkage map constructed with data from new markers and by merging four mapping resources. *Cytogenetic and genome research*, *111*(1), 5-15.

Philipsson, J., Andreasson, E., Sandgren, B., Dalin, G., & Carlsten, J. (1993). Osteochondrosis in the tarsocrural joint and osteochondral fragments in the fetlock joints in Standardbred trotters. II. Heritability. *Equine Veterinary Journal*,*25*(S16), 38-41.

Pizza, M., Scarlato, V., Masignani, V., Giuliani, M. M., Arico, B., Comanducci, M., ... & Rappuoli, R. (2000). Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science*,*287*(5459), 1816-1820.

Podliachouk, L. (1957) Les antigens de groupes sanguins des equids et leur transmission. hereditaire. Thesis, Univ. Paris.

Podliachouk, L. (1958) Les groupes sanguins des equids (cheval, mulet, ane). *Ann. Inst. Pasteur* 95: 7-22.

Pribyl, J., Rehout, V., Citek, J., & Pribylova, J. (2010). Genetic evaluation of dairy cattle using a simple heritable genetic ground. *Journal of the Science of Food and Agriculture*, *90*(11), 1765-1773. PMID: 20564310

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*. 81. <http://pngu.mgh.harvard.edu/purcell/plink/>

Raadsma, H. W. (2004). Quantitative trait loci mapping in dairy cattle: review and meta-analysis. *Genet. Sel. Evol*, *36*, 163-190.

Rafalski, J. A. (2010). Association genetics in crop improvement. *Current opinion in plant biology*, *13*(2), 174-180.

Rapp, J. P. (1983). Genetics of experimental and human hypertension.*Hypertension Physiopathology and Treatment*, 582-598.

Rapp, J. P. (1983). A paradigm for identification of primary genetic causes of hypertension in rats. *Hypertension*, *5*(2 Pt 2), I198.

Rapp, J. P. (1987). Use and misuse of control strains for genetically hypertensive rats. *Hypertension*, *10*(1), 7-10.

Rapp, J. P. (1991). Dissecting the primary causes of genetic hypertension in rats. *Hypertension*, *18*(3 Suppl), I18.

Raudsepp, T., Gustafson-Seabury, A., Durkin, K., Wagner, M. L., Goh, G., Seabury, C. M., ... & Chowdhary, B. P. (2008). A 4,103 marker integrated physical and comparative map of the horse genome. *Cytogenetic and genome research*, *122*(1), 28-36.

Reith, W., Satola, S., Sanchez, C. H., Amaldi, I., Lisowska-Grospierre, B., Griscelli, C., Hadam M. R. & Mach, B. (1988) Congenital immunodeficiency with a regulatory defect in MHC class II gene expression lacks a specific HLA-DR promoter binding protein, RF-X. *Cell*, 53(6), 897-906.

Ritchie, R. J., Mattei, M. G., & Lalande, M. (1998). A large polymorphic repeat in the pericentromeric region of human chromosome 15q contains three partial gene duplications. *Human molecular genetics*, *7*(8), 1253-1260.

Rooney, J. R. (1969). 3 Congenital Equine Scoliosis and Lordosis. *Clinical orthopaedics and related research*, *62*, 25-30.

Rooney, J. R. and Robertson, J. (1996) Equine Pathology. Iowa State University Press. Ames. p 222.

Rooney, J. R., & Prickett, M. E. (1967). Congenital lordosis of the horse. *The Cornell Veterinarian*, *57*(3), 417-428.

Rozen, S. Skaletsky, H. J. (1998) Primer3. Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html. Accessed at: http://biotools.umassmed.edu/bioapps/primer3_www.cgi.

Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor, G. L., Nelson, C. R., Hariharan, I. K., ... & Zhong, F. (2000). Comparative genomics of the eukaryotes. *Science*, *287*(5461), 2204-2215.

Rudolph, J. A., Spier, S. J., Byrns, G., Rojas, C. V., Bernoco, D., & Hoffman, E. P. (1992) Periodic paralysis in quarter horses: a sodium channel mutation disseminated by selective breeding. *Nature genetics*, *2*(2), 144-147.

Ruwende, C. Y. R. I. L., & Hill, A. (1998). Glucose-6-phosphate dehydrogenase deficiency and malaria. *Journal of Molecular Medicine*, *76*(8), 581-588.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., ... & Altshuler, D. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409(6822), 928-933.

Sagné, C., Agulhon, C., Ravassard, P., Darmon, M., Hamon, M., El Mestikawy, S., ... & Giros, B. (2001). Identification and characterization of a lysosomal transporter for small neutral amino acids. *Proceedings of the National Academy of Sciences*, *98*(13), 7206-7211.

Samani, N. J., Erdmann, J., Hall, A. S., Hengstenberg, C., Mangino, M., Mayer, B., ... & Schunkert, H. (2007). Genomewide association analysis of coronary artery disease. *New England Journal of Medicine*, *357*(5), 443-453. doi: 10.1056/NEJMoa072366

Sandberg, K., & Juneja, R. K. (1978). Close linkage between the albumin and Gc loci in the horse. *Animal blood groups and biochemical genetics*, *9*(3), 169-173.

Sandberg, K. (1974) Linkage between the K blood group locus and the 6-PGD locus in horses. *Anim. Blood Groups Biochem. Genet*. 5: 137–141.

Sanders, S. J., Ercan-Sencicek, A. G., Hus, V., Luo, R., Murtha, M. T., Moreno-De-Luca, D., ... & Mane, S. M. (2011). Multiple recurrent de novo CNVs, including duplications of the 7q11. 23 Williams syndrome region, are strongly associated with autism. *Neuron*, *70*(5), 863-885.

Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., Brown, N. L., ... & Smith, M. (1978). The nucleotide sequence of bacteriophage φX174.*Journal of molecular biology*, *125*(2), 225-246. doi:10.1038/265687a0.

Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology*, *94*(3), 441-448.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences,74*(12), 5463-5467.

Schmid, C. W. (1998). Does SINE evolution preclude Alu function?. *Nucleic acids research*, *26*(20), 4541-4550.

Schunkert, H., Götz, A., Braund, P., McGinnis, R., Tregouet, D. A., Mangino, M., ... & Samani, N. J. (2008). Repeated replication and a prospective meta-analysis of the association between chromosome 9p21. 3 and coronary artery disease. *Circulation*, *117*(13), 1675-1684.. doi:10.1161/Circulationaha.107.730614

Schunkert, H., König, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H., ... & de Faire, U. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature genetics*, *43*(4), 333-338. doi:10.1038/ng.784

Seabright, M. (1971) A rapid banding technique for human chromosomes. Lancet11, 971–972.

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.

Sharon, D., Glusman, G., Pilpel, Y., Khen, M., Gruetzner, F., Haaf, T., & Lancet, D. (1999) Primate evolution of an olfactory receptor cluster: diversification by gene conversion and recent emergence of pseudogenes. *Genomics*, 61(1), 24-36.

Sheridan, C. M., Magee, R. M., Hiscott, P. S., Hagan, S., Wong, D. H., McGalliard, J. N., & Grierson, I. (2002). The role of matricellular proteins thrombospondin-1 and osteonectin during RPE cell migration in proliferative vitreoretinopathy. *Current eye research*, *25*(5), 279-285.

Sherman, E. L., Nkrumah, J. D., Murdoch, B. M., Li, C., Wang, Z., Fu, A., & Moore, S. S. (2008). Polymorphisms and haplotypes in the bovine neuropeptide Y, growth hormone receptor, ghrelin, insulin-like growth factor 2, and uncoupling proteins 2 and 3 genes and their associations with measures of growth, performance, feed efficiency, and carcass merit in beef cattle. *Journal of animal science*, *86*(1), 1-16.

Shiue, Y., Bickel, L. A., Caetano, A. R., Millon, L.V., Clark, R.S., Eggleston, M. L., Michelmore, R., Bailey, E., Guérin, G., Godard, S., Mickelson, J. R., Valberg, S. J., Murray, J. D. and Bowling,. A.T. (1999) A synteny map of the horse genome comprised of 240 microsatellite and RAPD markers. *Animal Genetics*. 30(1),1-10.

Shoulson, I., & Chase, T. N. (1975). Huntington's disease. *Annual review of medicine*, *26*(1), 419-426.

Somerville, M. J., Mervis, C. B., Young, E. J., Seo, E. J., del Campo, M., Bamforth, S., ... & Osborne, L. R. (2005). Severe expressive-language delay related to duplication of the Williams–Beuren locus. *New England Journal of Medicine*, *353*(16), 1694-1701.

Sponenberg, D. (2003) Equine Color Genetics. 2nd Edition. Ames, IA: Iowa State University Press. 46-49.

Sponenberg, D. (2003) Equine Color Genetics. 2nd Edition. Ames, IA: Iowa State University Press. 110-120.

Stein, G. S., Lian, J. B., Van Wijnen, A. J., Stein, J. L., Montecino, M., Javed, A., ... & Pockwinse, S. M. (2004). Runx2 control of organization, assembly and activity of the regulatory machinery for skeletal gene expression. *Oncogene*,*23*(24), 4315-4329. PubMed ID : 15156188

Stormont, C., & Suzuki, Y. (1964). Genetic systems of blood groups in horses.*Genetics*, *50*(5), 915.

Stormont C, Suzuki Y.(1963) Genetic Control of Albumin Phenotypes in Horses. *Experimental Biology and Medicine*, 114(3), 673-675.

Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., ... & Olson, M. V. (2000). Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature*, *406*(6799), 959-964.

Stromberg, B. (1979). A review of the salient features of osteochondrosis in the horse. *Equine veterinary journal*, *11*(4), 211-214.

Sturtevant, A. H. (1913). The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association. *Journal of experimental zoology*, *14*(1), 43-59.

Sturtevant, A. H., & Beadle, G. W. (1939). An introduction to genetics. *An introduction to genetics.*

Sumner, A. T., Evans, H. J., & Buckland, R. A. (1971). New technique for distinguishing between human chromosomes. *Nature*, *232*(27), 31-32.

Sutton, W. S. (1902) On the morphology of the chromomsome group in Brachystola magna. *Biological Bulletin* 4, 24-39.

Takahashi, Y., Kou, I., Takahashi, A., Johnson, T. A., Kono, K., Kawakami, N., ... & Ikegawa, S. (2011). A genome-wide association study identifies common variants near LBX1 associated with adolescent idiopathic scoliosis. *Nature genetics*, *43*(12), 1237-1240.

Takeda, S., & Matsuoka, M. (2008). Genetic approaches to crop improvement: responding to environmental and population changes. *Nature Reviews Genetics*,*9*(6), 444-457.

Tanksley, S. D., Young, N. D., Paterson, A. H., & Bonierbale, M. W. (1989). RFLP mapping in plant breeding: new tools for an old science. *Nature Biotechnology*, *7*(3), 257-264.

Terry, R. B., Archer, S., Brooks, S., Bernoco, D., & Bailey, E. (2004). Assignment of the appaloosa coat colour gene (LP) to equine chromosome 1.*Animal genetics*, *35*(2), 134-137.

Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., ... & Venter, J. C. (1997). The complete genome sequence of the gastric pathogen Helicobacter pylori. *Nature*, *388*(6642), 539-547.

Tozaki T., Mashima S., Hirota K., Miura N., Choi-Miura N.H., Tomita M. (2001) Characterization of equine microsatellites and microsatellite-linked repetitive elements (eMLREs) by efficient cloning and genotyping methods. *DNA Research*. 8(1), 33-45.

Trommershausen-Smith, A. (1978) Linkage of tobiano coat spotting and albumin markers in a pony family. *Journal of Heredity* 69: 214-216.

Vidaud, D., Vidaud, M., Bahnak B.R., Siguret V., Laurian Y., Meyer D., ... & Lavergne J.M. (1992) Haemophilia B due to a de novo insertion of a human-specific Alu subfamily member within the coding region of the factor IX gene. *European Journal of Human Genetics*. 1(1), 30-36.

Villagómez D.A.F. Lear T.L. Chenier T. Lee S. McGee R.B. Cahill J. Foster R.A. Reyes E. · St John E. · King W.A. (2011) Equine Disorders of Sexual Development in 17 Mares Including XX, SRY-Negative, XY, SRY-Negative and XY, SRY-Positive Genotypes. *Sex Development* 5:16–25 doi:10.1159/000322811.

Visscher P.M., McEvoy B., Yang J. (2010) From Galton to GWAS: quantitative genetics of human height. *Genetic Research* (Camb) 92: 371–379. doi: 10.1017/S0016672310000571.

Wade, C. M., Giulotto, E., Sigurdsson, S., Zoli, M., Gnerre, S., Imsland, F., ... & Swinburne, J. E. (2009). Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science*, *326*(5954), 865-867.

Waldeyer-Hartz, H. W. G. (1888) Über Karyokinese und ihre Beziehungen zu den Befruchtungsvorgängen. Archiv für mikroskopische Anatomie und Entwicklungsmechanik. 32: 1-122.

Wang, X., Miller, D. C., Harman, R., Antczak, D. F., & Clark, A. G. (2013). Paternally expressed genes predominate in the placenta. *Proceedings of the National Academy of Sciences*, *110*(26), 10705-10710.

Ward, K., Ogilvie, J., Argyle, V., Nelson L, Meade M, Braun J and Chettier R. (2010) Polygenic inheritance of adolescent idiopathic scoliosis: A study of extended families in Utah. *American Journal of Medical Genetics*. 152A: 1178–1188. doi: 10.1002/ajmg.a.33145

Ward, K., Ogilvie, J. W., Singleton, M. V., Chettier, R., Engler, G., & Nelson, L. M. (2010) Validation of DNA-based prognostic testing to predict spinal curve progression in adolescent idiopathic scoliosis. *Spine*, *35*(25), E1455-E1464.

Watabe, H., Valencia, J. C., Le Pape, E., Yamaguchi, Y., Nakamura, M., Rouzaud, F., ... & Hearing, V. J. (2008). Involvement of dynein and spectrin with early melanosome transport and melanosomal protein trafficking. *Journal of Investigative Dermatology*, *128*(1), 162-174.

Watabe, H., Valencia, J. C., Yasumoto, K. I., Kushimoto, T., Ando, H., Muller, J., ... & Hearing, V. J. (2004). Regulation of tyrosinase processing and trafficking by organellar pH and by proteasome activity. *Journal of Biological Chemistry*, *279*(9), 7971-7981.

Weber, J., May, P. (1989) Abundant Class of Human DNA Polymorphisme Which Can Be Typed Using the Polymerase Chain Reaction. *American Journal of Human Genetics*. 44:388-396.

Weber, J. L. & May, P. E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics*. 44, 388−396.

Weber, J.L.& May, P.E. (1989)Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics.* 44, 388−396.

Weiss, M. C., Green, H. (1967) Human–mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. Proceedings of the National Academy of Sciences of the USA.;58:1104–1111

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., ... & Lathrop, M. (1992) A second-generation linkage map of the human genome. *Nature*, 359(6398), 794-801.

Weitkamp, L. and Allen, P. (1979) Evolutionary conservation of equine Gc alleles and of mammalian Gc/albumin linkage. *Genetics* 92:1347-1354.

WHO Scientific Group. (1995) Primary immunodeficiency diseases: report of a WHO scientific group. *Clinical Experimental Immunology*, 99, pp. 1–24.

WHO. World Health Organization. (2011) Top 10 Causes of Death. http://who.int/mediacentre/factsheets/fs310/en/.

Wilson, E. B. (1911). The sex chromosomes. Mikrosk Anat Entwicklungsmech. 77:249–271.

Winter, R. B., Moe, J. H., Lonstein, J. E. (1983) A review of family histories in patients with congenital spinal deformities. *Orthopaedic transactions*. 7:32.

Wise, C., Barnes, R., Gillum, J., Herring, J., Bowcock, A., Lovett, M. (2000) Localization of Susceptibility to Familial Idiopathic Scoliosis. Spine. 25, 18; 2372-2380.

Wood, H.S., Ke, X., Nuttall, T., McEwan, N., Ollier, W. E., Carter, S. D. (2009) Genome-wide association analysis of canine atopic dermatitis and identification of disease related SNPs. *Immunogenetics*. 61: 765-772. doi: 10.1007/s00251-009-0402-y.

Wreden, C.C., Johnson, J., Tra,n C., Sea,l R. P., Copenhagen, D.R., *et al.*, (2003) The H+-coupled electrogenic lysosomal amino acid transporter LYAAT1 localizes to the axon and plasma membrane of hippocampal neurons. *Journal of Neuroscience*. 15;23(4): 1265-75.

Zimin, A. V., Delcher, A. L., Florea, L., Kelley, D. R., Schatz, M. C., Puiu, D., ... & Salzberg, S. L. (2009) A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biology*, 10(4), R42.

VITA

Deborah Cook

Education

Doctor of Philosophy: Expected August 2014
Major: Veterinary Science
University of Kentucky, Lexington, Kentucky

Bachelor of Science: May 2006
Major: Recombinant Genetic Technology
Minor: Biochemistry
Western Kentucky University, Bowling Green KY

Associate in Science: May 2004.
Owensboro Community and Technical College, Owensboro KY

Positions Held

Research Assistant, University of Kentucky Department of Veternary Science
August 2006 to July 2010:
Supervisor: Professor Ernest Bailey, ebailey@uky.edu; 859-278-4897 x81105.

Western Kentucky University Biotechnology Center, Assistant Coordinator/Student Worker:
Fall 2004 to July 2006:
Professors: Sigrid Jacobshagen, sigrid.jacobshagen@wku.edu; (270) 745-5994
Claire Rinehart, Claire.Rinehart@wku.edu; (270) 745-6001.

AWARDS

USDA-NRSP8 Travel Award for Plant and Animal Genome Conference (2010)

USDA-NRSP8 Travel Award for Plant and Animal Genome Conference (2009)

Kentucky Opportunity Fellowship (Fall 2006)

Junior College Transfer Scholarship renewed (2005-2006)

Junior College Transfer Scholarship (2004-2005)

Deans List (fall 2002 through spring 2004)

Kentucky Community and Technical College System - All Academic Team Scholarship (2003)

USA Today, Kentucky Region, All Academic Team (2003)

Peer Reviewed Publications

Cook, D., Brooks, S., Bellone, R., Bailey, E. (2008) Missense Mutation in Exon 2 of SLC36A1 Responsible for Champagne Dilution in Horses. *PLoS Genetics* 4(9):e1000195.doi:10.1371/journal.pgen.1000195

Brooks, S. A., Gabreski, N., Miller, D., Brisbin, A., Brown, H. E., Streeter, C., Mezey, J., Cook, D., and Antczak, D. F. (2010) Whole-genome SNP association in the horse: identification of a deletion in myosin Va responsible for Lavender Foal Syndrome. *PLoS genetics, 6*(4), e1000909. doi:10.1371/journal.pgen.1000909

Cook, D., Gallagher, P. C. and Bailey, E. (2010) Genetics of Swayback in American Saddlebred Horses. *Animal Genetics* 41 (supl 2): 64-71.

Go , Y.Y., Bailey E., Cook D., Coleman S., MacLeod J., Chen K-C, Timoney P., and Balasuriya U.B. (2011) Genome-Wide Association Study Among Four Horse Breeds Identifies a Common Haplotype Associated with the In Vitro CD3+ T Cell Susceptibility/Resistance to Equine Arteritis Virus Infection. *Journal of Virology*. 85:13174-13184

Abstracts

Cook, D.G., Bailey E. (2010) Genetics of Extreme Lordosis in American Saddlebred Horses. *Plant and Animal Genome Conference XVIII (P621), p 222.*

Cook D, Patrick Gallagher, Ernest Bailey (2009) Illumina Equine SNP50 Bead Chip Investigation of Adolescent idiopathic lordosis among American Saddlebred Horses. *Journal of Equine Veterinary Science* 29: 315-316.

Cook, D. and Bailey E. (2009) Whole genome association study of extreme lordosis in the American Saddlebred Horses. *Plant and Animal Genome Conference XVIII (P173).*

Presentations

Cook, D. G. (2009) Illumina equine snp50 bead chip investigation of adolescent idiopathic lordosis among American Saddlebred horses. Equine Science Society. Keystone, CO.

Cook, D. G. (2009) Horse Genetics: A Look at Coat Dilution: Midway College. Genetics class invited speaker. Midway, KY.

Cook, D. G. (2008) Horse Genetics: Applications in Equine Research: Midway College. Genetics class invited speaker. Midway KY.

Cook, D. G. (2008) Champagne Horses and Coat Color Genes: Owensboro Community and Technical College. Phi Theta Kappa invited speaker. Owensboro, KY.