2023

MOLECULAR GENETIC STUDIES OF HORSES, ESPECIALLY WITH REFERENCE TO AGGRECAN AND DWARFISM

John Edmund Eberth
University of Kentucky, johneberth1971@gmail.com
Author ORCID Identifier:
https://orcid.org/0000-0002-1894-2317
Digital Object Identifier: https://doi.org/10.13023/etd.2023.276

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.

John Edmund Eberth, Student

Dr. Ernest Bailey, Major Professor

Dr. Martin Nielsen, Director of Graduate Studies

MOLECULAR GENETIC STUDIES OF HORSES, ESPECIALLY WITH REFERENCE TO *AGGRECAN* AND DWARFISM

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
John Edmund Eberth
Lexington, Kentucky
Director: Dr. Ernest Bailey, Professor of Genetics
Lexington, Kentucky
2023

https://orcid.org/0000-0002-1894-2317

ABSTRACT OF DISSERTATION

MOLECULAR GENETIC STUDIES OF HORSES, ESPECIALLY WITH REFERENCE TO *AGGRECAN* AND DWARFISM

This work consists of studies on dwarfism in Miniature horses and a study on breakdowns in Thoroughbreds in relation to gene *aggrecan*. A correction of the description and designation of *D3* dwarf variant in *aggrecan* (*ACAN*) from the master's thesis *Chondrodysplasia-like dwarfism in the Miniature horse* (2013). Commercial sequencing showed previous sequencing reads presented an artifact and not a single base deletion. Analysis showed a single base missense mutation in exon 8 identified as *D3** was the actual cause. *Multiple alleles of ACAN associated with chondrodysplastic dwarfism in Miniature horses* by Eberth et al. (2018), corrected the error from the 2013 thesis.

An investigation was conducted of American Miniature horses for the presence of a dwarfism haplotype in *ACAN* previously discovered in Miniature horses of Brazil. Andrade et al. (2020) found a haplotype associated with a dwarf phenotype in the Miniature horses of Brazil, however the variant is non-causative. The Brazil Miniature population showed complete association of dwarfism to the haplotype. The American Miniature population did show a high correlation to dwarfism but did not show complete association of a dwarf phenotype and haplotype.

A study was conducted which focused on the genomic investigation for putative *D5* dwarfism allele in Miniature horses. Eberth *et al*. (2018) described eight dwarf phenotype Miniature horses which only possessed 1 copy or no copies of the four mutations discovered. Whole genome sequencing was performed on two dwarfs possessing unknown genotypes. *In silico* analysis using PredictSNP showed 3 SNPs as likely deleterious to the function of the protein. A population of 464 Miniatures previously genotyped for dwarf variants *D1, D2, D3*, *D4,* and Brazil haplotype were tested with custom TaqMan assays designed for the three putative mutations. Variant g. 95258999 on Chromosome 1 shows independent segregation to dwarf variants *D1, D2, D3, D4* and Brazil haplotype. We propose this variant is associated with a dwarf phenotype in the Miniature horse named *D5*.

An investigation was conducted into changes in breed standard head phenotypes and dwarfism in the Miniature horse. Facial structure of the Miniature horse has changed significantly due to the desires of breeders since the breed inception. Morphometric measurements using a landmark based method were obtained on 106 normal phenotype Miniature horses to compare size variance of specific facial features in relation to dwarf

carrier status. Measurement results showed no correlation of specific facial feature size or shape to carrier state.

A study was conducted examining catastrophic injury in Thoroughbred training or racing and copy number variations in the *ACAN* VNTR region. A randomly selected group of 45 Thoroughbred horses and 17 horses that succumbed to injury on the racetrack were compared specifically for the sequence length of this region in *aggrecan*. Both groups presented significant variation in sequence length for this region of *aggrecan*. When the distribution of alleles between the two groups was compared, no significant differences were observed (P=0.699). This evidence suggests genetic variation for *aggrecan* and its ability to bind CS chains is not a major factor causing racetrack injury of Thoroughbred horses.

KEYWORDS: Dwarfism, *Aggrecan*, *Equus Caballus*, Miniature Horse, Genetics, Thoroughbred

John Edmund Eberth

June 21, 2023

MOLECULAR GENETIC STUDIES OF HORSES, ESPECIALLY WITH REFERENCE TO *AGGRECAN* AND DWARFISM

By
John Edmund Eberth

_____Ernest Bailey_____
Director of Dissertation

_____Martin Nielsen_____
Director of Graduate Studies

_____June 21, 2023_____
Date

DEDICATION

This is dedicated to my wife, Melinda, and my three daughters, Mariah, Lauren, and Marissa. Their never-ending love and support made this work possible. I have been blessed with a family that patiently helped me fulfill a dream.

Melinda, you are my rock. I could always trust you to keep everything around me on an even keel so I could complete this dream. Your ability to run our family smoothly and take care of the girls is a testament to your love for me knowing this work would end "soon". Your patience with me during this endeavor has been nothing short of amazing. Please know you are as much a part of this manuscript as I am. Thank you for all you are and do for us.

To my girls, each of you have always shown interest in what I was doing and asking how it was progressing, giving me support and emotional boosts to help me keep focused. My heart has felt your love and understanding, thank you for your patience. I am so proud to have children with your caring qualities.

"Patience is not simply the ability to wait - it's how we behave while we're waiting."
-Joyce Meyer

ACKNOWLEDEGMENTS

I would like to first thank all the individuals who supported or helped me through my research and education. I am forever grateful to each one of you.

To you, Dr. Ernest Bailey, I must convey my utmost gratitude, and appreciation. I could not have accomplished this research without your unwavering support and encouragement, and at times gentle prodding. Your patience with me during this work is a testament to you as an advisor but more importantly and trusted friend. Your willingness to share knowledge and life experiences will be deeply cherished. You motivated me to be better, be exceptional, not just as a scientist but as a human.

I would also like to thank the members of my committee, Dr. David Fardo, Dr. Kathy Graves, and Dr. Ted Kalbfleisch. I am extremely appreciative of their willingness to contribute their time and advice to my research. Their mentoring and care to ensure my work would be something to be proud of, Thank You, I am truly grateful.

I would like to thank my fellow member and associate of the Bailey Lab, Navid Yousefi-Mashouf. Navid, thank you for all your interest in my work as I am grateful for you sharing your findings with me and all my questions about your work. You are a great friend and colleague, and I enjoyed the numerous rants and frustrations we shared during our time together in the lab. I am truly grateful for your help at times when I was at my wits end.

My sincere thanks go to all my friends and colleagues at the Gluck Equine Research Center, especially to the Department Chair and Gluck Director, Dr. Dan Howe and Director of Graduate Studies, Dr. Martin Nielsen. I also would like to give special thanks to Gluck Genetic Testing Lab, Dr. Kathy Graves, Jeff Boyles, and Karin Davies for their assistance. Finally, special thanks go to Dr. James MacLeod for his mentoring advice and passionate assistance in my research endeavors.

I would like to thank Mr. Wayne Hipsley. He not only has been a true friend but a mentor in all my professional and educational endeavors. His advice has been instrumental in my success in business and professional ventures in all areas of the equine industry, which is challenging to say the least. His professional guidance has been vital to my growth.

I would like to thank the late Dr. Scott Bennett, who unfortunately passed in 2020. I had the pleasure of his acquaintance since I was 10 years old. Since my childhood, he had been an invaluable mentor in the realm of equine medicine and science, not only working under him as a veterinary medicine intern during my undergraduate studies but assisting in my graduate research as well as an advisor in my professional careers in the equine industry. In my opinion, his keen knowledge and expertise in equine medicine was unmatched.

I would like to give a special thanks to Liza MacMillan whose generous gift made it possible to accomplish some of the work in this dissertation. I admire her passion for the care of Miniature horses afflicted with dwarfism and keen interest in dwarfism research.

Finally, I would like to thank my family for their committed support over the many, many years it has taken to accomplish this work. My parents and siblings for their love and encouraging support during all the highs and lows of graduate school and research. I am eternally grateful for the love and devotion of my wife, Melinda. I could not have accomplished any of this without her by my side. I am grateful for the support from my three daughters, Mariah, Lauren, and Marissa. I am forever thankful for their patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I
Introduction

Characteristics of the Miniature horse

Stature among horses is usually the product of many genes (Makvandi-Nejad *et al.* 2012; Signer-Hasler *et al.* 2012; Tetens *et al.* 2013; Metzger *et al.* 2013). Frischknecht and co-workers (2015) identified a variant in the gene *HMGA2* that is common among Shetland ponies and other pony breeds but absent among full sized horses. A QTL unique to Miniature horses that followed a dominant mode of inheritance was discovered by Al Abri and co-workers (2018). Horses with at least one copy of this variant were 4.064 cm (1.6 inches) taller on average than horses homozygous for the recessive genotype. Miniature horses are a breed of horse selected for diminutive size. The foundation stock includes many pony breeds with significant influence from the American Shetland Pony. Though full-sized horses are typically greater than 58 inches at the withers, Miniature horses cannot exceed 34 inches at the withers for registration in the American Miniature Horse Association (AMHA 2023) or 38 inches at the withers for registration in the American Miniature Horse Registry (AMHR 2023). Any color combination and any phenotype of Miniature is acceptable to be registered if the height requirement is met. An example of a Miniature horse is shown below (Figure 1.1).

Figure 1.1. Adult Miniature horse with a height of approximately 32 inches at the withers.

Characteristics of dwarfism in Miniature horses

Breed constraints define Miniature horses to possess a phenotype with proportionate small stature. Unfortunately, diminutive size can be achieved by genetic variants causing conditions characterized as dwarfism. Dwarfism not only affects stature but also is associated with developmental defects that adversely affect the horse's health and lead to disproportionate size reduction. Dwarfism in Miniatures is recognized by exhibiting several of the following characteristics: severely shortened stature, shortened limbs relative to overall body size, bowed forelegs, shortened neck, enlarged cranium, flat faces with large bulging eye sockets with prominent eyes, low nasal bridge, severe under bite, retruded muzzle, cleft palate and protruding tongue and a large abdominal hernia, or embryonic loss (Naviaux 1999; Frankeny *et al.* 2003; Tyson *et al.* 2004; Kelmer *et al.* 2007; Eberth *et al.* 2009; Eberth 2013; Watanabe *et al.* 2014; Metzger *et al.* 2016; Eberth *et al.* 2018; Andrade *et al.* 2020). An example of the dwarf phenotype in Miniature horses is shown in Figure 1.2.

Figure 1.2. Adult chondrodysplastic dwarf Miniature horse.

Identification of genetic variants associated with dwarfism in livestock

In domesticated animals, variants in at least 14 genes have been reported to cause conditions characterized as both proportionate and disproportionate dwarfism (reviewed in Boegheim *et al.* 2017). Variants responsible for disruptions of normal body growth in hormonal pathways, signaling pathways or DNA regulation are in genes *growth hormone*

1 (*GH1), growth hormone receptor (GHR)* (Hull *et al*. 1993; Agarwal *et al*. 1994; Liu *et al*. 1999)*, ring finger protein 11 (RNF11)* (Shembade *et al*. 2009; Sartelet *et al.* 2012), *cGMP-dependent protein kinase II (PRKG2)* (Koltes *et al*. 2009)*, EvC ciliary complex subunit 2 (EVC2)* (Takeda *et al*. 2002)*, Indian Hedgehog (IHH)* (Jin *et al*. 2016)*, and gon-4 like (GON4L)* (Schwarzenbacher *et al.* 2016). Inheritance patterns of these mutations are mostly autosomal recessive, with the exceptions of *GHR* being z-linked recessive in chickens and *IHH* as autosomal semi-dominant. Variants responsible for disruptions of bone and cartilage development or the extra-cellular matrix components are in genes *type X collagen* (*COL10A1)* (Nielsen *et al.* 2000), *short stature homoebox (SHOX)* (Rafiti *et al*. 2016)*, cytokine receptor-like factor 2 (CRLF2)* (Rafiti *et al.* 2016)*, aggrecan (ACAN)* (Harper *et al*. 1998; Cavanagh *et al*. 2007; Eberth 2013; Eberth *et al*. 2018)*, type II collagen (COL2A1)* (Daetwyler *et al*. 2014)*, beta-1,4-galactosyltransferase 7 (B4GALT7)* (Orr *et al*. 2010; Leegwater *et al*. 2016)*, and solute carrier family 13 member 1 (SLC13A1)* (Thompson *et al*. 2005; Zhao *et al*. 2012). Inheritance patterns of these mutations are mostly autosomal recessive with the exceptions of *COL10A1* and *COL2A1* as autosomal dominant. Carriers of *ACAN* mutations for dwarfism in Dexter cattle display a mild form of dwarfism and may be considered as partial dominant (Cavanagh *et al*. 2007).

There may be as many as 400 types of proportionate or disproportionate dwarfisms in humans (Warman *et al*. 2010). Disproportionate dwarfisms or skeletal dysplasias are the most common. Alterations in genes have been described to cause various types of dwarfism in humans. These defects can be in genes responsible for pituitary gland hormones, thyroid gland hormones, hormone production and/or recognition, metabolism, cartilage development and bone growth plate development (Irie *et al*. 1995; Shiang *et al.* 1994; Bellus *et al.* 1995; Hendriks-Stegeman *et al*. 2001; Bottner *et al.* 2004; Pannier *et al.* 2009; Tompson *et al.* 2009; Stattin *et al*. 2010; Bochukova *et al.* 2012; van Mullem *et al.* 2012). Many of the genes mentioned to cause dwarfisms in livestock were previously described as the source for dwarfisms in humans.

A genome-wide association study implicated the gene *ACAN* associated with dwarfism in Miniature horses (Eberth *et al.* 2009). Mutations in *ACAN* cause dwarfism

among people (Tompson *et al.* 2009; Dateki, 2017; Hauer *et al.* 2017) and domestic animals including mice (Watanabe *et al.* 1994), cattle (Cavanagh *et al.* 2007; Struck *et al.* 2018) and chickens (Li *et al.* 1993). In all species, the sequence of the gene and structure of the protein are similar but not identical. In horses, the gene contains 17 exons spanning 65,416 bp on ECA1 in Ecab 3.0 (Primary assembly Chr 1: 95,255,522-95,320,938 reverse strand). A highly conserved large complex VNTR (Variable Number Tandem Repeat) array of 63 bp within exon 12 produces a specific structure for the *aggrecan* protein to function within the extra-cellular matrix (ECM) that forms articular cartilage. This region has significant copy number variability within and between individuals. The highly conserved array of ~2500-4500 bases consist of a 63-nucleotide tandem repeat array (unique sequences of 31 nucleotides and 32 nucleotides in tandem) which is sequentially repeated 39-72 times. This structure of nucleotides makes using Sanger short read sequencing results unreliable.

Chondrodysplastic dwarfism in Miniature horses was associated with four variants identified in exons of *ACAN* (Eberth 2013). A deletion of a single base in exons 3 (*D1*) caused a frameshift and presumed loss of function of the protein and causes lethality in homozygous form. A missense mutation in exon 7 (*D2*) is a single base substitution causing changes in amino acid. The exon 7 (*D2*) codes for the G1 binding domain of *ACAN* and the mutation is presumed to cause a decrease in function of the intricate binding interaction of link protein and G1 domain. Another mutation is a deletion of 21 bases in exon 15 (*D4*) causing a loss of 7 amino acids in the resulting protein. Initially the variant described as *D3* was reported and presumed to be a single base deletion in exon 12 associated with dwarfism (Eberth 2013). Subsequently, that conclusion was found to be incorrect and a variant in exon 8 was identified, associated with that class of dwarfism and designated *D3** (Eberth *et al.* 2018). This topic is the subject of Chapter II in this dissertation. Homozygosity or compound heterozygotes for any of the four variants designated *D1, D2, D3 and D4,* resulted in dwarfism and segregated independently. Table 1.1 describes the four mutations as annotated in Ecab 3.0.

Table 1.1. *ACAN* variants associated with dwarfism phenotypes in Miniature horses.

Allele	Phenotype	Variant Position (Ecab 3.0)	Exon	cDNAVariants ENSECAT00000040213.1	Protein Variant ENSECAP00000032890	dbSNP#
Acan-N	normal	none				
Acan-D1	dwarf	95,291,270	3	c.1075A>del	p.K82fx	ss2137497771
Acan-D2	dwarf	95,284,530	7	c.2100G>A	p.V424M	ss2137497770
*Acan-D3**	dwarf	95,282,140	8	c.2343G>C	p.A505P	ss2137497769
Acan-D4	dwarf	95,257,479-95,257,499	15	c.8463-8483del	p.F2545-C2551del	ss2137497768

The variant designated *D3** was first reported by Metzger *et al.* (2016). They identified a variant in *ACAN* (ECA1: g.95,282,140G>C of Ecab 3.0) associated with dwarfism in American Shetland Ponies. Since the Miniature foundation stock includes American Shetland Ponies and today many Ponies which mature at or under the height limit of 34" or 38" are also registered as Miniature horses, it is not a surprise the variant discovered by Metzger would be present in the Miniature population.

Dwarfism in horses as the result of variants in other genes

Rafati and co-workers (2016) described a dwarfism condition in Shetland Ponies referred to as skeletal atavism with a recessive mode of inheritance. This dwarfism, as shown in Figure 1.3, is characterized by altered development of the distal limb bones. Complete development of fibula and ulna results in splayed legs affecting locomotion. A malformed pelvic limb, short femora and tibia are also observed in affected individuals, unlike chondrodysplasia-like dwarfism described in Miniature horses. Their work described two deletions, one of ~160-180kb encompassing the entire coding region of the *Short Stature Homeobox (SHOX)* gene encoding, a protein known to be involved in skeletal development, and a partial deletion of the gene *cytokine receptor-like factor 2* (*CRLF2*) located downstream of *SHOX.* The second deletion of ~60-80kb includes downstream coding elements for *SHOX* and the partial deletion of *CRFL2*. Dwarfs displayed homozygosity and compound heterozygosity for these deletions. Allele frequencies of *Del-1* and *Del-2* were estimated to be 0.061 and 0.012, respectively within the Swedish Shetland pony population. Diseased Miniatures displaying the same dwarf phenotype were tested and found to possess the same deletions seen in the Swedish Shetland ponies, but allele frequency is unknown in Miniature horses.

Leegwater and co-workers (2016) described a splice site mutation in *beta, 4 galactosyltransferase 7 (B4GALT7)* causing disproportionate dwarfism among Fresian horses that also had a recessive mode of inheritance as seen in Figure 1.4. A missense mutation ECA14 g.4535550C>T in Ecab 2.0 causes an amino acid substitution (p.Arg17Lys). The mutation occurs at the last nucleotide of the first exon which impedes proper splicing of the primary transcript. As a result, the level of B4GALT7 mRNA from a dwarf is only 2% compared to normal levels. The protein is one of the enzymes that synthesize a linker between protein and glycosaminoglycan moieties of proteoglycans within the extracellular matrix. Mutations in *B4GALT7* in humans are associated with growth retardation and ligament laxity as common manifestations (Mosher *et al*. 2019).

Figure 1.3. Adult Miniature horse with skeletal atavism.

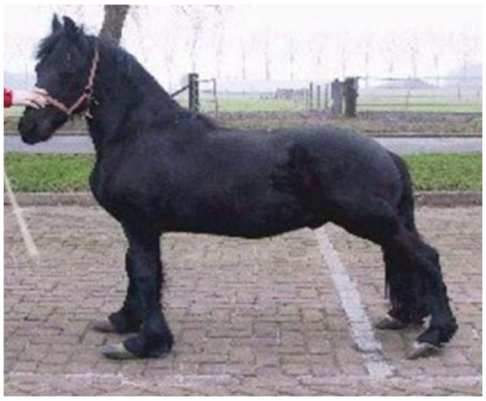


Figure 1.4. Fresian dwarf from Leegwater *et al.* 2016.

Evidence for the presence of yet other variants of ACAN causing dwarfism

Eberth *et al.* (2018) reported 7 phenotypic dwarf Miniature horses possessing only a single copy of a dwarf associated *ACAN* variant and 1 chondrodysplastic dwarf with none of the variants, suggesting the presence of additional variants that cause dwarfism. Table 1.2 lists phenotypes and genotypes for horses in that study.

Table 1.2. *ACAN* genotypes for 196 AMHA horses including 138 normal individuals, 41 chondrodysplastic dwarfs and 17 aborted dwarf fetuses from Eberth *et al.* (2018).

	Phenotypes		
Genotypes	**Term Live Dwarf**	**Aborted Dwarf Fetus**	**Normal**
N/N	1	0	60
N/D1	0	0	27
N/D2	4	0	45
*N/D3**	1	0	4
N/D4	2	0	2
D1/D1	0	3	0
D1/D2	0	9	0
*D1/D3**	0	2	0
D1/D4	1	1	0
D2/D2	17	1	0
*D2/D3**	8	1	0
D2/D4	5	0	0
D3/D3**	0	0	0
D3/D4*	2	0	0
D4/D4	0	0	0
Totals	**41**	**17**	**138**

Andrade *et al.* (2020) reported a haplotype associated with dwarfism in Miniature horses of Brazil (Ecab 3.0 ECA1: g.95271115A > T) in *ACAN*. Genetic testing of a Brazil Miniature population showed that none had more than one of the dwarf alleles *D1*, *D2*, *D3**, *D4* or g.95271115A > T. Investigation by Andrade and co-workers into breeds of large horses revealed the presence of g.95271115A > T including normal phenotyped homozygotes. This suggests the dwarf haplotype is in linkage disequilibrium with the causative mutation yet to be discovered. Here within this dissertation this haplotype will be referred to as the Brazil haplotype to simplify its reference due to it being non-causative. The Brazil variant was not listed as a SNP in Eberth (2013) thesis within the American Miniature horse population used in that study. In retrospect, none of the horses sequenced in that study possessed the variant associated with the Brazil haplotype. The Miniature horses of Brazil may have experienced unique genetic pressures and other equine breed influences which may have caused this variant to become prevalent. A Miniature horse population in the U.S. was investigated for the presence of the Brazil haplotype. This is the subject of Chapter III in this dissertation.

Chondrodysplastic dwarfs continue to be discovered from Miniature horses previously tested as non-carriers for dwarf mutations *D1-D4*. Regions of *ACAN* previously not explored are prime candidates to harbor unknown mutations such as the VNTR, UTRs

and non-coding regions of the gene. Whole genome sequencing is the most informative and cost-effective approach to provide those sequence details.

Computational tools for prediction of consequence from a sequence variant are used extensively in genetic disease research. Numerous web-based tools are freely available which use computational algorithms to predict the effect of nucleotide or amino acid substitutions on protein function. The list includes, VEP (McLaren 2016); MutPred (Li *et al.* 2009); nsSNPAnalyzer (Bao *et al.* 2005); PolyPhen-1 (PPH-1) (Ramensky *et al.* 2002); PolyPhen-2 (PPH-2) (Adzhubei *et al.* 2010); SNAP (Bromberg *et al.* 2007); MAPP (Stone *et al.* 2005); PANTHER (Thomas *et al.* 2004); PhD-SNP (Capriotti *et al.* 2006); SIFT (Ng *et al.* 2003); and SNPs&GO (Calabrese *et al.* 2009).

The objective of these tools is to predict a specific substitution as neutral or deleterious by relying on a variety of computational parameters with basis from studies in evolutionary, structural, physiology or chemical characteristics of a protein or parts of a protein. Utilizing machine learning techniques or artificial intelligence, these tools develop the probability predictions centered from training on datasets with annotated mutations. Flaws in dataset usage have been shown to result in over-fitness of these tools giving illegitimate high-performance estimates. This causes significant obstacles for researchers to have unbiased comparisons of an individual tool's predictive power. Most of these tools have been developed using different analysis principles, datasets, and machine learning methods. Some browser tools have been combined to give a single consensus prediction with improved accuracy, which have resulted in a few consensus classifiers such as CONDEL (González-Pérez *et al.* 2011); PON-P (Olatubosun *et al.* 2012); and Meta-SNP (Capriotti *et al.* 2013); which perform better than individual tools. However, PredictSNP consensus classifier (Bendl *et al.* 2014), trained from fully independent datasets, outperforms other consensus classifiers. It utilizes eight web-based tools but the best performing six results are presented in a table format. Predictions from the individual computational tools are supplemented by annotations from Protein Mutant Database, (Kawabata *et al.* 1999) and UniProt database (Apweiler *et al.* 2004).

Numerous *in silico* tools exist for interpreting WGS data to reveal the existence and prevalence of VNTR regions (Warburton *et al*. 2008; Bongartz 2019; Chu *et al*. 2018; Liao *et al*. 2023; Ikemoto *et al*. 2023). However, inquiry of copy number variation to assess for deleterious effects on a mature protein using *in silico* prediction (whether in coding or non-coding regions) is yet to be available. Only association of copy number variation to a phenotype through genetic studies of multiple individuals can reveal these occurrences as significant. Evaluation is especially difficult with large complex VNTR arrays with variable copy number haplotypes within and among many individuals. The most apparent and simplest method is to evaluate repeat size and sequence consistency or alterations and reading frame retention by hand. In cases of copy number variation in the VNTR region of *ACAN*, repeat size uniformity and reading frame was assessed by hand using Ecab 3.0 in Ensembl. INDEL variants of less than 6 bases were assessed by VEP web tool in Ensembl.

Investigation for new ACAN dwarf variants

A total of 19 dwarfs of unknown parentage were collected possessing unknown alleles. Two dwarfs chosen for WGS possessed one copy of the *D2* variant and an unknown dwarf allele. The choice of two dwarfs possessing the D2 variant was to assess a more complete haplotype for the *D2* allele but at the same time minimize potential confounding variants from other haplotypes due to the possibility of the existence of more than one unknown dwarf allele. This rationale was also to confirm variant g.95284530 C>T, aa. V424M as the cause for the *D2* phenotype. *In silico* SNP prediction analysis calculates the *D2* variant to have a neutral effect to possibly deleterious. If a different variant in a region previously unexamined is the true cause of the *D2* phenotype, WGS would likely provide possible answers with the two samples chosen. This is the subject of Chapter IV in this dissertation.

Phenotype variation associated with dwarf carrier status in Miniature horses

Eberth *et al.* (2018) calculated the carrier rate in Miniature horses for *D1-D4* at approximately 26%. This was notably high for a disease trait which had severe physical abnormalities and significantly shorted life span for the dwarfs. It was surmised it was

possible a physical characteristic expressed by carriers was desirable to breeders. Subsequently, Graves *et al.* (2020) showed adult carriers of *ACAN* dwarf mutations in Miniature horses have a statistically significant shorter height non-diseased phenotype than non-carriers which suggests the mutations display some amount of partial dominance. Small mature height is a highly desired characteristic to Miniature breeders. However, this height difference of 1.43 inches was small in relation to all other features of Miniatures. A slight reduction in height may not be a strong enough selection criteria to drive breeders to use carriers more than non-carriers if that is the only characteristic affected by a dwarf allele. In Andrade *et al.* (2020) the carrier rate of the Brazil population including the Brazil haplotype for dwarf alleles was estimated to be 47.7%. High recessive disease carrier rates, as seen in Miniatures world-wide, may be due to those carriers also expressing other characteristics partially affected by the recessive variant which are desired and selected for breeding purposes.

The foundation of the American Miniature horse began in the mid-20th century by breeders of American Shetland ponies (Naviaux 1999). Not surprisingly, the pedigrees of Miniatures have substantial evidence of American Shetland Pony influence. The American Shetland phenotype is not considered as refined or delicate in the shape or size of the head when compared to a Miniature horse phenotype. Miniatures have been selected to have a more refined, exotic, or dished head. Ideal breed standard descriptions in registries of Miniature horses use terms such as "dished" or "Arabian type" as descriptors for the desired shape and appearance of the head (AMHA 2023; AMHR 2023).

Selection for specific characteristics is a well-established technique when breeding domesticated animals which results in unique phenotypes. Specialized physical features of an animal may be due to a genetic mutation which produces a desired characteristic but also a diseased state. When this mutation is expressed in homozygous form it results in an extreme version of the phenotype which may have severe abnormalities or be lethal. Some well described examples in horses would be HYPP in Quarter horses and the overo coat color. Heterozygotes of both diseases display a desired phenotype, muscle hypertrophy in Quarter horses and a specific white color pattern in overo horses. In homozygous form,

HYPP causes severe muscle paralysis which can lead to death, and homozygous overo results in completely white foal with incomplete nerve development in the lower GI tract which is lethal (Rudolph *et al.* 1992; Metallinos *et al.* 1998; Santschi *et al.* 1998; Yan *et al.* 1998).

Dwarfism affects height significantly due to abnormal long bone growth, but other bones responsible for various physical characteristics such as the head can be significantly altered as well. *Aggrecan* is a constitutive component of the extra-cellular matrix comprising cartilage which matures into bone (Kiani *et al.* 2002). In dwarfisms which cause changes in bone structural composition, these diseases manifest as changes in morphological shape and proportion of the individual. Morphometric measuring techniques are used extensively in paleontology and pathology to describe an organism's phenotype and to distinguish disease. The Landmark based method is used extensively in human fetal and newborn examinations to look for early signs of growth abnormalities. Today by ultrasound, Landmark based measurement guidelines are used in humans to determine the existence of dwarfism and disproportionate physical characteristics of the neonatal foetus (Krakow *et al.* 2009).

The Miniature breed has undergone significant selection pressure for a more exotic, dished face since its foundation. Considering Graves and co-workers (2020) found mature height to be affected by dwarf variants in *ACAN*, could other structures such as the head be affected by these mutations in a carrier? This is the subject of Chapter V in this dissertation.

ACAN and skeletal soundness

Diseases of the musculoskeletal system can be complex and multigenic. Environmental factors such as age, occupation, nutritional state, weight, and participation in activities can also have significant effect on musculoskeletal competence. These factors alone or in combination challenge the cushioning effect of cartilage (Battié and Videman, 2006). Cartilage failure presented as Lumbar disc degeneration among people may be caused by a lower VNTR polymorphism in CS1 (Mashayekhi *et al.* 2010; Kawaguchi *et*

al. 1999; Cong *et al.* 2010; Solovieva *et al.* 2007). In people the CS1 domain consists of a series of 13-34 tandem repeats of 57 base-pairs each encoding 19 amino acids (Doege *et al.* 1991; Doege *et al.* 1997). The most frequent number of repeats is 26 in people. The lower numbers of repeats were associated with pathology. This evidence alone cannot be interpreted as disease variants in ACAN. Other genetic variants of different genes may also be involved in pathology and show statistical significance with association to disc degeneration (Mayer *et al.* 2013).

Musculoskeletal injuries are the predominant causes of death in racehorses. In Thoroughbreds they account for approximately 83% of deaths within the racing industry worldwide (Samol *et al.* 2021). Considerable interest has been focused on the traditional timetable utilized by trainers for young Thoroughbreds and their racing career longevity. Typical training timeframe starts at age 2 with peak racing ability during year 4, however, some may race to the age of 10 (Gramm *et al.* 2010). Repetitious peak physical stress at such a young age may be detrimental to long term racing success and increase the chance of injury (Crawford *et al.* 2021). For decades regulatory and racing industry supported groups as well as social factions have expressed concern over catastrophic injuries in Thoroughbred horse racing (Heleski *et al.* 2020). As early as 1990, breeders were questioning the possibility of perpetuating inferior genetics due to advances in medications allowing injury prone horses to continue to perform (Hovdey 1999). Increased restrictions of legal therapeutic drug use by individual states in the US and countries worldwide significantly decreased overall drug use. Individual country as well as world-wide meta-analysis for risk factors associated with breakdowns showed that track surface condition may be more significant than surface type (Maeda *et al.* 2016; Hitchens *et al.* 2019). These efforts to understand the causes of catastrophic events and mitigate their occurrences have not eliminated the problem.

Eberth *et al.* (2018) sequenced exons only of *ACAN*; however, the highly conserved large complex VNTR array of exon 12 did not produce reliable results using Sanger short read sequencing. Long range PCR amplification was performed to determine the extent of variability. The results showed significant copy number variability within and between

individuals. The region of ~2300-4500 base pairs comprise a 63-nucleotide tandem repeat that is sequentially repeated 36-72 times. The VNTR produces a specific protein structure for its function within the extra-cellular matrix in cartilage (Figure 1.5).

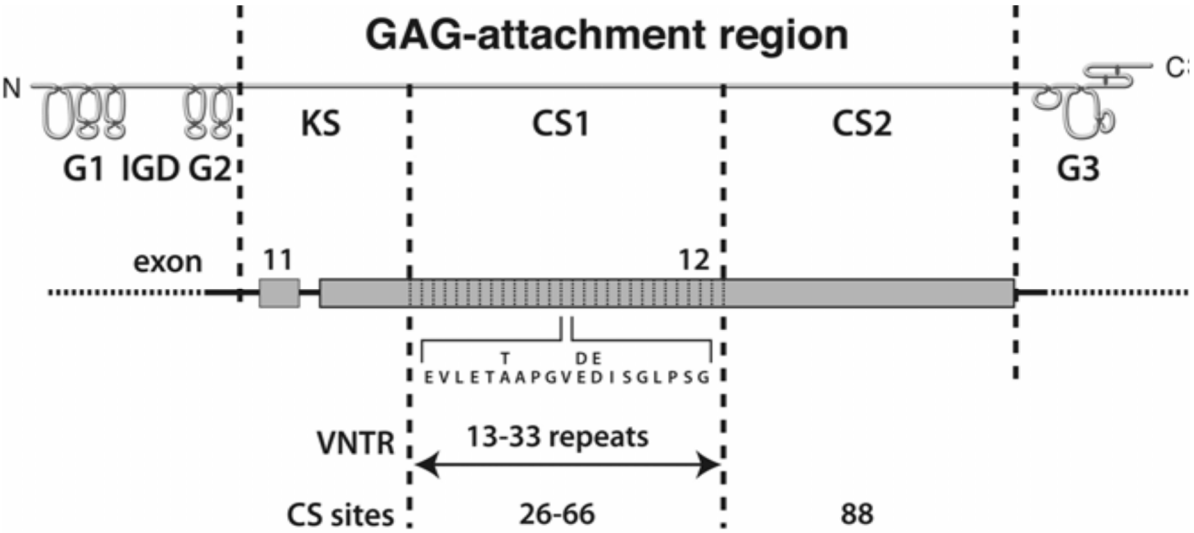


Figure 1.5. *Aggrecan* structure in relation to the variable number of tandem repeat (VNTR) polymorphism repeat of exon 12.

The essential goal in breeding racehorses is to select for early maturation and precocity while simultaneously choosing a lighter thus more fragile body frame. Consequently, the most competitive horses would be those with the highest risk of musculoskeletal damage and require increased time between competitions. This result would exist regardless of the environmental conditions mentioned above if breeders were unknowingly perpetuating an inferior heritable condition for racing. Considering the evidence in people and associations with pathology, do polymorphisms in *aggrecan* predispose horses to incompetent cartilage that experiences the extreme physical stress of racing? This is the subject of Chapter VI in this dissertation.

Aggrecan structure and function

These investigations concern the gene *aggrecan* (*ACAN),* a large proteoglycan essential to the formation of articular cartilage. Therefore, a brief overview of the genetics and biology of *ACAN* is pertinent to this introduction. The *ACAN* core protein is composed

of three globular domains designated G1, G2 and G3. Between these domains several keratin sulphate (KS) chains are attached and a large region of chondroitin sulphate chains (CS) between the G2 and G3 domains allowing post-translational glycosaminoglycan (GAG) side chain attachment. When forming the extra cellular matrix (ECM), *ACAN* interacts with link protein, a small glycoprotein which has a similar structural design to G1 and helps stabilize the overall ECM structure formation. This structure similarity allows the G1 globular domain to bond to hyaluronan and link protein non-covalently (Heinegård *et al.* 1974). This is part of the tertiary structure complex within the ECM.

Situated between the G2 domain and G3 domain is a large sequence modified by KS side chains and CS chains. Each *ACAN* contains ~ 100 chondroitin sulfate chains but fewer keratan sulfate chains (up to 60). These chains combine to make up 90% of *ACAN*'s mass (Kiani *et al.* 2002). The G2 has tandem repeats which make it homologous to G1 and link protein. Globular domain 2 (G2) is separated from G1 by an interlobular domain (IGD) which is involved in processing (Fosang *et al.* 1989; Paulsson *et al.* 1987). Located on the C-terminus end within the G3 domain are two epidermal growth factor-like sections, EGF1 and EGF2. Other sites include a carbohydrate recognition domain (CRD), complement binding protein (CBP)-like sites, a folded immunoglobulin region and tandem repeats of proteoglycan (Perkins *et al.* 1989). Figure 1.6 shows the globular and domain structures of *ACAN*.

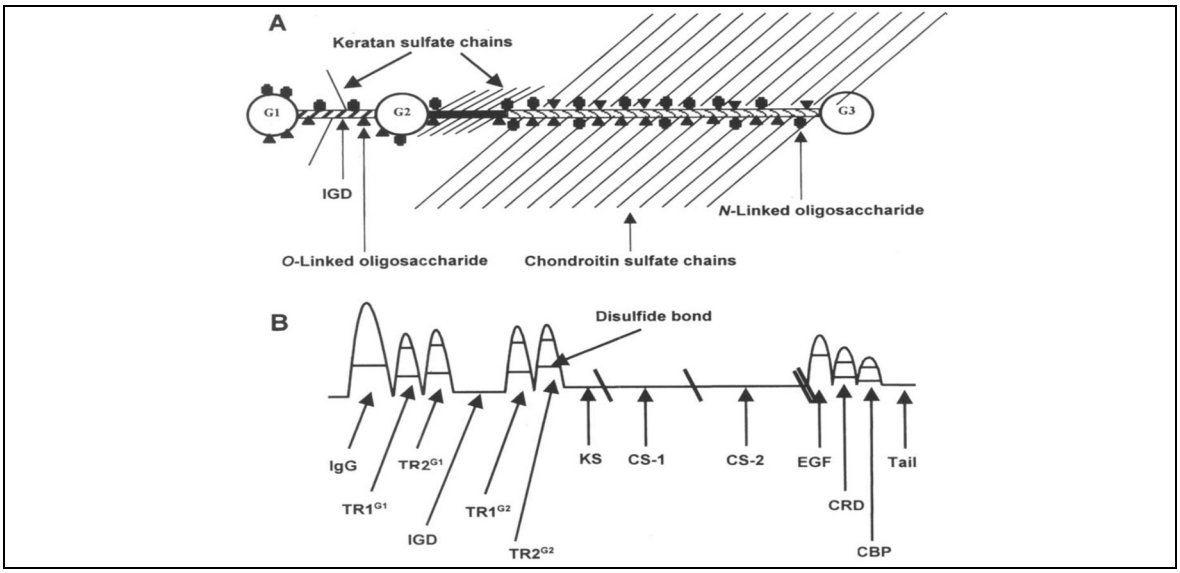


Figure 1.6. *Aggrecan* structure. **A**. Globular protein and attached GAG chain structure. **B**. Protein domain structure. Folded modules: IgG, immunoglobulin fold: TR, tandem repeats: EGF, epidermal growth factor-like module: CRD, carbohydrate recognition domain: CBP, complement binding protein-like module. Extended domains: IGD, interglobular domain: KS, keratan sulfate attachment domain: CS-1 and CS-2, chondroitin sulfate attachment domains (Kiani *et al.* 2002).

DNA sequence conservation is very high within the three globular domains (G1, G2, and G3) of known *ACAN* sequences across species, including humans. The regions of chondroitin sulfate chains and keratan sulfate chains are less conserved across species. Interestingly, there is a correlation between exon number sequence and structure domain layout of the mature *ACAN* protein. Exon 1 encodes an untranslated 5' sequence and exon 2 contains a translation start codon site for mRNA synthesis. The G1 domain is divided into sections of A, B, B' and correspond to subsequent exons 3, 4 - 5, and 6, respectively. Link protein exons 3, 4-5, 6 correspond to *ACAN* G1 sections in sequence similarity confirming binding site affinities (Valhmu *et al.* 1995). The G1 domain is comprised of three modules: an immunoglobulin fold module and two copies of an hyaluronan (HA)-binding link module, this is also referred to as the PTR or proteoglycan tandem repeat. The G1 domain also contains two cysteine-rich motifs that form disulfide bonds which are involved in *ACAN* interactions with HA to form larger complexes (Watanabe *et al.* 1998).

The rod-shaped structure separating the G1 and G2 domains is the interglobulin domain (IGD). The IGD is encoded only by exon 7 with structure and sequence unique in

the *ACAN* molecule. During pathological cartilage degradation in humans, this sequence is the region of proteolytic attack by Aggrecanase-1 and the site of *ACAN* turnover physiologically (Tortorella *et al.* 1999). Matrix metalloproteinases (MMP) and Aggrecanase-1 are involved in *ACAN* turnover in both diseased and normal cartilage (Lark *et al.* 1997).

The G2 domain is encoded by exons 8, 9, 10. There are two tandem repeats of proteoglycan that are about 67% similar in sequence to the G1 PTR repeats, however, it shows no functional ability to bind with Hyaluronan or link protein due to karatan sulfate side chains (Fosang *et al.* 1991). *ACAN* is the only proteoglycan family member that has this G2 sequence and structure domain with keratan sulfate and chondroitin chains to make it highly glycosylated. The G2 domain functions as a quality control domain to produce a mature functional *ACAN* protein (Kiani *et al.* 2001).

Following the G2 domain is the (KS) keratan sulfate domain encoded by exons 11 and 12. The sequence of amino acids in exon 12 make up a group of tandem repeat chains which varies in number greatly across species. Rats, mice, and chicken (Krueger *et al.* 1990) lack this repeat sequence region, however; bovine (Antonsson *et al.* 1989), human (Doege *et al.* 1997) and horse (this dissertation chapter 6) contain these sequences in various copy numbers. The complete function of keratan sulfate chains is not known, however it is suggested that they may contribute to tissue development, processing, and distribution. The one important characteristic of the KS chain is the binding of water molecules which significantly enhances the ECM's ability to handle load bearing stress (Kiani *et al.* 2001).

The chondroitin sulfate (CS) region is *ACAN*'s largest domain solely encoded by exon 12 in humans. The domain includes dipeptide repeats that contain serine-glycine residues and are separated by acidic and hydrophobic residues (Krueger *et al.* 1990). CS chain recognition sites for attachment have been suggested to be S-G-X-G (Bourdon *et al.* 1987) or (D/E)-X-S-G (Krueger *et al.* 1990). These recognition sequence sites are dependent on chaperone proteins and the presence of specific enzymes that are necessary

for post-translational modifications. The chief role of *ACAN* as a structural ECM proteoglycan is to hold water which is accomplished by the large number of negatively charged CS chains. The glycosaminoglycan (GAG) chains are vital in *ACAN* processing modifications because of their similar properties for holding water. This role in *ACAN* processing and secretion has been shown by the addition of GAG chains to the CS region. Without GAG modified CS sequences, CS containing constructs are not secreted (Kiani *et al.* 2001).

GAG chain modification is dependent on the G3 domain. This domain is highly complex involving alternative splicing of exons post-transcriptionally. In human *ACAN,* the G3 domain consists of two alternatively spliced EGF-like domains encoded by exons 13 and 14. Encoded by exons 15, 16, and 17 is a C-type Lectin-like domain, also called LEC-like domain (also referred to as CRD, carbohydrate recognition domain, module). Encoded by exon 18 is an alternatively spliced CRP-like domain (also referred to as CBP, compliment binding protein-like, module) (Doege *et al.* 1991; Fülöp *et al.* 1993). In chickens, the C-type Lectin-like CRD is encoded by exons 14, 15, 16, 17 and the CBP/CRP domain encoded by exon 18 (Kiani *et al.* 2002). Without the G3 domain, *ACAN* cannot be modified by GAG and the unmodified core protein cannot be secreted by the cell (Kiani *et al.* 2001). In overall function of *ACAN* processing, G3 is an important factor in the control of quality *ACAN* secretion and breakdown through degradation pathways. The unique feature of the G3 inter-domain dependent regulation and control is the GAG modification is separate from product secretion and the modification of the GAG matrix formation overrides its effect to the secretion of *ACAN* but not vice-versa (Kiani *et al.* 2002).

Chapter II

Reassignment of *D3* variant for dwarfism to *D3**

Summary

In the *Master Thesis* (Eberth 2103) the variant designated *D3* (Ecab 3.0, chr1 g.95270423 C>del) was reported as a cause of dwarfism in Miniature horses. Subsequent design of a commercial TaqMan® assay and testing of *D3* haplotype dwarfs and a larger, random population demonstrated that none possessed the deletion at the location in exon 12. Samples with the proposed *D3* deletion were re-sequenced and the original results found to be in error. Analysis of exons in *ACAN* of the dwarf horses that possessed only a single known variant (*D1*, *D2* and *D4*) led to identification of a single base missense mutation in exon 8 (Ecab 3.0, chr1: 95,282,140 G>C) associated with a dwarf phenotype. Most of the horses previously identified as possessing *D3* had this variant and, therefore, it was newly designated *D3**. The gene frequency of *D3** was 0.035 based on testing 200 randomly selected Miniature horses. This missense mutation may alter the structural and possibly functional integrity of the interglobular domain encoded by exon 8 involved in the physiological turnover of *aggrecan*. *D3** was reported along with *D1*, *D2* and *D4* (Eberth *et al.* 2018).

Introduction

A genome wide association study implicated *aggrecan* (*ACAN*) as a candidate gene (Eberth 2009). The cause of chondrodysplastic dwarfism in Miniature horses was reported to be due to four recessive genetic variants in exons of *ACAN* that were found to segregate independently (Eberth 2013). Those variants were designated *D1, D2, D3* and *D4*. Horses homozygous for those variants, or compound heterozygotes for those variants, were reported to be nonviable or affected with a form of chondrodysplastic dwarfism. Subsequently, development of a TaqMan® assay for the *D3* variant and testing the carriers of *D3* as well as a random population of miniature horses failed to detect the *D3* variant (unpublished data). This Chapter describes the work to identify a new variant, designated *D3** that was subsequently reported (Eberth *et al.* 2018). The variant known as *D3** was

first reported by Metzger and coworkers (2016) based on discovery of a dwarf Shetland pony found to be homozygous for this variant.

Materials and Methods

Horses

A total of 221 Miniature horses were used in the study including 73 dwarfs, 10 of which were previously identified dwarfs with the *D3* haplotype, 76 normal horses identified as non-carriers of any of the dwarfism *ACAN* variants associated with dwarfism, 72 normal horses that were genotyped as carriers of *D1*, *D2* or *D4.* Additionally, 6 horses from 2 sets of sire-dam-offspring nuclear families that included a dwarf offspring and one parent with a *D1*, *D2* or *D4* allele and another parent possessing an unknown dwarfism causing allele. Also included was a population of 361 Miniature horses received by the Gluck Equine Research Center Genetics Testing Laboratory from horse owners. One horse submitted to the lab from Miniature owners displayed a dwarf phenotype, however the remaining horses displayed a normal phenotype. Other breeds which have not reported the occurrence of dwarf phenotypes, were also tested which included 28 Thoroughbred, 22 American Standardbred, 4 American Saddlebred, 4 Tennessee Walking horses, 4 Arabian, 1 Hackney pony and 1 Caspian horse. IACUC 2022-4161 was approved in connection with this study of experimental animals.

DNA Isolation

DNA from blood or tissue was extracted using Puregene whole blood DNA extraction kits (Gentra Systems Inc., Minneapolis, MN) and Puregene tissue DNA extraction kits according to published protocols. Hair samples submitted were processed using 7-10 hair bulbs according to the method described by Locke *et al.* (2002). The hair bulbs were placed in 100 µl lysis solution containing 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 (ThermoFisher Waltham, MA.) and incubated at 60°C for 45 minutes followed by 95°C for 45 minutes to deactivate the proteinase K. Aliquots of the DNA samples were made for working dilutions at concentrations of approximately 150 ng/µl.

Sequencing

ACAN primers from Eberth (2013) were utilized for this study listed in Table 2.1. 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, 150 ng genomic DNA from hair, blood, or tissue extractions, 0.2 µl FastStart Taq DNA polymerase (Perkin Elmer) and 50 nM of each primer. Each exon template PCR product was quantified on a 2% agarose gel run in 1X TBE solution. PCR amplicons were prepared for sequencing per Eurofins (Louisville KY) requirements and shipped overnight for sequencing.

Table 2.1. Primers used for splice variant ACAN-201 to sequence exons 1-17. Lower case sequences indicate intron location. Upper case indicates location in exons. Exon numbering is according to Ensembl Genome browser on Ecab build 3.0.

Exon #	Forward Primer Sequence	Reverse Primer Sequence
Exon 2	gtgacctttgccctcactgt	gcacccagaatccagtcttc
Exon 3	tgggtggtcctctctagcac	tttccacaggtgaagcaaca
Exon 4	gtgcctgacctgctctatcg	agctcagtgctggtcaacg
Exon 5	cctgagtgtcacatcccactt	gagtggtagtggggtgaagg
Exon 6	gcctgctttgtccttcacag	aaaacagccccctattccac
Exon 7	gggctgagccgctaaagtt	aggccaagttccttccactt
Exon 8	gtctctccttctcgccctct	aagcctgacccttgagactg
Exon 9	agaacaggccctcattctgc	aggtaatgccctctcctcgt
Exon 10	gtgccacctgcctctgtc	cagaagtgggttctggagga
Exon 11	aggaggaaccttcaccacct	cgcccaagccatacgaac
Exon 12a	ggagcagtttctaatcccaca	ACTGAGGTCCTCTGCTCCAG
Exon 12b	TATCTCTGCAGTGGGCTCAG	AGGCAGTGGGCTCTAGATGA
Exon 12c	GAGTGGAGGACCTTGGTGAA	GAGGCAGTGGGCTCTAAATG
Exon 12d	TGCCTCCGGAGTAGAGGAC	TGACGACTTCCACCAATGTC
Exon 12e	GGCAAGCTCCTGAAGCAAGT	TCCAGATGTTGTCCCACTGA
Exon 12f	GTGGACTGTCCTCTGGACAAC	CCCTTCTCCTGCTTCTTGG
Exon 12g	TCCTCTGGAGCTGAGACTGG	ATCAGGGGACCCAGAAGC
Exon 12h	CCCAGCTTGTTGAGTCCAGT	ctgaactacccaaaccccttt
Exon 13	cttcaagcccctgacctgt	ctgctgacttctggcaagtg
Exon 14	cccaaacccacatcttctct	atgccgtccgaatgtatctc
Exon 15	ctgccctctgctcacctct	ggagccgaagtcttgattct
Exon 16	cacaggagccctttctgaag	cagggaggaggaggtgct
Exon 17	ccgagggctcactaggattt	gacgaagtgtcggtgatctg

*Variant analysis of sequence reads for investigation of D3**

ACAN exon sequence files received from Eurofins (Louisville KY) were compared among dwarf and non-dwarf individuals. Sequence data for exon regions was compared to the reference sequence for *Equus caballus* 3.0 (Primary assembly Chr 1: 95,255,522-95,320,938 reverse strand) specific to *ACAN* region, reference from whole genome shotgun sequence using Ensembl exon sequence viewer. Computer programs Vector NTI and Codon-Code Aligner (https://www.codoncode.com/aligner/) were utilized for variant analysis of sequence results.

Custom TaqMan® Assays

Custom TaqMan® assays were designed utilizing ThermoFisher Assay design tool (https://www.thermofisher.com/order/custom-genomic-products/tools/genotyping/) for presumed *D3* variant in exon 12 and the *D3** variant in exon 8 shown in Figure 1.

A. Custom TaqMan® assay Exon 12 variant *D3* VIC C FAM* (T)
CACACCTCNGGNCCANATNTGGNCCTGAGCACCAGCCTCCCAGAGTCTGAGTGGACCCC
GCAGATGCAGCGCC[C/*]TGCAGAGGCGCTCCTAGAAATCGAGTCCTCAAGCCCCCTG
TACTCANGAGAAGAGACCCCAACAGCCGAAACAGCCGTCTCCCCGACAGAGGCTTCCAT
CCCAGCT

B. Custom TaqMan® assay Exon 8 variant *D3** VIC G FAM C
GTGTTGTGTTCCACTACCGCCCGGGCTCCGACCGATACTCTCTGACCTTTGAGGAGGCA
CAGCAG[G/C]CCTGCCTGCNNACNGNGGCGGTCATNGCCTCGCCNGAGCAGCTCCAGG
CCGCCTACGAAGCGGGCTACGAGCAGTGTGACGCCGGCTGGCTGAGTGACCAGACCGTC
AG

Figure 2.1. **A.** Custom TaqMan® sequence submitted ThermoFisher to distinguish proposed *D3* mutation. **B.** Custom TaqMan® sequence submitted to ThermoFisher to distinguish *D3** mutation.

Custom Primers for Sanger sequencing Exon 12 variant

A smaller region of exon 12 in *ACAN* was chosen to design custom primers to sequence the area surrounding variant g.95270423 C>del. Figure 2 illustrates the genomic location with the proposed *D3* mutation highlighted in red. Sequencing primers were designed using Primer3plus online design tool https://www.primer3plus.com/index.html. Primers are highlighted in yellow and listed.

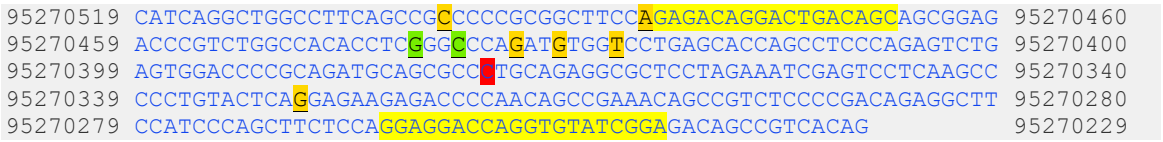


Figure 2.2. Genomic location of proposed *D3* mutation highlighted in red. Forward and Reverse primers listed. Forward Primer GAGACAGGACTGACAGCAGC. Reverse Primer TCCGATACACCTGGTCCTCC.

RESULTS

TaqMan® Assay results for D3 in exon 12

Eberth (2013) reported 10 dwarf horses possessing the *D3* variant (Table 1). A TaqMan® assay was designed to test the variant *D3* in a large, random population of miniature horses as well as the entire set of horses assembled to investigate dwarfism in miniature horses (Eberth, 2013; Eberth, this study 2023). None of the horses were found to possess this variant, even though 10 had previously been identified as possessing a deletion based on Sanger sequencing (Tables 2.2, 2.3).

Table 2.2. Genotypes of horses identified possessing *D3* variant and results from TaqMan® assay.

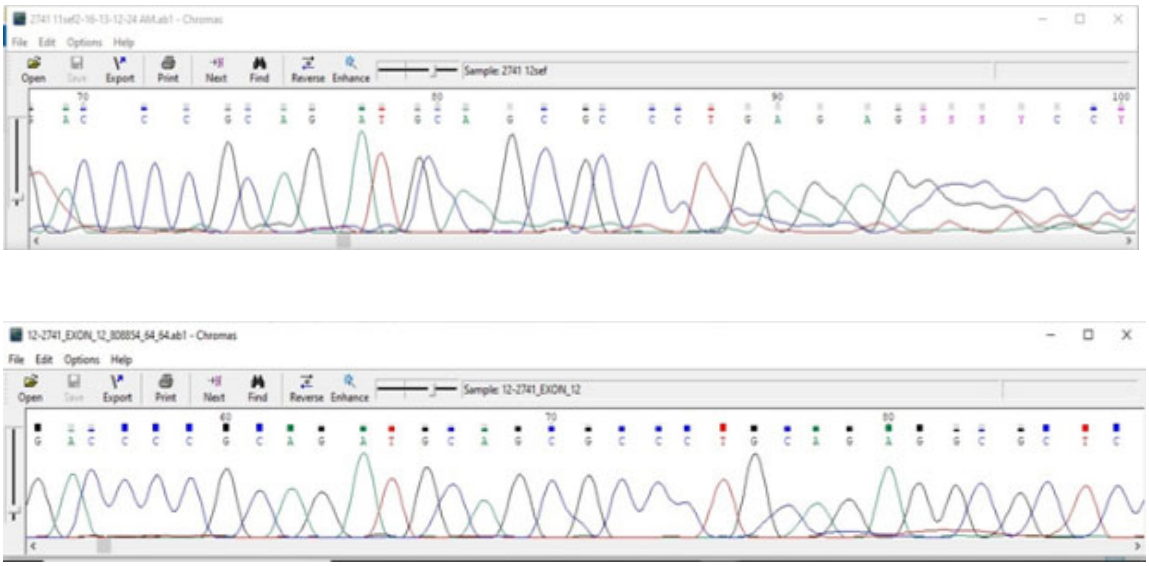
Acc. #	Phenotype	Genotype from DNA sequence	Genotype after TaqMan® for *D3*
2673	Dwarf	*D2/D3*	*D2/N*
2674	Dwarf	*D2/D3*	*D2/N*
2695	Dwarf	*D2/D3*	*D2/N*
2697	Dwarf	*D2/D3*	*D2/N*
2702	Dwarf	*D2/D3*	*D2/N*
2703	Dwarf	*D2/D3*	*D2/N*
2710	Dwarf	*D2/D3*	*D2/N*
2741	Dwarf	*D4/D3*	*D4/N*
2742	Dwarf	N/*D3*	*N/N*
2747	Dwarf	*D2/D3*	*D4/N*

Table 2.3. Results for *D3* TaqMan® assay on 221 Miniature horses.

Genotypes	Population size	*D3* TaqMan® results
Non-carriers, non-dwarf	76	0
Non-dwarf carriers D1, D2,D4	72	0
N/N dwarf	1	0
N/D1 dwarf	0	0
N/D2 dwarf	4	0
N/D3 dwarf*	1	0
N/D4 dwarf	2	0
D1/D1	3	0
D1/D2	9	0
D1/D3	1	0
D1/D4	2	0
D2/D2	31	0
D2/D3	8	0
D2/D4	10	0
D3/D4	0	0
D4/D4	1	0
Totals	**221**	**0**

Sequencing Artifact

Resequencing for the region on exon 12 presumed to harbor the deletion was conducted on 10 horses (Acc.# 2673, 2674, 2695, 2697, 2702, 2703, 2710, 2741, 2742,2745). Figure 2.3A and B shows sequence comparisons from Eberth (2013) (Figure 2.3A) and the new sequence (Figure 2.3B). As can be seen in Figure 2.3A, the sequence became unreadable at position 90 and this was observed in all the samples identified as possessing variant *D3*. This was interpreted as a deletion leading to a frameshift in the sequence. When the amplicons were re-sequenced following the results from the TaqMan® assay, the sequence was well rendered through this region, as can be seen in Figure 2.3B.



Figures 2.3A 2.3B. Sequence reads of exon 12 from a compound heterozygous dwarf. **Top 2.3A** sequence is forward from Eberth (2013) and the **Bottom 2.3B** Sequence read from Eurofins of the same dwarf acc. #2741.

Re-examination of Variants from Eberth 2013

Exon variants identified by Sanger sequencing in the horses identified with *D3* in the initial (Eberth, 2013) study were re-examined. Table 2.4 lists the variants reported for the horses listed in Table 2.2. Of note in the Table is the variant in exon 8 g.95,282,140 G>C in ECAB 3.0. This variant was overlooked because it appeared on the haplotype thought to possess the deletion variant characterized as *D3*. At the population level, the *D3* variant appeared to explain the cases of dwarfism clearly due to a deleted nucleotide and the variant in exon 8 was disregarded as deleterious. Variant g. 95,256,565 G > C was not present in two dwarfs with sequence reads (#2676 and #2745) and was not present in the two nuclear sire-dam-offspring families in Table 2.5.

Table 2.4. Exon variants identified with *D3* from Eberth (2013) present in the 10 dwarf samples.

Allele	Position	Exon	cDNA	rs#	Protein	SIFT score
ACAN-N	95,291,302	3	c.1155C>T	rs395927284	p.P71P	syn
ACAN-N	95,286,209	6	c.1809G>A	rs395462341	p.A289A	syn
ACAN-N	95,284,551	7	c.2191G>C	rs68461203	p.P417A	0.59 tolerated
ACAN-N	95,284,467	7	c.2275C>T	rs393751006	p.V445I	syn
ACAN-N	95,282,140	8	c.2455C>G	rs1095048823	p.A505P	0 deleterious
ACAN-N	95,282,129	8	c.2466C>G	rs3433087105	p.A508P	0.11 tolerated
ACAN-N	95,275,174	12	c.3348C>T	rs1141098330	p.T802T	syn
ACAN-N	95,274,967	12	c.3555C>T	rs11399992128	p.G871G	syn
ACAN-N	95,274,878	12	c.3644G>A	rs394162401	p.A901V	1 tolerated
ACAN-N	95,271,190	12	c.7332C>T	rs3443782452	p.S2130S	syn
ACAN-N	95,270,484	12	c.8038T>C	rs684661199	p.R2366G	0.51 tolerated
ACAN-N	95,270,440	12	c.8082C>T	rs1149835638	p.S2380S	syn
ACAN-N	95,256,565	16	c.7412G>C		p.C2582S	0 deleterious
ACAN-N	95,256,474	16	c.8778G>A	rs1147286743	p.T2612T	syn

Analysis of sire-dam-offspring sets

Resequencing was conducted on 10 horses from the original study and for two families (sire-dam and offspring) segregating for dwarfism but possessing an unknown dwarfism variant. The results confirmed identification and presence of the exon 8 variant, g.95,282,140 G>C. This variant was designated *D3**. Table 2.5 shows results from testing two families segregating for dwarfism including the *D3** variant.

Table 2.5. Genotype results for two sets of nuclear families for *D1, D2, D3** and *D4*.

Family 1	TaqMan® results
Sire #2677	*N/D3**
Dam #2675	*N/D4*
dwarf foal#2676	*D4/D3**
Family 2	
Sire #2736	*N/D3**
Dam #2737	*N/D1*
Dwarf foal#2683	*D1/D3**
Samples Previously *N/N*	
Acc# 2814	*N/D3**
Acc# 2830	*N/D3**

D3 sequence analysis in exon 8*

Figure 2.4AB below shows sequence reads of exon 8 and the location of g.95282140 G>C missense variant in ChromasPro sequence viewer. Top sequence (Figure 2.4A) is dwarf Acc# 2676 of sire-dam-offspring set from Table 2.5. Dwarf #2676 is heterozygous for g.95282140 G>C with a complete genotype of *D3*/D4*. The bottom sequence (Figure 2.4B) is the dam (Acc# 2675) of dwarf #2676. Dam #2675 was genotyped N/D4 and possesses the WT variant at g.95282140.

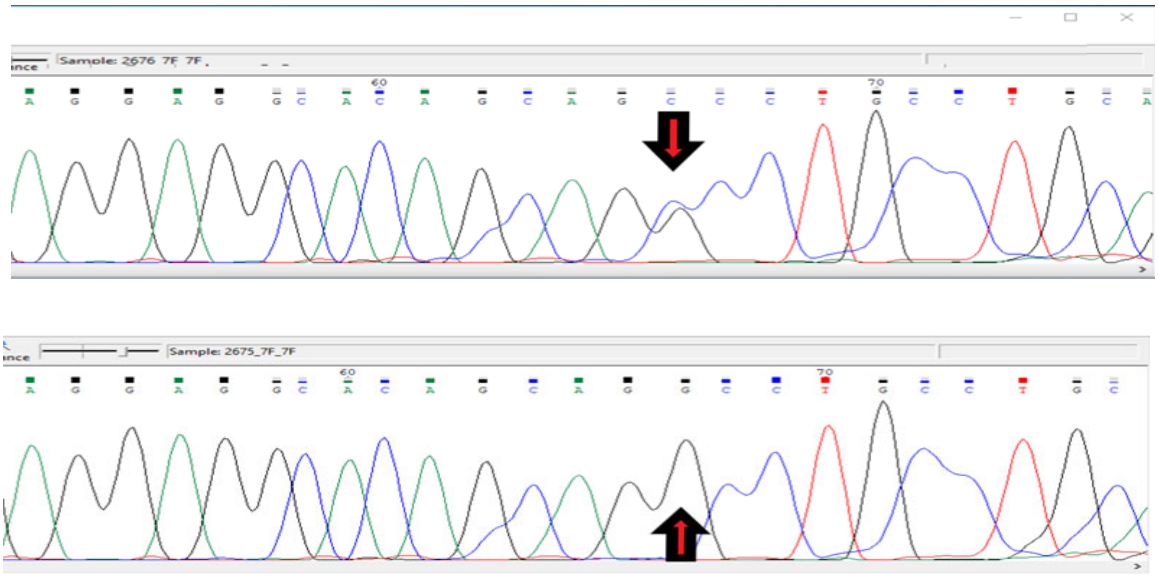


Figure 2.4A 2.4B. Sequence reads of exon 8 of rs1095048823 variant. **2.4A** (top sequence) from *D3*/D4* dwarf (Acc. #2676). **2.4B** (bottom sequence) from Dam (Acc. #2675) of dwarf #2676. Dam #2675 did not possess rs1095048823 and was genotyped *N/D4.*

Exon 8 D3 genotype variant in-silico analysis*

SIFT analysis (Kumar *et al.* 2009) was performed by Variant Effect Predictor (VEP) online tool within the Ensembl genome database http://www.ensembl.org/info/docs/tools/vep/index.html. SIFT score for g.95282140 G>C resulted in a value of 0 or "deleterious". PredictSNP internet tool (Bendl *et al.* 2014), a consensus classifier for SNP variant effects on protein function https://loschmidt.chemi.muni.cz/predictsnp/ predicted the missense variant with an 87% probability of causing a "deleterious" effect. This consensus analysis included five other

independent analysis tools that all predicted the missense mutation to be deleterious, with PhD-SNP 88%, PolyPhen-1 74%, PolyPhen-2 81%, SIFT 79% and SNAP 72% shown in Table 2.6. The single base missense mutation produces a change of amino acid alanine to proline. This may impact the structural and possibly functional integrity of the interglobular domain encoded by exon 8 which is involved in the physiological turnover of the *aggrecan* molecule.

Table 2.6. *In silico* analysis results for g.95282140 G>C

Analysis	Location	Mutation	Amino Acid	Score	Probability
SIFT	g.95282140	G > C	A505P	0 Deleterious	79%
PredictSNP	g.95282140	G > C	A505P	Deleterious	87%
PhD-SNP	g.95282140	G > C	A505P	Deleterious	88%
PolyPhen-1	g.95282140	G > C	A505P	Deleterious	74%
PolyPhen-2	g.95282140	G > C	A505P	0.9 Deleterious	81%
SNAP	g.95282140	G > C	A505P	Deleterious	72%

TaqMan® results of D3 dwarfs from Eberth (2013)*

A custom TaqMan® assay for exon 8 variant g.95,282,140 G>C (Figure 2.1) was designed to test the dwarf individuals which possessed the putative *D3* deletion as well as other dwarf individuals which possessed an unknown variant within the study population. Table 2.7 identifies the horses identified previously identified with *D3* and/or which were subsequently identified with the variant exon 8 variant g.95,282,140 G>C.

Table 2.7. Horses Reported in Eberth (2013) and Eberth *et al.* (2018) with TaqMan® assay results for g.95,282,140 G>C, (*D3*)*.

Acc. #	Phenotype	Eberth 2013 exon 12	Eberth 2018 exon 8
2673	Dwarf	*D2/D3*	*D2/D3**
2674	Dwarf	*D2/D3*	*D2/D3**
2676	Dwarf	*D4/N*	*D4/D3**
2683	Dwarf	*D1/N*	*D1/D3**
2694	Dwarf	*D2/N*	*D2/N*
2695	Dwarf	*D2/D3*	*D2/D3**
2697	Dwarf	*D2/D3*	*D2/D3**
2698	Dwarf	*N/N*	*N/N*
2701	Dwarf	*D2/N*	*D2/N*
2702	Dwarf	*D2/D3*	*D2/D3**
2703	Dwarf	*D2/D3*	*D2/D3**
2710	Dwarf	*D2/D3*	*D2/D3**
2741	Dwarf	*D4/D3*	*D4/N*
2742	Dwarf	*D3/N*	*D3*/N*
2745	Dwarf	*D4/D3*	*D4/N*
2746	Dwarf	*D2/N*	*D2/N*
2747	Dwarf	*D2/N*	*D2/D3**
2906	Dwarf	*D2/N*	*D2/N*
2911	Dwarf	*D2/N*	*D2/D3**

In genotype summation Table 2.8, none of the 74 non-Dwarf, Miniature horses which carried a copy of *D1*, *D2* or *D4* possessed the exon 8 missense variant, g.95,282,140 G>C. Four of the 78 non-carriers possessed one copy of the missense variant. SNP g.95,282,140 G>C showed independent segregation from dwarf variants *D1*, *D2*, *D4*. Table 2.8 includes the two nuclear family sets indicating genotype and TaqMan® results.

Table 2.8. TaqMan® results for 227 Miniatures for SNP g.95,282,140 G>C. This table includes the two nuclear families.

Genotypes/phenotypes	Count	*D3** TaqMan®
Non-carrier N/N	74	0
Carriers, D3/N, Previously N/N*	4	4 D3*/N
Carriers D1/N, D2/N, ,D4/N	74	0
N/N dwarf	1	0
N/D1 dwarf	0	0
N/D2 dwarf	4	0
N/D3 dwarf*	1	1 *N/D3**
N/D4 dwarf	2	0
D1/D1 dwarf	3	0
D1/D2dwarf	9	0
*D1/D3*dwarf*	2	2 *D1/D3**
D1/D4 dwarf	2	0
D2/D2 dwarf	31	0
D2/D3 dwarf*	8	8 *D2/D3**
D2/D4dwarf	10	0
D3/D4 dwarf*	1	1 *D3*/D4*
D4/D4 dwarf	1	0
Totals	**227**	**16**

Genotyping of other horse breeds

No samples from the large breed population of 64 horses possessed the missense variant which included 28 Thoroughbred, 22 American Standardbred, 4 American Saddlebred, 4 Tennessee Walking horses, 4 Arabian, 1 Hackney pony and 1 Caspian horse, shown in Table 2.9. Genotyping results presented evidence the exon 8 variant was associated with a dwarf phenotype and thus assigned *D3**.

Table 2.9. TaqMan® results for 64 large breed horses for SNP g.95,282,140 G>C.

Breed	Population size	*D3** TaqMan® results
Thoroughbred	28	0
Standardbred	22	0
Saddlebred	4	0
Tennessee Walker	4	0
Arabian	4	0
Hackney	1	0
Caspian	1	0
Totals	**64**	**0**

*Estimation of allele frequency for D3**

Calculations of allele frequency were performed for g.95,282,140 G>C from a population of 361 horses submitted for testing by owners of Miniature horses to the Gluck

Equine Research Center Genetics Testing Laboratory. One of the horses submitted for testing was phenotypically identified as a dwarf. The remaining horses had normal phenotypes. Among the normal horses, none were found with more than 1 of the alleles *D1- D4*. The allele frequency for *D3** was estimated to be 0.017 (Eberth *et al.* 2018).

Discussion

While investigating dwarfism in the Miniature horse, Eberth (2013) discovered SNPs that did not appear annotated in the reference sequence. A list of SNPs was proposed as non-causative of any equine disease at that time within *ACAN*. From the samples and sequence data compiled, it was not evident that variant g.95,282,140 G>C exon 8 sequence exhibited association with any dwarf phenotype investigated. Different horses were used at times when sequencing individual exons. Various numbers of DNA samples were sequenced for each exon to initially look for mutations causing dwarfism. The sequence data was a composite of partial gene sequence on the various individuals. During that process, a repeatable sequencing artifact which appeared to be a single base deletion was observed and labelled as *D3*. Subsequent work genotyping a larger population than used in Eberth (2013) showed the proposed deletion in exon 12 did not exist. Re-examination of sequence data and sequence analysis of sire-dam-offspring samples revealed a SNP previously labelled as non-causative was the cause for the *D3* dwarf phenotype. Genotype testing of g.95,282,140 G>C on the larger population as well as large breed horses showed conclusively it was responsible for the *D3* phenotype and was unique to the Miniature breed. This variant was assigned a designation as *D3**. This correction to Eberth (2013) thesis was noted in the publication of the body of work in *Multiple alleles of ACAN associated with chondrodysplastic dwarfism in Miniature horses* by Eberth *et al.* (2018). This new designation and location for *D3** was corroborated with work done by Metzger *et al.* (2016) in Shetland ponies. Publications by Miniature horse and American Shetland breed experts describe historical pedigree data which document a significant influence of the foundation of the Miniature breed was from American Shetland ponies.

Chapter III

Testing American Miniature horses for Dwarfism haplotype reported for Brazil Miniature horses

Summary

Metzger *et al.* (2016) and Eberth *et al*. (2018) reported four mutations (*D1, D2, D3** and *D4*) in *ACAN* associated with dwarfism in Miniatures and Shetlands. Eberth *et al*. (2018) observed 8 dwarf miniatures with only one of these alleles and proposed the existence of at least one other variant in *ACAN* was responsible for dwarfism. Andrade *et al*. (2020) identified dwarf Miniature horses in Brazil that did not have any of the previously reported variants and described a new missense variant in *ACAN* in exon 12, ECA1 g.95271115A > T associated with dwarfism. Of 347 Miniature horses tested by Andrade and co-workers, the Brazil variant was found in 272 as carriers and heterozygous in 13 dwarf foals and homozygous in 1 dwarf foal (allele frequency of 0.112). This variant was also found in other breeds of large horses with no history of dwarfism, and therefore not likely to be causative. In this study, a population of 464 American Miniatures, previously genotyped for dwarf variants *D1, D2, D3* and *D4,* were tested with a custom TaqMan® assay designed for the Brazil dwarf variant. The population included 91 non-dwarf carriers, 281 non-dwarf non-carriers, 73 dwarf individuals homozygous or compound heterozygous for *D1-D4*, and 19 dwarfs possessing one or no copies of the *D1-D4* variants. TaqMan® results demonstrated 6 dwarfs possessing one or none of the *D1-D4* variants also possessed one copy of Brazil dwarfism variant. No non-dwarf carrier of *D1-D4* possessed a copy of the Brazil variant. Seven non-dwarf non-carriers of D1-D4 possessed one copy of the Brazil variant. One compound heterozygous dwarf possessed one copy of the Brazil variant. One non-carrier of *D1-D4* possessed two copies of the Brazil haplotype and expressed a normal phenotype. In conclusion, The Brazil variant does show a high association with a dwarf phenotype but is not in complete LD with the unknown putative variant causing dwarfism in the American Miniature population. The allele frequency of the Brazil variant was 0.0172 in this study compared to 0.112 in the Brazil Miniature horse population.

Introduction

Eberth *et al*. (2018) reported 7 dwarf Miniature horses with only one of the dwarf alleles and proposed the existence of at least one other variant in *ACAN* was responsible for dwarfism. Recently, Andrade *et al*. (2020) identified dwarf Miniature horses in Brazil that did not have any of the previously reported variants and described a new missense variant in *ACAN* in exon 12, ECA1 g.95271115A > T associated with dwarfism. The variant was found in 15 of 347 Miniature horses (frequency of 0.0172). This variant was also found in other breeds of large horses with no history of dwarfism, and therefore not likely to be causative. The purpose of this study was to investigate the presence and frequency of missense SNP in *ACAN*, Ecab 3.0 ECA1: g.95271115A > T associated with a chondrodysplasia-like dwarfism phenotype in American Miniature horses and determine to what extent this haplotype might account for the unknown dwarfism alleles of Miniature horses.

Materials and Methods

Horses

A population of 464 American Miniatures were used in this study, of which 358 were previously genotyped for dwarf variants *D1, D2, D3* and *D4*. This population included 91 non-dwarf carriers, 175 non-dwarf non-carriers (94 from the Gluck Equine Research Center Genetics Testing Laboratory, 62 from Eberth *et al*. (2018), and 19 from work within this dissertation), 73 dwarf individuals which were homozygous or compound heterozygous for *D1-D4*, and 17 dwarfs genotyped with possessing one copy and 2 dwarfs with none of the *D1-D4* variants. Also included, were 106 Miniatures horses of unknown genotype from Chapter V of this dissertation. Other horses used to determine allele presence and frequency, included whole genome sequence data was available from three sources: 101 Japanese Thoroughbreds (Tozaki *et al*. 2010), 534 horses of diverse breeds (Durward-Akhurst *et al.* 2021) and 230 Thoroughbred horses (Bailey, Kalbfleisch and Petersen, unpublished data). IACUC 2022-4161 was approved in connection with this study of experimental animals.

DNA Isolation

DNA from blood was extracted using Puregene whole blood DNA extraction kits (Gentra Systems Inc., Minneapolis, MN) and Puregene tissue DNA extraction kits according to published protocols. Hair samples submitted were processed using 7-10 hair bulbs according to the method described by Locke *et al*. (2001). The hair bulbs were placed in 100 µl lysis solution containing 1X FastStart Taq Polymerase PCR buffer (Roche), 2.5 mM MgCl (Roche), 0.5% Tween 20 (JT Baker, Phillipsburg, NJ) and 0.01 mg proteinase K (ThermoFisher Waltham, MA.) and incubated at 60°C for 45 minutes followed by 95°C for 45 minutes to deactivate the proteinase K. Aliquots of the DNA samples were made for working dilutions at concentrations of approximately 150 ng/µl.

Genotyping

The Miniature horse population was genotyped using a custom TaqMan® assay designed through the ThermoFisher website using the custom TaqMan® design tool https://www.thermofisher.com/order/custom-genomic-products/tools/genotyping/. Figure 3.1 provides illustrations of the genomic location of the Brazil haplotype variant and the custom TaqMan® assay utilized for testing. Genotyping was performed on Applied Biosystems Step One Plus Real Time PCR system.

A. Brazil haplotype genomic location in Ecab 3.0

```
95271179 AACTCAGTGGTGCCCATTCTGGAGCACCGGCCGTGTCTGGAGACCATTCGGGATTTTCGG 95271120
95271119 ACTTAAGCGGGCTGCCGTCAGGGCTGGTGGAGCCCAGTGGAGAGCCATCGAGTACTCCAC 95271060
```

B. Brazil haplotype Assay – VIC-A FAM-T

```
AACTCAGTGGTGCCCATTCTGGAGCACCGGCCGTGTCTGGAGACCATTCGGGATTTTCG
GACTT[A/T]AGCGGGCTGCCGTCAGGGCTGGTGGAGCCCAGTGGAGAGCCATCGAGTA
CTCCAC
```

Figure 3.1. A. Brazil dwarf haplotype SNP highlighted in red shown in Ecab 3.0. **B.** Sequence of custom TaqMan® assay.

Results

Genotyping Brazil variant in American Miniature horses

A population of 464 Miniatures were tested for the Brazil dwarf variant. Of the 464 miniature horses, 15 possessed the Brazil variant. Specifically, the variant showed differences in presence/absence and frequency within different Miniature groups in this study. Among the non-dwarf non-carriers, the Brazil allele was not present in the population from Eberth *et al*. (2018). Among other non-carrier non-dwarf samples collected for specific work within this dissertation (Chapters II & IV), no samples possessed a copy. Within the population of 94 tested from the Gluck Equine Research Center Genetics Testing Laboratory, one individual (Acc# H14-04034) was found to be homozygous for the Brazil variant and another 3 possessed a single copy, shown in Table 3.1. This horse was investigated for registration, height, and produce record. Records show it was registered at 32" tall at maturity, possessed multiple show records that would distinguish it as show quality. Multiple offspring were recorded and registered as sired by this horse. This evidence provides a rational assumption that the horse and offspring of record passed the photo examination for possessing normal phenotypes and that none displayed dwarfism characteristics. According to registration rules, a dwarf cannot be registered to either registry and a photo record is used to determine a dwarf phenotype during the registration process (AMHA 2023; AMHR 2023).

Table 3.1. Non-dwarf non-carriers of *D1-D4* genotype of Brazil variant g.95271115A > T.

Genotype	Source	Phenotype	Brazil SNP A/T	Brazil SNP T/T
N/N-62	Eberth *et al*. 2018	Normal	0	0
N/N-94	Gluck Testing Lab	Normal	3	1
N/N-19	Eberth (2023) Diss. Ch 2 & 4	Normal	0	0
Total-175			3	1

Within the population of 106 Miniatures collected for Chapter V of this dissertation, four individuals possessed a single copy as shown in Table 3.2. These horses were collected from 6 breeding farms across multiple states within the US. Their genotype status was unknown for all dwarfism variants at the time of collection.

Table 3.2. Population with unknown genotype from Eberth (2023) Chapter V.

Genotype	Source	Phenotype	Brazil SNP A/T	Brazil SNP T/T
Unknown- 106	Eberth 2023 Dissertation Ch V	Non-dwarf Normal	4	0

None of the 91 samples tested previously as carriers of dwarf variants *D1-D4*, possessed a copy of the Brazil variant, shown in Table 3.3. These horses were from an earlier dwarfism study, (Eberth *et al*. 2018) as well as studies within this dissertation.

Table 3.3. Non-dwarf carriers of *D1-D4* genotype of Brazil variant g.95271115A > T. Sources for these were from Eberth *et al.* (2018), Eberth (2023 Dissertation Ch 2 & 4).

Genotype	Phenotype	Brazil SNP A/T	Brazil SNP T/T
N/D1- 31	Non-dwarf Normal	0	0
N/D2- 51	Non-dwarf Normal	0	0
N/D3- 4*	Non-dwarf Normal	0	0
N/D4- 5	Non-dwarf Normal	0	0
Total - 91		0	0

Within the population of dwarf horses genotyped as homozygotes or compound heterozygotes for *D1-D4*, the Brazil variant was found in one individual (Acc.# 2710) with a genotype *D2/D3** (Table 3.4). The genotype was confirmed for Acc.# 2710 by TaqMan® and sequence data for *D2* and *D3**.

Table 3.4. Homozygous and compound heterozygous dwarfs of *D1-D4* and genotype of Brazil variant g.95271115 A>T. Sources for these were from Eberth *et al.* (2018), Eberth (2023) Dissertation Ch 2 & 4.

Genotype ***D1-D4***	Phenotype	Brazil SNP A/T	Brazil SNP T/T
D1/D1- 3	Dwarf	0	0
D1/D2- 9	Dwarf	0	0
D1/D3- 2*	Dwarf	0	0
D1/D4- 2	Dwarf	0	0
D2/D2- 30	Dwarf	0	0
D2/D3- 14*	Dwarf	1	0
D2/D4- 10	Dwarf	0	0
D3/D3*- 0*	Dwarf	0	0
D3/D4- 2*	Dwarf	0	0
D4/D4- 1	Dwarf	0	0
Total- 73		1	0

Among the 19 dwarf horses with one or none of the *D* (dwarf) alleles, 6 possessed one copy of the Brazil variant, shown in Table 3.5. Five of those 6 dwarf individuals could be genotyped completely with this variant. Thirteen dwarf horses remained that possess

an unknown dwarf allele. Table 3.6 gives are summary of the American Miniature horse population tested for the Brazil variant.

Table 3.5. 19 dwarf horses possessing one or no *D1-D4* mutations and genotype for Brazil dwarf variant.

Acc. #	*D1-D4*	Brazil SNP	Genotype
2694	*N/D2*	*N/Brazil*	*D2/Brazil*
2698	*N/N*	*N/N*	*N/N*
2701	*N/D2*	*N/N*	*N/D2*
2741	*N/D4*	*N/N*	*N/D4*
2742	*N/D3**	*N/N*	*N/D3**
2745	*N/D4*	*N/N*	*N/D4*
2746	*N/D2*	*N/N*	*N/D2*
2906	*N/D2*	*N/N*	*N/D2*
3545	*N/N*	*N/Brazil*	*N/Brazil*
3547	*N/D2*	*N/N*	*N/D2*
3563	*N/D3**	*N/N*	*N/D3**
3566	*N/D2*	*N/N*	*N/D2*
3567	*N/D2*	*N/Brazil*	*D2/Brazil*
3568	*N/D2*	*N/N*	*N/D2*
3578	*N/D2*	*N/Brazil*	*D2/Brazil*
3579	*N/D2*	*N/N*	*N/D2*
3591	*N/D2*	*N/Brazil*	*D2/Brazil*
3596	*N/D2*	*N/Brazil*	*D2/Brazil*
3612	*N/D4*	*N/N*	*N/D4*

Table 3.6. 445 Miniatures of all genotypes possessing *N-D4* and results for Brazil variant. This does not include individuals from Table 3.5.

Genotype *D1-D4*	Phenotype	Brazil SNP A/T	Brazil SNP T/T
N/N- 281	Normal	7	1
N/D1- 31	Normal	0	0
N/D2- 51	Normal	0	0
N/D3- 4*	Normal	0	0
N/D4- 5	Normal	0	0
D1/D1- 3	Dwarf	0	0
D1/D2- 9	Dwarf	0	0
D1/D3- 2*	Dwarf	0	0
D1/D4- 2	Dwarf	0	0
D2/D2- 30	Dwarf	0	0
D2/D3- 14*	Dwarf	1	0
D2/D4- 10	Dwarf	0	0
D3/D3*- 0*	Dwarf	0	0
D3/D4- 2*	Dwarf	0	0
D4/D4- 1	Dwarf	0	0
Totals- 445		**8**	**1**

Presence of Brazil variant in other breeds

Whole genome sequence data was available for horses from three sources: 101 Japanese Thoroughbreds (Tozaki *et al.* 2010), 534 horses of diverse breeds (Durward-Akhurst *et al.* 2021), and 230 Thoroughbred horses (Bailey, Kalbfleisch and Petersen, unpublished data). The Brazil variant was found in all three studies' datasets, shown in Table 8. Dwarfism has not been reported for Thoroughbred horses or most of the breeds included in the study by Durward-Akhurst *et al*. (2021).

Table 3.7. Allele frequency of Brazil variant g.95271115 A>T in other horse populations.

Breed	Study	Number	Brazil haplotype
Diverse breeds	Durward-Akhurst *et al*. 2021	534	0.027
Japanese Thoroughbreds	Tozaki *et al*. 2010	101	0.119
American Thoroughbreds	Bailey,Kalbfleisch and Petersen, unpublished	230	0.065

Discussion

Chondrodysplastic dwarfism in Miniatures has been shown to involve multiple alleles in *ACAN* segregating independently within populations world-wide. These alleles can combine to produce homozygous and compound heterozygous phenotypes which present a range of physical malformations dependent on the variants inherited by the individual. Previous works have documented dwarf individuals which possess unknown variants presenting phenotypes like those seen in chondrodysplastic dwarfism. Since the publications of Metzger *et al.* (2016) and Eberth *et al*. (2018), Andrade *et al*. (2020) discovered a haplotype in *ACAN* associated with a dwarf phenotype in Miniatures of Brazil.

The Brazil variant was found in 14 out of 464 Miniature horses in this study in which 6 of these were dwarfs. One normal Miniature horse was homozygous for this variant, demonstrating that this variant cannot be responsible for dwarfism in Miniature horses. Indeed, in separate studies the variant was identified among Thoroughbred and other horse breeds not known to harbor dwarfism causing variants. However, it was observed in 6 of the 19 dwarf horses thought to possess as yet unidentified variants for dwarfism. Together, with the observations of Andrade *et al*. (2020) suggests that this

variant is in linkage disequilibrium with a mutation that causes dwarfism in both Brazil and the USA.

As mentioned in Eberth *et al*. (2018) and Andrade *et al*. (2020), the VNTR region of exon 12 was not completely sequenced in those studies. The extensive copy number variation in this region is quite complex producing unreliable short read Sanger sequences. This area may hold the true cause of the Brazil dwarf phenotype and possibly other dwarf variants. Future studies using long read sequencing should be able to identify the causative variant in this haplotype.

Andrade and co-workers (2020) described other unique genetic influences which may have occurred to the Brazil Miniature horse population. These influences are unlike the American Miniature horses, which may account for the high dwarf allele frequencies they encountered as well as the complete association of the haplotype to a dwarf phenotype. This is corroborated in the history of the American Miniature horse (Naviaux 1999). The importation of American Miniatures into Brazil has been somewhat restricted due to purchase and shipping costs. However, a unique Miniature horse breed was developed in Argentina called the Falabella Miniature horse as early as the late 19th century. The history of the Falabella Miniature horse is somewhat obscure with the founder claiming many breeds were used in its development (Naviaux 1999). Interestingly, this horse breed was imported in small numbers to America, but its phenotype is not considered desirable to breeders today. The Falabella Miniature would be an easily accessible Miniature horse, for breeders in Brazil, to cross with the American Miniatures imported to Brazil.

The Brazil dwarf variant within the Miniature horses of the United States does not show complete association to a diseased phenotype. This may be due to genetic pressure differences experienced by the population of Miniatures in Brazil versus the population in the United States. The allele frequency of all dwarf variants within the Brazil population is significantly high (47.7%) in relation to the normal genotype. The US population as well shows a significant dwarf allele frequency but not to the percentage of that in Brazil. Breeding practices can be restricted by international importation laws and costs. This may

be a partial explanation to the significantly high dwarf allele frequency rate within the Brazil population in relation to the population in the US.

Chapter IV
Genomic Investigation for additional, novel Dwarfism associated haplotypes in Miniature Horses

Summary

Whole genome sequencing was performed on two chondrodysplastic dwarfs heterozygous for the *D2* variant and possessing an unknown dwarf variant. Whole genome sequence analysis revealed 9 intronic SNPs, 1 intron deletion, 6 exonic SNPs that altered amino acids, and 1 coding deletion within *ACAN* unique to at least one or both dwarfs. Both individuals possessed one copy of an in-frame deletion located in the VNTR region. *In silico* analysis using VEP (Variant Effect Predictor) and PredictSNP showed 3 SNPs projected to have high probability as deleterious to the function of the protein, one in each exon 13, 14 and 16 at 79%, 87% and 72% respectively. Dwarf 2906 possessed one copy in exon 13, g.95258999 C>T aa. 2479 Alanine>Threonine. Dwarf 2701 possessed one copy in Exon 14 g.95258406, A>C aa. 2525 Tryptophan > Glycine. Both dwarfs possessed one copy in exon 16 g.95256565, G>C aa. 2582 Cysteine > Serine. A population of 464 Miniatures previously genotyped for dwarf variants *D1, D2, D3*, *D4* and the Brazil dwarf variant were tested with three custom TaqMan® assays designed for the three mutations. The population included 99 normal phenotyped carriers, 273 Normal phenotyped non-carriers, 73 dwarf individuals homozygous or compound heterozygous for *D1-D4*, and 19 dwarfs genotyped with possessing one copy or none of *D1-D4* variants. TaqMan® results showed variant g.95256565 G>C to have extensive distribution within the population with high correlation to the dwarf variant assigned *D2;* however, exceptions were seen. TaqMan® results also showed genotyped carriers of *D1-D4* did not possess a copy of SNPs g.95258999 C>T or g.95258406 A>C. Four non-carriers possessed one copy of variant g.95258999 in exon 13. No genotyped homozygous or compound heterozygous dwarfs possessed a copy of SNPs g.95258999 C>T or g.95258406 A>C. Six heterozygous dwarfs with an unknown variant possessed one copy of g.95258999 C>T. One dwarf negative for *D1-D4* and Brazil dwarf variants possessed two copies of g.95258999 C>T. Dwarf 2701 was the only sample which possessed a single copy of SNP g.95258406 A>C. Variant g.95284530 C>T (assigned *D2* SNP) continues to show 100% linkage disequilibrium with *D2* dwarf phenotype. SNP g.95258999 C>T shows independent segregation to dwarf

variants *D1, D2, D3, D4* and the Brazil haplotype. When SNP g.95258999 is combined with another dwarf variant, a dwarf phenotype is present thus given the designation *D5*.

Introduction

Previous work by Eberth *et al.* (2018) described 4 variants of *ACAN* associated with dwarfism in Miniature horses. The trait had a recessive mode of inheritance, demonstrated by the observation that almost all affected dwarfs were homozygotes or compound heterozygotes for the dwarfism causing variants. However, there were also seven dwarfs that possessed a single copy of one of the variants, and one dwarf which possessed none, suggesting the existence of additional variants associated with dwarfism. Further evidence was provided by Andrade *et al.* (2020) who reported a novel variant of *ACAN* associated with chondrodysplastic dwarfism haplotype in Miniature horses of South America, distinct from *D1-D4*. In the previous chapter of this dissertation, evidence was given for the presence of this haplotype among Miniature horses in North American, but it did not account for all the putative dwarfism alleles that had not yet been identified. The purpose of this study was to investigate the presence of additional dwarfism causing alleles. The approach was to use whole genome sequencing of two dwarf horses possessing a known and a putative unknown variant for dwarfism. Genomic analysis was restricted to the region of *ACAN* which contains variants associated with chondrodysplastic dwarfism in the Miniature horse.

Materials and Methods

Horses

Whole genome sequencing: Two horses, 2701 and 2906, were selected for whole genome sequencing. They were chosen because each had a dwarf phenotype but had been genotyped as *D2/N* as described previously (Eberth *et al.* 2018). The objective in choosing these horses was to compare and confirm the *D2* haplotype as well as to identify new variants causing dwarfism.

Miniature horse population: To estimate variant frequencies, hair samples from 195 Miniature horses were selected from among the samples submitted to the Gluck Equine Research Center Genetic Testing laboratory. These were normal phenotype horses

submitted for testing for the dwarfism causing *ACAN* variants and previously genotyped as non-carriers of *D1-D4* and the Brazil variant.

Miniature horse dwarf study: The remaining 269 individuals were acquired for research purposes as having produced a dwarf, possessed a dwarf phenotype, or related to a dwarf producer. The dwarf research project population included 99 normal phenotyped carriers, 78 normal phenotyped non-carriers, 79 dwarf individuals which were homozygous or compound heterozygous for *D1-D4* and Brazil variant, and 13 dwarfs genotyped with possessing one copy or none of the *D1-D4* variants or the Brazil variant.

Other horses: To determine allele presence and frequency in other breeds, whole genome sequence data was available from three sources: 101 Japanese Thoroughbreds (Tozaki *et al.* 2010), 534 horses of diverse breeds (Durward-Akhurst *et al.* 2021) and 230 Thoroughbred horses (Bailey, Kalbfleisch and Petersen, unpublished data). Also, a set of 93 large breed horses from the Gluck Equine Research Center genomics research laboratory reference DNA collection. This included 20 Standardbreds, 20 Arabians, 25 Thoroughbreds, 12 American Saddlebreds, 10 American Quarter horses and 6 Tennessee Walking horses. IACUC 2022-4161 was approved in connection with this study of experimental animals.

DNA isolation

DNA from blood was extracted using Puregene whole blood DNA extraction kits (Gentra Systems Inc., Minneapolis, MN) and Puregene tissue DNA extraction kits according to published protocols. Hair samples submitted were processed using 7-10 hair bulbs according to the method described by Locke *et al.* (2002). The hair bulbs were placed in 100 µl lysis solution containing 1X FastStart Taq Polymerase PCR buffer (Roche), 2.5 mM MgCl (Roche), 0.5% Tween 20 (JT Baker, Phillipsburg, NJ) and 0.01 mg proteinase K (ThermoFisher Waltham, MA.) and incubated at 60°C for 45 minutes followed by 95°C for 45 minutes to deactivate the proteinase K. Aliquots of the DNA samples were made for working dilutions at concentrations of approximately 150 ng/µl.

Whole genome sequencing

As noted above, DNA samples from 2701 (*D2/N*) and 2906 (*D2/N*), were submitted to Psomagen Inc (Rockville, MD) to perform Whole Genome Sequencing. Genomic quality standard minimum of 500 ng of total DNA with an DNA Integrity Number (DIN) of 7.0 or higher was provided from each horse. Genomic DNA was evaluated by Psomagen using a Bioanlayzer instrument. Illumina TruSeq DNA PCR-Free kit was utilized for genomic library preparation. Illumina NovaSeq sequencing instrument was used to produce paired-end short-read sequences (read size of 150 bp). This generated 50 Giga-base of sequence data per horse. Expected sequencing coverage was to be on average 20X, based on the genomic size of 2.5 Gb EqCab 3.0 reference assembly. Psomagen provided sequence data as raw FASTQ files.

Analysis of Whole Genome Sequence

Sequence adapters were removed from sequence reads using TrimGalore software. BWA aligner (Li *et al*. 2010) was used to map sample reads to the equine reference genome assembly EqCab 3.0 to examine Chromosome 1 sequence only containing the gene *ACAN*. Genome Analysis Toolkit (GATK) (McKenna *et al*. 2010) was used to call the variants to generate final VCF files for each sample.

Genomic region specific for ACAN

Utilizing IGV, (Integrated Genome Viewer 2013), the genomic region (Primary assembly Chr 1: 95,255,522-95,320,938 reverse strand) specific to *ACAN* including UTRs, intron and exon regions were examined to discriminate between the two dwarf samples, two Saddlebred horses, equine genome reference horse Twilight VCF file as well as databases of whole genome sequence of 101 Japanese Thoroughbreds (Tozaki *et al.* 2010), 534 horses of diverse breeds (Durward-Akhurst *et al.* 2021) and 230 Thoroughbred horses (Bailey, Kalbfleisch and Petersen, unpublished data) to compile a list of variants unique to the two dwarfs. Reference genome Ecab 3.0 in Ensembl was also used to compare dwarf variants to other breed annotations to compile a final list of unique variants in the two dwarfs. Sanger sequence reads from Eberth *et al.* (2018) were aligned and analysis performed using Codon Code aligner https://www.codoncode.com/aligner/index.htm.

Variant analysis of coding regions

Functional assessment of unique variants in coding regions were performed using genomic database Ensembl web-based bioinformatic analysis tool Variant Effect Predictor (VEP) http://www.ensembl.org/info/docs/tools/vep/index.html for SIFT calculations (McLaren 2016). To support the SIFT calculations, consensus classifier PredictSNP https://loschmidt.chemi.muni.cz/predictsnp/ was used to examine the variants. PredictSNP is a web-based consensus classifier of eight independent analysis tools for altered protein function due to sequence or amino acid change. These tools include MAPP (Stone *et al.* 2005), nsSNPAnalyzer (Bao *et al.* 2005), Panther (Thomas *et al.* 2004), PHD-SNP (Capriotti *et al.* 2006), PolyPhen-1 (PPH-1) (Ramensky *et al.* 2002), PolyPhen-2 (PPH-2) (Adzhubei *et al.* 2010), SIFT (Ng *et al.* 2003) and SNAP (Bromberg *et al.* 2007). The results are given in a table as predictive probabilities for each individual tool of a neutral or deleterious effect as well as an overall probability calculated using proprietary algorithms.

To assess variation within the VNTR, the only option was to evaluate repeat size consistency and reading frame retention by hand using Ecab 3.0 in Ensembl. No *in silico* tools explored could evaluate large VNTRs and the potential effects on a mature protein. INDEL variants in coding regions were assessed by VEP web tool in Ensembl.

Variant analysis of non-coding regions

In silico prediction tools for non-coding regions are not as prevalent or as comprehensive relative to predictive tools for coding regions. Almost all non-coding tools used for sequence analysis reference the human genome and studies regarding non-coding regions. However, the Variant Effect Predictor (VEP) tool (McLaren 2016) within Ensembl web-based genome database not only provides some predictive analysis of non-coding regions but can do it using any reference genome it has annotated in the database. Because the sequences at the two ends of an intron are relatively conserved and the prominent *cis*-acting elements of splicing, these sequence characteristics have been the emphasis of most web-based tools. Therefore, this study restricted non-coding regions to

SNP variants within splicing consensus regions (−3 to +8 at the 5′ splice site and −12 to +2 at the 3′ splice site) as defined by Burge *et al*. (1999).

Genotyping

During the study, variants were identified and tested in populations of horses. Genotyping was conducted using custom TaqMan® assays designed through the ThermoFisher website using the custom TaqMan® design tool https://www.thermofisher.com/order/custom-genomic-products/tools/genotyping/. Genotyping assays were performed on Applied Biosystems Step One Plus Real Time PCR system. The three custom assays were designed from equine reference sequence Ecab 3.0 in Ensembl. Shown in Figure 4.1 are genomic locations in Ecab 3.0 for the three assays and sequences submitted to Thermofisher custom assay design tool.

A. g.95258999 G>A Exon 13 genomic coordinates

```
95259082 ACCAGGAGCTGTGTGAGAATGGCTGGACCAAGTTCCAGGGCCACTGTTACCGCTACTTTC      95259023
95259022 CCGACCGCGAGACCTGGGTGGACGCCGAGAGCAGGTGTCGGGAGCAGCAGTCACACCTGA      95258963
95258962 GCAGCATCGTCACTCCCGAGGAGCAGGAGTTTGTCAACA                           95258924
```

B. g.95258999 G>A Exon 13 Assay VIC G FAM A

```
ACCAGGAGCTGTGTGAGAATGGCTGGACCAAGTTCCAGGGCCACTGTTACCGCTACTTTCCCGACCGCGAG
ACCTGGGTGGAC[G/A]CCGAGAGCAGGTGTCGGGAGCAGCAGTCACACCTGAGCAGCATCGTCAC
```

C. g.95258406 T>G Exon 14 genomic coordinates

```
95258468 ACAATGCCCAAGACTACCAGTGGATCGGCCTGAATGACAGGACCATCGAAGGGGACTTCC    95258409
95258408 GCTGGTCAGATGGACACTCCTTG                                        95258386
95258385 gtgagttctgcaggggcactggggagagttggcagcccagaggagggagatgaggacccc    95258326
```

D. g.95258406 G>C Exon 14 Assay VIC T FAM G

```
ACAATGCCCAAGACTACCAGTGGATCGGCCTGAATGACAGGACCATCGAAGGGGACTTCCGC
[T/G]GGTCAGATGGACACTCCTTGgtgagttctgcaggggcac
```

E. g.95256565 G>C Exon 16 genomic coordinates

```
95256613 gcgcatcacaggagccctttctgaagactgtctgtccctcag                    95256572
95256571TGGCCTGCGGAGACCCCCCTGTGGTGGAGCACGCCAGGACCTTTGGGCGGAAGAAGGCGC    95256512
```

F. g.95256565 G>C Exon 16 Assay VIC G FAM C

```
GCGCATCACAGGAGCCCTTTCTGAAGACTGTCTGTCCCTCAGTGGCCT[G/C]
CGGAGACCCCCCTGTGGTGGAGCACGCCAGGACCTTTGGGCGGAAGAAGGCGC
```

Figure 4.1. Highlighted in yellow are the variants discovered and locations for the custom TaqMan® assays, All sequences are in reverse reading frame; **A.** g.95258999 C>T p.A2479T Exon 13; **B.** g.95258999 C>T Assay VIC G FAM A; **C.** g.95258406 G>C p.W2525G Exon 14; **D.** g.95258406 G>C Assay VIC T FAM G; **E.** g.95256565 G>C p.C2582S Exon 16; **F.** g.95256565 G>C Assay VIC G FAM C

Results

Whole Genome Sequencing

Analysis of whole genome sequence was specific to the region of *ACAN* including all UTRs, exons, and introns (Ecab 3.0 Primary assembly chr1: 95,255,522-95,320,938 reverse strand). Whole genome analysis identifying alternative variants was performed in the forward reading frame. This study concentrated on variants within *ACAN* when combined with other dwarf variants which results in a dwarf phenotype. Variants outside of *ACAN* were not investigated for possible deleterious effects as it relates to chondrodysplasia-like dwarfism in Miniature horses. A cartoon flow chart in Figure 4.2 demonstrates the simplified process determined would best elicit new variants which could be associated with a dwarf phenotype.

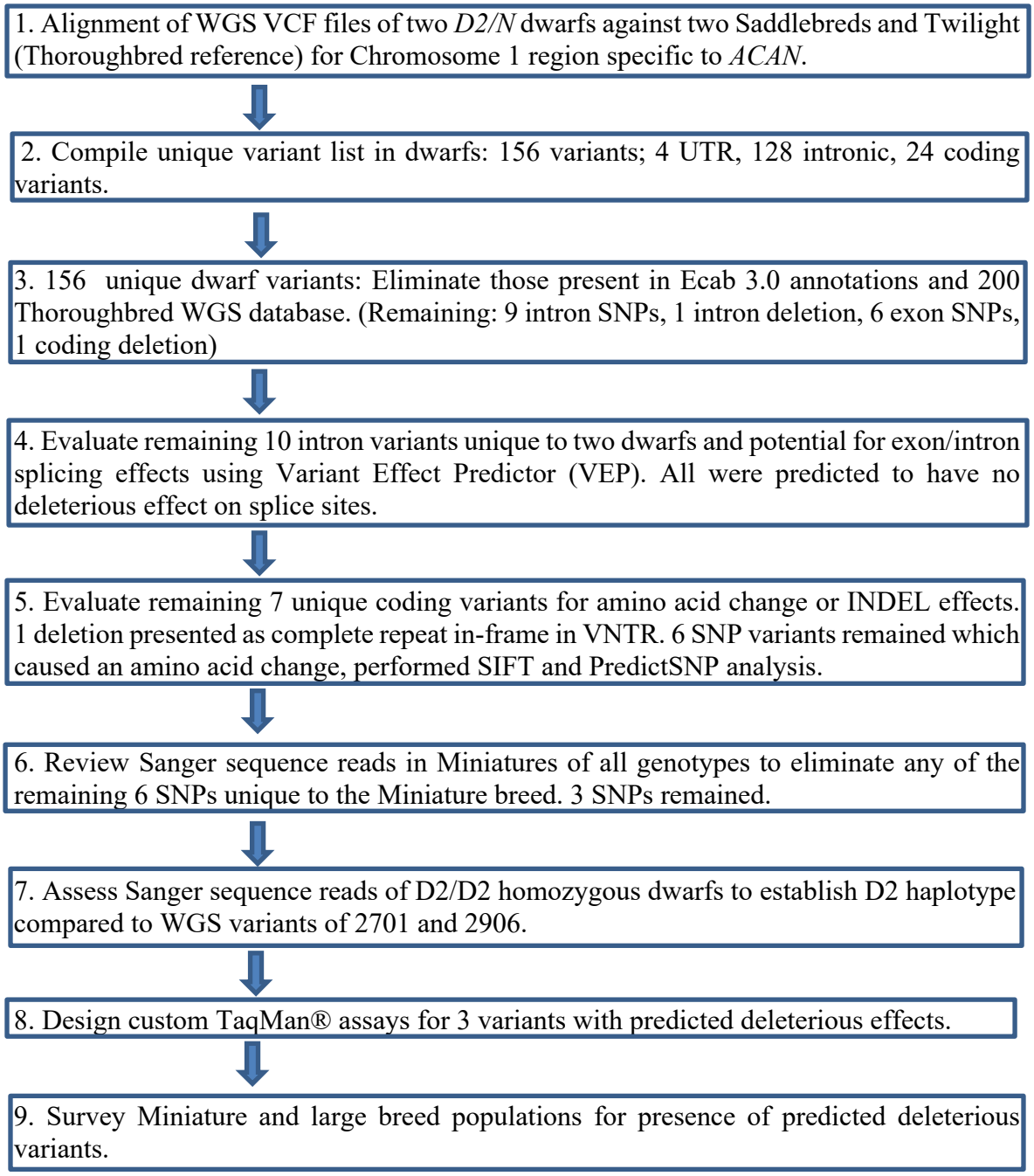


Figure 4.2. A cartoon flow chart of the process used to eliminate unique variants possessed by two *D2/N* dwarfs after initial analysis of aligned VCF files with Twilight and two Saddlebreds. All unique dwarf variants were also examined against Ecab 3.0 or a WGS database of 200 Thoroughbreds.

Whole genome VCF files of the two dwarfs were compared to VCF files of two Saddlebred horses and the VCF file of the Thoroughbred Twilight. Table 4.1 lists the number of variants when the two dwarf horses were compared to the VCF files of Twilight

and two Saddlebreds. This preliminary comparison was to cull potentially common variants from two breeds with no history of dwarfism. The results produced a total of 156 unique variants possessed by either or both dwarfs.

Table 4.1. Unique variants possessed by each dwarf compared to Twilight and two Saddlebreds. Results totaled 156 unique variants in 2701 and 2906.

Variant location in *ACAN*	**2701**	**2906**
introns	104 SNPs, 13 INDELs	90 SNPs, 10 INDELs
exons	19 SNPs, 1 deletion	22 SNPs, 1 deletion
5' UTR	2 SNPs	4 SNPs
3' UTR	0	0

The 156 unique variants with specific genomic locations and copy number for each dwarf are listed in Supplemental Table 4.1. These variants were subsequently compared to variants seen in WGS databases of 534 horses of diverse breeds and 331 Thoroughbreds using IGV (Integrated Genome Viewer 2013; Durward-Akhurst *et al*. 2021; Tozaki *et al*. 2010; Bailey, Kalbfleisch and Petersen, unpublished data). The results of this comparison produced 32 unique variants possessed by both or either dwarf: 9 SNPs and 1 INDEL within introns and 23 SNPs and 1 deletion within exons. The remaining 17 unique variants were examined according to gene location and ascertain potential to cause a negative impact to the mature protein.

Table 4.2. 156 unique variants possessed by each dwarf compared to other breeds in Ecab 3.0 and WGS of 230 Thoroughbreds. Results produced 17 unique variants possessed by 2701 and 2906.

Variant location in *ACAN*	**2701**	**2906**
introns	9 SNPs, 1 deletion	9 SNPs
exons	5 SNPs, 1 deletion	5 SNPs, 1 deletion
5' UTR	0	0
3' UTR	0	0

WGS variants in non-coding regions

After comparing the VCF sequence files to Twilight and two Saddlebred VCF files, the remaining 156 unique dwarf variants were examined for presence of variation in WGS database of 230 Thoroughbreds. This was performed to ascertain all variants in *ACAN* unique to only the two dwarfs. A total of 9 SNPs were found unique to the two dwarfs in non-coding regions as shown in Table 4.3. These SNPs were examined for potential

interference of conserved splice sites using VEP (Variant Effect Predictor) web tool within the Ensembl genome database internet browser http://www.ensembl.org/info/docs/tools/vep/index.html. Assessment showed all 9 were predicted to have no deleterious effects on the mature protein (Burge *et al.* 1999; Cartegni *et al.* 2002; Choi *et al*. 2012; Jian *et al.* 2014).

Table 4.3. Intronic regions contained 9 SNPs unique to both dwarfs, taken from Supplemental Table 1, which are not annotated in Ensembl Ecab 3.0 or present in WGS database of 230 Thoroughbreds. Reference nucleotides and alternative variants possessed by each dwarf 2701 and 2906 and genomic locations are shown in forward reading frame.

#CHROM	POS	REF	ALT	2701	2906
chr1	95268662	G	A	G/A	G/A
chr1	95269519	G	A	G/A	G/A
chr1	95281618	T	G	T/G	T/G
chr1	95291513	A	G	A/G	A/G
chr1	95299780	G	C	G/C	G/C
chr1	95300373	A	T	A/T	A/T
chr1	95300602	T	C	T/C	T/C
chr1	95310855	C	T	C/T	C/T
chr1	95313448	G	C	G/C	G/C

A total of 15 intron INDELs were unique to either or both dwarfs from Supplementary Table 4.1. These were examined for presence in Ecab 3.0 and a WGS database of 230 Thoroughbreds as shown in Table 4.4. One INDEL remained unique to one dwarf, 2701. This INDEL was investigated for potential to cause interference of conserved splice sites using VEP web tool. This INDEL is located within a string of Adenosine nucleotides which is annotated for copy number variation in Ecab 3.0 downstream from the location. Four SNPs located in the 5' UTR and unique to the two dwarfs are taken from Supplementary Table 4.1 and shown in Table 4.5. All 4 variants were found to be present in Ecab 3.0.

Table 4.4. Intron regions contained 15 INDELs unique to both or either dwarf. Variants highlighted in yellow are annotated in Ensembl Ecab 3.0 with no deleterious effects or are present in WGS database of 230 Thoroughbreds. Variant highlighted in grey is unique to 2701 dwarfs and not annotated in Ecab 3.0 nor are present in WGS of 230 Thoroughbreds. Copy number of possessing alternative variants are shown for each dwarf 2701 and 2906. Reference and alternate variants are shown in the forward reading frame.

#CHROM	POS	REF	ALT	2701	2906
chr1	95280925	CT	C	C/C	C/C
chr1	95288613	CT	C	CT	CT/C
chr1	95292049	CG	C	CG/C	GC/C
chr1	95296240	TTTC	T	T/T	TTTC/TTTC
chr1	95296246	C	CT	C/CT	CT/CT
chr1	95300282	TG	T	T/T	TG/T
chr1	95303036	TC	T	TC/T	TC/TC
chr1	95312374	C	CG	CG/CG	C/CG
chr1	95313314	A	AT	A/AT	A/AT
chr1	95313514	G	GCC	G/GCC	GCC/GCC
chr1	95315420	CT	CTT,C	CTT/C	CT/CTT
chr1	95316183	AG	A	AG/A	AG/AG
chr1	95319305	A	ACGGGG	A/ACGGGG	A/A
chr1	95319318	CAG	C	CAG/C	CAG/CAG
chr1	95319828	CG	C	CG/CG	CG/C

Table 4.5. UTR regions contained 4 variants unique both or either dwarf, taken from Supplemental Table 1. All variants are present in Ensembl Ecab 3.0. Reference and alternative variants possessed by each dwarf 2701 and 2906 are shown in forward reading frame.

#CHROM	POS	REF	ALT	2701	2906
chr1 UTR	95320242	G	A	G/G	G/A
chr1 UTR	95320330	A	T	T/T	T/T
chr1 UTR	95320434	G	A	G/A	G/A
chr1 UTR	95320896	G	C	G/G	G/C

While variants in any of these regions can have an impact on phenotype, the variants most likely to have an effect are those in exons. Therefore, we disregarded all but the exon variants and investigated them for predicted effect on the protein.

WGS Variants in ACAN coding regions

Deletion in VNTR of exon 12

The deletion spanning 95274456-95274520 in exon 12 in Table 4.6 is a complete tandem repeat in-frame removed within the VNTR. This region has been shown to be

variable in copy number within an individual and between individuals, which is annotated in Ensembl as well as described in this dissertation, (Chapter VI). The understanding is the structure and function of the CS1 binding region is dependent upon the repeated sequence. The sequence contains a tandem repeat array of 63 bases (unique sequences of 31 and 32 in tandem) which is then repeated 40 times in the reference genome of the horse. The biological significance of the tandem repeat is its determinant of protein structure and function working as a single unit repeated a variable number of times in the human as shown in Figure 4.4. From that observation of the VNTR array in humans, it has since been shown to be variable in other species as well. This INDEL is not considered a deleterious variant from the analysis performed at this time. Below in Figure 4.3, a cartoon displays the sequence layout and genomic location of the repeat and alternative variants which represent the deletion in heterozygous form in the two dwarf samples. The computer analysis algorithm of the VCF files for variant calling divided the deletion into two parts; however, alignment of the sequences produced a continuous segment within the genome. The nucleotide C representing the large segment in red is a SNP variant annotated in Ecab 3.0 for the A nucleotide indicated with the arrow.

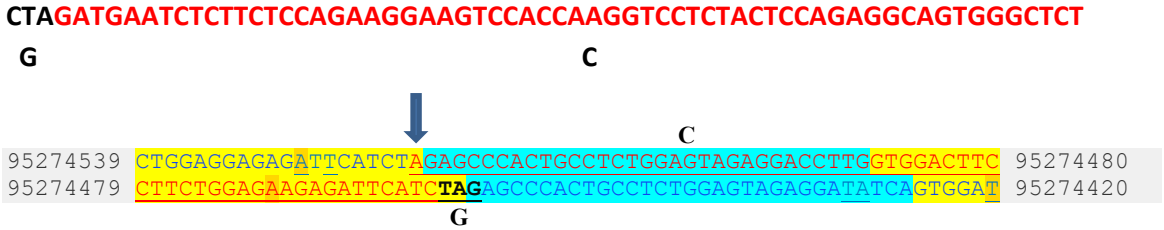


Figure 4.3. TOP: Visual representation of the deleted conserved repeat represented by nucleotides in the two dwarfs in forward reading frame, G in black and C in red. **BOTTOM:** Genomic location and color-coded representation of tandem repeat sequences highlighted in yellow and blue with repeat sequence deletion underlined in reverse reading frame. From Ensembl Ecab 3.0. *ACAN* is located on the reverse strand of the equine genome.

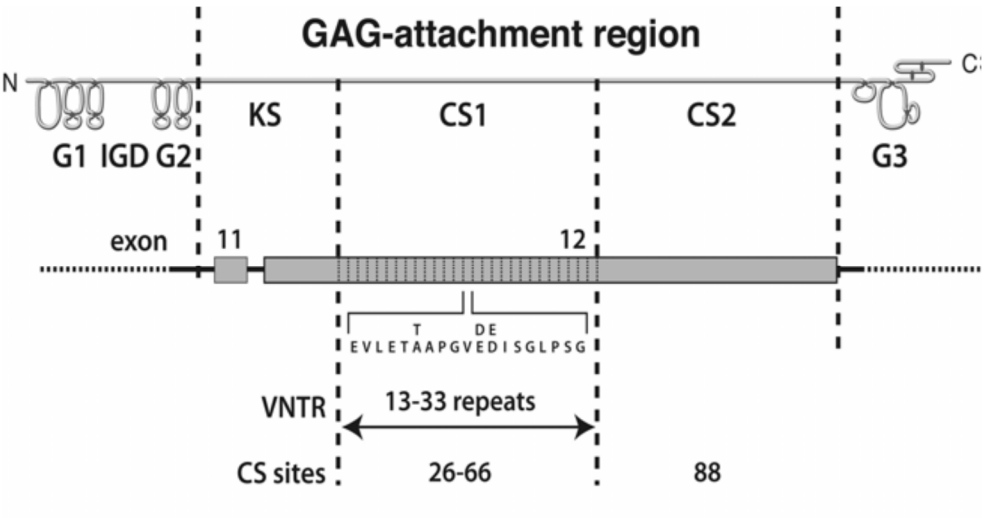


Figure 4.4. Human *aggrecan* structure in relation to the variable number of tandem repeat (VNTR) array polymorphism repeat of exon 12.

SNPs in coding regions of ACAN

After the elimination of UTR and intron variants for effects in splicing, we concentrated on variants in coding regions. A total of 23 SNPs and 1 deletion unique to both or either dwarf from the Supplemental Table 4.1 is listed below in Table 4.6. Variants which altered amino acids or were INDELs were considered as most likely candidates for further investigation for dwarfism. These criteria resulted in the removal of 12 synonymous SNPs with 11 variants remaining.

Table 4.6. 23 SNPs and 1 deletion located in exons unique to both or either dwarf. Variant change of reference amino acid noted with its location. Haplotype results of the two dwarf samples 2701 and 2906 are shown for each variant. Reference nucleotides and alternative variants are shown in forward reading frame.

ACAN	**POS**	**REF**	**ALT**	**AA#**	**2701**	**2906**
chr1 exon 16	g.95256565	C	G	C2582S	C/G	C/G
chr1 exon 14	g.95258406	A	C	W2525G	A/C	A
chr1 exon 13	g.95258999	C	T	A2479T	C	C/T
chr1 exon 12	g.95270440	C	T	syn	C/T	C/T
chr1 exon 12	g.95270484	T	C	R2366G	C	C
chr1 exon 12	g.95270801	G	A	A2361V	G/A	G/A
chr1 exon 12	g.95270812	C	A	syn	C/A	C/A
chr1 exon 12	g.95271190	C	T	syn	C/T	C/T
chr1 exon 12	g.95273047	C	T	syn	C	C/T
chr1 exon 12	g.95273221	A	C	syn	A/C	A/C
chr1 exon 12	g.95273545	G	A	syn	G/A	G/A
chr1 exon 12	g.95273562	A	G	S1340P	A/G	A/G
chr1 exon 12	g.95274016	G	T	Syn	G	G/T
chr1 exon 12	g.95274097	A	G	Syn	A/G	A/G
chr1 exon 12	g.95274456	CTA	C	DEL	CTA/C	CTA/C
chr1 exon 12	g.95274459	GATGAATCTCTTCTCCA GAAGGAAGTCCACCAA GGTCCTCTACTCCAGAG GCAGTGGGCTCT	G	DEL	REF/G	REF/G
chr1 exon 12	g.95274849	C	T	G911R	C/T	C/T
chr1 exon 12	g.95274878	G	A	A901V	G/A	G/A
chr1 exon 12	g.95274967	T	C	Syn	T/C	T/C
chr1 exon 7	g.95284467	C	T	S445I	C/T	C/T
chr1 exon 7	g.95284510	G	A	Syn	G/A	G/A
chr1 exon 7	g.95284530	C	T	V424M	C/T	C/T
chr1 exon 7	g.95284551	G	C	P417A	G/C	G/C
chr1 exon 6	g.95286209	G	A	Syn	G/A	G/A
chr1 exon 3	g.95291302	C	T	Syn	C	C/T

The remaining 11 SNPs which altered amino acids were subjected to SIFT analysis using the Ensembl web tool VEP with results shown in Table 4.7. These 11 variants were investigated for presence in other breeds in Ecab 3.0 of Ensembl and WGS database of 230 Thoroughbreds. Six variants were present in Ecab 3.0 or WGS of 230 Thoroughbreds.

Table 4.7. Genomic locations for 11 SNPs in *ACAN* which alter amino acids with SIFT scores. Variants highlighted in yellow were present in Ecab 3.0 or WGS of 230 Thoroughbreds. Reference nucleotides and alternate variants are in forward reading frame.

ACAN	Position	Reference	Alternative	AA position	SIFT Score
chr1 exon 16	95256565	C	G	C2582S	0 deleterious
chr1 exon 14	95258406	A	C	W2525G	0 deleterious
chr1 exon 13	95258999	C	T	A2479T	0 deleterious
chr1 exon 12	95270484	T	C	R2366G	0.53 tolerated
chr1 exon 12	95270801	G	A	A2361V	0.3 tolerated
chr1 exon 12	95273562	A	G	S1340P	0.02 deleterious
chr1 exon 12	95274849	C	T	G911R	0.51 tolerated
chr1 exon 12	95274878	G	A	A901V	1.0 tolerated
chr1 exon 7	95284467	C	T	S445I	0.40 tolerated
chr1 exon 7	95284530	C	T	V424M	0.1 tolerated
chr1 exon 7	95284551	G	C	P417A	1.0 tolerated

Table 4.8 lists 5 SNPs which were not present in Ecab 3.0 from Ensembl or WGS of 230 Thoroughbreds. An addition to this list is g.95284530 which is annotated in Ecab 3.0 as the *D2* dwarf variant. It was retained in final the list due to its association with the *D2* dwarf phenotype which both dwarf samples possess one copy, and it is unique to the Miniature horse.

Table 4.8. Six coding variants identified from WGS in two heterozygous *D2* dwarf samples which alter amino acids with SIFT scores. Alternative variants are not present in other breeds. Reference nucleotides and alternate variants are shown in froward reading frame.

ACAN	Position	Reference	Alternative	AA position	2701	2906	SIFT Score
chr1 exon 16	95256565	C	G	C2582S	C/G	C/G	0 deleterious
chr1 exon 14	95258406	A	C	W2525G	A/C	A	0 deleterious
chr1 exon 13	95258999	C	T	A2479T	C	C/T	0 deleterious
chr1 exon 12	95270801	G	A	A2361V	G/A	G/A	0.3 tolerated
chr1 exon 12	95274849	C	T	G911R	C/T	C/T	0.51 tolerated
chr1 exon 7	95284530	C	T(*D2*)	V424M	C/T	C/T	0.1 tolerated

PredictSNP analysis

Six variants unique to Miniature horses from Table 4.8, which caused an amino acid change, were submitted to PredictSNP *in silico* analysis for potential to cause impaired protein function. Three of the variants were considered to have a neutral effect by overall probability. Three variants showed high probabilities of deleterious effects over the six best performing analysis tools. *In silico* analysis of *D2* variant V424M using PredictSNP shows across five proprietary analysis tools that it has a 74% probability to have a neutral effect on the mature protein. Table 4.9 shows probabilities of non-reference variants to

cause deleterious effects. Variant g.95256565 C>G (forward strand) previously has not shown association with a dwarf phenotype, however *in silico* analysis and haplotype association warranted further investigation.

Table 4.9. Six SNP variants possessed by 2701 and 2906 indicating PredictSNP probabilities for deleterious effect of protein function. Highlights in blue are predicted to be neutral with respective probabilities. Highlights in yellow are predicted to be deleterious with respective probabilities. Variants g.95258999, g.95258406, and g.95256565 show 79%, 87% and 72% probabilities for deleterious effect.

Mutation	PredictSNP	MAPP	PhD-SNP	PolyPhen-1	PolyPhen-2	SIFT	SNAP
g.95284530	74%	NA	83%	67%	45%	65%	55%
g.95274849	75%	NA	83%	67%	61%	76%	62%
g.95270801	74%	74%	72%	67%	61%	53%	58%
g.95258999	79%	86%	86%	59%	81%	79%	50%
g.95258406	87%	82%	86%	74%	81%	79%	85%
g.95256565	72%	62%	88%	67%	54%	79%	72%

SNP evaluation with Sanger Sequence reads of Miniature horses

Six variants unique to Miniatures remained as candidates for further investigation for association to a dwarf phenotype. Sanger sequence reads from Eberth *et al*. (2018) were used to assess their presence in genotyped dwarfs and carriers. The rationale was to eliminate variants that are present in carriers of other dwarf mutations which are of normal non-dwarf phenotype. This will show the variant in question as unlikely to be deleterious if a normal non-dwarf Miniature already possesses a dwarf mutation and possessed a copy of the variant in question.

Variant g. 95284530 C>T (*D2*) (froward reading frame) was not present in Thoroughbreds and is associated with a dwarf phenotype. It is annotated in Ensembl as associated with a dwarf phenotype *D2* we previously reported (Eberth et al. 2018). The Miniature populations have previously been genotyped for this variant and have shown to be in complete LD with a dwarf phenotype and segregates independently from all other dwarf variants. SIFT score 0.1 (tolerated) and PredictSNP score of 74% probability of neutral effect.

Variant g. 95274849 C>T was not present in Thoroughbreds or Ecab 3.0 but was present in Miniatures. Analysis of sequence reads of carriers showed normal carriers homozygous for the alternative variant. SIFT analysis also shows it to be tolerated with a score of 0.51 and PredictSNP score of 75% probability of neutral effect.

Variant g.95270801 G>A was not present in Thoroughbreds or Ecab 3.0 but was present in Miniatures. Sequence read alignment and analysis showed this variant to be rare in the Miniature population, however it showed heterozygosity in two homozygous *D2* dwarf samples. It was not present in sequences of carriers or non-carriers. All samples of all genotypes were not completely sequenced. This variant showed a SIFT score of 0.3 (tolerated) and PredictSNP score of 74% probability of neutral effect.

Variant g.95256565 C>G (forward reading frame) was not present in Thoroughbreds or Ecab 3.0 but was present in Miniatures. Analysis of sequence reads showed high correlation to homozygous *D2* dwarf samples but showed heterozygosity in a compound heterozygous dwarf genotyped as *D1*/*D4*. Not all samples had Sanger sequences available to assess genotypes. This variant was a candidate to investigate further with a SIFT score of 0 and PredictSNP score of 72% probability of deleterious effect.

Variants g.9525899 C>T and g.95258406 A>C (froward reading frame) were not present in Thoroughbreds or Ecab 3.0. In Miniature horse sequence reads, three dwarfs possessing an unknown dwarf variant from Eberth *et al.* (2018) showed heterozygosity at g.9525899 C>T and no other samples sequenced possessed the alternate variant. Analysis of all sequence reads available from Eberth *et al*. (2018) for variant g.95258406 A>C showed only one sample heterozygous for the alternate variant, sample 2701. Both g.9525899 C>T and g.95258406 A>C were candidates to investigate further with a SIFT score of 0 (deleterious) and PredictSNP score of 79% and 87% probability of being deleterious. Three custom TaqMan® assays were designed and tested on Miniature and other breed populations for variants g.9525899 C>T aa. A2479T, g.95258406 A>C aa.W2525G and g.95256565 C>G aa. C2582S.

Investigation for variation in D2 dwarf haplotype

A *D2* dwarf haplotype was identified using Sanger sequence reads of eight homozygous *D2* dwarfs (acc#2699, 2700, 2705, 2706, 2707, 2708, 2709, 2789) from Eberth *et al*. (2018) compared to the 24 WGS variants from the two heterozygous *D2* dwarfs shown in Table 4.10. Only coding exons were available in samples from Eberth *et al*. (2018). The VNTR region in exon 12 could not be completely assessed due to unreliable sequence reads. Not all individuals of *D2/D2* genotype were completely sequenced for all *ACAN* exons in Eberth *et al*. (2018). Synonymous variants were included to give as complete haplotype as possible from the samples sequenced. Variants which altered amino acids were compared to the copy number results of the two dwarfs 2701 and 2906 to investigate for potential cause of the *D2* phenotype. Sanger sequence reads of carriers of the *D2* variant were examined for heterozygosity for the *D2/D2* haplotype. Ten variants, highlighted in yellow, are present in homozygous form with the *D2* phenotype including the *D2* variant g.95284530 C>T (forward reading frame). The data in this table are consistent with a minimum number of 2 haplotypes for *D2* when considering these variants.

Table 4.10. *D2* haplotype generated from exon Sanger reads with sample count and genotypes listed at each location. This haplotype was compared to WGS of two *D2/N* heterozygous dwarfs with all unique coding and non-coding variants generating a consensus *D2* haplotype. Reference nucleotides and alternate variants are shown in forward reading frame.

Genomic Position	Ref Seq	Genotype 2701	Genotype 2906	Genotype #1 *D2/D2*	Genotype #2 *D2/D2*	Consensus Haplotype
g.95256565	C	C/G	C/G	2: G/G		G
g.95258406	A	A/C	A	1: A/A	1: A/C	A
g.95258999	C	C	C/T	1: C/C	1: C/T	C
g.95268662	G	G/A	G/A	INTRON		A
g.95269519	G	G/A	G/A	INTRON		A
g.95270440	C	C/T	C/T	2: C/C	1: C/T	C
g.95270484	T	C	C	3: C/C		C
g.95270801	G	G/A	G/A	2: A/A	1: G/A	A
g.95270812	C	C/A	C/A	3: A/A		A
g.95271190	C	C/T	C/T	1: T/T	1: C/T	T
g.95273047	C	C	C/T	VNTR		C
g.95273221	A	A/C	A/C	VNTR		C
g.95273545	G	G/A	G/A	VNTR		A
g.95273562	A	A/G	A/G	VNTR		G
g.95274016	G	G	G/T	VNTR		G
g.95274097	A	A/G	A/G	VNTR		G
g.95274456	CTA	CTA/C	CTA/C	VNTR		C
g.95274459	GATGAATCTCTTCTCCAG AAGGAAGTCCACCAAGGT CCTCTACTCCAGAGGCAG TGGGCTCT	REF/G	REF/G	VNTR		G
g.95274849	C	C/T	C/T	3: T/T		T
g.95274878	G	G/A	G/A	3: A/A		A
g.95274967	T	T/C	T/C	3: C/C		C
G.95281618	T	T/G	T/G	INTRON		G
g.95284467	C	C/T	C/T	1: C/C	7: C/T	C
g.95284510	G	G/A	G/A	8: A/A		A
g.95284530	C	C/T	C/T	8: T/T		T
g.95284551	G	G/C	G/C	1: G/G	7: C/C	G/C
g.95286209	G	G/A	G/A	4: G/G		G
g.95291302	C	C	C/T	5: C/C		C
g.95291513	A	A/G	A/G	INTRON		G
g.95299780	G	G/C	G/C	INTRON		C
g.95300373	A	A/T	A/T	INTRON		T
g.95300602	T	T/C	T/C	INTRON		C
g.95310855	C	C/T	C/T	INTRON		T
g.95313448	G	G/C	G/C	INTRON		C

Genotypes of heterozygous *D2* dwarfs 2701 and 2906 at genomic locations homozygous in *D2/D2* dwarfs were compared as shown in Table 4.9. Whole genome unique variants in non-coding regions and the VNTR were added to uncover the most informative *D2* haplotype possible from the two heterozygous dwarfs. The only two SNPs that altered amino acids present in both heterozygous dwarfs and homozygous dwarfs and not present in other breeds were g.95284530 C>T and g.95256565 C>G. All other SNPs

were either synonymous or were present in other breeds. Variants g.9525899 C>T and g.95258406 A>C were not present in other breeds but were not possessed by both 2701 and 2906. The VNTR deletion is the only INDEL present in both dwarfs as a single copy and may be a variant in the *D2* haplotype; however, as of this analysis it does not appear deleterious.

Custom TaqMan® assay results

Genotyping for variant g.95256565 G>C aa. C2582S

A Miniature population totaling 247 individuals was tested for SNP variant g.95256565 G>C (reverse reading frame). The first group included 19 dwarfs of unknown parentage which possessed one or no dwarf variants *D1-D4*. Variant g.95256565 G>C (reverse reading frame) showed complete association to the *D2* variant, shown in Table 4.11. Previous sequencing by Eberth *et al*. (2018) showed no association to a dwarf phenotype. A population of 247 Miniatures of all genotypes was genotyped for variant g.95256565 G>C (reverse reading frame), shown in Table 4.12.

Table 4.11. 19 dwarf horses possessing one or no *D1-D4* and genotype results for variants Brazil haplotype identified as *D* and g.95256565 G>C. Highlighted in yellow shows complete association with heterozygous *D2* dwarfs possessing an unknown variant. Genotype assay and variant calls are shown in reverse reading frame.

Acc. #	*D1-D4*	Brazil haplotype	95256565 G/C
2694	*N/D2*	*N/D*	G/C
2698	*N/N*	*N/N*	G/G
2701	*N/D2*	*N/N*	G/C
2741	*N/D4*	*N/N*	G/G
2742	*N/*D3*	*N/N*	G/G
2745	*N/D4*	*N/N*	G/G
2746	*N/D2*	*N/N*	G/C
2906	*N/D2*	*N/N*	G/C
3545	*N/N*	*N/D*	G/G
3547	*N/D2*	*N/N*	G/C
3563	*N/*D3*	*N/N*	G/G
3566	*N/D2*	*N/N*	G/C
3567	*N/D2*	*N/D*	G/C
3568	*N/D2*	*N/N*	G/C
3578	*N/D2*	*N/D*	G/C
3579	*N/D2*	*N/N*	G/C
3591	*N/D2*	*N/D*	G/C
3596	*N/D2*	*N/D*	G/C
3612	*N/D4*	*N/N*	G/G

Table 4.12. 247 Miniatures of all genotypes possessing *N-D4* plus Brazil variant and genotype count for variant g.95256565 G>C. Results from Table 10 are included. Blue highlights show samples that demonstrate exceptions for *D2* association with g.95256565 G>C. Genotype assay and variant calls are shown in reverse reading frame.

Genotype	Phenotype	95256565 G/G	95256565 G/C	95256565 C/C
N/N- 74	Normal	72	2	0
N/D1- 27	Normal	27	0	0
N/D2- 44	Normal	2	42	0
N/D3- 4*	Normal	4	0	0
N/D4- 2	Normal	2	0	0
N/Brazil-0	Normal	0	0	0
N/N-1	Dwarf	1	0	0
N/Brazil-1	Dwarf	1	0	0
N/D1-0	Dwarf	0	0	0
N/D2-12	Dwarf	0	12	0
N/D3-2	Dwarf	2	0	0
N/D4-2	Dwarf	2	0	0
D1/D1- 3	Dwarf	3	0	0
D1/D2- 9	Dwarf	0	7	2
D1/D3- 2*	Dwarf	2	0	0
D1/D4- 2	Dwarf	1	1	0
D2/D2- 30	Dwarf	3	2	25
D2/D3- 14*	Dwarf	1	13	0
D2/D4- 10	Dwarf	1	9	0
D3/D3*- 0*	Dwarf	0	0	0
D3/D4- 2*	Dwarf	2	0	0
D4/D4- 1	Dwarf	1	0	0
D1/Brazil-0	Dwarf	0	0	0
D2/Brazil-5	Dwarf	0	5	0
D3/Brazil-0	Dwarf	0	0	0
D4/Brazil-0	Dwarf	0	0	0
Totals- 247		**127**	**93**	**27**

Results from Table 4.12 present g.95256565 G>C as highly correlated with variant *D2* and dwarf phenotype for all Miniatures in this study. Blue highlights within Table 4.12 emphasize the groups which showed samples where g.95256565G>C did not correlate to the *D2* variant or phenotype. Exceptions to the presence of g.95256565 G>C with dwarf variant *D2* in homozygous *D2* dwarfs was 5 out of 30, with two possessing one copy of g.95256565 G>C and 3 individuals possessing the reference allele for g.95256565. Individuals with one copy of variant *D2* totaled 94; however, 88 possessed a copy of g.95256565 G>C. Four individuals heterozygous for *D2* possessed two copies of the reference allele of g.95256565. Two individuals with one copy of *D2* possessed two copies of the alternate allele for g.95256565. Individuals which do not possess a copy of dwarf

variant *D2* totaled 123 with 3 of those possessing one copy of the alternate allele for g.95256565.

Whole genome sequence of 101 Japanese Thoroughbreds (Tozaki *et al.* 2010), 534 horses of diverse breeds (Durward-Akhurst *et al.* 2021) and 230 Thoroughbred horses (Bailey, Kalbfleisch and Petersen, unpublished data) were examined for the g.95256565 G>C. No horses possessed a copy of the G>C variant. A set of 93 horses of large breeds was genotyped for variant g.95256565 G>C. This included 20 Standardbreds, 20 Arabians, 25 Thoroughbreds, 12 American Saddlebreds, 10 American Quarter horses and 6 Tennessee Walking horses. Table 4.13 shows genotype results of large breed horses. Variant g.95256565 G>C was present in two Standardbred horses.

Table 4.13. Set of large breeds genotyped for variants g.95256565 G>C. Genotype assay and variant calls are shown in reverse reading frame.

Large Breed horses	**95256565 G/C**
20 Standardbred	2-G/C
20 Arabian	0
356 Thoroughbred	0
12 ASB	0
10 QH	0
6 TWH	0
534 diverse breeds	0
Totals-728	2

Genotype results for variants g.95258406 G>C aa. W2525G

A Miniature population totaling 247 individuals was tested for SNP variant g.95258406 A>C. This included 19 dwarfs of unknown parentage which possessed one or no dwarf variants *D1-D4* and Brazil haplotype. Only one sample, 2701 possessed a copy of the variant g.95258406 A>C highlighted in yellow shown in Table 4.14. Previous sequencing by Eberth *et al*. (2018) showed no association to a dwarf phenotype.

Table 4.14. Genotype and phenotype of 247 Miniature horses and results of g.95258406 A>C. Sample 2701 was the only sample which possessed a single copy.

Genotype *D1-D4*	Phenotype	95258406 A/A	95258406 A/C	95258406 C/C
N/N- 74	Normal	74	0	0
N/D1- 27	Normal	27	0	0
N/D2- 44	Normal	44	0	0
N/D3- 4*	Normal	4	0	0
N/D4- 2	Normal	2	0	0
N/Brazil-0	Normal	0	0	0
N/N-1	Dwarf	1	0	0
N/Brazil-1	Dwarf	1	0	0
N/D1-0	Dwarf	0	0	0
N/D2-12	Dwarf	11	1	0
N/D3-2*	Dwarf	2	0	0
N/D4-2	Dwarf	2	0	0
D1/D1- 3	Dwarf	3	0	0
D1/D2- 9	Dwarf	9	0	0
D1/D3- 2*	Dwarf	2	0	0
D1/D4- 2	Dwarf	2	0	0
D2/D2- 30	Dwarf	30	0	0
D2/D3- 14*	Dwarf	14	0	0
D2/D4- 10	Dwarf	10	0	0
D3/D3*- 0*	Dwarf	0	0	0
D3/D4- 2*	Dwarf	2	0	0
D4/D4- 1	Dwarf	1	0	0
D1/Brazil-0	Dwarf	0	0	0
D2/Brazil-5	Dwarf	5	0	0
D3/Brazil-0*	Dwarf	0	0	0
D4/Brazil-0	Dwarf	0	0	0
Totals- 247		**246**	**1**	**0**

Whole genome sequence of 101 Japanese Thoroughbreds (Tozaki *et al.* 2010), 534 horses of diverse breeds (Durward-Akhurst *et al.* 2021) and 230 Thoroughbred horses (Bailey, Kalbfleisch and Petersen, unpublished data) were examined for g.95258406 A>C. No horses possessed a copy of the A>C variant. A population of breeds of large horses were genotyped for g.95258406 A>C shown in Table 4.15.

Table 4.15. Large breeds genotyped for variants g.95258406 A>C.

Large Breed horses	95258406 A/C
20 Standardbred	0
20 Arabian	0
356 Thoroughbred	0
12 ASB	0
10 QH	0
6 TWH	0
534 diverse breeds	0
Totals- 728	0

Genotype results for SNP g.95258999 C>T aa.A2479T

Nineteen dwarf individuals possessing an unknown disease variant are listed in Table 4.16 below with genotype results for Brazil dwarf haplotype and SNP g.95258999 C>T. The alternate variant, designated *D5*, showed independent segregation to *D1-D4* as well as to the Brazil dwarf variant. Four dwarf individuals possessed an unknown variant yet to be associated with a phenotype. Table 4.17 shows the results of a population of 464 Miniatures. Four horses previously genotyped as non-carriers possessed a single copy of g.95258999 C>T. One dwarf previously genotyped as *N/N* possessed two copies of the alternative variant g.95258999 C>T. Six dwarf horses previously genotyped as heterozygous *N/D* possessed one copy of the alternative variant g.95258999 C>T.

Table 4.16. 19 dwarfs possessing one or no dwarf alleles *D1-D4* and Brazil haplotype. Dwarfs possessing an unknown variant are highlighted in yellow.

Acc. #	*D1-D4*	Brazil haplotype	95258999 C>T	Dwarf Genotype/haplotype
2694	*N/D2*	*N/Brazil*	C/C	*D2/Brazil*
2698	*N/N*	*N/N*	T/T	*D5/D5*
2701	*N/D2*	*N/N*	C/C	*N/D2*
2741	*N/D4*	*N/N*	C/T	*D4/D5*
2742	*N/D3**	*N/N*	C/C	*N/D3**
2745	*N/D4*	*N/N*	C/C	*N/D4*
2746	*N/D2*	*N/N*	C/C	*N/D2*
2906	*N/D2*	*N/N*	C/T	*D2/D5*
3545	*N/N*	*N/Brazil*	C/C	*N/Brazil*
3547	*N/D2*	*N/N*	C/C	*N/D2*
3563	*N/D3**	*N/N*	C/T	*D3*/D5*
3566	*N/D2*	*N/N*	C/T	*D2/D5*
3567	*N/D2*	*N/Brazil*	C/C	*D2/Brazil*
3568	*N/D2*	*N/N*	C/T	*D2/D5*
3578	*N/D2*	*N/Brazil*	C/C	*D2/Brazil*
3579	*N/D2*	*N/N*	C/T	*D2/D5*
3591	*N/D2*	*N/Brazil*	C/C	*D2/Brazil*
3596	*N/D2*	*N/Brazil*	C/C	*D2/Brazil*
3612	*N/D4*	*N/N*	C/C	*N/D4*

Table 4.17. Genotype and phenotype of 464 Miniature horses and results of g.95258999 C>T. Samples from Table 4.14 are included. Samples highlighted in yellow show presence of copy number for alternative variant.

Genotype	Phenotype	95258999 C/C	95258999 C/T	95258999 T/T
N/N- 273	Normal	269	4	0
N/D1- 31	Normal	31	0	0
N/D2- 51	Normal	51	0	0
N/D3- 4*	Normal	4	0	0
N/D4- 5	Normal	5	0	0
N/Brazil-7	Normal	7	0	0
Brazil/Brazil-1	Normal	1	0	0
N/N-1	Dwarf	0	0	1
N/Brazil-1	Dwarf	1	0	0
N/D1-0	Dwarf	0	0	0
N/D2-7	Dwarf	3	4	0
N/D3-2*	Dwarf	3	1	0
N/D4-3	Dwarf	2	1	0
D1/D1- 3	Dwarf	3	0	0
D1/D2- 9	Dwarf	9	0	0
D1/D3- 2*	Dwarf	2	0	0
D1/D4- 2	Dwarf	2	0	0
D2/D2- 30	Dwarf	30	0	0
D2/D3- 14*	Dwarf	14	0	0
D2/D4- 10	Dwarf	10	0	0
D3/D3*- 0*	Dwarf	0	0	0
D3/D4- 2*	Dwarf	2	0	0
D4/D4- 1	Dwarf	1	0	0
D1/Brazil-0	Dwarf	0	0	0
D2/Brazil-5	Dwarf	5	0	0
D3/Brazil-0*	Dwarf	0	0	0
D4/Brazil-0	Dwarf	0	0	0
Totals- 464		**453**	**10**	**1**

Genotype results of large horse breed individuals for variant g.95258999 C>T

Whole genome sequence of 101 Japanese Thoroughbreds (Tozaki *et al.* 2010), 534 horses of diverse breeds (Durward-Akhurst *et al.* 2021) and 230 Thoroughbred horses (Bailey, Kalbfleisch and Petersen, unpublished data) were examined for g.95258999 C>T. No horses possessed a copy of the alternate variant. A collection of 93 horses of large breeds was genotyped for variant g.95258999 C>T. This included 20 Standardbreds, 20 Arabians, 25 Thoroughbreds, 12 American Saddlebreds, 10 American Quarter horses and 6 Tennessee Walking horses. Table 4.18 shows genotype results with no large breed individuals possessing the alternate variant.

Table 4.18. Large breeds genotyped for variants g.95258999 C>T.

Large Breed horses	95258999 C/T
20 Standardbred	0
20 Arabian	0
356 Thoroughbred	0
12 ASB	0
10 QH	0
6 TWH	0
534 diverse breeds	0
Totals- 728	**0**

Of the 19 dwarf horses which had *N* designation before, 13 can be accounted for with variants, *D5*, Brazil haplotype variant, g.95258999 C>T, and g.95258406 A>C shown in summary Table 4.19. Six dwarf horses remain, highlighted in yellow, which possess one copy of various dwarf variants suggesting yet another dwarf allele independently segregating in the Miniature horse exists.

Table 4.19. Summary of 19 dwarfs possessing an unknown dwarf variant and genotype results. Highlighted in yellow are dwarfs possessing an unknown variant.

Acc. #	*D1-D4, N*	*Genotype Summary*
2694	*N/D2*	*D2/Brazil*
2698	*N/N*	*D5/D5*
2701	*N/D2*	g.95258406 A>C */D2*
2741	*N/D4*	*D4/D5*
2742	*N/D3**	*N/D3**
2745	*N/D4*	*N/D4*
2746	*N/D2*	*N/D2*
2906	*N/D2*	*D2/D5*
3545	*N/N*	*N/Brazil*
3547	*N/D2*	*N/D2*
3563	*N/D3**	*D3*/D5*
3566	*N/D2*	*D2/D5*
3567	*N/D2*	*D2/Brazil*
3568	*N/D2*	*D2/D5*
3578	*N/D2*	*D2/Brazil*
3579	*N/D2*	*D2/D5*
3591	*N/D2*	*D2/Brazil*
3596	*N/D2*	*D2/Brazil*
3612	*N/D4*	*N/D4*

Discussion

Chondrodysplastic dwarfism in Miniatures has been demonstrated to involve multiple alleles in *ACAN* segregating independently within populations world-wide. These alleles can combine to produce homozygous and compound heterozygous phenotypes which present a range of physical malformations dependent on the variants inherited by the individual. Previous works have documented dwarf individuals which possess unknown variants presenting phenotypes like those seen in chondrodysplastic dwarfism. This work involved a collection of dwarf individuals with unknown parentage which possessed unknown variants to explain their phenotype.

Since the publications of Metzger *et al*. (2016) and Eberth *et al.* (2018), Andrade *et al.* (2020) discovered a haplotype in *ACAN* associated with a dwarf phenotype in Miniatures of Brazil. This variant is present in American Miniatures; however, it is not in complete linkage disequilibrium with the causative mutation (see Chapter III of this dissertation). This variant can still be utilized to account for dwarf individuals that do not possess other dwarf variants. Five dwarfs possessing an unknown variant also possessed one copy of the Brazil haplotype and one copy of a *D1-D4* variant and no copies of g.95258999 or g.95258406. These five dwarfs are clarified by their genotypes. One dwarf from the 19 dwarfs possessing an unknown allele, possessed a single copy of the Brazil haplotype.

Whole genome sequencing analysis revealed both 2701 and 2906 possessed one copy of a deletion within the VNTR of exon 12. This deletion was presented as in-frame and totaled 63 nucleotides which is a complete repeat within the VNTR. This region is documented in Ecab 3.0 as well as within this dissertation as being variable in copy number within an individual and within the horse species. Algorithmic analysis of whole genome short read sequences is utilized to determine copy number of genetic regions to ascertain the presence of deletions in heterozygous form. Alignment of these reads in non-repeated regions are straight forward. In large highly conserved repeated regions with variable copy numbers, alignment may be difficult to visualize. From this work utilizing the available analysis tools, the deletion detected in both individuals does not appear to be abnormal or

out of frame. It does not preclude it from future investigation for the cause of a dwarf phenotype. The VNTR region remains a complex area for genomic analysis. This region, and the in-frame deletion both dwarfs possess as heterozygotes, may contain the true cause of the *D2* phenotype. This region may be resolved in the future with better genomic analysis tools.

Variant g.95256565 G>C in exon 16 showed linkage disequilibrium to the *D2* dwarf phenotype and *D2* variant g.95,284,530 C>T. However, in the Eberth *et al.* (2018) population 14 exceptions were present for g.95256565 G>C, where the designated *D2* variant g.95,284,530 C>T has shown complete association to a dwarf phenotype. Large horse breed genotyping revealed two Standardbreds as heterozygous for the variant g.95256565 G>C. The presence of g.95256565 G>C in 2 Standardbred individuals does not preclude this variant as deleterious to cause malformations, no physical information was available for these two horses. Yet, no other breeds have shown to be afflicted with chondrodysplastic dwarfism other than the Miniature horse and Shetland Ponies. It has been described previously that the Miniature horse has significant genetic influence from the Shetland pony. Studies show *ACAN* affects bone growth and maturation, and this variant (g.95256565 G>C) may have some other effect to growth which does not manifest as dwarfism, such as reduced size, considering it is not present in the other breeds tested except two Standardbreds. It is conceivable this SNP causes a phenotype which would have an intragenic pleiotropic or additive effect on the overall phenotype of carriers or dwarfs. This interaction was not within the realm of this study to ascertain phenotype variations of *D2* carriers or *D2* dwarfs which did not possess g.95256565 G>C to the those that did. Future investigation may be considered to distinguish phenotypic variation for g.95256565 G>C and g. 95,284,530 C>T.

PredictSNP analysis grades g.95,284,530 C>T (*D2*) with a 74% probability of a neutral effect with a SIFT score of 0.1 and tolerated. Whereas variant g.95256565 G>C with a 72% probability of a deleterious effect and a SIFT score of 0 and deleterious. Whole genome sequencing analysis of 2701 and 2906 did not produce another variant with a more robust association or cause to the *D2* phenotype. From this study, g.95,284,530 C>T

remains the most reliable marker for *D2* genotyping. *In silico* analysis of sequence variants to determine deleterious effects on the function of a mature protein can be imprecise and are used more as a guide. Using various *in silico* tools, the current *D2* designate in exon 7 has shown to likely have a neutral effect to the function of mature *aggrecan*. However, this designate continues to be in complete linkage disequilibrium with a dwarf phenotype. There is a possibility it may not be the cause of the dwarf phenotype associated with it, but another mutation has yet to be discovered which is better suited to explain the phenotype.

Variant g.95258406 A>C is unique to sample 2701 from the evidence collected in this work. No other Miniature or horse of a large breed possessed a copy of this variant from 1,392 genotypes investigated. This variant may have deleterious effects and is the cause for the phenotype expressed by 2701; however, no other association could be ascertained from this work. Future investigation of this variant may be warranted to determine if it is associated with a diseased phenotype.

Genotype results showed g.95258999 C>T has complete association with a dwarf phenotype. Carriers of variant g.95258999 C>T do not possess any other dwarf variant. Carriers of *D1-D4* and the Brazil haplotype variant do not possess variant g.95258999 C>T. Dwarf 2698 possesses two copies of g.95258999 C>T and no other dwarf variant. Six individuals expressing a dwarf phenotype and possessing one dwarf allele also possess one copy of g.95258999 C>T. Genotype results demonstrated g.95258999 C>T has independent distribution to *D1-D4* and Brazil dwarf variants in 464 individuals. Four non-carriers of *D1-D4* or the Brazil SNP show as carriers for g.95258999 C>T. We were not able to show a familial inheritance pattern with the samples in this study. Allele frequency of g.95258999 C>T in the population of 464 is 0.013 with a carrier rate of 0.025. We designated this dwarf variant as *D5*.

From the evidence presented in this study, at least one and possibly more unknown dwarf variants are segregating within the Miniature population. As mentioned earlier, the VNTR region of exon 12 is quite complex and future improvements to whole genome sequencing and analysis may be able to better explain the variations and cause of disease.

The SNPs annotated in this study that cause an amino acid change at the protein level have the potential to have other effects within the Miniature horse that we are not aware of at this time. Considering Miniature horses are the smallest of the equines and GWAS studies have implicated *ACAN* as a contributor to stature in organisms, the diminutive size could be a result from a unique mutation(s) in *ACAN* that does not result in a diseased animal. There are possibly other dwarf types that exist in Miniature horses that were not seen or acquired for this dwarfism study. If other chondrodysplastic dwarf phenotypes are found to exist, some of the variants that are present as non-causative at this time should be considered for investigation.

Supplementary Table 4.1. Alignment results of VCF sequence files for dwarfs 2701 and 2906 against VCF file sequences of two Saddlebreds and Twilight. 157 variants unique to either or both dwarfs 2701 and 2906.

#CHROM	POS	REF	ALT	2701	2906
chr1 exon 16	95256565	C	G	CG	CG
chr1 exon 14	95258406	A	C	AC	AA
chr1 exon 13	95258999	C	T	CC	CT
chr1	95261078	T	C	CC	CC
chr1	95261136	G	T	TT	TT
chr1	95261248	C	T	TT	TT
chr1	95261673	T	C	CC	CC
chr1	95261885	G	T	GT	GG
chr1	95262110	A	G	GG	GG
chr1	95262408	T	C	CC	CC
chr1	95262500	A	G	GG	GG
chr1	95262984	C	G	GG	GG
chr1	95263246	A	G	GG	GG
chr1	95263547	A	C	CC	CC
chr1	95264235	A	C	CC	CC
chr1	95264570	C	T	TT	TT
chr1	95264764	G	A	AA	AA
chr1	95265119	G	C	CC	CC
chr1	95266653	C	T	TT	TT
chr1	95267700	T	C	CC	CC
chr1	95268164	C	T	TT	CT
chr1	95268329	A	G	AG	AA
chr1	95268662	G	A	GA	GA
chr1	95268850	T	C	CC	CC
chr1	95268933	C	A	CA	CC
chr1	95269352	A	G	AG	AG
chr1	95269501	G	A	GA	GA
chr1	95269519	G	A	GA	GA
chr1	95269909	T	C	TC	TC
chr1 exon 12	95270440	C	T	CT	CT
chr1 exon 12	95270484	T	C	CC	CC
chr1 exon 12	95270801	G	A	GA	GA
chr1 exon 12	95270812	C	A	CA	CA
chr1 exon 12	95271190	C	T	CT	CT
chr1 exon 12	95273047	C	T	TT	CT
chr1 exon 12	95273221	A	C	AC	AC
chr1 exon 12	95273545	G	A	GA	GA
chr1 exon 12	95273562	A	G	AG	AG

Supplementary Table 4.1. (*continued*) Alignment results of VCF sequence files for dwarfs 2701 and 2906 against VCF file sequences of two Saddlebreds and Twilight. 157 variants unique to either or both dwarfs 2701 and 2906.

chr1 exon 12	95274016	G	T	GG	GT
chr1 exon 12	95274097	A	G	AG	AG
chr1 exon 12	95274456	CTA	C	CTA/C	CTA/C
chr1 exon 12	95274459	GATGAATCTCTTCTCC AGAAGGAAGTCCACC AAGGTCCTCTACTCCA GAGGCAGTGGGCTCT	G	GATGAATCTCTTCTCC AGAAGGAAGTCCACC AAGGTCCTCTACTCCA GAGGCAGTGGGCTCT/G	GATGAATCTCTTCTCC AGAAGGAAGTCCACC AAGGTCCTCTACTCCA GAGGCAGTGGGCTCT/G
chr1 exon 12	95274849	C	T	CT	CT
chr1 exon 12	95274878	G	A	GA	GA
chr1 exon 12	95274967	T	C	TC	TC
chr1	95276279	T	C	TC	TC
chr1	95277207	T	C	TC	TC
chr1	95278229	A	C	AC	AC
chr1	95278836	T	C	TC	TC
chr1	95279540	C	G	CG	CG
chr1	95280925	CT	C	C/C	C/C
chr1	95281330	A	C	AC	AC
chr1	95281618	T	G	TG	TG
chr1 exon 7	95284467	C	T	CT	CT
chr1 exon 7	95284510	G	A	GA	GA
chr1 exon 7	95284530	C	T	CT	CT
chr1 exon 7	95284551	G	C	GC	GC
chr1	95285481	A	G	GG	GG
chr1	95285845	G	T	GT	GT
chr1 exon 6	95286209	G	A	GA	GA
chr1	95287103	G	A	GA	GA
chr1	95288099	G	T	GT	GT
chr1	95288235	G	C	GC	GC
chr1	95288367	G	C	GG	GC
chr1	95288613	CT	C	CT/CT	CT/C
chr1	95288937	G	A	AA	AA
chr1	95289900	C	A	CA	CA
chr1	95290356	C	A	CA	CA
chr1	95290406	T	A	TA	TA
chr1	95290568	C	T	CT	CC
chr1	95290931	C	T	CT	CT
chr1	95290957	G	A	AA	AA
chr1 exon 3	95291302	C	T	CC	CT
chr1	95291513	A	G	AG	AG
chr1	95291540	T	C	TT	TC

Supplementary Table 4.1. (*continued*) Alignment results of VCF sequence files for dwarfs 2701 and 2906 against VCF file sequences of two Saddlebreds and Twilight. 157 variants unique to either or both dwarfs 2701 and 2906.

chr1	95291965	G	A	GA	GG
chr1	95292049	CG	C	CG/C	CG/C
chr1	95293327	A	G	AG	AG
chr1	95293373	C	A	CA	CC
chr1	95293571	G	A	GA	GA
chr1	95293739	T	C	CC	CC
chr1	95293853	G	A	GA	GA
chr1	95296240	TTTC	T	TTTC/T	TTTC/TTTC
chr1	95296246	C	CT	C/CT	CT/CT
chr1	95297169	G	A	GG	GA
chr1	95297311	T	C	TC	CC
chr1	95298156	C	T	CT	CC
chr1	95299780	G	C	GC	GC
chr1	95299806	G	A	GA	GG
chr1	95300282	TG	T	"-/-"	TG/T
chr1	95300289	T	G	GG	TG
chr1	95300373	A	T	AT	AT
chr1	95300469	T	G	GG	TG
chr1	95300602	T	C	TC	TC
chr1	95300722	G	A	GA	GA
chr1	95300982	T	A	AA	TA
chr1	95302081	A	T	TT	AT
chr1	95302189	C	T	CT	CC
chr1	95302214	A	G	GG	AG
chr1	95302220	G	T	TT	GT
chr1	95302221	C	T	TT	CT
chr1	95302508	T	C	TC	TT
chr1	95303036	TC	T	TC/T	TC/TC
chr1	95305516	A	G	AG	AG
chr1	95305626	G	A	GA	GA
chr1	95305657	A	T	AT	AA
chr1	95305700	G	A	GA	GG
chr1	95306428	G	A	GA	GA
chr1	95306489	A	G	AG	AG
chr1	95306498	A	G	AG	AG
chr1	95306668	C	T	CT	CC
chr1	95307925	A	G	GG	AG
chr1	95308944	T	C	TC	TC

Supplementary Table 4.1. (*continued*) Alignment results of VCF sequence files for dwarfs 2701 and 2906 against VCF file sequences of two Saddlebreds and Twilight. 157 variants unique to either or both dwarfs 2701 and 2906.

chr1	95309524	G	A	GA	GA
chr1	95309544	C	T	CT	CT
chr1	95310263	A	G	GG	AG
chr1	95310831	C	A	AA	CA
chr1	95310855	C	T	CT	CT
chr1	95311082	T	G	GT	GT
chr1	95312216	G	A	AA	AA
chr1	95312374	C	CG	CG/CG	C/CG
chr1	95312838	T	C	TC	TC
chr1	95312942	T	C	TC	TT
chr1	95313068	C	A	CA	CA
chr1	95313314	A	AT	A/AT	A/AT
chr1	95313448	G	C	GC	GC
chr1	95313514	G	GCC	G/GCC	GCC/GCC
chr1	95314566	C	G	CC	CG
chr1	95315367	T	G	TG	TG
chr1	95315420	CT	CTT,C	CTT/C	CT/CTT
chr1	95315772	C	T	CT	CC
chr1	95315782	C	G	CG/CG	CC
chr1	95315843	G	A	GA	GG
chr1	95316183	AG	A	AG/A	AG/AG
chr1	95316317	T	C	TC	CC
chr1	95317235	G	A	GA	GA
chr1	95317390	T	C	TC	TC
chr1	95317491	C	T	CT	CC
chr1	95317712	T	C	CC	CC
chr1	95317928	G	A	GG	GA
chr1	95318079	A	G	GG	GG
chr1	95318458	C	T	CC	CT
chr1	95319260	G	A	GA	GG
chr1	95319305	A	ACGGGG	A/ACGGGG	A/A
chr1	95319307	A	C	AC	AA
chr1	95319311	A	C	AC	AA
chr1	95319315	A	C	AC	AA
chr1	95319318	CAG	C	CAG/C	CAG/CAG
chr1	95319339	C	T	CT	CT
chr1	95319513	T	G	TT	CT
chr1	95319828	CG	C	CG/CG	CG/C

Supplementary Table 4.1. (*continued*) Alignment results of VCF sequence files for dwarfs 2701 and 2906 against VCF file sequences of two Saddlebreds and Twilight. 157 variants unique to either or both dwarfs 2701 and 2906.

chr1	95320026	C	G	CC	CG
chr1	95320040	C	T	CT	CC
chr1 UTR	95320242	G	A	GG	GA
chr1 UTR	95320330	A	T	TT	TT
chr1 UTR	95320434	G	A	GA	GA
chr1 UTR	95320896	G	C	GG	GC

Chapter V

Changes in breed standard head phenotypes and dwarfism in the Miniature horse

Summary

Facial structure of the Miniature horse has changed significantly due to the desires of breeders since the breed's inception. Facial features described as "straight", or "pony" were the prevailing type in the foundation of the breed. Today, the most desired type is a smaller more refined "dished" or "Arabian type" facial structure. The desired "dished" head is a neotenic effect commonly seen in other breeds of domesticated animals. In people, spatial measurements of head and facial landmarks in-utero and as infants and adults are used to determine abnormal growth or diseased state. Sometimes in animals, a desired phenotype can be the result of a genetic variant that may also produce a diseased state. One such example is the amount of muscling in Quarter horses with HYPP. In the case of Miniature horses, chondrodysplasia-like dwarfism is caused by recessive mutations in *aggrecan*, the main constituent of articular cartilage. Carriers of dwarf variants have a statistically significant but slightly shorter mature height than non-carriers. Dwarf Miniatures have significant structural abnormalities of joints and bones including the head and facial features. Morphometric measurements using a landmark based method were obtained on 106 normal phenotype Miniature horses to compare size variance of specific facial features in relation to carrier status of dwarf variants. Not all dwarf variants were represented in the carriers in this study. Measurements for each facial category for 15 carriers were compared to 91 non-carriers by Mann-Whitney U-tests for statistical calculations (P value significance set at $p<0.05$, two-tailed hypothesis). Measurement results showed no correlation of specific facial features, size, or shape to carrier state.

Introduction

Dwarfism is not a desirable characteristic of miniature horses and is associated with multiple health defects (Eberth *et al.* 2018). Therefore, the high carrier rate for dwarfism associated gene variants among Miniature horses is surprising (Eberth *et al.* 2018; Andrade *et al*. 2020; this dissertation Chapter IV). While two alleles cause dwarfism and health

problems, a single copy may confer characteristics that breeders find desirable. In such cases, heterozygotes may be preferentially selected for breeding. Indeed, Graves *et al*. (2020) reported that non-dwarf carriers of the ACAN variants *D1, D2, D3* and *D4,* had a slightly reduced stature at the withers (1.43 inches). However, this is not a large difference and breeders may be influenced by traits related to the perception of an attractive head. Therefore, this study was initiated to investigate the possibility that other phenotypic differences exist between carriers and non-carriers of dwarfism causing alleles of *ACAN*. In cats, the Scottish-fold breed is based on carrier status for a variant in the gene *TRPV4* affecting cartilage and causing a unique ear-fold phenotype, among other characteristics (Gandolfi *et al.* 2016). In horses, carriers of a variant in the gene *EDNRB* exhibit a white spotting pattern, Frame Overo, that is attractive to some breeders; homozygotes for this variant exhibit Overo Lethal White Foal Syndrome and die shortly after birth due to multiple congenital defects (Metallinos *et al.* 1998; Santschi *et al*. 1998; Yan *et al.* 1998). The genetic disease HYPP in Quarter horses occurs when heterozygote highly muscled parents, which have sodium channel defects, produce a double muscled diseased phenotype offspring with muscle paralysis. Heterozygotes present a desired phenotype as highly muscled but with minimal disease expression (Rudolph *et al*. 1992).

In humans, mutations in *ACAN* have been shown to produce not just diseased phenotypes with abnormal bone growth but heterozygous carriers of some mutations present with shortened stature and other phenotypic characteristics when compared to non-carriers (Stattin *et al.* 2010; Nilsson *et al.* 2014; Quintos *et al*. 2015; van der Steen *et al*. 2016; Dateki *et al*. 2017). One example is spondyloepimetaphyseal dysplasia SEMD type (Tompson *et al.* 2009) showed heterozygous carriers of the recessive mutation presented with a shortened stature and other phenotypic characteristics such as shortened phalanges and facial features. Recently, Sentchordi-Montané *et al*. (2018), described previously uncharacterized variants in *ACAN* with identification of heterozygous mutations in human individuals with short stature, minor skeletal defects and mild facial dysmorphisms.

The focus of this study was on the face. Since the inception of the Miniature horse, the breed standard for the face has changed. The most common facial structures of

founding Miniature bloodstock were described as “straight”, reflecting standard characteristics of American Shetland and American Hackney Pony breeds. These pony breeds had significant genetic influence on the early Miniature horse breed phenotype. Today the desired head phenotype is a “dished” or “Arabian type” (referring to the refined facial structure of the Arabian horse) (AMHA 2023).

Ideal head phenotype for the Miniature horse

The rule book for the AMHA provides descriptions of ideal head structure and type for the Miniature horse (AMHA 2023), quoted here. “The head is beautiful, triangular in shape and comparatively small in proportion to the length of neck and body. The forehead is broad with large, prominent eyes. The eyes are set well apart and are placed approximately 1/3 the distance from the poll to muzzle. The distance between the muzzle and eyes is comparatively short. A profile may be straight or slightly dished below the eyes, blending into large nostrils on a small, refined muzzle”. Figure 5.1 shows examples of the two types as described.

A. **B.**

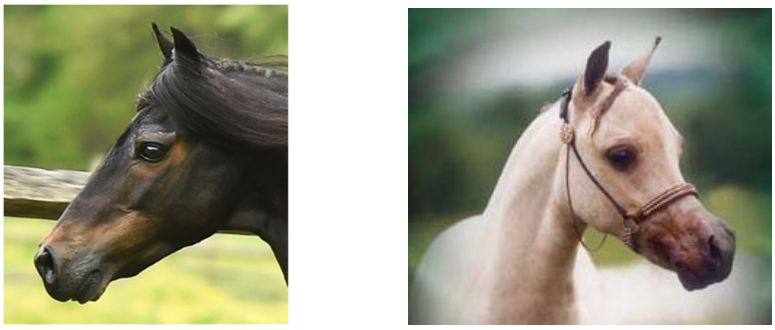


Figure 5.1 Head types of AMHA horses. **A.** Straight profile Shown in **B.** The desired dished profile on the right.

Physical features of a severely shortened nasal bridge and enlarged craniums with large wide-set eyes are common features seen in dwarf humans and in dwarf horses. In normal Miniatures a substantially dished head with a short maxilla, small muzzle and large eyes set wide apart is highly desired and some popular bloodlines are known for these characteristics.

Morphometric measurements to determine disease

Morphometric methods are utilized to describe and compare shapes of organisms or of structural parts of organisms. Conventional use of morphometric methods has been used in forensic sciences and paleobiology to look for shape or structural distinctions of different organisms, fossils, etc. (Rohlf *et al*. 1993). Morphometric measurement methods are used extensively in characterizing physical features of human heads and faces with regards to disease (Krakow *et al*. 2009; Kolar *et al*. 1997; Farkas 1994). Methodologies for craniofacial anthropometry has evolved since the 20th century by establishment of specific techniques and norms with manual measurement tools. Today, computerized video imagery allows much more accurate measurement for morphometric evaluation (Krakow *et al*. 2009; Deutsch *et al*. 2012; Drake *et al*. 2011). Landmark-based methods can be utilized for both two-dimensional (picture) and three-dimensional specimens. Figure 5.2 shows the use of landmark based method to measure the human face in a two-dimensional scenario.

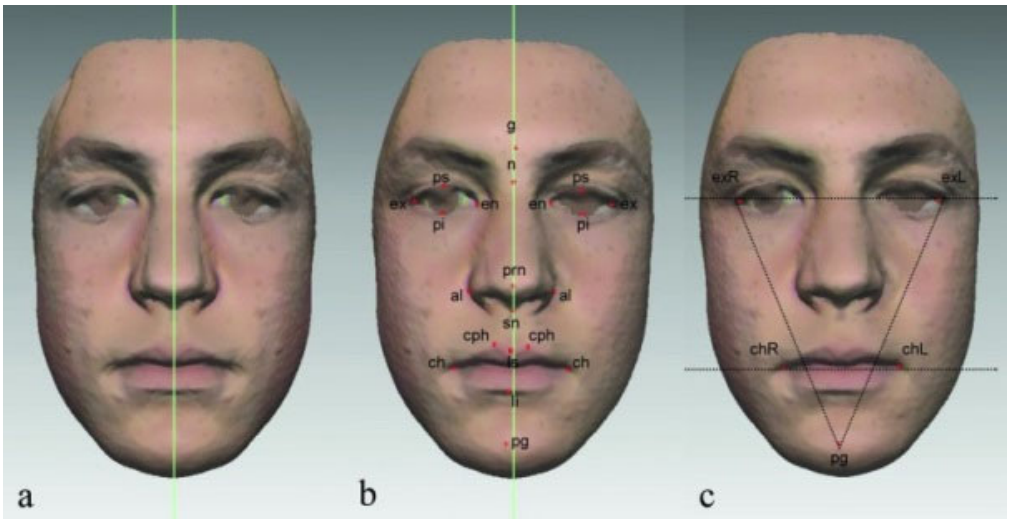


Figure 5.2. Landmark reference points of measurement used in human morphometric measurements.

Features in human heads due to disease have specific measurable points of reference and disproportionate values compared to non-disease proportion measurements. Many landmark features of the face are continuous, quantitative traits. Measurement values are considered abnormal when they are above or below 2 S.D. from the mean. A continuous trait, like ear size, is quite different from a discontinuous trait such as a chin

dimple since it is considered either present or absent. Digital Calliper measurements and ratio measurements are used on heads to determine bone growth abnormalities. Measurements are used in humans to determine the existence of dwarfism and disproportionate physical characteristics with neonatal, infant children and adults (Krakow *et al*. 2009). These comparative measurements were performed on Miniature horses as a measurement protocol.

The purpose of this study was to investigate the effect Miniature horse dwarf variants have on specific structural features of the head using landmark reference points measured by digital calipers.

Objectives of this study

1. Develop a protocol to measure and record distances of specifically selected facial features of Miniature horses by means of the landmark measurement technique used in human facial evaluations.
2. Investigate the genotype of the individuals measured for the dwarf variants *D1*, *D2*, *D3*, *D4*, Brazil dwarf haplotype and putative *D5*.
3. Investigate each category of measurement for statistical significance in relation to carrier status.

Materials and Methods

Horses

A total of 106 mature Miniature horses were selected possessing a non-dwarf phenotype and aged two years and older. Mature horses were used because horses complete over 95% of their growth in the first two years, thus the head size and shape has completed its growth. Horses were located on various breeding farms in Kentucky, Indiana, and Florida. Hair samples were collected from individuals for DNA isolation and tested for *D1, D2, D3, D4,* Brazil haplotype and putative *D5* dwarf variants. Fifteen

Miniatures were genotyped as carriers of a dwarf variant and 91 genotyped as non-carriers. IACUC 2022-4161 was approved in connection with this study of experimental animals.

Morphometric analysis of facial characteristics

Digital calliper measurements were utilized on horses for the measurement protocol. Seven measurements were performed as shown in Figure 5.3. Measurements obtained were, (E) ear length, (Y) eye width, (F) frontal bone length, (M) maxilla length, (W) width between eye medial commissures, (R & L) length between medial commissures and tip of muzzle. Ear measurements were taken from the intertragic notch of the ear to the ear tip. Eye measurements were taken from the lateral commissure to the medial commissure. Frontal bone length was taken from the pole to the cross-section of W between the two medial commissures of the eyes. Maxilla length was taken from the cross-section of W to the tip of the muzzle. Width of cranium (W) was measured as the distance between the two medial commissures. The distance between the left and Right commissures to the muzzle (R & L) were measured from each commissure to the center of the muzzle. Each measurement was duplicated, all measurements are the average of two separate measurements.

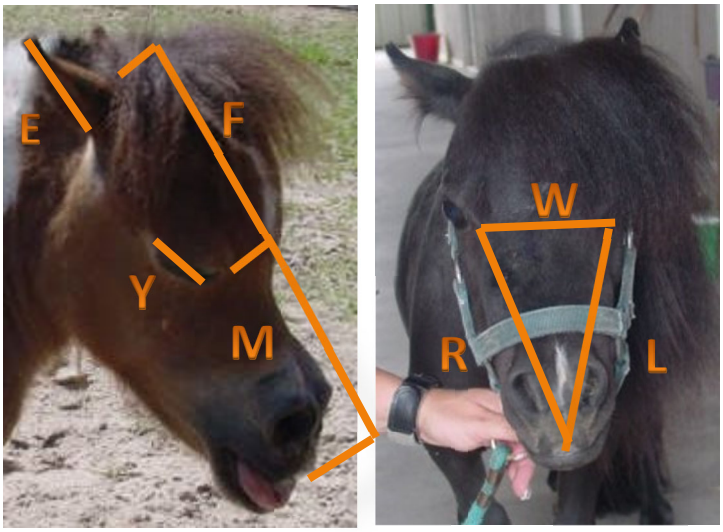


Figure 5.3. Measurement locations. E =Ear length; Y= Eye width; F = Frontal bone length (from pole to W); M= Maxilla length; W= width between eye medial commissure; R= length between right medial commissure and tip of muzzle; L= length between left medial commissure and tip of muzzle.

DNA isolation

Hair samples submitted were processed using 7-10 hair bulbs according to the method described by Locke *et al*. (2002). The hair bulbs were placed in 100 µl lysis solution containing 1X FastStart Taq Polymerase PCR buffer (Roche), 2.5 mM MgCl (Roche), 0.5% Tween 20 (JT Baker, Phillipsburg, NJ) and 0.01 mg proteinase K (ThermoFisher Waltham, MA.) and incubated at 60°C for 25 minutes followed by 94°C for 25 minutes to deactivate the proteinase K. Aliquots of the DNA samples were made for working dilutions at concentrations of approximately 150 ng/µl.

Genotyping for dwarf variants

Custom TaqMan® assays were obtained for dwarf variants *D2*, *D3,* Brazil variant, and *D5*. These were performed on Applied Biosystems Step One Plus qPCR. Custom amplicon size assays with labelled primers were obtained for dwarf variants *D1* and *D4*. PCR reactions were performed on MJ Research PTC-200 thermocycler. Labelled amplicon fragments were analysed on Applied Biosystems 3500 sequencer (Figure 5.4).

A. Exon 7 variant *D2* VIC G FAM A

GTGAAGCTCGAGGCAATGTGATCCTCACAGTGAAACCCATCTTTGGCGTCTCCCCCACC
ATCCTGGAGCCTGGGGAGNCCTTCACGTCTGTCCCTGGC[G/A]TGGGGACCACCGCCT
TCCCNGAGGCGGAGAACGAGACTGGAGCGGCCACCAGGCCCTGGG

B. Exon 8 variant *D3** VIC G FAM C

GTGTTGTGTTCCACTACCGCCCGGGCTCCGACCGATACTCTCTGACCTTTGAGGAGGCA
CAGCAG[G/C]CCTGCCTGCNNACNGNGGCGGTCATNGCCTCGCCNGAGCAGCTCCAGG
CCGCCTACGAAGCGGGCTACGAGCAGTGTGACGCCGGCTGGCTGAGTGACCAGACCGTC
AG

C. g.95271115 A>T Brazil haplotype Exon 12 VIC-A FAM-T

AACTCAGTGGTGCCCATTCTGGAGCACCGGCCGTGTCTGGAGACCATTCGGGATTTTCG
GACTT[A/T]AGCGGGCTGCCGTCAGGGCTGGTGGAGCCCAGTGGAGAGCCATCGAGTA
CTCCAC

D. g.95258999 G>A Exon 13 *D5* VIC G FAM A

ACCAGGAGCTGTGTGAGAATGGCTGGACCAAGTTCCAGGGCCACTGTTACCGCTACTTT
CCCGACCGCGAGACCTGGGTGGAC[G/A]CCGAGAGCAGGTGTCGGGAGCAGCAGTCAC
ACCTGAGCAGCATCGTCAC

E. Exon 15 size variant *D4*

ACAN D4 FORWARD

5′-[6FAM]CTTGTCGTGCAGGTGAAGTG-3′

ACAN D4 REVERSE

5′-GCGGTGAGGCCAGTTCTTTG-3′

F. Exon 2 size variant *D1*

ACAN D1 FORWARD

5′-[6FAM]TCTTGGTAGGCGCTGTTGAC-3′

ACAN D1 REVERSE

5′-AACTCGCTGAGTGTCAGCAT-3′

Figure 5.4. Custom Assays for dwarf variants. **A**. *D2*. **B**. *D3*. **C**. Brazil haplotype**. D**. *D5*. **E**. *D4*. **F**. *D1*.

Statistical Tests: Mann-Whitney U-tests (P value significance set at 0.05, two-tailed) https://www.socscistatistics.com/tests/mannwhitney/default2.aspx were performed on each measurement category of carrier against each measurement category of non-carrier.

Results

Genotyping results for carrier status were combined with digital caliper measurements shown as grouped in Table 5.1 for carriers and Table 5.2 for non-carriers.

Table 5.1. 15 Non-dwarf dwarf carrier Miniature horses with dwarf genotype status and head measurements. Location of measurement identifiers. E =Ear length; Y= Eye width; F = Frontal bone length (from pole to W); M= Maxilla length; W= width of cranium between eye medial commissures; R= length between right medial commissure and tip of muzzle; L= length between left medial commissure and tip of muzzle.

Horse	Head Measurement Locations								Genotype N/D
ID	E	Y	F	M	W	R	L	R + L AVE	D1-D5 or DB (Brazil)
1	92.6	33.4	180.5	230.5	122.2	238.8	238.8	238.8	N/D1
24	96.8	29.7	188.7	254.6	132.4	249.7	249.7	249.7	N/D4
25	89.3	37.4	175.8	214.9	121.3	226.1	226.1	226.1	N/D2
27	93.9	36.6	171.5	229.8	124.9	226.2	226.2	226.2	N/D2
28	84.1	33.4	162.6	234.3	120.2	237.5	237.5	237.5	N/DB
37	77.8	35.2	164.4	223.8	118.4	224.8	224.8	224.8	N/D2
38	82.4	35	153.2	209.4	119.4	218.9	218.9	218.9	N/D2
50	82.1	37.2	172.3	235.3	122.2	224.5	224.5	224.5	N/DB
52	90.8	38.7	184.1	235.5	125.5	231.1	231.1	231.1	N/DB
53	82.7	36.3	184.6	222.2	124.2	225.2	225.2	225.2	N/D1
58	87.6	34.7	171.2	212.3	124.1	219.2	219.2	219.2	N/D1
73	89.6	35.8	172.3	222	130.2	225	225	225	N/D2
74	90.9	35.5	183.6	203.3	128.8	223.1	223.1	223.1	N/DB
86	90.2	30.7	190.4	245.3	137.5	257.3	257.3	257.3	N/D4
103	88.9	32.4	169.5	189.7	121.9	204	204	204	N/D2

Table 5.2. 91 Non-dwarf non-carrier Miniature horses with dwarf genotype status and head measurements. Location of measurement identifiers. E =Ear length; Y= Eye width; F = Frontal bone length (from pole to W); M= Maxilla length; W= width of cranium between eye medial commissures; R= length between right medial commissure and tip of muzzle; L= length between left medial commissure and tip of muzzle.

Horse	Head Measurement Locations								Genotype N/N
ID	E	Y	F	M	W	R	L	R + L AVE	N-normal
2	97.3	32.4	185.7	221.7	134.9	228.7	228.7	228.7	N
3	98.9	29.9	184.1	217.9	131.8	231	231	231	N
4	91.5	36.5	182.5	208.4	136	234.9	234.9	234.9	N
5	95.7	40.3	197.1	218.2	124	223.1	223.1	223.1	N
6	93.1	35	181.2	235	124.6	235.4	235.4	235.4	N
7	102.8	36.3	182.8	242.2	121.5	244.2	244.2	244.2	N
8	89.2	33.3	187	203.8	120.8	220.4	220.4	220.4	N
9	85.2	36.5	189.1	219.3	126.7	233.2	233.2	233.2	N
10	89.5	35.8	173.1	213.2	124.5	223.1	223.1	223.1	N
11	86.4	34.5	183.2	227.2	132.2	230.9	230.9	230.9	N
12	88.6	35.5	171	230.3	132.4	232.4	232.4	232.4	N
13	103	34.4	185.4	250.3	131.7	253.3	253.3	253.3	N
14	99.4	32.2	177.3	229.2	117.9	235.8	235.8	235.8	N
15	85	29.5	183.8	222.2	121.4	224.3	224.3	224.3	N
16	94	38.9	188.2	230.4	128.1	234.1	234.1	234.1	N
17	81.9	28.2	154.3	221.6	116.3	223.1	223.1	223.1	N
18	87.4	37.8	134.5	235.8	118.1	238.2	238.2	238.2	N
19	88.3	32.8	179	236.1	125.7	237.1	237.1	237.1	N
20	85.9	36.8	183.7	218.4	123.8	224.4	224.4	224.4	N
21	98.6	36.3	189.5	230.3	129.7	233.1	233.1	233.1	N
22	88.8	34.8	166.6	197.8	120.6	218.3	218.3	218.3	N
23	82.5	34.5	144.4	225.3	125.1	237.8	237.8	237.8	N
26	91.4	34.5	189.7	242.8	122.7	248.2	248.2	248.2	N
29	83.8	39.5	185.2	227.5	124.7	227	227	227	N
30	74.7	34	154	186	125	192.5	192.5	192.5	N
31	78.3	31.9	167.6	234.5	127.1	228.1	228.1	228.1	N
32	87.6	36.9	168.7	230	124.7	227.5	227.5	227.5	N
33	80.5	29.1	177.8	214.8	125.8	210.8	210.8	210.8	N
34	78	35.4	164.9	216.8	123.7	220.7	220.7	220.7	N
35	89.2	33.8	180.3	222.6	135.58	226.7	226.7	226.7	N
36	95.7	39	187	224.1	132.5	221.8	221.8	221.8	N
39	72.9	37.9	176.8	214.68	131.66	210.5	210.5	210.5	N
40	89.9	33.9	190.3	212.8	133.1	224.1	224.1	224.1	N
41	87.3	35	171.1	210.5	126.8	214.5	214.5	214.5	N
42	80.8	36.8	185.5	251	127.7	255.4	255.4	255.4	N

Table 5.2. ***(continued)*** 91 Non-dwarf non-carrier Miniature horses with dwarf genotype status and head measurements. Location of measurement identifiers. E =Ear length; Y= Eye width; F = Frontal bone length (from pole to W); M= Maxilla length; W= width of cranium between eye medial commissures; R= length between right medial commissure and tip of muzzle; L= length between left medial commissure and tip of muzzle.

43	85.4	34.6	179.6	244.4	139.3	248.9	248.9	248.9	N
44	85	35.4	162.1	216.8	123	229.9	229.9	229.9	N
45	87.7	33.7	168.5	214.5	120.3	220.7	220.7	220.7	N
46	82.4	28	168.2	219.6	120	209.6	209.6	209.6	N
47	93.2	34.2	174.1	212.4	126.2	225.1	225.1	225.1	N
48	86.1	33.4	171.9	210.5	122.8	214.5	214.5	214.5	N
49	90.8	41.4	174.1	232.9	129.5	235.5	235.5	235.5	N
51	100.9	36.8	178.6	251.3	132.8	245.1	245.1	245.1	N
54	88	38.6	172	241.6	138	239.7	239.7	239.7	N
55	84.9	34.8	172.5	239.6	129.4	243.5	243.5	243.5	N
56	93.2	35.1	198.3	248.5	128.43	246.5	246.5	246.5	N
57	92.7	35	185.4	236.2	128.7	242.4	242.4	242.4	N
59	94.1	32.1	167.3	215.5	116.6	206.5	206.5	206.5	N
60	102.1	39.4	177.8	214.8	113.5	211.2	211.2	211.2	N
61	82.6	33.7	173.5	210.5	127.6	225.4	225.4	225.4	N
62	82.8	32.1	182.8	210.3	117.6	216.5	216.5	216.5	N
63	95.7	34.3	174.5	232.2	137.9	234.8	234.8	234.8	N
64	89.1	36.1	182.3	223.2	124.2	228.1	228.1	228.1	N
65	82.3	34.5	160.5	227.5	122.5	223.6	223.6	223.6	N
66	94.5	37.1	186.3	238.6	132.1	237.4	237.4	237.4	N
67	93.7	33.5	181.2	238.4	129	236.3	236.3	236.3	N
68	95.8	35	204.7	242.6	135.5	251.6	251.6	251.6	N
69	75.9	38.6	174.5	231.3	135.1	133.7	233.7	183.7	N
70	81.5	36.5	178.2	227.5	119.2	139.5	239.5	189.5	N
71	98.3	39	183	244	122.5	252	252	252	N
72	90.2	35.4	187	219	123.8	234	234	234	N
75	100.5	35.8	189	227	139.5	246	246	246	N
76	98.5	35.7	175.8	215	123.3	227.2	227.2	227.2	N
77	80.7	35.9	164.4	213.1	120.7	225.6	225.6	225.6	N
78	97.2	39.6	186.7	225.6	128.7	251	251	251	N
79	85.8	28.9	161.7	193.9	107.5	199.5	199.5	199.5	N
80	94.4	36.7	198.3	240.2	129.5	250.6	250.6	250.6	N
81	100.2	32.6	178.7	214.3	129.4	222.1	222.1	222.1	N
82	96.1	35.8	184	234	121	239	239	239	N
83	85.5	34.5	176.8	216.4	118.9	228.4	228.4	228.4	N
84	79.2	31.2	162.5	205	107.7	213	213	213	N
85	92	33.6	178	224.4	123.4	233.2	233.2	233.2	N

Table 5.2. ***(continued)*** 91 Non-dwarf non-carrier Miniature horses with dwarf genotype status and head measurements. Location of measurement identifiers. E =Ear length; Y= Eye width; F = Frontal bone length (from pole to W); M= Maxilla length; W= width of cranium between eye medial commissures; R= length between right medial commissure and tip of muzzle; L= length between left medial commissure and tip of muzzle.

87	85.5	34.8	171.6	199.8	120	208	208	208	N
88	89.6	34.6	181.4	222.7	117.2	229.6	229.6	229.6	N
89	94	38.7	163.5	223.3	128.8	219.7	219.7	219.7	N
90	88.6	37.3	172	215	122.9	222	222	222	N
91	93.6	25.7	194.3	225.1	121.3	239.1	239.1	239.1	N
92	87.2	34.9	165.1	221.3	122.4	232.6	232.6	232.6	N
93	88.5	33.4	162.5	195.3	109.1	206.3	206.3	206.3	N
94	84.6	37.3	178.4	220.9	131.6	230.1	230.1	230.1	N
95	92.9	26.2	191	235	121.3	243	243	243	N
96	100.9	32.1	181.8	218.6	112.2	243.6	243.6	243.6	N
97	95.2	32.7	184.7	223	120.3	241	241	241	N
98	98	33	168	223	125.3	230	230	230	N
99	82.9	32.3	163.7	192.1	111.6	210.9	210.9	210.9	N
100	90.1	34.3	180.5	229	135.4	243	243	243	N
101	86.8	35.6	177.8	208	126.5	229	229	229	N
102	83.5	34.9	184.1	213.1	132.7	231	231	231	N
104	95.7	36	184.5	244.9	132.9	253.9	253.9	253.9	N
105	91.1	30.9	173.7	223.9	128.3	236	236	236	N
106	88.3	31.9	159.5	199.1	113.1	217	217	217	N

Mann-Whitney U-tests were performed on a free use website https://www.socscistatistics.com/tests/mannwhitney/default2.aspx for statistical calculations (two-tailed hypothesis). U-tests were completed for each measurement category of carrier against each measurement category of non-carrier. Other U-tests were completed using ratios of specific measurement categories. Ratios were calculated for the F-frontal and M-maxilla measurements for both carriers and non-carriers. Those ratio calculations were then subjected to U-test. Other ratio calculations performed were for W-cranium to M-maxilla, W-cranium to F-Frontal, W-cranium to R + L average length. The average length for each horse as a single measurement. This was used due to slight measurement variations taken at time of collection. Sum calculations were performed for the sum of W-cranium + F-frontal, and the sum of F-frontal and M-maxilla. The sums for each group, carriers, and non-carriers, were then subjected to U-tests. Table 5.3 gives a

summary of the U-test score, Z-score, p-value, and significance for each test run on each category of measurement.

Table 5.3. Summary of Mann-Whitney U-test results for each measurement category for carriers versus non-carriers. Calculations were run with P-value set at $p<0.05$ for significance and two-tailed hypothesis.

Category	U score	Z-score	p-value	significance
E-ear length carrier vs non-carrier	589.5	0.83844	0.4009	not significant
Y-eye width carriers vs non-carrier	649	0.29912	0.76418	not significant
F-frontal carriers vs non-carrier	599.5	0.7478	0.45326	not significant
M-maxilla carrier vs non-carrier	650.5	-0.28552	0.77182	not significant
W-cranium carrier vs non-carrier	617.5	0.58464	0.56192	not significant
R+L average length carrier vs non-carrier	527.5	0.494	0.62414	not significant
ratio of cranium/frontal carrier vs non-carrier	609	-0.66169	0.50926	not significant
ratio of cranium/maxilla carrier vs non-carrier	616	0.59824	0.5485	not significant
ratio cranium/ave length carrier vs non-carrier	677	0.04532	0.96012	not significant
Ratio of frontal/maxilla carrier vs non-carrier	530	1.37777	0.16758	not significant
sum of frontal and maxilla carrier vs non-carrier	661.5	0.18582	0.8493	not significant
sum cranium and frontal carrier vs non-carrier	585.5	0.8747	0.3843	not significant

Values of the measurements obtained for carriers (Table 5.4) and non-carriers (Table 5.5) for specific characteristics, as well as calculations for ratios, and sums are indicated.

Table 5.4. Values of the measurements and ratios for carriers. Columns represent: 1: ear length, 2: eye width, 3: frontal, 4: maxilla, 5: cranium, 6: ratio frontal/cranium, 7: ratio cranium/maxilla, 8: ratio frontal/maxilla, 9: sum of frontal and maxilla, 10: sum of cranium and frontal. Horse sample is given as ID number.

ID	1	2	3	4	5	6	7	8	9	10
1	92.6	33.4	180.5	230.5	122.2	0.67700831	0.530152	0.78308	411	302.7
24	96.8	29.7	188.7	254.6	132.4	0.701642819	0.520031	0.741163	443.3	321.1
25	89.3	37.4	175.8	214.9	121.3	0.689988623	0.564449	0.818055	390.7	297.1
27	93.9	36.6	171.5	229.8	124.9	0.728279883	0.543516	0.746301	401.3	296.4
28	84.1	33.4	162.6	234.3	120.2	0.739237392	0.513017	0.693982	396.9	282.8
37	77.8	35.2	164.4	223.8	118.4	0.720194647	0.529044	0.734584	388.2	282.8
38	82.4	35	153.2	209.4	119.4	0.779373368	0.570201	0.731614	362.6	272.6
50	82.1	37.2	172.3	235.3	122.2	0.709228091	0.519337	0.732257	407.6	294.5
52	90.8	38.7	184.1	235.5	125.5	0.681694731	0.532909	0.781741	419.6	309.6
53	82.7	36.3	184.6	222.2	124.2	0.672806067	0.558956	0.830783	406.8	308.8
58	87.6	34.7	171.2	212.3	124.1	0.724883178	0.58455	0.806406	383.5	295.3
73	89.6	35.8	172.3	222	130.2	0.755658735	0.586486	0.776126	394.3	302.5
74	90.9	35.5	183.6	203.3	128.8	0.701525054	0.633546	0.903099	386.9	312.4
86	90.2	30.7	190.4	245.3	137.5	0.722163866	0.560538	0.776192	435.7	327.9
103	88.9	32.4	169.5	189.7	121.9	0.719174041	0.642594	0.893516	359.2	291.4

Table 5.5. Values of the measurements and ratios for non-carriers. Columns represent: 1: ear length, 2: eye width, 3: frontal, 4: maxilla, 5: cranium, 6: ratio cranium/frontal, 7: ratio cranium/maxilla, 8: ratio frontal/maxilla, 9: sum of frontal and maxilla, 10: sum of cranium and frontal. Horse sample is given as ID number.

ID	1	2	3	4	5	6	7	8	9	10
2	97.3	32.4	185.7	221.7	134.9	0.726440495	0.60848	0.837618	407.4	320.6
3	98.9	29.9	184.1	217.9	131.8	0.715915263	0.604865	0.844883	402	315.9
4	91.5	36.5	182.5	208.4	136	0.745205479	0.652591	0.87572	390.9	318.5
5	95.7	40.3	197.1	218.2	124	0.629122273	0.568286	0.9033	415.3	321.1
6	93.1	35	181.2	235	124.6	0.687637969	0.530213	0.771064	416.2	305.8
7	102.8	36.3	182.8	242.2	121.5	0.664660832	0.501652	0.754748	425	304.3
8	89.2	33.3	187	203.8	120.8	0.645989305	0.592738	0.917566	390.8	307.8
9	85.2	36.5	189.1	219.3	126.7	0.670015865	0.577747	0.862289	408.4	315.8
10	89.5	35.8	173.1	213.2	124.5	0.719237435	0.583959	0.811914	386.3	297.6
11	86.4	34.5	183.2	227.2	132.2	0.721615721	0.581866	0.806338	410.4	315.4
12	88.6	35.5	171	230.3	132.4	0.774269006	0.574902	0.74251	401.3	303.4
13	103	34.4	185.4	250.3	131.7	0.710355987	0.526169	0.740711	435.7	317.1
14	99.4	32.2	177.3	229.2	117.9	0.664974619	0.514398	0.77356	406.5	295.2
15	85	29.5	183.8	222.2	121.4	0.660500544	0.546355	0.827183	406	305.2
16	94	38.9	188.2	230.4	128.1	0.680658874	0.55599	0.81684	418.6	316.3
17	81.9	28.2	154.3	221.6	116.3	0.753726507	0.524819	0.6963	375.9	270.6
18	87.4	37.8	134.5	235.8	118.1	0.878066914	0.500848	0.570399	370.3	252.6
19	88.3	32.8	179	236.1	125.7	0.702234637	0.532402	0.758153	415.1	304.7
20	85.9	36.8	183.7	218.4	123.8	0.673924878	0.56685	0.841117	402.1	307.5
21	98.6	36.3	189.5	230.3	129.7	0.684432718	0.563178	0.82284	419.8	319.2
22	88.8	34.8	166.6	197.8	120.6	0.723889556	0.609707	0.842265	364.4	287.2
23	82.5	34.5	144.4	225.3	125.1	0.86634349	0.55526	0.640923	369.7	269.5
26	91.4	34.5	189.7	242.8	122.7	0.646810754	0.505354	0.781301	432.5	312.4
29	83.8	39.5	185.2	227.5	124.7	0.673326134	0.548132	0.814066	412.7	309.9
30	74.7	34	154	186	125	0.811688312	0.672043	0.827957	340	279
31	78.3	31.9	167.6	234.5	127.1	0.758353222	0.542004	0.714712	402.1	294.7
32	87.6	36.9	168.7	230	124.7	0.73918198	0.542174	0.733478	398.7	293.4
33	80.5	29.1	177.8	214.8	125.8	0.707536558	0.585661	0.827747	392.6	303.6
34	78	35.4	164.9	216.8	123.7	0.750151607	0.570572	0.760609	381.7	288.6
35	89.2	33.8	180.3	222.6	135.58	0.751968941	0.609075	0.809973	402.9	315.88
36	95.7	39	187	224.1	132.5	0.70855615	0.591254	0.834449	411.1	319.5
39	72.9	37.9	176.8	214.7	131.66	0.744683258	0.613285	0.823551	391.5	308.46
40	89.9	33.9	190.3	212.8	133.1	0.699421965	0.62547	0.894267	403.1	323.4
41	87.3	35	171.1	210.5	126.8	0.741087084	0.602375	0.812827	381.6	297.9
42	80.8	36.8	185.5	251	127.7	0.688409704	0.508765	0.739044	436.5	313.2
43	85.4	34.6	179.6	244.4	139.3	0.775612472	0.569967	0.734861	424	318.9
44	85	35.4	162.1	216.8	123	0.75879087	0.567343	0.747694	378.9	285.1

Table 5.5. ***(continued)*** Values of the measurements and ratios for non-carriers. Columns represent: 1: ear length, 2: eye width, 3: frontal, 4: maxilla, 5: cranium, 6: ratio cranium/frontal, 7: ratio cranium/maxilla, 8: ratio frontal/maxilla, 9: sum of frontal and maxilla, 10: sum of cranium and frontal. Horse sample is given as ID number.

ID	1	2	3	4	5	6	7	8	9	10
45	87.7	33.7	168.5	214.5	120.3	0.713946588	0.560839	0.785548	383	288.8
46	82.4	28	168.2	219.6	120	0.713436385	0.546448	0.765938	387.8	288.2
47	93.2	34.2	174.1	212.4	126.2	0.724870764	0.594162	0.81968	386.5	300.3
48	86.1	33.4	171.9	210.5	122.8	0.714368819	0.583373	0.816627	382.4	294.7
49	90.8	41.4	174.1	232.9	129.5	0.743825388	0.556033	0.747531	407	303.6
51	100.9	36.8	178.6	251.3	132.8	0.74356103	0.528452	0.710704	429.9	311.4
54	88	38.6	172	241.6	138	0.802325581	0.571192	0.711921	413.6	310
55	84.9	34.8	172.5	239.6	129.4	0.750144928	0.540067	0.71995	412.1	301.9
56	93.2	35.1	198.3	248.5	128.4	0.647655068	0.516821	0.797988	446.8	326.7
57	92.7	35	185.4	236.2	128.7	0.694174757	0.544877	0.784928	421.6	314.1
59	94.1	32.1	167.3	215.5	116.6	0.696951584	0.541067	0.776334	382.8	283.9
60	102.1	39.4	177.8	214.8	113.5	0.638357705	0.528399	0.827747	392.6	291.3
61	82.6	33.7	173.5	210.5	127.6	0.735446686	0.606176	0.824228	384	301.1
62	82.8	32.1	182.8	210.3	117.6	0.643326039	0.559201	0.869234	393.1	300.4
63	95.7	34.3	174.5	232.2	137.9	0.79025788	0.593885	0.751507	406.7	312.4
64	89.1	36.1	182.3	223.2	124.2	0.681294569	0.556452	1.467794	306.5	306.5
65	82.3	34.5	160.5	227.5	122.5	0.763239875	0.538462	0.705495	388	283
66	94.5	37.1	186.3	238.6	132.1	0.70907139	0.553646	0.780805	424.9	318.4
67	93.7	33.5	181.2	238.4	129	0.71192053	0.541107	0.760067	419.6	310.2
68	95.8	35	204.7	242.6	135.5	0.661944309	0.558533	0.843776	447.3	340.2
69	75.9	38.6	174.5	231.3	135.1	0.774212034	0.58409	0.754431	405.8	309.6
70	81.5	36.5	178.2	227.5	119.2	0.668911336	0.523956	0.783297	405.7	297.4
71	98.3	39	183	244	122.5	0.669398907	0.502049	0.75	427	305.5
72	90.2	35.4	187	219	123.8	0.662032086	0.565297	0.853881	406	310.8
75	100.5	35.8	189	227	139.5	0.738095238	0.614537	0.832599	416	328.5
76	98.5	35.7	175.8	215	123.3	0.701365188	0.573488	0.817674	390.8	299.1
77	80.7	35.9	164.4	213.1	120.7	0.734184915	0.566401	0.771469	377.5	285.1
78	97.2	39.6	186.7	225.6	128.7	0.689341189	0.570479	0.827571	412.3	315.4
79	85.8	28.9	161.7	193.9	107.5	0.664811379	0.554409	0.833935	355.6	269.2
80	94.4	36.7	198.3	240.2	129.5	0.653050933	0.539134	0.825562	438.5	327.8
81	100.2	32.6	178.7	214.3	129.4	0.724118635	0.603826	0.833878	393	308.1
82	96.1	35.8	184	234	121	0.657608696	0.517094	0.786325	418	305
83	85.5	34.5	176.8	216.4	118.9	0.672511312	0.549445	0.817006	393.2	295.7
84	79.2	31.2	162.5	205	107.7	0.662769231	0.525366	0.792683	367.5	270.2
85	92	33.6	178	224.4	123.4	0.693258427	0.549911	0.793226	402.4	301.4
87	85.5	34.8	171.6	199.8	120	0.699300699	0.600601	0.858859	371.4	291.6
88	89.6	34.6	181.4	222.7	117.2	0.646085998	0.526269	0.814549	404.1	298.6

Table 5.5. ***(continued)*** Values of the measurements and ratios for non-carriers. Columns represent: 1: ear length, 2: eye width, 3: frontal, 4: maxilla, 5: cranium, 6: ratio cranium/frontal, 7: ratio cranium/maxilla, 8: ratio frontal/maxilla, 9: sum of frontal and maxilla, 10: sum of cranium and frontal. Horse sample is given as ID number.

ID	1	2	3	4	5	6	7	8	9	10
89	94	38.7	163.5	223.3	128.8	0.787767584	0.576803	0.732199	386.8	292.3
90	88.6	37.3	172	215	122.9	0.714534884	0.571628	0.8	387	294.9
91	93.6	25.7	194.3	225.1	121.3	0.624292331	0.538872	0.863172	419.4	315.6
92	87.2	34.9	165.1	221.3	122.4	0.741368867	0.553095	0.746046	386.4	287.5
93	88.5	33.4	162.5	195.3	109.1	0.671384615	0.558628	0.832053	357.8	271.6
94	84.6	37.3	178.4	220.9	131.6	0.737668161	0.595745	0.807605	399.3	310
95	92.9	26.2	191	235	121.3	0.635078534	0.51617	0.812766	426	312.3
96	100.9	32.1	181.8	218.6	112.2	0.617161716	0.513266	0.831656	400.4	294
97	95.2	32.7	184.7	223	120.3	0.651326475	0.539462	0.828251	407.7	305
98	98	33	168	223	125.3	0.745833333	0.561883	0.753363	391	293.3
99	82.9	32.3	163.7	192.1	111.6	0.681734881	0.580947	0.85216	355.8	275.3
100	90.1	34.3	180.5	229	135.4	0.750138504	0.591266	0.78821	409.5	315.9
101	86.8	35.6	177.8	208	126.5	0.711473566	0.608173	0.854808	385.8	304.3
102	83.5	34.9	184.1	213.1	132.7	0.720803911	0.622712	0.863914	397.2	316.8
104	95.7	36	184.5	244.9	132.9	0.720325203	0.54267	0.753369	429.4	317.4
105	91.1	30.9	173.7	223.9	128.3	0.738629822	0.573024	0.775793	397.6	302
106	88.3	31.9	159.5	199.1	113.1	0.709090909	0.568056	0.801105	358.6	272.6

Sorted values by rank are shown in Table 5.6. This provides a visual evaluation to show carrier measurements obtained are dispersed throughout all characteristic category measurement ranges among the non-carrier samples. As can be seen, no measurement range or calculation can distinguish carriers from non-carriers for the facial characteristics examined.

Table 5.6. Sorted values of the measurements and ratios for carriers and non-carriers. Carrier values are highlighted in yellow. Columns represent: 1: ear length, 2: eye width, 3: frontal, 4: maxilla, 5: cranium, 6: ratio frontal/cranium, 7: ratio cranium/maxilla, 8: ratio frontal/maxilla, 9: sum of frontal and maxilla, 10: sum of cranium and frontal.

1	2	3	4	5	6	7	8	9	10
72.9	25.7	134.5	186	107.5	0.617161716	0.500848	0.570399	306.5	252.6
74.7	26.2	144.4	189.7	107.7	0.624292331	0.501652	0.640923	340	269.2
75.9	28	153.2	192.1	109.1	0.629122273	0.502049	0.693982	355.6	269.5
77.8	28.2	154	193.9	111.6	0.635078534	0.505354	0.6963	355.8	270.2
78	28.9	154.3	195.3	112.2	0.638357705	0.508765	0.705495	357.8	270.6
78.3	29.1	159.5	197.8	113.1	0.643326039	0.513017	0.710704	358.6	271.6
79.2	29.5	160.5	199.1	113.5	0.645989305	0.513266	0.711921	359.2	272.6
80.5	29.7	161.7	199.8	116.3	0.646085998	0.514398	0.714712	362.6	272.6
80.7	29.9	162.1	203.3	116.6	0.646810754	0.51617	0.71995	364.4	275.3
80.8	30.7	162.5	203.8	117.2	0.647655068	0.516821	0.731614	367.5	279
81.5	30.9	162.5	205	117.6	0.651326475	0.517094	0.732199	369.7	282.8
81.9	31.2	162.6	208	117.9	0.653050933	0.519337	0.732257	370.3	282.8
82.1	31.9	163.5	208.4	118.1	0.657608696	0.520031	0.733478	371.4	283
82.3	31.9	163.7	209.4	118.4	0.660500544	0.523956	0.734584	375.9	283.9
82.4	32.1	164.4	210.3	118.9	0.661944309	0.524819	0.734861	377.5	285.1
82.4	32.1	164.4	210.5	119.2	0.662032086	0.525366	0.739044	378.9	285.1
82.5	32.1	164.9	210.5	119.4	0.662769231	0.526169	0.740711	381.6	287.2
82.6	32.2	165.1	210.5	120	0.664660832	0.526269	0.741163	381.7	287.5
82.7	32.3	166.6	212.3	120	0.664811379	0.528399	0.74251	382.4	288.2
82.8	32.4	167.3	212.4	120.2	0.664974619	0.528452	0.746046	382.8	288.6
82.9	32.4	167.6	212.8	120.3	0.668911336	0.529044	0.746301	383	288.8
83.5	32.6	168	213.1	120.3	0.669398907	0.530152	0.747531	383.5	291.3
83.8	32.7	168.2	213.1	120.6	0.670015865	0.530213	0.747694	384	291.4
84.1	32.8	168.5	213.2	120.7	0.671384615	0.532402	0.75	385.8	291.6
84.6	33	168.7	214.3	120.8	0.672511312	0.532909	0.751507	386.3	292.3
84.9	33.3	169.5	214.5	121	0.672806067	0.538462	0.753363	386.4	293.3
85	33.4	171	214.7	121.3	0.673326134	0.538872	0.753369	386.5	293.4
85	33.4	171.1	214.8	121.3	0.673924878	0.539134	0.754431	386.8	294
85.2	33.4	171.2	214.8	121.3	0.67700831	0.539462	0.754748	386.9	294.5
85.4	33.4	171.5	214.9	121.4	0.680658874	0.540067	0.758153	387	294.7
85.5	33.5	171.6	215	121.5	0.681294569	0.541067	0.760067	387.8	294.7
85.5	33.6	171.9	215	121.9	0.681694731	0.541107	0.760609	388	294.9
85.8	33.7	172	215.5	122.2	0.681734881	0.542004	0.765938	388.2	295.2

Table 5.6. *(continued)* Sorted values of the measurements and ratios for carriers and non-carriers. Carrier values are highlighted in yellow. Columns represent: 1: ear length, 2: eye width, 3: frontal, 4: maxilla, 5: cranium, 6: ratio frontal/cranium, 7: ratio cranium/maxilla, 8: ratio frontal/maxilla, 9: sum of frontal and maxilla, 10: sum of cranium and frontal.

1	2	3	4	5	6	7	8	9	10
85.9	33.7	172	216.4	122.2	0.684432718	0.542174	0.771064	390.7	295.3
86.1	33.8	172.3	216.8	122.4	0.687637969	0.54267	0.771469	390.8	295.7
86.4	33.9	172.3	216.8	122.5	0.688409704	0.543516	0.77356	390.8	296.4
86.8	34	172.5	217.9	122.5	0.689341189	0.544877	0.775793	390.9	297.1
87.2	34.2	173.1	218.2	122.7	0.689988623	0.546355	0.776126	391	297.4
87.3	34.3	173.5	218.4	122.8	0.693258427	0.546448	0.776192	391.5	297.6
87.4	34.3	173.7	218.6	122.9	0.694174757	0.548132	0.776334	392.6	297.9
87.6	34.4	174.1	219	123	0.696951584	0.549445	0.780805	392.6	298.6
87.6	34.5	174.1	219.3	123.3	0.699300699	0.549911	0.781301	393	299.1
87.7	34.5	174.5	219.6	123.4	0.699421965	0.553095	0.781741	393.1	300.3
88	34.5	174.5	220.9	123.7	0.701365188	0.553646	0.78308	393.2	300.4
88.3	34.5	175.8	221.3	123.8	0.701525054	0.554409	0.783297	394.3	301.1
88.3	34.5	175.8	221.6	123.8	0.701642819	0.55526	0.784928	396.9	301.4
88.5	34.6	176.8	221.7	124	0.702234637	0.55599	0.785548	397.2	301.9
88.6	34.6	176.8	222	124.1	0.707536558	0.556033	0.786325	397.6	302
88.6	34.7	177.3	222.2	124.2	0.70855615	0.556452	0.78821	398.7	302.5
88.8	34.8	177.8	222.2	124.2	0.70907139	0.558533	0.792683	399.3	302.7
88.9	34.8	177.8	222.6	124.5	0.709090909	0.558628	0.793226	400.4	303.4
89.1	34.8	177.8	222.7	124.6	0.709228091	0.558956	0.797988	401.3	303.6
89.2	34.9	178	223	124.7	0.710355987	0.559201	0.8	401.3	303.6
89.2	34.9	178.2	223	124.7	0.711473566	0.560538	0.801105	402	304.3
89.3	35	178.4	223.2	124.9	0.71192053	0.560839	0.806338	402.1	304.3
89.5	35	178.6	223.3	125	0.713436385	0.561883	0.806406	402.1	304.7
89.6	35	178.7	223.8	125.1	0.713946588	0.563178	0.807605	402.4	305
89.6	35	179	223.9	125.3	0.714368819	0.564449	0.809973	402.9	305
89.9	35	179.6	224.1	125.5	0.714534884	0.565297	0.811914	403.1	305.2
90.1	35.1	180.3	224.4	125.7	0.715915263	0.566401	0.812766	404.1	305.5
90.2	35.2	180.5	225.1	125.8	0.719174041	0.56685	0.812827	405.7	305.8
90.2	35.4	180.5	225.3	126.2	0.719237435	0.567343	0.814066	405.8	306.5
90.8	35.4	181.2	225.6	126.5	0.720194647	0.568056	0.814549	406	307.5
90.8	35.4	181.2	227	126.7	0.720325203	0.568286	0.816627	406	307.8
90.9	35.5	181.4	227.2	126.8	0.720803911	0.569967	0.81684	406.5	308.1
91.1	35.5	181.8	227.5	127.1	0.721615721	0.570201	0.817006	406.7	308.5

Table 5.6. *(continued)* Sorted values of the measurements and ratios for carriers and non-carriers. Carrier values are highlighted in yellow. Columns represent: 1: ear length, 2: eye width, 3: frontal, 4: maxilla, 5: cranium, 6: ratio frontal/cranium, 7: ratio cranium/maxilla, 8: ratio frontal/maxilla, 9: sum of frontal and maxilla, 10: sum of cranium and frontal.

1	2	3	4	5	6	7	8	9	10
91.4	35.6	182.3	227.5	127.6	0.722163866	0.570479	0.817674	406.8	308.8
91.5	35.7	182.5	227.5	127.7	0.723889556	0.570572	0.818055	407	309.6
92	35.8	182.8	229	128.1	0.724118635	0.571192	0.81968	407.4	309.6
92.6	35.8	182.8	229.2	128.3	0.724870764	0.571628	0.82284	407.6	309.9
92.7	35.8	183	229.8	128.4	0.724883178	0.573024	0.823551	407.7	310
92.9	35.8	183.2	230	128.7	0.726440495	0.573488	0.824228	408.4	310
93.1	35.9	183.6	230.3	128.7	0.728279883	0.574902	0.825562	409.5	310.2
93.2	36	183.7	230.3	128.8	0.734184915	0.576803	0.827183	410.4	310.8
93.2	36.1	183.8	230.4	128.8	0.735446686	0.577747	0.827571	411	311.4
93.6	36.3	184	230.5	129	0.737668161	0.580947	0.827747	411.1	312.3
93.7	36.3	184.1	231.3	129.4	0.738095238	0.581866	0.827747	412.1	312.4
93.9	36.3	184.1	232.2	129.4	0.738629822	0.583373	0.827957	412.3	312.4
94	36.5	184.1	232.9	129.5	0.73918198	0.583959	0.828251	412.7	312.4
94	36.5	184.5	234	129.5	0.739237392	0.58409	0.830783	413.6	313.2
94.1	36.5	184.6	234.3	129.7	0.741087084	0.58455	0.831656	415.1	314.1
94.4	36.6	184.7	234.5	130.2	0.741368867	0.585661	0.832053	415.3	315.4
94.5	36.7	185.2	235	131.6	0.74356103	0.586486	0.832599	416	315.4
95.2	36.8	185.4	235	131.66	0.743825388	0.591254	0.833878	416.2	315.6
95.7	36.8	185.4	235.3	131.7	0.744683258	0.591266	0.833935	418	315.8
95.7	36.8	185.5	235.5	131.8	0.745205479	0.592738	0.834449	418.6	315.88
95.7	36.9	185.7	235.8	132.1	0.745833333	0.593885	0.837618	419.4	315.9
95.7	37.1	186.3	236.1	132.2	0.750138504	0.594162	0.841117	419.6	315.9
95.8	37.2	186.7	236.2	132.4	0.750144928	0.595745	0.842265	419.6	316.3
96.1	37.3	187	238.4	132.4	0.750151607	0.600601	0.843776	419.8	316.8
96.8	37.3	187	238.6	132.5	0.751968941	0.602375	0.844883	421.6	317.1
97.2	37.4	187	239.6	132.7	0.753726507	0.603826	0.85216	424	317.4
97.3	37.8	188.2	240.2	132.8	0.755658735	0.604865	0.853881	424.9	318.4
98	37.9	188.7	241.6	132.9	0.758353222	0.606176	0.854808	425	318.5
98.3	38.6	189	242.2	133.1	0.75879087	0.608173	0.858859	426	318.9
98.5	38.6	189.1	242.6	134.9	0.763239875	0.60848	0.862289	427	319.2
98.6	38.7	189.5	242.8	135.1	0.774212034	0.609075	0.863172	429.4	319.5
98.9	38.7	189.7	244	135.4	0.774269006	0.609707	0.863914	429.9	320.6
99.4	38.9	190.3	244.4	135.5	0.775612472	0.613285	0.869234	432.5	321.1

Table 5.6. *(continued)* Sorted values of the measurements and ratios for carriers and non-carriers. Carrier values are highlighted in yellow. Columns represent: 1: ear length, 2: eye width, 3: frontal, 4: maxilla, 5: cranium, 6: ratio frontal/cranium, 7: ratio cranium/maxilla, 8: ratio frontal/maxilla, 9: sum of frontal and maxilla, 10: sum of cranium and frontal.

1	2	3	4	5	6	7	8	9	10
100.2	39	190.4	244.9	135.6	0.779373368	0.614537	0.87572	435.7	321.1
100.5	39	191	245.3	136	0.787767584	0.622712	0.893516	435.7	323.4
100.9	39.4	194.3	248.5	137.5	0.79025788	0.62547	0.894267	436.5	326.7
100.9	39.5	197.1	250.3	137.9	0.802325581	0.633546	0.903099	438.5	327.8
102.1	39.6	198.3	251	138	0.811688312	0.642594	0.9033	443.3	327.9
102.8	40.3	198.3	251.3	139.3	0.86634349	0.652591	0.917566	446.8	328.5
103	41.4	204.7	254.6	139.5	0.878066914	0.672043	1.467794	447.3	340.2

Discussion

The use of morphometric measurement techniques to establish measurement criteria which can distinguish a diseased state in comparison to a non-diseased state is a well-established protocol in humans, especially in early development of children (Krakow *et al.* 2009; Kolar *et al*. 1997; Farkas 1994). When an inherited diseased trait presents in homozygous form in offspring, the initial consideration is for the mutation to be recessive in nature. However, in some circumstances the heterozygous carriers present a phenotype which may be termed a mild form or have some characteristic affected by the recessive trait. In cases of partial dominance of a recessive diseased trait, morphometric measurement techniques of those individuals can show differences which may be statistically significant to ones that do not possess the diseased trait. Recessive disease mutations in *ACAN* have been demonstrated to have a measurable effect on height and other phenotypic characteristics in horses and humans (Graves *et al.* 2020; Sentchordi-Montané *et al*. 2018). This study investigated specific facial measurements of Miniature horses and compared those to their carrier status for dwarfism. With the measurements performed here, we did not identify a statistically significant difference in any measurement between carriers of dwarfism and non-carriers for the characteristics chosen.

There were several limitations to this study. The number of carriers was small, only 15. However, no trend was observed with this small number of samples that suggests a larger study might become significant. The small numbers also prevented us from comparing the different types of carriers. It remains possible that one type of carrier, for example genotype *D2/N*, might be affected while genotype *D3/N* would be unaffected. Ideally, a more in-depth analysis of each dwarf variant with 10-20 carrier individuals would give a clearer picture of the effects each mutation may have on other features of Miniature horses besides a slight decrease in height. However, the results from this study did not provide preliminary evidence supporting continuation of the premise outlined in this study.

Chapter VI

Catastrophic Injury in Thoroughbred Training or Racing and Copy Number Variations in the *ACAN* VNTR Region

Summary

Aggrecan is a large aggregating proteoglycan vital to the formation of articular cartilage. Variation in people for the length of the region of *aggrecan,* called the CS1 domain, may be an important factor in early joint degeneration and early onset osteoarthritis. This region is characterized by a series of repeat units that determine the number of chondroitin sulfate chains in the region.

Therefore, we investigated the variation of *aggrecan* in Thoroughbred horses that suffered catastrophic breakdown, with reference to the region specific to binding chondroitin sulfate chains. A randomly selected group of 45 Thoroughbred horses and 17 horses that succumbed to injury on the racetrack were compared specifically for the sequence length of this region in *aggrecan*. Both groups presented significant variation in sequence length for this region of *aggrecan*. A total of 6 alleles were observed in both groups representing between 36 and 44 repeat units. When the distribution of alleles between the two groups was compared, no significant differences were observed (P=0.484). This evidence suggests genetic variation for *aggrecan* and its ability to bind CS chains is not a major factor causing racetrack injury of Thoroughbred horses.

Introduction

Horse racing produces significant stress on the musculoskeletal system of horses which can result in injury or death (Georgopoulos *et al*. 2016; Hitchens *et al*. 2019; Gramm *et al*. 2010; Maeda *et al*. 2016; Crawford *et al*. 2021). Cartilage is a key component of the musculoskeletal system and *aggrecan* (*ACAN*) is a vital component of cartilage. Studies in humans showed *aggrecan* turnover, lifespan and function are impaired in patients with osteoarthritis and rheumatoid arthritis (Dean *et al*. 1989; Wolfe *et al*. 1993). Cartilage failure presented as Lumbar disc degeneration among people may be caused by a lower

VNTR polymorphism in CS1 (Mashayekhi *et al*. 2010; Kawaguchi *et al*. 1999; Cong *et al*. 2010; Solovieva *et al*. 2007). In people the CS1 domain consists of a series of 13-34 tandem repeats of 57 base-pairs each encoding 19 amino acids (Doege *et al.* 1991; Doege *et al*. 1997). The most frequent number of repeats is 26. The lower numbers of repeats were associated with pathology (Battié *et al*. 2006; Roughley *et al*. 2006; Mayer *et al*. 2013). Because of the observations of early cartilage failure in people associated with *ACAN* length variants of this region, we considered it appropriate to investigate the extent of variation and possible pathology in racehorses in athletic events.

The purpose of this study was to investigate variation in the *aggrecan* CS1 domain and its potential to cause injury to racehorses.

Objectives of this study

1. Develop an assay for detection of genetic variation of the *ACAN* VNTR region in horses.
2. Investigate the extent of variation of this segment in Thoroughbred horses.
3. Investigate the distribution of variants among horses which suffered catastrophic outcomes while racing or training for races.

Materials and Methods

Horses

This study involved use of DNA from 62 Thoroughbred horses, including 45 controls and 17 horses (cases) that were identified as having catastrophic breakdowns while racing. The 45 controls were selected from an archive of 200 Thoroughbred samples collected from Kentucky farms in the 1980s and chosen based on the high quality and length of DNA fragments, as described below.

The DNA samples from 17 cases were selected from among 89 horses with catastrophic outcomes as described from necropsy reports and provided by the Kentucky

Horse Racing Commission. The KHRC had collected tissue samples from horses which had experienced catastrophic injuries while racing or training for racing in Kentucky which resulted in death, either directly from the incident or because of loss of utility. DNA was extracted from tissues of all 89 horses as described below, however the quality and length of DNA fragments were suitable for only 17.

DNA isolation

This study required long lengths of DNA. For the randomly selected controls, DNA had previously been isolated from peripheral blood cells using Puregene whole blood DNA extraction kits (Gentra Systems Inc., Minneapolis, MN) or phenol/chloroform extraction method. Over 200 DNA samples were tested, and many were found deficient in long length DNA strands. This may be an age effect since we found that DNA samples isolated within the last 5 years were found to be more likely to have long-length DNA strands. For this study we were able to use DNA from 45 control horses.

DNA was isolated from tissues of the 89 horses that had experienced racetrack catastrophic injury resulting in death. The tissues used included spleen, kidney, and muscle. The tissue samples weighing approximately 50 mg were processed by mechanical pulverization/grinding then treated using Puregene tissue DNA extraction kits (Gentra Systems Inc., Minneapolis, MN) according to published protocols. Problems were encountered with this procedure using frozen tissue that must be extensively processed to extract DNA. This is a standard DNA extraction protocol appropriate for most genetic studies and DNA was easily obtained. However, most of the DNA was fragmented into lengths less than 2000 base pairs. In this case, we wished to investigate DNA regions exceeding 2000 base pairs. We developed a modified method to the published Puregene tissue protocol to process the tissues with greater care. Mechanical pulverization was eliminated to process the tissue and instead performed mincing with a scalpel. This was followed by 2 days of incubation in water bath at 55 °C with 15 µl proteinase K instead of the recommended overnight timeframe. The protocol was followed as published for the remainder of DNA isolation. Despite this modified treatment, tissue samples that were older or that experienced freezer burn and tissues other than muscle did not contain DNA

of sufficient quality for use in the assay. Suitable quality DNA was obtained from muscle tissue of 17 horses identified as experiencing catastrophic racetrack injury. These samples produced an assay of 34 alleles.

DNA Quality and PCR amplification of regions exceeding 1000 base pairs

Amplification of a large DNA region, exceeding 2000 base-pairs, utilizes significantly more PCR reagents as well as a unique program of amplification to overcome the complexity of large amplicons being made. Previous studies in our laboratory mainly consisted of producing amplicon DNA fragments less than 1000 base-pairs in length and usually only 200-500 base-pairs long. While DNA can be readily extracted from many biological sources and stored in a variety of forms, it is challenging to extract and store long sections or strands of DNA. Isolated DNA gradually breaks down into lengths of less than 2000 base pair and shorter. Extracted DNA stored in our reference archive over 10 years ago tended to comprise shorter fragments. We experienced a difficulty in achieving and analyzing 2000+ base pair DNA fragments extracted from various sources. DNA extracted from tissue using aggressive maceration procedures proved to have short length DNA fragments and were not useful. The most reliable source of DNA from tissues was recently collected muscle which had not experienced thawing and refreezing.

PCR primers

Several sets of primers were investigated for amplification of the region defined as CS1 which contains the VNTR region of Exon 12. Amplification was difficult due to the existence of repetitive DNA and its effect on PCR. Primers were designed using *Primer3Plus* online tool (https://www.primer3plus.com/index.html) and produced by Integrated DNA Technologies (https://www.idtdna.com/). Primer set selected which produced measurable assay results are below in Figure 6.1. These primers were located outside the repeat region and were predicted to produce a product size of 2520 bases from the predicted sequence for Twight in Ecab 3.0 online reference sequence.

Forward primer: TTAGGTGGTCCACGACTCCT

Reverse primer: GGCAACCCACTAAGGTCCTC

Figure 6.1. PCR primers for ACAN VNTR region

Long Range PCR conditions

PCR Template for amplified in 50 µl PCR reactions using 1X PCR buffer with 2.0 mM $MgCl_2$, 200 µM of each dNTP, 1 µl genomic DNA from hair, blood, or tissue extractions, 0.2 µl FastStart Taq DNA polymerase (Perkin Elmer) and 50 nM of each primer. Each exon template PCR product was quantified on a 1% agarose gel. PCR DNA Template was amplified in 50 µl PCR reactions. PCR reaction cocktail included Taq polymerase Enzyme blend with proof-reading, double concentrated reaction buffer 3mM $MgCl_2$, 0.4 mM of each dNTP, 150 ng genomic DNA from tissue extractions, 2.5 Taq DNA polymerase units / 50 ul reaction and 300 nM of each primer.

PCR l program entailed initial denaturation step of 2 minutes at 94C followed by an initial set of 10 cycles of 10 seconds at 94C, 70 seconds at 58C, 150 seconds at 72C, followed by PCR 2 program as set of 20 cycles of 15 seconds at 94C, 30 seconds at 58C and 150 seconds at 72 C with 5 seconds added each cycle, followed by a final 7 minutes at 72C, then held at 4C when completed.

Electrophoresis

Template PCR products were tested on a 1% agarose gel run at 90 volts for 2.5 hours. This allowed for the large amplicons to have more uniform migration thus producing bands that may be close in size to be delineated for documentation. The fragment size reference ladder was Bioventures Inc. Biomarker large fragment size ladder was 3000 bp, 2500bp, 2000bp, 1500bp, 1000bp.

Statistical Analysis of Alleles

The frequencies of the ACAN alleles were compared for the control population and the injury population using chi-square analysis.

Results

Evaluation of VNTR Amplicons

Variation in fragment length of *ACAN* was readily apparent from gel electrophoresis (Figure 6.2). The precise size of the fragments could not be determined due to the limited resolution of the agarose gels. However, the estimated sizes ranged from ~2300 base pairs to ~2800 base pairs, corresponding to repeat length of 36 to 44. The amplicons were organized into 6 groups and assigned allele designates as E, H, K, N, Q and T corresponding to base pair lengths of approximately 2300, 2400, 2500, 2600 , 2700 and 2800 respectively.

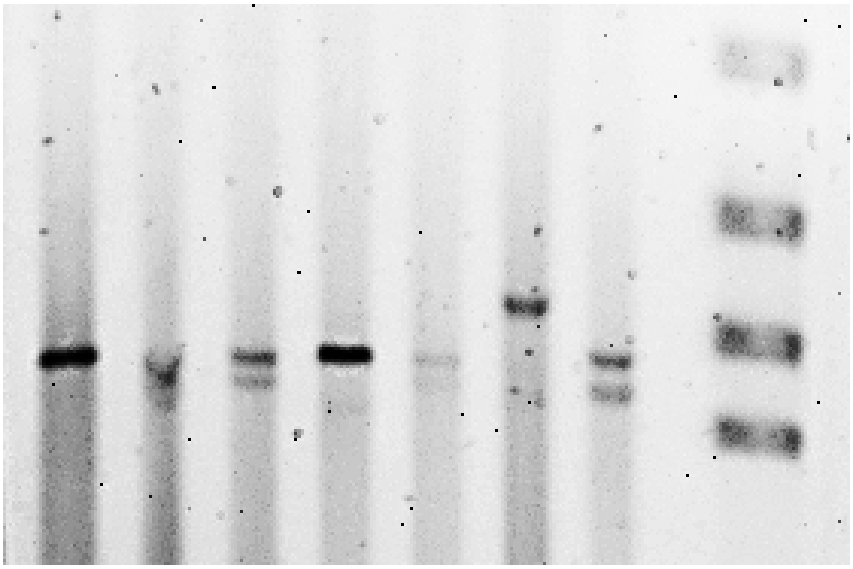


Figure 6.2. Gel image of Exon 12 VNTR region variation among Thoroughbred horses. Ladder biomarker in far right with bands top to bottom, 1500, 2000, 2500, 3000 bp size. Genotypes scored from left to right: KK (2400), KN (2400/2700), KN (2400/2700), KK (2400), KN (2400/2700), EE (2300), KN (2400/2700).

Variation of VNTR region in Thoroughbred horse

Of the 45 horses tested, 6 alleles for the ACAN VNTR were observed to be present. The differences in length among the six alleles spanned an estimated 500 base pairs. This would be equivalent to a difference of 8 repeats between the shortest and longest variant. The distribution of lengths was roughly a normal distribution with most alleles falling

between 2500 and 2700 base pairs (Table 6.1). This may not encompass all possible alleles in Thoroughbreds. Additional alleles may exist. This method, however, does detect the most common alleles and can determine the relative VNTR sizes.

Table 6.1: Alleles of *ACAN* VNTR found among normal Thoroughbred horses.

Allele Designation	Product Length (base pairs)	Number of alleles (N=90)
E	2300	4
H	2400	3
K	2500	18
N	2600	24
Q	2700	30
T	2800	11

Distribution of variants among horses which suffered catastrophic outcomes

The same alleles found among the control horses were observed among the horses which had experienced injuries. One notable exception was the shortest VNTR was not observed in the injured samples. Table 6.2 lists the results for testing 17 horses.

Table 6.2: Alleles of *ACAN* found among Thoroughbred horses experiencing catastrophic racetrack injuries.

Allele Designation	Product Length (base pairs)	Number of alleles (N=34)
E	2300	0
H	2400	2
K	2500	7
N	2600	9
Q	2700	10
T	2800	6

Comparison of ACAN alleles in the control population and the injury population of Thoroughbred horses

The observed versus the expected occurrences of the alleles among the horses which had experienced racetrack injuries were compared and evaluated using the chi-square test. Expected numbers were based on the frequencies observed in the control population (shown in Table 6.3). The P value for this comparison was P=0.484 and was not significant. This indicates that the two populations had the same distribution of alleles. No VNTR sizes were associated with catastrophic injury in Thoroughbreds. Due to the

lack of samples in the E allele group, allele groups E and H were combined to calculate the Chi-square to represent the two groups.

Table 6.3. Chi-Square Test for homogeneity between the control and injury population for the occurrence of ACAN VNTR alleles. For Chi-square 0.77 with 4 degrees of freedom, two tailed, P=0.484.

Allele	Observed Number	Expected Number	Chi-Square
E	0	1.36	N/A
H	2	2.38	0.06
K	7	6.80	0.01
N	9	9.18	0.01
Q	10	11.22	0.13
T	6	4.42	0.56
		Sum of Chi-Square	0.77
		P	0.484

Discussion

The premise for this study was the hypothesis that variation existed for the *ACAN* VNTR array region and that individuals experiencing racetrack injury would, on average, show a skewed distribution of alleles. In humans, individuals with a lower number of VNTR repeats were predisposed to cartilage injuries (Doege *et al*. 1997; Kawaguchi *et al*. 1999; Solovieva *et al*. 2007; Mashayekhi *et al*. 2010; Cong *et al*. 2010; Mayer *et al*. 2013). We saw no evidence for this in Thoroughbred racehorses based on the comparisons described in Table 3. The distribution of *ACAN* VNTR alleles among the controls and the injured horses were almost identical and statistically indistinct. These results showed variation exists among horses for the copy number of the VNTR array region in *ACAN*. However, based on these data, it does not appear that any allele(s) of the VNTR region have a negative functional impact between these populations. Admittedly this is a small study, and a larger study may identify an impact. However, these results suggest that, if an impact exists, it is not likely to have a large impact.

Variation for VNTR region in ACAN in horses

This work proved the existence of variation for the VNTR array region of the *ACAN* gene in horses. From the samples tested, we discovered six alleles based on length of PCR amplicons. More alleles likely may exist that were not present in the two groups examined;

however, the six discovered are the most prevalent in Thoroughbreds. Also, some alleles may exist outside the VNTR size range found and those would be detected if more horses were tested. It is likely alleles not produced in this study are uncommon because a sufficient number of horses to detect common alleles were tested. Variation to this extent is characteristic of highly conserved repetitive sequence elements. This unique VNTR array is in an exon of a gene and is quite large. Its complexity of two unique sequences of 31 nucleotides and 32 nucleotides in tandem repeated anywhere from 36 to 44 times, from samples in this study, makes the protein produced by this gene one of the largest molecules in the body. However, the assay resolution is 500 bp which makes distinguishing a single 63 base tandem repeat size difference challenging. It is possible tests with greater resolution may lead to subdivision of some of the currently identified size alleles within the range we found. For example, the categories designated K, N and E may include multiple alleles which cannot be distinguished on these gels.

Repetitive sequence length variations arise due to errors in alignment of DNA during meiosis when unequal crossing-over events occur resulting in one sister chromatid having more repeat sequences and less repeats in the other sister chromatid. A similar but less complex type of VNTR are Microsatellite DNA markers which are repeats of dinucleotides which can have 4 to 20 alleles within relatively small regions of less than 100 base pairs. These sequences usually are in non-coding regions so there is no functional consequence to length variations of these markers. In this case of the VNTR array in *ACAN*, it occurs in a coding region and has a functional and structural use in the protein within cartilage. This complex repeat region is very large and errors in DNA pairing during meiosis would be common. The resulting variation for sequence length results in broad differences in size and weight of the subsequent protein. The presence of copy number variation in this VNTR is anticipated but for the variation to have a range of 10-15% of the repeat region is somewhat surprising in this study among Thoroughbreds, where the repeat range of allele sizes seen among horse breeds can be up to 100% (Eberth unpublished data). Indeed, this is seen in microsatellites where alleles can vary by 100% in the number of repeat units. The restricted range of repeats size alleles is likely due to those sizes being present in the foundation individuals of the Thoroughbred and thus have been perpetuated.

The existence of repeat size alleles within other horse breeds presenting a range of 36-72 repeats in a coding exon shows the functional diversity of this large complex repeat array. There also may be functional reasons the *aggrecan* protein in how it interacts with other proteins to create the ECM within cartilage. Larger or smaller number of repeats may result in restriction or loss of function in the horse considering the wide range of sizes seen between breeds, for instance a Miniature versus a Percheron. Perhaps this is because significant gain or loss of repeats can affect the cushioning ability and lifespan of the molecule or structure competency of the matrix. The length of the VNTR fell within a very narrow range for 80% of the control population. This implies that there is strong functional selection for length of the region and that very long and very short repeats are not compatible with the Thoroughbred phenotype desired, unlike what is seen in people with less repeats in the VNTR (Kawaguchi *et al*. 1999) which are not under strict selection of a phenotype. We did not see extreme phenotypes in the Thoroughbred but presumably such variants exist. It may be due to the relatively small number of animals tested or those individuals may be lost through culling in selection by breeders. It is noteworthy, no large draft breeds were examined in this project. It is unknown if the largest breeds of this species may possess larger number of repeats of the VNTR, or in a multi-breed survey what the range of repeats exist functionally in the VNTR region of *ACAN* within the species. Since this study, methods have been developed to amplify long DNA sequences more efficiently, including PAC-Bio® Long-Range DNA sequencing. Application of these techniques may allow us to resolve the length of the alleles more accurately.

REFERENCES

Al Abri, M. A., Posbergh, C., Palermo, K., Sutter, N. B., Eberth, J., Hoffman, G. E., & Brooks, S. A. (2018). Genome-wide scans reveal a quantitative trait locus for withers height in horses near the ANKRD1 gene. *Journal of Equine Veterinary Science*, *60*, 67-73.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., ... & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature methods*, *7*(4), 248-249.

Agarwal, S.K., Cogburn, L.A., Burnside, J., 1994. Dysfunctional growth hormone receptor in a strain of sex-linked dwarf chicken: evidence for a mutation in the intracellular domain. Journal of Endocrinology 142, 427–434.

AMHA. American Miniature Horse Association (2023) *Official Rule Book of American Miniature Horse Association*, *25th Edition* (2023). Alvarado, TX: American Miniature Horse Association.

AMHR. American Miniature Horse Registry (2023). *The American Miniature Horse Registry Official Rule Book,* (2023). Morton IL: American Miniature Horse Registry.

de Andrade, Danilo Giorgi Abranches, et al. "Evaluation of a new variant in the aggrecan gene potentially associated with chondrodysplastic dwarfism in Miniature horses." *Scientific reports* 10.1 (2020): 1-10.

Antonsson, P., Heinegård, D., & Oldberg, A. (1989). The keratan sulfate-enriched region of bovine cartilage proteoglycan consists of a consecutively repeated hexapeptide motif. *Journal of Biological Chemistry*, *264*(27), 16170-16173.

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., ... & Yeh, L. S. L. (2004). UniProt: the universal protein knowledgebase. *Nucleic acids research*, *32*(suppl_1), D115-D119.

Bao, L., Zhou, M., & Cui, Y. (2005). nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic acids research*, *33*(suppl_2), W480-W482.

Battié MC, Videman T. Lumbar disc degeneration: epidemiology and genetics. *J Bone Joint Surg Am.* 2006 88 Suppl 2:3-9.

Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E. D., Zendulka, J., ... & Damborsky, J. (2014). PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS computational biology*, *10*(1), e1003440.

Bellus GA, McIntosh I, Smith EA, Ayslworth AS, Kaitila I, Horton WA, Greenhaw FA, Hecht JT, Francomano CA. (1995). A recurrent mutation in the tyrosine kinase domain of fibroblast growth factor receptor 3 causes hypochondroplasia. *Nature Genetics,* 10:357–359.

Bochukova, E., Schoenmakers, N., Agostini, M., Schoenmakers, E., Rajanayagam, O., Keogh, J. M., & Chatterjee, K. (2012). A mutation in the thyroid hormone receptor alpha gene. *New England Journal of Medicine*, *366*(3), 243-249.

Boegheim, I. J., Leegwater, P. A., van Lith, H. A., & Back, W. (2017). Current insights into the molecular genetic basis of dwarfism in livestock. *The Veterinary Journal*, *224*, 64-75.

Bongartz, P. (2019). Resolving repeat families with long reads. *BMC bioinformatics*, *20*, 1-11.

Bottner A, Keller E, Kratzsch J, Stobbe H, Weigel JF, Keller A, Hirsch W, Kiess W, Blum WF, Pfaffle RW. PROP1 mutations cause progressive deterioration of anterior pituitary function including adrenal insufficiency: a longitudinal analysis. *Journal of Clinical Endocrinology & Metabolism*. 2004; 89:5256–65.

Bromberg, Y., & Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on function. *Nucleic acids research*, *35*(11), 3823-3835.

Bourdon, M. A., Krusius, T., Campbell, S., Schwartz, N. B., & Ruoslahti, E. (1987). Identification and synthesis of a recognition signal for the attachment of glycosaminoglycans to proteins. *Proceedings of the National Academy of Sciences*, *84*(10), 3194-3198.

Burge, C. B., T. Tuschl, and P. A. Sharp. "The RNA world." *Center for Cancer Research and Department of Biology Massachusetts Institute of Technology* (1999).

Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L., & Casadio, R. (2009). Functional annotations improve the predictive score of human disease-related mutations in proteins. *Human mutation*, *30*(8), 1237-1244.

Capriotti, E., Calabrese, R., & Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*, *22*(22), 2729-2734.

Capriotti, E., Altman, R. B., & Bromberg, Y. (2013). Collective judgment predicts disease-associated single nucleotide variants. *BMC genomics*, *14*, 1-9.

Cartegni, Luca, Shern L. Chew, and Adrian R. Krainer. "Listening to silence and understanding nonsense: exonic mutations that affect splicing." *Nature reviews genetics* 3.4 (2002): 285-298.

Cavanagh, J. A., Tammen, I., Windsor, P. A., Bateman, J. F., Savarirayan, R., Nicholas, F. W., & Raadsma, H. W. (2007). Bulldog dwarfism in Dexter cattle is caused by mutations in ACAN. *Mammalian Genome*, *18*, 808-814.

Choi Y., Sims G.E., Murphy S., Miller J.R. & Chan A.P. (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS ONE 7, e46688.

Chu, C., Pei, J., & Wu, Y. (2018). An improved approach for reconstructing consensus repeats from short sequence reads. *BMC genomics*, *19*, 9-17.

Cong L, Pang H, Xuan D, Tu GJ. Association between the expression of aggrecan and the distribution of aggrecan gene variable number of tandem repeats with symptomatic lumbar disc herniation in Chinese Han of Northern China. *Spine* 2010 35(14):1371-6.

Crawford, Kylie L. et al. "Survival Analysis of Training Methodologies and Other Risk Factors for Musculoskeletal Injury in 2-Year-Old Thoroughbred Racehorses in Queensland, Australia." Frontiers in veterinary science 8 (2021): 698298–698298.

Daetwyler, H.D., Capitan, A., Pausch, H., Stothard, P., van Binsbergen, R., Brøndum, R. F., Liao, X., Djari, A., Rodriguez, S.C., Grohs, C., et al., 2014. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. Nature Genetics 46, 858–865.

Dateki, S. (2017) *ACAN* mutations as a cause of familial short stature. Clinical Pediatric Endocrinology 26, 119-125.

Dateki, S, Nakatomi, A, Watanabe, S, Shimizu, H, Inoue, Y, Baba, H, et al.Identification of a novel heterozygous mutation of the Aggrecan gene in a family with idiopathic short stature and multiple intervertebral disc herniation. J Hum Genet 2017

Dean, D.D., Martel-Pelletier, J., Pelletier, J.P., Howell, D.S. and Woessner, J.F., 1989. Evidence for metalloproteinase and metalloproteinase inhibitor imbalance in human osteoarthritic cartilage. *The Journal of clinical investigation*, *84*(2), pp.678-685.

Deutsch, C. K., Shell, A. R., Francis, R. W., & Bird, B. D. (2012). The Farkas system of craniofacial anthropometry: methodology and normative databases. In *Handbook of Anthropometry* (pp. 561-573). Springer, New York, NY.

Doege KJ, Sasaki M, Kimura T, Yamada Y. Complete coding sequence and deduced primary structure of the human cartilage large aggregating proteoglycan, aggrecan. Human-specific repeats, and additional alternatively spliced forms. *J Biol Chem*. 1991 266(2):894-902.

Doege KJ, Coulter SN, Meek LM, Maslen K, Wood JG. 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. *J Biol Chem*. 1997 272(21):13974-9.

Drake, Abby Grace (2011). "Dispelling dog dogma: An investigation of heterochrony in dogs using 3D geometric morphometric analysis of skull shape". *Evolution & Development*. 13 (2): 204–213.

Durward-Akhurst, S. A., Schaefer, R. J., Grantham, B., Carey, W. K., Mickelson, J. R., & McCue, M. E. (2021). Genetic variation and the distribution of variant types in the horse. *Frontiers in genetics*, *12*, 2428.

Eberth, J.E., Swerczak, T.S. & Bailey, E. (2009) Investigations of dwarfism among Miniature horses using the Illumina horse SNP50 Bead Chip. Journal of Equine Veterinary Science 29, 315.

Eberth, J. (2013) Chondrodysplasia-like dwarfism in the Miniature Horse. MSc Thesis, University of Kentucky. Lexington, KY.

Eberth, J. E., Graves, K. T., MacLeod, J. N., & Bailey, E. (2018). Multiple alleles of ACAN associated with chondrodysplastic dwarfism in Miniature horses. *Animal genetics*, *49*(5), 413-420.

Farkas, Leslie G., ed. *Anthropometry of the Head and Face*. Lippincott Williams & Wilkins, 1994.

Fosang, A. J., & Hardingham, T. E. (1989). Isolation of the N-terminal globular protein domains from cartilage proteoglycans. Identification of G2 domain and its lack of interaction with hyaluronate and link protein. *Biochemical Journal*, *261*, 801-809.

Fosang, A. J., & Hardingham, T. E. (1991). 1-C-6 epitope in cartilage proteoglycan G2 domain is masked by keratan sulphate. *Biochemical Journal*, *273*, 369-373.

Frankeny, R. L., & Duren, S. (2003). *Miniature Horses: A Veterinary Guide for Owners and Breeders*. Trafalgar Square Pub.

Frischknecht, M., Jagannathan, V., Plattet, P., Neuditschko, M., Signer-Hasler, H., Bachmann, I., ... & Leeb, T. (2015). A non-synonymous HMGA2 variant decreases height in Shetland ponies and other small horses. *PloS one*, *10*(10), e0140749.

Fülöp, C., Walcz, E., Valyon, M., & Glant, T. T. (1993). Expression of alternatively spliced epidermal growth factor-like domains in *aggrecan*s of different species. Evidence for a novel module. *Journal of Biological Chemistry*, *268*(23), 17377-17383.

Gandolfi B, Alamri S, Darby WG, Adhikari B, Lattimer JC, Malik R, Wade CM, Lyons LA, Cheng J, Bateman JF, McIntyre P, Lamandé SR, Haase B. A dominant TRPV4 variant underlies osteochondrodysplasia in Scottish fold cats. Osteoarthritis Cartilage. 2016 Aug;24(8):1441-50. doi: 10.1016/j.joca.2016.03.019. Epub 2016 Apr 6. PMID: 27063440.

Georgopoulos, S.P.; Parkin, T.D.H. Risk factors associated with fatal injuries in Thoroughbred racehorses competing in flat racing in the United States and Canada. J. Am. Vet. Med Assoc. 2016, 249, 931–939.

González-Pérez, A., & López-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *The American Journal of Human Genetics*, *88*(4), 440-449.

Gramm M, Marksteiner R. The effect of age on thoroughbred racing performance. *J Equine Sci*. 2010;21(4):73-78. doi:10.1294/jes.21.73

Graves, K. T., Eberth, J. E., & Bailey, E. (2020). Heterozygotes for ACAN dwarfism alleles in horses have reduced stature. *Animal genetics*, *51*(3), 420-422.

Harper, P.A., Latter, M.R., Nicholas, F.W., Cook, R.W., Gill, P.A., 1998. Chondrodysplasia in Australian Dexter cattle. Australian Veterinary Journal 76, 199–202.

Hauer, N. N., Sticht, H., Boppudi, S., Büttner, C., Kraus, C., Trautmann, U., ... & Thiel, C. T. (2017). Genetic screening confirms heterozygous mutations in ACAN as a major cause of idiopathic short stature. *Scientific Reports*, *7*(1), 12225.

Heinegård, D., & Hascall, V. C. (1974). Aggregation of cartilage proteoglycans III. Characteristics of the proteins isolated from trypsin digests of aggregates. *Journal of Biological Chemistry*, *249*(13), 4250-4256.

Heleski, C., Stowe, C. J., Fiedler, J., Peterson, M. L., Brady, C., Wickens, C., & MacLeod, J. N. (2020). Thoroughbred Racehorse Welfare through the Lens of 'Social License to Operate—With an Emphasis on a US Perspective. *Sustainability*, *12*(5), 1706.

Hendriks-Stegman, B., Augustijn, K., Bakker, B., Holthuizen, P., Van Der Vliet, P. & Jansen, M. (2001) Combined pituitary hormone deficiency caused by compound heterozygosity for two novel mutations in the POU domain of the *PIT1*/*POU1F1*. *Journal of Clinical Endocrinology and Metabolism*, 86, 1545 – 1550.

Hitchens, P.L.; Morrice-West, A.V.; Stevenson, M.A.; Whitton, R.C. Meta-analysis of risk factors for racehorse catastrophic musculoskeletal injury in flat racing. Vet. J. 2019, 245, 29–40.

Hovdey, J. (1999). *Cigar: America's Horse*. Eclipse Press.

Hull, K.L., Fraser, R.A., Marsh, J.A., Harvey, S., 1993. Growth hormone receptor gene expression in sex-linked dwarf Leghorn chickens—evidence against a gene deletion. Journal of Endocrinology 137, 91–98.

Ikemoto, K., Fujimoto, H., & Fujimoto, A. (2023). Localized assembly for long reads enables genome-wide analysis of repetitive regions at single-base resolution in human genomes. *Human Genomics*, *17*(1), 1-16.

Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 14, 178-192 (2013).

Irie, Y., Tatsumi, K., Ogawa, M., Kamijo, T., Preeyasombat, C., Suprasongsin, C. & Amino, N. (1995) A novel E250X mutation of the *PIT1* gene in a patient with combined pituitary hormone deficiency. *Endocrinology Journal*, 42, 351 – 354.

Jian, Xueqiu, Eric Boerwinkle, and Xiaoming Liu. "In silico prediction of splice-altering single nucleotide variants in the human genome." *Nucleic acids research* 42.22 (2014): 13534-13544.

Jin, S., Zhu, F., Wang, Y., Yi, G., Li, J., Lian, L., Zheng, J., Xu, G., Jiao, R., Gong, Y., Hou, Z., Yang, N., 2016. Deletion of Indian hedgehog gene causes dominant semi-lethal creeper trait in chicken. Scientific Reports 6, 30172.

Kawabata, T., Ota, M., & Nishikawa, K. (1999). The protein mutant database. *Nucleic acids research*, *27*(1), 355-357.

Kawaguchi Y, Osada R, Kanamori M, Ishihara H, Ohmori K, Matsui H, Kimura T. Association between an aggrecan gene polymorphism and lumbar disc degeneration. *Spine* 1999 24(23):2456-60.

Kelmer, G., & Wilson, D. A. (2007). Bilateral metatarsophalangeal arthrodesis for severe congenital flexor tendon laxity and metatarsophalangeal luxation in a miniature horse. *Equine Veterinary Education*, *19*(10), 547-550.

Kiani C, Lee V, Cao L, Chen L, Wu Y, Zhang Y, Adams ME, Yang BB. (2001). Roles of *aggrecan* domains in biosynthesis, modification by glycosaminoglycans and product secretion. *Biochemical Journal*, *354*, 199-207.

Kiani C, Chen L, Wu YJ, Yee AJ, Yang BB. 2002. Structure and function of aggrecan. Cell Research12: 19–32.

Kolar, John C., and Elizabeth M. Salter. *Craniofacial anthropometry: Practical measurement of the head and face for clinical, surgical, and research use*. Charles C Thomas Pub Limited, 1997.

Koltes, J.E., Mishra, B.P., Kumar, D., Kataria, R.S., Totir, L.R., Fernando, R.L., Cobbold, R., Steffen, D., Coppieters, W., Georges, M., et al., 2009. A nonsense mutation in cGMP-dependent type II protein kinase (PRKG2) causes dwarfism in American Angus cattle. Proceedings of the National Academy of Sciences of the United States of America 106, 19250–19255.

Krakow, D., Lachman, R. S., & Rimoin, D. L. (2009). Guidelines for the prenatal diagnosis of fetal skeletal dysplasias. *Genetics in Medicine*, *11*(2), 127.

Krueger, R. C., Fields, T. A., Hildreth, J., & Schwartz, N. B. (1990). Chick cartilage chondroitin sulfate proteoglycan core protein. I. Generation and characterization of peptides and specificity for glycosaminoglycan attachment. *Journal of Biological Chemistry*, *265*(20), 12075-12087.

Kumar P., Henikoff S. & Ng P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature Protocols 4, 1073-1081.

Lark, M. W., Bayne, E. K., Flanagan, J., Harper, C. F., Hoerrner, L. A., Hutchinson, N. I., & Lohmander, L. S. (1997). *Aggrecan* degradation in human cartilage. Evidence for both matrix metalloproteinase and *aggrecan*ase activity in normal, osteoarthritic, and rheumatoid joints. *Journal of Clinical Investigation*, *100*(1), 93.

Leegwater, P. A., Vos-Loohuis, M., Ducro, B. J., Boegheim, I. J., van Steenbeek, F. G., Nijman, I. J., ... & Schurink, A. (2016). Dwarfism with joint laxity in Friesian horses is associated with a splice site mutation in B4GALT7. *Bmc Genomics*, *17*(1), 1-9.

Li, B., Krishnan, V. G., Mort, M. E., Xin, F., Kamati, K. K., Cooper, D. N., ... & Radivojac, P. (2009). Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*, *25*(21), 2744-2750.

Li, H., Schwartz, N. B., & Vertel, B. M. (1993). cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core protein and identification of a stop codon in the aggrecan gene associated with the chondrodystrophy, nanomelia. *Journal of Biological Chemistry*, *268*(31), 23504-23511.

Li H, Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010; 26(5): 589–95

Liao, W. W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., ... & Paten, B. (2023). A draft human pangenome reference. *Nature*, *617*(7960), 312-324.

Liu, J., Boyd, C.K., Kobayashi, Y., Chase, C.C., Hammond, A.C., Olson, T.A., Elsasser, T.H., Lucy, M.C., 1999. A novel phenotype for Laron dwarfism in miniature Bos indicus cattle suggests that the expression of growth hormone receptor 1A in liver is required for normal growth. Domestic Animal Endocrinology 17, 421–437.

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

Maeda, Yousuke, Michiko Hanada, and Masa-Aki Oikawa. "Epidemiology of racing injuries in Thoroughbred racehorses with special reference to bone fractures: Japanese experience from the 1980s to 2000s." *Journal of equine science* 27.3 (2016): 81-97.

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.

Mashayekhi F, Shafiee G, Kazemi M, Dolati P. Lumbar disk degeneration disease and aggrecan gene polymorphism in northern Iran. *Biochem Genet*. 2010 48(7-8):684-9.

Mayer, Jillian E., et al. "Genetic polymorphisms associated with intervertebral disc degeneration." *The Spine Journal* 13.3 (2013): 299-317.

McKenna A, Hanna M, Banks E, et al., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010; 20(9): 1297–303.

McLaren, W. "The Ensembl Variant Effect Predictor Genome biology 17, 122." (2016).

Metallinos, D. L., Bowling, A. T., & Rine, J. (1998). A missense mutation in the endothelin-B receptor gene is associated with Lethal White Foal Syndrome: an equine version of Hirschsprung disease. *Mammalian Genome*, *9*(6), 426-431.

Metzger, J., Schrimpf, R., Philipp, U., & Distl, O. (2013). Expression levels of LCORL are associated with body size in horses. *PLoS One*, *8*(2), e56497.

Metzger J., Gast A.C., Schrimpf R., Rau .J, Eikelberg D., Beineke A., Hellige M., & Distl O.(2016) Whole-genome sequencing reveals a potential causal mutation for dwarfism in the Miniature Shetland pony. Mammalian Genome 8, 143-151.

Mihalic Mosher, T., Zygmunt, D. A., Koboldt, D. C., Kelly, B. J., Johnson, L. R., McKenna, D. S., ... & McBride, K. L. (2019). Expansion of B4GALT7 linkeropathy phenotype to include perinatal lethal skeletal dysplasia. *European Journal of Human Genetics*, *27*(10), 1569-1577.

van Mullem, A., van Heerebeek, R., Chrysis, D., Visser, E., Medici, M., Andrikoula, M., Tsatsoulis, A., Peeters, R., Visser, T. J. (2012) Clinical phenotype and mutant TR-alpha-1. (Letter) *New England Journal of Medicine,* 366: 1451-1453.

Naviaux, Barbara. *Miniature Horses - Their Care, Breeding And Coat Colors*. 1999.

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research*, *31*(13), 3812-3814.

Nielsen, V.H., Bendixen, C., Arnbjerg, J., Sørensen, C.M., Jensen, H.E., Shukri, N.M., Thomsen, B., 2000. Abnormal growth plate function in pigs carrying a dominant mutation in type X collagen. Mammalian Genome 11, 1087–1092.

Nilsson, O, Guo, MH, Dunbar, N, Popovic, J, Flynn, D, Jacobsen, C, et al. Short stature, accelerated bone maturation, and early growth cessation due to heterozygous aggrecan mutations. J Clin Endocrinol Metab 2014;99: E1510–8.

Pannier, S., Martinovic, J., Heuertz, S., Delezoide, A. L., Munnich, A., Schibler, L., & Legeai-Mallet, L. (2009). Thanatophoric dysplasia caused by double missense FGFR3 mutations. *American Journal of Medical Genetics Part A*, *149*(6), 1296-1301.

Paulsson, M., Morgelin, M., Wiedemann, H., Beardmore-Gray, M., Dunham, D., Hardingham, T., & Engel, J. (1987). Extended and globular protein domains in cartilage proteoglycans. *Biochemical Journal*, *245*, 763-772.

Perkins, S. J., Nealis, A. S., Dudhia, J., & Hardingham, T. E. (1989). Immunoglobulin fold and tandem repeat structures in proteoglycan N-terminal domains and link protein. *Journal of Molecular Biology*, *206*(4), 737-748.

Olatubosun, A., Väliaho, J., Härkönen, J., Thusberg, J., & Vihinen, M. (2012). PON-P: Integrated predictor for pathogenicity of missense variants. *Human mutation*, *33*(8), 1166-1174.

Orr, N., Back, W., Gu, J., Leegwater, P., Govindarajan, P., Conroy, J., Ducro, B., Van Arendonk, J.A., MacHugh, D.E., Ennis, S., et al., 2010. Genome-wide SNP association-based localization of a dwarfism gene in Friesian dwarf horses. Animal Genetics 41 (Suppl. 2), 2–7.

Quintos, JB, Guo, MH, Dauber, A. Idiopathic short stature due to novel heterozygous mutation of the aggrecan gene. J Pediatr Endocrinol Metab 2015;28: 927–32.

Rafati, N., Andersson, L. S., Mikko, S., Feng, C., Raudsepp, T., Pettersson, J., ... & Rubin, C. J. (2016). Large deletions at the SHOX locus in the pseudoautosomal region are associated with skeletal atavism in Shetland Ponies. *G3: Genes, Genomes, Genetics*, *6*(7), 2213-2223.

Ramensky, V., Bork, P., & Sunyaev, S. (2002). Human non-synonymous SNPs: server and survey. *Nucleic acids research*, *30*(17), 3894-3900.

Rohlf, F. James, and Leslie F. Marcus. "A revolution morphometrics." *Trends in ecology & evolution* 8.4 (1993): 129-132.

Roughley, P., et al. "The involvement of aggrecan polymorphism in degeneration of human intervertebral disc and articular cartilage." *Eur Cell Mater* 11.1 (2006): 1-7.

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.

Samol, M. A., Uzal, F. A., Blanchard, P. C., Arthur, R. M., & Stover, S. M. (2021). Sudden death caused by spinal cord injury associated with vertebral fractures and fetlock failure in a Thoroughbred racehorse. *Journal of Veterinary Diagnostic Investigation*, *33*(4), 788-791.

Santschi, E. M., Purdy, A. K., Valberg, S. J., Vrotsos, P. D., Kaese, H., & Mickelson, J. R. (1998). Endothelin receptor B polymorphism associated with lethal white foal syndrome in horses. *Mammalian Genome*, *9*, 306-309.

Sartelet, A., Druet, T., Michaux, C., Fasquelle, C., Géron, S., Tamma, N., Zhang, Z., Coppieters, W., Georges, M., Charlier, C., 2012. A splice site variant in the bovine RNF11 gene compromises growth and regulation of the inflammatory response. PLoS Genetics 8, e1002581.

Schwarzenbacher, H., Wurmser, C., Flisikowski, K., Misurova, L., Jung, S., Langenmayer, M.C., Schnieke, A., Knubben-Schweizer, G., Fries, R., Pausch, H., 2016. A frameshift mutation in GON4L is associated with proportionate dwarfism in Fleckvieh cattle. Genetics Selection Evolution 48, 25.

Sentchordi-Montané, L., Aza-Carmona, M., Benito-Sanz, S., Barreda-Bonis, A. C., Sánchez-Garre, C., Prieto-Matos, P., ... & Heath, K. E. (2018). Heterozygous aggrecan variants are associated with short stature and brachydactyly: description of 16 probands and a review of the literature. *Clinical endocrinology*, *88*(6), 820-829.

Shiang R, Thompson LM, Zhu YZ, Church DM, Fielder TJ, Bocian M, Winokur ST, Wasmuth JJ. (1994). Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia. *Cell*, *78*(2), 335-342.

Shembade, N., Parvatiyar, K., Harhaj, N.S., Harhaj, E.W., 2009. The ubiquitin-editing enzyme A20 requires RNF11 to downregulate NF-kB signalling. EMBO Journal 28, 513–522.

Signer-Hasler, H., Flury, C., Haase, B., Burger, D., Simianer, H., Leeb, T., & Rieder, S. (2012). A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PloS one*, *7*(5), e37282.

Solovieva S, Noponen N, Männikkö M, Leino-Arjas P, Luoma K, Raininko R, Ala-Kokko L, Riihimäki H. Association between the aggrecan gene variable number of tandem repeats polymorphism and intervertebral disc degeneration. *Spine* 2007 32(16):1700-5.

Stattin, EL, Wiklund, F, Lindblom, K, Onnerfjord, P, Jonsson, BA, Tegner, Y, et al. A missense mutation in the aggrecan C-type lectin domain disrupts extracellular matrix interactions and causes dominant familial osteochondritis dissecans. Am J Hum Genet 2010;86: 126–37.

van der Steen, M, Pfundt, R, Maas, SJ, Bakker-van Waarde, WM, Odink, RJ, Hokken-Koelega, AC. ACAN gene mutations in short children born SGA and response to growth hormone treatment. J Clin Endocrinol Metab 2016 (e-pub ahead of print 6 Oct 2016;doi:).

Stone, E. A., & Sidow, A. (2005). Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome research*, *15*(7), 978-986.

Struck, Ann-Kathrin, *et al.* "A recessive lethal chondrodysplasia in a miniature zebu family results from an insertion affecting the chondroitin sulfate domain of aggrecan." *BMC genetics* 19.1 (2018): 1-9.

Takeda, H., Takami, M., Oguni, T., Tsuji, T., Yoneda, K., Sato, H., Ihara, N., Itoh, T., Kata, S.R., Mishina, Y., et al., 2002. Positional cloning of the gene LIMBIN responsible for bovine chondrodysplastic dwarfism. Proceedings of the National Academy of Sciences of the United States of America 99, 10549–10554.

Tetens, J., Widmann, P., Kühn, C., & Thaller, G. (2013). A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in G erman W armblood horses. *Animal genetics*, *44*(4), 467-471.

Thomas, P. D., & Kejariwal, A. (2004). Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proceedings of the National Academy of Sciences*, *101*(43), 15398-15403.

Thompson, K.G., Blair, H.T., Linney, L.E., West, D.M., Byrne, T., 2005. Inherited chondrodysplasia in Texel sheep. New Zealand Veterinary Journal 53, 208–212.

Tompson, S. W., Merriman, B., Funari, V. A., Fresquet, M., Lachman, R. S., Rimoin, D. L., ... & Krakow, D. (2009). A recessive skeletal dysplasia, SEMD aggrecan type, results from a missense mutation affecting the C-type lectin domain of aggrecan. *The American Journal of Human Genetics*, *84*(1), 72-79.

Tortorella MD, Burn TC, Pratta MA, Abbaszade I, Hollis JM, Liu R, Rosenfeld SA, Copeland RA, Decicco CP, Wynn R, Rockwell A, Yang F, Duke JL, Solomon K, George H, Bruckner R, Nagase H, Itoh Y, Ellis DM, Ross H, Wiswall BH, Murphy K, Hillman MC, Jr., Hollis GF, Arner EC, *et al*. (1999). Purification and cloning of *aggrecan*ase-1: a member of the ADAMTS family of proteins. *Science*, *284*(5420), 1664-1666.

Tozaki, T., Miyake, T., Kakoi, H., Gawahara, H., Sugita, S., Hasegawa, T., ... & Nakano, Y. (2010). A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the MSTN gene. *Animal genetics*, *41*, 28-35.

Tyson, R., Graham, J. P., Colahan, P. T., & Berry, C. R. (2004). Skeletal atavism in a miniature horse. *Veterinary Radiology & Ultrasound*, *45*(4), 315-317.

Valhmu, W. B., Palmer, G. D., Rivers, P. A., Ebara, S., Cheng, J. F., Fischer, S., & Ratcliffe, A. (1995). Structure of the human *aggrecan* gene: exon-intron organization and association with the protein domains. *Biochemical Journal*, *309*, 535-542.

Warburton, P. E., Hasson, D., Guillem, F., Lescale, C., Jin, X., & Abrusan, G. (2008). Analysis of the largest tandemly repeated DNA families in the human genome. *BMC genomics*, *9*, 1-18.

Warman ML, Cormier-Daire V, Hall C, Krakow D, Lachman R, et al. 2011. Nosology and classification of genetic skeletal disorders: 2010 revision. Am. J. Med. Genet. A 155A:943–68

Watanabe, H., Kimata, K., Line, S., Strong, D., Gao, L. Y., Kozak, C. A., & Yamada, Y. (1994). Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the aggrecan gene. *Nature genetics*, *7*(2), 154-157.

Watanabe, H., Yamada, Y., & Kimata, K. (1998). Roles of aggrecan, a large chondroitin sulfate proteoglycan, in cartilage structure and function. *Journal of biochemistry*, *124*(4), 687-693.

Watanabe, M. J., Alonso, J. M., Yamada, A. L. M., Rodrigues, K. A., Hussni, C. A., & Alves, A. L. G. (2014). Phenotypic findings of dwarfism in newborn pony: cases report. *Ars Veterinaria*, *30*(1), 1-4.

Wolfe, G., MacNaul, K.L., Buechel, F.F., McDonnell, J., Hoerrner, L., Lark, M., Moore, V. and Hutchinson, N., 1993. Differential in vivo expression of collagenase and stromelysin in synovium and cartilage: In human RA and OA patients and in two animal models of acute inflammatory arthritis. *Arthritis Rheum*, *36*, pp.1540-7.

Yan, G. C., Croaker, D., Zhang, A. L., Manglick, P., Cartmill, T., & Cass, D. (1998). A dinucleotide mutation in the endothelin-B receptor gene is associated with lethal white foal syndrome (LWFS); a horse variant of Hirschsprung disease (HSCR). *Human molecular genetics*, *7*(6), 1047-1052.

Zhao, X., Onteru, S.K., Piripi, S.A., Thompson, K.G., Blair, H., Garrick, D.J., Rothschild, M.F., 2012. In a shake of a lamb's tail: using genomics to unravel a cause of chondrodysplasia in Texel sheep. Animal Genetics 43, 9–18.

VITA

John Edmund Eberth

Education

University of Kentucky 2014-2023
Lexington, Kentucky, USA
Doctor of Philosophy in Veterinary Science
College of Agriculture, Department of Veterinary Science
Maxwell H. Gluck Equine Research Center

University of Kentucky 1996-2013
Lexington, Kentucky, USA
Master of Science in Veterinary Science
College of Agriculture, Department of Veterinary Science
Maxwell H. Gluck Equine Research Center

Western Kentucky University 1995-1996
Bowling Green, Kentucky, USA
College of Agriculture, Department of Animal Science

DePauw University 1990-1995
Greencastle, Indiana, USA
Bachelor of Arts in Biology

Professional Positions

Graduate Research Assistant, University of Kentucky – August 1996 to May 2023

Geneticist, Equus Analytics LLC. – August 2018 to June 2023

Licensed AMHA Judge, AMHA – February 2008 to August 2015

President, Arion Management Inc. – January 1998 to December 2012

Equine Trainer, Little King Farm Inc. – May 1989 to August 1996

Veterinary Technician, Snodgrass Veterinary Clinic – August 1995 to June 1996

Intern, U.C. Davis Veterinary Genetics Testing Lab – December 1994 to February 1995

Intern, Equine Services PSC. – December 1993 – February 1994

Intern, Louisville Zoo – December 1992 – February 1993

Awards and Honors

Chairman Equine Genetics Committee**,** AMHA – February 2006 to February 2011

Science Award for High School Senior Achievement – May 1990

Publications

Thieulent, C.J., Carossino, M., Balasuriya, U.B., Graves, K., Bailey, E., **Eberth, J**., Canisso, I.F., Andrews, F.M., Keowen, M.L. and Go, Y.Y., 2022. Development of a TaqMan® Allelic Discrimination qPCR Assay for Rapid Detection of Equine CXCL16 Allelic Variants Associated With the Establishment of Long-Term Equine Arteritis Virus Carrier State in Stallions. *Frontiers in genetics*, p.826.

Graves, K.T., **Eberth, J.E**. and Bailey, E., 2020. Heterozygotes for ACAN dwarfism alleles in horses have reduced stature. *Animal Genetics*, *51*(3), pp.420-422.

Mitz, Carl, Stone, W., Miller, R., Felix, S., Christensen, J., Deluhery, J., Carlson, N., **Eberth, J**., Green, R., Lindholm, A., 2019. The Mitz Journal of Equine Dental Condition.

Tanaka, J., Leeb, T., Rushton, J., Famula, T.R., Mack, M., Jagannathan, V., Flury, C., Bachmann, I., **Eberth, J**., McDonnell, S.M. and Penedo, M.C.T., 2019. Frameshift variant in MFSD12 explains the mushroom coat color dilution in Shetland ponies. *Genes*, *10*(10), p.826.

Eberth, J.E., Graves, K.T., MacLeod, J.N. and Bailey, E., 2018. Multiple alleles of ACAN associated with chondrodysplastic dwarfism in Miniature horses. *Animal genetics*, *49*(5), pp.413-420.

Al Abri, M.A., Posbergh, C., Palermo, K., Sutter, N.B., **Eberth, J**., Hoffman, G.E. and Brooks, S.A., 2018. Genome-wide scans reveal a quantitative trait locus for withers height in horses near the ANKRD1 gene. *Journal of Equine Veterinary Science*, *60*, pp.67-73.

Regatieri, I.C., **Eberth, J.E**., Sarver, F., Lear, T.L. and Bailey, E., 2016. Comparison of DMRT 3 genotypes among American Saddlebred horses with reference to gait. *Animal genetics*, *47*(5), pp.603-605.

Rafati, N., Andersson, L.S., Mikko, S., Feng, C., Raudsepp, T., Pettersson, J., Janecka, J., Wattle, O., Ameur, A., Thyreen, G. and **Eberth, J**., 2016. Large deletions at the SHOX locus in the pseudoautosomal region are associated with skeletal atavism in Shetland Ponies. *G3: Genes, Genomes, Genetics*, *6*(7), pp.2213-2223.

Sarkar, S., Bailey, E., Go, Y.Y., Kalbfleisch, T., Chelvarajan, L., Cook, F., **Eberth, J**., Bhowmik, M., Artuishin, S., Timoney, P.J. and Balasuriya, U.B., 2016. Variants of CXCL16 gene is associated with in vitro CD3+ T cell susceptibility to equine arteritis virus infection. *Nat Commun*.

Sarkar, S., Bailey, E., Go, Y.Y., Cook, R.F., Kalbfleisch, T., **Eberth, J**., Chelvarajan, R.L., Shuck, K.M., Artiushin, S., Timoney, P.J. and Balasuriya, U.B., 2016. Allelic variation in CXCL16 determines CD3+ T lymphocyte susceptibility to equine arteritis virus infection and establishment of long-term carrier state in the stallion. *PLoS genetics*, *12*(12), p.e1006467.

Sarkar, S., Chelvarajan, L., Go, Y.Y., Cook, F., Artuishin, S., Mondal, S., Anderson, K., **Eberth, J**., Timoney, P.J., Kalbfleisch, T.S. and Bailey, E., 2016. Equine Arteritis Virus Uses Equine CXCL16 (EqCXCL16) as an Entry Receptor 2. *Receptor*, *2*(3), p.4.

Sarkar, S., Chelvarajan, L., Go, Y.Y., Cook, F., Artiushin, S., Mondal, S., Anderson, K., **Eberth, J**., Timoney, P.J., Kalbfleisch, T.S. and Bailey, E., 2016. Equine arteritis virus uses equine CXCL16 as an entry receptor. *Journal of virology*, *90*(7), pp.3366-3384.

Eberth, J.E., 2013. Chondrodysplasia-like dwarfism in the miniature horse. MS Thesis. University of Kentucky.

John E. Eberth. (2011) It's All in the Genes. *International Show Horse*. Blue Ribbon Edition. **2.** P. 162.

JE. Eberth, T. Swerczak, E. Bailey. (2009) Investigation of Dwarfism Among Miniature Horses using the Illumina Horse SNP50 Bead Chip. *Journal of Equine Veterinary Science*. **29**(5). P.315.

Abstracts

Eberth J. & Bailey E. (2010) Genetics of dwarfism in Miniature Horses. ISAG, Edinburgh, Scotland. July 25-30, 2010.

Regatieri, Inaê & **Eberth, John** & Sarver, Fred & Bailey, Ernest. (2015) Preliminary study of DMRT3 variation and association with performance gait for American Saddlebred horses. ADSA®-ASAS Joint Annual Meeting. Orlando, FL. July 2015.

Sofia, M., L. Andersson, Thyreen Gunilla, Dalin Göran, **Eberth John E**., Bailey E., Andersson Leif, Lindgren Gabriella, Rubin Carl-Johan. (2013) Identification of the locus causing skeletal atavism in Shetland ponies. Proceedings of the International Havemeyer Horse Genome Workshop, Azores, Portugal; July 10-12, 2013. p.54.

Bailey, E., Easley, J, **Eberth, J.** (2013) GWAS for congenital brachygnathia of Thoroughbred Horses. Proceedings of the International Havemeyer Horse Genome Workshop, Azores, Portugal; July 10-12, 2013. p.136.

Presentations at Scientific Meetings

JE Eberth. (May 2009) *Investigation of Dwarfism Among Miniature Horses using the Illumina Horse SNP50 Bead Chip.* 2009 Equine Science Society Symposium, Keystone, CO.

Oral Presentations

JE Eberth. (April 2023) *Molecular Genetic Studies of Horses, Especially With Reference to Aggrecan and Dwarfism.* Doctor of Philosophy Dissertation Defense Seminar. University of Kentucky, Lexington, KY.

JE Eberth. (July 2013) *Chondrodysplasia-like Dwarfism in the Miniature Horse*. Master's Thesis Defense Seminar. University of Kentucky, Lexington, KY.

JE Eberth. (May 2009) *Investigation of Dwarfism Among Miniature Horses using the Illumina Horse SNP50 Bead Chip.* 2009 Equine Science Society Symposium, Keystone, CO.

JE Eberth. (April 2008) *Preliminary Investigation of Candidate Genes for Type 1 Dwarfism in Miniature Horses*. Veterinary Science Graduate Seminar. University of Kentucky, Lexington, KY.

Outreach Presentations

JE Eberth. (March 2022) *Dwarfism in Miniature Horses.* Podcast Episode 2896. Horse Radio Network- Horses in the Morning

JE Eberth. (February 2018) *Genetics of Dwarfism in the Miniature Horse*. Miniature Horse Club of the England. London, England.

JE Eberth. (February 2018) *Nutrition, Fitness and Management of the Miniature Horse*. Miniature Horse Club of the England. London, England.

JE Eberth. (January 2013) *Genetics of Dwarfism in the Miniature Horse*. Miniature Horse Club of the Czech Republic. Prague, Czech Republic.

JE Eberth. (January 2013) *Equine Coat Colors*. Miniature Horse Club of the Czech Republic. Prague, Czech Republic.

JE Eberth. (January 2013) *Genetics of Dwarfism in the Miniature Horse*. Miniature Horse Club of the Netherlands. Antwerp, Flanders, Belgium.

JE Eberth. (January 2013) *Equine Coat Colors*. Miniature Horse Club of the Netherlands. Antwerp, Flanders, Belgium.

JE Eberth. (June 2009) *Investigation of Dwarfism Among Miniature Horses using the Illumina Horse SNP50 Bead Chip.* American Miniature Horse Association Board of Directors Meeting 2009. Fort Worth, TX.

JE Eberth. (February 2009) *Genetics of the Equine Oral Cavity*. International Association of Equine Dentistry. 2009 Annual Convention. Orlando, FL.

JE Eberth. (February 2008) *Dwarfism in Miniature Horses*. Miniature Horse Club of Ontario. Toronto, Ontario. Canada.

JE Eberth. (February 2007) *Dwarfism in Miniature Horses*. American Miniature Horse Association 2007 National Convention. Ft Lauderdale, FL.

John Edmund Eberth

June 21, 2023
Date