The prevalence of severe combined immunodeficiency, lavender foal syndrome and cerebellar abiotrophy in Arabian horses in South Africa

by

Carolynne Jane Tarr

Submitted in partial fulfillment of the requirements for the degree of

Magister Scientiae

Department of Production Animal Studies

Faculty of Veterinary Science

University of Pretoria

Onderstepoort

Supervisor

Dr. Cindy Harper

Co-supervisors

Prof. Alan Guthrie

Prof. Peter Thompson

DECLARATION

I, Carolynne Jane Tarr, do hereby declare that the research presented in this dissertation was conceived and executed by myself and, apart from the normal guidance from my supervisors, I have received no assistance.

Neither the substance, nor any part of this dissertation, has been submitted in the past or is to be submitted for a degree at this University or any other University.

This dissertation is presented in partial fulfillment of the requirements for the degree MSc in Production Animal Studies.

I hereby grant the University of Pretoria free license to reproduce this dissertation, in part or as a whole, for the purpose of research or continuing education.

Signed:

Carolynne Jane Tarr

Date:
20 June 2012

ACKNOWLEDGEMENTS

I would like to express my gratitude to the Veterinary Genetics Laboratory for providing me with the opportunity to perform this study.

Sincere thanks go to my principal supervisor, Dr. Cindy Harper, and co-supervisors, Prof. Alan Guthrie and Prof. Peter Thompson, for their input, guidance, and that most-valued commodity - time.

I am indebted to the staff of the Veterinary Genetics Laboratory, in particular Anette Ludwig, for technical assistance during the course of this project.

I thank the Arab Horse Society of South Africa for supplying lists of data used in this study and for kindly answering all my queries.

Special thanks go to my family, for their support and encouragement, and to my dearest Christopher, who makes it all worthwhile.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SCID	Severe combined immunodeficiency
LFS	Lavender foal syndrome
CA	Cerebellar abiotrophy
DNA	Deoxyribonucleic acid
DNA-PK	DNA-dependent protein kinase
DNA-PKcs	DNA-dependent protein kinase catalytic subunit
RAG	Recombinase Activating Gene
V(D)J	Variable (Diversity) Joining
ATP	Adenosine triphosphate
SNP	Single nucleotide polymorphism
8-oxoG	Seven,8-dihydro-8-oxoguanine
PCR	Polymerase chain reaction
bp	base pair(s)
CE	Capillary electrophoresis
AHSSA	Arab Horse Society of South Africa
VGL	Veterinary Genetics Laboratory, University of Pretoria
ISAG	International Society of Animal Genetics
CI	Confidence intervals

ABSTRACT

The prevalence of carriers of three genetic disorders, severe combined immunodeficiency, lavender foal syndrome and cerebellar abiotrophy, in registered, purebred Arabians in South Africa was assessed. Genotyping for the three disorders was performed on individuals randomly selected from two populations: purebred Arabian horses born in South Africa during the intervals 1 August 2004 to 31 July 2005 and 1 August 2009 to 31 July 2010, in line with physiological breeding seasons. This permitted an estimation of the change in prevalence of each disorder between 2004 and 2009, during which time compulsory testing for SCID, and selective breeding based on the results thereof, was performed. The prevalence of lavender foal syndrome and cerebellar abiotrophy for the 2009 breeding season was found to be 11.7% (95% confidence interval [CI] 7.80 – 16.37) and 5.1% (95% CI 2.56 – 8.69) respectively, with no statistically significant change in the prevalence of these disorders between 2004 and 2009. Utilizing a larger sample size, the prevalence of severe combined immunodeficiency was found to have decreased significantly from 6.4% (95% CI 6.05 – 6.78) in 2004 to 3.4% (95% CI 2.67 – 4.23) in 2009 ($P < 0.0001$). These results will encourage the genetic screening of Arabian horses intended for breeding purposes in order to prevent the birth of clinically affected individuals. This study also highlights the usefulness of genetic testing as a tool to decrease the prevalence of specific genetic disorders within animal populations.

CHAPTER 1: Introduction

The Arabian horse has had a close association with man since ancient times and is renowned for its beauty and intelligence. As one of the oldest breeds in existence, Arabian bloodlines have been used to develop most of the modern breeds of horses found today, lending refinement, speed and strength.

A small number of Arabian stallions were imported from the East into South Africa during the period 1800 to 1870 and were bred to local mares. However, the first Arabian horse of importance in South Africa is considered to be the stallion Azrek, imported in 1891 from Crabbet Park Stud, England, by Sir Cecil John Rhodes. Azrek was sadly not bred to any purebred Arabian mares in this country, his genetics being lost to mixed breeds as was the case with many of the first Arabians in South Africa. During the 1940's, the earliest pure-bred Arabian studs were established, notably Orange Valley Farm, owned by G.C. Kock, and Rynheath, owned by E.M. Hind. The Olford Arabian Stud is the oldest Arabian stud in South Africa today, founded by Betty Arnold in 1948.

The breed is known to be afflicted with a number of genetic disorders. Although much progress has been made, these diseases are often misunderstood and misdiagnosed, by breeders and veterinarians alike. The occurrence of genetic disorders on stud farms has also been stigmatized, with breeders tending to withhold such information from the public and potential clients.

The three genetic disorders studied here are inherited in an autosomal recessive manner. A carrier of the disease-causing allele shows no clinical signs. When breeding two carriers, the offspring have a one-in-four chance of being genetically homozygous normal, a two-in-four chance of being a carrier of the disorder, and a one-in-four chance of developing the disease. When breeding a carrier to a homozygous normal individual, the offspring from that mating have a 50% chance of being carriers of the mutation and a 50% chance of not carrying the mutation.

Through genetic testing for these disorders, carrier-to-carrier matings can be avoided, effectively preventing the birth of clinically affected foals. This prevents the loss of revenue to the breeder due to the loss of a foal, veterinary consultation and possibly attempted treatment, and costs incurred during the breeding and maintenance of the mare during the pregnancy. Genetic testing also prevents the suffering of an affected foal. Removing carriers from the breeding population could be used to eliminate

these disorders from the gene pool more rapidly; however, this could lead to the loss of desirable traits present in the carrier population.

An unpublished study done in 2004 estimated the prevalence of severe combined immunodeficiency (SCID) in South African Arabians to be 8.3%. Since 2005, a test-and-breed policy has been in place with regards to SCID (C Harper, Veterinary Genetics Laboratory, Onderstepoort, pers. comm., 2011). No formal prevalence studies have been performed for lavender foal syndrome (LFS) and cerebellar abiotrophy (CA) in South Africa.

This study assessed the current prevalence of SCID, LFS and CA in South African Arabians in order to further our understanding of these disorders and to quantify the impact of SCID, LFS and CA on the Arabian horse population in South Africa. This will be used to promote the genetic testing of breeding animals in order to reduce the overall prevalence of these mutations in the population.

The objectives of the current study were:

- Assessment of the current carrier prevalence of SCID, LFS and CA in Arabian horses in South Africa.
- Retrospective analysis of the change in prevalence of SCID carriers between 2004 and 2009, during which time a test-and-breed policy instituted by the Arab Horse Society of South Africa was in effect.
- Retrospective analysis of the change in prevalence of LFS and CA carriers between 2004 and 2009, during which time no test-and-breed policy was active, for comparison purposes against SCID.

CHAPTER 2: Literature Review

2.1 Severe Combined Immunodeficiency

Severe combined immunodeficiency (SCID) is a primary immunodeficiency disorder first described in horses in 1973, following reports of increased susceptibility of Arabian foals to fatal adenoviral infections [1].

Similar immunodeficiency disorders, each caused by a unique genetic mutation, are recognized in humans, mice and dogs. SCID in C.B-17 BALB/c mice [2] and Jack Russell Terriers [3], as in horses, is inherited as an autosomal recessive trait and is caused by mutations in the gene coding for DNA-dependent protein kinase (DNA-PK). SCID in the Bassett Hound [4] and Cardigan Welsh Corgi [5] is X-linked and is caused by mutations in the gene IL2RG. This results in the decreased functioning of interleukins, proteins which are critical in the communication and migration of lymphoid cells. Recently, SCID caused by a defect in the gene coding for Recombinase Activating Gene (RAG) 1 has been identified in the Frisian Water Dog [6]. SCID in humans occurs due to a number of possible genetic mutations, involving in particular the genes IL2RG, RAG1 or RAG2 [2].

Equine SCID is associated with horses of Arabian breeding [7]. The disorder has been diagnosed in an Appaloosa foal and, more recently, in a Caspian filly, based on clinical parameters and post mortem findings. Arabian ancestors are thought to be the source of SCID in the Appaloosa foal [8]. Genetic testing failed to confirm the diagnosis of SCID in the Caspian foal. This raises the possibility of an unidentified mutation, not detected by the current genetic test, having been the cause of SCID in this case [9].

Foals affected with SCID are incapable of producing antigen-specific immune responses due to a lack of functional B and T lymphocytes, which results in a profound susceptibility to infectious agents [7]. Although seemingly normal at birth, foals succumb to infections by the age of five months despite intensive veterinary care [10]. The age at which opportunistic infections arise in SCID-affected foals is likely to be influenced by the adequacy of passive transfer of maternal antibodies via colostrum, the rate of catabolism and dilution of maternal antibodies and the level of pathogen challenge to which foals are exposed. Correction of the disorder has been achieved through bone marrow transplantation from a

histocompatible donor; however the procedure is impractical in equine medicine and is not recommended [10].

SCID is inherited as an autosomal recessive trait [11, 12]. The genetic basis of the condition is a five base-pair deletion in the gene coding for DNA-dependent protein kinase catalytic subunit (DNA-PKcs) [13], located on equine chromosome nine [14].

DNA-PK, found in virtually all tissues [2], plays a crucial role during lymphocyte development, where gene segments encoding Variable (V), Diversity (D) and Joining (J) regions are randomly re-assorted in order to provide a diverse spectrum of antigen-specific T-cell receptors and B-lymphocyte surface immunoglobulin receptors. RAG1 and RAG2, found only in lymphoid tissues, cleave DNA at target sites adjacent to the immune receptor gene segments [2]. DNA-PK is one of at least seven proteins required for nonhomologous repair of the double strand break and completion of the V(D)J recombination process [15]. Two new DNA joints are formed: coding joints and signal joints. Coding joints, between V and (D) and between (D) and J regions, form part of the specific antigen binding regions of T-lymphocyte receptors and B-lymphocyte surface immunoglobulin receptors. Signal joints are formed by the circularization of excised segments of DNA [2]. It has been widely believed that signal joint formation renders signal ends inert. Signal ends could potentially take part in aberrant joining. However, more recent work suggests that signal joints may function in transposition events and thereby contribute to genomic variation in lymphocytes [16].

DNA-PK consists of 3 subunits, of which two subunits, with molecular masses of 80 and 70 kd, make up the DNA-binding, heterodimeric complex Ku. The 80 kd subunit binds first to the end of a ligated section of double stranded DNA. DNA-PKcs, the third subunit of DNA-PK, then joins Ku, resulting in the DNA-PK becoming a functional kinase. Proper function of DNA-PK is essential for coding and signal joint formation and hence V(D)J recombination in horses. The SCID mutation results in truncation of the DNA-PKcs protein with a loss of 967 amino acids from the C-terminus, resulting in a functionally inactive product. The subsequent failure of V(D)J recombination results in the elimination of lymphocyte precursors and a severe paucity of functional T- and B lymphocytes [2].

Anecdotally, DNA-PK may also have tumor suppressor properties due to its DNA repair functions. Cells deficient in DNA-PK activity have been shown to be hypersensitive to ionizing radiation [17] and

individuals heterozygous for the equine SCID mutation were found to have an increased susceptibility to equine sarcoids [18].

Until the advent of a genetic test for SCID, diagnosis was based upon the presence of a severe lymphopaenia (< 1000 lymphocytes/µl), undetectable serum IgM levels after 14 – 30 days of age, by which time maternal IgM is undetectable [19] and hypoplasia of the spleen or thymus, supported by the appropriate signalment [20]. DNA-based testing provides a more definitive method of diagnosis as well as a means of identifying carriers of the SCID mutation.

A prevalence study of SCID in the United States of America, using 250 randomly selected Arabian horses, estimated the frequency of carriers at 8.4% [21]. A similar study in the United Kingdom, utilizing a random sample of 106 Arabian horses, estimated the frequency of SCID carriers at 1 to 5% (95% confidence interval) [7]. Prevalence studies performed in Morocco and Brazil estimated the prevalence of SCID carriers in Arabian horses in these countries to be 7% and 1.5% respectively [22]. No SCID carriers were detected in studies of Arabian horses in Slovenia [23], Iran [24] or Poland [22]. In South Africa, an unpublished study in 2004, based on the random sampling of approximately 200 Arabian horses, estimated the prevalence of SCID carriers at 8.3% (C Harper, Veterinary Genetics Laboratory, Onderstepoort, pers. comm., 2011).

2.2 Lavender Foal Syndrome

Lavender foal syndrome (LFS), also known as coat colour dilution lethal [25], is a disorder of Arabian foals, particularly those of Egyptian descent [26]. The disease manifests from birth, following a normal or assisted delivery and, generally, a normal gestation period. LFS is characterised by a dilute coat colour, described as pink, pewter or lavender, and a range of neurological signs including recumbency, opisthotonus, paddling movements and extensor rigidity. Clustering of clinical signs resembles tetanic seizure activity. Affected foals are unable to stand and nurse despite a strong suck reflex [25, 27]. Intermittent, spontaneous nystagmus and strabismus were noted by Page *et al* [27].

Haematology, serum biochemistry and clinical pathology assays show no clinically specific changes [27]. The disorder is invariably lethal, with attempts at treatment proving futile and ending in humane euthanasia. At post mortem, an abnormal coat colour is frequently the only macroscopic lesion observed [25, 27]. Although Fanelli [25] described the abnormal clumping of melanin in hair roots and shafts on histopathology of skin biopsies, Page *et al.* [27] found no consistent microscopic abnormalities in skin samples from three affected foals [27], and also noted no specific macro- or microscopic lesions of the central nervous system which could be attributed to LFS.

A link between LFS and benign epilepsy of Arabian foals has been postulated and warrants further research [25]. The latter is a rare neurological disorder associated with foals of Arabian breeding. Benign epilepsy is characterized by intermittent seizures affecting foals up to the age of twelve months, by which time the condition spontaneously resolves. Seizure activity is variable and may range from abnormal mouth movements to generalized seizures involving muscle spasms, recumbency and thrashing of the limbs [28].

Recent studies have determined the genetic basis of LFS to be a single base-pair deletion in the MYO5A gene (c.4459delC), which codes for the protein myosin-Va. The mutation (p.Arg1487AlafsX13) alters the reading frame of MYO5A and results in the incorporation of a premature stop codon [29] [26].

Myosins form a large superfamily of proteins which are able to bind actin and produce movement via adenosine triphosphate (ATP)-hydrolysis. Myosins consist of three functional domains: the NH_2-terminal motor domain, which interacts with actin and binds ATP; the neck domain, which binds to light chains or calmodulin; and the COOH-terminal tail domain [30]. Possible functions of the myosin tail domain include cargo-binding and localization within the cell [31]. The structure of myosin-V, which exists as a two-headed dimer, differs from other myosin classes by the presence of an extended neck region containing six tandem repeats, termed IQ motifs. These are thought to bind to six light chains or calmodulin. A stalk region, between the neck and globular tail, exhibits a coiled-coil formation [32].

At least 30 classes of myosins have been described based on phylogenetic analyses [33]. The first myosin discovered, myosin II, is known as a “conventional” myosin and is able to form large bipolar filaments which function in muscle contraction. Myosins discovered subsequently were termed “unconventional” myosins and have traditionally been thought to function as short-range transporters of membranous

organelles along actin filaments, in conjunction with kinesins or dyneins, which act as long-range transporters of organelles along microtubules [34]. However, recent research indicates that unconventional myosins may have alternative roles such as tethering of organelles to F-actin in the cell periphery, organisation of actin filaments during endo- and exocytosis and regulation of transcription through influences on spindle function and cell division [33].

The exact roles of myosin-Va are not completely understood. Studies performed on mice which have mutations in MYO5A, and exhibit a dilute coat colour, suggest that the abnormal coat colour occurs as a result of defective transport of pigment granules from melanocytes to the keratinocytes of the hair shaft. Melanosomes therefore accumulate abnormally within melanocytes. In the cerebellum of dilute mice, extension of smooth endoplasmic reticulum into Purkinje cell dendritic spines does not occur as normal, which may account for seizure activity observed in these animals due to disruption of calcium ion homeostasis. Preparations of purified synaptic vesicles have been found to contain myosin-Va, suggesting a role in synaptic vesicle dynamics [31].

The prevalence of LFS in the Arabian horse population of South Africa is unknown [25, 27]. The frequency of carriers of the mutant LFS allele in the Egyptian Arabian and non-Egyptian Arabian populations in the United States of America is estimated to be 10.3% and 1.8% respectively [26].

2.3 Cerebellar Abiotrophy

Cerebellar abiotrophy (CA), a progressive neurological disorder characterised by the degeneration of cerebellar Purkinje cells, has been described in a number of animal species including cats, dogs, cattle and chickens [35]. Affected animals are normal at birth and for a variable period thereafter, following which neurological signs develop. Equine CA is associated with horses of Arabian breeding and was first documented in a pure-bred Arabian colt in 1966 [36]. The putative genetic mutation responsible for equine CA is present at low frequencies in three non-Arabian horse breeds – notably breeds founded or developed using Arabian stock. The potential therefore exists for any breed with Arabian horses in its ancestry to be affected by CA [37].

Clinical signs of CA usually develop between the ages of 6 weeks and 4 months, and include ataxia, hypermetria, intention head tremors and the absence of a menace response [35]. Prior to the development of genetic testing, a preliminary diagnosis of CA was confirmed by post-mortal histopathological examination of the brain. Lesions are limited to the cerebellum, with the most striking abnormality being a severe paucity of Purkinje cells. Where present, Purkinje cells tend to be small and shrunken, with degenerative nuclear changes such as karyolysis and pyknosis. The granular layer neurons are depleted, with a mild reduction in the layer's thickness. The molecular layer exhibits an increase in cellularity and is approximately half to one third of its normal thickness in severely-affected areas. Macroscopically, no obvious reduction in size of the cerebellum is noted [36]. Antemortal magnetic resonance imaging of the head and central nervous system of a CA-affected foal revealed no obvious abnormalities [38].

CA has an autosomal recessive mode of inheritance [39]. The genetic mutation responsible for the disorder is believed to be a single nucleotide polymorphism (SNP) (G → A) located in exon 4 of TOE1 (Target Of Early growth response 1) on equine chromosome two, resulting in the incorporation of arginine instead of histidine at this position [35]. Both arginine and histidine are small, polar, positively-charged amino acids, raising the question of the significance of this SNP on the structure and function of the product of TOE1. TOE1 may function in cell cycle arrest and tumour suppression in association with Early growth response 1 (Egr1) and p53. Anecdotally, mutations in p53 are estimated to contribute to approximately 50% of human cancers [40].
The 5' end of TOE1 overlaps with the 5' end of MUTYH (MutY Homolog) on the complementary strand of DNA. The site of the CA SNP in TOE1 is approximately 1200 base pairs upstream from the 5' end of MUTYH, adjacent to a potential binding site for GATA-2, a transcription factor involved in expression of MUTYH. Using quantitative PCR analysis, cerebellar MUTYH levels were shown to increase in proportion to the age of the animal. In CA-affected individuals, expression of MUTYH increased with age but remained significantly lower than the levels of expression found in normal equines [35].

Oxidative damage to DNA by reactive oxygen species produced during cell metabolism is thought to play a role in degenerative diseases, including brain dysfunction. Seven,8-dihydro-8-oxoguanine (8-oxoG) is produced through oxidative damage to DNA. This molecule is particularly hazardous due to its ability to mispair with adenine, resulting in incorrect base incorporation during DNA replication. MUTYH, a post-replication DNA glycosylase, is able to remove adenines mismatched with guanines or 8-oxoG,

protecting the integrity of the DNA molecule [41]. Four isoforms of MUTYH have been detected, three of which are exclusive to the brain. Expression of MUTYH occurs predominantly in the neurons of the hippocampus, cortex and cerebellum, being particularly abundant in cerebellar Purkinje neurons [42].

Post-mitotic brain neurons are predisposed to oxidative DNA damage due to their longevity and high metabolic activity [43]. The nuclear DNA of these cells no longer replicates. Conversely, mitochondrial genomes replicate continuously, the mean mitochondrial life-span being three to four weeks. Mitochondrial DNA is also prone to oxidative damage since these genomes are not protected by histone proteins and exist in a highly oxidative environment [44]. Studies performed on isoforms of MUTYH in the rat suggest that MUTYH is involved primarily in post-replicative DNA repair in the rat embryo and neonate, when neurons are proliferating rapidly. Once neuron replication ceases, MUTYH localises mainly in the mitochondria, where it functions in repair of mitochondrial DNA damage [45].

An interesting finding by Brault *et al* [35] was the presence of clinically normal individuals with homozygous affected genotypes for CA. This finding may illustrate a large amount of variability in the age of onset and severity of the disease. A recent case report described the late diagnosis of CA in a six year old Arabian mare [46]. It is possible that associated genes or genetic mutations play a role in the clinical manifestation of CA.

Although CA is not fatal, affected individuals are often euthanased due to their inability to perform as riding or sports horses and the potential harm they may cause to themselves and others. Brault and Penedo [47] estimated the carrier frequency of CA to be 19.7%, although it is unclear which Arabian populations were sampled and whether this value represented an unbiased estimate. No formal prevalence study for CA has been performed.

2.4 Laboratory Methodologies

2.4.1 Polymerase chain reaction

Polymerase chain reaction (PCR) was first described by Kary B. Mullis in 1986 [48]. The PCR was a major breakthrough in genetic research due to its ability to produce vast quantities of target DNA. The process

is simple, involving (i) denaturation of double-stranded DNA, (ii) annealing of two oligonucleotide primers bordering the DNA segment of interest, and (iii) extension of a complementary strand of DNA, from each primer, by the enzyme DNA polymerase. The process is repeated in a number of cycles, each cycle resulting in a doubling of the quantity of targeted DNA. The quantity of copied DNA therefore increases exponentially. Early PCR protocols made use of the heat-labile Klenow fragment of *Escherichia coli* DNA polymerase I, which was destroyed by the denaturing step and had to be replenished with each cycle. The heat-stable DNA polymerase of *Thermus aquaticus* was an improvement to this enzyme [49], and remains in use today.

Multiplex PCR involves the addition of more than one pair of primers to the PCR mixture, amplifying two or more target sequences of DNA in one reaction. Less sample DNA and reagents are consumed by performing a multiplex PCR analysis than would be required to perform a PCR for each primer pair individually. The potential for contamination and the amount of labour involved is also minimized [50].

The genetic mutations responsible for SCID and LFS involve base pair deletions. The PCR tests for these diseases make use of forward and reverse primers, flanking the deletion sites. Wild-type and mutant alleles are differentiated based on the size difference of the PCR fragments. For SCID, the PCR fragment representing the mutant allele is 5 base pairs (bp) smaller than the fragment representing the wild-type allele. For LFS, the PCR fragment representing the mutant allele is 1 bp smaller than the fragment representing the wild-type allele.

The genetic mutation responsible for CA is a SNP (G → A). To detect this SNP, two allele-specific forward (sense) primers are used, one complementary to the wild-type allele at the 3'-end (5' – GTA ATG CTG CCA GGA CAC **G**), and the other complementary to the mutant allele (5' – GTA ATG CTG CCA GGA CCG **A**), together with a common reverse (anti-sense) primer. The last nucleotide complements either the wild-type or mutant allele, representing the SNP in question. Two mismatches are also included in the design of each primer at the 3'-end adjacent to the site of the SNP. The wild-type primer is able to bind sufficiently to wild-type allele template DNA, despite mismatches of the penultimate and third last nucleotides, to allow extension of this primer by *Taq* polymerase. In the presence of the mutant allele, mismatches at the 3'-end prevent extension of the wild-type primer. The mutant primer works in a similar manner. Primers designed in this way exhibit the specificity required to provide an accurate

genetic test for CA. The wild-type and mutant allele primers are fluorescently labeled in two different colours in order to distinguish between the two alleles.

2.4.2 Capillary electrophoresis

Electrophoresis involves the separation of charged fragments through the action of an electric field. Capillary electrophoresis (CE) allows automation of the process of separation and detection of fragments of DNA. The narrow capillary dissipates heat efficiently through its high surface-to-volume ratio. This enables the use of high electric fields which shortens processing times. A laser beam excites fluorescent labels attached to DNA fragments. Detection of fluorescence with time forms an electropherogram, consisting of peaks associated with each DNA fragment [51].

Due to the nature of the above genetic tests, their sensitivity and specificity is extremely high, provided that the sample has amplified sufficiently well. The presence of positive controls, showing their expected genotypes, verifies the success of each PCR reaction performed. The negative control should show no peaks on capillary electrophoresis, verifying the absence of gross contamination of the PCR plate. Minor well-to-well contamination is, however, possible. This is minimized by the use of sound laboratory practices during the setting up of each PCR plate, including absolute attention to sample placement and careful, accurate pipetting of samples.

The visualization of a well-defined peak showing the appropriate fragment length demonstrates the presence of the corresponding allele. The visualization of two well-defined peaks, differing in length or fluorescent colour as described above (section 2.4.1), is easily recognized to represent the presence of both the wild-type and mutant allele.

The addition of an extra non-template nucleotide (usually adenine, hence the term “A plus”) to the 5’ end of DNA fragments can result in genotyping errors due to variation in the lengths of DNA fragments produced by PCR, particularly when the artifact occurs inconsistently. The addition of a segment of non-complementary nucleotides to the 5’ end of reverse primers, known as “PIG-tailing”, encourages consistent addition of one extra adenine nucleotide to the DNA fragment [52]. However, in the case of genetic testing for disorders such as SCID, LFS and CA, this artifact is not reported to be problematic.

A possible source of error could be the failure of one allele to amplify, thereby giving the appearance of a homozygous genotype despite the individual being a heterozygote for the disorder. This could occur due to a laboratory error or the occurrence of a null allele, usually caused by a mutation within the primer binding site which prevents amplification of the allele [53]. Although not documented, the impact of dropped or null alleles is believed to be negligible. Alleles with imperfect amplification caused by poor DNA extraction technique or DNA samples of inferior quality may be difficult to read with confidence. This is prevented by the re-extraction or exclusion of samples showing unsatisfactory amplification. The criteria selected for discarding samples showing unsatisfactory amplification included size standard failure and the absence of clearly distinguishable fragment peaks.

2.4.3 Spectrophotometry of DNA extracts

Spectrophotometry is used to characterize the concentration and purity of nucleic acid samples. Analysis of the absorbance of light waves through a DNA sample is used to calculate the concentration of the sample, using the absorbance through pure water or buffer as a reference point. The absorbance spectrum of light at wavelengths of 260 and 280 nm is compared, the ratio between the two (260/280) providing an indication of the purity of a sample of nucleic acids. A 260/280 absorbance ratio of 1.8 – 2.0 indicates satisfactory purity of the DNA [54]. In this study, DNA extracts from the 2004 breeding season which failed to amplify on initial PCR, and which were suspected of being of inferior quality, underwent spectrophotometry using the NanoDrop 1000 version 3.8.1 (Thermo Fisher Scientific, United States of America). Samples found to have a 260/280 absorbance ratio of below 1.5 were excluded from further PCR analysis. Original sample materials were re-extracted for these individuals.

2.5 Statistical methodologies: The hypergeometric distribution

When randomly sampling individuals from a population with a given disease prevalence, the probability of sampling a diseased individual is equal to the prevalence of the disease in the population. If sampling is done with replacement, or the population size is infinite, then the probability of sampling a diseased individual remains constant with repeated sampling and the number of diseased individuals detected in

a given sample size follows a binomial distribution. However, if sampling is done without replacement and the population size is finite, then the probability of selecting a diseased individual changes with each successive individual sampled and the number of diseased individuals follows a hypergeometric distribution. Statistical inferences (confidence intervals or hypothesis testing) regarding the prevalence of disease can be made using either the binomial or the hypergeometric distribution, and results will be very similar if the population size is large and only a very small proportion of the total population is sampled. However, in finite populations, and particularly when the proportion sampled is not vastly smaller than the total population, it is preferable to use hypergeometric-based methods. These take into account the population size and will produce more accurate confidence intervals and more power to detect differences between proportions. Sample size calculation, confidence interval calculation and statistical significance testing for differences between proportions can be done using StatCalc version 2.0 [55].

CHAPTER 3: Materials and Methods

3.1 Materials

The DNA profiling of Arabian horses facilitates parentage verification which is a prerequisite for registration with the Arab Horse Society of South Africa (AHSSA). Arabian horse owners and breeders submit hair, blood or tissue samples to the Veterinary Genetics Laboratory (VGL) (Faculty of Veterinary Science, University of Pretoria) routinely, for this purpose. The VGL is one of the primary laboratories responsible for parentage verification of Arabian horses in South Africa (C Harper, Veterinary Genetics Laboratory, Onderstepoort, pers. comm., 2011).

All samples received at the VGL undergo a DNA extraction process. Both the original sample and the extracted DNA sample are catalogued and stored indefinitely. Hair samples are stored at room temperature. Blood is stored in the original collection tube at -20°C or on FTA™ paper. Tissue and extracted DNA samples are stored at -20°C.

This study made use of DNA extracts as well as tissue, blood and hair samples in storage at the VGL.

3.2 Experimental design

3.2.1 Selection of samples

This study utilised two target populations: purebred Arabian horses born in South Africa between 1 August 2004 and 31 July 2005 (2004 breeding season); and purebred Arabian horses born in South Africa between 1 August 2009 and 31 July 2010 (2009 breeding season). Lists of individuals born in South Africa during these periods were provided by the AHSSA, representing the study populations. A random sample of individuals from each breeding season was screened for SCID, LFS and CA, in order to limit the number of individuals tested.

Details including horse name, stud prefix, birth date, VGL sample number and VGL extraction number were entered into a Microsoft Access database. Access to this database was limited to the researcher and respective supervisors in order to protect the confidentiality of results.

The sample size required from each population was estimated using the prevalence results of a pilot study of 112 individuals from the 2009 breeding season. The pilot study was arbitrarily intended to include a minimum of 100 individuals. The sampling frame was randomly sorted and the first 150 foals on the list were selected for testing. This allowed for sample failures during PCR and sequencing which was expected to be approximately 30%.

Preliminary prevalence results for SCID, LFS and CA from the pilot study were then used to calculate the final sample size required to estimate the population proportions, within 95% confidence intervals at a 5% allowable error. This was calculated using the software programme StatCalc. Further individuals from the 2009 breeding season as well as the required number of individuals from the 2004 breeding season were then tested in order to reach or exceed the required sample size.

Samples from the 2009 breeding season which failed to amplify on initial PCR underwent a second PCR. Due to the increased age and potential degradation of samples from the 2004 breeding season, samples which failed to amplify on initial PCR underwent spectrophotometry using the NanoDrop 1000 Version 3.8.1 (Thermo Fisher Scientific, United States of America). These samples were all shown to be of an unsatisfactory purity (260/280 absorbance ratio below 1.5) and were therefore re-extracted from the original hair or blood sample.

Any DNA extract which could not be located or failed to amplify on PCR within a maximum of 2 attempts was replaced by selection of the next consecutive sample along the list of random numbers, until the required number of horses was genotyped.

3.2.2 Controls

Template and non-template controls were included in each PCR performed. The non-template (or “negative”) control well contained all the required PCR components with the

deliberate omission of sample DNA. Template (or "positive") controls used were a Thoroughbred horse which was homozygous normal for all three disorders, an Arabian horse known to be a carrier of both SCID and LFS, a CA-carrier and a CA-affected foal. Although a Thoroughbred horse was used as a homozygous normal control, any Arabian horse known to be homozygous normal for all three disorders would have been suitable. More stringent control checks were placed on the markers for CA due to this disorder not having been tested for previously in the VGL, whereas the genetic tests for SCID and LFS had been in use for some time and had been previously validated.

The positive controls utilized in this study are described in Table 1, which shows the control ID, species, breed, original sample type and disease marker genotype. Each positive control sample was extracted from the original sample for use in this project.

TABLE 1: List of positive control samples included in each PCR performed showing control identification, breed, original sample type and disease marker genotypes.

Control identification	**Breed**	**Original sample type**	**Disease marker genotype**		
			SCID	**LFS**	**CA**
Disease negative	Thoroughbred	Blood	WW	NN	NN
SCID and LFS carrier	Arabian	Blood	DW	LN	NN
CA carrier	Arabian	Blood	WW	NN	CN
CA affected	Arabian	Tissue	WW	NN	CC

3.2.3 DNA extraction methods

a) DNA extraction from hair samples

The DNA extraction protocol for hair samples involved digestion of the hair with NaOH to release the DNA. Approximately six hairs were cut as close to the root as possible and collected in a 1.5 ml Eppendorf tube. The roots were incubated in 200 mM NaOH (Sigma-Aldrich, United States of America) at 97°C for 15 minutes. The solution was then neutralized by the addition of 200 mM HCl (Merck, South Africa) and 100 mM Tris-HCl (Promega, United States of America; Merck, South Africa). The extracted DNA was immediately ready for use in PCR.

b) DNA extraction from blood samples

For extraction of DNA from blood, blood was collected in ethylenediaminetetraacetic acid (EDTA) blood collection tubes. The red blood cells were lysed using 10mM NaCl (Promega, United States of America) and 10 mM EDTA (Merck, South Africa) at a pH of 7. This was followed by centrifugation at 10 000 revolutions per minute for 2 minutes. The supernatant was discarded leaving behind a proteinaceous pellet containing white blood cells. The cells were lysed by incubation at 56°C for 2 hours in a solution containing 10 mM Tris-HCl (pH 8), 100 µM EDTA, 50 mM NaCl, 1% sodium dodecyl sulphate (SDS) (BDH Laboratory Supplies, United Kingdom) and 0.11 mg/ml proteinase-K (Sigma-Aldrich, United States of America). The samples were heated to 94°C for 10 minutes to inactivate the proteinase-K. DNA was extracted with phenol:chloroform:isoamyl alcohol 25:24:1 saturated with 10mM Tris, ph 8.0, 1mM EDTA (PCIA;Sigma-Aldrich, United States of America). The phenol component removed proteinaceous material and chloroform separated the phenol from the aqueous phase containing DNA. Cold absolute ethanol (Merck, South Africa) and centrifugation was used to precipitate the DNA. The DNA pellet was washed in 70% ethanol before it was rehydrated in Tris-EDTA (TE) buffer (Gentra, United States of America).

c) DNA extraction from tissue samples

The method of extraction for tissue samples involved overnight digestion of approximately 50 mg of finely chopped tissue at 56°C in a solution containing 10 mM Tris-HCl (pH 8), 100

µM EDTA, 1% sodium dodecyl sulphate (SDS) and 0.11 mg/ml proteinase-K. The samples were heated to 94°C for 10 minutes, followed by the addition of PCIA. As for blood samples, this was followed by the addition of cold absolute ethanol, centrifugation and finally washing in 70% ethanol before rehydration in TE buffer.

d) DNA extraction from FTA™ paper

FTA™ cards (Whatman Ltd, United Kingdom) were previously used by the VGL for the long term storage of received blood samples. FTA™ paper is produced by the addition of specific reagents to filter paper which lyse cells and protect nucleic acids from bacterial growth and environmental degradation. Extraction of DNA from FTA™ paper essentially aims to remove all chemicals and waste products from the paper through washing, leaving only DNA attached. DNA was extracted from discs of FTA™ paper 2 mm in diameter containing a dried blood sample. The discs were washed in a solution containing 100mM Tris base (Promega, United States of America) and 0.1% SDS. The samples were gently agitated in solution for 30 minutes. Samples were then washed for 30 minutes in a second solution containing 5 M guanidine thiocyanate (Sigma-Aldrich, United States of America). This was followed by 3 washes in distilled water and a final wash in 95% ethanol, each for 5 minutes. The discs were air-dried for 2 hours, followed by the addition of distilled water and heating to 90°C for 10 minutes to release the DNA.

3.2.4 Polymerase chain reaction

To determine the SCID, LFS and CA genotypes of each individual in the sample sets, polymerase chain reaction (PCR) was performed using two panels. Panel A contained primers for CA as described by Brault and Penedo (2011). Panel B consisted of a multiplex-PCR panel, including primers for SCID [7], LFS [29] and five polymorphic markers chosen from markers recommended by the International Society of Animal Genetics (ISAG) and used routinely by the VGL for identification and parentage verification of horses. Primers were obtained from Life Technologies (South Africa) and were rehydrated, aliquoted and stored at -84°C until use. Table 2 shows the primers used, their marker identification and fluorescent label, the fragment sizes produced by each primer, the sequence of each forward and reverse primer, and the primer concentration utilised in PCR.

TABLE 2: List of primers used in panels A and B including primer label, fluorescent dye, fragment size range, sequences of forward and reverse primers (5' – 3') and the primer concentration utilised in PCR.

Panel	Marker identification	Flourescent label	Fragment sizes or range	Forward primer sequence (5' – 3')	Reverse primer sequence (5' – 3')	Primer concentration (mg/µl)
A	CA_N	Vic	80	GTAATGCTGCCAGGACACG	CTCATACCTTGTCTGGCTGCT	0.15
A	CA_CA	Fam	80	GTAATGCTGCCAGGACCGA	CTCATACCTTGTCTGGCTGCT	0.7
B	SCID	Pet	169 / 174	AAGTTGGTCTTGTCATTGAGC	TTTGTGATGATGTCATCCCAG	0.15
B	ACDF (LFS)	Pet	159 / 160	AGAATGAGGCTGAAGCCCTC	GTGATCTCATGCTGCAGGCT	0.23
B	HMS3	Ned	256 – 282	AGTGCAACCCCAAACATCAG	GCCACCTCACTCCACTATAA	0.14
B	LEX3	Fam	202 – 246	AGTGCTGAGACTTCTGAGAG	ATTAGGCAACGGTCAGAAGG	0.5
B	HMS2	Ned	209 – 239	TGCTAAAAGCTTGCAGTCGA	AAGACACACGGTGGCAACTG	0.2
B	HTG10	Ned	85 – 119	CGCCCCCACACTCCATAAAT	AGTGACTTATTGTGGCGA	0.33
B	ASB2	Pet	223 – 261	GTGTCGTTTCAGAAGGTCAACC	TCTCTTTGCGCACTTCCCAG	0.55

Primer mixes for each panel were made up separately. PCR reactions were carried out in a total volume of 11 µl per sample consisting of 0.4 µl primer mix, 3.6 µl molecular grade water, 5 µl Kapa 2G Robust HotStart ReadyMix (Kapa Biosystems, South Africa) and 2 µl of DNA extract at a concentration of approximately 100 ng/µl.

Panels A and B were run on two different PCR programmes, each on a 9800 Fast Thermal Cycler (Life Technologies, South Africa).

Panel A was run according to the PCR protocol described by Brault and Penedo (2011). Panel B required an initial activation step of 95°C for 5 minutes, followed by 35 cycles of 95°C for 45 seconds, 56°C for 15 seconds and 72°C for 15 seconds. A final elongation step of 72°C for 10 minutes was followed by indefinite holding of the samples at 4°C.

Positive and negative controls were included in each sample tray which underwent PCR. Following PCR, the products of Panels A and B for each sample were combined.

3.2.5 Capillary electrophoresis

For each reaction, 10 µl of Hi-Di™ Formamide (Life Technologies, South Africa) were combined with 0.25 µl of GeneScan™ - 500 LIZ™ Size Standard (Life Technologies, South Africa) and 1 µl of the combined PCR products of one sample. Following denaturation of the samples at 95°C for 2 minutes, capillary electrophoresis was performed using the ABI 3130 *XL* Genetic Analyzer (Life Technologies, South Africa), with an injection time of 15 seconds and a voltage of 15 kV.

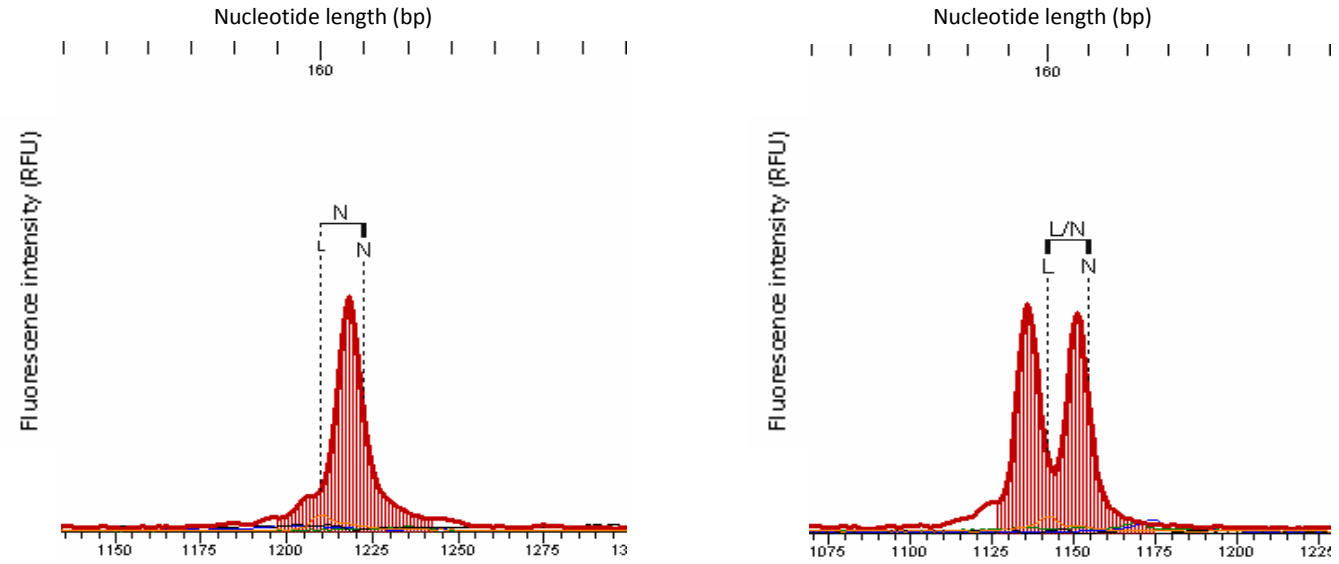
3.2.6 Genotype evaluation

Fragment sizes produced for each marker were recorded and analysed on the software programme STRand version 2.4.49 (University of California, Davis, United States of America) [56]. The SCID, LFS and CA genotype of each horse in the sample set was read and entered into the relevant Access database. Any sample which failed to show clear genotype distinction for all three disorders was excluded from the database. This was done in order to

minimise the possibility of reading errors as samples which failed to amplify for one disorder tended to show poor amplification for all three disorder markers. Results were read and recorded on two separate occasions for comparison purposes in order to minimise errors.

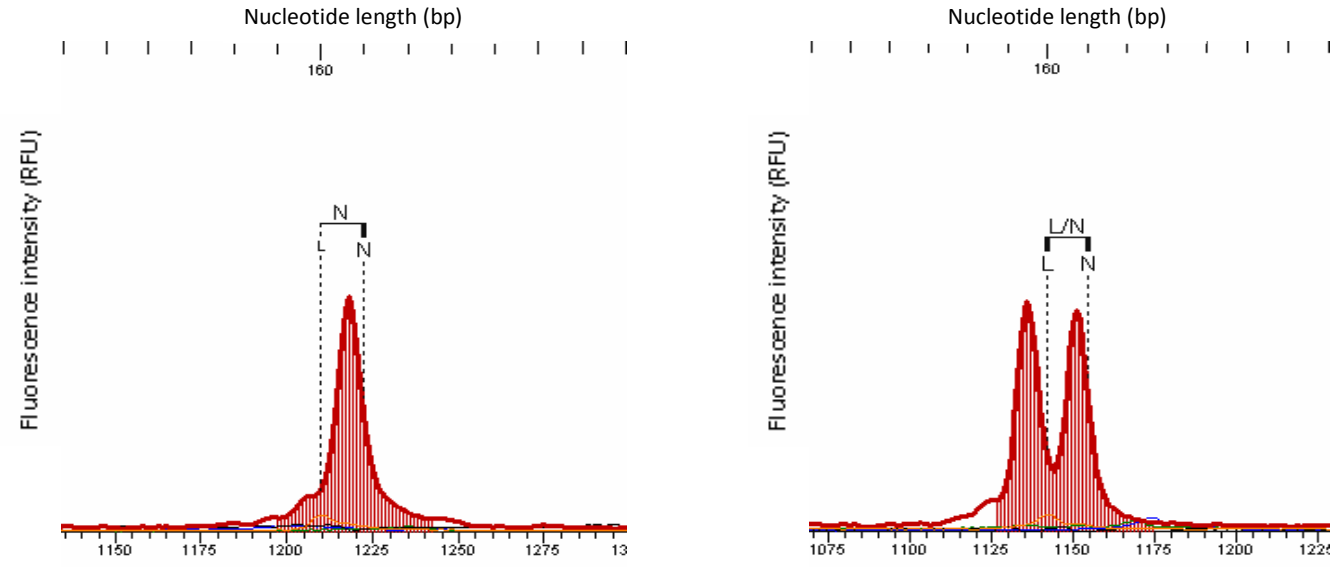
For SCID, the annotation WW was assigned to homozygous normal individuals while DW indicated a carrier of the mutant SCID allele. For LFS, NN indicated a homozygous normal individual, LN representing a carrier of the mutant LFS allele. Similarly, for CA a genotype of NN was used to represent homozygous normal individuals, CN representing a carrier of the CA mutant allele. DD, LL and CC would represent homozygous affected individuals for SCID, LFS and CA respectively, although only CC was seen in the CA-affected control.

The abbreviations used to denote SCID and LFS genotypes are currently in use in the VGL. The abbreviations NN, CN and CC were chosen to denote CA genotypes in order to provide clear distinction from the genotypes for SCID and LFS. Figures 1 to 3 illustrate electropherograms detected during the course of this study, representing homozygous normal and heterozygous carrier genotypes for SCID and LFS, as well as homozygous normal, heterozygous carriers and homozygous affected genotypes for CA. No SCID- or LFS-affected individuals were detected during the course of this study; hence no electropherogram images for these genotypes are shown. The upper x-axis of the electropherogram corresponds to nucleotide length (of a fragment produced during PCR) in base pairs (bp). The lower x-axis corresponds to the scan number of the instrument used. The y-axis corresponds to the intensity of fluorescence, measured in relative fluorescent units (RFU).



SCID (WW) SCID (DW)

FIGURE 1 Electropherogram images showing the possible genotypes detected for severe combined immunodeficiency during this study, where WW represents a homozygous normal individual and DW represents a heterozygous carrier of the mutant allele.



LFS (NN) LFS (LN)

FIGURE 2 Electropherogram images showing the possible genotypes detected for lavender foal syndrome, where NN represents a homozygous normal individual and LN represents a heterozygous carrier of the mutant allele.

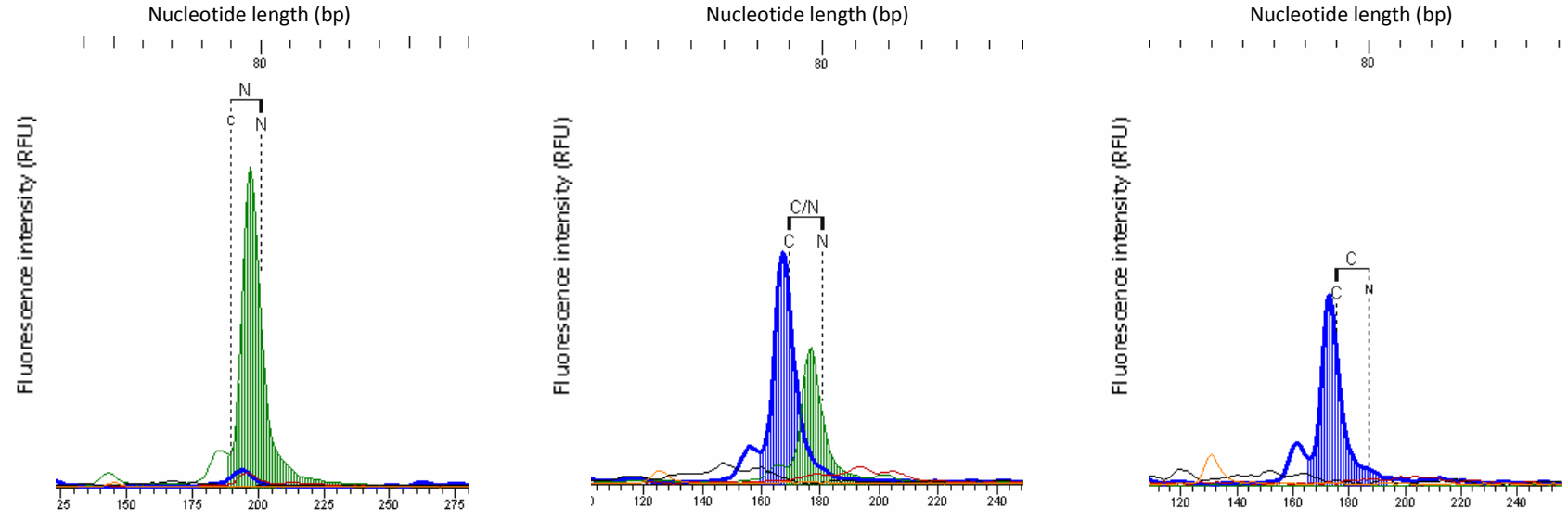


FIGURE 3 Electropherogram images showing the possible genotypes detected for cerebellar abiotrophy, where NN represents a homozygous normal individual, CN represents a heterozygous carrier of the mutant allele and CC represents a CA-affected individual.

3.2.7 Additional data collection for severe combined immunodeficiency

Testing for SCID has been performed routinely by the VGL on most Arabians profiled since 2004. In addition to the PCR testing performed above, all available SCID genotypes for members of the two populations were retrieved from the VGL database. All results obtained in this study through the course of testing using multiplex panel B were checked against SCID results for these individuals on the VGL database, where available. Not all horses selected for testing in this study had a SCID result available on the VGL database, hence the need for the inclusion of SCID within the multiplex panel.

Identification of individuals in the dataset was based on the stud prefix and horse's name, or the horse's name and birth date. The additional data provided by the VGL were used to calculate a more accurate estimate of the prevalence of SCID for each population and to compare the prevalence between the two populations (see below).

3.2.8 Statistical analysis of results

The prevalence of carriers of SCID, LFS, and CA were calculated, along with exact hypergeometric 95% confidence intervals using StatCalc, for each of the two foal populations. For each disease, the prevalence of carriers amongst Arabians born during the 2004 breeding season was compared with that amongst Arabians born during the 2009 breeding season using a two-tailed *E* test based on estimated *P*-values for the comparison of proportions in two finite populations [57], as implemented in StatCalc. A *P*-value of less than 0.05 was considered statistically significant.

CHAPTER 4: Results

4.1 Preliminary results: pilot study

Prevalence results from an initial study of 112 foals from the 2009 breeding season are shown in Table 3. These were used in order to calculate the final sample size requirement for the study. Table 3 shows the sample size required in order to estimate the prevalence of heterozygotes for each disorder with 95% confidence and with an allowable error (half-width of exact hypergeometric 95% confidence interval) of 5%, calculated using the software programme StatCalc.

TABLE 3: Preliminary prevalence results for SCID, LFS and CA showing the number of individuals initially tested, the number of results obtained, the proportion of heterozygotes detected (prevalence) as well as the 95% confidence intervals (CI) and the sample size required in order to achieve the required precision (CI half-width of 5%).

Genetic disorder	SCID	LFS	CA
Total population size	898	898	898
Number of individuals tested	150	150	150
Number of results obtained	112	112	112
Number of heterozygotes	5	15	5
Prevalence (%)	4.5	13.4	4.5
95% CI* (%)	1.6; 9.8	7.9; 20.7	1.6; 9.8
Required sample size for CI half-width of 5%	77	178	77

* Exact hypergeometric 95% confidence interval

The final sample size required in order to estimate the carrier prevalence for all three disorders at the required precision was therefore a minimum of 178 individuals per breeding season.

4.2 Final results

4.2.1 Sample size and sampling fractions

4.2.1.1 2004 breeding season

Samples tested in this population group included DNA extracts and blood stored on FTA™ paper. A total of 88 FTA™ paper extractions were performed. Due to the age of stored DNA extracts and deterioration in their quality, as shown on spectrophotometry, samples which failed to amplify on initial PCR were re-extracted from the original sample received. A total of 33 hair samples and 1 blood sample was extracted. No further PCR testing or re-extractions were performed once the required number of sample results had been reached. Sampling details for individuals born during the 2004 breeding season are shown in Table 4. Reasons for "non-response" included individuals not profiled through the VGL and therefore not found on the VGL database, discarded DNA extracts or original samples and extracts which failed to amplify on PCR.

TABLE 4: Sampling details for the 2004 breeding season population subset.

Total number of Arabians born in South Africa in 2004	826
Total number of individuals selected for analysis	247
Number of sample results obtained	203
"Non-response" rate	17.8%
Sampling fraction	24.6%
Total number of SCID results available from VGL database	800
Sampling fraction for SCID	96.9%

4.2.1.2 2009 breeding season

All samples tested from this population group were DNA extracts held in storage at the VGL and extracted previously upon receipt of the original sample material. No original samples underwent re-extraction. Sampling details for the 2009 breeding season population subset are shown in Table 5. Reasons for "non-response", similarly, included individuals not found on the VGL database, discarded DNA extracts and extracts which failed to amplify on PCR.

TABLE 5: Sampling details for the 2009 breeding season population subset.

Total number of Arabians born in South Africa in 2009	898
Total number of individuals selected for analysis	275
Number of sample results obtained	197
"Non-response" rate	28.4%
Sampling fraction	21.9%
Total number of SCID results available from VGL database	699
Sampling fraction for SCID	77.8%

4.2.2 Statistical analysis of results

4.2.2.1 Severe combined immunodeficiency

The data for SCID determined through testing during the course of this study, as the multiplex panel B, as well as data generated by the VGL during the routine profiling of Arabians, are shown in Table 6. Exact hypergeometric 95% confidence interval and results of a hypergeometric-based two-tailed test for differences between proportions, using the software programme Statcalc, are also shown. All SCID genotypes determined during this research project were compared to results in the VGL database. No discrepancies were observed.

TABLE 6: Prevalence of severe combined immunodeficiency carriers amongst Arabian horses born during the 2004 and 2009 breeding seasons, showing data generated during the current research project as well as data generated routinely by the Veterinary Genetics Laboratory (VGL), Onderstepoort

	Data generated during current study		**Data generated during routine profiling of Arabian horses by the VGL**	
Breeding season	**2004**	**2009**	**2004**	**2009**
Total population size	826	898	826	898
Number of individuals tested	203	197	800	699
Observed heterozygotes	14	10	51	24
Proportion of heterozygotes (p) (%)	6.9	5.1	6.4	3.4
95% CI* (%)	4.00; 10.65	2.56; 8.69	6.05; 6.78	2.67; 4.23
P-value**	0.38		<0.0001	

* Exact hypergeometric 95% confidence interval

**Where H_0: $p_{2004} = p_{2009}$ and H_a: $p_{2004} \neq p_{2009}$

Testing of a random sample of 203 horses from the 2004 breeding season estimated the prevalence to be 6.9% (95% confidence interval [CI] 4.0 - 10.7). In comparison, results of the testing, previously, of 800 individuals from the same population increased the precision of the estimated prevalence (6.4%) to between 6.1 and 6.8% at the 95% confidence level.

Testing of a random sample of 197 individuals from the 2009 breeding season generated an estimated prevalence of 5.1% (95% CI 2.6 - 8.7). The testing of 699 individuals from the same population group provided an estimated prevalence of 3.4% (95% CI 2.7 - 4.2).

Based on these results, the prevalence of carriers of SCID amongst Arabian foals has decreased between the 2004 breeding season and the 2009 breeding season, this decrease being statistically significant ($P < 0.0001$).

4.2.2.2 Lavender foal syndrome

The genotyping results for LFS and the statistical comparison of the proportion of carriers of LFS among Arabians born during the 2004 and 2009 breeding seasons are shown in Table 7.

TABLE 7: Prevalence of lavender foal syndrome carriers amongst Arabian foals born in South Africa during the 2004 and 2009 breeding seasons.

Breeding season	2004	2009
Total population size	826	898
Sample size	203	197
Observed heterozygotes	27	23
Proportion of heterozygotes (p) (%)	13.3	11.7
95% CI* (%)	9.32; 18.04	7.80; 16.37
P-value**	0.58	

* Exact hypergeometric 95% confidence interval

**Where H_0: $p_{2004} = p_{2009}$ and H_a: $p_{2004} \neq p_{2009}$

The prevalence of LFS in foals born during the 2004 breeding season was found to be 13.3% (95% CI 9.3 - 18.0). The prevalence of LFS in foals born during the 2009 breeding season was found to be 11.7% (95% CI 7.8 - 16.4). There was no statistically significant difference in the prevalence of LFS between the two breeding seasons (*P* = 0.58).

4.2.2.3 Cerebellar abiotrophy

The genotyping results for CA and the statistical comparison of the proportion of carriers of CA among Arabians born during the 2004 and 2009 breeding seasons are shown in Table 8.

TABLE 8: Prevalence of cerebellar abiotrophy carriers amongst Arabian foals born in South Africa during the 2004 and 2009 breeding seasons.

Breeding season	2004	2009
Total population size	826	898
Sample size	203	197
Observed heterozygotes	10	10
Proportion of heterozygotes (p) (%)	4.9	5.1
95% CI* (%)	2.54; 8.35	2.56; 8.69
P-value**	0.95	

* Exact hypergeometric 95% confidence interval

**Where H_0: $p_{2004} = p_{2009}$ and H_a: $p_{2004} \neq p_{2009}$

The prevalence of CA in foals born during the 2004 breeding season was found to be 4.9 (95% CI 2.5 - 8.4). The prevalence of CA in foals born during the 2009 breeding season was found to be 5.1 (95% CI 2.6 - 8.7). There was no statistically significant difference in the prevalence of CA between the two breeding seasons (*P* = 0.95)

CHAPTER 5: Discussion

This study utilised two populations of Arabian foals, born within limits defined by physiological breeding seasons. These time frames were selected in order to minimize the risk of one mare producing two foals within one breeding season, thereby contributing her genotype to the dataset twice. Populations of foals born over the period of one year were chosen in order to provide a satisfactory representation of all breeding, purebred Arabian horses in South Africa. The use of two sampling frames allowed for comparisons to be made of the prevalence of each disease between the two time periods.

The DNA profiling and parentage verification of Arabian horses is a prerequisite for registration with the Arab Horse Society of South Africa (AHSSA). Arabian horse owners and breeders submit hair, blood or tissue samples routinely for this purpose. The VGL is the primary laboratory responsible for parentage verification of Arabian horses in South Africa (C Harper, Veterinary Genetics Laboratory, Onderstepoort, pers. comm., 2011). This statement was confirmed during this study, as 95.3% and 77.8% of Arabians born during the 2004 and 2009 breeding seasons respectively were found to have been profiled by the VGL. The difference in these two values may be due to the delay between the birth of foals and DNA profiling. This study found that a foal may be profiled up to seven years after birth in South Africa. It is therefore likely that more Arabians from the 2009 breeding season will be profiled by the VGL over time.

All relevant individuals listed with the AHSSA underwent random selection. Reasons for non-response included individuals not matched to the VGL database, discarded samples and samples which failed to amplify on PCR. Individuals profiled by a competing laboratory, individuals for which profiling had not yet been requested from the VGL and name or birth date mismatches due to typing errors or omissions could result in a horse not being detected on the VGL database. Failure to amplify on PCR occurred primarily as a result of deterioration in the quality of a DNA extract, due to degeneration of DNA over time.

The carrier status of an individual is unlikely to affect the probability of inclusion or exclusion of that individual from the study population, as carrier status is only detected through genetic testing. The external validity of this study is therefore likely to be good and should not have been a source of bias. Similarly, the carrier status of an individual is unlikely to

affect the probability of non-response of that individual. The internal validity of this study is therefore likely to be good and should not have been a source of bias.

Although 5 microsatellite markers were included in panel B, they were not essential to this study. These markers are useful to confirm the identity of an individual when required, such as during routine testing for SCID and LFS by the VGL. These could be used in subsequent studies in which the exact identification of an individual must be confirmed.

The current prevalence of carriers of SCID was found to be 3.4% (95% CI: 2.67 - 4.23). Table 9 lists the prevalence of carriers of SCID in the United States of America, the United Kingdom, Morocco, Brazil, Poland, Slovenia and Iran.

TABLE 9: The prevalence of carriers of SCID among Arabian horses in the United States of America, the United Kingdom, Morocco, Brazil, Poland, Slovenia and Iran.

Country	**Prevalence of SCID carriers**	**Reference**
United States of America	8.4%	[21]
United Kingdom	1 – 5%*	[7]
Morocco	7.0%	[22]
Brazil	1.5%	[22]
Poland	0.0%	[22]
Slovenia	0.0%	[23]
Iran	0.0%	[24]

* 95% CI

The decrease in the prevalence of SCID since the initiation of a test-and-breed policy was shown to be statistically significant ($P < 0.0001$). Any lethal recessive allele is likely to decrease slowly in prevalence over time, if the presence of the allele confers neither increased nor decreased fitness for breeding, and assuming random mating and an infinite population size. If a disease allele is associated with factors which decrease fitness of the heterozygote for breeding, the prevalence of that allele would be expected to drop more rapidly. Should an allele be associated with factors which increase a carrier's fitness for

breeding, that allele would be expected to increase in prevalence over time, eventually reaching a stable equilibrium [58]. In this case, the ability to identify a carrier based on genetic testing may decrease that individual's fitness for breeding, depending on the breeder's decision based on the test result. The decrease in the prevalence of carriers of SCID may have been accelerated by breeders opting to exclude carrier individuals from breeding programmes completely. While this hastens removal of a disease allele from the population, it can result in an increase in inbreeding due to a decrease in the size of the available breeding population. This carries the risk of the removal of desired traits from the population as well as the surfacing of further deleterious recessive alleles. A study by Ding *et al* [18] found an increase in the risk for sarcoids in carriers for SCID. Further study is required to determine positive and negative associations with carrier status for all three genetic disorders described here.

Assuming random mating and an infinite population size, and approximating the current prevalence of SCID to 3.5%, approximately 3 SCID-affected foals would be expected per 10000 conceptions. However, Arabian breeders in South Africa are well aware of the disease and the availability of a genetic test, therefore fewer cases of SCID would be expected.

The current prevalence of carriers of LFS in Arabian horses in South Africa was found to be 11.7% (95% CI 7.8 - 16.4). Estimates of the prevalence of carriers of LFS in the United States of America were 10.3% in Egyptian Arabians and 1.8% in non-Egyptian Arabians. The prevalence of LFS in South Africa was found to be higher than expected. This could indicate the strong influence of Egyptian Arabian bloodlines on breeding in South Africa. It could be expected that approximately one in ten Arabian horses in South Africa carries the mutant allele for LFS. Assuming random mating and an infinite population size, and approximating the prevalence to 12%, approximately 4 LFS-affected foals would be expected per 1 000 conceptions. However, within particular breeding lines, the prevalence of LFS could vary significantly.

The current prevalence of carriers of CA in Arabian horses in South Africa was found to be 5.1% (95% CI 2.6 - 8.7). This is lower than the prevalence of CA in the United States of America, estimated to be 19.7% by Brault *et al* [47]. Unlike the study done by Brault *et al,* no homozygous affected individuals were detected during the course of this study. Assuming random mating and an infinite population size, approximately 6 CA-affected foals would be

expected per 10 000 conceptions. As for LFS, this could vary substantially within particular breeding lines.

There was no statistical evidence of a change in prevalence of carriers of LFS or CA between 2004 and 2009. This finding may have remained true should we have been able to test a larger proportion of the study population. Without the availability of genetic testing, no selective breeding based on an individual's genotype for LFS or CA could be carried out during this period as disease alleles are effectively "hidden" in the heterozygote form. The birth of an affected foal is the only indication of the carrier status of the foal's dam and sire. The decrease in prevalence of SCID over this time frame was found to be statistically significant ($P < 0.0001$). Selective breeding based on an individual's genotype for SCID was carried out during this period due to the availability of a genetic test for SCID. Therefore, genetic testing is a valuable tool, both in terms of preventing the birth of affected foals and in achieving a decrease in the frequency of disease alleles in a population.

No homozygous affected individuals were detected during the course of this study. The detection of LFS and SCID-affected individuals was deemed highly unlikely as these foals are likely to perish prior to DNA profiling and registration with the AHSSA. Similarly, most CA affected foals would be found to be unsuitable for sport and breeding purposes within the first 6 months of life, and would therefore not be registered with the AHSSA. Breeders are not forced to register all foals born to registered mares (H Labuschagne, South African Stud Book, pers. comm., 2012). All horses excluded from registration are likely to be horses which are considered to fall below breed standards or which perish prior to registration. Although this includes individuals affected with SCID, LFS or CA, the impact of homozygous-affected individuals on the results of this study is likely to be marginal due to the low incidence of each disease.

Assuming that a mutant allele arose originally by chance in one individual, the prevalence of carriers for all three disorders, world-wide, has increased over time. A study by Leroy and Bemaung [59] investigated the dissemination and purging of deleterious alleles in domestic animal populations using simulations of common breeding practices. This study, based on several dog breeds, found that the use of popular sires increased the risk of dissemination of genetic disorders, while line breeding and the breeding of closely related individuals tended to decrease the risk of dissemination of genetic disorders. Further study is required to

investigate the origin and dissemination of these three genetic disorders in South Africa. This includes the impact of influential sires used during the development of the Arabian breed in South Africa and the prevalence of these disorders among specific breeding lines.

CHAPTER 6: Conclusion

The objectives of the current study, namely to determine the prevalence of carriers of SCID, LFS and CA in pure-bred, registered Arabians in South Africa and to assess any change in the prevalence of carriers of each disease between 2004 and 2009, were successfully addressed.

The results of this study help to quantify the impact of SCID, LFS and CA on Arabian horses in South Africa and will be used to demonstrate the importance of genetic testing in order to prevent the birth of affected foals. This will be done by making these results available to breeders of Arabian horses.

No statistically significant difference in the prevalence of carriers of LFS or CA for the two breeding seasons could be shown. Aided by the availability of a larger sample size, the prevalence of carriers of SCID was shown to have decreased significantly between 2004 and 2009, during which time a genetic test for this disorder was available. This result highlights the value of genetic testing as a tool to decrease the load of deleterious alleles in a population and reduce the prevalence of carriers. This information will be used in the future to promote the genetic testing of Arabian horses as well as other domestic animal species.

References

1. McGuire, T.C. and Poppie, M.J. (1973) Hypogammaglobulinemia and thymic hypoplasia in horses: a primary combined immunodeficiency disorder. *Infect. Immun.* **8,** 272-277.

2. Perryman, L.E. (2004) Molecular pathology of severe combined immunodeficiency in mice, horses, and dogs. *Vet. Pathol.* **41,** 95-100.

3. Bell, T.G., Butler, K.L., Stickle, J.E., Ramos-Vara, J.A. and Dark, M.J. (2002) Autosomal recessive severe combined immunodeficiency of Jack Russell terriers. *J Vet Diagn Invest.* **14,** 194-204.

4. Jezyk, P.F., Felsburg, P.J., Haskins, M.E. and Patterson, D.F. (1989) X-linked severe combined immunodeficiency in the dog. *Clin. Immunol. Immunopathol.* **52,** 173-189.

5. Somberg, R.L., Pullen, R.P., Casal, M.L., Patterson, D.F., Felsburg, P.J. and Henthorn, P.S. (1995) A single nucleotide insertion in the canine interleukin-2 receptor gamma chain results in X-linked severe combined immunodeficiency disease. *Vet. Immunol. Immunopathol.* **47,** 203-213.

6. Verfuurden B., Wempe F., Reinink P., van Kooten P.J.S., Martens E., Gerritsen R., Vos J.H., Rutten V.P.M.G. and Leegwater P.A. (2011) Severe combined immunodeficiency in Frisian Water Dogs caused by a *RAG1* mutation. *Genes and Immunity.* **12,** 310-313.

7. Swinburne J., Lockhart L., Scott M. and Binns M.M. (1999) Estimation of the prevalence of severe combined immunodeficiency disease in UK Arab horses as determined by a DNA-based test. *The Veterinary Record.* **145,** 22-23.

8. Perryman, L.E., Boreson, C.R., Conaway, M.W. and Bartsch, R.C. (1984) Combined immunodeficiency in an Appaloosa foal. *Vet. Pathol.* **21,** 547-548.

9. Larson, J., Buechner-Maxwell, V., Crisman, M.V., LeRoith, T. and Witonsky, S. (2011) Severe combined immunodeficiency in a Caspian filly. *J. Vet. Intern. Med.* **25,** 954-958.

10. Finno, C.J., Spier, S.J. and Valberg, S.J. (2009) Equine diseases caused by known genetic mutations. *Vet. J.* **179,** 336-347.

11. Poppie, M.J. and McGuire, T.C. (1977) Combined immunodeficiency in foals in Arabian breeding: evaluation of mode of inheritance and estimation of prevalence of affected foals and carrier mares and stallions. *J. Am. Vet. Med. Assoc.* **170,** 31-33.

12. Perryman, L.E. and Torbeck, R.L. (1980) Combined immunodeficiency of Arabian horses: confirmation of autosomal recessive mode of inheritance. *J. Am. Vet. Med. Assoc.* **176,** 1250-1251.

13. Shin, E.K., Perryman, L.E. and Meek, K. (1997) A kinase-negative mutation of DNA-PK(CS) in equine SCID results in defective coding and signal joint formation. *J. Immunol.* **158,** 3565-3569.

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

15. Ruis, B.L., Fattah, K.R. and Hendrickson, E.A. (2008) The catalytic subunit of DNA-dependent protein kinase regulates proliferation, telomere length, and genomic stability in human somatic cells. *Mol. Cell. Biol.* **28,** 6182-6195.

16. Neiditch, M.B., Lee, G.S., Huye, L.E., Brandt, V.L. and Roth, D.B. (2002) The V(D)J Recombinase Efficiently Cleaves and Transposes Signal Joints. *Mol. Cell.* **9,** 871-878.

17. Wiler, R., Leber, R., Moore, B.B., VanDyk, L.F., Perryman, L.E. and Meek, K. (1995) Equine severe combined immunodeficiency: a defect in V(D)J recombination and DNA-dependent protein kinase activity. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 11485-11489.

18. Ding, Q., Bramble, L., Yuzbasiyan-Gurkan, V., Bell, T. and Meek, K. (2002) DNA-PKcs mutations in dogs and horses: allele frequency and association with neoplasia. *Gene.* **283,** 263-269.

19. Davis, J.L. and Jones, S.L. (2003) Equine Primary Immunodeficiencies. *Compendium on Continuing Education for the Practising Veterinarian.* **25,** 548-557.

20. Perryman, L.E. (2000) Primary immunodeficiencies of horses. *Vet. Clin. North Am. Equine Pract.* **16,** 105-116, vii.

21. Bernoco, D. and Bailey, E. (1998) Frequency of the SCID gene among Arabian horses in the USA. *Anim. Genet.* **29,** 41-42.

22. Piro, M., Benjouad, A., Tligui, N.S., El Allali, K., El Kohen, M., Nabich, A. and Ouragh, L. (2008) Frequency of the severe combined immunodeficiency disease gene among horses in Morocco. *Equine Vet. J.* **40,** 590-591.

23. Zavrtanik, J., Mesarič, M. and Majdič, G. (2005) Genetic monitoring for severe combined immunodeficiency carriers in horses in Slovenia. *Slovenian Veterinary Research.* **42,** 37-41.

24. Seyedabadi, H.R., Afraz, F., Banabazi, M.H. and Emrani, H. (2006) Identification of SCID carriers among Iranian Arab horses using a test based on PCR for DNA-PKcs gene. *8th World Congress on Genetics Applied to Livestock Production, August 13-18, Belo Horizonte, Brazil.*

25. Fanelli, H.H. (2005) Coat colour dilution lethal ('lavender foal syndrome'): a tetany syndrome of Arabian foals. *Equine Veterinary Education.* **17,** 260-263.

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

27. Page, P., Parker, R., Harper, C., Guthrie, A. and Neser, J. (2006) Clinical, clinicopathologic, postmortem examination findings and familial history of 3 Arabians with lavender foal syndrome. *J. Vet. Intern. Med.* **20,** 1491-1494.

28. McAuliffe, S. and Slovis, N. (2011) Nervous System. In: *Equine Clinical Medicine, Surgery and Reproduction,* Eds G.A. Munroe and J.S. Weese. Manson Publishing, London.

29. Bierman, A., Guthrie, A.J. and Harper, C.K. (2010) Lavender foal syndrome in Arabian horses is caused by a single-base deletion in the MYO5A gene. *Anim. Genet.* **41,** 199-201.

30. Sellers, J.R. (2000) Myosins: a diverse superfamily. *Biochim. Biophys. Acta.* **1496,** 3-22.

31. Reck-Peterson, S.L., Provance, D.W.,Jr, Mooseker, M.S. and Mercer, J.A. (2000) Class V myosins. *Biochim. Biophys. Acta.* **1496,** 36-51.

32. Cheney, R.E., O'Shea, M.K., Heuser, J.E., Coelho, M.V., Wolenski, J.S., Espreafico, E.M., Forscher, P., Larson, R.E. and Mooseker, M.S. (1993) Brain myosin-V is a two-headed unconventional myosin with motor activity. *Cell.* **75,** 13-23.

33. Woolner, S. and Bement, W.M. (2009) Unconventional myosins acting unconventionally. *Trends Cell Biol.* **19,** 245-252.

34. Langford, G.M. (1995) Actin- and microtubule-dependent organelle motors: interrelationships between the two motility systems. *Curr. Opin. Cell Biol.* **7,** 82-88.

35. Brault, L.S., Cooper, C.A., Famula, T.R., Murray, J.D. and Penedo, M.C. (2011) Mapping of equine cerebellar abiotrophy to ECA2 and identification of a potential causative mutation affecting expression of MUTYH. *Genomics.* **97,** 121-129.

36. Dungworth, D.L. and Fowler, M.E. (1966) Cerebellar hypoplasia and degeneration in a foal. *Cornell Vet.* **56,** 17-24.

37. Brault, L.S. and Penedo, M.C. (2011) The frequency of the equine cerebellar abiotrophy mutation in non-Arabian horse breeds. *Equine Vet. J.* **43,** 727-731.

38. Ferrell, E.A., Gavin, P.R., Tucker, R.L., Sellon, D.C. and Hines, M.T. (2002) Magnetic resonance for evaluation of neurologic disease in 12 horses. *Vet. Radiol. Ultrasound.* **43,** 510-516.

39. Brault, L.S., Famula, T.R. and Penedo, M.C. (2011) Inheritance of cerebellar abiotrophy in Arabians. *Am. J. Vet. Res.* **72,** 940-944.

40. Sperandio, S., Tardito, S., Surzycki, A., Latterich, M. and de Belle, I. (2009) TOE1 interacts with p53 to modulate its transactivation potential. *FEBS Lett.* **583,** 2165-2170.

41. Gu, Y., Parker, A., Wilson, T.M., Bai, H., Chang, D.Y. and Lu, A.L. (2002) Human MutY homolog, a DNA glycosylase involved in base excision repair, physically and functionally interacts with mismatch repair proteins human MutS homolog 2/human MutS homolog 6. *J. Biol. Chem.* **277,** 11135-11142.

42. Englander, E.W., Hu, Z., Sharma, A., Lee, H.M., Wu, Z.H. and Greeley, G.H. (2002) Rat MYH, a glycosylase for repair of oxidatively damaged DNA, has brain-specific isoforms that localize to neuronal mitochondria. *J. Neurochem.* **83,** 1471-1480.

43. Lee, H.M., Wang, C., Hu, Z., Greeley, G.H., Makalowski, W., Hellmich, H.L. and Englander, E.W. (2002) Hypoxia induces mitochondrial DNA damage and stimulates expression of a DNA repair enzyme, the Escherichia coli MutY DNA glycosylase homolog (MYH), in vivo, in the rat brain. *J. Neurochem.* **80,** 928-937.

44. Parker, A.R. and Eshleman, J.R. (2003) Human MutY: gene structure, protein functions and interactions, and role in carcinogenesis. *Cell Mol. Life Sci.* **60,** 2064-2083.

45. Lee, H.M., Hu, Z., Ma, H., Greeley, G.H.,Jr, Wang, C. and Englander, E.W. (2004) Developmental changes in expression and subcellular localization of the DNA repair glycosylase, MYH, in the rat brain. *J. Neurochem.* **88,** 394-400.

46. Foley, A., Grady, J., Almes, K., Patton, K. and Davis, E. (2011) Cerebellar abiotrophy in a 6-year-old Arabian mare. *Equine Veterinary Education.* **23,** 130-134.

47. Brault, L.S. and Penedo, M.C. (2011) The frequency of the equine cerebellar abiotrophy mutation in non-Arabian horse breeds. *Equine Vet. J.* **43,** 727-731.

48. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb. Symp. Quant. Biol.* **51,** 263-273.

49. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science.* **239,** 487-491.

50. Markoulatos, P., Siafakas, N. and Moncany, M. (2002) Multiplex polymerase chain reaction: a practical approach. *J. Clin. Lab. Anal.* **16,** 47-51.

51. Righetti, P.G., Gelfi, C. and D'Acunto, M.R. (2002) Recent progress in DNA analysis by capillary electrophoresis. *Electrophoresis.* **23,** 1361-1374.

52. Brownstein, M.J., Carpten, J.D. and Smith, J.R. (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques.* **20,** 1004-1010.

53. Oddou-Muratorio, S., Vendramin, G.G., Buiteveld, J. and Fady, B. (2009) Population estimators or progeny tests: what is the best method to assess null allele frequencies at SSR loci? *Conservation Genetics.* **10,** 1343-1347.

54. Wu, Q., Chen, M., Buchwald, M. and Phillips, R.A. (1995) A simple, rapid method for isolation of high quality genomic DNA from animal tissues. *Nucl. Acids Res.* **23,** 5087-5088.

55. Krishnamoorthy, K. (2006) Handbook of Statistical Distributions with Applications. Chapman, Boca Raton

56. Toonen, R.J. and Hughes, S. (2001) Increased throughput for fragment analysis on an ABI PRISM 377 automated sequencer using a membrane comb and STRand software. *BioTechniques.* **31,** 1320-1324.

57. Krishnamoorthy, K. and Thomson, J. (2002) Hypothesis Testing About Proportions in Two Finite Populations. *The American Statistician.* **56,** 215-222.

58. Thompson, P.N., Heesterbeek, J.A.P. and van Arendonk, J.A.M. (2006) Changes in disease gene frequency over time with differential genotypic fitness and various control strategies. *Journal of Animal Science.* **84,** 2629-2635.

59. Leroy, G. and Baumung, R. (2011) Mating practices and the dissemination of genetic disorders in domestic animals, based on the example of dog breeding. *Anim. Genet.* **42,** 66-74.