UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Refinement and validation of a microsatellite based identification and parentage testing panel in horses

by

Anandi Bierman

Submitted in partial fulfillment of the requirements for the degree of

Magister Scientiae

Supervisor
Dr. Cindy Harper
Co-supervisors
Prof. Martin Schulman
Prof. Alan Guthrie

Department of Production Animal Studies
Faculty of Veterinary Science
University of Pretoria
Onderstepoort
2010

DECLARATION

I, Anandi Bierman, 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:
Anandi Bierman (candidate)

Date:

ACKNOWLEDGEMENTS

I have been so fortunate to be surrounded by family and friends who have provided me with unconditional support and encouragement throughout all the great milestones of my life, and then some…

Thank you so much to my supervisors Dr. Cindy Harper, Prof. Alan Guthrie and Prof. Martin Schulman. I cannot begin to express how much I have learnt and grown during the course of this project and I thank you for your patience and guidance.

Thank you to the breeders who offered samples for this project: Anette Struchtemeier, Thornhill Quarter Horse Stud, Wehlia de Villiers, Nickweb Clydesdale Stud and the Onderstepoort Teaching Animal Unit.

I am grateful for the financial assistance provided by the Veterinary Genetics Laboratory and the University of Pretoria.

Lastly I would like to thank my friends and colleagues at the VGL for always being good for a laugh!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The power of microsatellite markers lies in their ability to identify. Whether it is the identification of genes and associating them with known phenotypes or identifying and discerning individuals from one another, the role they play in the genetic field has been immense. Parentage testing of horses today is done via molecular means as opposed to serology. Microsatellite marker panels are decided upon by bodies such as the International Society for Animal Genetics (ISAG) in order to uphold international genotyping standards. The current horse microsatellite marker panel is not fully characterized and many markers are amplified by primers originally designed for linkage studies and were never intended for multiplex PCR analysis. The aim of this study was to refine and validate the current marker panel used for horses through sequencing of the repeat elements and flanking regions as well as the design of new primers for the setup of a marker panel incorporating more microsatellites and better primers. Sequencing of microsatellite flanking regions revealed that much variation lies within the regions flanking a microsatellite repeat element. Sequencing of the repeat element showed that not all markers are simple repeats, as was previously thought. The primers used to amplify microsatellite markers for horses were re-designed in the course of this study, utilizing knowledge gained from flanking region variation and repeat element length. New primers and known allele sizes allowed for the implementation of a nomenclature system in horses based on repeat element length as opposed to alphabet letters. By incorporating more markers into the panel it was hoped that a greater discriminatory power would be achieved. Measures of genetic diversity such as Observed Heterozygosity and Polymorphism Information Content showed negligible differences between the two panels however genotyping data from the old ISAG panel of nine markers showed that the probability of excluding an individual in a parentage test was better when using more markers.

CHAPTER 1: INTRODUCTION

1.1 Background & Problem Statement

There are more horses in the world today than in the era before mechanization, when horses were the driving force behind all transport and agriculture (Wilk, 1999). Managing such a large population and keeping track of its individuals requires a system of identification that is unbiased and easily exchanged internationally.

An individual horse can be priceless and therefore identification of the individual is important (Dimsoski, 2003). Discerning one horse from another used to be based on appearances alone but for example, brands can be covered and white feet painted. In one instance, Hill (2002) found errors in the assignment of founding mares of the Thoroughbred population even though it remains one of the most comprehensive pedigrees of all domestic animals and parentage verification by blood typing has been compulsory since the 1980's (Hill et al., 2002). A molecular approach was decided upon as it cannot be altered and it is much more informative than blood typing. Today microsatellite markers are indices of genetic variation and provide a unique 'fingerprint' (or profile) for each individual animal. At the meeting of the International Society for Animal Genetics (ISAG) in New Zealand in 1998 a minimum marker panel for international exchange was decided upon. Nine microsatellite markers were chosen and standardized based on an alphabetical nomenclature system. Regular worldwide comparison tests are held every two years under the auspices of ISAG (Bowling, 2001).

Genetic markers quickly replaced blood typing which shows low levels of variation between individuals (Shriver et al., 1995). Microsatellite markers are, however, not without their own shortcomings. The primers currently used to amplify these markers were originally designed for linkage studies and have inefficient annealing temperatures, GC contents and sequences that make them less than ideal for genotyping purposes (INRA Biotechnology Laboratories; Horsemap database, http://locus.jouy.inra.fr/ and NCBI GenBank, www.ncbi.nih.nlm.gov).

The international comparison tests held by ISAG have shown low success rates. During the 2005/2006 test only 28% of participating laboratories achieved 100% genotyping success on all samples. This figure dropped to 25% of the laboratories achieving 100% success in the 2007/2008 test. Generally, these low success rates are due to genotyping errors as assays and manual sample handling can never be absolutely reliable (Bonin et al., 2004). The nature of DNA itself is also one of the main reasons for genotyping errors. Equine microsatellites consist of dinucleotide repeat elements that create stutter bands during the amplification of DNA fragments or Polymerase Chain Reaction (PCR). Stutter bands can make the calling of actual alleles difficult. Mutations in microsatellite flanking regions can lead to failure of primer binding, leading to null alleles and can cause false assignment of homozygosity. Grimaldi (1997) found, upon sequencing of a CA repeat microsatellite, that differences between alleles were more complex than previously thought. The authors found that much of the variability in microsatellites is due to variation in the repeat's flanking regions and not only to variation in repeat length (Grimaldi & Crouau-Roy, 1997). These 'flanking region' alleles might be identical in state but not by descent and their occurrence is referred to as allelic homoplasy (Narkuti & Oraganti, 2008).

Inbreeding complicates individual identification as more inbred populations will have less genetic variation present to be identified by microsatellite markers. This has created a need for more informative markers in order to distinguish between closely related individuals. Excluding the nine markers recommended by ISAG, there are additional microsatellite markers used by many laboratories to get around the problems caused by low genetic variability. Many laboratories also use self designed primer modifications for the recommended markers in order to obtain better probabilities of exclusion.

The probability of exclusion (PE) is the probability of making an accurate exclusion of parentage when an individual and its two parents are analyzed. According to Ozkan (2009) this figure should be greater than 0.999. Informativeness of a marker is determined by the number of alleles it has as well as the frequency distribution of those alleles in a population (Ozkan et al., 2009). Other statistical measures of variability include Heterozygosity (He) and Polymorphism Information Content (PIC). Many population studies have been done in horses and the levels of inbreeding and genetic variation are known for many breeds (Achmann et al., 2004 & Canon et al., 2000) yet there is no collective data on the horse marker panel currently used.

As more animal genotyping is done, especially for forensic purposes, the need for standardized minimum guidelines becomes evident (Budowle et al., 2005). The International Society for Forensic Genetics (ISFG) has published recommendations on the use of DNA-based typing in forensic genetics (Morling et al., 2002) and propose a sequence-length based nomenclature as is used for human DNA-typing (Budowle et al., 2005). Genotyping of many animal species are starting to follow these guidelines as it serves as a good model for animal-typing. Positive controls should be used with each genotyping run and can consist of any previously typed sample although commercially available cell lines have the advantage of being internationally available as well as being an inexhaustible source of DNA.

1.2 Research Questions & Objectives

This study aims to develop and redefine an international microsatellite marker panel for horses that is based upon a solid scientific background. The new panel should have greater statistical significance than its predecessor: increased informativeness and a high probability of exclusion are essential as is accurate data on parameters such as allele frequencies, inbreeding coefficients and Hardy-Weinberg equilibrium. Sequence data generated during the course of this study will elucidate much about sequence variation in the regions flanking the microsatellite repeats.

The objectives of this study are as follows:

- Sequence microsatellite flanking regions of two problem markers; HTG10 and HMS3, and three markers known to work well in most instances; VHL20, HMS7 and HTG4.
- Sequence microsatellite alleles.
- Define a sequence-based nomenclature system with alleles differentiated by microsatellite repeat length
- Design new primers for 16 microsatellite markers and the Amelogenin sexing marker.
- Optimize primer concentrations and PCR parameters of new primers in one multiplex PCR.
- Compare the difference in informativeness between the 16-plex marker panel and the 9-plex ISAG marker panel.
- Determine the informativeness of a hypothetical 12-plex marker panel that could be implemented by ISAG, selecting the best markers to use for this panel from the 16 markers used in this project.

1.3 Thesis Statement

Through the implementation of a sequence-based nomenclature system, recommended by the ISFG, and the addition of seven microsatellite markers to the nine markers recommended by ISAG this project will aim to redefine and improve upon the microsatellite marker panel used for parentage testing in horses.

1.4 Delineations and Limitations

- Statistical analyses of the marker panels (old and new) will mostly be based on samples from the Thoroughbred and Arabian horse breeds as samples from other breeds are limited.
- Sequencing of microsatellite flanking regions will be limited to five markers only due to the costs involved.
- The microsatellites chosen for the new horse microsatellite marker panel are based on the nine ISAG-recommended markers as well as markers prescribed by ISAG for additional use. No other markers were considered.
- The marker Lex003 was not included in the statistical validation of the marker panel for the populations of Thoroughbreds and Arabs as it is present on the X chromosome and therefore only one copy is ever present in stallions. This would create a skewed Heterozygosity estimate.

1.5 Definitions and Abbreviations

The term ‘horse microsatellite marker panel’ is used throughout, though this term is analogous to the terms ‘genotyping panel’ or ‘parentage testing panel’.

Whenever the microsatellite marker panel is referred to as ‘current’ or ‘old’, it refers to the panel of nine markers and their published primers as recommended by ISAG. The microsatellite marker panel designed in the course of this study consists of 16 microsatellite markers and the Amelogenin sex marker and will be referred to as the ‘new horse microsatellite marker panel’.

AFLP – Amplified Fragment Length Polymorphism
He – Heterozygosity
HWE – Hardy Weinberg Equilibrium
ISAG – International Society for Animal Genetics
ISFG – International Society for Forensic Genetics
ISSR – Inter Simple Sequence Repeat
PIC – Polymorphic Information Content
RAPD – Random Amplified Polymorphic DNA
RFLP – Restriction Fragment Length Polymorphism
SNP – Single Nucleotide Polymorphism
STR – Short Tandem Repeat

1.6 Results Overview

Chapter 3

Allele specific sequencing and the application of a numerical nomenclature system.

This chapter will focus on the sizes of the repeat elements of alleles sequenced for each locus and the implementation of a new numerical nomenclature system based on the sequenced sizes of repeat elements.

Chapter 4

Sequencing microsatellite flanking regions, primer design and panel validation for a new multiplex panel. This chapter will discuss any sequence anomalies observed between breeds and individual animals focusing on their significance to primer binding and allele sizing. The principles and parameters behind primer design for the new microsatellite marker panel will be discussed. The setup and validation of the new marker panel will be discussed and the final primer concentrations and PCR parameters to be employed will be given.

Chapter 5

Locus information and population data analysis for nine ISAG markers compared to the 17-plex panel in 100 Arabian and Thoroughbred horses. A brief marker status report for the newly designed microsatellite marker panel will be given indicating fragment sizes and dinucleotide repeat lengths. An analysis of population data for 100 Thoroughbred and 100 Arabian horse samples genotyped using the old marker panel of nine ISAG markers as well as the new microsatellite marker panel using 16 microsatellites, will be done using CERVUS 3.0.3 (Tristan Marshall; Fieldgenetics Ltd. www.fieldgenetics.com). The analysis will include allele frequencies and number of effective alleles, probability of non-exclusion, Polymorphism Information Content, Heterozygosity as well as Hardy-Weinberg equilibrium.

CHAPTER 2: LITERATURE REVIEW

2.1. Microsatellites and Other Genetic Markers

Microsatellites are short segments of DNA that consist of repeating nucleotides and are therefore also referred to as Short Tandem Repeats (STRs). The repeat units can range from two to six base pair motifs and the entire microsatellite can range in size from fewer than ten to hundreds of bases; depending on the number of repeat units. The nature of the repeat element can be categorized as simple $(AC)_n$; compound (two or more microsatellites found in close proximity) or complex (containing repeat units of several nucleotides), either of which may be interrupted or not (Kofler et al. 2008). Microsatellites are inherited in a Mendelian fashion and CA-repeats are the most common motif in most mammalian genomes. Microsatellites are abundant and evenly spread throughout the genome although they are often associated with non-coding DNA.

A marker is defined by Meudt (2007) as an amplified locus that is informative in that it shows polymorphism between individuals and it can be visualized by, for example, a gel-based method such as Amplified Fragment Length Polymorphism (AFLP) (Meudt & Clarke, 2007). Genetic markers can be classified according to two types: genes with known functions (Type I) and anonymous DNA fragments (Type II). Type II markers include marker systems such as AFLP, Random Amplified Polymorphic DNA (RAPD) and microsatellites (Emara & Kim, 2003).

Whole genome markers such as Restriction Fragment Length Polymorphism (RFLP) and AFLP rely on the digestion of genomic DNA by restriction enzymes to produce variable patterns between individuals. These approaches are gel based and therefore not suited to high throughput genotyping (Beuzen et al., 2007). Being dominant markers, it is also impossible to distinguish between homozygotes and heterozygotes. This makes the discriminatory power of these methods less than ideal, although RFLP has been used successfully for parentage analysis in Thoroughbred horses (Takagi et al., 1995). RAPD is another dominant marker system which is not whole genome-based but relies on the detection of polymorphisms with a few nucleotide mismatches (Beuzen et al., 2000). Another repeat-based marker is the Inter Simple Sequence Repeat (ISSR) which produces similar profiles to RAPD but uses primers anchored to microsatellite sequences (Njiru et al., 2007). Minisatellites were the markers used during the initial stages of human forensic DNA fingerprinting. These markers consist of repetitive units ranging in size from 6 to hundreds of bases (Tamaki & Jeffreys, 2005). Minisatellites are effective in human identification but have shown poor power of discrimination in horses (Anunciacao & Astolfi-Filho, 2000).

Traditionally, DNA separated in a polyacrylamide gel would have been visualized by autoradiography or silver staining. Variation from one gel to another, lane-to-lane variation and crooked gels meant that size standards were never completely accurate. A decision would usually have to be made as to which allele (repeat number) corresponded to which band on the Southern Blot autoradiograph. Comparing data from one autoradiograph to another or between different laboratories, therefore, proved very difficult (Bennett, 2000). PCR, high resolution media and fluorescence technology have resolved this problem and have enabled scientists to analyze up to twenty microsatellites in a single gel lane.

Microsatellites are co-dominant which means that heterozygotes can be distinguished from homozygotes. In addition, microsatellites can have many different alleles at a single locus. These traits make them highly polymorphic and ideal in identification. Microsatellite polymorphism is generated through interplay of a high mutation rate and polymerase slippage during DNA replication (Beuzen et al., 2000 & Lee & Cho, 2006). The polymerase slippage that generates microsatellites and the variation thereof occurs as a result of mispairing between the template DNA strand and the newly synthesized strand during DNA replication. The unpaired segment of DNA loops out, resulting in the gain of a repeat unit if the loop is on the new DNA strand or the loss of a repeat unit if the loop is on the template strand. Due to this mechanism the abundance of a repeat unit size increases as the size of the repeat decreases, i.e. mononucleotide repeats are the most abundant and pentanucleotide repeats are the rarest. In general, such loops generated through polymerase slippage are repaired by the DNA mismatch repair systems in the cell. Therefore, the microsatellite’s observed mutation rate and the actual rate at which slippage occur in the cell are not the same. This is an extremely useful phenomenon: if repair was too efficient and the mutation rate was too low, microsatellites would be very rare. If repair mechanisms didn’t exist and mutation rates were too high, there would be too much variation from one generation to the next (Bennett, 2000).

The same principle responsible for the generation of microsatellite variation is also one of the major hindrances in the amplification of microsatellites. Stutter or shadow bands are unwanted artefacts generated during PCR amplification of microsatellites, especially dinucleotide repeats. These bands appear along with the band corresponding to an expected fragment in increments of the repeat size of the specific marker. In genotyping of horses where dinucleotide repeats are used, the stutter bands occur in increments of 2bp larger or smaller than the expected allele. This often leads to uncertainty as to the exact size of an allele or whether the allele is homozygous or heterozygous (Scotti et al., 1999). Slipped strand mispairing occurring during the PCR or artefactual 'recombination' caused by out-of-register annealing of truncated PCR products could explain this phenomenon. Sequencing data generated by Hauge (1993) rule out PCR recombination and are in favor of slipped strand mispairing as the cause (Hauge & Litt, 1993).

Microsatellite alleles are differentiated by the size of the repeat element, but the size of a microsatellite allele can also be influenced by mutations in the flanking regions around the repeat element. This often creates alleles of the same apparent length but with different sized repeat elements, a phenomenon known as allelic homoplasy. Such alleles might be identical in state, but not by descent. In addition, flanking region mutations could abolish primer binding, creating non-amplifying or null alleles and a false heterozygote deficiency in a population (Narkuti & Oraganti, 2008). It is important to test the inheritance of potential genetic markers against known parent-offspring relationships as this helps detect null alleles, allele dropout and other scoring difficulties.

Currently there is a lack of data relating to microsatellite structure at the intraspecific level and at a population level there is not enough sequence data to demonstrate microsatellite complexity and size homoplasy (Grimaldi & Crouau-Roy, 1997). Sequencing is an important tool in revealing size homoplasy in genotyped homozygotes. Most often the homoplasy is due to polymorphic sites in the flanking regions but it can also be caused by base changes in the microsatellite repeat itself. MacAvoy (2008) reported that sequencing revealed discrepancies in allele sizing of up to 5bp between sequence size and electrophoresis fragment size (MacAvoy et al., 2008).

Microsatellites have often been described as 'junk DNA' due to their apparent lack of function within the genome (Bowling, 2001) but recent studies are discovering that there is more to these repetitive sequences than meets the eye. Microsatellites are known to form secondary structures such as hairpins, Z-DNA and B-DNA. These structures in the genome have been shown to affect DNA replication through hairpin structures stopping DNA extension. Secondary structures in promoter regions of genes have also been shown to affect gene expression (Chistiakov et al., 2006). Dinucleotide repeat sequences have also been found to be preferred sites for recombination and this correlates well with their ability to form secondary structures.

It is suspected that microsatellite density, rather than a specific motif, determines the relationship with recombination (Guo et al., 2009). Microsatellites are distributed throughout the entire genome but there are certain trends as to where they might be clustered or found in higher densities. Microsatellite densities have been found to be higher on the X chromosome. In the genome of *Drosophila melanogaster* it seems that base composition has an effect on microsatellite density while the ends of chromosome arms, centromeric and pericentromeric regions show a reduction in microsatellites (Guo et al., 2009). The localization of microsatellites to particular regions of the chromosome is indicative of their role in the organization of chromosome structure. Telomere-associated repeats have been implicated in the maintenance of stability of the telomeres during DNA replication and secondary structures formed in the centromeres aid in centromeric chromatin compactness (Chistiakov et al.,2006).

The marker system likely to replace microsatellites in many applications is Single Nucleotide Polymorphisms (SNPs), which is the most abundant polymorphism found in most organisms. Their abundance makes them useful, despite their limited polymorphic content (as singular polymorphisms). A SNP is the substitution of one nucleotide for another and therefore it is a bi-allelic marker as opposed to microsatellite loci which can have multiple alleles. SNPs have low mutation rates, short amplicon sizes, are distributed evenly throughout the genome and are suitable for high throughput analysis. The drawbacks are that many more SNPs are needed to achieve the same discriminatory power as a panel of microsatellites and also that the cost efficiency is not yet well established (Sobrino et al., 2005).

There are, however, many good reasons to increase interest in the use of SNPs for genetic analysis. SNPs are abundant and distributed throughout the genome. The frequency of SNPs in the Equine Genome is estimated to be one every 732bp (Shubitowski et al., 2001). The abundance of SNPs provides more potential markers associated with genes or loci of interest thus increasing the density and efficacy of linkage maps. SNPs are inherited in a more stable fashion than microsatellites making them more suitable for long term selection. With the advent of DNA microarray technology and solid state platforms such as SNP-chips, SNPs are proving more suitable for truly high throughput genotyping (Beuzen et al., 2000).

2.2. Applications of Microsatellite Markers

Identification of individuals used to be based on blood typing. These serological tests for erythrocyte antigens have many drawbacks in that they show low levels of polymorphism and heterozygosity and are not suitable for high throughput analysis (Alosi et al., 1980). In contrast, microsatellite loci can have multiple alleles and are superior to serology as they permit accurate inclusion as well as exclusion (Anunciacao & Astolfi-Filho, 2000). Another drawback was that blood grouping antisera was never made available commercially and individual laboratories had to keep animals whose blood could be used to make up a series of reagents by alloimmunization. Sample use was also limited to fresh blood which created problems when no veterinary staff was available or when samples needed to be transported internationally (Bowling, 2001).

The advent of genetic markers solved the problems posed by serology. Microsatellites as genetic markers have many uses ranging from identification to gene discovery and phylogenetics. Many markers originally intended for gene mapping were eventually applied to identification (Dimsoski, 2003). The application of microsatellites for the identification of individuals has proven to be one of the most significant, making it suitable for use in studies on reproductive success, kinship and discerning questionable parentage (Lee & Cho, 2006).

Forensic Genetics began in the 1980’s with minisatellites and their detection by multi locus probes (Jeffreys et al.,1985). The system worked well as the probability of two individuals having matching profiles was so low that it would only occur in monozygotic twins. On analysis of the putative offspring and parents, a minimum of three mismatches are required to report as an exclusion of parentage (Narkuti & Oraganti, 2008). In due course, because of the lack of sensitivity offered by multi locus probes, single locus probes were used instead. These targeted a single minisatellite and required less DNA than the RFLP-based Southern Blotting technique of multi locus probes (Tamaki & Jeffreys, 2005). Subsequently, microsatellites were set to replace minisatellites as the marker of choice in forensic genetics (Tamaki & Jeffreys, 2005).

Today microsatellites are routinely used to genotype various animal species (Anunciacao & Astolfi-Filho, 2000; Luikart et al., 1999; Mukesh et al., 2009) and regular comparison tests are held under the auspices of ISAG. The advantage of microsatellite testing lies in its ease of use: any sample containing the animal’s DNA can be used and it is quick and efficient with modern automated technologies (Ozkan et al., 2009). Microsatellite-based identification has increased the efficiency of breeding programs which has led to increased profit (Luikart et al., 1999). Animal breeding in advanced countries relies heavily on quantitative genetics as it has the power to maximize the response to selection (Beuzen et al., 2000).

Microsatellites are not the only markers used for linkage mapping and the mapping of Quantitative Trait Loci (QTL). They are used very often and have elucidated many candidate genes for disease (Dierks et al., 2007), visible phenotypic variation such as coat color (Locke et al., 2001 & Swinburne et al., 2002) as well as economically important traits (Tozaki et al., 2007). Thoroughbred horses, in particular, are proving to be valuable models in the search for genes that govern physical performance traits.

Appropriately spaced, it only requires 430 microsatellites to provide a sufficient whole genome Linkage Disequilibrium map in the horse (Tozaki et al., 2005). Sequence homology between human and horse chromosomes means that equine chromosomes can be used in comparative mapping as well. This approach provides insight into QTLs in humans from known equine sequences and vice versa. The horse-human comparative map generated by Tozaki (2007) provides a platform for the study of performance genes using equine microsatellites. Thoroughbreds in particular can serve as a good model as they are used for racing all over the world and accurate performance and pedigree data are available to facilitate the breeding of potentially better racehorses (Tozaki et al., 2007).

Genetic diversity can be measured through the frequencies of genotypes and alleles, the proportion of polymorphic loci, observed and expected heterozygosity as well as allelic diversity. Knowledge of the variation within and amongst breeds can aid in the establishment of conservation priorities for rare and endangered breeds and species (Aberle et al., 2004). Genetic markers also provide estimates of the times of divergence of different species; useful information in population and phylogenetic studies that enable researchers to look back into the history of a species in order to determine its future through conservation efforts (Achmann et al., 2004; Marletta et al., 2006).

Canon et al. states that conservation genetics aims to preserve genetic variability in a population as this is linked to the viability of that population. Theoretically, fragmented subpopulations are important in maintaining variation as it reduces the loss of alleles. The authors refer to Celtic horse populations but the principle applies equally well to Thoroughbreds or Arabians in South Africa that are divided into subpopulations between different stud farms, bloodlines or provinces of the country (Canon et al., 2000).

The usefulness of genetic markers and their application in population genetics lies in the observation that if one can manage the rate of inbreeding or the effective population size, a general framework for managing genetic resources becomes possible. It can restrict inbreeding depression, the probability of losing beneficial alleles and the risk of extinction. There is no major difference between selection and conservation programs, especially when farmed animals are considered. Selection programs aim to maximize performance for economically important traits and impose restrictions on inbreeding; conservation programs aim to minimize inbreeding and some selection is imposed to avoid decreased performance that make the breed valuable (Toro et al., 2009).

2.3. Microsatellites and Their Use in Genotyping

Genotyping errors often go unnoticed in many studies as they appear to be inconspicuous at first. Unfortunately, errors can never be completely eradicated as tests are never completely reliable and human error such as manual sample handling should always be taken into account. Genotyping errors are generally due to allelic dropout, false alleles and contamination. Genotyping errors due to technical causes are best documented as follows: amplification artefacts, biochemical anomalies, electrophoresis discrepancies, surrounding temperatures, materials and protocols used or template DNA quality and quantity. These errors at the population level affect allele frequencies and the accurate discrimination of different genotypes. False allele frequencies create a false excess of heterozygotes, false departure from Hardy Weinberg equilibrium or false inbreeding coefficients, not to mention false parentage results.

The success of genotyping relies greatly on the initial setup of a good marker panel. Extensive validation of population groups and the use of numerous markers lay a good foundation (Sobrino et al., 2005). A good marker panel should be accurate, effective and economical. Rapid return of results, established reference standards and ease of information transfer between laboratories are also essential. With increasing demand for forensic applications, a marker panel cannot be limited by sample type and should preferably be highly automated (Bowling, 2001). A marker panel that can be amplified in as few PCR reactions as possible and analyzed by automated means saves time and money (Dimsoski, 2003). The addition of more loci to a panel will increase its power of discrimination (Ozkan et al., 2009). The development cost of new and applicable microsatellite markers is relatively high, especially in non-commercially utilized species. The development of locus specific markers requires the isolation and characterization of the locus, sequencing, primer design and synthesis (Hayden & Sharp, 2001).

Microsatellites have reading-rules that govern how the sequence should be read as it influences the designation of the repeat motif. Generally, for microsatellites found in the coding segments of genes or in introns, the encoding strand is used to assign the number and type of repeats to the microsatellite. For other microsatellites, the sequence first described in the literature should be used for defining the repeat element. The sequence motif should be designated on the first 5' nucleotide that can make up a repeat motif (Budowle et al., 2005). A partial deletion or insertion in or around the repeat element of a microsatellite can lead to intermediate or variant alleles. These alleles are designated in nomenclature systems as the number of perfect repeats followed, after a full stop, by the number of partial repeats represented (Myres et al., 2009).

There are problems associated with sharing genotyping data generated by different electrophoretic platforms. For this reason, internal control samples and size standards are used when genotyping samples in order to ensure consistency between runs. Though an alphabetical nomenclature system is exchangeable between laboratories the fact is that an international standard should be maintained for all organisms genotyped for paternity or forensic purposes. Laboratory standards are moving toward sequence-based nomenclature as recommended by the ISFG (van Asch et al., 2008).A universal nomenclature system would aid effective data sharing especially since allele designation for new markers is not automatically available. The use of allelic ladders for standardization and allele designation is recommended by the ISFG, though this system is not always practical when working with di-nucleotide repeats and currently there is no allelic ladder commercially available for horses (van De Goor et al. 2009).

2.4. Molecular Methodologies

DNA extraction

DNA extraction from hair samples mainly requires breaking the hair follicle open to expose the DNA in the cell nuclei. 200mM Sodium Hydroxide (NaOH; Sigma®-Aldrich, St. Louis; MO) is used in an alkaline lysis step and hair roots are incubated at 97°C for 15 min. The solution is then brought to a more neutral pH using an acidic buffer consisting of 200mM Hydrochloric Acid (HCl; Saarchem, MERCK, Midrand; Gauteng) and 100mM Tris-HCl, pH 8.5 (Tris base; Promega, Madison; WI) before it is ready to use for PCR.

Extracting DNA from blood samples requires washing un-clotted blood in 10mM Sodium Chloride (NaCl; Promega, Madison; WI) and 10mM Ethylene Diamine Tetra Acetic Acid (EDTA, pH 7; Promega, Madison; WI) to break open red blood cells which then stay in solution when the sample is centrifuged. DNA-containing white blood cells and proteins are present in the pellet which is incubated at 56°C for 2 hours in a solution of 10mM Tris-HCl (pH 8), 10mM EDTA, 50mM NaCl, 20 % Sodium Dodecyl Sulphate (SDS; BDH Laboratory Supplies, Poole; England) and 20µm/ml Proteinase-K (Sigma®-Aldrich, St. Louis; MO). This solution serves to buffer the reaction, break down proteins and lyse the white blood cells to expose the DNA in the cell nucleus. The samples are subjected to treatment with Phenol-Chloroform (PCIA; Sigma®-Aldrich, St. Louis; MO) which rids the released DNA of contaminating proteins. Phenol and Chloroform degrades proteins by separating the liquid or aqueous phase from the organic phase (Sambrook, 1989). Finally DNA is precipitated and washed with 96% and 70% Ethanol (EtOH) respectively. Extracted DNA pellets are resuspended in Tris-EDTA (TE; Promega, Madison; WI).

Polymerase Chain Reaction

Polymerase Chain Reaction (PCR), invented by Kary B. Mullis (Mullis & Faloona, 1987) is a method of synthesizing or replicating specific segments of DNA. Two oligonucleotide primers flank the sequence of interest and through repeated cycles of heat denaturing, primer annealing and primer extension by DNA polymerase, the DNA segment is copied and accumulates exponentially (Innis et al., 1990).

Multiplex PCR is a variant of PCR in which two or more loci are amplified simultaneously in the same reaction**.** Multiplex PCR was first described in 1988. In multiplex PCR, as more loci are added, the pool of enzyme and nucleotides becomes a limiting factor. More time is needed for polymerase to complete synthesis and typical difficulties encountered include uneven or lack of amplification and difficulties in reproducing results (Henegariu et al., 1997).

Several components are required for a successful PCR reaction and primers are probably the most notable. The success of a multiplex PCR depends on criteria such as annealing temperatures of the primers, primer length and GC content. If primers are too short they won't be specific enough while a GC content that is too high will slow the denaturing of the DNA. Other sequence anomalies such as 3' complimentarity or runs of three or more similar bases could lead to the formation of primer dimers or other secondary structures, leaving less primer and other reagents available to aid the amplification of the DNA segment of interest. The buffer added to a PCR reaction maintains the pH and prevents nicking and depurination of the DNA while the concentration of Magnesium Chloride (MgCl) affects primer binding and enzyme activity. Deoxynucleotide Triphosphates (dNTPs) are the building blocks used for the synthesis of new strands of DNA by *Taq* Polymerase (Innis et al., 1990).

Sequencing

DNA can be sequenced to determine the order and type of nucleotides in a strand. Most sequencing today is based on Sanger Dideoxy sequencing, derived by Frederick Sanger in 1977. Samples are PCR amplified with the addition of fluorescently labeled dideoxynucleotide triphosphates (ddNTPs) which lack Hydroxyl groups at the 2' and 3' carbons. When DNA polymerase adds nucleotides to a growing chain it can add either a ddNTP or a dNTP to the growing strand. Once a ddNTP is in place, another nucleotide cannot be added because of the Hydroxyl groups. This creates fragments of different sizes, ending in different bases. Different sized fragments are separated by capillary electrophoresis and pass the laser beam, which picks up the fluorescence of each ddNTP, from smallest to largest (Fairbanks & Andersen, 1999). Computer software constructs electropherograms based on the fluorescence of the ddNTPs and the sequence of a DNA strand can be read based on the different colored peaks observed.

Capillary electrophoresis

Traditionally, gel electrophoresis involves the separation of DNA strands in a gel-based medium by utilizing the negative charge of DNA and its attraction or repulsion by electric current. Capillary electrophoresis uses the same principle although the gel is replaced by fine capillaries and polymer. Separate buffer reservoirs contain an electrode connected to the power supply. Samples are injected onto the capillary by temporarily replacing one of the buffer reservoirs with a sample reservoir and applying an electric potential. Separated fragments can be detected by the laser through the capillary wall and data is processed in the form of an electropherogram which depicts sample migration versus fluorescence (Grossman & Colburn, 1992). Capillary electrophoresis has many applications including fragment analysis; genotyping and SNP typing (Sanchez et al., 2008).

CHAPTER 3: ALLELE SPECIFIC SEQUENCING AND THE APLICATION OF A NUMERICAL NOMENCLATURE SYSTEM

Introduction

Microsatellites are the subject of extensive study and a publication by Schlotterer (2000) illustrates the confounding data that so many studies have yielded over the course of the past 20 years. A table from this report depicts the lack of consensus over matters of microsatellite evolution ranging from debates over their functionality to directional evolution and even whether or not there truly is selection against longer length alleles. Despite the dinucleotide motifs being prone to producing stutter products that make allele calling difficult, microsatellites have been used successfully to genotype horses internationally for many years and many laboratories have extensive collections of data from these markers.

The characteristics of a good marker panel are simple: polymorphic markers; fast, economical use and accuracy. International standardization and data exchange depend greatly on accuracy between laboratories. Authors agree that using standard controls such as ATCC cell lines (www.atcc.org) and establishing the precise lengths of marker alleles along with a proper nomenclature system are crucial in allowing standardization of genotypic data for international data exchange (Lipinski et al., 2007).

With the international scale of genotyping as well as the increase in forensic cases, a nomenclature system that follows international trends as well as a system that is based on sound sequence data, is required. A study by van De Goor (2009) suggests the use of a nomenclature system based on that of the ISFG. The ISFG requires the repeat sequence motif of an STR to be defined and distinguishes between simple and complex repeats. To date no data has been published on the sequence characteristics of the microsatellite markers used for genotyping horses, except for a single paper by van De Goor et al. (2009).

The ISFG requires that loci used for parentage testing or individual identification have no palindromic sequences in the flanking regions, do not lie within coding regions of chromosomes and should preferably occupy separate chromosomes (van Asch et al., 2008). ISAG, through its Equine Working Group, standardizes the markers used for parentage testing in horses and organizes international comparison tests between laboratories that do parentage testing in animals (Lee & Cho, 2006). A minimum panel of nine loci was agreed upon at the 1998 ISAG meeting in New Zealand. The choice of markers was based on work done through international collaboration and the chosen markers have demonstrated consistent results among laboratories participating in ISAG's comparison tests (Bowling, 2001). Kits for parentage testing, based on these loci, have been developed by Applied Biosystems (StockMarks®) (Bozzini et al., 1996; Marklund, Ellegren, Eriksson, Sandberg, & Andersson, 1994) and Finnzymes Diagnostics (The Equine Genotypes™ Panel 1.1).

The current nomenclature system used for genotyping horses is based on alphabet letters assigned to internationally standardized alleles. Fragment-sizing is automated and allele calling is based on a binning process where maximum and minimum values for inclusion into a particular bin are set. This chapter will discuss the sizes of repeat elements or alleles extrapolated from the sequencing results obtained for three alleles each of 16 microsatellite loci. The number of repeat motifs will constitute a new numerical nomenclature system rather than an alphabetical system.

Methods and Materials

Sequencing Marker Alleles

Samples were obtained from storage at the Veterinary Genetics Laboratory (VGL) and consisted of Thoroughbred horse samples previously genotyped by the VGL and homozygous for the alleles chosen to be sequenced. The M or middle allele was sequenced for every locus. In addition, two alleles, one large and one small within the known allelic range were sequenced for every one of the 16 loci. These alleles were selected as per availability.

Novel sequencing primers for sequencing of the microsatellite repeat elements were designed using FastPCR (Primer Digital Ltd. Version 5.2.118; 2008) or Primer Designer 4 (SciEd Central Version 4.20) and based on GenBank sequences (www.ncbi.nih.nlm.gov). Primers were designed so as to anneal approximately 40bp from the repeat element. Primers for sequencing were obtained from Integrated DNA Technologies; Whitehead Scientific (Pty) Ltd. PCR reactions were carried out using 1x PCR Gold Buffer and 1.5mM MgCl (Applied Biosystems; Roche, Branchburg; New Jersey), 0.5mM dNTP mix (Thermo Fisher Scientific Inc., Epsom; Surrey) and 2.5U Super-Therm GOLD Taq Polymerase (Southern Cross Biotechnology, Claremont; Cape Town) in 20µl reaction volumes. The PCR conditions on the Veriti 96-Well Thermal Cycler (Applied

Biosystems, Warrington; UK) were as follows: 95°C for 10min; 35 cycles consisting of 95°C for 45sec, 60°C for 1min 15sec, 72°C for 2min and a final extension of 72°C for 30min. After the final extension step, samples were held at 4°C until used. PCR products were purified using the MSB® Spin PCRapace kit by Invitek (Invisorb®; Berlin LOT BP080040; Ref 10202203).

Sequencing reactions were carried out using ABI PRISM® dGTP BigDye® Terminator Cycle Sequencing Kit (Applied Biosystems, Warrington; UK) and reactions were cleaned using a Sodium Acetate (Amresco®, Solon; Ohio) and Ethanol-based cleanup before being sequenced on the ABI 3130x Genetic analyzer in 10µl Hi-Di™ Formamide (Applied Biosystems, Warrington; UK). Results were visualized using Applied Biosystems Sequencing Analysis software v5.2.

HTG7, HTG6, CA425 and HMS2 had to be additionally genotyped in order to match the sequence sizes obtained to fragment sizes of a particular allele as no genotyping data was available for these loci. Published primers for these markers were obtained from HorseMap (www.locus.jouy.inra.fr) and samples were genotyped in singleplex PCR reactions using 1x PCR Gold Buffer and 1.5mM MgCl (Applied Biosystems; Roche, Branchburg; New Jersey), 0.5mM dNTP mix (Thermo Fisher Scientific Inc., Epsom; Surrey) and 2.5U Super-Therm GOLD Taq Polymerase (Southern Cross Biotechnology, Claremont; Cape Town) in 20µl reaction volumes. The PCR conditions on the Veriti 96-Well Thermal Cycler (Applied Biosystems, Warrington; UK) were as follows: 95°C for 10min; 35 cycles consisting of 95°C for 45sec, 60°C for 1min 15sec, 72°C for 2min and a final extension of 72°C for 30min. After the final extension step, samples were held at 4°C until used. Genotyping was performed on the 3130xl Genetic Analyzer (Applied Biosystems). Samples were mixed with HiDi-Formamyde (Applied Biosystems) and LIZ-500 size standard. Genotyping data was analyzed on STRand version 2.3.94 (University of California).

Conversion to a Numerical Nomenclature System

Alphabetical allele designators were replaced by the sequenced sizes of the microsatellite repeat elements where numbers would indicate the number of dinucleotide repeats. Sizes of the alleles that were not sequenced were extrapolated based on the dinucleotide nature of the markers.

Results

Sequencing Microsatellite Alleles

Novel sequencing primers were designed in order to sequence through microsatellite repeat elements; the primer sequences are listed in table 3.1 with their respective annealing temperatures and GC content. A list of the genotyping primers for determining the fragment sizes of CA425, HMS2, HTG6 and HTG7 is given in table 3.2. Three different sized alleles were successfully sequenced for all loci and their sizes and repeat motifs are given in table 3.3.

Table 3.1: Sequencing primers designed for sequencing repeat elements of each of the 16 microsatellite markers.*

Marker	**Primer sequence**	**Primer length**	**Tm**	**GC%**	**Fragment length**
AHT4	TACCCAGAGTCGGAGAGCAA	20	56.8°C	55.00%	232bp
	AGCTCCATTCAAGGCAACGTG	21	58.0°C	52.40%	
AHT5	TTCTCTGCTCGCAGATGCAG	20	57.1°C	55.00%	202bp
	GAGTGCAGGCTAAGGAGGCTCAG	23	61.3°C	60.90%	
ASB2	GTGTCGTTTCAGAAGGTCAACC	22	56.1°C	50.00%	280bp
	TCTCTTTGCGCACTTCCCAG	20	57.4°C	55.00%	
HMS6	GTTGAACTGTGTGAAGCTGCCA	22	58.1°C	50.00%	222bp
	TGGAGAGCAACAAAACTCCC	20	55.1°C	50.00%	
ASB17	AAACACAGCCTGCCACCTA	19	56.4°C	52.60%	282bp
	AAGGTCTTGCAGATGGTGCCTC	22	59.4°C	54.50%	
HTG6	CAGATCTCTGGGCATAGAGCA	21	55.8°C	52.40%	245bp
	CTTCCAAAGCAAACCCAAGATC	22	54.8°C	45.50%	
HTG7	ATGGCAGTAGCTGAGGTTTGG	21	57.1°C	52.40%	203bp
	AAAGTGTCTGGGCAGAGCTG	20	57.3°C	55.00%	
HMS2	TGCTAAAAGCTTGCAGTCGA	20	54.5°C	45.00%	238bp
	AAGACACACGGTGGCAACTG	20	57.9°C	55.00%	
ASB23	AGGCCAACTCTCCGTTATGC	20	57.1°C	55.00%	279bp
	TGTAGCTGTGACCCACACAG	20	56.5°C	55.00%	
VHL20	GAACTCTGTGTGGTCAATGG	20	53.5°C	50.00%	190bp
	ATACCGCTCATTGGTGCCCA	20	58.7°C	55.00%	
LEX003	AGTGCTGAGACTTCTGAGAG	20	60.0°C	50.00%	129bp
	ATTAGGCAACGGTCAGAAGG	20	61.0°C	50.00%	
CA425	TGTGCTGCGTTCCTACTGCG	20	64.0°C	55.00%	291bp
	TTTGTTGCCGAAGACCCACC	20	65.0°C	55.00%	

HTG10	CGCCCCCACACTCCATAAAT	20	57.0°C	55.00%	1002bp
	AGTGACTTATTGTGGCGA	18	50.0°C	44.00%	
HTG4	ATGTTATTGTGTGGTGCTCT	20	51.0°C	40.00%	805bp
	ATAAAGAACAGGGAGAACGC	20	52.0°C	45.00%	
HMS7	CAGATTGGTGGTTGCCAGAG	20	56.0°C	55.00%	789bp
	GCATCTGGTCCGTCCTACTA	20	55.0°C	55.00%	
HMS3	AGTGCAACCCCAAACATCAG	20	55.0°C	50.00%	597bp
	GCCACCTCACTCCACTATAA	20	53.0°C	50.00%	

* Fragment sizes are based on primer binding to GenBank reference sequences.

Table 3.2: Additional genotyping primers for determining fragment sizes of markers not in routine use.

Marker	**Primer sequence**	**Primer length**	**Tm**	**GC%**	**Fragment length of M allele**
CA425	AGCTGCCTCGTTAATTCA	18	58°C	44%	240bp or 19 dinucleotide repeats
	CTCATGTCCGCTTGTCTC	18	59°C	55%	
HTG6	CCTGCTTGGAGGCTGTGATAAGAT	24	66°C	50%	84bp or 18 dinucleotide repeats
	GTTCACTGAATGTCAAATTCTGCT	24	62°C	37%	
HTG7	CCTGAAGCAGAACATCCCTCCTTG	24	67°C	54%	118bp or 17 dinucleotide repeats
	ATAAAGTGTCTGGGCAGAGCTGCT	24	68°C	50%	
HMS2	CTTGCAGTCGAATGTGTATTAAATG	25	60°C	36%	222bp or 15 dinucleotide repeats
	ACGGTGGCAACTGCCAAGGAAG	22	69°C	59%	

Table 3.3: Characteristics and sizes of the repeat elements of the microsatellite markers sequenced

Locus	**GenBank Accession no.**	**Repeat element**	**Alleles sequenced (alphabetical nomenclature)**	**Allele sizes (no. dinucleotide repeat elements)**
AHT4	NW_001867395.1	Compound; $(AC)_n AT(AC)_n$	J;M;O	27;30;32
AHT5	NW_001875797.1	Simple; $(GT)_n$	J;M;N	16;19;20
ASB2	NW_001867379.1	Simple; $(GT)_n$	K;M;Q	18;20;24
HMS3	NW_001867432.1	Compound; $(TG)_2(CA)_2TC(CA)_n\ GA(CA)_5$	I;M;P	21;25;28
HMS6	NW_001867412.1	Simple; $(GT)_n$	L;M;P	14;15;18
HMS7	NW_001867387.1	Compound; $(AC)_2(CA)_n$	J;M;O	16;19;21
HTG4	NW_001867432.1	Complex; $(TG)_n\ AT(AG)_5\ AAG(GA)_5\ ACAG(AGGG)_3$	K;M;P	30;32;35
HTG10	NW_001867391.1	Simple and compound; $(TG)_n$ and $TATC(TG)_n$	L;M;O	20;21;23
VHL20	NW_001867407.1	Simple; $(TG)_n$	J;M;R	14;17;22
ASB17	NW_001867402.1	Simple; $(AC)_n$	G;M;Q	14;20;24
ASB23	NW_001867411.1	Simple and compound; $(TG)_n$ and $(TG)_n\ TT(TG)_4$	I;M;V	17;21;30
CA425	NW_001867400.1	Simple; $(GT)_n$	J;M;N	16;19;20
HMS2	NW_001867364.1	Compound; $(CA)_n\ (TC)_2$	K;M;P	18;20;23
HTG6	NW_001867379.1	Simple; $(TG)_n$	G;M;P	12;18;21
HTG7	NW_001867413.1	Simple; $(GT)_n$	K;M;O	15;17;19
LEX003	NW_001877047.1	Simple; $(TG)_n$	F;M;P	13;20;23

A Numerical Nomenclature System

With the successful sequencing of microsatellite alleles, actual allele sizes were matched to alphabetical allele designations (table 3.4) and a nomenclature system based on the number of dinucleotide repeat elements found within the microsatellite repeat was set up. The sizes of alleles not sequenced were extrapolated based on allele motilities during genotyping and considering the dinucleotide nature of the repeat element.

Table 3.4: Conversion of alphabetical nomenclature to repeat – based nome

	Letter nomenclature for alleles																					
Marker	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**
AHT4						24	25	26	**27**	28	29	**30**	31	**32**	33	34	35	36					
AHT5							14	15	**16**	17	18	**19**	**20**	21	22	23	24						
ASB2	9	10	11	12	13	14	15	16	17	**18**	19	**20**	21	22	23	**24**	25	26	27	28			
HMS3						19	20	**21**	22	23	24	**25**	26	27	**28**	29	30	31	32				
HMS6							10	11	12	13	**14**	**15**	16	17	**18**	19							
HMS7									**16**	17	18	**19**	20	**21**	22	23	24	25					
HTG4							27	28	29	**30**	31	**32**	33	34	**35**	36	37						
HTG10			12	13	14	15	16	17	18	19	**20**	**21**	22	**23**	24	25	26	27	28	29			
VHL20								13	**14**	15	16	**17**	18	19	20	21	**22**	23	24				
ASB17			11	12	13	**14**	15	16	17	18	19	**20**	21	22	23	**24**	25	26	27	28	29		
ASB23						15	16	**17**	18	19	20	**21**	22	23	24	25	26	27	28	29	**30**	31	
CA425							14	15	**16**	17	18	**19**	**20**	21	22	23	24						
HMS2							10	11	12	**13**	14	**15**	16	17	**18**	19	20	21	22	23	24		
HTG6					11	**12**	13	14	15	16	17	**18**	19	20	**21**	22	23						
HTG7							12	13	14	**15**	16	**17**	18	**19**	20	21	22						
LEX003				12	**13**	14	15	16	17	18	19	**20**	21	22	**23**	24	25	26	27	28	29	30	31

Discussion

In order to develop a repeat-based nomenclature system van De Goor et al. took a selection of alleles from 35 populations consisting of different breeds found in Europe (n = 9094) and sequenced for 17 polymorphic microsatellite loci. In their study, as in this one, a proposal is presented for repeat-based allele nomenclature in horses on the basis of sequenced alleles. Both studies sequenced three alleles from each locus. Van De Goor et al. sequenced the most prevalent alleles in a variety of different breeds while this study focused on the middle or 'M' allele and two randomly chosen alleles, larger and smaller than the 'M' allele, respectively from Thoroughbred horses only (van De Goor et al., 2009).

Primer extension can be slowed or interrupted by repetitive elements and therefore primers for sequencing through microsatellite repeat elements were designed to anneal close to the microsatellite repeats (table 3.1) in order to sequence through these elements and accurately determine their size. CA425, HMS2, HTG6 and HTG7 were additionally genotyped (table 3.2) in order to support sequence data with fragment lengths observed.

Sequencing revealed compound repeat elements for AHT4, HMS3, HMS7, HTG10, ASB23 and HMS2 and a complex repeat for HTG4 (table 3.3). This is in agreement with the work published by van De Goor (2009) and with reference sequences obtained from GenBank. Allele designation for compound repeats would be interpreted as follows where $(CA)_{10}GA(CA)_5$ is equal to allele 16. For HTG10 an insertion of T after the repeat element was observed for some samples resulting in a sequence of $TATC(TG)_n$*T*CCGG. This would, however, not have an effect on the interpretation of allele size. ASB23 displayed a variant repeat sequence of $(TG)_5AT(TG)_{18}TT(TG)_4$ in a sample with the allele 'U' or 29 repeat units. This variant was only observed in a single sample.

Only HMS3 and ASB23 had single base differences in their repeat elements that were not depicted in GenBank reference sequences; however these anomalies were described by van De Goor (2009). HTG10 and ASB23 were found to have compound or simple repeat elements, an important aspect to consider as alleles that are the same size might not be identical by descent due to the different natures (compound or simple) of the repeat element. MacAvoy et al. found the sequencing of parental microsatellite alleles were important in the initial stages of setting up a microsatellite panel for genotyping of Greenshell^{TM} mussels. Sequence differences in same size alleles were observed for many homozygotes, revealing the presence of more than one lineage (MacAvoy, Wood, & Gardner, 2008). The complexity of the microsatellite repeat element is also important when primer binding is considered. HTG4 has a complex microsatellite pattern of 41bp after its ‘TG’ dinucleotide repeat which must be taken into account when sizing the allele. If primer binding sites were located within this region, it would result in the inability to determine the true allele size of the locus.

Sequenced sizes of the repeat elements correlated well with the fragment lengths for the same alleles observed in STRand version 2.3.94 (University of California). HTG4 had the largest M allele consisting of 32 repeat elements [$(TG)_{15}$ $AT(AG)_5$ $AAG(GA)_5$ $ACAG(AGGG)_3$] while the smallest M allele was observed for HMS2 [$(CA)_{13}(TC)_2$] and HMS6 [$(GT)_{15}$] with 15 dinucleotide repeats. All compound repeats had flanking repetitive elements with fixed sizes and a larger variable repeat element that made up the bulk of the microsatellite. AHT4 is an exception to this rule in that it has a variable ‘AT’ in the middle of the main repeat element of ‘AC’. There appears to be no fixed position for the ‘AT’ sequence and the length of the flanking ‘AC’ sequences are variable. Insufficient samples were sequenced to determine whether this variability is applicable to individuals with same sized alleles or not.

Table 3.4 depicts the conversion of the alphabetical nomenclature system to a numerical, repeat-based system. No intermediate alleles were observed and the sizes of all alleles could be extrapolated based on the three alleles sequenced for each locus. We propose a system of allele calling based on the size of the repeat element as recommended by the ISFG. The numerical nomenclature system will be based on the sequenced length of the repeat element rather than on the fragment length derived from capillary electrophoresis. Hill (2008) show the importance of determining the makeup of a microsatellite sequence when numerical nomenclature systems are used. Upon sequencing human microsatellite alleles, several loci were found to contain compound repeats not previously observed (Hill et al., 2008). Allele ranges and reference alleles had to be adjusted to accommodate the extra bases found in compound repeats and thus allow accurate allele sizing.

Size differences incurred between different machines from different laboratories will not pose a problem as an in-lane size standard (LIZ-500; Applied Biosystems) is run in addition to a positive control with a known genotype. Any discrepancies in allele calling can be corrected by adjusting the bins to suit the positive control. Some authors have suggested the use of an allelic ladder for fragment sizing (van De Goor et al., 2009) however this could prove difficult as the stutter products created with dinucleotide repeat elements would overshadow the allelic ladder.

CHAPTER 4: SEQUENCING MICROSATELLITE FLANKING REGIONS, PRIMER DESIGN AND PANEL VALIDATION FOR A NEW MULTIPLEX GENOTYPING PANEL

Introduction

Genotyping errors generally go unnoticed unless a parentage is excluded as a result thereof. As datasets increase, so too will their associated errors (Bonin et al., 2004). Many errors occur during the scoring of alleles and Luikart (1999) observed that many markers could not be included in their panel simply due to scoring difficulty. Inaccuracies in scoring might be caused by null alleles. Dakin et al. define null alleles as any allele that consistently fails to amplify to detectable levels (Dakin & Avise, 2004). This lack of amplification is most often due to mutations, especially in the primer binding sites. Characterizing polymorphisms in microsatellite flanking regions could aid in solving this problem. MacAvoy (2008) found that sequencing also revealed much size homoplasy in parental alleles due to differences in flanking regions or in the repeat element itself. Sequencing also revealed discrepancies in allele sizing for certain loci (MacAvoy et al., 2008).

Markers used for parentage testing in humans are thoroughly studied before being used commercially. Many reports describe sequence anomalies such as deletions in microsatellite flanking regions (Divne et al., 2010 & Park et al., 2008) that lead to null alleles or discrepancies in allele sizing. A study by Deucher et al. describes a case in which the profile from a fetus did not match that of its mother in a study on maternal cell contamination (Deucher et al., 2010). Designing primers that bind externally to the old primers showed that the initial mismatch in profiles was due to sequence variation in the old primer binding sites and a resultant lack of amplification of the STR.

A good genotyping panel should be cost effective, fast to set up and easy to visualize. Amplifying all markers in a multiplex PCR and visualization through capillary electrophoresis is quick, easy and automated. Microsatellite markers in horses have been used extensively in varying combinations. The original primers were intended, mostly, for linkage studies and were never designed to work together in a multiplex PCR. In horses there are many instances where these markers fail to amplify the product effectively or where alleles are difficult to discern. The original published primers are used most often; however some laboratories have redesigned primers for certain loci in order to obtain better amplification or easier allele calling. With a better understanding of sequence anomalies in the flanking regions, new primers can be designed that will provide a consistent and reliable primer set for genotyping.

Microsatellite flanking regions are becoming an important topic in many studies (MacAvoy et al., 2008) aimed at a better understanding of microsatellite structure in and around the repeat element. In this study, primers were designed to sequence the flanking regions of only five microsatellite markers in horses. Markers were chosen that have been associated with difficulties in allele calling (HMS3 and HTG10). Additional markers known to amplify well and read easily were also selected for comparison (VHL20, HTG4, and HMS7). Similarities in areas of sequence variation were observed in the five markers sequenced and therefore not all 17 loci were sequenced as it can be assumed that areas of sequence variation will be similar in these markers as well.

Methods and materials

Sequencing Flanking Regions

Microsatellite flanking regions were sequenced for the markers HTG10, VHL20, HMS7, HMS3 and HTG4 only. Samples from a variety of different horse breeds were used as listed in table 4.1. Samples consisted of either fresh blood in EDTA, or hair pulled from the tail collected manually from the Onderstepoort Teaching Animal Unit (OTAU) and various owners volunteering to participate in the project. Some samples were available as extracts stored at the VGL, Onderstepoort.

DNA was extracted from hair by incubating six hair roots for 15 minutes at 97°C in a solution of 200mM Sodium Hydroxide (NaOH; Sigma®-Aldrich, St. Louis; MO) after which a solution of 200mM Hydrochloric Acid (HCl; Saarchem, MERCK, Midrand; Gauteng) and 100mM Tris-HCl, pH 8.5 (Tris base; Promega, Madison; WI) was added. Samples were stored at -20°C until used.

Table 4.1: Sample List

Breed	**No. samples**	**Origin**	**Sample type**
Thoroughbred (Tb)	16	VGL	DNA extract
Arab (Ar)	10	VGL	DNA extract
Warmblood (Wb)	10	VGL	DNA extract
Nooitgedacht (N)	20	OTAU	Hair
Quarter Horse (QH)	6	Volunteer	Hair
Appaloosa (Ap)	5	VGL	DNA extract Blood
Clydesdale (Cl)	6	VGL Volunteer	Blood Hair
Welsh Pony (We)	10	Volunteer	Hair
Lusitano (Lu)	17	Volunteer	Hair
American Saddle Horse (Sa)	10	VGL	DNA extract
Friesian (Fr)	13	VGL	DNA extract
Miniature horse (M)	20	Volunteer	Hair

DNA was extracted from 500µl blood using a Phenol-Chloroform (PCIA; Sigma®-Aldrich, St. Louis; MO) based extraction method with Ethanol precipitation. Prior to Phenol-Chloroform extraction, blood was washed twice in Red Blood Cell Lysis solution (10 mM NaCl, 10mM EDTA, pH 7; Promega, Madison; WI) followed by incubation of the pellet at 56°C for 2 hours in White Blood Cell Lysis solution (10mM Tris-HCl, pH 8; 10mM EDTA and 50mM NaCl; Promega, Madison; WI), 20% Sodium Dodecyl Sulphate (SDS; BDH Laboratory Supplies, Poole; England) and Proteinase-K (Sigma®-Aldrich, St. Louis; MO). Extracted DNA was resuspended in 50µl Tris-EDTA (TE; Promega, Madison; WI) buffer and kept at -20°C.

Novel primers for sequencing the microsatellite flanking regions were designed using FastPCR (Primer Digital Ltd. Version 5.2.118; 2008) or Primer Designer 4 (SciEd Central Version 4.20) in order to anneal 100bp to 400bp from the repeat element (table 4.2). GenBank reference sequences were used for primer design. Primers for sequencing were obtained from Integrated DNA Technologies; Whitehead Scientific (Pty) Ltd. PCR reactions were carried out using 1x PCR Gold Buffer and 1.5mM MgCl (Applied Biosystems; Roche, Branchburg; New Jersey), 0.5mM dNTP mix (Thermo Fisher Scientific Inc., Epsom; Surrey) and 2.5U Super-Therm GOLD Taq Polymerase (Southern Cross Biotechnology, Claremont; Cape Town) in 20µl reaction volumes. The PCR conditions on the Veriti 96-Well Thermal Cycler (Applied Biosystems, Warrington; UK) were as follows: 95°C for 10min; 35 cycles consisting of 95°C for 45sec, 60°C for 1min 15sec, 72°C for 2min and a final extension of 72°C for 30min. After the final extension step, samples were held at 4°C until used. PCR products were purified using the MSB® Spin PCRapace kit by Invitek (Invisorb®; Berlin LOT BP080040; Ref 10202203).

Sequencing reactions were carried out using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Warrington; UK) and reactions were cleaned using a Sodium Acetate (Amresco®, Solon; Ohio) and Ethanol-based cleanup before being sequenced on the ABI 3130x Genetic analyzer in 10µl Hi-Di^{TM} Formamide (Applied Biosystems, Warrington; UK). Results were visualized using Applied Biosystems Sequencing Analysis software v5.2.

Multiplex Panel Design

The strategy used to design primers for multiplex PCR was based on a report by Schoske (2003). GenBank sequences were used in order to design the primers (www.ncbi.nih.nlm.gov). The choice of primer binding sites on the GenBank reference sequence was influenced by the sequenced size of the M allele and the flanking region sequences obtained in this project. Markers with the most repeats in the M allele or with the greatest number of alleles made up the largest fragments in the multiplex PCR. In addition primers were designed so as not to anneal closer than 20bp from the dinucleotide repeat as this might encourage slippage and sequencing results showed these regions to be highly polymorphic (MacAvoy et al., 2008). Markers were spatially arranged in size so as not to overlap when labeled with the same fluorescent dye. and fragment sizes were made large enough to account for any possible, undiscovered alleles at the ends.

Fluorescently-labeled primers for genotyping (Applied Biosystems, Warrington; UK) were designed using either Primer Designer 4 (Primer Designer 4 by SciEd Central Version 4.20) or FastPCR (FastPCR Professional by Primer Digital Ltd. Version 5.2.118). Forward primers were labeled with one of four different dyes; 6-FAM, VIC, NED or PET and 7bp AB tail sequences (Applied Biosystems; P/N: 4304979) were added to the 5' end of reverse primers. All primers were made up to a concentration of 100µM stock solution in TE buffer (Promega, Madison; WI) and further diluted to 10µM or 5µM working solution with molecular grade water. The size standard LIZ-500 (Applied Biosystems) was used with all genotyping reactions. Primers were designed in order to meet the following criteria:

- GC content 45% to 50%
- Tm 55°C to 60°C
- Primer length 18bp to 22bp
- 3' end must not contain 3 or more consecutive G or C bases
- Test to avoid self complimentarity, primer dimers or hairpins (Autodimer software used; www.cstl.nist.gov/biotech/strbase; Vallone & Butler, 2004)

Laboratory Validation

Newly designed multiplex genotyping primers were amplified in singleplex PCR reactions and separated by Agarose gel electrophoresis (on a 1% Agarose gel, Amresco®; Solon, Ohio) as well as Capillary Electrophoresis (3130xl Genetic Analyzer; Applied Biosystems) in order to ascertain primer quality and functionality. 0.4µM of each primer pair was used per sample in 20µl reaction volumes with 2xBuffer; 3mM MgCl, 1mM dNTP mix and 5U Taq Polymerase per sample on a PCR program of 95°C for 10min followed by 35 cycles of 95°C for 45sec; 60°C for 1min 15sec; 72°C for 2min; final extension step of 72°C for 30min and held at 4°C until used. Capillary electrophoresis was carried out using virtual filter set G5 to differentiate between the spectral compositions of the different fluorescent dyes and samples were run using module FragmentAnalysis36_POP7.

Once all primers were verified to be working, a multiplex PCR reaction was set up using all primers together at uniform concentrations of 0.1µM per primer in 10µl reaction volumes. Five randomly chosen Thoroughbred samples were used and the amplified fragments were run on the 3130xl Genetic Analyzer (Applied Biosystems) with LIZ-500 size standard and HiDi-Formamyde (Applied Biosystems). Genotyping data was analyzed on STRand version 2.3.94 (University of California).

Results

Sequencing flanking regions

Novel primers were designed in order to sequence the microsatellite flanking regions of five microsatellite markers. The primer sequences with corresponding melting temperatures and GC content are listed in Table 4.2.

The newly designed primers were designed to sequence between 100 and 400bp of flanking sequence and therefore product sizes range from 597bp to 1098bp. Table 4.3 depicts the primer binding regions in relation to the repeat element as well as the sequence anomalies observed when sequences were compared to GenBank reference sequences (www.ncbi.nlm.nih.gov). HTG4 and VHL20 were the only loci that did not have any sequence variation in the bases immediately flanking the repeat element (table 4.3). HTG4 contains mononucleotide repeats, especially of Thymidine in the 5' region flanking the repeat element.

Table 4.2: Sequencing primers designed for sequencing of flanking regions for five microsatellite markers.

Marker	**Primer sequence**	**Primer length**	**Tm**	**GC%**	**Fragment size of M allele**
VHL20	AGATGAACAGTTAGAGAGCGG	21	53.7°C	47.60%	1098bp
	ACTTCCTCACCACCCTCATA	20	54.4°C	50.00%	
HTG10	CGCCCCCACACTCCATAAAT	20	57.0°C	55.00%	1002bp
	AGTGACTTATTGTGGCGA	18	50.0°C	44.00%	
HTG4	ATGTTATTGTGTGGTGCTCT	20	51.0°C	40.00%	805bp
	ATAAAGAACAGGGAGAACGC	20	52.0°C	45.00%	
HMS7	CAGATTGGTGGTTGCCAGAG	20	56.0°C	55.00%	789bp
	GCATCTGGTCCGTCCTACTA	20	55.0°C	55.00%	
HMS3	AGTGCAACCCCAAACATCAG	20	55.0°C	50.00%	597bp
	GCCACCTCACTCCACTATAA	20	53.0°C	50.00%	

Table 4.3: Sequence anomalies observed in microsatellite flanking regions when compared to GenBank sequences.

Marker	Sequence
HTG10	**cgcccccacactccataaat**tactgaaggagtaaaaagacccacatttagtctttgg(a)^{1}aataccttaatttgctaagtaagataatacagaattg(aaaaca)^{1}aaaacaaaccaaggagaaatgcttttgttgaatatgacctcactatttcactagtgctacaatgtgtgtaag^{8}aaagaaataaagta^{5}aa$^{5;4;8}$aattttacatcagaaa$^{2;6}$atatagagtggaataggccagtttggtcctttgaacatgttatttcagttagaatgattagaggatacaatggcttaaatggaattgca$^{2;5;6;8;9}$ttaaaaacttcagccatggaacaagtttcatgtatttagaatgtaaattgttaatgcattttta^{8}ttgggctttttattctgatctgtcacatttgaattaactgacttt(atc)$^{1;2;5;4;6;7;8;9}$**(tg)$_n$t**$^{2;8+gc}$gc$^{1;2;5;4;7;8}$cgggggtggggcgggaattgg^{4}ccatttgtaataaactttatttcaaggctttccatatgacattaggaagtgtttaactaaagaaaatgttttattgctttaagaca^{11}ggttcttatacagatgtattttcatcatcctttagagt^{2}gctctt^{2}gtttgggacaaagttc$^{3;7;11}$tgagtcatatccagaaatgagaaataga$^{2;3;4}$cagagggccctatagag^{11}ctgatgttaatattttaagggttgttaatattttatcacaatgtta(t)tttt(g)$^{1;2;3;4;5;6;7;8;10;11;12}$tttgcattgtatgtgacactgaatggtaaatttttaaaatgtgcttttaa^{7}aatgtg^{8}atttttattgttgaaaaatcagattggcttaataatt^{7}ttcttcat$^{3;4;7;8}$ggataattttgttagcttacatcta$^{3;5;4}$ctctctaaattgaaagtcta^{8}ct$^{2;5;4;8;11}$agtacaaatgggtgatctgtaaataaggtttaattttctttatcgaggtataattgacatataacattatgttagtttcaggtgtacaacataatgatttgatatttatacatattgcaatatga**tcgccacaataagtcact**
HTG4	**atgttattgtgtggtgctct**ccaatccctgcacactttttcctttgtttccacagctcaacaaactcaactttcccttatactcctttccaatgagc^{8}tact^{2}atcaa^{2}tttttg^{7}tttttttaaagacaaattcctatatttttttgttttttaaagatgatttcctatatttattgttgtatttattt^{2}atttttagaaattt^{8}cacgtt^{1}tttgtaactgtgccagtatttttattga^{2}gg^{9}tattgtgtcag$^{1;6}$tt$^{5;6;9}$ggggc^{5}tctggttagaacat$^{2;5;6;7;8;9;10}$g$^{5;6;7}$acacagg^{6}ttctgaatggtctgt$^{2;8}$c$^{6;10}$cttag$^{1;2;4;5;6;7;8;10}$ctttattttc^{5}cca^{2}taactcccatcattta^{2}c^{9}tata^{2}atattccttctatc^{6}tcag$^{1;5;6;8;9}$tcttgattgcagg^{7}a^{1}caatgagca^{2}ggaagg$^{1;6;9}$ccag^{6}ggtttccagaggtt(**tg)$_n$atagagagag**^{2}**agaaggagagagagaacagagggagggaggg**agagctgctccagaaagccaaggactaaaatac$^{1;2;4;5;6;7;8;10;11}$at$^{1;2;4;5;6;7;8;10;11}$ga$^{3;9}$cca^{9}cattagctcttgctgctttaacttgttcata$^{1;2;4;5;6;7;8;10;11}$gtgt^{2}tcaatcaacattcaagggttatagctattcacaaaggagc^{3}caaggtttgaa^{3}gattgggaaag$^{1;2;3;4;5;6;8;10;11}$aaggc^{2}cacatttagacattaaaatctggtctcatat^{5}tcaacccatttttt^{5}ccag$^{2;6}$aa^{4}aca^{5}taggtgagatttatttttatgaagatgctc$^{2;8}$atatatag$^{1;2}$ttttcaacaaatgtcaacata^{2}caattcagt$^{2;5}$taaacaagaaccaagaaggttgccttgtctcatagggttgcaaaaactga**gcgttctccctgttctttat**
HMS3	**agtgcaaccccaaacatcag**atagtgataatgccgtggaaaaaataaagcgtggta^{1}agatgg^{1}ccatgtcatgccagatg^{11}c^{1}aattgatc$^{2;7;8}$gttattttatggg^{11}gg^{1}ct^{11}gtt^{11}gctgc(gggaaaa^{2}g)^{2}tac^{1}acaa$^{1;7;11}$cc^{1}ttggaaatactgatgtatgactttcattgttcattcaagatctgcatgaagaaagaaatgacagcaacaaacatcagtcagaagctgcgaa^{1}ccatttcatctcagttcc^{11}taacacattaatttcatagtttttcctttatttatgaatcaatgctaaaccctccccatcctca^{11}ctttttcactttgttttgtgattcataaaggggatggagga^{12}ccatggatgccagcacg**tgtgcacatcc**$^{1;9}$**a(ca)$_n$ga(ca)$_5$**$^{1;4}$atctag^{2}aaagc$^{1’2}$tgttttcttgttatgtgacaaagag$^{5;7;11}$ttggggctttagagcaagagggaccaggattggaatctcaatttgg$^{4;7}$c^{2}cac^{1}tgaataactttgggaaatcat$^{1;2;7}$ttaagctctctaaacctctgtt^{1}gccca^{1}tttgtaacgtgtaatgatttctgcccatggga**ttatagtggagtgaggtggc**
HMS7	**tgagtatggttgtaaaagggc**agcatgttaagtgtccatcaatgggcgaatgaatgaagaaaaggttgtacacacacacacacacacacacacacacacacagtggaatactattcacccataaaaaagatggaaatcctgccatttgtgacaacatagatgagccttgagggcattaagctaagtgaaataagtcagacagagaaagacaaataccatatgatcgcaatcctatgtggaatcttaaaaacaaaacaaagaacaaaaaatgagctcacagataacagagaacagattgg^{2}tggttgccagaggcaggggttggttggtgggtgaaatgaa^{2}tgaaagtggtcaaaagttataaacttccagttataaaataaataagtcccgggg^{2}atgtaatgcactgcgtgggcaccatagttaatatactgtattgtatatt^{4}tgaaagt^{4}tgctaa$^{2;4}$aagagtagatc^{4}ttaaa^{4}agtt^{4}ctcatcacaagaaaaaaattgtaactttgtgtggtgat^{4}g^{4}gatgttaactagac^{4}ttaagt^{4}gtggt^{4}gatcatttca^{2}caatatacatacacg^{4}tctaatcactatgttgtacacctga^{4}cactaatataatg^{4}ttatacacat^{4}tatatg^{4}tca^{4}tttatatca^{4}cttttttttagg^{4}acagcatgagaggtc^{4}tttgtggtaatgaaactgttcttgaaacataccttga$^{2;4}$ctgttgtggtagatacatgaa^{2}cccag^{4}acgtgacaaaattgcatagaactaaatacac$^{1;2;4}$**(ca)$_n$**ttagtacatgtaatactggtgaaatccaaataagattggtggatggtatcaacatgagtttcctggttgt^{4}gatattttgct^{4}gtag$^{2;4}$ttt^{4}a^{1}taagatgttaccactggaggacgctgggtgaagggt^{4}a^{2}cacaggactg^{4}ctc^{2}tgtattcttacaact^{4}gcctgtgaacctacaattatcccaaaatttaaaagtttaattaaaacaagtagcaaccagaacatattccttctaaatcaactgacctttcagaa**aggttcattgagggctgta**

VHL20	**agatgaacagttagagagcgg**taaggatgcactgtaagggtaaacacaaatcaaaaaagcaaaaaacaaattctctgtgttaataagaaggaaagagatccccct^{2}ccacccg^{1}ct^{1}ttcttatagtatttactttaa^{1}aaaacgtgta^{4}agttctttctctgtctcttgaaaatgtatataaatttcttcaaag$^{1;3;5;6;7;9;11}$caaaataagcctttggctactcttaagact^{4}cag^{8}aactatctctctgaaatgtag^{4}ccatcaaggaagg^{3}t^{3}agcccca^{9}ctatctcccagtttctttggaaggatgaaaacctaacttt^{1}gctggatgcctcattccaaaactactcctgtcataaagatc^{11}tgatgagc^{8}ttacttttcctttggattaagctaattagctaacagatagtcaattcccat^{11}tgccagttgaattgaggatgaactctgtgtggt^{5}ca$^{1;5;7;9}$atggt^{2}gctgtcaagtcctctt^{2}acttgaagactagctattgt$^{1;8}$ttatctt**(tg)**$_{n}$ctgagg$^{8;11}$aagattctccct^{8}gagtta^{8}acatct^{8}gggcta^{11}atcttcctc$^{7;8}$tattttgtatatggctt^{4}gctgccacagcatgg^{11}gcaccaatgag^{1}cggtatgggt^{12}ccgtgcccaggaacggaacccaggcccctgaggcagt^{9}gcgc$^{1;2;3;4;5;7;8;9;10;11;12}$gctgaacttaaccacaag^{4}gccaccaggccagcccttacctattgtttatcttaaggacatttaagtaatgggtt^{4}gt^{7}atc^{12}tgcttggctatataacagggtgaggtttcttctgtttttgcaatcttttaggggattacctgtgatgtgaatcacctctggattgacacttgttcaataacaaacctttttctctttc^{7}tcttctacttggtggagaggtttactgggttgagaggagatttacttttaat$^{2;4;5;8}$tatatttccccaacagtgttcaa^{4}tattgttgctaacaacgttaaacttt$^{1;5}$atctgagccatgtgttttgaaaaagcagcaactgttaagaaatatcctcaccttttgtgttctt$^{1;5;8;9}$ggaaagggctgatcaagaaggaccaccacaaattctgaaataagtcctgtccgtat**acttcctcaccaccctcata**

Key: Deletions are indicated by the base(s) in brackets; insertions are indicated by '+'; base substitutions are marked with a numerical superscript specific to the breed in which it was observed. Original, published primers are underlined and sequencing primers for flanking regions as well as repeat elements are in bold.

Legend for breed specific numbering of base substitutions: Thoroughbred (1); Nooitgedacht (2); Quarter Horse (3); Namibian Warmblood (4); Lusitano (5); Arabian Horse (6); Warmblood (4); Welsh Ponies (7); American Saddle Horse (8); Friesian Horse (9); Miniature Ponies (10); Clydesdale Horse (11); Appaloosa (12)

Multiplex Panel Design

Novel primers were designed for 16 microsatellite loci and the sexing marker, Amelogenin. The marker sizes as well as fluorescent label are depicted in table 4.4. Reverse primers were tailed with a 7bp sequence (AB tail; Applied Biosystems). Markers were designed so as not to overlap in size unless different fluorescent labels could be used. Fragment sizes were decided upon based on the size of the repeat element as well as the number of alleles. Markers HTG10, HMS3, AHT4 and CA425 that had the greatest number of alleles or the largest repeats were designed to be the largest fragments. Table 4.5 depicts the primer binding sites for each locus in relation to the repeat element based on sequence data obtained from GenBank (www.ncbi.nlm.nih.gov).

Table 4.4: Newly designed primer sequences for 16 microsatellite markers and sexing marker, Amelogenin, with corresponding fragment sizes and fluorescent labels for use in a single multiplex panel for genotyping horses.

Marker	**Primer sequence**	**Frament size (M allele)**	**Smallest Fragment**	**Largest fragment**	**Fluorescent label**
VHL	F: ggtcaatggtgctgtcaagtcc R: GTGTCTTgcccagatgttaactcaggga	119 bp	107 bp	144 bp	6FAM
HTG4	F: cttgattgcaggacaatgag R: GTGTCTTaagcagcaagagctaatgtg	176 bp	162 bp	197 bp	
HMS2	F: gaatgctaaaagcttgcagtcg R: GTGTCTTcaactgccaaggaagccacta	227 bp	213 bp	256 bp	
HTG10	F: cttcagccatggaacaag R: GTGTCTTtgtcccaaacaagagcactc	307 bp	285 bp	336 bp	
CA425	F: ggctactgcaactttcagca R: GTGTCTTtagcatttggacagccccaa	378 bp	364 bp	399 bp	
ASB17	F: taaccaggcagcagtcagga R: GTGTCTTctgtggagtttgagtatcgctg	139 bp	117 bp	168 bp	VIC
AMEL	F: aacaccaccagccaaacctc R: GTGTCTTttccagaggcaggtcaggaag cat		170 bp	194 bp	
HTG6	F: cttcatgagcttcctgcttgg R: GTGTCTTtctggaatccctctcagctc	228 bp	210 bp	249 bp	
HMS3	F: atggaggaccatggatgcca R: GTGTCTTtcgaaagtgagctagccacctc	275 bp	259 bp	300 bp	
AHT4	F: ccccaactgagaatgtttggca R: GTGTCTTctccattcaaggcaacgtgg	373 bp	357 bp	396 bp	
HMS6	F: gaactgtgtgaagctgccagta R: GTGTCTTagttttccagctccatcttgtgaa gtg	180 bp	166 bp	199 bp	NED
HMS7	F: catgaacccagacgtgacaa R: GTGTCTTacagagcagtcctgtgtacc	216 bp	206 bp	239 bp	
ASB2	F: aggtcaacctctcggctattgc R: GTGTCTTtctttgcgcacttcccagaa	265 bp	239 bp	292 bp	
LEX	F: agggtacatctaaccagtgctg R: GTGTCTTgttcacatgcttcaccttggca	314 bp	294 bp	347 bp	
ASB23	F: aaggcagcatttgaacccagg R: GTGTCTTacagtcctgtagctgtgaccca	365 bp	349 bp	396 bp	
HTG7	F: ggtttggcaatacttcctggga R: GTGTCTTtaaagtgtctgggcagagctgc	190 bp	176 bp	211 bp	PET
AHT5	F: cttctctgctcgcagatgca R: GTGTCTTagcacccaagtttccagaggta	289 bp	275 bp	310 bp	

Table 4.5: Layout of newly designed genotyping primers on sequence and in relation to the microsatellite repeat

AHT4 NW_001867395.1
CCCCAACTGAGAATGTTTGGCAAAGAAATCATTTGATTAGAGAAGGTACGGGTTTCTGTGTTATAAGAAGTCAAATATGAACAGCAGGAATTCCGGAGGCCACAGAGGGAAGAGTTCCAAAGAGGCATTGGGAGGCAGCTGGCTACCCAGAGTCGGAGAGCAACCGCCTGAGCAAGGAAGTCCTAGCCTTAGGAATAAAATTGGCAGAAT(AC)$_n$AT(AC)$_n$AGAGCTGCTAGAAGAGCTGGGGCTGACCCAGGGTAAACTCTCTGGGAGCCTTATTATTTCGGGAAGGTGTTGTAAAGACCAGCCC**CCACGTTGCCTTGAATGGAG**CT

AHT5 NW_001875797.1
C**TTCTCTGCTCGCAGATGCA**GCCCGAAAACCTCCCAGCGGTGTCCCAGCGCCTCCAACCAGCCACGGACACATCCCTGCCTGCACTGCCCCTCTCCCCTC(GT)$_n$ATGTTTGGAGGATCCCCCAAGACATGTGGGAGGGGGCGAGGGCTGAGCCTCCTTAGCCTGCACTCCCCCCACCCCCCATGTCCTCGGAGGCTTAGAGGGGGAAGTCAGAGTACTAGACGCTAGCTCCTC**TACCTCTGGAAACTTGGGTGCT**

ASB2 NW_001867379.1
AGGTCAACCTCTCGGCTATTGCCTCAATTTTACTCTTTGGGATCTCCTTCCTGTAGTTTAAGCTTCTGAATC(GT)$_n$AGACATTGGGAACATTAGCTAAGAGTCTCAATTCTCAAATTTTTGTTTCTCAAACTTTTTCCTCACTTTGAATGACAGAGACTTAACTCCTATCAGAGAACTCAGTTTTGTTACATTTAACAAGCAAAATAAC**TTCTGGGAAGTGCGCAAAGA**

HMS3 NW_001867432.1
ATGGAGGACCATGGATGCCAGCACG(TG)$_2$(CA)$_2$TC(CA)$_n$GA(CA)$_5$ATCTAGAAAGCTGTTTTCTTGTTATGTGACAAAGAGTTGGGGCTTTAGAGCAAGAGGGACCAGGATTGGAATCTCAATTTGGCCACTGAATAACTTTGGGAAATCATTTAAGCTCTCTAAACCTCTGTTGCCCATTTGTAACGTGTAATGATTTCTGCCCATGGGATTATAGTGGAGT**GAGGTGGCTAGCTCACTTTCGA**

HMS6 NW_001867412.1
GAACTGTGTGAAGCTGCCAGTATTCAACCATTGGCACTTTTTTGTGGTTTATCTTAAAAATTATTCTTCAAATCAGAAACCCATATAGAATTATATGTAAGGACGAGTAA(GT)$_n$AACTTTTGAGTTA**CACTTCACAAGATGGAGCTGGAAAACT**

HMS7 NW_001867387.1
CATGAACCCAGACGTGACAAAATTGCATAGAACTAAAT(AC)$_2$(CA)$_n$TTAGTACATGTAATACTGGTGAAATCCAAATAAGATTGGTGGATGGTATCAACATGAGTTTCCTGGTTGTGATATTTTGCTGTAGTTTATAAGATGTTACCACTGGAGGACGCTGGGTGAAG**GGTACACAGGACTGCTCTGT**

HTG4 NW_001867432.1
CTTGATTGCAGGACAATGAGCAGGAAGGCCAGGGTTTCCAGAGGTT(TG)$_n$AT(AG)$_5$ACAG(AGGG)3AGAGCTGCTCCAGAAAGCCAAGGACTAAAATACATGAC**CACATTAGCTCTTGCTGCTT**

HTG10 NW_001867391.1
CTTCAGCCATGGAACAAGTTTCATGTATTTAGAATGTAAATTGTTAATGCATTTTATTGGGCTTTTTATTCTGATCTGTCACATTTGAATTAACTGACTTTATC(TG)$_n$CCGGGGGTGGGGCGGGAATTGGCCATTTGTAATAAACTTTATTTCAAGGCTTTCCATATGACATTAGGAAGTGTTTAACTAAAGAAAATGTTTTATTGCTTTAAGACAGGTTCTTATACAGATGTATTTTCATCATCCTTTA**GAGTGCTCTTGTTTGGGACA**

VHL20 NW_001867407.1
GGTCAATGGTGCTGTCAAGTCCTCTTACTTGAAGACTAGCTATTGTTTATCTT(TG)$_n$CTGAGGAAGATTC**TCCCTGAGTTAACATCTGGGC**

HTG6 NW_001867379.1
CTTCATGAGCTTCCTGCTTGGAGGCTGTGATAAGATAC(CA)$_n$AATGCTAAAGAGCAGAATTTGACATTCAGTGAACTGACACAAGGAAGGGCACAGACAGTACTGAGATATAGGGAAAAGTCTTTGATCTTGGGTTTGCTTTGGAAGATTTCACAAAGGGGATGAGGCTTGCATGGGGCCTTGAGGAT**GAGCTGAGAGGGATTCCAGA**

HTG7 NW_001867413.1
GGTTTGGCAATACTTCCTGGGAAGAGGCAGGGAGGGAGGTAATAGGATCTGATCCAAGAGAGGTAGTGGCCAGCCTGAAGCAGAACATCCCTCCTTGTCGCA(GT)$_n$GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTCTGTTAGGGGGAGGACAGGGTGGAAGAGTCCGTGTA**GCAGCTCTGCCCAGACACTTTA**

HMS2 NW_001867364.1
GAATGCTAAAAGCTTGCAGTCGAATGTGTATTAAATGACTGTATTTGCTATGAAAAACTGGAACCTCTGTTCTTAATGAATCCTTTATGGAACATATAGTTATGTTTT(CA)$_n$(TC)$_2$CTGATGAGAAGCAGTACTCTTGTAAGAAATTATTTTTTTCTTTGAAAGATTTGGAAAAGGGGTG**TAGTGGCTTCCTTGGCAGTTG**

ASB17 NW_001867402.1
TAACCAGGCAGCAGTCAGGATCTCCACCGGAAGAGTCT(AC)$_n$CCCACTTAATTTTCAAGGTACAAAGGTACCGCCCTCCAT**CAGCGATACTCAAACTCCACAG**

ASB23 NW_001867411.1
AAGGCAGCATTTGAACCCAGGCTCCAGAGCCCCACAACTCACAACATGCTTATCAGTAGGCTCTGCCGTCCAGCTACCCAGGCCAACTCTCCGTTATGCTCAGCCTTTATCTCCCCTTTTCAACTTTTATGCAACTTGCAGGTGGAGGAGGTTTGTAATTGGAATGGAATGTATGAAATGCAAGGATGAAGAGGGCAGCAGGTTGGGAAGGAGGCTGGACTCCCGAGC(TG)$_n$TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGGTAGAGGTTGCAGGTGTTAAAAATGACTTCTCATCTAACCCACCAGGGCAAGAGCATGTCCCCCCGGGAGCTGTG**TGGGTCACAGCTACAGGACTGT**

LEX3 NW_001877047.1
AGGGTACATCTAACCAGTGCTGAGACTTCTGAGAGACACTCACTC(TG)$_n$TTTATCCAATATTATGTTTGGGTTTTTTTAATCTTTTATTTTAATCCGTTGCCAGTCTTCCTCCTTTTTTTCCTTCCCAAAACCCCCCAGTACTTAGTTGTATATCCTAGTTGTAGGTTCTAGTTCTTCTATGTGAGACACCTCCTCAGCATGGCTTGATGAGCAGTGCTAGGTCCGTGTCCAGGATCCAAATGGTCAAAACCTGAGGC**TGCCAAGGTGAAGCATGTGAAC**

CA425 NW_001867400.1
GGCTACTGCAACTTTCAGCAGTTTCCTCACTCTCTGGGTCTGCTGTTCTCTCGCTGCCCAGCAAATCGGTGAAGGGAGCCGACCTAGTCATCCACGCTCACGTTTGTTCTAGAAGCATTTATTGAGAACCTGTGGAATTCTCGGCCTACGATCATGAACTTTCATGAGCACAGCTGCCTCGTTAATTCAGAAGTGTGTGCTGCGTTCCTACTGTGGGGATGGCAGGGTTCCTCCTGCTGGGGCAGGCTGGGCTCTGCTCGCAGGGAGCCGAC(GT)$_n$GGACCCAGCCCGTGGTCAGGGGCTTTGCTGGGGGCACTTGAGCTCTGC**TTGGGGCTGTCCAAATGCTA**

Key: Primer binding sites are depicted in bold

Table 4.6: Comparison of GC content and melting temperature (Tm) between the old, published primers and new primers designed in this study.

Marker	**GC content old (Forward & Reverse)**	**GC content new (Forward & Reverse)**	**Tm old (Forward & Reverse)**	**Tm new (Forward & Reverse)**	**Other problems associated with original published primers**
AHT4	55% & 52%	50% & 55%	66°C & 56°C	57°C & 56°C	Sequence runs in forward and reverse primers
AHT5	65% & 70%	55% & 50%	69°C & 70°C	57°C & 57°C	Reverse primer dimers. False priming and sequence runs
ASB2	40% & 40%	54% & 50%	59°C & 58°C	58°C & 56°C	Sequence runs forward and reverse primers
HMS3	37% & 37%	55% & 54%	61°C & 61°C	58°C & 58°C	Sequence runs forward and reverse primers. Reverse primer dimers
HMS6	45% & 45%	50% & 44%	64°C & 63°C	57°C & 59°C	Forward primer dimers
HMS7	36% & 41%	50% & 55%	62°C & 61°C	54°C & 56°C	Primer dimers & sequence runs Forward and reverse primers
HTG4	45% & 63%	45% & 45%	63°C & 67°C	58°C & 60°C	Forward primer dimers & sequencing runs in reverse primer
HTG10	25% & 73%	50% & 50%	56°C & 77°C	57°C & 55°C	Sequencing runs forward and reverse primers
VHL20	41% & 47%	54% & 52%	60°C & 63°C	58°C & 56°C	Forward and reverse primer runs
HTG6	50% & 37%	52% & 55%	66°C & 62°C	56°C & 55°C	Reverse primer runs
HTG7	54% & 50%	50% & 54%	67°C & 68°C	57°C & 59°C	Reverse primer dimers and sequence runs
HMS2	36% & 59%	45% & 52%	60°C & 69°C	55°C & 57°C	Sequence runs in forward primer
ASB17	60% & 60%	55% & 50%	65°C & 64°C	57°C & 55°C	Reverse primer dimers and sequence runs
ASB23	37% & 33%	52% & 54%	60°C & 55°C	58°C & 59°C	Sequence runs in forward and reverse primers
Lex003	45% & 40%	50% & 50%	62°C & 59°C	56°C & 58°C	Forward primer dimers and sequencing runs in reverse primer
CA425	44% & 55%	50% & 50%	58°C & 59°C	55°C & 56°C	

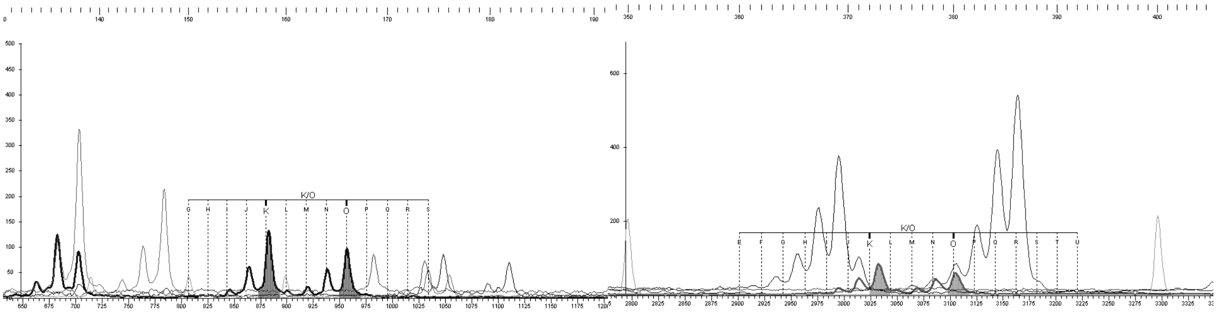
Table 4.6 lists the melting temperatures and GC content of the published primers and compares them to that of the newly designed primers. Other observed problems associated with the published primers are also described. The multiplex genotyping panel is run optimally using good quality DNA extracts (100-500ng/µl; absorbency ratios above 1.6). 0.5 - 1µl of DNA is required for PCR. Table 4.7 depicts the concentrations per sample of reagents for a successful multiplex PCR using the new panel. The final PCR program that was used started with a denaturation step at 95°C for 10min followed by 35 cycles of 95°C for 45sec; 60°C for 1min 15sec; 72°C for 2min and a final extension step of 72°C for 30min. Samples were held at 4°C until used.

Table 4.7: Optimum primer and reagent concentrations for 16 microsatellite markers and Amelogenin in a multiplex PCR reaction using the primers designed in this study.

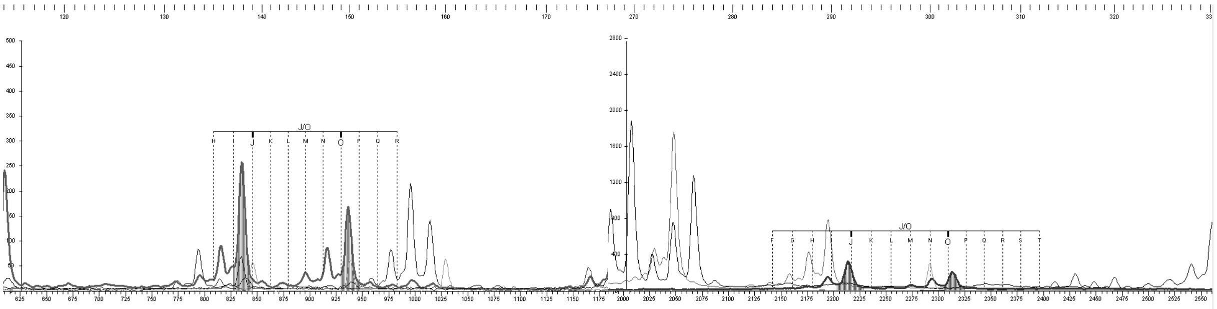
Primer or reagent	**Stock concentration**	**Concentration in 10µl reaction volume**
Lex003	5µM	0.125µM
HMS3	5µM	0.18µM
ASB23	5µM	0.06µM
AHT4	5µM	0.06µM
AHT5	5µM	0.125µM
CA425	5µM	0.125µM
ASB2	5µM	0.125µM
HTG10	5µM	0.125µM
HMS7	5µM	0.125µM
HTG6	5µM	0.125µM
HMS2	5µM	0.25µM
HTG7	5µM	0.06µM
HMS6	5µM	0.125µM
AMEL	5µM	0.03µM
HTG4	5µM	0.25µM
VHL20	5µM	0.06µM
ASB17	5µM	0.06µM
PCR Buffer	10x	2x
MgCl	25mM	3mM
dNTP	20mM	1mM
Taq Polymerase	1000U	5U

Multiplex Panel Validation

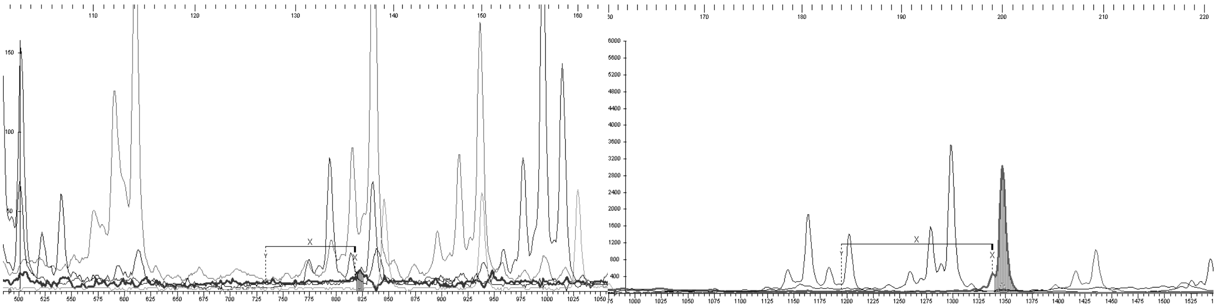
STRand software version 2.3.94 (University of California) was used for the analysis of genotyping results and electropherograms of each locus were compared for the old and new marker panels. Figure 4.1 shows the electropherograms of each locus for a single sample genotyped using both panels. The images depict the fluorescence of the peak on the Y axis and the fragment size in base pairs on the top X axis. Selected markers are highlighted within their respective bins with the genotype of that locus depicted alphabetically above each bin.



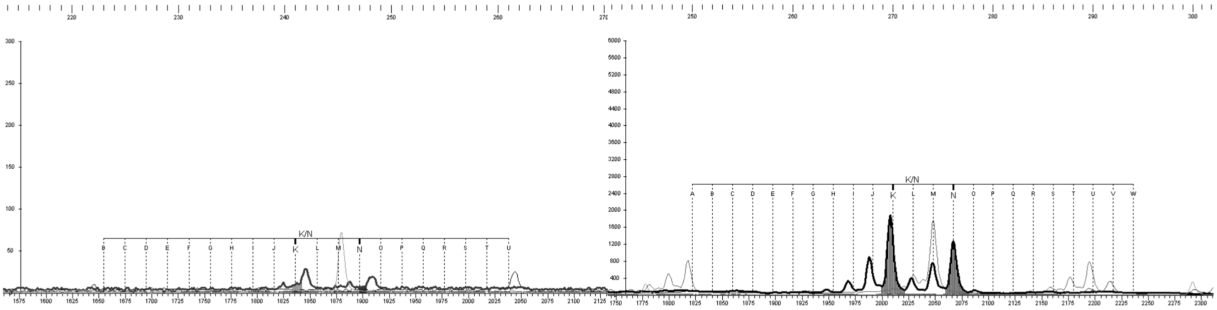
AHT4



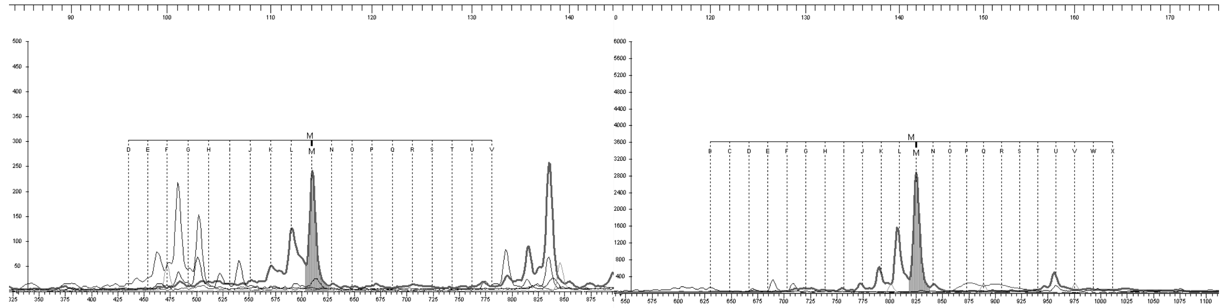
AHT5



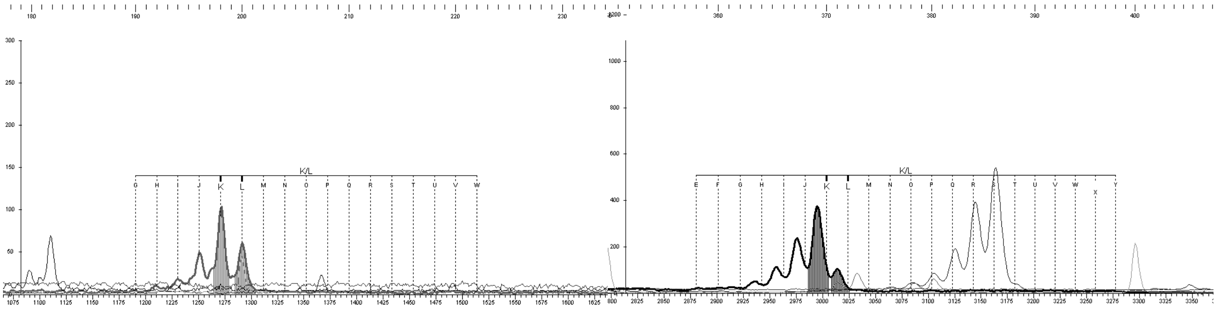
AMEL



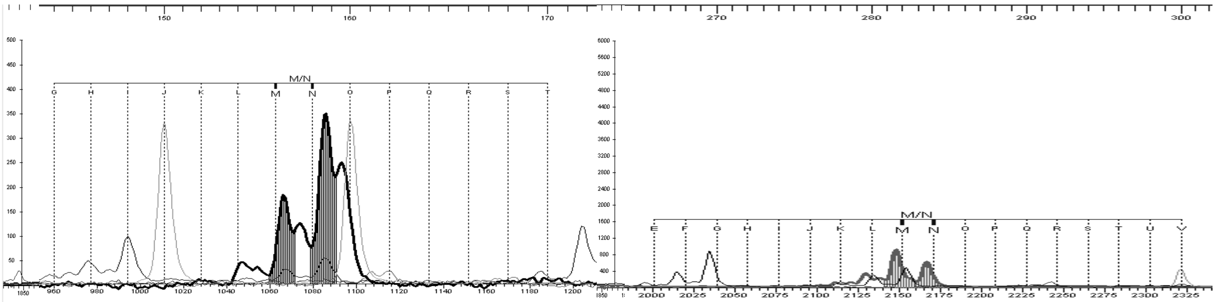
ASB 2



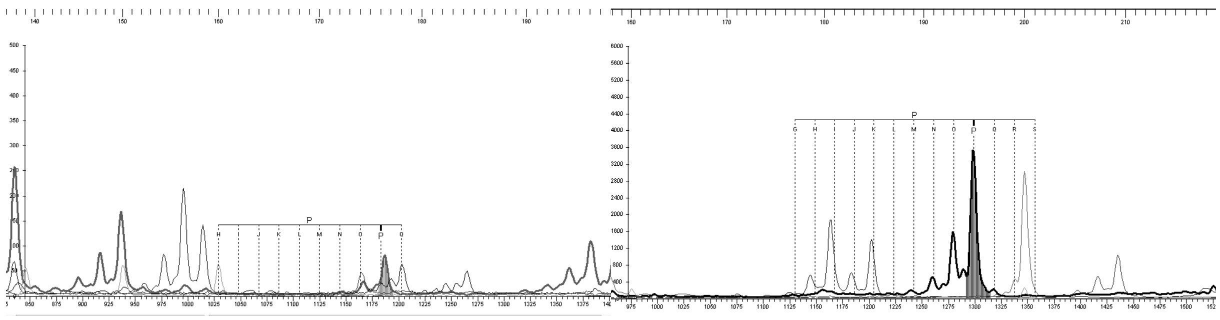
ASB17



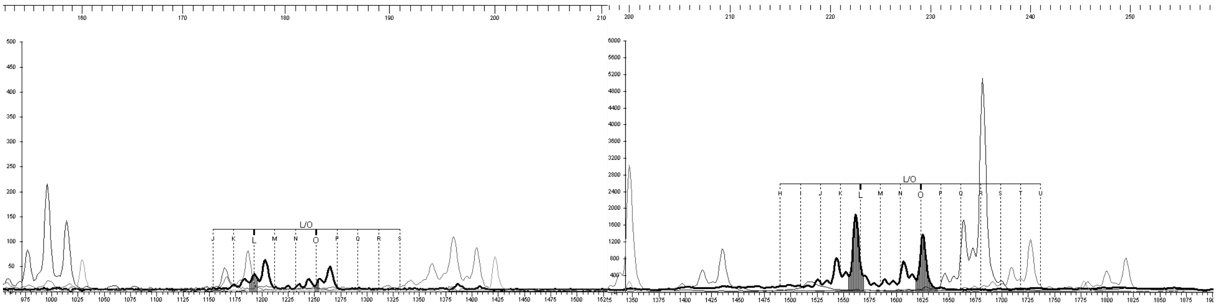
ASB23



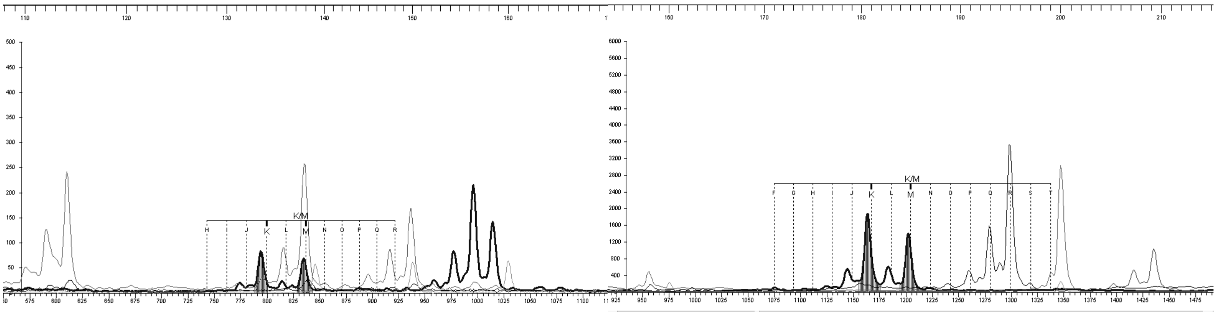
HMS3



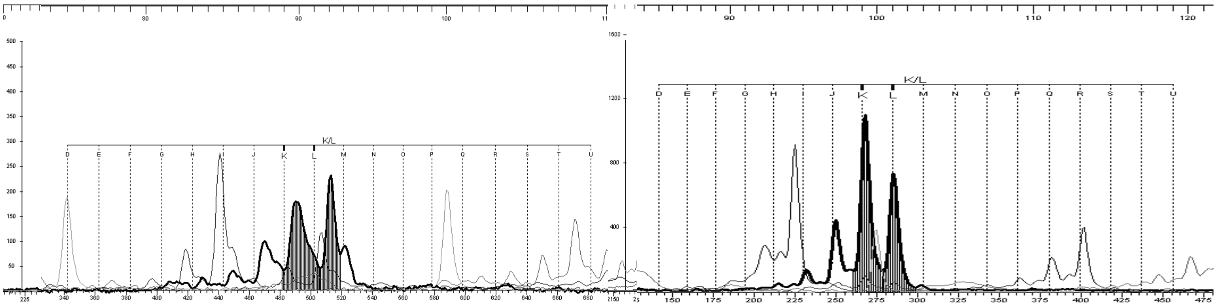
HMS6



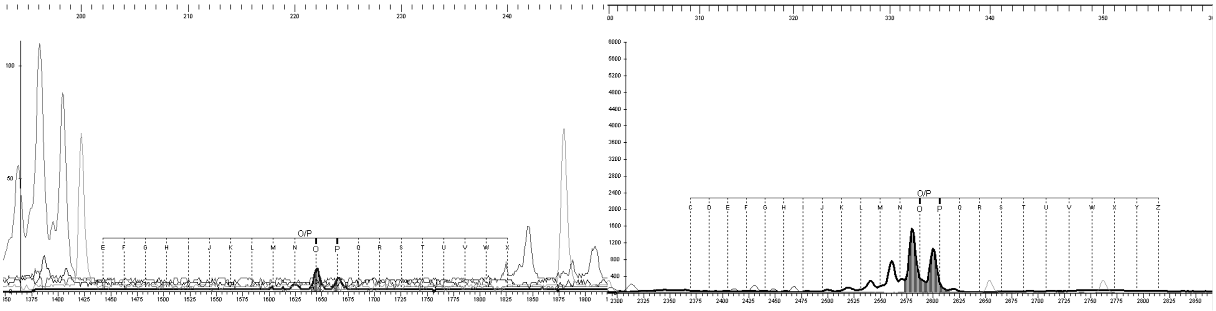
HMS7



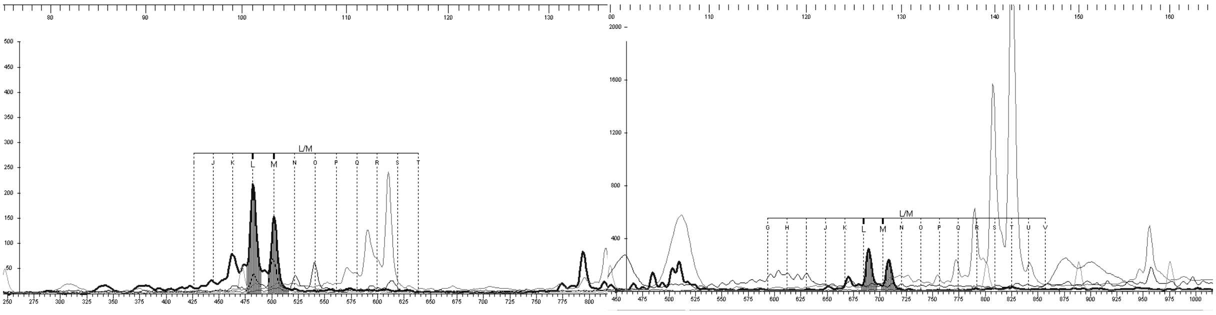
HTG4



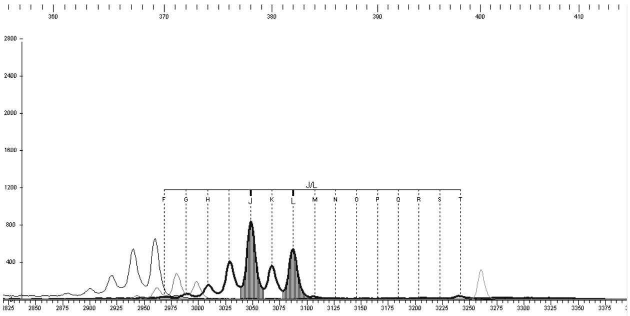
HTG10



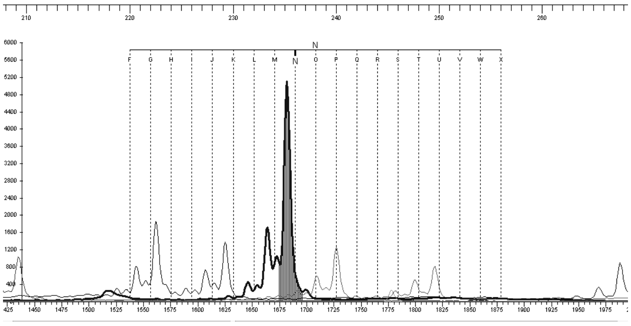
LEX003



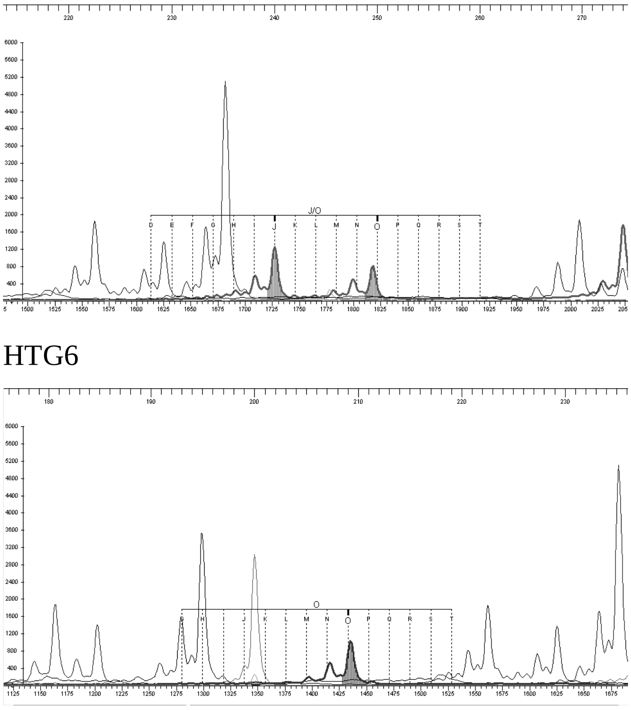
VHL20



CA425



HMS2



HTG6

HTG7

Figure 4.1: Electropherogram images of the loci amplified using the original, published primers (left) and compared to images of loci amplified using the newly designed primers from this study (right).

Discussion

A better understanding of the sequences flanking microsatellite repeat elements is crucial if a better understanding of genotyping and its associated problems are to be obtained. Higher variability in a microsatellite locus as a result of flanking region mutations can lead to problems in primer binding and ultimately null alleles or allelic dropout (MacAvoy, Wood, & Gardner, 2008). Variation in the flanking regions also explain size homoplasy between alleles which can cause variability in the running time of genotyping and have shown that similar sized alleles might be identical in state but not by descent; a crucial factor when parentage testing is considered. Vowles et al. state that flanking regions of microsatellite repeats up to 50bp show high levels of convergent evolution and even short repeats have altered mutational rates in the flanking regions (Vowles & Amos 2004). Therefore it is not only the microsatellite repeat which is important but the flanking regions which serve as primer binding sites as well.

No sequence variant was found to be associated with all individuals of a specific breed tested and no individual was found to have all of the sequence anomalies exhibited for a certain locus. Sequence variation was observed in the forward primer binding site of the old published primers for HTG4 for breeds as diverse as Thoroughbred, Lusitano, Arabian, American Saddle Horse and Welsh Pony. HTG10 exhibited primer binding site mutations for breeds such as Nooitgedach, Lusitano, Arabian and Welsh Ponies while primer binding site sequence variants were observed in VHL20 for breeds such as Nooitgedacht, American Saddle Horse and Clydesdale.

HMS3 showed sequence variation in both the forward and reverse primer binding sites as well as the bases flanking the 5'end of the repeat element which was found preferentially in Friesian horses. Repeat elements other than the microsatellite dinucleotide repeats were found in the 5' region of HMS7. These repeats should be noted as they are unfavorable for primer binding and could complicate genotyping. This repeat element isn't depicted in table 4.3 as a new forward primer was designed in order to circumvent it.

Many authors describe problems associated with genotyping data. In particular, the deviation of HMS3 from Hardy-Weinberg equilibrium is a phenomenon which is observed frequently (Aberle et al., 2004; Achmann et al., 2004). HTG10 and HMS3 are prone to genotyping errors or possibly null alleles when heterozygous individuals are genotyped with the old published primers. Allele calling is also difficult for these markers as alleles often resemble 'stutter' products and are overlooked unless a comparison to parents can be made. Often, individuals would then be typed as homozygotes, creating a false heterozygote deficiency. Sequencing of the flanking regions suggests poor primer binding as a probable cause of the problems discussed.

HTG10 shows no sequence variation in the forward primer binding site, however, the reverse primer binding site for HTG10 stretches into the repeat element and there is sequence variation in this region. The forward primer binding site of HMS3 shows a sequence difference though this was only observed in some Thoroughbred samples however; the sequence variance was at the 3' end of the primer. The reverse primer binding site also has a single sequence variant observed in Lusitano, Welsh Ponies and Clydesdale.

A significant difference between this study and a similar study done by van De Goor (2009) lies in the primers used for the multiplex PCR panel. van De Goor (2009) used the original published primers while this study redesigned primers for every microsatellite locus, taking into account repeat element sizes and sequence variations in the flanking regions. The decision to re-design the primers was based on the observation that genotyping errors and null alleles had been observed in many laboratories and in published studies (Aberle et al., 2004; Achmann et al., 2004) especially with regard to markers such as HMS3 and HTG10. This was mostly due to these original primers not being intended for genotyping work but initially developed for linkage studies (Marklund et al. 1994; Tozaki et al. 2005).

The newly designed primers (table 4.4 and 4.5) have a much more uniform melting temperature and GC content across all primers and have been tested for dimer and hairpin formation as well as the presence of sequence runs. The original published primers were not intended to be used together in a single multiplex PCR and have extreme melting temperatures and GC content that make primer binding too specific (in the presence of very high annealing temperatures) and slow denaturation of double stranded DNA (in the presence of very high GC contents). In addition, the published primers often contain runs of repetitive sequences and are prone to dimer formation (table 4.6).

Problems associated with genotyping when using the original published primers have been solved by the re-design of the primers which work well enough together to be run in a single multiplex PCR. Primer design for multiplex PCR was based on recommendations given by Schoske (2003). Difficulties encountered using the newly designed primers and microsatellite panel mainly concern amplification of the larger fragments such as AHT4, CA425 and AHT5. An increase in primer concentration, annealing time or extension time increases the amplification of these markers. In addition, it was found that increasing the number of PCR cycles above 35, decreasing the annealing temperature or increasing the annealing and extension times above 2min provided no advantages as far as product concentration was concerned.

Optimum run parameters for the new multiplex PCR for genotyping horses were obtained. Pipetting very small volumes of primer proved difficult and inaccurate; therefore primer stock concentrations were decreased from 10µM to 5µM. Annealing and extension times of the PCR program were increased as this was found to increase amplification of larger markers such as CA425 and AHT4.

Electropherograms of loci amplified using the newly designed and original published primers show an improvement where interpretation of the marker panel is concerned (figure 4.1). The sex marker, Amelogenin shows improved amplification using the new primers and a clearer peak is produced. Though Amelogenin and ASB17 are labeled with the same fluorescent label and lie within 2bp of one another on the electropherogram, interpretation of these markers have not posed a problem in any of the samples analyzed throughout this study. It should be taken into account that though ASB17 is listed as spanning up to 168bp in size, this includes the 7bp tail region as well as 4bp added for putative undiscovered alleles. The same principle applies to HMS7 and ASB2 which are both labeled with NED and appear to overlap at 239bp (table 4.4).

Reading genotypes of dinucleotide markers is often complicated by stutter products and calling the correct allele is difficult. Computer software will often select the largest peak as the correct allele, but this is not always correct. Tozaki et al. support this statement as they found that almost all microsatellites in their study produced stutter products (Tozaki et al. 2001) which made automated scoring difficult. They resorted to manual allele calling, selecting the peak with the strongest signal and the lowest electrophoretic mobility.

An improvement is seen in the allele calling of problem markers such as HMS3 and HTG10. For HMS3 the genotype MN used to prove particularly difficult to read and figure 4.1 depicts how the peaks resembled the allele N with stutter products. When typed using the new primers, it is clear that the genotype is actually MN. The same is seen for the genotype KL in HTG10. Using the old primers it appears as if only the L allele is present, with accompanying stutter products. However, when the new primers are used it is clear that the genotype is KL.

Calling the correct allele can be further complicated when certain markers are amplified with tailed primers while others are not. ASB2, HMS3 and HMS7 are not tailed in the old marker panel (figure 4.1) and interpretation of these loci is difficult as the peak with the highest signal is not always the true allele. Adding Tail sequences to the reverse primers clears up the problem and the peaks with the strongest signal can be read as the correct genotype.

CHAPTER 5: LOCUS INFORMATION AND POPULATION DATA ANALYSIS: INFORMATIVE VALUES OF THE CURRENT AND NEW MICROSATELLITE PANELS

Introduction

Initially intended for mapping purposes, microsatellite markers or STRs have proven useful in genotyping for identification purposes or parentage testing (Dimsoski, 2003). Thoroughbred horses, in particular, are useful models for parentage testing studies as the Thoroughbred has had a closed studbook since 1791 (Weatherby, 1791) and improvement of the breed relies heavily on accurate pedigree data (Luikart et al., 1999).

The use of DNA for parentage testing is unique in that any sample containing the animal's DNA can be used and PCR technology makes testing economical and effective. Markers used in parentage testing must be highly informative and, taken together, have a small probability of two individuals having the same genotype (Ozkan et al., 2009). The informativeness of a locus can be determined by statistical expressions such as the number of alleles, allele frequency, Heterozygosity (He), Polymorphic Information Content (PIC) and Probability of Exclusion (PE).

During incidents of inbreeding or population bottlenecks, the number of alleles is generally reduced faster than the heterozygosity (Aberle et al., 2004). Plante et al. found the average number of alleles to be 5.5 in a population of 50 Thoroughbred horses using 12 microsatellites (Plante et al., 2007). Aberle (2004) found the average number of alleles in an Arab population of 25 to be 4.37 when using 31 microsatellites. Heterozygosity of a marker is another good indicator of genetic variation. In an attempt to fix desirable traits in a population individuals are often inbred which ultimately causes a reduction in heterozygosity and a loss of rare alleles.

Allele frequencies calculated from genotypic data are important since the informativeness of a locus depends on the number of alleles observed and the frequency of distribution of the alleles (Ozkan et al., 2009). Polymorphic loci are those for which the most common allele has a frequency < 0.95 while an allele with a frequency < 0.05 is considered to be rare (Hartl & Clark, 2007).

The Polymorphism Information Content (PIC) is another measure of variation similar to heterozygosity and is calculated from allele frequencies. A high PIC value is indicative of a locus with high informativeness. For linkage mapping Dierks et.al selected markers with PIC values > 0.5 as markers with values below this level are insufficient for paternity testing. In their study, the average PIC value was 0.49 with a maximum of 0.83 (Dierks et al., 2007).

The Probability of Exclusion (PE) for effective parentage testing should be > 0.999 (Ozkan et al., 2009). Castagnasso (2007) describe a power of exclusion of 0.993 for single parent exclusion in farm-bred jumping horses using 12 microsatellite markers, while a PE of 0.999 was attained for a double exclusion when the mother, offspring and putative sires are known. CERVUS 3.0.3 (Tristan Marshall; Fieldgenetics Ltd. www.fieldgenetics.com) calculates non exclusion probabilities: the probability of not excluding an unrelated candidate parent or parent pair at a locus (NE-1P, NE-2P, NE-PP) or the probability that the genotypes at a locus of two random individuals or siblings do not differ enough for them to be differentiated (NE-I, NE-SI). Incorporating more markers can increase the PE of a marker panel.

Methods and Materials

Population Data Generation and Analysis

The informativeness of the old (ISAG-9) and new (17-plex) marker panels was analyzed using CERVUS 3.0.3 (Tristan Marshall; Fieldgenetics Ltd. www.fieldgenetics.com). Allele frequencies, Probability of Non-Exclusion, deviations from Hardy Weinberg equilibrium (HWE), observed heterozygosity (He) and Polymorphism Information Content (PIC) were determined for all loci except Amelogenin and Lex003 in sample populations of 100 randomly selected Thoroughbred and 100 Arabian horse samples. Lex003 was not considered as it is found on the X chromosome and therefore always homozygous in stallions which will be expressed as a false heterozygote deficiency.

Setup and Statistical Validation of a Hypothetical Marker Panel

Markers not present in the ISAG panel (ASB17, HMS2, HTG6, HTG7, CA425 and ASB23) were analyzed in combinations of three in order to determine the best set to add to and increase the informativeness of the existing ISAG panel. The following statistics were calculated from the population data: allele frequency, the mean number of alleles per locus, observed heterozygosity (He), Polymorphism Information Content (PIC) and non-exclusion probability for the different marker combinations were calculated and compared.

Results

Population Data

Only two markers deviated from Hardy Weinberg Equilibrium. HMS2 in Thoroughbreds showed a significant deviation at the 5% level with 1 degree of freedom. HTG6 in Arabian horses showed a significant deviation at the 0.1% level with 3 degrees of freedom. The estimated frequency of null alleles for these markers is summarized in table 5.6.

The allele frequencies for 15 loci of the two horse breeds studied are depicted in figure 5.1. The number of alleles per locus for each population as well as the mean observed heterozygosity, PIC and Non-Exclusion Probability are summarized in tables 5.1 to 5.4.

The highest allele frequency was observed for locus ASB23, allele S or 24 with a frequency >0.8. Allele frequencies as low as 0.005 were observed in almost all loci and were equally distributed between the two sample populations. These were not considered significant as allele frequencies as low as 0.005 are often associated with genotyping errors. ASB17 and ASB2 had the greatest number of alleles while HTG7 had the least.

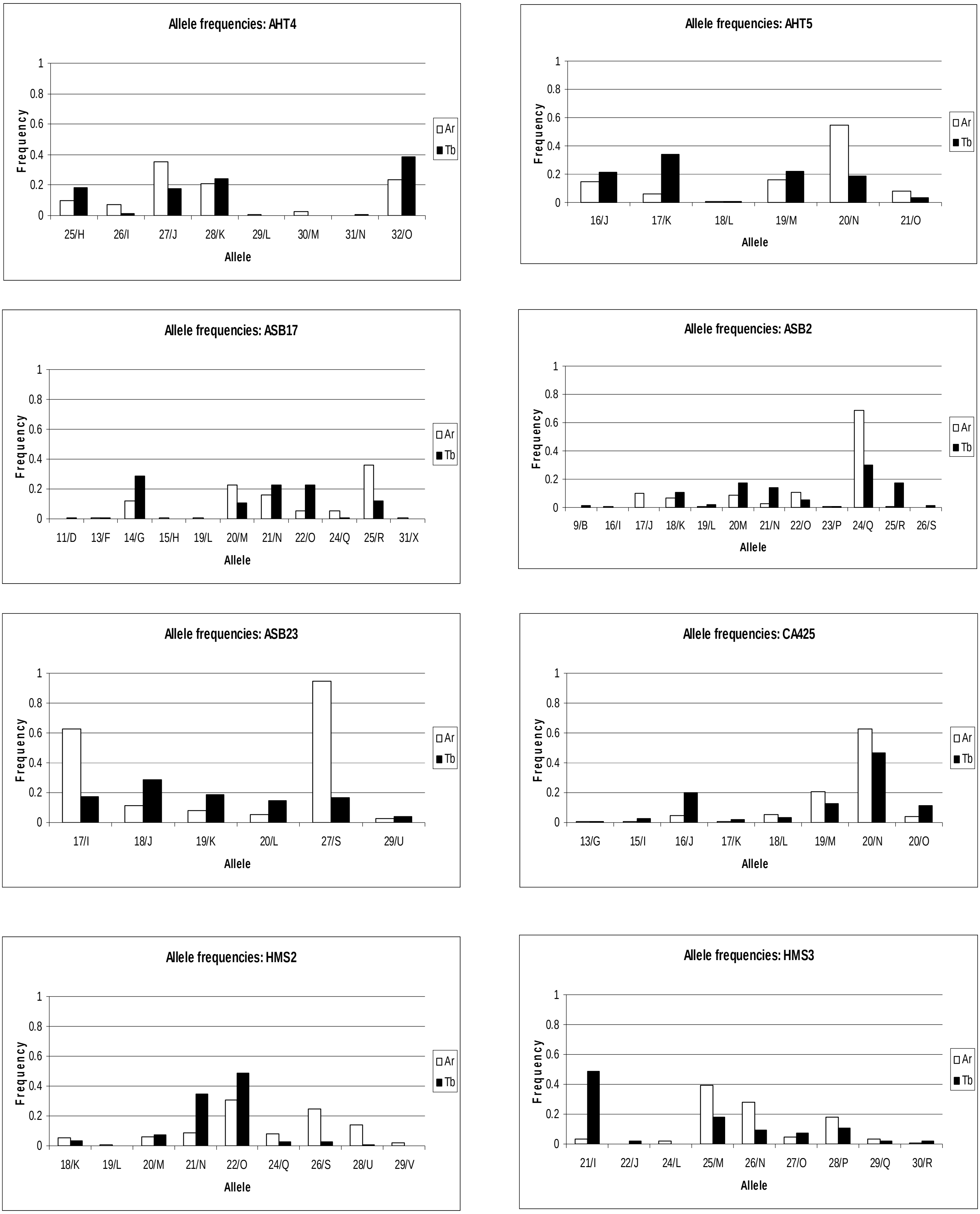


Figure 5.1: Allele frequencies across all loci for the sample populations of Thoroughbred and Arabian horse (continued overleaf).

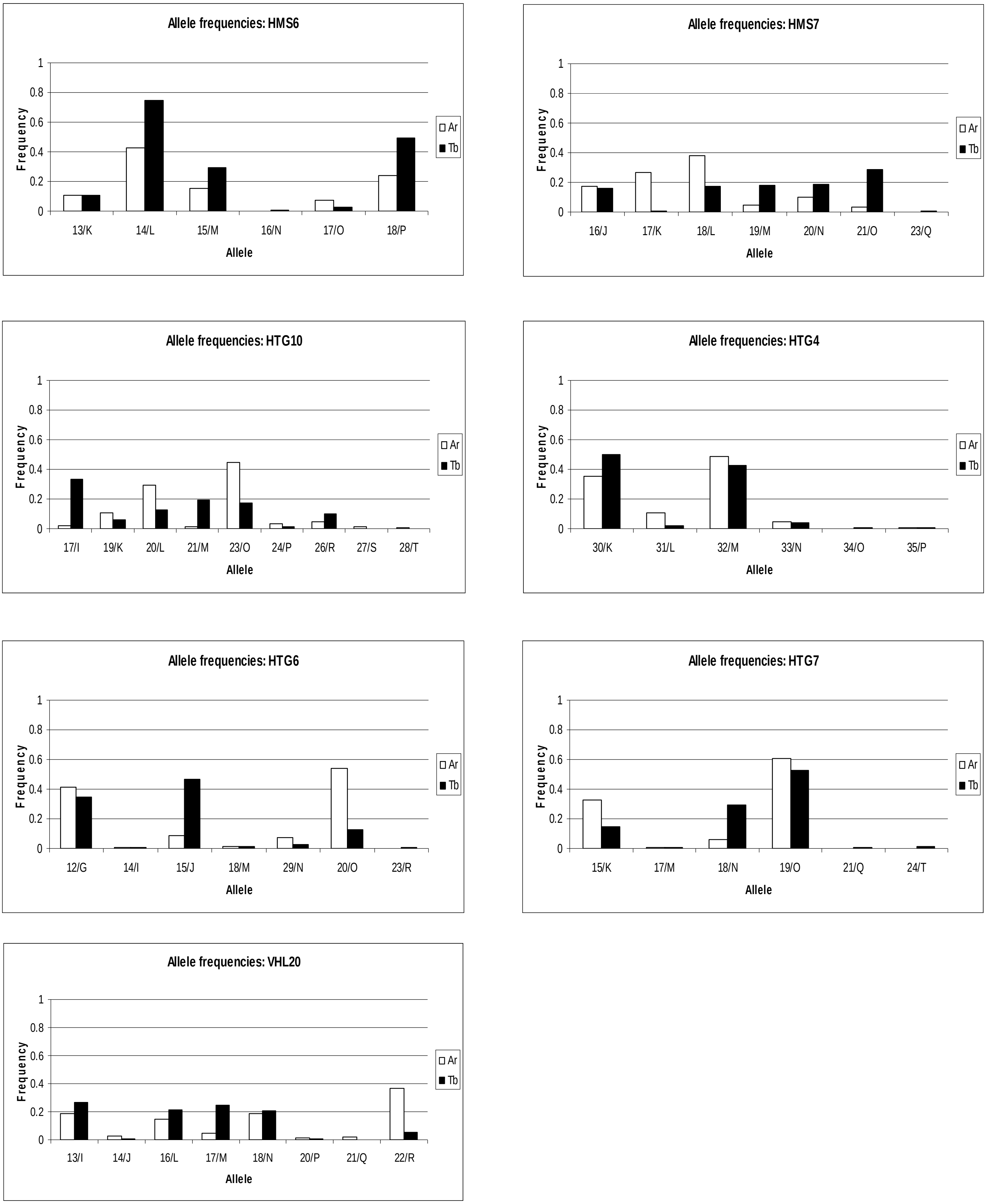


Figure 5.1: Allele frequencies across all loci for the sample populations of Thoroughbred and Arabian horse.

Table 5.1: The number of alleles for each locus, the mean number of alleles for the sample populations of Thoroughbred (n = 100) and Arabian horses (n = 100) and the PIC values and Observed Heterozygosity for those two sample populations

Marker	No. alleles Ar	No. alleles Tb	PIC Ar	PIC Tb	He Ar	He Tb
AHT4	7	6	0.722	0.686	0.74	0.73
AHT5	6	6	0.608	0.713	0.56	0.79
ASB17	10	8	0.743	0.753	0.74	0.80
ASB2	10	10	0.485	0.792	0.45	0.76
ASB23	6	6	0.551	0.773	0.55	0.80
CA425	8	8	0.637	0.676	0.57	0.72
HMS2	9	7	0.782	0.571	0.82	0.44
HMS3	8	8	0.686	0.675	0.72	0.63
HMS6	5	6	0.679	0.596	0.70	0.63
HMS7	6	7	0.703	0.783	0.70	0.79
HTG10	9	7	0.649	0.763	0.63	0.76
HTG4	5	6	0.551	0.475	0.56	0.46
HTG6	6	7	0.666	0.575	0.66	0.59
HTG7	4	6	0.437	0.546	0.50	0.59
VHL20	8	7	0.734	0.739	0.70	0.80
Mean no. alleles per locus	7.87	7.27	-	-	-	-

Table 5.2: Mean observed heterozygosities for the old panel of nine ISAG markers as well as the new marker panel of 15 microsatellite markers for both sample populations

Mean expected heterozygosity	
Arabian: 15 markers	0.6861
Arabian: 9 markers	0.6907
Thoroughbred: 15 markers	0.7228
Thoroughbred: 9 markers	0.7373

Table 5.3: Mean Polymorphic Information Content (PIC) for the old panel of nine ISAG markers as well as the new marker panel of 15 microsatellite markers for both sample populations

Mean PIC	
Arabian: 15 markers	0.6422
Arabian: 9 markers	0.6464
Thoroughbred: 15 markers	0.6745
Thoroughbred: 9 markers	0.6914

Table 5.4: The combined non-exclusion probability for the old panel of nine ISAG markers as well as the new marker panel of 15 microsatellite markers for both sample populations.

	Arabian population; 15 markers	Arabian population; 9 markers	Thoroughbred population; 15 markers	Thoroughbred population; 9 markers
Combined NE-1P:	5x10^{-3}	4 x10^{-2}	2 x10^{-3}	2 x10^{-2}
Combined NE-2P:	7.05 x10^{-3}	3 x10^{-3}	2.69 x10^{-5}	1 x10^{-3}
Combined NE-PP:	8 x10^{-8}	6.4 x10^{-5}	2 x10^{-8}	1.45 x10^{-5}
Combined NE-I:	9.25 x10^{-14}	2 x10^{-8}	8.29 x10^{-15}	1.57 x10^{-9}
Combined NE-SI:	4.79 x10^{-6}	6.13 x10^{-4}	2.18 x10^{-6}	3.24 x10^{-4}

* Combined non-exclusion probability for first parent exclusion (NE-1P), second parent exclusion (NE-2P), parent pair exclusion (NE-PP), individual exclusion (NE-I) and sib-identity (NE-SI).

The null allele frequency for each locus (table 5.5) was calculated more as an estimate of genotyping errors than of true non amplifying alleles.

Table 5.5: Estimated null allele frequencies of all 15 loci for both sample populations.

Locus	Null allele frequency estimate Ar	Null allele frequency estimate Tb
AHT4	0.009	0.0048
AHT5	0.0705	-0.023
ASB17	0.0254	-0.0127
ASB2	0.0769	0.0376
ASB23	0.0294	0.0017
CA425	0.0768	-0.0041
HMS2	-0.0173	0.1854
HMS3	0.0042	0.0596
HMS6	0.0191	0.0123
HMS7	0.0287	0.0129
HTG10	0.0475	0.0194
HTG4	0.0566	0.1121
HTG6	0.0378	0.0348
HTG7	0.0171	0.0155
VHL20	0.0456	-0.0187

Setup and Statistical Validation of A Hypothetical Marker Panel

For hypothetical genotyping panels of twelve markers the mean observed heterozygosity, number of alleles per locus, PIC and non-exclusion probabilities are summarized in figures 5.2 to 5.4 and table 5.6.

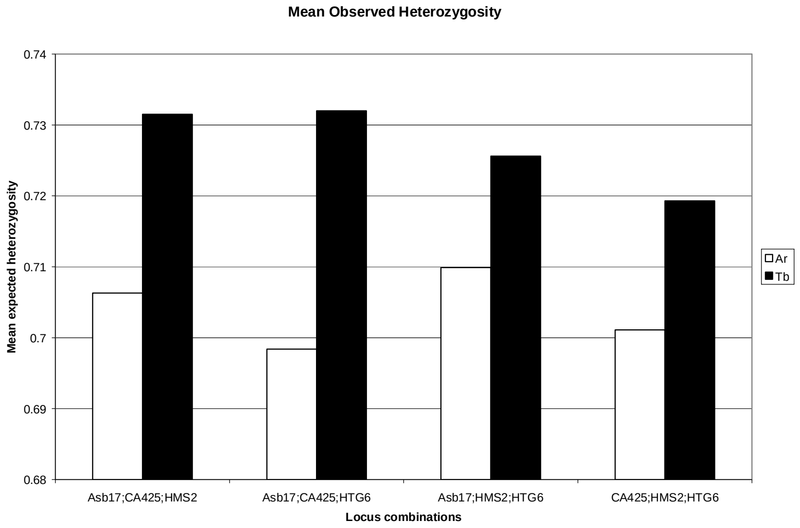


Figure 5.2: Mean observed heterozygosity in the sample populations of Thoroughbred and Arabian horses for hypothetical genotyping panels using the nine ISAG loci and combinations of ASB17, CA425, HMS2 and HTG6.

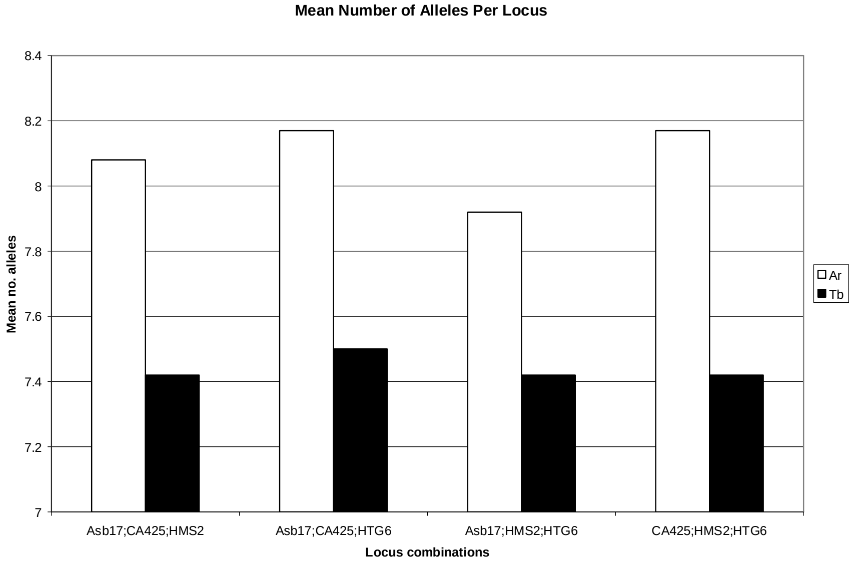


Figure 5.3: Mean number of alleles per locus in the sample populations of Thoroughbred and Arabian horses for hypothetical genotyping panels using the nine ISAG loci and combinations of ASB17, CA425, HMS2 and HTG6.

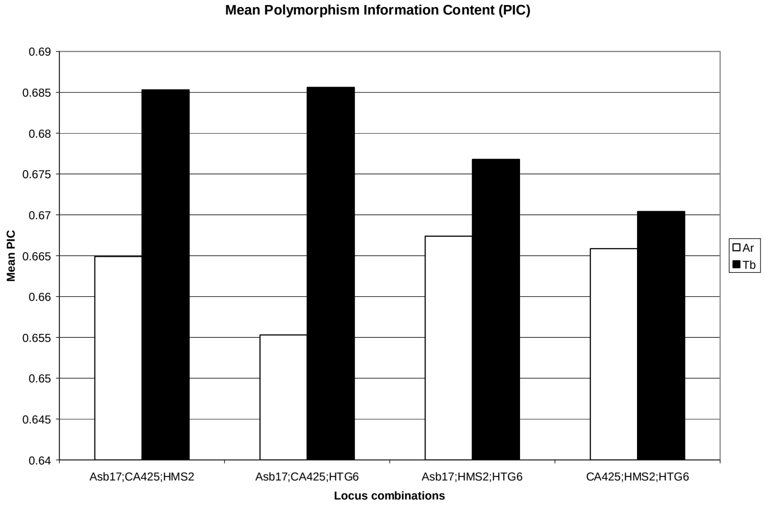


Figure 5.4: Mean Polymorphic Information Content in the sample populations of Thoroughbred and Arabian horses for hypothetical genotyping panels using the nine ISAG loci and combinations of ASB17, CA425, HMS2 and HTG6.

Table 5.6: Combined non-exclusion probabilities in the sample populations of Thoroughbred and Arabian horses for hypothetical genotyping panels using the nine ISAG loci and combinations of ASB17, CA425, HMS2 and HTG6.

ASB17;CA425;HMS2	**Ar**	**Tb**
NE-1P	x10^{-2}	6.67 x10^{-3}
NE-2P	2.90 x10^{-4}	1.67 x10^{-4}
NE-PP	9.60 x10^{-7}	4.80 x10^{-7}
NE-I	1.11 x10^{-11}	2.89 x10^{-12}
NE-SI	3.85 x10^{-5}	2.49 x10^{-5}
ASB17;CA425;HTG6		
NE-1P	1 x10^{-2}	6 x10^{-3}
NE-2P	4 x10^{-4}	1.68 x10^{-4}
NE-PP	1.68 x10^{-6}	4.90 x10^{-7}
NE-I	2.32 x10^{-11}	2.86 x10^{-12}
NE-SI	4.52 x10^{-5}	2.48 x10^{-5}
ASB17;HMS2;HTG6		
NE-1P	1 x10^{-2}	7 x10^{-3}
NE-2P	2.78 x10^{-4}	2.08 x10^{-4}
NE-PP	9.40 x10^{-7}	7.20 x10^{-7}
NE-I	1.01 x10^{-11}	4.77 x10^{-12}
NE-SI	3.63 x10^{-5}	2.81 x10^{-5}
CA425;HMS2;HTG6		
NE-1P	1 x10^{-2}	8 x10^{-3}
NE-2P	3.51 x10^{-4}	2.49 x10^{-4}
NE-PP	1.32 x10^{-6}	9.20 x10^{-7}
NE-I	1.74 x10^{-11}	7.15 x10^{-12}
NE-SI	4.27 x10^{-5}	3.17 x10^{-5}

* Combined non-exclusion probability for first parent exclusion (NE-1P), second parent (NE-2P), parent pair (NE-PP), individual (NE-I) and sib identity (NE-SI).

ASB23 and HTG7 were excluded from the markers available for addition to the hypothetical 12-marker panel due to low heterozygosities and a low number of alleles. The number of alleles for each locus as well as the heterozygosity and non-exclusion probability for these two markers is summarized in table 5.7.

Table 5.7: Number of alleles, observed Heterozygosity and Non Exclusion Probability for ASB23 and HTG7.

	ASB23	HTG7
Number of alleles	6 in Arabian horses 6 in Thoroughbreds	4 in Arabian horses 6 in Thoroughbreds
Observed heterozygosity	0.58 in Arabian horses	0.521 in Arabian horses
NE-1P	0.804 in Arabian horses 0.571 in Thoroughbreds	0.865 in Arabian horses 0.804 in Thoroughbreds

Discussion

Plante (2007) found the average number of alleles to be 5.5 in a population of 50 Thoroughbred horses using 12 microsatellites. Aberle (2004) found the average number of alleles in an Arab population of 25 to be 4.37 when using 31 microsatellites. These results differ from the average number of alleles observed in this study which show that the mean number of alleles per locus to be 7.27 for 15 markers (Lex003 and Amelogenin excluded) in a population of 100 Thoroughbred horses and 7.87 in a population of 100 Arab horses (table 5.1). The marked difference in these figures is most probably due to the fact that this study incorporates more markers than that of Plante (2007) and more samples than that of both Aberle (2004) and Plante (2007). It illustrates the value of why direct comparisons between studies should be undertaken with care because of differing sample sizes and the differences in markers used.

Results from this study showed that there is a marked difference between Thoroughbreds and Arabian horses as to which alleles have the highest frequency per locus. Based on data generated by CERVUS 3.0.3 (Tristan Marshall; Fieldgenetics Ltd. www.fieldgenetics.com) for 15 loci in Arabian and Thoroughbred populations it would appear that all loci are polymorphic in both populations since the allele frequencies lay far below 0.95 (figure 5.1). There are marked differences in the allele frequencies of the two populations and often alleles that are present in low frequencies in one population are completely absent in the other. Rare alleles i.e. alleles with a frequency < 0.005 were observed in almost all markers although their distribution was equal between the two sample populations.

The mean Observed Heterozygosity (He) for the Arabian population is 0.6861 when 15 markers are considered while the Thoroughbred population has a mean He of 0.7228 (table 5.2). The heterozygosity in these populations is comparable to the variation observed in other horse breeds and shows no genetic impoverishment (Aberle, Hamann, Drogemuller, & Distl, 2004; Achmann et al., 2004; Plante et al., 2007). The difference in heterozygosity between the nine ISAG markers and 15 microsatellite markers is not significant and both panels show sufficient genetic variation to be used successfully for parentage testing in horses.

The lower heterozygosity observed in the Arabian population might be due to inbreeding but the presence of rare alleles in the Arabian population and alleles unique to Arabian horses that are absent in Thoroughbred horses indicate that inbreeding has not been severe. The mean number of alleles would also be low in a population where the inbreeding coefficient is high and this is not the case in the Arabian population. Heterozygosity is only a good indicator of inbreeding when such events occur very frequently and even then at least 200 loci are required for accurate assumptions to be made (van Eldik et al., 2006). As far as the heterozygosity of individual markers is concerned (table 5.1) the markers ASB2, HMS2 and HTG4 show marked differences between the two populations studied with ASB2 having insufficient heterozygosity for parentage testing in Arabs while HMS2 and HTG4 would prove insufficient in Thoroughbreds.

Deviations from HW may be due to genotyping errors, null alleles or inbreeding (Ozkan et al., 2009). Only HMS2 in Thoroughbreds and HTG6 in Arabians show a deviation from HWE at a significance level of 5% and 0.1% respectively. Given that the deviations are observed mostly at $P < 0.001$ it would appear that these deviations are not due to chance alone (Hartl & Clark, 2007) and might be attributed to genotyping errors or possibly due to inbreeding or even selection acting upon a nearby gene. Provided the frequency of null alleles or genotyping errors at a locus is not too high, this marker need not be excluded from the parentage testing panel. The markers that show significant deviation from HWE have null allele frequencies of 0.1854 for HMS2 in Thoroughbreds and 0.0378 for HTG6 in Arabian horses. The frequency for HMS2 is one of the highest frequencies of all loci and might indicate that the deviation is due to genotyping errors. The frequency of HTG6 is closer to zero and indicates an absence of genotyping errors and the deviation might, therefore, be attributed to selection or inbreeding.

The mean PIC values for the old and new marker panels in Thoroughbreds and Arabian horses range from 0.64 to 0.69. The differences in value between the old and new panels are negligible (table 5.3). The PIC values obtained in this study are > 0.5 and indicate that the marker panels have enough variation to be used in parentage testing. There is no significant difference between the PIC values for the nine ISAG marker panel and 15 microsatellite markers indicating that both panels are suitable for genotyping horses. Individually, all markers have PIC values > 0.5 in both populations except for HTG7 in the Arab population and HTG4 in Thoroughbreds (table 5.1).

The combined non-exclusion probabilities for the new panel show lower figures than those for the panel containing 9 ISAG markers (table 5.4). This is the case in both sample populations and is indicative that the use of more markers gives more discriminatory power to the panel.

Excluding Lex003, which is found on the X chromosome, there are six markers added to this new panel that are not present in the current ISAG panel. A new marker panel can be recommended containing the nine ISAG markers as well as three new markers that are found to be the most informative of the six originally added in this study. ASB23 and HTG7 were discarded first due to their small number of alleles, low heterozygosities in the Arab populations and high non exclusion probabilities (table 5.7). The remaining markers; ASB17, CA425, HMS2 and HTG6 were analyzed in combinations of three in order to determine the best set to add to the ISAG panel.

Considering the mean number of alleles per locus, the greatest number of alleles in both populations of horses was obtained using a marker combination of ASB17, CA425 and HTG6 (table 5.2). The highest expected heterozygosity (figure 5.2) in Thoroughbreds was obtained using the marker combination of ASB17, CA425 and HTG6, although this produced the lowest heterozygosity in Arabs. Considering this, the best marker combination to use in order to obtain satisfactory heterozygosity figures would be ASB17, CA425 and HMS2. The highest PIC value was obtained using the markers ASB17, CA425 and HMS2 (figure 5.4) in the Arabian population although for Thoroughbreds this was the second best combination. The differences in Non-exclusion probabilities for the different marker combinations were marginal and insufficient to render a decision being made based on these figures (table 5.6). HMS2 and HTG6 deviate from HWE in Thoroughbreds and Arabian horses respectively. HTG6 shows a lower expected null allele frequency than HMS2 and its deviation might, therefore, not be due to genotyping errors, making it the more likely candidate for addition to a genotyping panel. Therefore the loci ASB17, CA425 and HTG6 would be a good choice of markers to add to the existing ISAG panel of nine markers if a greater probability of exclusion was required.

At the 2010 ISAG meeting in Edinburgh it was decided to add the markers ASB17, ASB23 and HMS2 to the nine ISAG markers (personal communication from Dr. CK Harper, Veterinary Genetics Laboratory, Onderstepoort). Despite the deviation from HWE for HMS2, this marker can still be used effectively in the genotyping panel if novel primers were designed to possibly decrease the number of genotyping errors which seem to affect its accordance with HWE. Similarly, ASB17 will be a good marker to use as it has a large number of alleles and a mean heterozygosity of 0.77 for the two horse populations tested. ASB23 and HTG7 were purposefully discarded as possible markers because of their small number of alleles, low heterozygosities and high probability of non exclusion. In my opinion a marker such as HTG6 or CA425 would be better suited to replace ASB23 in the ISAG panel.

CHAPTER 6: GENERAL CONCLUSIONS

A recent study published by van De Goor (2009) proposes a nomenclature system that will serve as the standard in equine genotyping. Their study is very similar to this project and therefore serves as an important benchmark to which this study can be compared. The authors of that study confirm that no details have been published about the allele structure or DNA sequence variation within and flanking microsatellite repeats, respectively. Therefore, this study contributes to what is already known about microsatellite sequences. Van De Goor (2009) also emphasize the need for a numerical nomenclature system which would allow for more effective data sharing and that is based on the principles of the human repeat-based system recommended by the ISFG and can be used in legal case work.

Many laboratories have redesigned primers for the horse genotyping markers in order to overcome difficulties in genotyping or data analysis. Generally, however, the original published primers are used. Primers were redesigned in this study for 16 microsatellite markers aided by sequencing data of repeat elements and flanking regions. These primers can be used in a single multiplex PCR with a size range of 400bp as markers with the same fluorescent dye do not overlap in size. The new primers provide a genotyping panel with greater accuracy and less chance of primer binding problems as their design is based on sound sequence data.

Population data of 100 Thoroughbred and 100 Arabian horses indicated that the addition of more markers to the ISAG panel of nine markers did have a positive effect on the ability of the panel to exclude individuals in parentage verification tests. No increase was observed, concerning genetic variability, with the addition of more markers to the panel as was indicated by similar values obtained for He and PIC in both panels.

In this study, seven markers were added to the nine ISAG markers. Some markers are less than ideal due to chromosomal location (Lex003), low numbers of alleles and low heterozygosity (ASB23 and HTG7) and deviations from HWE due to genotyping errors (HMS2). The remaining markers; ASB17, HTG6 and CA425 can be recommended for addition to the ISAG panel in order to increase the probability of exclusion.

There is a need for detailed population studies for application in forensic casework. Genetic pedigree structures, power of identity and breed assignment are required. Such studies are currently being undertaken and van De Goor et al. state that results will be made available in the near future. Future research should consider population studies and statistical validation of the microsatellite marker panel in South African horse populations and indigenous breeds such as the Nooitgedacht and Boerperd. The true nature of microsatellites is only beginning to be understood. Solid state platforms and new marker systems such as SNPs will inevitably play a greater role in the future of animal genotyping. For the moment, however, the microsatellite marker panels in use are well characterized and allow easy and efficient DNA-based identification of horses world wide.

LIST OF REFERENCES

Aberle, K. S., Hamann, H., Drogemuller, C., & Distl, O. (2004). Genetic diversity in German draught horse breeds compared with a group of primitive, riding and wild horses by means of microsatellite DNA markers. Animal Genetics, 35, 270-277.

Achmann, R., Curik, I., Dovc, P., Kavar, T., Bodo, I., Habe, F., et al. (2004). Microsatellite diversity, population subdivision and gene flow in the Lipizzan horse. Animal Genetics, 35, 285-292.

Alosi, S., Balbo, M., Biagi, G., Braend, M., Catalano, A., Catarsini, O., et al. (1980). Current Status of Equine Blood Typing. In Proceeding II International Seminary. Centro Studi Gruppi Sanguigni del Cavallo (UNIRE), Pisa.

Anunciacao, C. E., & Astolfi-Filho, S. (2000). Paternity test in “mangalarga-marchador” equines by DNA-fingerprinting. Plasmid, 35(10), 2007-2015.

Bennett, P. (2000). Demystified . . . Microsatellites. Journal of Clinical Pathology: Molecular Pathology, 53, 177-183.

Beuzen, N. D., Stear, M. J., & Chang, K. C. (2000). Molecular markers and their use in animal breeding. The Veterinary Journal, 160, 42-52.

Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochman, C., Taberlet, P., et al. (2004). How to track and assess genotyping errors in population genetics studies. Molecular Ecology, 13, 3261-3273.

Bowling, A. T. (2001). Historical development and application of molecular genetic tests for horse identification and parentage control. Livestock Production Science, 72, 111-116.

Bozzini, M., Fantin, D., Ziegle, J., van Haeringen, H., Jacobs, W., Ketchum, M., et al. (1996). Automated equine paternity testing. Animal Genetics, 27.

Budowle, B., Garofano, P., Hellman, A., Ketchum, M., Kanthaswamy, S., Parson, W., et al. (2005). Recommendations for animal DNA forensic and identity testing. International Journal of Legal Medicine, 119, 295-302.

Canon, J., Checa, M. L., Carleos, C., Vega-Pla, J. L., Vallejo, M., Dunner, S., et al. (2000). The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data. Animal Genetics, 31, 39-48.

Castagnasso, E. E., Kienast, M. E., Garcia, P. P., & Giovambattista, G. (2007). A Case of Multiple Assignments (Paternity⁄Maternity) in an Equine-Out Breeding System. Journal of Forensic Science, 52(4), 889-890.

Chistiakov, D. A., Hellemans, B., & Volckaert, F. A. (2006). Microsatellites and their genomic distribution, evolution, function and applications: A review with special reference to fish genetics. Aquaculture, 255, 1 - 29.

Dakin, E. E., & Avise, J. C. (2004). Microsatellite null alleles in parentage analysis. Heredity, 93, 504-509.

Deucher, A., Chiang, T., & Schrijver, I. (2010). Consultations in Molecular Diagnostics: Rare sequence variation in the genome flanking a short tandem repeat locus can lead to a question of nonmaternity. Journal of Molecular Diagnostics, 12(3), 384-389.

Dierks, C., Lohring, K., Lampe, V., Wittwer, C., Drogemuller, C., Distl, O., et al. (2007). Genome-wide search for markers associated with osteochondrosis in Hanoverian warmblood horses. Mammalian Genome, 18, 739-747.

Dimsoski, P. (2003). Development of a 17-plex microsatellite polymerase chain reaction kit for genotyping horses. Croatian Medical Journal, 44(3), 332-335.

Divne, A., Edlund, H., & Allen, M. (2010). Forensic analysis of autosomal STR markers using Pyrosequencing. Forensic Science International: Genetics, 4, 122-129.

Emara, M. G., & Kim, H. (2003). Genetic Markers and their Application in Poultry Breeding. Poultry Science, 82, 952-957.

Fairbanks, D. J., & Andersen, W. R. (1999). Genetics: The continuity of life. (G. Carlson). United States: Brooks/Cole Publishing Company.

Grimaldi, M., & Crouau-Roy, B. (1997). Microsatellite Allelic Homoplasy Due to Variable Flanking Sequences. Journal of Molecular Evolution, 44, 336-340.

Grossman, P., & Colburn, J. (1992). Capillary electrophoresis: theory and practice. (P. Grossman & J. Colburn). California: San Diego, California: Academic Press.

Guo, W., Ling, J., & Li, P. (2009). Consensus features of microsatellite distribution: Microsatellite contents are universally correlated with recombination rates and are preferentially depressed by centromeres in multicellular eukaryotic genomes. Genomics, 93, 323-331.

Hartl, D., & Clark, A. (2007). Principles of population genetics. Sunderland: Sinauer Associates.

Hauge, X. Y., & Litt, M. (1993). A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. Human Molecular Genetics, 2(4), 411-415.

Hayden, M. J., & Sharp, P. J. (2001). Targeted development of informative microsatellite (SSR) markers. Nucleic Acids Research, 29(8).

Henegariu, O., Heerema, N. A., Dlouhy, S. R., Vance, G. H., & Vogt, P. H. (1997). Multiplex PCR: Critical Parameters and Step-by-Step Protocol. BioTechniques, 23, 504-511.

Hill, E. W., Bradley, D. G., Al-barody, M., Ertugrul, O., Splan, R. K., Zakharov, I., et al. (2002). History and integrity of thoroughbred dam lines revealed in equine mtDNA variation. Animal Genetics, 33, 287-294.

Hill, C.R. et al., 2008. Characterization of 26 MiniSTR Loci for Improved Analysis of Degraded DNA Samples. Journal of Forensic Science, 53(1), 73-80.

Innis, M. A., Gelfand, D. H., Snisky, J. J., & White, T. J. (1990). PCR Protocols: A guide to methods and applications. San Diego: San Diego Academic Press.

Jeffreys, A., Wilson, V., & Thein, S. (1985). Individual-specific 'fingerprints' of human DNA. Nature, 316.

Kofler, R., Schlotterer, C., Luschutzky, E., & Lelley, T. (2008). Survey of microsatellite clustering in eight fully sequenced species sheds light on the origin of compound microsatellites. BMC Genomics, 9(612), 1-14.

Lee, S., & Cho, G. (2006). Parentage testing of thoroughbred horses in Korea using microsatellite DNA typing. Journal of Veterinary Science, 7, 63–67.

Lipinski, M. J., Amigues, Y., Blasi, M., Broad, T. E., Cherbonnel, C., Cho, G. J., et al. (2007). An international parentage and identification panel for the domestic cat (*Felis catus*). Animal Genetics, 38, 371-377.

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

Luikart, G., Biju-Duval, M., Ertugrul, O., Zagdsuren, Y., Maudet, C., Taberlet, P., et al. (1999). Power of 22 microsatellite markers in fluorescent multiplexes for parentage testing in goats (*Capra hircus*). Animal Genetics, 30(August), 431-438.

MacAvoy, E. S., Wood, A. R., & Gardner, J. P. (2008). Development and evaluation of microsatellite markers for identification of individual Greenshell™ mussels (Perna canaliculus) in a selective breeding programme. Aquaculture, 274, 41 - 48.

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

Marletta, D., Tupac-Yupanqui, I., Bordonaro, S., Garcia, D., Guastella, A. M., Criscione, A., et al. (2006). Analysis of genetic diversity and the determination of relationships among western Mediterranean horse breeds using microsatellite markers. Journal of Animal Breeding and Genetics, 123, 315-325.

Meudt, H. M., & Clarke, A. C. (2007). Almost Forgotten or Latest Practice? AFLP applications, analyses and advances. Trends in Plant Science, 12(3). doi: 10.1016/j.tplants.2007.02.001.

Morling, N., Allen, R. W., Carracedo, A., Geada, H., Guidet, F., Hallenberg, C., et al. (2002). Paternity Testing Commission of the International Society of Forensic Genetics: recommendations on genetic investigations in paternity cases. Forensic Science International, 129, 148-157.

Mukesh, M., Sodhi, M., Kataria, R. S., & Mishra, B. P. (2009). Livestock Use of microsatellite multilocus genotypic data for individual assignment assay in six native cattle breeds from north-western region of India. Livestock Science, 121, 72-77.

Mullis, K. B., & Faloona, F. A. (1987). Specific Synthesis of DNA in Vitro via a Polymerase-Catalyzed Chain Reaction. Methods in Enzymology, 155, 335-350.

Myres, N. M., Ritchie, K. H., Lin, A. A., Hughes, R. H., Woodward, S. R., Underhill, P. A., et al. (2009). Y-chromosome short tandem repeat intermediate variant alleles DYS392.2, DYS449.2, and DYS385.2 delineate new phylogenetic substructure in human Y-chromosome haplogroup tree. Croatian Medical Journal, 50, 239-249.

Narkuti, V., & Oraganti, N. M. (2008). De novo deletion at D13S317 locus: A case of paternal-child allele mismatch identified by microsatellite typing. Clinica Chimica Acta, 403, 264-265.

Njiru, Z. K., Constantine, C. C., Gitonga, P. K., & Reid, S. A. (2007). Genetic variability of *Trypanosoma evansi* isolates detected by inter-simple sequence repeat anchored-PCR and microsatellite. Veterinary Parasitology, 147, 51-60.

Ozkan, E., Soysal, M. I., Ozder, M., Koban, E., Sahin, O., Inci, T., et al. (2009). Evaluation of parentage testing in the Turkish Holstein population based on 12 microsatellite loci. Livestock Science, 124, 101-106. doi: 10.1016/j.livsci.2009.01.004.

Park, M. J., Shin, K., Kim, N. Y., Yang, W. I., Cho, S., Lee, H. Y., et al. (2008). Characterization of Deletions in the DYS385 Flanking Region and Null Alleles Associated with AZFc Microdeletions in Koreans. Journal of Forensic Science, 53(2), 331-334.

Plante, Y., Vega-pla, J. L., Lucas, Z., Colling, D., de March, B., Buchanan, F., et al. (2007). Genetic diversity in a feral horse population from Sable Island, Canada. Journal of Heredity, 98(6), 594-602.

Sambrook, J. (1989). Molecular cloning: A laboratory manual. New York: Cold Spring Harbor Laboratory Press.

Sanchez, J. J., Balogh, K., Berger, B., Bogus, M., Butler, J. M., Fondevila, M., et al. (2008). Forensic typing of autosomal SNPs with a 29 SNP-multiplex—Results of a collaborative EDNAP exercise. Forensic Science International: Genetics, 2, 176-183.

Schlotterer, C. (2000). Evolutionary dynamics of microsatellite DNA. Chromosoma, 109, 365-371.

Schoske, R., Vallone, P. M., Ruitberg, C. M., & Butler, J. M. (2003). Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci. Anals of Bioanalytical Chemistry, 375, 333-343.

Scotti, I., Paglia, G., Magni, F., & Morgante, M. (1999). Microsatellite markers as a tool for the detection of intra- and interpopulational genetic structure. (E. Gillet).

Shriver, M. D., Jin, L., Boerwinkle, E., Deka, R., & Chakraborty, R. (1995). A Novel Measure of Genetic Distance for Highly Polymorphic Repeat Loci Tandem. Molecular Biological Evolution, 12(5), 914-920.

Shubitowski, D. M., Venta, P. J., Douglass, C. L., & Ewart, S. L. (2001). Polymorphism identification within 50 equine gene-specific sequence tagged sites. Animal Genetics, 32, 78-88.

Sobrino, B., Brion, M., & Carracedo, A. (2005). SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Science International, 154, 181-194.

Swinburne, J. E., Hopkins, A., & Binns, M. M. (2002). Assignment of the horse grey coat colour gene to ECA25 using whole genome scanning. Animal Genetics, 33, 338-342.

Takagi, K., Tsuchiya, K., Tsumagari, S., Kitazima, Y., Takeishi, M., Yoshii, T., et al. (1995). Application of parent/offspring pedigree in Thoroughbred horses with DNA fingerprint. Journal of Reproduction and Development, 41(1), 1-5.

Tamaki, K., & Jeffreys, A. J. (2005). Human tandem repeat sequences in forensic DNA typing. Legal Medicine, 7, 244-250. doi: 10.1016/j.legalmed.2005.02.002.

Toro, M. A., Fernandez, J., & Caballero, A. (2009). Molecular characterization of breeds and its use in conservation. Livestock Science, 120, 174 - 195.

Tozaki, T., Kakoi, H., Mashima, S., Hirota, K., Hasegawa, T., Ishida, N., Miura, N., Choi-Miura, N. & Tomita, M. (2001). Population Study and Validation of Paternity Testing for Thoroughbred Horses by 15 Microsatellite Loci. Journal of Veterinary Medical Science, 63(11), 1191-1197.

Tozaki, T., Hirota, K., Hasegawa, T., Tomita, M., & Kurosawa, M. (2005). Prospects for whole genome linkage disequilibrium mapping in Thoroughbreds. Gene, 346, 127 - 132.

Tozaki, T., Swinburne, J., Hirota, K., Hasegawa, T., Ishida, N., Tobe, T., et al. (2007). Improved resolution of the comparative horse–human map: Investigating markers with in silico and linkage mapping approaches. Gene, 392, 181 - 186.

Vallone, P. M., & Butler, J. M. (2004). AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques, 37, 226-231.

Vowles, E.J. & Amos, W. (2004). Evidence for Widespread Convergent Evolution around Human Microsatellites. PLOS Biology, 2(8).

Weatherby. (1791). An introduction to the general studbook. London: Weatherby & Sons.

Wilk, A. (1999). National Geographic: Horses. P.A. Gallo & co. (Pty)Ltd.

van Asch, B., Alves, C., Pereira, F., Gusmao, L., & Amorim, A. (2008). A new autosomal STR multiplex for canine genotyping. Forensic Science International:Genetics Supplement Series, 1, 628-629.

van De Goor, L. P., Panneman, H., & van Haeringen, A. (2009). A proposal for standardization in forensic bovine DNA typing: allele nomenclature of 16 cattle-specific short tandem repeat loci. Animal Genetics, 40, 630-636.

van De Goor, L. P., Panneman, H., & van Haeringen, W. A. (2009). A proposal for standardization in forensic equine DNA typing: allele nomenclature for 17 equine-specific STR loci. Animal Genetics, 41, 122-127.

van Eldik, P., Van Der Waaij, E. H., Ducro, B., Kooper, A. W., Stout, T. A., Colenbrander, B., et al. (2006). Possible negative effects of inbreeding on semen quality in Shetland pony stallions. Theriogenology, 65, 1159-1170.