10-3-2016

Profiling Populations Using Neutral Markers, Major Histocompatibility Complex Genes and Volatile Organic Compounds as Modeled in Equus caballus Linnaeus

Ketaki Deshpande
Florida International University, kdesh002@fiu.edu

DOI: 10.25148/etd.FIDC001190

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

PROFILING POPULATIONS USING NEUTRAL MARKERS, MAJOR HISTOCOMPATIBILITY COMPLEX GENES AND VOLATILE ORGANIC COMPOUNDS AS MODELED IN *EQUUS CABALLUS* LINNAEUS

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOLOGY

by

Ketaki Deshpande

2016

To: Dean Michael R. Heithaus
College of Arts, Sciences and Education

This dissertation, written by Ketaki Deshpande, and entitled, Profiling Populations Using Neutral Markers, Major Histocompatibility Complex Genes and Volatile Organic Compounds as Modeled in *Equus Caballus* Linnaeus having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

_____
Kenneth Furton

_____
Robert Lickliter

_____
Jeffrey Wells

_____
Eric von Wettberg

_____
DeEtta Mills, Major Professor

Date of Defense: October 3, 2016

The dissertation of Ketaki Deshpande is approved.

_____
Dean Michael R. Heithaus
College of Arts, Sciences and Education

_____
Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2016

DEDICATION

I dedicate this dissertation to my parents Dr. Vijay Deshpande and Dr. Urmila Deshpande as well as my late grandmother Padmaja Deshpande. You will forever be my inspiration and strength. I will always strive to make you proud.

ACKNOWLEDGMENTS

This dissertation could not have been completed without the help and support from my committee members, colleagues, friends and family. It is my great pleasure to acknowledge people who have given me guidance, help and encouragement throughout my journey as a PhD student. I would like to thank my PhD advisor, Dr. DeEtta Mills, whose support, guidance and encouragement cannot be overstated. I am grateful to her for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this dissertation. The joy and enthusiasm she has for research was contagious and motivational for me, even during tough times in the Ph.D. pursuit. I am also thankful for the excellent example she has provided as a successful mentor and professor.

Besides my advisor, I would like to thank the rest of my advisory committee members Dr. Kenneth Furton, Dr. Jeffrey Wells, Dr. Robert Lickliter, and Dr. Eric vonWettberg for their important suggestions and remarks in this research, but also for the hard questions, which incentivized me to widen my research from various perspectives. I thank them for their time and effort in service on my doctoral committee despite their already heavy loads of responsibility. I would like to thank the Department of Biology for providing me teaching assistantship for five years. I would like to gratefully acknowledge the Conservation Genetics Resources Journal for allowing me to reprint a publication as Chapter 2 of this dissertation.

A good support system is important to surviving and staying sane in grad school. My time at Florida International University was made enjoyable in large part due to my friends and colleagues that became a part of my life. My lab mates have contributed

immensely to my personal and professional time at Florida International University. I thank Beatrice Kallifatidis, Dr. Natalie Damaso, Lauren Martin, Evelyn Perez, Melissa Villareal, Ivy Cheung, Julian Mendel and Christina Burns for their stimulating discussions, friendships as well as good advice and especially for your consideration during the last few months by allowing me unlimited use of lab equipment so I may finish my experiments on time. My special appreciation goes to all the friends (too many to list here but you know who you are!) for their invaluable assistance, support and friendship that I needed during the period of completing my PhD. I am enormously indebted to Dr. Priyanka Kushwaha, and Dr. Deepthi Nori who have offered me great help both psychologically and physically during the rough periods of graduate school. These two friends formed the core of my research time at Florida International University and I am glad that they will remain an integral part of my life forever.

Finally, my deepest gratitude to my parents, my brother and sister-in-law, whose support and constant encouragement helped me through the hard times of this program. My sincere appreciation is expressed to them for their patience, warm support, love, understanding, and inspiration. Without their blessings and encouragement, I would not have been able to finish this work.

ABSTRACT OF THE DISSERTATION

PROFILING POPULATIONS USING NEUTRAL MARKERS, MAJOR HISTOCOMPATIBILITY COMPLEX GENES AND VOLATILE ORGANIC COMPOUNDS AS MODELED IN *EQUUS CABALLUS* LINNAEUS

by

Ketaki Deshpande

Florida International University, 2016

Miami, Florida

Professor DeEtta Mills, Major Professor

Assessing the genetics of wild animal populations aims to understand selective pressures, and factors whether it be inbreeding or adaptation, that affect the genome. Although numerous techniques are available for assessing population structure, a major obstacle in studying wild populations is obtaining samples from the animals without having to capture them, which can lead to undue distress and injury. Therefore, biologists often use non-invasive sampling methods (i.e., collection of feces, hair) to extract host DNA. In this study, new DNA extraction protocols were developed that improved the quality and quantity of DNA obtained from fecal matter. Fecal samples aged up to Day 6 as well as field samples with unknown days since defecation were successful in individualization of the contributors using microsatellites and were further used to demonstrate kinship. Neutral markers such as short tandem repeat, and mitochondrial D-loop sequences are used for assessing relatedness and evolutionary relationships and can mutate without detrimental effects on the organism. Loci, such as the major histocompatibility complex (MHC), adapt more rapidly under selective pressure such as

parasite load, or resistance to diseases and support natural selection processes. Analysis of the neutral microsatellites in Big Summit feral horse population demonstrated a population lacking diversity and trending towards being an inbred population. However, examination of the MHC genes showed maintenance of greater variation that may be the result of selection pressures. The MHC similarity and lower genetic demarcation between geographically separated horse populations further indicated effect of selection pressures in preserving diversity at the MHC genes. Although such molecular markers are used in profiling populations, the current study was also successful in demonstrating the use of individual odor profiles as an additional profiling tool. Volatile organic compounds (VOC) obtained from hair of domestic horses were able to individualize horses as well as differentiate between horse breeds and display kinship. The relation of genetics to odor phenotype is of interest as the inherent polymorphic nature of MHC genes has the potential to generate unique combinations of genotypes that presumably produce distinct odor phenotypes. Subsequently, this study was able to show a significant correlation between MHC genotypes and VOC odor profiles in horses. Understanding the relationship between MHC and odor using domestic horses with known relatedness provides evidence that these same correlations may be applicable to wild equids and dictates their harem hierarchal social structure.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

FIGURE PAGE

LIST OF ABREVIATIONS AND ACRONYMS

AFLP.....Amplified Fragment Length Polymorphisms

AMOVA.....Analysis of Molecular Variance

AML.....Appropriate Management Level

ANOSIM.....Analysis of Similarities

AQHA.....American Quarter Horse Association

BLM.....Bureau of Land Management

bp.....Base pair

CAP.....Constrained Analysis of Principal Coordinates

CAR.....Carboxen

CD8.....Cluster of Differentiation 8

DEPC.....Diethylpyrocarbonate

D-loop.....Displacement loop

DNA.....Deoxyribonucleic acid

dsDNA.....double stranded DNA

DVB.....Divinylbenzene

EDTA.....Ethylene diamine tetracetic acid

ELA.....Equine Leucocyte Antigens

F_{IS} .....Fixation index

FS.....Full Sibling

GC…….....Gas-Chromatography

H2.....Histocompatibility -2 gene

H_{Exp}.....Expected Heterozygosity

H_{Obs}.....Observed Heterozygosity

HCA.....Hierarchical Cluster Analysis

HMA.....Herd Management Area

HS.....Half Sibling

HWE.....Hardy Weinberg Equilibrium

ISAG.....International Society of Animal Genetics

KPSI.....Kilo Pound per Square Inch

log.....Logarithm

MHC.....Major Histocompatibility Complex

mtDNA.....Mitochondrial DNA

MS.....Mass Spectroscopy

ND-PULSE.....No Disk Pulse

NJ.....Neighbor Joining

nMDS.....Non-Metric Multidimensional Scaling

PBS.....Phosphate Buffered Saline

PCoA.....Principal Coordinate Analysis

PCR .....Polymerase Chain Reaction

PCT.....Pressure Cycling Technology

PDMS.....Polydimethylsiloxane

PI.....Probability of Identity

PO.....Parent Offspring

POP.....Performance Optimized Polymer

PSI.....Pound per Square Inch

Q+PCT.....Qiagen protocol modified plus PCT

RAPD.....Random Amplification of Polymorphic DNA

RFLP.....Restriction Fragment Length Polymorphism

RFU.....Relative Fluorescent Units

U.....Unrelated

UNK.....Unknown

UPGMA.....Unweighted Pair Group Method with Arithmetic Mean

USDA-FS.....United States Department of Agriculture – Forest Services

SE.....Standard Error

SIMPER.....Similarity Percentage Analysis

SNP.....Single Nucleotide Polymorphisms

SPME.....Solid-phase Microextraction

STR.....Short Tandem Repeat

SSCP.....Single Strand Conformational Polymorphism

VOC.....Volatile Organic Compound

WTR.....Well to Read Distance

CHAPTER 1

1. INTRODUCTION

1.1. Conservation genetics

Conservation genetics rests firmly on the principle that inbreeding depression, resulting from mating between close relatives, and lower levels of genetic diversity will reduce fitness and long-term adaptability in natural populations (Frankham, 1995). Understanding the causes and consequences of genetic diversity in natural populations, requires methods to quantify observed or expected diversity within and between populations (Ljungqvist, et al., 2010).

1.1.1. Major issues in conservation genetics

The major issue in conservation biology is gathering a deeper understanding of the impactful and natural forces causing species decline within a time frame that can help stop extinctions. Conservation genetics looks specifically at the deleterious effects of inbreeding and the loss of genetic diversity that hamper the ability of individuals to adapt in response to environmental change. Patterns in genetic diversity are the result of combined historical and contemporary effects on evolutionary forces, including genetic drift, gene flow and natural selection. Fragmentation of populations, reduction in gene flow, effects of random processes overriding natural selection and inbreeding may further lead to increased expression of deleterious recessive traits. Additionally, the effects of small population size are of foremost concern in conservation biology since endangered species have small or declining populations (Frankham, 1996; Randi, et al., 2000). Such

populations suffer from inbreeding and loss of genetic diversity resulting in increased extinction risks. Consequently, a major objective of genetic management is to minimize inbreeding and consequential loss of genetic diversity. Overall, the central questions in conservation genetics are: (1) the degree to which genetic bottlenecks, genetic drift and low effective population size (Lande & Barrowclough, 1987) will reduce genetic diversity within a population; and (2) the impact of this reduction on the population's long-term viability (Frankham, 1995).

1.2. Sampling in Conservation

Sampling methods in wildlife studies can be grouped into three categories; destructive, non-destructive and non-invasive sampling. Destructive sampling is where the animal is killed to obtain the tissues necessary for genetic analysis, while non-destructive sampling includes capture of the animal where a biopsy is done or an invasive blood sample is drawn (Taberlet & Luikart, 1999). Recently, non-invasive genetic sampling has become a widely accepted method for studying wildlife populations that does not stress the animals in any way (Taberlet, et al., 1999).

1.2.1. Non-invasive sampling

The method entails collecting biological samples that have been left behind by the animal as a source of host DNA that can be collected without having to catch or disturb the animal. Non-invasive sampling methods involving DNA extracted from feces have been used for genetic studies in primates (Morin, et al., 2001), bats (Vege & McCracken, 2001), coyotes (Kohn, et al., 1999), elephants (Wasser, et al., 2004) and bears (Taberlet,

et al., 1997) but DNA recovery is often difficult and limits the success rate of using fecal matter as a source of host DNA.

1.2.2. Advantages of Non-invasive sampling

Non-invasive sampling and extraction of viable DNA from hair, feces, saliva, eggshell membranes, feathers and even urine (Beja-Pereira, et al., 2009) have been done, with feces and hair being the most commonly used sources (Broquet, et al., 2007). Fecal samples often contain a high number of sloughed intestinal epithelial cells, which contain the DNA of the organism (De, et al., 2015). Feces also provide information on dietary habits and health status of the animal. Furthermore, non-invasive sampling is essential in cases where particular species are highly elusive, endangered or spread over vast territories.

1.2.3 Limitations of non-invasive sampling

While non-invasive sampling has clear advantages, there are still problems associated with its use (Broquet & Petit, 2004). The current limitations are low quantities and quality of DNA (Bonin, et al., 2004), contamination by non-target DNA from the animals' diet, and fecal samples often contain polymerase inhibitors which dramatically affects the polymerase chain reaction (PCR). Often these drawbacks lead to low amplification rates, and high genotyping errors, such as false alleles and allelic dropout (Taberlet, et al., 1999).

1.3. Current methods in conservation

Advances in genetic techniques and instrumentation allow for rapidly screening of large numbers of individuals and has revolutionized the application of DNA technologies for wildlife management. Different types of DNA markers are used in conservation genetics: mitochondrial DNA (mtDNA) sequences (Ashley, et al., 1990), microsatellites, Single Nucleotide Polymorphisms (SNPs) (Vignal, et al., 2002), Random Amplification of Polymorphic DNA (RAPDs) (DeSalle & Amato, 2004), Amplified Fragment Length Polymorphisms (AFLPs) (Bensch & Åkesson, 2005; Vos et al., 1995), Restriction Fragment Length Polymorphisms (RFLPs), and Single Strand Conformational Polymorphisms (SSCPs) (Piggott & Taylor, 2003).

1.3.1. Mitochondria

The typical vertebrate mitochondrial genome (mtDNA) is a circular, haploid genome of ~17 000 base pairs (bp) encoding for genes involved primarily in cellular adenosine triphosphate production (Wolstenholme, 1992). The mitochondrial genome is maternally inherited and typically lacks recombination. In animals, a relatively high rate of mutation and high copy number, compared with many nuclear DNA markers, make mtDNA sequences very useful in many analyses (Ishida, et al., 1994).

1.3.2. Short Tandem Repeats (STR)

Short tandem repeats or STRs (also called microsatellites) are sequence of repeated nucleotides, usually between 75 and 300 base pairs (bp) long. These are found in non-coding regions of the genome and are considered ‘selectively neutral’ (Queller &

Goodnight, 1989; Queller, et al., 1993). Most STR markers are highly polymorphic with a mutation rate of $\sim 10^{-5}$ to 10^{-3} mutations per locus per generation (Weber & Wong, 1993). The STR markers have been employed to analyze population structure within species, detect differences between closely related species, identify past population bottlenecks (Franks, et al., 2011), provide an effective population size estimation (census), assign individuals to populations (Hansen, et al., 2002) and estimate migration and gene flow (Falush, et al., 2003). Additionally they are also used for individual identification (genetic assignment), parentage and kinship analyses (Smith, et al., 2000) as well as quantifying population genetic diversity.

Population genetic studies often focus on such neutral markers rather than the adaptive, genetic variation in functional genes. Selection pressures do not directly target neutral markers such as mitochondrial displacement loop (D-loop), microsatellites or single nucleotide polymorphism. These markers are used to quantify genetic diversity (Aberle, et al., 2004; Valera, et al., 2005), track dispersal patterns (Pope, 1992), measure inbreeding (Luis, et al., 2007; Pusey & Wolf, 1996), determine relationships (Tozaki et al., 2001) as well as understand population structure and history (Zechner, et al., 2002). These properties make STRs ideal genetic markers for conservation genetics and wildlife management. However, neutral markers do not provide information on adaptive changes that affect genes under selection, such as the highly variable major histocompatibility complex (MHC). Therefore, studying the distribution of neutral markers within populations brings forth the question: Will genetic diversity be similar throughout the genome, or are some loci more affected than others? (Frankham, 1996; Madsen, et

al.,1999). Emerging evidence shows that patterns of variation and divergence in adaptive traits are not always associated with the same variation in neutral markers and several studies (Hughes, 1991; Sommer, 2005) have questioned the validity of using only neutral markers for development of conservation strategies. Therefore, the use of genetic markers linked to adaptive traits, including genes involved in immune defense, reproduction and some physiological functions is suggested (Hughes, 1991; Hughes & Yeager, 1998; O'Brien, 1994).

1.4. Major Histocompatibility Complex markers

The Major Histocompatibility Complex (MHC) of vertebrates is a highly polymorphic multi-gene family that plays a pivotal role in the vertebrate immune system by encoding a collection of immune and non-immune related molecules (Snell, 1981). The MHC family includes two major subfamilies, Class I and Class II. The polymorphic Class I MHC molecules are glycoproteins expressed on the surface of all nucleated somatic cells and present peptides to cytotoxic T lymphocytes. Class II MHC molecules have a much more restricted expression pattern, in that they are expressed primarily on antigen-presenting cells of the immune system. In mammals, Class I and Class II genes are located in different regions on the same chromosome and are separated by the Class III subfamily containing genes involved in non immune function (Klein, 1986).

1.4.1. Evolution of MHC Polymorphisms

Class I and Class II genes have extraordinarily high levels of polymorphism (Hughes & Yeager, 1998). Two hypotheses have been proposed to explain this level of

MHC polymorphism: (i) pathogen-driven selection and (ii) MHC-based mate choice (Potts & Wakeland, 1993; Wedekind, et al., 1995). The pathogen-driven selection favors genetic diversity of the MHC genes through heterozygote advantage (Doherty & Zinkernagel, 1975) and frequency dependent selection (Slade & McCallum, 1992) and takes into account the central role of MHC in the vertebrate immune system. In a population exposed to an array of pathogens, it would be advantageous for an individual to be heterozygous at MHC loci since heterozygosity will be able to present a broad range of antigens and thus help resist various pathogens (Doherty & Zinkernagel, 1975). These selection pressures subsequently serve as the underlying reason for the MHC-based mate selection hypothesis. Mating preferences of individuals with dissimilar MHC would preferentially produce MHC heterozygous offspring that should have increased fitness. Such mating preferences are possible since MHC polymorphism have been suggested to influence individual odor profiles (Boyse, et al., 1982) that, in turn, are used to select a mate (Eizaguirre, et al., 2009; Penn, 2002).

1.4.2. MHC, odor and mate choice

Most animals can distinguish between individuals on the basis of a number of aspects, which include sex, reproductive status, familiarity, herd membership, kinship and individual identity (Thorn & Hurst, 2004). A number of studies have revealed that the MHC is associated with the expression of both odorous and non-odorous olfactory cues identified as volatile organic compounds (VOC) that can be detected by individuals of various species. Extensive studies have been carried out in rodents (Gheusi, et al., 1997; Yamazaki, et at., 1999; Zhang & Zhang, 2011), fish (Gerlach & Lysiak, 2006;

Olsson, et al., 2003; Reusch, et al., 2001), birds (Leclaire, et al., 2012), primates (Setchell, et al., 2010) and even humans (Wedekind, et al., 1995; Wedekind & Furi, 1997). These studies indicate that individuals have a distinct body-odor type, which is determined, at least in part, by the inherited MHC alleles, although the underlying mechanism is still unknown and direct associations have not been made. While a detailed pathway of MHC associated odor production is unclear, different hypotheses have been suggested to describe how MHC genes impact individual odor. These hypotheses include-

(i) The MHC molecule hypothesis, which suggests fragments of MHC molecules in biological fluids produce the odorous compounds (Ferst, et al., 1998; Singh, et al., 1987).

(ii) The peptide hypothesis proposes MHC molecules may alter metabolites or peptides in urine, which are responsible for the odorous compounds (Singer, et al., 1997; Yamaguchi, et al., 1981).

(iii) The microflora hypothesis suggests MHC genes may modify odor by influencing the population of commensal microbes that are determined by specific MHC alleles (Singh et al., 1990).

(iv) The carrier hypothesis proposes that MHC molecules could be changed to volatile aromatics (Pearse-Pratt, et al., 1992).

(v) The peptide-microflora hypothesis suggests that MHC molecules influence odor by changing the peptides that are available to commensal microbes for subsequent metabolism (Penn & Potts, 1998; Wedekind & Penn, 2000).

The production of odor, however, is not limited to one factor and can be attributed to skin oils, sweat and VOCs emanating from the skin surface. However, in different laboratory strains of mice and rats, MHC variation has been shown to influence the expression of urinary volatile and secondary metabolites in chemical signaling and thereby modulating behavioral patterns, both socially and reproductively. Studies on mice include mating preferences of histocompatibility-2 gene (H2) associated diversity in the AKR-H-2b congenic mouse strain that suggested a preference to mates with heterozygous genes (Yamazaki, et al., 1976; 1999). Odor identification in unconditioned, wild versus congenic mice depicted that classical MHC genes affect an individual's VOC profiles that subsequently can be detected by wild unconditioned mice. Studies have also looked at the influence of fetal MHC in house mice. It was observed that in a pool of genetically identical pregnant females with 9-18 day old fetuses of varied MHC type, the male mice were attracted to pregnant females carrying fetuses with dissimilar MHC genes than their own genotype. Studies have also been conducted on human populations to understand whether there is an influence or preference of odor and mate choice. These studies infer that in highly inbred populations such as the North American Hutterite, MHC disassortative selection may be adaptive so as to avoid inbred population structure (Wedekind, et al., 1995). Ober, et al., (1997) studied classical human leucocyte antigen (HLA) types for 400 couples from the Hutterite community and found significantly fewer HLA matches between husbands and wives than expected when taking into account the social structure of Hutterites (Ober, et al., 1997).

In most of these studies, behavioral responses to complex body secretions (urine in mice, and sweat on male worn T-shirts in the human studies) were recorded, but the chemistry of these secretions and the mechanisms involved in their formation have received relatively little attention. Yamaguchi et al., (1981) found that mice can recognize the body odor of a potential mate with dissimilar MHC using urine samples. They found different, mostly unidentified, carboxylic acids to be important discriminators whose relative abundances were associated with different MHC types (Yamaguchi, et al., 1981). They concluded that the type of body odor results from a complex odor of several acids, with the relative abundances differing, and not their presence/absence contributing to the body-odor differences in mice.

1.5. Equids as a model system

Equids have a harem social structure (1-2 stallions, many mares), where stallions actively defend all members of their band against predators and other stallions. In the wild, young bachelor males, ostracized from their natal herds by the dominant stallion, commonly form their own temporary associations until they can successfully 'steal' mares from another herd or defeat the alpha stallion and take over a herd. During mating season, stallions act more aggressively to keep the mares within the herd and away from the bachelors.

Currently under the Wild Horse and Burro Act, America's wild feral horses are being protected and managed on herd management areas (HMA) or restricted public rangelands. Human encroachment and mandated management strategies are designed to maintain a certain population size on rangelands and limit inter-herd migrations among

these herds. Some smaller and geographically isolated HMA herds that are no longer able to roam freely through unrestricted migration corridors, cannot maintain complete random mating, and have limited mate choice, which forces mating between closely related individuals. Of concern are some herds that are showing signs of distress in terms of low survival of offspring, and higher mortality rates and birth defects (Ashley, 2004). Individuals in these isolated herds are becoming more genetically homogeneous which could make them more susceptible to infectious diseases and deleterious inherited traits, leading to inbreeding depression (Hedrick 2000). The question then follows that if such populations of wild horses are inbred and have similar odor signatures, how does that affect the herd social and breeding structure and subsequent mate choice? Thus, it is essential to understand the underlying factors that may govern mate choice in these populations further affecting social behavior and population structure.

Behavioral and observational studies of domestic and feral horses have revealed that horses communicate in many ways, including visual displays, sounds and through smell (Rubenstein, 1981; Rubenstein & Hack, 1992). Horses have a more highly developed sense of smell than humans, and they use their ability to distinguish different odors that are necessary to communicate and survive as part of a herd (Hothersall, et al., 2010). Horses use odor cues to identify other horses, particularly when a mare picks her foal of a large herd. During mating season, the stallion constantly checks mares to detect the ones in heat (estrus) using odor cues driven by hormonal changes in the mares (Marinier, et al., 1988; Stahlbaum & Houpt, 1989). Additionally, horses demonstrate scent-marking behavior, which is used by other horses for conspecific identification and

communication of dominance and a stallion's territory. Behavioral studies showed that horses exhibited a great interest in sniffing urine or feces to identify others. Mutual sniffing of the body and grooming between horses are also commonly observed during greetings (Rubenstein & Hack, 1992) and sexual behavior (Stahlbaum & Houpt, 1989).

1.6. Objectives of study

This project aimed to: (a) use non-invasive samples and improved DNA extraction techniques in order to assess genetic relatedness between different populations of horses—wild and domestic; (b) identify genetic structure in wild and domestic horses using both neutral markers such as STRs, mtDNA and adaptive markers like the MHC genes; and (c) identify correlation or patterns between adaptive MHC genes and odor production (VOCs) that could be used for individual identification and kin recognition. To date, no study has looked at neutral and adaptive markers across numerous equine breeds and correlated these independent data sets for a better understanding of selective adaptation and inbreeding analyses. These inclusive, concatenated analyses can better support major conservation strategies for America's wild horses or any managed wildlife populations.

1.6.1. AIM 1: Improved DNA profiles from aged horse feces using pressure cycling technology (PCT)

Hypothesis: There will be no difference in DNA yield obtained from Qiagen® DNA Stool Mini Kit and modified extraction method using PCT.

The first aim involved the development of a non-invasive sampling technique of fecal matter from domestic horses, to provide a viable protocol for an improved DNA extraction. Aged fecal samples as well as three pasture samples with unknown days since defecation were used to identify the contributors and analyze kinship. Development of the method should prove to be an extremely useful and reliable method for conservation or forensic cases where fecal matter may be the only sample available for analyses.

1.6.2. AIM 2: Genetic structure within the wild horse populations inhabiting Oregon's herd management areas (HMA)

Hypothesis: Genetic diversity will occur at neutral markers in wild horses because they can mate at random.

The second aim involved the analysis of genetic diversity using a non-invasive collection method to assess the genetic health of Big Summit herd and determine their genetic fitness. Understanding the genetic diversity for the isolated Big Summit herd of feral horses and the genetic structure of feral horses from nearby HMAs would enable enhanced conservation and management strategies.

1.6.3. AIM 3: Differentiation of three horse populations determined by Major Histocompatibility Complex

Hypothesis: There will be no difference in MHC genetic distance between geographically separated horse populations

The third aim involved characterization of seven MHC loci to study breed differentiation between three geographically separated horse populations. Evaluating MHC loci in subpopulations would help understand how genetic variation is maintained across distant populations. Additionally, understanding the mechanism of selective pressures would enhance the knowledge needed for maintaining diversity in managed populations.

1.6.4. AIM 4: The equine volatilome is unique but can also reflect kinship: a possible mechanism for inbreeding avoidance?

Hypothesis: The volatilomes of closely related horses will not be any more similar than the profiles of unrelated horses

The fourth aim involved identification of unique odor types from body hair amongst related and unrelated domestic horses. Evidence for equine body volatiles that are presumably used for individual discrimination and intra-species communication is lacking; therefore, this study examined the VOC profiles of domestic horses to determine the components of individual odor and whether these profiles can reliably indicate a degree kinship. The odor profiles generated by chemical analysis help understand the role of odor in individual identification, kin recognition, and its influence of subsequent mate selection in wild equids.

1.6.5. AIM 5: The correlation between genetics, body volatiles and relatedness: major histocompatibility complex genes and odor profiles as modeled in *Equus caballus*.

Hypothesis: MHC genotype does not influence VOC odor signatures; therefore odor similarity does not reflect genetic similarity or relatedness in horses.

The fifth aim involved the investigation of relationship between MHC genotype, relatedness and VOC odor profiles in domestic horses of two breeds. The study examined whether MHC genotype may influence odor signals, and that odor similarity may further reflect genetic kinship in horses.

1.7. Rationale of study

Currently in conservation, the management techniques often include assessment of microsatellites in order to quantify heterozygosity in populations but without really considering all the scientific implications toward conservation. Diversity within microsatellites alone would not ensure fitness within a population. There is evidence that genetic diversity in adaptive traits is not always linked with variation in these microsatellites (Ujvari & Belov, 2011; Sommer, 2005). Subsequently the use of these markers for development of conservation strategies is still disputed. In light of these shortcomings and the known association of MHC with unique odor types, it is imperative to study odor effects on mate selection, especially in highly inbred populations that can impact individual and overall herd fitness. The MHC genes can further be used to expand the repertoire of selection criteria to ensure the science behind conservation management can maintain healthy populations.

Linking MHC genetic data to VOCs, the combination of which could influence mate choice, social behavior, and overall fitness to maintain outbreeding and limit inbreeding will add to the basic knowledge of genotype influencing individual body odor and relationships.

1.8. References

Aberle, K. S., Hamann, H., Drögemüller, 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(4), 270-277.

Ashley, M. V., Melnick, D. J., & Western, D. (1990). Conservation genetics of the black rhinoceros (*Diceros bicornis*), I: Evidence from the mitochondrial DNA of three populations. Conservation Biology, 4(1), 71-77.

Ashley, M. C. (2004). Population genetics of feral horses: Implications of behavioral isolation. Journal of Mammalogy, 85(4), 611-617.

Beja-Pereira, A., Oliveira, R., Alves, P. C., Schwartz, M. K., & Luikart, G. (2009). Advancing ecological understandings through technological transformations in noninvasive genetics. Molecular Ecology Resources, 9(5), 1279-1301.

Bensch, S., & Åkesson, M. (2005). Ten years of AFLP in ecology and evolution: Why so few animals? Molecular Ecology, 14(10), 2899-2914.

Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C., & Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. Molecular Ecology, 13(11), 3261-3273.

Boyse, E. A., Beauchamp, G. K., Yamazaki, K., Bard, J., & Thomas, L. (1982). Chemosensory communication. A new aspect of the major histocompatibility complex and other genes in the mouse. Oncodevelopmental Biology and Medicine: The Journal of the International Society for Oncodevelopmental Biology and Medicine, 4(1-2), 101-116.

Broquet, T., Ménard, N., & Petit, E. (2007). Noninvasive population genetics: A review of sample source, diet, fragment length and microsatellite motif effects on amplification success and genotyping error rates. Conservation Genetics, 8(1), 249-260.

Broquet, T., & Petit, E. (2004). Quantifying genotyping errors in noninvasive population genetics. Molecular Ecology, 13(11), 3601-3608.

De, B. C., Patra, M. C., Kumar, S., Brahma, B., Goutam, D., Jaiswal, L., . . . De, S. (2015). Noninvasive method of DNA isolation from fecal epithelial tissue of dairy animals. Animal Biotechnology, 26(3), 211-216.

DeSalle, R., & Amato, G. (2004). The expansion of conservation genetics. Nature Reviews Genetics, 5(9), 702-712.

Doherty, P. C., & Zinkernagel, R. M. (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature, 256, 50-52

Eizaguirre, C., Yeates, S. E., Lenz, T. L., Kalbe, M., & Milinski, M. (2009). MHC-based mate choice combines good genes and maintenance of MHC polymorphism. Molecular Ecology, 18(15), 3316-3329.

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics, 164(4), 1567-1587.

Ferstl, R., Eggert, F., & Muller-Ruchholtz, W. (1998). Major histocompatibility complex-associated odours. Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association, 13(5), 1117-1119.

Frankham, R. (1995). Conservation genetics. Annual Review of Genetics, 29(1), 305-327.

Frankham, R. (1996). Relationship of genetic variation to population size in wildlife. Conservation Biology, 10(6), 1500-1508.

Franks, S. J., Pratt, P. D., & Tsutsui, N. D. (2011). The genetic consequences of a demographic bottleneck in an introduced biological control insect. Conservation Genetics, 12(1), 201-211.

Gerlach, G., & Lysiak, N. (2006). Kin recognition and inbreeding avoidance in zebrafish, *Danio rerio*, is based on phenotype matching. Animal Behaviour, 71(6), 1371-1377.

Gheusi, G., Goodall, G., & Dantzer, R. (1997). Individually distinctive odours represent individual conspecifics in rats. Animal Behaviour, 53(5), 935-944.

Hansen, M. M., Ruzzante, D. E., Nielsen, E. E., Bekkevold, D., & Mensberg, K. D. (2002). Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (Salmo trutta) populations. Molecular Ecology, 11(12), 2523-2535.

Hedrick, P. W. (2000). Applications of population genetics and molecular techniques to conservation biology. Genetics, Demography and Viability of Fragmented Populations, 4, 113– 125.

Hothersall, B., Harris, P., Sörtoft, L., & Nicol, C. J. (2010). Discrimination between conspecific odour samples in the horse (*Equus caballus*). Applied Animal Behaviour Science, 126(1), 37-44.

Hughes, A. L. (1991). MHC polymorphism and the design of captive breeding programs. Conservation Biology, 5(2), 249-251.

Hughes, A. L., & Yeager, M. (1998). Natural selection at major histocompatibility complex loci of vertebrates. Annual Review of Genetics, 32(1), 415-435.

Ishida, N., Hasegawa, T., Takeda, K., Sakagami, M., Onishi, A., Inumaru, S., . . . Mukoyama, H. (1994). Polymorphic sequence in the D-loop region of equine mitochondrial DNA. Animal Genetics, 25(4), 215-222.

Klein, J. (1986). Natural history of the major histocompatibility complex Wiley.

Kohn, M. H., York, E. C., Kamradt, D. A., Haught, G., Sauvajot, R. M., & Wayne, R. K. (1999). Estimating population size by genotyping faeces. Proceedings. Biological Sciences / the Royal Society, 266(1420), 657-663.

Lande, R., & Barrowclough, G. F. (1987). Effective population size, genetic variation, and their use in population management. Viable Populations for Conservation, 87-123.

Leclaire, S., Merkling, T., Raynaud, C., Mulard, H., Bessiere, J. M., Lhuillier, E., . . . Danchin, E. (2012). Semiochemical compounds of preen secretion reflect genetic

make-up in a seabird species. Proceedings.Biological Sciences / the Royal Society, 279(1731), 1185-1193.

Ljungqvist, M., Åkesson, M., & Hansson, B. (2010). Do microsatellites reflect genome-wide genetic diversity in natural populations? A comment on Väli et al.(2008). Molecular Ecology, 19(5), 851-855.

Luis, C., Cothran, E. G., & Oom Mdo, M. (2007). Inbreeding and genetic structure in the endangered Sorraia horse breed: Implications for its conservation and management. The Journal of Heredity, 98(3), 232-237.

Madsen, T., Shine, R., Olsson, M., & Wittzell, H. (1999). Conservation biology: Restoration of an inbred adder population. Nature, 402(6757), 34-35.

Marinier, S., Alexander, A., & Waring, G. (1988). Flehmen behaviour in the domestic horse: Discrimination of conspecific odours. Applied Animal Behaviour Science, 19(3), 227-237.

Morin, P. A., Chambers, K. E., Boesch, C., & Vigilant, L. (2001). Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). Molecular Ecology, 10(7), 1835-1844.

Ober, C., Weitkamp, L. R., Cox, N., Dytch, H., Kostyu, D., & Elias, S. (1997). HLA and mate choice in humans. The American Journal of Human Genetics, 61(3), 497-504.

O'Brien, S. J. (1994). A role for molecular genetics in biological conservation. Proceedings of the National Academy of Sciences of the United States of America, 91(13), 5748-5755.

Olsson, M., Madsen, T., Nordby, J., Wapstra, E., Ujvari, B., & Wittsell, H. (2003). Major histocompatibility complex and mate choice in sand lizards. Proceedings. Biological Sciences / the Royal Society, 270 Suppl 2, S254-6.

Pearse-Pratt, R., Schellinck, H., Brown, R., & Roser, B. (1992). Evolutionary and immunological implications of the role of the MHC in olfactory signaling. Chemical Signals in Vertebrates 6 (167-173) Springer.

Penn, D. J. (2002). The scent of genetic compatibility: Sexual selection and the major histocompatibility complex. Ethology, 108(1), 1-21.

Penn, D., & Potts, W. (1998). How do major histocompatibility complex genes influence odor and mating preferences? Advances in Immunology, 69, 411-436.

Piggott, M. P., & Taylor, A. C. (2003). Remote collection of animal DNA and its applications in conservation management and understanding the population biology of rare and cryptic species. Wildlife Research, 30(1), 1-13.

Pope, T. R. (1992). The influence of dispersal patterns and mating systems on genetic differentiation within and between populations of the red howler monkey (*Alouatta seniculus*). Evolution, 46(4) , 1112-1128.

Potts, W. K., & Wakeland, E. K. (1993). Evolution of MHC genetic diversity: A tale of incest, pestilence and sexual preference. Trends in Genetics, 9(12), 408-412.

Pusey, A., & Wolf, M. (1996). Inbreeding avoidance in animals. Trends in Ecology & Evolution, 11(5), 201-206.

Queller, D. C., & Goodnight, K. F. (1989). Estimating relatedness using genetic markers. Evolution, 43(2), 258-275.

Queller, D. C., Strassmann, J. E., & Hughes, C. R. (1993). Microsatellites and kinship. Trends in Ecology & Evolution, 8(8), 285-288.

Randi, E., Lucchini, V., Christensen, M. F., Mucci, N., Funk, S. M., Dolf, G., & Loeschcke, V. (2000). Mitochondrial DNA variability in Italian and East European wolves: Detecting the consequences of small population size and hybridization. Conservation Biology, 14(2), 464-473.

Reusch, T. B., HaÈberli, M. A., Aeschlimann, P. B., & Milinski, M. (2001). Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. Nature, 414(6861), 300-302.

Rubenstein, D. I. (1981). Behavioural ecology of island feral horses. Equine Veterinary Journal, 13(1), 27-34.

Rubenstein, D. I., & Hack, M. A. (1992). Horse signals: The sounds and scents of fury. Evolutionary Ecology, 6(3), 254-260.

Setchell, J. M., Vaglio, S., Moggi-Cecchi, J., Boscaro, F., Calamai, L., & Knapp, L. A. (2010). Chemical composition of scent-gland secretions in an old world monkey (*Mandrillus sphinx*): Influence of sex, male status, and individual identity. Chemical Senses, 35(3), 205-220.

Singer, A. G., Beauchamp, G. K., & Yamazaki, K. (1997). Volatile signals of the major histocompatibility complex in male mouse urine. Proceedings of the National Academy of Sciences of the United States of America, 94(6), 2210-2214.

Singh, P. B., Brown, R. E., & Roser, B. (1987). MHC antigens in urine as olfactory recognition cues. Nature 327, 161-164.

Singh, P. B., Herbert, J., Roser, B., Arnott, L., Tucker, D. K., & Brown, R. E. (1990). Rearing rats in a germ-free environment eliminates their odors of individuality. Journal of Chemical Ecology, 16(5), 1667-1682.

Slade, R. W., & McCallum, H. I. (1992). Overdominant vs. frequency-dependent selection at MHC loci. Genetics, 132(3), 861-864.

Smith, D. G., Kanthaswamy, S., Viray, J., & Cody, L. (2000). Additional highly polymorphic microsatellite (STR) loci for estimating kinship in rhesus macaques (*Macaca mulatta*). American Journal of Primatology, 50(1), 1-7.

Snell, G. D. (1981). Studies in histocompatibility. Science (New York, N.Y.), 213(4504), 172-178.

Sommer, S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Frontiers in Zoology, 2(1), 1.

Stahlbaum, C. C., & Houpt, K. A. (1989). The role of the flehmen response in the behavioral repertoire of the stallion. Physiology & Behavior, 45(6), 1207-1214.

Taberlet, P., Camarra, J., Griffin, S., Uhres, E., Hanotte, O., Waits, L., . . . Bouvet, J. (1997). Noninvasive genetic tracking of the endangered Pyrenean brown bear population. Molecular Ecology, 6(9), 869-876.

Taberlet, P., & Luikart, G. (1999). Non-invasive genetic sampling and individual identification. Biological Journal of the Linnean Society, 68(1-2), 41-55.

Taberlet, P., Waits, L. P., & Luikart, G. (1999). Noninvasive genetic sampling: Look before you leap. Trends in Ecology & Evolution, 14(8), 323-327.

Thom, M. D., & Hurst, J. L. (2004). Individual recognition by scent. In Annales Zoologici Fennici (765-787). Finnish Zoological and Botanical Publishing Board.

Tozaki, T., Kakoi, H., Mashima, S., Hirota, K., Hasegawa, T., Ishida, 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.

Ujvari, B., & Belov, K. (2011). Major histocompatibility complex (MHC) markers in conservation biology. International Journal of Molecular Sciences, 12(8), 5168-5186.

Valera, M., Molina, A., Gutiérrez, J. P., Gómez, J., & Goyache, F. (2005). Pedigree analysis in the Andalusian horse: Population structure, genetic variability and influence of the Carthusian strain. Livestock Production Science, 95(1), 57-66.

Vege, S., & MCCracken, G. F. (2001). Microsatellite genotypes of big brown bats (*Eptesicus fuscu*s: Vespertilionidae, Chiroptera) obtained from their feces. Acta Chiropterologica, 3(2), 237-244.

Vignal, A., Milan, D., SanCristobal, M., & Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution, 34(3), 1.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., . . . Kuiper, M. (1995). AFLP: A new technique for DNA fingerprinting. Nucleic Acids Research, 23(21), 4407-4414.

Wasser, S. K., Shedlock, A. M., Comstock, K., Ostrander, E. A., Mutayoba, B., & Stephens, M. (2004). Assigning African elephant DNA to geographic region of origin: applications to the ivory trade. Proceedings of the National Academy of Sciences of the United States of America, 101(41), 14847-14852.

Weber, J. L., & Wong, C. (1993). Mutation of human short tandem repeats. Human Molecular Genetics, 2(8), 1123-1128.

Wedekind, C., & Furi, S. (1997). Body odour preferences in men and women: Do they aim for specific MHC combinations or simply heterozygosity? Proceedings. Biological Sciences / the Royal Society, 264(1387), 1471-1479.

Wedekind, C., & Penn, D. (2000). MHC genes, body odours, and odour preferences. Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association, 15(9), 1269-1271.

Wedekind, C., Seebeck, T., Bettens, F., & Paepke, A. J. (1995). MHC-dependent mate preferences in humans. Proceedings. Biological Sciences / the Royal Society, 260(1359), 245-249.

Wolstenholme, D. R. (1992). Animal mitochondrial DNA: Structure and evolution. International Review of Cytology, 141, 173-216.

Yamaguchi, M., Yamazaki, K., Beauchamp, G. K., Bard, J., Thomas, L., & Boyse, E. A. (1981). Distinctive urinary odors governed by the major histocompatibility locus of the mouse. Proceedings of the National Academy of Sciences of the United States of America, 78(9), 5817-5820.

Yamazaki, K., Singer, A., Curran, M., & Beauchamp, G. K. (1999). Origin, functions, and chemistry of H-2 regulated odorants. Advances in Chemical Signals in Vertebrates (pp. 173-180) Springer.

Zechner, P., Sölkner, J., Bodo, I., Druml, T., Baumung, R., Achmann, R., . . . Brem, G. (2002). Analysis of diversity and population structure in the Lipizzan horse breed based on pedigree information. Livestock Production Science, 77(2), 137-146.

Zhang, Y. H., & Zhang, J. X. (2011). Urine-derived key volatiles may signal genetic relatedness in male rats. Chemical Senses, 36(2), 125-135.

CHAPTER 2

2. IMPROVED DNA PROFILES FROM AGED HORSE FECES USING PRESSURE CYCLING TECHNOLOGY

This chapter has been published:

Deshpande, K., Villarreal, M., & Mills, D. K. (2016). Improved DNA profiles from aged horse feces using pressure cycling technology. *Conservation Genetics Resources,* 1-9.

2.1. Introduction

Fecal matter sampling has received much attention from molecular biologists as a source of host DNA. Feces contain hosts cells shed from epithelial lining of the digestive system plus a complex mixture of gut microorganisms, food particles, digestive enzymes, bile salts and mucus (Kohn & Wayne 1997; Fernando et al., 2003; Piggott & Taylor 2003; Wehausen et al., 2004), many of which act as inhibitors for the polymerase chain reaction (Kohn & Wayne 1997). These factors plus the age of the feces at collection time often make it difficult to obtain a complete genetic profile of the contributing individual.

Feces is a less considered but potentially significant item of evidence that can be encountered at crime scenes (Brettell et al., 2009) and can be used to prove the defendant had been present at the scene of the crime. For instance, there are reported cases where DNA has been extracted from fecal stains found on the clothing of rape victims, feces recovered from places that have been burglarized and dog feces found on the sole of the suspect's sneakers as well as clothing have helped link suspects to the crime (Norris &

Bock 2000; Johnson et al., 2005; https://www.vgl.ucdavis.edu/forensics/links.php). There is also an increase in forensic use of animal DNA to determine the identity of samples to species or assigning samples to an individual or to kin (Linacre et al., 2011). Apart from its forensic application, fecal sampling is commonly used in wildlife studies for identification of species, assigning genotypes to an individual, population, or geographic origin, gender, estimate population size, study diet and feeding patterns, and to conduct disease surveys (Waits & Paetkau 2005). Also, fecal matter (scat) provides a non-invasive sample when capture of an animal is not possible.

There are, however, some disadvantages of using fecal samples. These include: (i) inefficient recovery of host cells and low-quality, degraded host DNA (Bonin et al., 2004; McKelvey & Schwartz 2004); (ii) samples are highly contaminated by non-target DNA of the food source and gut microbes (Bradley & Vigilant 2002; Pompanon et al., 2005); and (iii) complex mixtures of inhibitors which can lead to no or low amplification rates and result in the introduction of artifacts and genotyping errors such as allelic drop-out or drop-in. Even with these disadvantages, fecal matter may provide the only sample available from which to obtain a genetic profile (Kohn 2010). Numerous commercial extraction kits have been evaluated for their cell lysis efficiency and high-yield of clean DNA in order to increase PCR performance (Nechvatal et al., 2008; Yu & Morrison, 2004; Anderson and Lebepe-Mazur, 2003). The commercial QIAamp® Stool Kit by Qiagen has been shown to have high extraction efficiencies and a relatively high DNA yield that works well in downstream applications (Nechvatal et al., 2008).

Pressure cycling technology (PCT) is a cell lysis method that employs hydrostatic pressure in rapid succession of low ambient and ultra-high pressures to induce mechanical stress on cells and disrupt biomolecular interactions (Gross et al., 2008a; Tao et al., 2003). The high-pressure destabilizes the cell membrane, rupturing and releasing cellular components including DNA, RNA, proteins, and lipids (Gross et al., 2008a). PCT has been used successfully for diverse sample types including blood, bone (Yuan et al., 2011), animal and plant tissue (Harrington et al., 2004; Okubara et al., 2007), insects, small organisms and microbes (Tao et al., 2003; Smejkal et al., 2006; Garrett et al., 2002) and cell lines (Gross et al., 2008b). More recently, it has been shown to improve differential DNA extractions from cellular mixtures, which are commonly encountered in forensic casework (Nori & McCord 2015). In the present study, the ability of pressure cycling to differentially disrupt the epithelial cells was crucial in obtaining increased host DNA from the horse fecal samples.

The objectives of this study were: (i) to use PCT to improve the differential host cell lysis and increase DNA yield from fecal matter; (ii) to obtain DNA profiles of individual horses from aged fecal samples using this PCT optimized method; and (iii) to test the methods on fecal matter gathered from a pasture as representative samples, where the time of deposition and contributor of the feces were unknown and to "match" the sample with the horse of origin.

2.2. Methodology

2.2.1. Sources of fecal samples

Ten domestic horses that had previously been DNA typed using hair samples provided the DNA reference profiles. Fecal samples were collected from each horse's stall within an hour of defecation. The outside of fecal boluses were sampled with three separate cotton swabs moistened with 1X Phosphate buffered saline (PBS), and all swabs were extracted within an hour of collection. After the initial (fresh) sampling on day zero, the same fecal samples were placed in air permeable containers and kept at an average ambient outdoor temperature of 17°C. The same swab collection protocol was used for each aged sample on days 2, 4 and 6 (N=30 per time point) in order to determine the temporal span within which a fecal sample could be sampled and still provide a useful DNA profile.

2.2.2 Pasture study

To test whether a DNA profile could be obtained from an unknown contributor where days since defecation were not known and be matched to that individual, five random samples from a nearby pasture were collected. Swabs were collected in triplicate (N=15) as described previously and DNA was extracted within an hour of collection. The horses that had access to that pasture were a subset of the ten horses used in this study and each individual had a known reference genotype. The physical condition and texture of each fecal sample from the field were recorded (Table 1).

Table 1. DNA quantitation and amplified alleles for five unknown pasture samples. At the time of fecal bolus swabbing on sampling days the average ambient outside temperature was 17°C.

Unknown Samples (UNK)	Sample description	Average DNA Yield (ng/µL)	Amplified # of Alleles	Alleles not amplified
UNK 1	Semi dry exterior, moist inside	19.40 ±2.4	12	
UNK 2	Dry exterior with moist patches	12.03 ±3.1	11	HMS3
UNK 3	Moist bolus but disintegrating	15.90 ±3.3	11	
UNK 4	Dried hard bolus, lighter in weight	3.30 ±0.7	4	HMS6, HMS3, HTG7, HTG6
UNK 5	Extremely dry bolus both exterior and interior	5.90 ±1.9	6	HMS6, HMS3, HTG7

2.2.3 Cell lysis and DNA Extraction

Samples were extracted using the following protocol for all swabs (N=135). To assess if pressure cycling could enhance host cell lysis and DNA yields from fecal samples, a comparison was first made to a modified Qiagen QIAmp® DNA Stool Mini Kit (Qiagen, Valencia, CA) extraction. The modifications to the kit protocol were as follows: Each swab was placed in a 2 mL microfuge tube and incubated overnight in 1 mL of Buffer ASL with 25 µL of proteinase K at 55° C in a thermo-shaker (Eppendorf AG, Hamburg, Germany) set at 300 rpm (Archie et al., 2003). After overnight incubation, only half of an InhibitEX tablet was added to the tube instead of a full tablet (Renan et al. 2012). Since proteinase K was added to the overnight digestion, it was omitted during sample digestion in Buffer AL (Archie et al. 2003). To enhance the binding of DNA to the Qiagen column, 1 µL of carrier RNA along with Buffer AL was added prior to incubation at 70° C. Finally, the DNA was eluted in 50 µL of Buffer AE after incubation at room temperature for five minutes prior to the final centrifugation step.

2.2.4. Pressure Cycling Technique (PCT)

To assess if pressure cycling could enhance host cell lysis, a PCT step was incorporated into the extraction protocol described above. After the overnight incubation step, the swab (including the Buffer ASL that was incubated overnight with swab) was placed into a no-disk PULSE tube (ND-PULSE tubes, Pressure Biosciences Inc., South Easton, MA). The tube was sealed with a cap on one end of the PULSE tube and a moveable ram on the other end. Each swab was subjected to 50 rapid, repeated pressure cycles of 15,000 psi in the Barocycler NEP2320 (Pressure Biosciences Inc., South Easton, MA). Each cycle provided 20 seconds of high pressure followed by 10 seconds of atmospheric pressure. The purpose of a lower pressure (15k psi) was to optimize lysis of the horse epithelial cells and minimize the lysis of the plant and microbial cells. After the PCT procedure, the samples were extracted according to the modifications listed above. The recovered DNA was quantified using Qubit™ 2.0 Fluorometer (Table 2) to provide a comparison of the DNA yield between the two extraction methods, even though it was recognized that the DNA represented a mixture of genomic DNA from horse, plant and microbes. Quantification data were used to normalize extracted DNA concentrations to 1 ng for PCR amplification.

Table 2. A comparison of the two DNA extraction methods based on the average DNA yield for Day 0 samples using Qubit™ 2.0 Fluorometer.

Sample	Average yield in ng/ µL	
	Modified QIAmp® DNA Stool Mini Kit protocol	Modified QIAmp® DNA Stool Mini Kit protocol with PCT
1	4.30 ±1.8	22.72 ±1.6
2	5.78 ±0.4	28.67 ±1.3
3	4.31 ±1.6	27.29 ±1.3
4	7.77 ±0.2	29.98 ±1.2
5	5.98 ±0.4	20.81 ±1.1
6	4.76 ±1.4	26.20 ±1.5
7	5.04 ±3.3	25.18 ±1.7
8	4.93 ±0.4	22.82 ±1.3
9	6.68 ±1.9	27.37 ±2.6
10	5.01 ±2.5	22.15 ±1.5
Unknown 1	4.16 ±0.1	19.40 ±2.4
Unknown 2	3.70 ±0.7	12.03 ±3.1
Unknown 3	3.78 ±1.7	15.90 ±3.3
Unknown 4	0.46 ±0.9	3.30 ±0.7
Unknown 5	1.97 ±2.2	5.90 ±1.9

2.2.5. PCR amplification and fragment analysis

Using fluorescently labeled forward primers, we amplified six equine loci (all dinucleotides; VHL20, HTG4, HTG6, HMS6, HTG7, and HMS3 (Guérin et al., 1994; Ellegren et al., 1992; Marklund et al., 1994; Van Haeringen et al., 1994) in one multiplex PCR after optimizing relative primer concentrations (Table 3). The six markers were a subset from the markers included in the StockMarks® for Equine 17-plex Genotyping Kit. They were selected not only because they were shorter markers that are best suited for degraded DNA but also to correctly compare sample profiles with the commercially available equine genotyping kit. The final reaction mix contained: 1X PCR buffer, 2.5 mM $MgCl_2$, 0.25 mM dNTPs, 4 µL of 6-plex primer mix (Table 2), 1.0 Unit AmpliTaq Gold® polymerase, 1 ng DNA and DEPC water to volume (15 µl). PCR cycling

conditions were: 95°C for 10 minutes for one cycle, then 30 cycles of 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 60 seconds with a final extension at 72°C for 60 minutes (Dimsoski 2003). One microliter of the PCR product was added to 11.5 µL Hi-Di™ Formamide and 0.5 µL GeneScan™ 600 LIZ® Size Standard, denatured at 95°C for 3 minutes and then immediately placed on ice for 3 minutes. The samples were separated on ABI 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA) using a 36 cm capillary array, POP-7 polymer (Applied Biosystems), 1X electrophoresis running buffer with EDTA (Applied Biosystems) with a 36 cm well to read distance (WTR), using Module DS33, filter G5v2. Samples were analyzed using GeneMapper® 3.7 software (Applied Biosystems, Foster City, CA). Analysis parameters were set to local Southern size calling and the minimum analytical threshold was set to 50 relative fluorescent units (RFUs).

Table 3. Primer sequences and concentrations per reaction for the equine 6-plex STRs and their size range.

Locus	Fluorochrome	Primer Sequence (5'-3')	Final Conc. (µM)	Size range (bp)	Reference
VHL20	6-FAM	F: CAAGTCCTCTTACTTGAAGACTAG R: AACTCAGGGAGAATCTTCCTCAG	0.25	87-105	Van Haeringen et al., 1994
HTG4	6-FAM	F: CTATCTCAGTCTTGATTGCAGGAC R: CTCCCTCCCTCCCTCTGTTCTC	0.07	127-139	Ellegren et al., 1992
HTG6	VIC	F: GTTCACTGAATGTCAAATTCTGCT R: CCTGCTTGGAGGCTGTGATAAGAT	0.17	84-102	Ellegren et al., 1992
HMS6	VIC	F: GAAGCTGCCAGTATTCAACCATTG R: CTCCATCTTGTGAAGTGTAACTCA	0.19	151-169	Guérin et al., 1994
HTG7	NED	F: CCTGAAGCAGAACATCCCTCCTTG R: ATAAAGTGTCTGGGCAGAGCTGCT	0.22	118-128	Marklund et al., 1994
HMS3	NED	F: CCATCCTCACTTTTTCACTTTGTT R: CCAACTCTTTGTCACATAACAAGA	0.13	148-170	Guérin et al., 1994

At present, there is no allelic ladder for the equine STR profiling, so allele calls are based upon the internal size standard and known positive controls. The StockMarks® for Equine 17-plex Genotyping Kit equine control DNA was used as a positive control for each run. In addition, each profile from the fecal samples was compared to the reference genotype for each horse. Since the horses in this study were Quarter Horses, most had been DNA typed at an external laboratory according to the American Quarter Horse Association (AQHA) registration requirements and these external DNA typing profiles were used to verify all allele calls for this study. Allele sizes were converted to allele repeat numbers based on latest equine typing standardization (Van de Goor et al. 2009).

2.2.6 Analysis

Two-tailed t-test assuming unequal variance was used to analyze corresponding peak heights of the expected alleles for method comparison and were performed using Excel (Figure 1). To compare the extraction efficiency, the average peak heights in RFUs for each allele and standard deviation were calculated (Figure 2). The averaged percent of amplified alleles compared to the known number of alleles for each horse was calculated for all alleles obtained on Days 0, 2, 4 and 6 (Table 4). ML-Relate software (Kalinowski et al., 2006) was also used to check if the six loci could identify kinship for the ten domestic horses (Table 5). ML-RELATE calculates maximum likelihood to estimate relatedness and assigns relationship from co-dominant genetic data, e.g. microsatellites. Profiles obtained on Days 0, 2, 4 and 6 were used for kinship analysis (Figure 3)

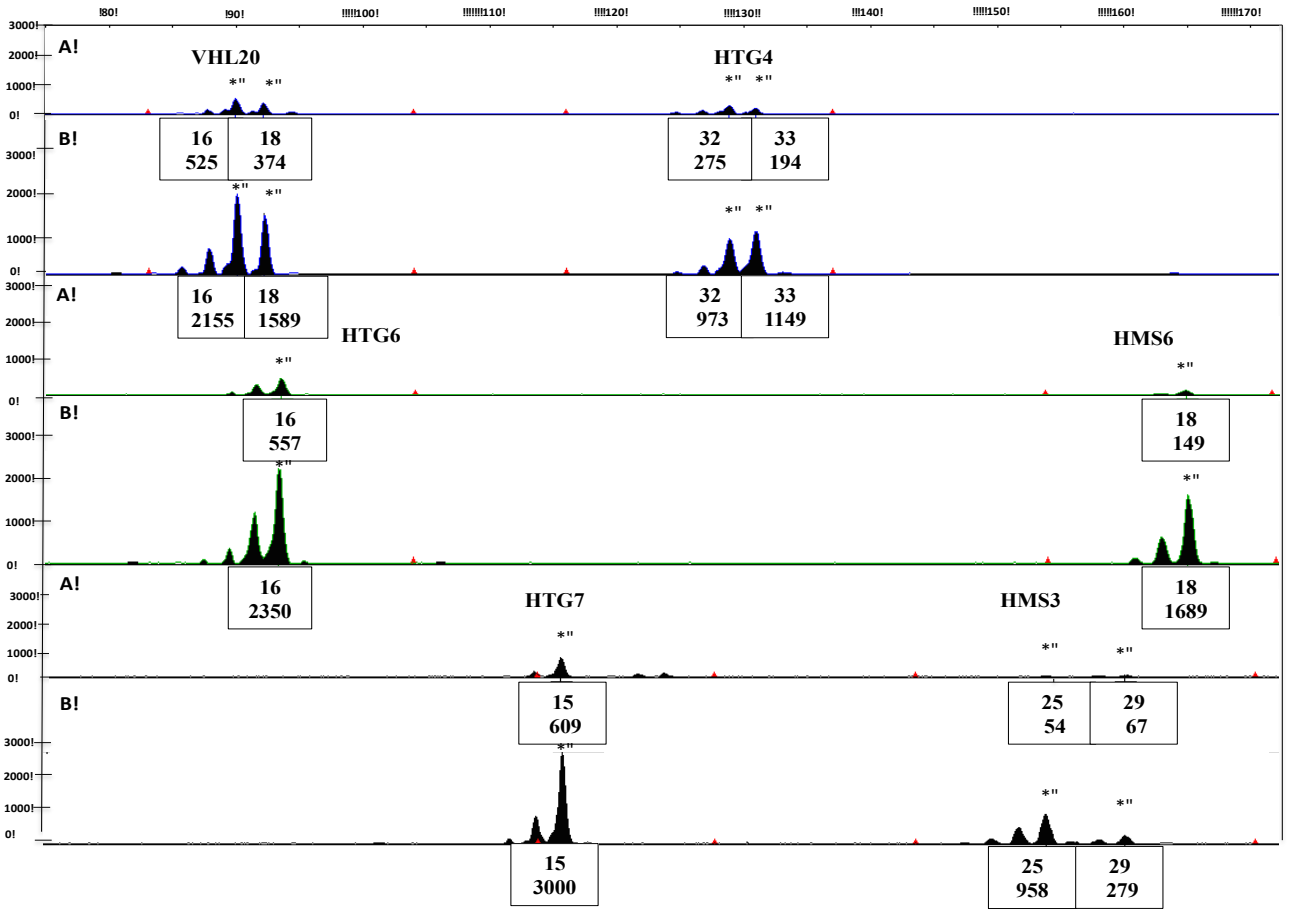

Figure 1. Example of an equine STR profile comparing DNA extraction techniques. (A) Modified Qiagen QIAmp®DNA Stool Mini Kit protocol only, and (B) Modified Qiagen QIAmp® DNA Stool Mini Kit protocol with PCT.
All nine alleles amplified matched the reference sample. The true peak is denoted by asterisk, while the repeat number and peak intensities are shown in the boxes below each peak.

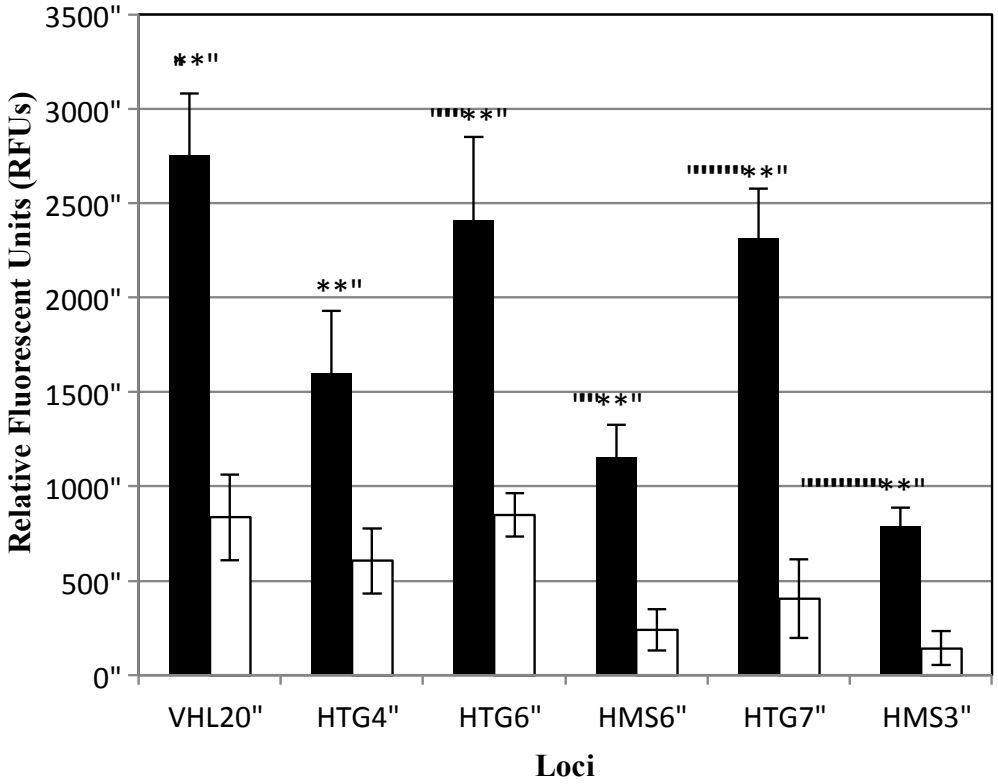

Figure 2. Comparison of PCT incorporated Qiagen modified DNA extraction protocol with Qiagen modified protocol for the six loci.

Table 4. DNA quantitation and number of amplified alleles for ten horses on days 0, 2, 4 and 6. At the time of fecal bolus swabbing on sampling days the average ambient outside temperature was 17°C.

Sample	Days	Average DNA Yield (ng/µL)	Amplified # of Alleles	% of Expected Alleles Amplified	Alleles not Amplified
Horse 1	0	22.72 ±1.6	12	100%	
	2	20.05 ±1.1	12	100%	
	4	18.60 ±1.7	10	83%	HMS3 (1 allele)
	6	11.40 ±1.2	10	83%	HMS3 (1 allele)
Horse 2	0	28.67 ±1.3	12	100%	
	2	24.99 ±2.1	12	100%	
	4	11.00 ±1.5	11	92%	HMS6
	6	10.52 ±0.9	5	42%	HMS6, HMS3, VHL20 HTG7
Horse 3	0	27.29 ±1.3	9	100%	
	2	26.85 ±0.9	9	100%	
	4	23.25 ±0.8	9	100%	
	6	11.37 ±1.0	9	100%	
Horse 4	0	29.98 ±1.2	12	100%	
	2	23.03 ±1.0	12	100%	
	4	11.31 ±0.9	8	67%	HMS6, HMS3
	6	10.78 ±0.7	8	67%	HMS6, HMS3
Horse 5	0	20.81 ±1.1	11	100%	
	2	20.20 ±0.4	11	100%	
	4	10 .17 ±1.1	11	100%	
	6	6.60 ±1.6	11	100%	
Horse 6	0	26.20 ±1.5	12	100%	
	2	22.46 ±1.9	12	100%	
	4	14.58 ±0.7	12	100%	
	6	14.10 ±1.9	12	100%	
Horse 7	0	25.18 ±1.7	12	100%	
	2	18.20 ±0.8	12	100%	
	4	11.56 ±2.4	12	100%	
	6	10.90 ±2.2	12	100%	
Horse 8	0	22.82 ±1.3	12	100%	

	2	19.64 ±1.1	10	83%	HMS3
	4	9.90 ±1.4	8	67%	HMS6, HMS3
	6	6.80 ±0.9	8	67%	HMS6, HMS3
Horse 9	0	27.37 ±2.6	11	100%	
	2	20.20 ±0.7	11	100%	
	4	15.80 ±0.6	10	91%	HMS3 (1 allele)
	6	7.40 ±1.9	10	91%	HMS6 (1 allele)
Horse 10	0	22.15 ±1.5	12	100%	
	2	19.53 ±1.4	11	92%	HMS6 (1 allele)
	4	11.24 ± 1.0	10	83%	HMS6, HMS3
	6	10.60 ±0.8	7	58%	HMS6, HMS3, VHL20 (1 allele)

Table 5. Kinship analysis for three unknown pasture samples based on the obtained genotype.

Partial profiles obtained from unknown pasture samples 4 and 5 were not used for kinship analysis. The five unknown fecal samples belonged to two horses (Horse 2 and Horse 8). Unknown fecal samples are listed here as in table as UNK 1, UNK 2 and UNK 3. The kinship abbreviations are HS – Half Sib, FS – Full Sib, PO – Parent Offspring, U – Unrelated, Self – Matched

Domestic horse	UNK 1	UNK 2	UNK 3
Horse 1	HS	U	U
Horse 2	Self	HS	HS
Horse 3	U	U	U
Horse 4	U	U	U
Horse 5	PO	PO	PO
Horse 6	U	U	U
Horse 7	HS	HS	HS
Horse 8	HS	Self	Self
Horse 9	U	U	U
Horse 10	FS	HS	HS

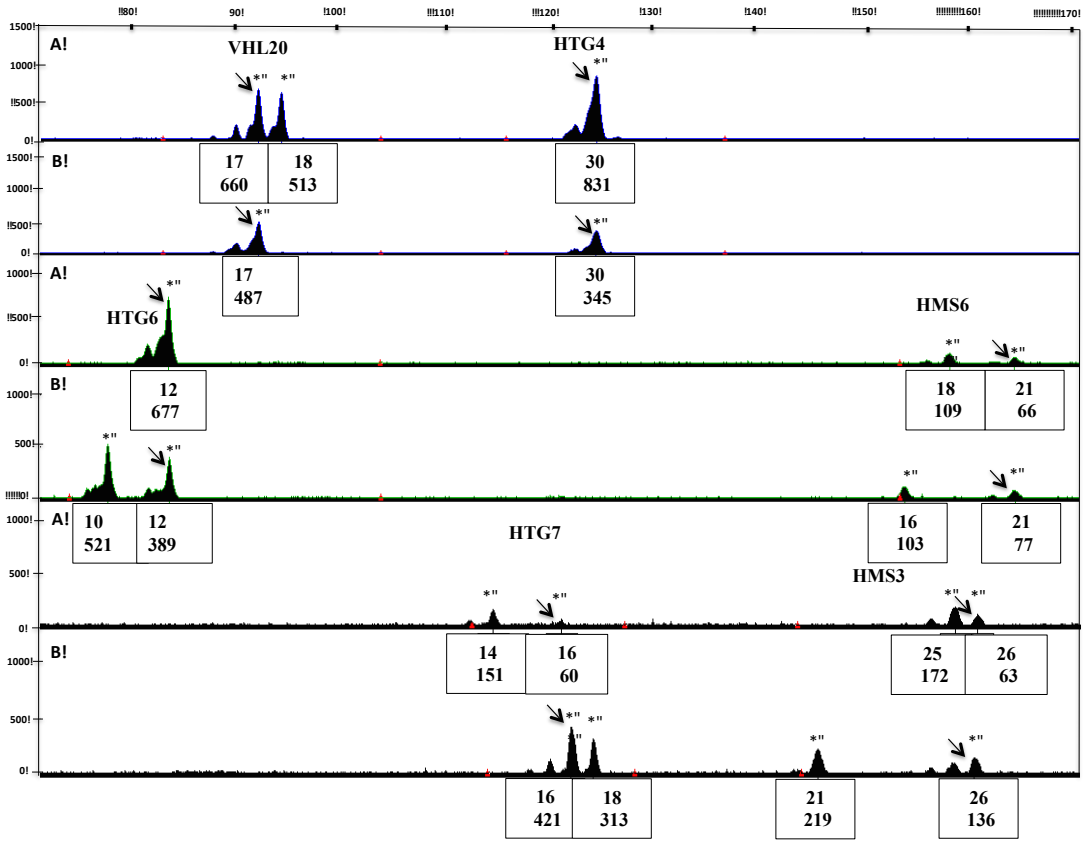

Figure 3. Equine profile showing relationship A Dam B Offspring. The true peak is denoted by asterisk, while represents the common alleles used to determine relationship. The repeat number and peak intensities are shown in the boxes below each peak.

2.3 Results

2.3.1 Extraction Comparison

Fecal samples from ten horse and five from the pasture samples were tested using the modified Qiagen QIAamp® stool kit protocol (Q) and Qiagen protocol modified plus PCT (Q+PCT) protocol. The Q+PCT method yielded the greatest concentrations of DNA from fresh and aged samples ($p<0.001$)(Table 1).

2.3.2 Profile Comparison

There was a 100% success rate obtained for full horse DNA profiles from fresh (Day 0) and fecal samples aged up to Day 2. Additionally, the profiles were identical to the reference samples verifying that extraction and purification methods were reliable and reproducible. For 90% of the samples, full horse profiles were obtained for Day 4 and 60% of the samples produced full profiles even six days post defecation. It was not unexpected that the percent of amplified alleles would decrease as the age of samples increased (Table 3).

For the full profiles obtained at Day 4, 100% of them could be used to assess kinship while 60% of the profiles obtained on Day 6 could still be used (Table 5). The M-L Relate data were designated as parent offspring, half or full sibling and unrelated. Profiles obtained from unknown pasture samples were determined to be fecal samples that matched horse 2 (Unknown 1, 3 and 5) and horse 8 (Unknown 2, 4) (Table 5). Using these "unknown" profiles, relationships of parent offspring and half sibs could be determined for unknown samples 1, 2 and 3. However, since only partial profiles were obtained for unknown pasture samples 4 and 5, relationship determination could not be done using the kinship software.

2.4 Discussion

Conventional methods or commercial kits for isolating DNA from feces extract total metagenomic DNA from the sample. However, an accurate individual genotype can be derived only when host cells are efficiently lysed and host DNA is not "diluted out" by

non-target organismal DNA in the sample. Preferably, this would involve selective disruption of host cells over other cell types in the sample and would release more target DNA versus non-host DNA. PCT has previously been compared to other standard DNA extraction methods for its ability to differentially lyse cells by utilizing pressure variation (Bradley et al., 2000; Okubara et al., 2007; Gross et al., 2008b). Gross et al. (2008b) applied PCT to selectively release mitochondria from rat kidney and skeletal muscle tissue without rupturing membrane protein complexes in the outer cellular membrane. Recently, differential extraction of sperm and vaginal epithelial cells was carried out using alkaline lysis combined with PCT. This method successfully separated male and female DNA from a mixture without compromising DNA recovery (Nori & McCord 2015). In this study, the sloughed, intact horse epithelial cells were more readily recovered from the surface of the feces with wetted swabs than from within fecal bolus, making surface swab sampling more amenable for host cell recovery.

The DNA extraction from fecal swabs using Q+PCT in this study dramatically increased the host DNA yield compared to only the modified Qiagen protocol as indicated in the three-fold increase of relative peak height intensities in the subsequent DNA profiles. The higher host DNA yield using Q+PCT extraction technique provided 100% amplification success for fresh as well as fecal samples aged up to four days and still useful partial profiles on Day 6 post defecation. PCT's ability to differentially lyse host cells provided a higher yield of horse DNA with minimal interference from other cellular macromolecules and microbial and plant DNA. Past studies using various DNA extraction techniques have demonstrated amplification success rates ranging from 70-

80% of brown bear scat samples (Bellemain et al., 2005), 27-77% profiles from wild boar samples (Kolodziej et al., 2012), and 40% of wolf scat (Creel et al., 2003) on samples with unknown age of defecation. The success of complete genotypes was after numerous replicates. While these studies generated individual DNA profiles, a significant loss of alleles or the addition of erroneous alleles in the sampled population often proved to be a problem requiring multiple replicates and limiting the information on individualization as well as number of individuals (Waits and Paetkau 2005). In the current study using Q+PCT, 100% of the profiles for Day 0, 2 and 90% at Day 4 showed no allelic dropouts or erroneous alleles. Only four of the samples on Day 6 showed a decrease in allele amplification of the larger amplicons (HTG4, HTG6 and HMS3) with some peaks dropping below the analytical threshold but no erroneous alleles were observed. These results were far better than those reported by Piggott (2005) and Santini et al. (2007) where allele amplification success of fecal samples from the brush-tailed rock-wallaby (*Petrogale penicillata*) and the red fox (*Vulpes vulpes*) (Piggott 2005) and wolf (*Canis lupus*) (Santini et al., 2007) declined rapidly within three days post-defecation and genotyping was not reliable or possible beyond one week. Ultimately, successful genotyping from complex samples depends on maximizing the concentration of host DNA, and minimizing non-target animal (food remnants) and plant DNA (Deuter et al., 1995; Flagstad et al., 1999; Fernando et al., 2003; Wehausen et al., 2004) and that goal was achieved in this current study using the Q+PCT extraction technique.

The use of genotypes obtained on Day 4 and 6 for successfully identifying relationship is important in cases where the days since defecation are unknown. The

profiles obtained from the five pasture fecal samples of unknown age were able to be matched to two of the reference horses. The ability to extract DNA and obtain complete or partial profiles from dry, a disintegrating fecal bolus would be crucial in the sampling of feces in the wild. Not only were the profiles from pasture samples able to identify the individual contributors but also three out of five profiles were also useful in determining kinship. In wild animal population studies, individualization would be crucial for population census. These DNA profiles could also provide valuable information on reproductive success; inbreeding, and genetic diversity and the data would be extremely helpful in formulating strategies in conservation and management.

In this study the Q+PCT method increased the DNA yield, which, in turn, improved PCR results and reduced common artifacts. In addition, it increased the ability to identify individuals even from fecal samples as old as six days post defecation. The method provided genotypes that could be used to identify the individual contributor and kinship from randomly chosen pasture samples. The future application of this optimized method could prove very useful in wildlife or domestic animal forensics cases where fecal matter may be the only available source of host DNA.

2.5 References

Anderson, K. L., & Lebepe-Mazur, S. (2003). Comparison of rapid methods for the extraction of bacterial DNA from colonic and caecal lumen contents of the pig. Journal of Applied Microbiology, 94(6), 988-993.

Archie, E. A., Moss, C. J., & Alberts, S. C. (2003). Characterization of tetranucleotide microsatellite loci in the African savannah elephant (*Loxodonta africana africana)*. Molecular Ecology Notes, 3(2), 244-246.

Bellemain, E. V. A., Swenson, J. E., Tallmon, D., Brunberg, S., & Taberlet, P. (2005). Estimating population size of elusive animals with DNA from hunter-collected feces: Four methods for brown bears. Conservation Biology, 19(1), 150-161.

Bonin, A., Bellemain, E., Bronken Eidesen, P., Pompanon, F., Brochmann, C., & Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. Molecular Ecology, 13(11), 3261-3273.

Bradley, D. W., Hess, R. A., Tao, F., Sciaba-Lentz, L., Remaley, A. T., Laugharn, J. A., & Manak, M. (2000). Pressure cycling technology: a novel approach to virus inactivation in plasma. Transfusion, 40(2), 193-200.

Bradley, B. J., & Vigilant, L. (2002). False alleles derived from microbial DNA pose a potential source of error in microsatellite genotyping of DNA from faeces. Molecular Ecology Notes, 2(4), 602-605.

Brettell, T. A., Butler, J. M., & Almirall, J. R. (2009). Forensic science. Analytical Chemistry, 81(12), 4695-4711.

Creel, S., Spong, G., Sands, J. L., et al. (2003). Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. Molecular Ecology, 12(7).

Deuter, R., Pietsch, S., Hertel, S., & Müller, O. (1995). A method for preparation of fecal DNA suitable for PCR. Nucleic Acids Research, 23(18), 3800-3801.

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

Ellegren H., Johansson M., Sandberg K. & Andersson L. (1992) Cloning of highly polymorphic microsatellites in the horse. Animal Genetics 23, 133-42.

Fernando, P., Vidya, T. N. C., Payne, J., et al. (2003). DNA analysis indicates that Asian elephants are native to Borneo and are therefore a high priority for conservation. PLoS Biology, 1(1), 110-115.

Flagstad, Ø., Røed, K., Stacy, J. E., & Jakobsen, K. S. (1999). Reliable noninvasive genotyping based on excremental PCR of nuclear DNA purified with a magnetic bead protocol. Molecular Ecology, 8(5), 879-883.

Garrett, P. E., Tao, F., Lawrence, N., Ji, J., Schumacher, R. T., & Manak, M. M. (2002). Tired of the same old grind in the new genomics and proteomics era? Targets, 1(5), 156-162.

Gross, V., Carlson, G., Kwan, A. T., Smejkal, G., Freeman, E., Ivanov, A. R., & Lazarev, A. (2008a). Tissue fractionation by hydrostatic pressure cycling technology: the unified sample preparation technique for systems biology studies. Journal of Biomolecular Techniques, 19(3), 189.

Gross, V., Lazarev, A. V., Lawrence, N. P., & Schumacher, R. T. (2008b). Isolation of mitochondria from cell cultures by PCT for proteomic analysis. Biotechniques, 45(1), 99.

Guérin, G., Bertaud, M., & Amigues, Y. (1994). Characterization of seven new horse microsatellites: HMS1, HMS2, HMS3, HMS5, HMS6, HMS7 and HMS8. Animal Genetics, 25(S2), 62-62.

Haeringen, H. V., Bowling, A. T., Stott, M. L., Lenstra, J. A., & Zwaagstra, K. A. (1994). A highly polymorphic horse microsatellite locus: VHL20. Animal Genetics, 25(3), 207-207.

Harrington, S., McCouch, S., Tao, F., Lawrence, N., & Schumacher, R. T. (2004). Use of pressure cycling technology (pct) for the release of dna from plants. In Annual Plant, Animal and Microbes Genomics Conference, San Diego, CA.

Johnson, D. J., Martin, L. R., & Roberts, K. A. (2005). STR-typing of human DNA from human fecal matter using the QIAGEN QIAamp® stool mini kit. Journal of Forensic Sciences, 50(4), 802.

Kalinowski, S. T., Wagner, A. P., & Taper, M. L. (2006). ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. Molecular Ecology Notes, 6(2), 576-579

Kohn, M. H., & Wayne, R. K. (1997). Facts from feces revisited. Trends in Ecology & Evolution, 12(6), 223-227.

Kohn, M. H. (2010). Noninvasive genome sampling in chimpanzees. Molecular Ecology, 19(24), 5328-5331.

Kolodziej, K., Theissinger, K., Brün, J., Schulz, H. K., & Schulz, R. (2012). Determination of the minimum number of microsatellite markers for individual genotyping in wild boar (*Sus scrofa*) using a test with close relatives. European Journal of Wildlife Research, 58(3), 621-628.

Linacre, A., Gusmao, L., Hecht, et al. (2011). ISFG: recommendations regarding the use of non-human (animal) DNA in forensic genetic investigations. Forensic Science International: Genetics, 5(5), 501-505.

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(S2), 19-23.

McKelvey, K. S., & Schwartz, M. K. (2004). Genetic errors associated with population estimation using non-invasive molecular tagging: problems and new solutions. Journal of Wildlife Management, 68(3), 439-448.

Nechvatal, J. M., Ram, J. L., Basson, M. D., Namprachan, P., Niec, S. R., Badsha, K. Z., ... & Kato, I. (2008). Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. Journal of Microbiological Methods, 72(2), 124-132.

Nori, D. V., & McCord, B. R. (2015). The application of alkaline lysis and pressure cycling technology in the differential extraction of DNA from sperm and epithelial cells recovered from cotton swabs. Analytical and Bioanalytical Chemistry, 1-10.

Norris, D. O., & Bock, J. H. (2000). Use of fecal material to associate a suspect with a crime scene: report of two cases. Journal of Forensic Sciences, 45(1), 184-187.

Okubara, P. A., Schroeder, K. L., Li, C., Schumacher, R. T., & Lawrence, N. P. (2007). Improved extraction of *Rhizoctonia* and *Pythium* DNA from wheat roots and soil samples using pressure cycling technology. Canadian Journal of Plant Pathology, 29(3), 304-310.

Piggott, M. P. (2005). Effect of sample age and season of collection on the reliability of microsatellite genotyping of faecal DNA. Wildlife Research, 31(5), 485-493.

Pompanon, F., Bonin, A., Bellemain, E., & Taberlet, P. (2005). Genotyping errors: causes, consequences and solutions. Nature Reviews Genetics, 6(11), 847-846.

Renan S, Speyer E, Shahar N, Gueta T, Templeton AR, Bar-David S (2012) A factorial design experiment as a pilot study for noninvasive genetic sampling. Molecular Ecology Resources 12(6):1040-1047

Santini, A., Lucchini, V., Fabbri, E., & Randi, E. (2007). Ageing and environmental factors affect PCR success in wolf (*Canis lupus*) excremental DNA samples. Molecular Ecology Notes, 7(6), 955-961.

Smejkal, G. B., Robinson, M. H., Lawrence, N. P., Tao, F., Saravis, C. A., & Schumacher, R. T. (2006). Increased protein yields from *Escherichia coli* using pressure-cycling technology. Journal of Biomolecular Techniques, 17(2), 173.

Tao, F., Lawrence, N. P., Miller, W., et al. (2003). Biological sample preparation system using Pressure Cycling Technology (PCT). In Advances in High Pressure Bioscience and Biotechnology II (pp. 413-417). Springer Berlin Heidelberg.

Van de Goor LHP, Panneman H, Van Haeringen WA (2009) A proposal for standardization in forensic equine DNA typing: allele nomenclature of 17 equine-specific STR loci. Animal Genetics, 41: 122-127.

Waits, L. P., & Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. Journal of Wildlife Management, 69(4), 1419-1433.

Wehausen, J. D., Ramey, R. R., & Epps, C. W. (2004). Experiments in DNA extraction and PCR amplification from bighorn sheep feces: the importance of DNA extraction method. Journal of Heredity, 95(6), 503-509.

Yu, Z., & Morrison, M. (2004). Improved extraction of PCR-quality community DNA from digesta and fecal samples. Biotechniques, 36(5), 808-813.

Yuan, M. Q., Tu, Z., & Li, C. X. (2011). A rapid and effective bone DNA extraction method using pressure cycling technology. Forensic Science and Technology, 4, 012.

CHAPTER 3

3. GENETIC STRUCTURE WITHIN THE WILD HORSE POPULATIONS INHABITING OREGON'S HERD MANAGEMENT AREAS

This manuscript has been submitted to: Journal of Heredity

3.1. Introduction

America's wild horses are descendants from European horses introduced during early explorations and colonization of the Americas (McGahern et al., 2006; Prystupa et al., 2012; Luis et al., 2006b). Because of their iconic value as a reminder of America's West, the Wild Free-Roaming Horses and Burros Act of 1971 and 1978 mandated the Bureau of Land Management (BLM) to aid in the protection, management, and control of wild horses and burro populations on a given rangeland. Considered an integral part of the existing natural ecosystems on the public lands, these wild horses are now protected from indiscriminate capture, branding, harassment, and death. The BLM management goals aim to achieve the Appropriate Management Level (AML) by employing methods of random capture and removal of horses (http://www.blm.gov/or/resources/whb/herd-manage.php). Although these approaches have maintained the desired herd size, they often overlook the close relationships of the active breeding gene pool within herds and may be putting existing sequestered herds at risk of a local population crash. Decreased genetic diversity can subsequently lead to inbreeding, lower fitness and ultimately to herd extinction. It is, therefore, vital that the management strategies for wild horses sequestered on public rangelands consider employing genetic analyses so as to maintain a

healthy, diverse gene pool (Beauclerc et al., 2010) at an optimal carrying capacity for the rangelands.

Within the State of Oregon, 19 Herd Management Areas (HMAs) have been established (Figure 4), all of which are under the management of the USDA-Forest Service (USDA-FS) or the BLM. The herd numbers in Oregon are increasing yearly on average by twenty percent (http://www.blm.gov/or/resources/whb/herd-manage.php).

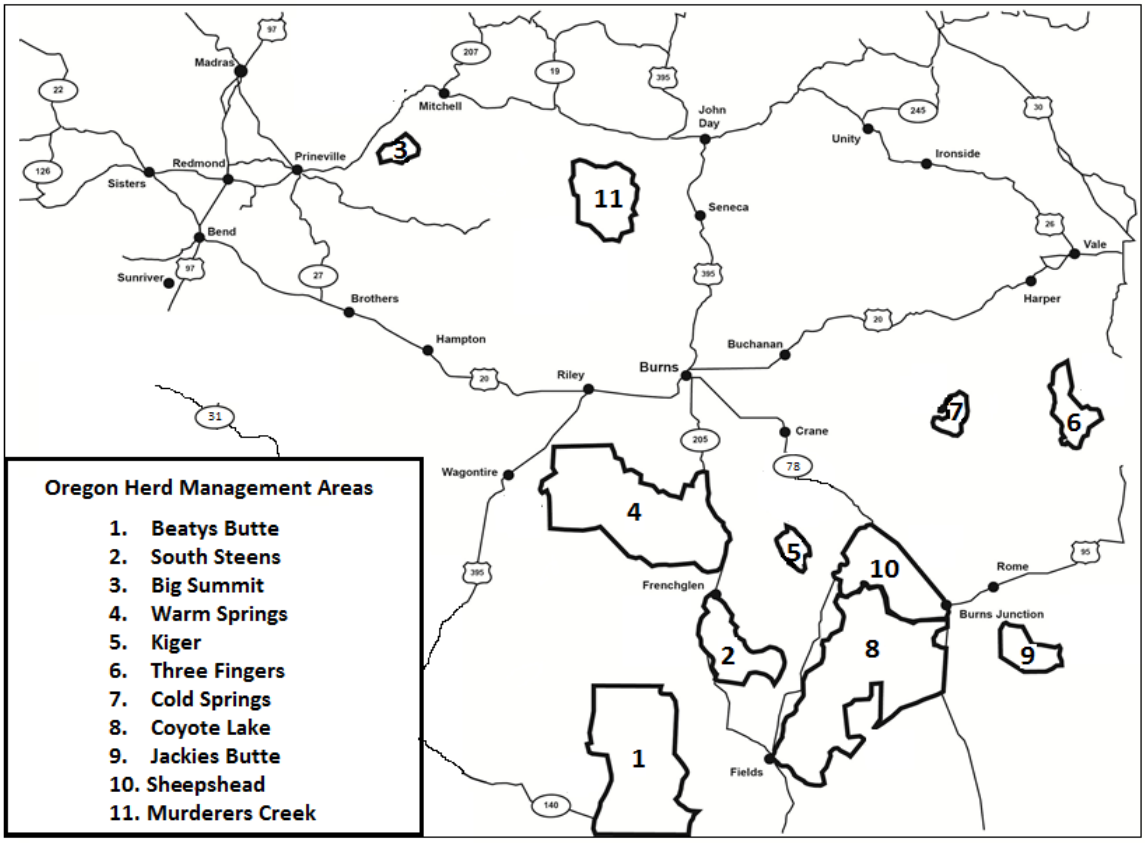

Figure 4. Geographical locations of HMA in Oregon used in this study. Adapted from (http://www.blm.gov/or/resources/whb/files/or_state_map_HMA.pdf).
Sites included in the study were Beatys Butte (1), South Steens (2), Big Summit (3), Warm Springs (4), Kiger (5), Jackies Butte (9) and Murderer's Creek (11). The numbers represent the HMA identifiers in Oregon.

The Big Summit HMA, within the Ochoco National Forest located in central Oregon, was the primary area of study. The rangeland covers approximately 27,300 acres of high desert valleys surrounded by conifer-covered mountains. The elevation ranges from 1200 to 2200 meters and is subject to cold snowy winters and hot summers. The overall forest terrain with steep ravines and cliffs hinders traditional aerial and accurate ground survey for census. Severe winters with few survivors (personal communication, Central Oregon Wild Horse Coalition), the encroachment of human development and fencing off of adjacent lands has further impacted this herd and eliminated any natural migration corridors between neighboring HMAs for several generations.

The 2011 census indicated the herd size to be approximately 55-70 individuals, a herd size significantly lower than the recommended minimum of 150-200 breeding individuals to maintain random mating (Cothran et al., 1998; Luis et al., 2006a) within a natural, free-roaming population. The amalgamation of all these factors have resulted in this HMA's wild horse population to be considered "high–risk" as classified by the Food and Agriculture Organization (FAO 1998; Luis et al., 2006a). The objective of the present study was to provide a genetic analysis using non-invasive (hair) collection methods to assess the genetic health of this small herd compared to horses within other HMAs. These analyses included the determination of genetic structure, distribution of observed genetic variation, phylogenetic analysis of HMA populations and ancestral breed assignment from microsatellite allele and mtDNA data.

3.2. Methodology

3.2.1. Ethics Statement

The protocol was approved under Institutional Animal Care and Use Committee (IACUC #15-001) of Florida International University. No specific permissions were required for sample collection. Hair samples were donated by horse owners and The United States Department of Agriculture-Forest Services (USDA-FS), Prineville, Oregon and mailed to Florida International University.

3.2.2. Sample Collection

A total of 70 individual (mane) hair samples were furnished by owners of captive Big Summit horses (24), other captive HMA horses (Jackies Butte (6), South Steens (7), Warm Springs (5), Beatys Butte (4), Kiger (6) and Murderers Creek (9)) as well as hair that was left behind on the bark of pine trees within the Ochoco Forest. These "Unknown" samples (9) were representative of the current 'wild' gene pool in Big Summit. Since samples were collected over a period of time, the mtDNA analysis does not include the unknown (9) samples. The published genotypic data of 19 domestic equine breeds (Lusitano (43), Koniks (50), Irish Cob (50), Welsh (50), Dartmore (22), Shire (29), Tennesse Walker (23), Andalusian (50), Fell (50), Connemara (40), Standardbred (50), Appaloosa (50), Fjord (50), Hackney (50), Thoroughbred (50), Icelandic (50), Dutchdraft (50), Haflinger (50), Kaspian (17)) documented in Van de Goor et al. 2010 were used for analysis.

3.2.3. DNA Extraction

Genomic DNA was extracted from hair samples using the hair protocol provided by QIAamp® DNA Mini Kit (Qiagen, Germantown, MD). Five to ten hair strands were used for samples originating from known horses. One hair root was used for unknown sources (those obtained from trees and fences) to insure the DNA isolated was from a single horse. DNA was quantified using the Qubit® dsDNA HS assay kit on the Qubit® 2.0 Fluorometer (Life Technologies, Grand Island, NY) and the average DNA yield was 5-20 ng depending on the number and quality of roots extracted.

3.2.4. Microsatellite amplification and fragment analysis

The DNA extracts were amplified using StockMarks® for Horse Equine Genotyping Kit (Applied Biosystems, Foster City, CA) per manufacturer's protocol and STRs were separated using an ABI Prism 310 capillary Genetic Analyzer (Applied Biosystems, Foster City, CA) and analyzed by GeneMapper® Software Version 4.0 (Applied Biosystems, Foster City, CA). Allele sizes were assigned a letter score following the International Society of Animal Genetics guidelines (ISAG) (Gill et al. 1994, 1997; Bär et al. 1997; Schneider et al. 1998) and were later converted to allele repeat numbers based on the updated equine typing standards (Van de Goor et al. 2009). The ISAG guideline allows the comparison of genotypes across various laboratories and eliminates binning errors. Domestic horses that had been DNA typed at an external laboratory according to the American Quarter Horse Association (AQHA) registration requirements were used as positive controls and these external DNA typed profiles as

well as the internal DNA standard within the kit were used to verify all allele calls for this study.

3.2.5. mtDNA amplification and sequencing

The DNA was amplified by polymerase chain reaction (PCR) using the mtDNA D-loop HVR1 region, specific equine primers (Forward: 5'-CTA GCT CCA CCA TCA ACA CC-3' and Reverse: 5'-ATG GCC CTG AAG AAA GAA CC-3') which amplified a 410 bp region. Each 20 µl PCR reaction contained: 10-50 ng DNA, 0.5 µM each primer, 2.5 mM $MgCl_2$, 0.2 mM dNTPs and 0.6 U GoTaq® Hot Start Polymerase (Promega, Madison, WI). The PCR cycles consisted of a denaturation step of 95 °C for 2 min, followed by 30 cycles of 95 °C for 1 min, 60 °C for 45 sec, and 72 °C for 1 min, with a final extension of 72 °C for 5 min on the C1000 Touch™ Thermal Cycler (Bio-Rad, Hercules, CA). The PCR products were purified using ExoSAP-IT® (USB, Cleveland, OH) and then sequenced using the Big Dye® Terminator version 3.1 sequencing kit (Applied Biosystems, Carlsbad, CA). Products were purified with 75% Isopropanol and dried down at 80 °C for 1 min. Hi-Di™ formamide (Applied Biosystems, Carlsbad, CA) (12.5 µL) was added and the products were subsequently denatured at 95 °C for 2 min, loaded onto an ABI Prism® 310 Genetic Analyzer (Applied Biosystems, Carlsbad, CA), separated and analyzed. DNA sequences were aligned using the software ClustalW (European Bioinformatics Institute, Cambridge, UK).

3.2.6. Statistical Analysis

Microsatellite data - Statistical analysis was carried out on 14 loci (Canon 2000; Luis et al. 2006b; Monies et al. 2010) that were present in all samples. Allele frequencies were calculated and analyzed using GeneAlEx (Peakall and Smouse 2006, 2012) and Genepop (Raymond and Rousset 1995). Frequencies were verified and deviations from Hardy Weinberg Equilibrium (HWE) were calculated using GeneAlEx and Hardy Weinberg (HW) check (Kalinowski 2006a). Individual relationships within the herd were estimated through ML-Relate (Kalinowski et al., 2006b), predicting the degrees of relatedness as parent/offspring (PO), full sibling (FS), half siblings (HS) or unrelated (U). Kinship likelihoods were calculated at 95% confidence interval (Kalinowski et al., 2006b; Wagner et al., 2006). Additionally, fixation indices (F_{IS}, F_{IT}, F_{ST}) were calculated using GenAlEx and a coordination plot was plotted using F_{ST} genetic distances. F_{IT} is the inbreeding coefficient of an individual relative to the total population. F_{IS} is the inbreeding coefficient of an individual relative to the subpopulation, and F_{ST} is the effect of subpopulations compared to the total population.

mtDNA sequences - The genetic variability among the sequences was measured through haplotype and nucleotide diversity using the program Arlequin version 3.5.2.2 (Excoffier and Lischer 2010) (http://cmpg.unibe.ch/software/arlequin35/). The probability of identity was calculated as the sum of the square of the frequency of each haplotype in that population ($PI=\Sigma q_k^2$ where q_k is the frequency of the k^{th} haplotype) (McGahern et al., 2006; Hill et al. 2002; Álvarez et al., 2012). In order to investigate population differentiation among the different HMAs studied, the population pairwise F_{ST}

values were computed under the Analysis of Molecular Variance (AMOVA) also using Arlequin version 3.5.2.2 (Excoffier and Lischer 2010). The statistical significance of the values was estimated through 1000 permutations.

Phylogenetic Analysis

Microsatellite data: Phylogenetic trees were generated using Nei's (1983) distance measure (D_a) (Nei et al., 1983; Takezaki and Nei 1996; Luis et al., 2006a) that were calculated using allele frequencies through POPTREE2 software's unrooted neighbor-joining tree

mtDNA sequences: The sequences were truncated to 388 bp, between the positions of 15459 and 15846 np, in accordance to the published nucleotide position numbering that follows Xu and Árnason (1994) (GenBank X79547), the mtDNA reference sequence of the horse. A maximum parsimony tree was constructed based on the comparisons of different sequence haplotypes that were identified, using the software, Network 4.611 (Fluxus Technology Ltd, Clare, UK) (Bandelt et al., 1999) (http://www.fluxus-engineering.com/sharenet.htm).

Population Substructure

Model-based clustering analysis using STRUCTURE 2.3.4 (Pritchard et al., 2000) was used to determine the admixture from ancestral breeds of the wild horses in the HMAs, assign Unknown hair samples to the most probable HMA, and assess the population structure of the HMAs. Systematic sampling (sampling interval=3) was used to normalize the individual domestic breed dataset, reducing population size to 50 or

fewer animals per breed to prevent a skewed clustering towards overrepresented breeds. An admixture model with correlated allele frequencies was adopted without prior delineation of populations to account for recent gene flow between breeds that share major alleles. The parameter for individual admixture alpha was set to be the same for all simulations with uniform prior probability distribution. Distinct populations (K) were estimated using 150,000 burn-ins, 350,000 Markov Chain Monte Carlo repetitions, and three independent runs for each K (1-30) (Evanno et al., 2005). The optimal K was then calculated using an *ad hoc* quantity (ΔK) and estimated as K=7 for 19 domestic breeds and seven HMAs horses.

3.3. Results

The unknown samples were verified (on the basis of collection site, F_{ST} genetic distance) to be from Big Summit and further microsatellite analysis assigned a total of 33 individuals as being part of the Big Summit population.

Polymorphism

Microsatellites: A total of 109 alleles were detected across 14 microsatellite markers in the Big Summit herd. The mean number of alleles ranged from 6-10 alleles per locus. Statistical results for H_{Obs}/H_{Exp}, and deviation from HWE for Big Summit and other HMA populations show reduced heterozygosity in the Big Summit populations (Tables 6, 7). Allele frequencies for the Big Summit HMA (Table 8) demonstrated a broad range of alleles for all 14 loci compared to the domestic equine breeds (Van de Goor et al., 2010). The HMA populations had a greater number of private alleles than

domestic populations with the isolated Big Summit HMA harboring the majority of the private alleles (Table 9). The Big Summit population expected heterozygosity was higher than the observed heterozygosity for all the markers tested. Except for AHT4, VHL20 and AHT5, 11 other markers were found to have significantly deviated from HWE ($P < 0.001$). The inbreeding coefficient F_{IS} for each marker in the Big Summit herd ranged between 0.08 for HMS2 to 0.46 for ASB23 (Table 10). The overall Big Summit inbreeding coefficient was 0.26 and demonstrated higher homozygosity compared to other HMA populations.

Table 6. Statistical results and indices for the Big Summit HMA based on 14 microsatellite loci.

Locus	Alleles	Ne	H_{Obs}	H_{Exp}	HWE *P* value	Significance	F_{IS}
AHT4	8	4	0.64	0.76	0.52	NS	0.16
HMS7	10	3	0.55	0.70	0.00	*P*<0.001	0.22
HTG4	6	3	0.39	0.70	0.00	*P* <0.001	0.44
VHL20	6	5	0.55	0.78	0.10	NS	0.30
AHT5	9	4	0.67	0.75	0.19	NS	0.11
HMS6	8	3	0.42	0.71	0.00	*P* <0.001	0.40
HMS3	10	5	0.58	0.81	0.00	*P* <0.001	0.29
HTG10	10	4	0.58	0.77	0.00	*P* <0.001	0.25
ASB17	9	6	0.64	0.84	0.00	*P* <0.001	0.24
CA425	7	3	0.52	0.67	0.00	*P* <0.001	0.23
ASB23	8	5	0.42	0.79	0.00	*P* <0.001	0.46
HMS2	6	4	0.67	0.72	0.00	*P* <0.001	0.08
ASB2	6	4	0.67	0.78	0.03	*P* <0.05	0.14
HTG7	6	3	0.45	0.63	0.00	*P* <0.001	0.28
Total	109		0.55	0.74			0.26

N= Number of individuals; HWE = Hardy Weinberg Equilibrium; F_{IS} = Fixation index (inbreeding coefficient); H_{Obs} = Observed Heterozygosity; H_{Exp} = Expected Heterozygosity

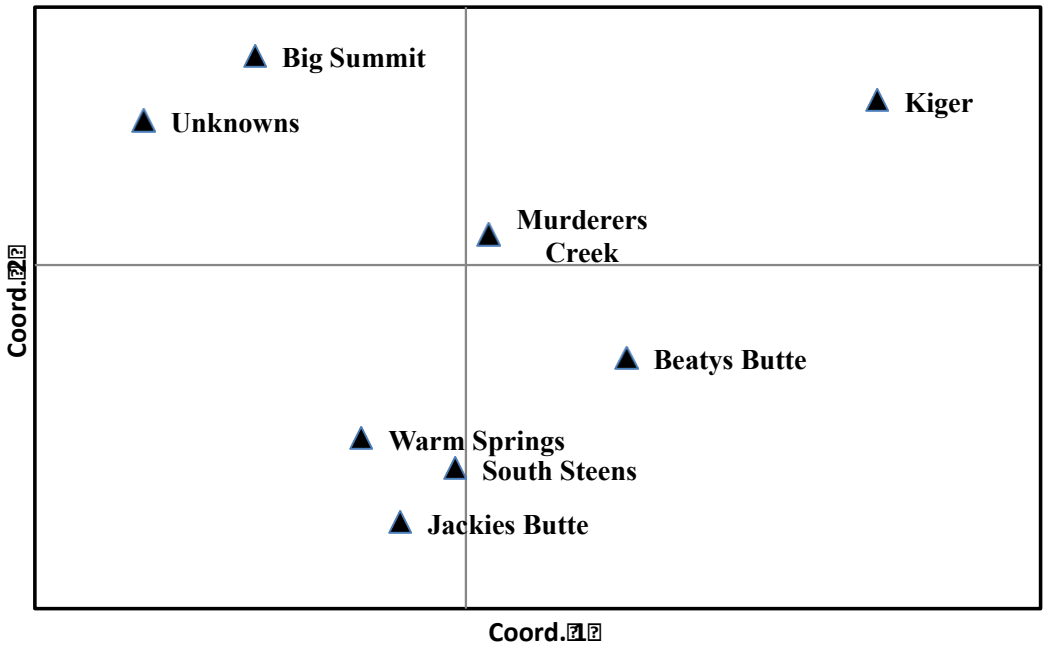

Figure 5. PCoA ordination plot based on F_{ST} genetic distance. The plot shows clustering of neighboring HMA compared to Big Summit and the unknown samples. The geographically distant Big Summit HMA plots closer to the adjacent Murderer's Creek HMA but away from the cluster of six geographically closer HMAs.

Table 7. Mean statistical results and indices for seven HMA populations based on 14 microsatellite loci. A high F_{IS} is representative of inbreeding.

Population	N (Individuals)	H_{Obs}	H_{Exp}	F_{IS}
Big Summit	33	0.55	0.74	0.26
Jackies Butte	6	0.69	0.75	0.07
Murderer's Creek	9	0.69	0.73	0.04
Beatys Butte	4	0.63	0.68	0.08
Warm Springs	5	0.64	0.69	0.06
South Steens	7	0.77	0.73	-0.04
Kiger	6	0.55	0.61	0.09

H_{Obs} = Observed Heterozygosity; H_{Exp} = Expected Heterozygosity, F_{IS} = Inbreeding Coefficient

Table 8. Allele frequencies observed in Big Summit HMA (33 individuals) for 14 microsatellite loci.

Alleles	AHT 4	HMS 7	HTG 4	VHL 20	AHT 5	HMS 6	HMS 3	HTG 10	ASB 17	CA 425	ASB 23	HMS 2	ASB 2	HTG 7
N	8	10	6	6	9	8	10	10	9	7	8	6	6	6
12	-	-	-	-	-	0.0	-	-	-	-	-	-	-	-
13	-	-	-	-	-	0.4	-	-	-	-	-	-	-	-
14	-	-	-	-	-	0.1	-	-	0.0	-	-	-	-	-
15	-	0.0	-	0.1	-	0.1	-	-	-	-	-	0.3	-	0.1
16	-	0.0	-	0.2	0.0	0.1	-	-	-	0.1	-	0.2	0.2	0.1
17	-	0.0	-	0.3	0.0	0.0	-	-	-	-	-	0.0	-	-
18	-	0.3	-	-	0.3	0.0	-	0.0	-	0.0	0.2	-	-	0.1
19	-	0.0	-	-	0.0	-	-	0.0	-	0.2	0.1	0.2	0.0	0.5
20	-	0.4	-	0.1	0.2	0.0	0.0	0.1	0.2	0.5	0.3	0.0	0.2	-
21	-	0.0	-	0.1	0.1	-	-	0.0	0.1	-	0.0	-	-	0.0
22	-	0.0	-	-	0.0	-	0.0	-	0.2	-	-	-	-	-
23	-	0.0	-	-	-	-	0.0	0.0	0.0	-	-	-	-	-
24	0.0	0.0	-	0.0	0.0	-	0.0	-	0.0	-	0.0	-	0.2	-
25	0.0	-	-	-	0.0	-	0.2	0.1	0.0	-	-	0.0	0.1	-
26	0.3	-	-	-	-	-	0.0	0.4	0.1	0.0	-	-	-	-
27	0.3	-	-	-	-	-	0.2	0.0	-	0.0	-	-	0.0	-
28	0.0	-	-	-	-	-	0.2	0.0	0.0	-	-	-	-	-
29	0.1	-	0.0	-	-	-	0.0	0.0	-	0.0	0.0	-	-	-
30	0.0	-	0.1	-	-	-	0.1	-	-	-	0.0	-	-	-
31	-	-	0.3	-	-	-	-	-	-	-	0.0	-	-	-
32	-	-	0.4	-	-	-	-	-	-	-	-	-	-	-
33	0.0	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	-	0.0	-	-	-	-	-	-	-	-	-	-	-
35	-	-	0.1	-	-	-	-	-	-	-	-	-	-	-

Table 9. Private allele list for Big Summit HMA based on 14 microsatellite loci.

Locus	Allele	Frequency
AHT4	30	0.05
HMS7	15	0.02
	24	0.02
AHT5	22	0.03
	25	0.02
HMS6	13	0.48
HMS3	22	0.03
	23	0.02
HTG10	25	0.12
	28	0.06
	29	0.02
ASB17	28	0.06
CA425	26	0.03
	29	0.02
ASB2	24	0.02
	31	0.05
HMS2	25	0.03
ASB2	16	0.24
	19	0.02
HTG7	21	0.03

mtDNA sequencing: Within all the HMA horses in the study, 36 variable nucleotide substitutions defining 16 different haplotypes (Appendix 1) were identified. Of the 36 nucleotide substitutions, 34 sites were transitions and two were single base deletions. Each of the 16 haplotypes identified were assigned to seven of the 18 major haplogroups (A-R) (Achilli et al., 2012).

Population Differentiation

The values for F_{IT}, F_{ST}, and F_{IS} which denote the inbreeding coefficient (I) relative to the total (T) population, the effect of subpopulations (S) compared to T, and I relative to S, respectively are shown in Table 5. Considering how F_{ST} is calculated, positive values indicate some degree of variation in the allele frequencies across subpopulations. The F_{ST} values (Table 11) exhibit moderate genetic variation in the allele frequencies across subdivisions. The variability at the group level between the seven populations assessed by population pairwise F_{ST} values is summarized in Table 6. The results indicated that there was statistically significant genetic differentiation between Big Summit and geographically close Murderer's Creek HMA (0.237). Big Summit showed the highest values when compared with Beatys Butte (0.609) and South Steens (0.550), indicating an even lower degree of gene flow between these HMA herds. All values were statistically significant ($P < 0.05$).

Table 10. Statistical results of F-Statistics over Big Summit and six HMA populations for each locus. A negative F_{IS} value indicates excess of heterozygosity while a positive value indicates a trend towards loss of heterozygosity. The F_{ST} values closer to 0 indicate little genetic differentiation, while values 0.25 to 1 indicate greater genetic differentiation.

Locus	F_{IT}	F_{IS}	F_{ST}
AHT4	0.074	-0.050	0.118
HMS7	0.031	-0.080	0.103
HTG4	0.243	0.167	0.092
VHL20	0.153	0.009	0.146
AHT5	0.070	-0.036	0.102
HMS6	0.333	0.203	0.163
HMS3	0.202	0.138	0.075
HTG10	0.108	-0.019	0.124
ASB17	0.403	0.306	0.140
CA425	0.342	0.224	0.153
ASB23	0.278	0.195	0.103
HMS2	0.308	0.041	0.279
ASB2	0.283	0.127	0.179
HTG7	0.077	-0.033	0.106

F_{IS} = (Mean H_{Exp} - Mean H_{Obs}) / Mean H_{Exp}; F_{IT} = (H_t - Mean H_{Obs}) / Ht; F_{ST} = (H_t - Mean H_{Obs}) / H_t; Mean H_{Exp} = Average H_e across the populations; Mean H_o = Average H_o across the populations; H_t = Total expected heterozygosity

Table 11. Population pairwise F_{ST} values (below diagonal) and the P value significance of F_{ST} (above diagonal) between seven populations for the mtDNA. A high value is indicative of a lesser degree of gene flow with complete differentiation among the populations. A "+" indicates $P < 0.05$ and a NS indicates there was no statistical significance.

	Big Summit	Kiger	Murderer's creek	Warm Spring	Jackies Butte	Beatys Butte	South Steens
Big Summit	0	+	+	+	+	+	+
Kiger	0.424	0	+	NS	NS	+	NS
Murderer's creek	0.237	0.190	0	NS	+	+	+
Warm Spring	0.341	-0.018	0.028	0	NS	NS	NS
Jackies Butte	0.463	0.208	0.151	-0.062	0	NS	NS
Beatys Butte	0.609	0.325	0.288	0.176	0.231	0	NS
South Steens	0.550	0.150	0.264	0.124	0.231	0.097	0

Genetic diversity and relationships

Genetic relationship data using ML-Relate for the captive Ochoco herd members (n= 33) reported four parent-offspring relations, 14 full siblings and 49 half siblings. The mtDNA genetic diversity estimates, including haplotype diversity (*h*), nucleotide diversity (π), and the probability index, found within the seven populations studied (Table 12) indicated that Warm Springs and Beatys Butte had the most diverse populations (*h*=1.00 ± 0.13 and *h*=1.00 ± 0.18, respectively) even with only a 4-5 horses sampled. These HMAs exhibited one haplotype per horse sampled in each population. Big Summit horses had the most conserved mtDNA haplotypes (*h*=0.51 ± 0.04) with 24 horses sampled. The analysis of the nucleotide diversity revealed that Warm Springs and Murderer's Creek exhibited the most variation (π= 0.026 ± 0.017 and π= 0.024 ± 0.014). On the other hand, Beatys Butte and South Steens (π=0.008 ± 0.006 and π= 0.009 ± 0.006,) demonstrated very limited nucleotide diversity, keeping in mind only 4-7 horses from these HMAs were sampled. The estimated PI of the haplotypes identified ranged from 0.20 to 0.51 for the HMAs in Oregon, and was 0.51 for the Big Summit HMA.

Table 12. Genetic diversity indices of mtDNA lineages.

HMA Sub-populations	# Individuals	# Haplotypes	Haplotype diversity (*h*) (±SE)	Nucleotide diversity (π) (±SE)	Probability of Identity (PI)
Big Summit	24	2	0.51 ± 0.04	0.012 ± 0.007	0.51
Warm Springs	5	5	1.00 ± 0.13	0.026 ± 0.017	0.20
Jackies Butte	6	3	0.73 ± 0.16	0.017 ± 0.011	0.39
Murderers Creek	9	6	0.89 ± 0.09	0.024 ± 0.014	0.21
Beatys Butte	4	4	1.00 ± 0.18	0.008 ± 0.006	0.25
Kiger	6	2	0.60 ± 0.13	0.019 ± 0.012	0.50
South Steens	7	3	0.67 ± 0.16	0.009 ± 0.006	0.43

h= haplotype diversity, π = nucleotide diversity, ±SE = standard error and PI= probability of identity.

Phylogenetic Analysis

The phylogenetic trees were constructed using 14 loci on Nei (1983) genetic distances and an unrooted Neighbor Joining (NJ) to compare the seven HMAs in Oregon to 19 domestic breeds (Nei et al., 1983). The phylogenetic tree indicated five distinct clusters of related breeds. The cold-blooded breeds Dutch Draft and Haflinger grouped together, the Warmblood breeds like the Thoroughbreds grouped closer to the Spanish breeds, while the Koniks, Fjord, Appaloosa, Tennessee Walker and Icelandic formed a separate branch. Although the HMA horses branched separately from the domestic breeds, the Big Summit HMA formed a separate cluster from the six HMA and domestic breeds. However, the HMA horses seem to be closer to Spanish breeds than to other domestic breeds.

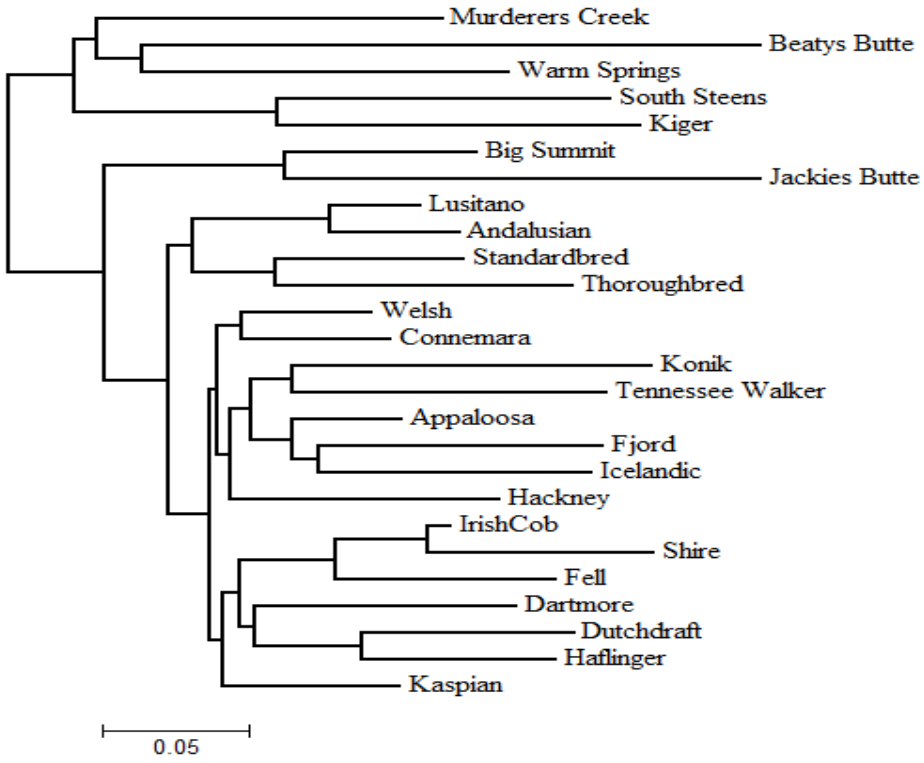

Figure 6. Phylogenetic tree of HMA and domestic equine breeds. NJ dendrogram constructed on POPTREE2 from Nei's (1983) D_a distances showing genetic relationships across seven HMAs in relation to 19 known equine breeds.

A median-joining network tree for mtDNA was constructed from the most parsimonious tree of all the sampled horses according to the identified haplotypes and significant clusters (Figure 7). There was a clear differentiation of all the haplogroups that were assigned, evidenced by their apparent clustering. Haplogroup I was exclusively composed of horses belonging to the Big Summit. Almost all other haplogroups had no clear geographic affiliation and were composed of a mixture of horses from different HMAs, with the exception of haplogroup Q1 and Q3 (Figure 7).

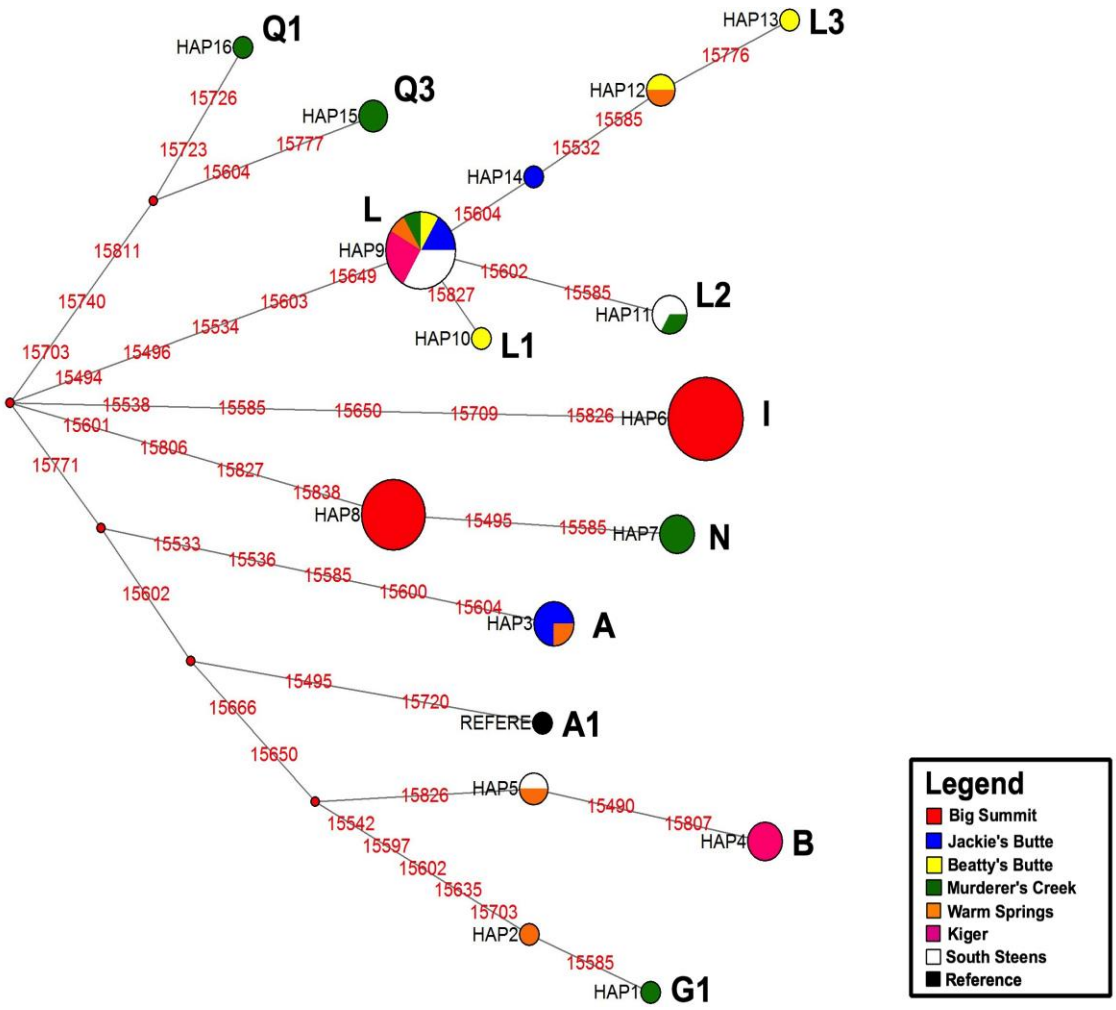

Figure 7. Median joining tree of wild horse mtDNA sequences. Haplogroups and haplotypes are designated accordingly. Circles are proportional to the number of horses they represent. Pie slices within circles indicate the HMA from which the horses came.

Population Substructure

Cluster analysis performed on the seven HMAs (Figure 8) showed distinct groupings that supports the phylogenetic tree (Figure 6). Dissimilar populations are divided by a black line and are numbered using the representative domestic breeds and HMAs (Figure 8). Each individual is a column divided into K colors, each color representing a breed cluster. In concurrence with the phylogenetic tree, Icelandic, Fjord and Koniks group together; Fell, Irish Cob and Shire as one cluster; the Spanish breeds, Andalusian and Lusitano, group together and the HMA populations group distinctively from the domestic breeds. Within the HMAs, two clusters are seen in the Big Summit HMA and the other six HMAs as subpopulations. The HMAs as a whole showed significant ancestral contributions from over five domestic breeds as observed with the presence of various colors (Figure 8). The major color (light blue), corresponding to Andalusian and Lusitano breed, was seen in the Big Summit, followed by orange, dark blue and others corresponding to Standardbred, Thoroughbred, Appaloosa, Icelandic and other breeds. The pink color in the group of six HMAs shows similarity to the Konik horses.

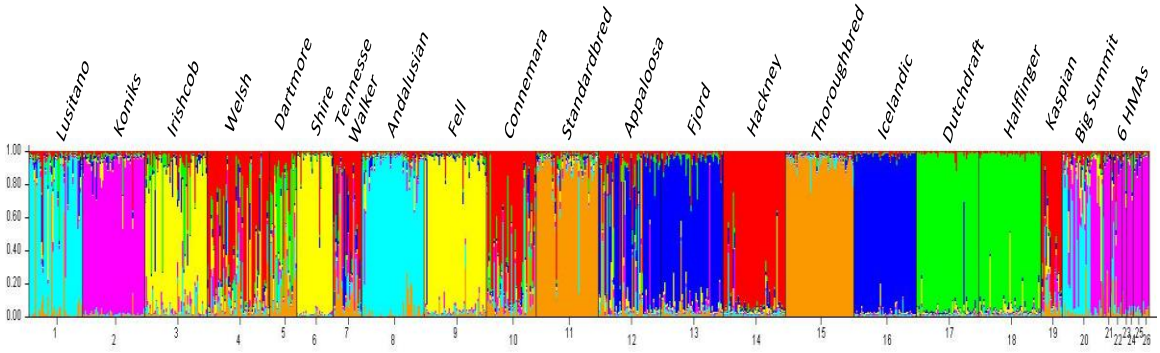

Figure 8. Cluster analysis for HMAs and 19 domestic equine breeds. Clustering for seven Oregon HMAs and 19 domestic breeds with K = 7 inferred clusters estimated by STRUCTURE 2.3.4 (Pritchard et al. 2000). A contribution of Spanish breeds (Andalusian and Lusitano domestic breeds) was seen in Big Summit as well as Koniks. Similarity to the Konik semi-feral European (Poland) horses, Standardbred and Thoroughbred were seen in other HMAs.

3.4. Discussion

The horses inhabiting the Ochoco National Forest are considered feral but managed as "wildlife". They are no doubt recent descendants of domestic horses that helped settle the Western USA and either escaped captivity or were emancipated by their owners during the settling of America. However, these horses are unlike most domestic breeds which are now managed and bred under heavily managed animal husbandry practices, with many breeds specifically line bred for certain characteristics. Previous genetic analysis (Cothran, 2011) of 12 horses from the Big Summit HMA using 12 microsatellites indicated inbreeding within the herd as well as a strong ancestral lineage to the Iberian-derived Spanish Andalusian breed, the North American gaited breeds, and Arabian breeds (www.blm.gov/or/districts/burns/wildhorse/). In that study, it was also concluded that the Big Summit horses were closely aligned with the Iberian breeds. The results of the current study also showed a strong ancestral contribution to the Big Summit

horses from the Iberian breeds, specifically the Andalusian and Lusitano breeds, supporting the Cothran (2011) study. Although contributions from other modern breeds were seen across all the Oregon HMAs, it was interesting to see a strong correlation to the semi-feral, Polish Konik breed within the six other HMA populations—a population of horses not associated with the settling of the West. This diverse admixture association may be a better indicator of the lack of human intervention and the ability to randomly mate than it is of the Konik contribution via bloodlines. In other words, if a population has a more diverse breeding pool plus an opportunity for mate selection via random mating that could be reflected in the ancestral diversity seen in some HMA populations. Therefore, the contribution from diverse ancestral breeds and the lack of human intervention or selective breeding practices, one would expect the Oregon wild horses to be more ancestrally diverse compared to the domestic breeds.

However, reproductive isolation, reduced herd sizes, geographical and physical barriers to migration corridors, and human settlement has sequestered the Big Summit herd on an "island" of natural land. The Big Summit small population appears to have been trapped in a fragmented habitat unable to migrate and breed outside of their local gene pool for several generations, mimicking the inbreeding seen in island populations. While the mating behavior of free-roaming horses should prevent them from breeding to close relatives, within a small herd horses have a higher probability of encountering close relatives and thus, increase the intra-family relationships. The Big Summit HMA analysis in this study displayed a large number of full and half siblings, indicative of close relatedness coupled with an apparent lack of recent gene flow between neighboring

HMAs. However, keeping in mind the small sample sizes, and because of their hierarchal social structure and serendipitous capture methods, these results may be reflective of existing family groups being captured together. Nevertheless, the overall fixation index of 0.26 and deviation from Hardy-Weinberg equilibrium at most of the DNA markers supported that inbreeding is rising from non-random mating and limited mate selection in the Big Summit HMA. Population pairwise F_{ST} values based on mtDNA data provided additional evidence that the Big Summit horses are experiencing restricted gene flow and show little contributions from other HMA herds. Similar findings have been seen in studies on Spanish Celtic horse breeds (Canon 2000) Portuguese horse breeds (Luis et al. 2006a) and Iberian breeds (Luis et al., 2006a) where distinct genetic differentiation and partition of the genetic variability and structure was observed within breeds (Kavar 2008). Conversely, genetic data cannot fully resolve geographic dispersal or whether or not physical migration corridors between herds prior to the initiation of the Act of 1971 and human encroachment were present. It is crucial to understand that analysis cannot definitively resolve whether the shared alleles are common by descent versus common by location. A number of private alleles were found in the Big Summit HMA although this may be due to a larger sample size from the Big Summit population compared to other HMA populations. However, a similar pattern has been seen in feral horses isolated and protected on the Sable Island, Canada where a large number of breed specific alleles were noted (Plante et al., 2007).

Evidence is mounting in conservation genetics where small insular populations, analogous to some of the Oregon HMAs, are greatly impacted by inbreeding and should

be candidates for genetic restoration if managed quickly (Nei et al., 1975; Ingvarsson 2001; Tallmon et al., 2004). Research shows bottlenecks, founder effects and genetic drift increase the risk of a decline in genetic variation and result in a decrease in heterozygosity at numerous loci. A cumulative loss of genetic variation can lead to the expression of deleterious alleles, which may become fixed in small populations such as the Big Summit herd (Nei et al., 1975; Westermeier et al., 1998). The loss, in turn, decreases fitness and increases the probability of extinction of small isolated populations. Inbreeding seems to be already impacting these horses as they appear more susceptible to diseases, have more physical deformities and the herd suffers from high infant morbidity and mortality—all phenomena that have been noticed in this herd (personal communication, USDA-FS personnel).

Other studies have previously demonstrated that destruction of habitat, sequestration and fragmentation of populations subsequently led to a bottleneck resulting in inbreeding and limited gene diversity (Frantzen et al., 2001; Ouborg et al., 2010). In the Losina breed, the loss of heterozygosity and increased homozygosity displayed a trend towards being out of HWE. They concluded the significant deviation from HWE for three loci (HTG10, HMS3 & AHT4 ($P < 0.01$)) could be a direct consequence of a small breeding population and differentiation of the breed over time (Canon et al. 2000). Other studies have implied that a substantial loss in genetic diversity reduces fitness, limits gene flow into the populations, and in turn, alters the adapting capacity of a population (Bryant et al., 1999; Beauclerc et al., 2010). The Big Summit herd could echo the Sorraia and Przewalski's horse populations, which were subjected to intense

bottlenecks, and displayed severe losses in the mean number of alleles and in the level of heterozygosity.

Genetic diversity has been shown to be an essential part of any functioning ecosystem (Croteau et al., 2012). The thorough understanding of what influences patterns of genetic diversity and population structure are essential and crucial in populations of wide-ranging species, such as the wild horses (Croteau et al., 2012). Providing genetic data on these Oregon herds will allow for informed management decisions where inbreeding can be minimized and genetic diversity maximized. Integrating scientific analyses with current management strategies is pivotal for saving these small herds of wild horses. The ultimate goal for current and future management, whether on a large-scale or small-scale should always place its focus on maintaining the health of the herds, gene flow, and the highest form of genetic diversity.

3.5. Acknowledgements

The authors thank the USDA-FS Prineville, Oregon and their personnel for collecting hair samples from the Big Summit range. Funding for this project was supported by the USDA-Forest Service: Challenge Cost-share agreement #NFS 09-CS-11060701-008 to DM. We are grateful to the members of the Central Oregon Wild Horse Coalition and especially Gayle Hunt for providing hair samples from the wild horses representing HMAs in Oregon. A special thanks to International Forensic Research Institute at Florida International University for the use of the Forensic DNA Profiling Facility. The work represents the point of view of the authors and may not be the official view of the USDA-FS.

3.6. References

Achilli A, Olivieri A, Soares P, Lancioni H, Kashani BH, Perego UA, Nergadze SG, Carossa V, Santagostino M, Capomaccio S, Felicetti M, Al-Achkar W, Penedo MCT, Verini-Supplizi A, Houshmand M, Woodward SR, Semino O, Silvestrelli M, Giulotto E, Pereira L, Bandelt H, Torroni A (2012) Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. Proceedings of the National Academy of Sciences 109:2449-2454

Álvarez I, Fernández I, Lorenzo L, Payeras L, Cuervo M, Goyache F (2012) Founder and present maternal diversity in two endangered Spanish horse breeds assessed via pedigree and mitochondrial DNA information. Journal of Animal Breeding Genetics 129:271-279

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Molecular Biology and Evolution 16(1), 37-48

Bär W, Brinkman B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr W, Olaisen B (1997) DNA recommendations - further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. Forensic Science International, 87(3), 179-4

Beauclerc, K. B., Johnson, B., & White, B. N. (2010). Genetic rescue of an inbred captive population of the critically endangered Puerto Rican crested toad (*Peltophryne lemur*) by mixing lineages. Conservation Genetics, 11(1), 21-32.

Bryant EH, Backus VL, Clark ME, Reed DH (1999) Experimental tests of captive breeding for endangered species. Conservation Biology, 13(6) 1487-1496

Canon J, Checa ML, Carleos C, Vega-Pla JL, Vallejo M, Dunner S (2000) The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data. Animal Genetics, 31(1), 39-48.

Cothran EG, Santos SA, Mazza MCM, Lear TL, Sereno JRB (1998) Genetics of the Pantaneiro horse of the Pantanal region of Brazil. Genetics and Molecular Biology, 21(3): 343-9.

Cothran EG (2011) Report to USDA-FS - Genetic Analysis of the Big Summit HMA, OR April 8, 2011

Croteau EK, Heist EJ, Nielsen CK, Hutchinson JR, Hellgren EC (2012) Microsatellites and mitochondrial DNA reveal regional population structure in bobcats (*Lynx rufus*) of North America. Conservation Genetics, 13(6), 1637-1651.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology, 14(8), 2611-2620.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources, 10(3), 564-567

FAO (1998) Secondary Guidelines for Development of National Farm Animal Genetic Resources Management Plans. Management of Small Populations at Risk. Initiative for Domestic Animal Diversity. FAO, Rome.

Frantzen, M. A. J., Ferguson, J. W. H., & De Villiers, M. S. (2001). The conservation role of captive African wild dogs (*Lycaon pictus*). Biological Conservation, 100(2), 253-260.

Gill P, Kimpton C, dAloja E, Andersen JF, Bär W, Brinkmann B, Holgersson S, Johnsson V, Kloosterman AD, Lareu MV, Nelleman L, Pfitzinger H, Phillips CP, Schmitter H, Schneider PM, Stenersen M (1994) Report of the European DNA profiling group (EDNAP) – towards standardisation of short tandem repeat (STR) loci. Forensic Science International, 65(1) 51-59.

"Herd Management Areas Oregon/Washington BLM." U.S. Department of the Interior Bureau of Land Management. http://www.blm.gov/or/resources/whb/herd-manage.php. Accessed 12 March 2014

Hill EW, Bradley DG, Al-Barody M, Ertugrul O, Splan RK, Zakharov I, Cunningham EP (2002) History and integrity of Thoroughbred dam lines revealed in equine mtDNA variation. Animal Genetics, 33(4), 287-294

Ingvarsson PK (2001) Restoration of genetic variation lost – the genetic rescue hypothesis. Trends Ecology Evolution, 16(2) 62–63

Kalinowski ST (2006a) HW-QUICKCHECK: an easy to use computer program for checking genotypes for agreement with Hardy-Weinberg expectations. Molecular Ecology, 6(4), 974-979

Kalinowski ST, Wagner AP, Taper ML (2006b) ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. Molecular Ecology Notes, 6(2) 576-579.

Kavar T and Dovc P (2008) Domestication of the horse: Genetic relationships between domestic and wild horses. Livestock Science, 116(1-3) 1-4.

Luis C, Juras R, Oom MM, Cothran EG (2006a) Genetic diversity and relationships of Portuguese and other horse breeds based on protein and microsatellite loci variation. Animal Genetics, 38(1), 20-27.

Luis C, Bastos-Silveira C, Cothran EG, Oom MM (2006b) Iberian origins of New World horse breeds. Journal of Heredity, 97(2), 107-113.

McGahern AM, Edwards CJ, Bower MA, Heffernan A, Park SDE, Brophy PO, Bradley DG, MacHugh DE, Hill EW (2006) Mitochondrial DNA sequence diversity in extant Irish horse populations and in ancient horses. Animal Genetics 37(5), 498-502

Monies D, Abu Al Saud N, Sahar N, Meyer BF (2010) Population studies and parentage testing for Arabian horses using 15 microsatellite markers. Animal Genetics, 42(2), 225-226

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Evolution, 29(1), 1-10.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. Journal of Molecular Evolution, 19(2), 153-70.

"Oregon BLM wild horse and burro population data." U.S. Department of the Interior Bureau of Land Management. http://www.blm.gov/or/resources/whb/files/2011foalcrop.pdf. Accessed 20 March 2013.

Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R. K., & Hedrick, P. W. (2010). Conservation genetics in transition to conservation genomics. Trends in Genetics, 26(4), 177-187

Peakall, R. and Smouse P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Resources. 6(1), 288-295.

Peakall, R. and Smouse P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28, 2537-2539.

Plante Y, Vega-Pla JL, Lucas Z, Colling D, De March B, Buchanan F (2007) Genetic diversity in a feral horse population from Sable Island, Canada. Journal of Heredity, 98(6), 594–602.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.

Prystupa JM, Hind P, Cothran EG, Plante Y (2012) Maternal lineages in native Canadian equine populations and their relationship to the Nordic and Mountain and Moorland pony breeds. Journal of Heredity, 103(3), 380-390

Raymond M, & Rousset F (1995) GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. Journal of Heredity, 86(3), 248-249.

Schneider HR, Rand S, Schmitter H, Weichhold G (1998) ACTBP2-nomenclature recommendations of GEDNAP. International Journal of Legal Medicine, 111(2) 97-100.

Takezaki N and Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics, 144(1) 389-399.

Tallmon DA, Luikart G, Waples RS (2004) The alluring simplicity and complex reality of genetic rescue. Trends in Ecology & Evolution, 19(9), 489–496.

Van de Goor LHP, Panneman H, Van Haeringen WA (2009) A proposal for standardization in forensic equine DNA typing: allele nomenclature of 17 equine-specific STR loci. Animal Genetics, 41(2), 122-127.

Van De Goor LHP, Van Haeringen WA, Lenstra JA (2010) Population studies of 17 equine STR for forensic and phylogenetic analysis. Animal Genetics, 42(6), 627-633.

Wagner AP, Creel S, Kalinowski ST (2006) Estimating relatedness and relationships using microsatellite loci with null alleles. Heredity, 97, 336–345

Westemeier, R. L., Brawn, J. D., Simpson, S. A., Esker, T. L., Jansen, R. W., Walk, J. W., ... & Paige, K. N. (1998). Tracking the long-term decline and recovery of an isolated population. Science, 282(5394), 1695-1698.

CHAPTER 4

4. GENETIC VARIATION OF MAJOR HISTOCOMPATIBILITY COMPLEX MICROSATELLITES: A COMPARISON OF DOMESTIC, KATHIAWARI AND FERAL HERD MANAGEMENT AREA HORSES

4.1. Introduction

Population genetic studies of wild animals often focus on neutral markers such as mitochondrial D-loop, microsatellites or single nucleotide polymorphism rather than the adaptive, genetic variation in functional genes. Neutral markers are used to quantify genetic diversity (Aberle, et al., 2004; Valera, et al., 2005), track dispersal patterns (Pope, 1992), measure inbreeding (Luis, et al., 2007; Pusey & Wolf, 1996), determine relationships (Tozaki et al., 2001) as well as understand population structure and history (Zechner et al., 2002). However, neutral markers do not provide information on adaptive changes that affect genes under selection, such as the highly variable major histocompatibility complex (MHC).

The MHC is an extremely polymorphic multi-gene family, primarily associated with the immune defense system. These genes encode for cell surface molecules which directly and indirectly play a critical role in immunological activity and more recently have been associated with producing unique odor types in vertebrates (Edwards & Hedrick, 1998; Knapp, Robson, & Waterhouse, 2006). The vertebrate MHC genes are classified into three classes (I, II and III) that are linked in a cluster on the same chromosome (Trowsdale, 1995). Class I is typically described as classical and non-classical genes. The polymorphic, classical MHC molecules present epitopes to the T-

cell receptors of CD8 T-lymphocytes while the non-classical molecules present peptide antigens but exhibit limited polymorphism. Class II molecules present antigens to CD4 lymphocytes while Class III molecules include secreted proteins associated with diverse immune and non-immune functions.

The maintenance of genetic diversity at markers under selection is one of the central goals in conservation biology because it is believed to contribute to both short and long-term survivability. Pathogen-mediated selection (Parham, Adams, & Arnett, 1995), negative assortative mating, heterozygote advantage, frequency-dependent selection and maternal-fetal interaction have been proposed as few of the mechanisms influencing balancing selection (Edwards & Hedrick, 1998; Hedrick, 1998). The patterns of variation within MHC genes are consistent with balancing selection where multiple alleles are maintained in a population in order to increase individual fitness and viability. This variation subsequently influences nucleotide polymorphisms, allele frequencies, and linkage disequilibrium at MHC loci (Hedrick & Thomson, 1983). Although there is debate regarding the relative importance of each mechanism, the role of MHC in pathogen resistance appears more likely since MHC molecules are actively involved in immune response (Brown & Eklund, 1994). An inherent problem in identifying the forces that drive and maintain MHC diversity is that population structure and selective mechanisms are unknown in most wild populations. Given the long history of investigation on MHC structure, function, and selective mechanisms, this gene complex is an extremely valuable candidate for studying adaptation in natural populations (Hedrick, 1994).

4.1.1. Equine Lymphocyte Antigen (ELA) genes

The equine MHC, referred to as the equine lymphocyte antigen (ELA) complex (Bailey, Marti, Fraser, Antczak, & Lazary, 2000; Marti et al., 1996), is localized on chromosome ECA20q14q22 (Ansari et al., 1988; Makinen et al., 1989). Comparative analyses of the equine MHC region have demonstrated significant conservation of gene content (Madden, 1995) and order (Gustafson et al., 2003), relative to other mammalian species. However, the genomic organization of ELA appears most similar to humans (Albright, et al., 1991) where the MHC exists as a single contiguous segment and lacks disruptions as is seen in ruminants, pigs, and cats (Kelley et al., 2005). The most striking observation of the ELA is the presence of two homologues of the Class II DQA locus distributed on two different chromosomes (5 and 20), a phenomenon which has not been observed in any other mammalian species (Fraser & Bailey, 1998). Furthermore, the Class II DRA locus has been shown to exhibit greater allelic diversity than in any other taxon (Albright-Fraser et al., 1996; Brown et al., 2004; Luís et al., 2005). The DRA and DQA equine MHC genes encode the α-chain of Class II molecules and are predicted to have antigen-presenting functions that are homologous to human leucocyte antigen genes (Kamath & Getz, 2011).

The diversity in equine MHC genes has previously been examined in only few domestic or captive equids (Albright-Fraser et al., 1996; Brown et al., 2004). The identification of microsatellites within the MHC in different horse populations provides the opportunity to perform comparative analyses and decipher gene diversity between populations. This study examines the diversity of the American horses that include (i)

domestic breeds, (ii) two different herds of feral horses managed within geographically distinct herd management areas (HMA) in Oregon and (iii) a geographically distant Kathiawari breed from India. When the horse was domesticated approximately 6000 years ago, humans began taking horses from the wild, and managing them under conditions that were convenient to humans (i.e., application of animal husbandry principles) (Goodwin, 2007). Most domesticated breeds share ancestral Arabian bloodlines thought to be one of the foundation lineages in the development of nearly every modern horse breed (Jansen et al., 2002; Vila et al., 2001). Horses have been selectively bred for centuries, initially for specific traits and then as pedigrees (stud books) were maintained, it led to specific breed designations as those seen in the USA domestic horses (Warmuth et al., 2012). The American domestic horses included in this study are Thoroughbreds, Paint, Appaloosa, and Quarter Horses. The feral Oregon horses are thought to be recent descendants of previously domesticated horses that escaped captivity or were emancipated by their owners (Ryden, 2005) during the settling of the western USA. Previous studies have shown that many of the wild horses are descended from Spanish breeds (Chapter 3). Unlike most domestic breeds, which are now bred under artificial selection conditions, the feral horses, for the most part, have been able to, somewhat randomly, select their mates over the last 100-200 years. The Kathiawari horse breed is found exclusively in the Indian subcontinent. However, owing to indiscriminate breeding and lack of sound breeding policies, the breed characteristics are being diluted by outbreeding and only few thousand true Kathiawari horses are now in existence (Kaura, 1961). The breed is historically presumed to be a mixture between the Arabian horse and other desert breeds with some influence from the Mongolian horse

(Gupta et al., 2014). High genetic variability at neutral microsatellites in Kathiawari horses has demonstrated high heterozygosity within the present day breed (Koringa, et al., 2008).

The current study examined genetic diversity using two geographically distant horse cohorts that include the Indian Kathiawari breed, and American breeds (domestic and feral). The study tested the hypothesis that there should be no difference at the MHC loci between geographically distinct breeds as they are the same species. Previous research on ELA has focused on identifying MHC haplotypes and genetic polymorphisms in the classical Class I and Class II genes using genome-wide association studies (McCue et al., 2012; Mittmann et al., 2010).

4.2. Methodology

4.2.1. Sample Collection

Hair samples were collected from 155 horses belonging to 47 domestic horses (Thoroughbred, Quarter Horses, Paint, Appaloosa), 35 Kathiawari and 73 HMA horses (36 from the Big Summit HMA and an additional 37 from a cluster of neighboring Herd Management Areas). Hair samples of feral horses were furnished by owners of captive horses or furnished by Forest Service personnel from hair found on trees or fences during census surveys.

4.2.3. DNA Extraction

Genomic DNA was extracted from hair samples using the hair protocol provided

by QIAamp® DNA Mini Kit (Qiagen, Germantown, MD). Five to twelve hair strands with bulbs were used for samples originating from known horses. One hair root was used for unknown sources (those obtained from trees and fences) to insure the DNA isolated was from a single horse. DNA was quantified using the Qubit® dsDNA HS assay kit on the Qubit® 2.0 Fluorometer (Life Technologies, Grand Island, NY). The average DNA yield was 5-20 ng depending on the number and quality of roots extracted.

4.2.4 Microsatellite amplification and fragment analysis

DNA was amplified using six MHC microsatellites (three from Class I and four from Class II) depicted in Fig 9. These included TAMU_305_93, ABGe_9019 belonging to Class I and COR113, COR112, COR114, UM011 belonging to Class II (Brinkmeyer-Langford, Cai, Gill, & Skow, 2013; Tseng, Miller, Cassano, Bailey, & Antczak, 2010). Fragment analysis was achieved using capillary electrophoresis on an ABI Prism 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA) and GeneMapper ® Software Version 4.0 (Applied Biosystems, Foster City, CA).

4.2.5 Statistical Analysis

Microsatellite data – Allele frequencies, number of alleles, observed (H_{Obs}) and expected heterozygosity (H_{Exp}), were calculated using GenAlex (Peakall & Smouse, 2006; Peakall & Smouse, 2012). Fixation indices (*F*) were also computed using the same software. GenAlex was used to evaluate pairwise genetic differentiation between populations, and departure from Hardy-Weinberg equilibrium, using Chi-square and sequential Bonferonni correction on loci. Analysis of molecular variance (AMOVA) was

performed to determine the amount of genetic variation attributable to within and between populations. To construct evolutionary relationships, an Unweighted Pair Group Method with Arithmetic Mean (UPGMA) tree was constructed with program POPTREE2 (Takezaki, Nei, & Tamura, 2010) based on F_{ST} genetic distance measures from allele frequency and 1000 bootstrap iterations.

4.3. Results

4.3.1. Identification and characterization of MHC I and II alleles

Seventy-six alleles were found over all populations and all markers. The number of alleles per locus ranged from 11 (TAMU_305_93, and ABGe_9019) to 15 (COR 114). Some rare alleles were found with a high frequency in the Appaloosa and Quarter Horse breeds within the American domestic population. The number of alleles, number of effective alleles = $[\frac{1}{(sum\ pi^2)}]$ (where pi is the frequency of the ith allele at a locus), observed and expected heterozygosity are shown in Table 13. Some significant heterozygote deficiencies after Bonferonni correction were found, for different loci and populations (Table 14). However, the mean heterozygosity deficit was observed in the feral and Big Summit populations (Table 15).

4.3.2. Population relationships and clustering

The overall F_{ST} values based on the six microsatellites ranged from 0 indicating no differentiation or the same population to 0.359 suggesting a larger genetic differentiation between populations. The domestic horses showed little differentiation

when compared to the feral HMA horses and a greater genetic distance of 0.325 with the Kathiawari breed (Table 16). Using the F_{ST} values, the American horse populations cluster together, but cluster away from the Kathiawari breed. Although the American horses group away from the Kathiawari breed, the two HMA populations do not group close to one another. They are equidistant from the domestic horse population (Figure 10).

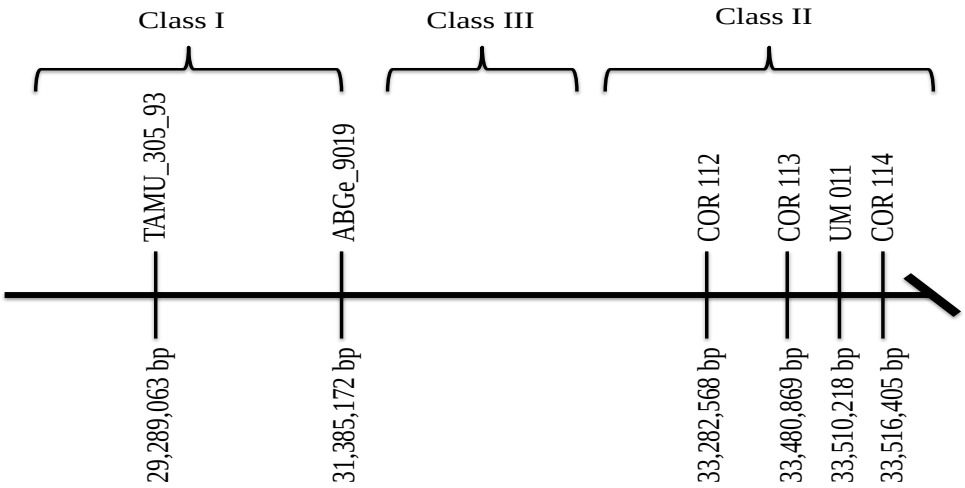

Figure 9. Schematic representation of six of Major Histocompatibility Complex microsatellite loci locations on ECA20. Positions of loci are indicated in base pairs (bp).

Table 13. Statistical results and indices across six MHC microsatellite loci for each population

Population	# of individuals	N_a	N_e	H_{Obs}	H_{Exp}	*F*
Domestic	47	11.17	6	0.85	0.83	-0.03
Kathiawari	35	9.33	6	0.82	0.83	0.01
Big Summit HMA	36	8.00	5	0.75	0.78	0.04
Other HMAs	37	9.33	6	0.73	0.84	0.13

Na= Number of Alleles, Ne= Number of Effective Alleles H_{Obs}= Observed Heterozygosity, H_{Exp}= Expected Heterozygosity, and *F*=Fixation Index

Table 14. Statistical results and indices based on six MHC microsatellite loci.

		TAMU_ 305_93	ABGe_ 9019	COR113	COR112	COR114	UM011
Domestic (N=47)	N_a	9	11	11	12	13	11
	N_e	5	7	8	5	7	4
	H_o	0.83	0.96	0.77	0.91	0.87	0.77
	H_e	0.80	0.86	0.88	0.81	0.85	0.76
	F	-0.03	-0.11	0.13	-0.13	-0.03	-0.01
	HWE Sig	NS	NS	**	NS	***	NS
Kathiawari (N=35)	N_a	9	11	9	9	8	10
	N_e	5.51	6.19	5.57	5.47	5.29	7.54
	H_o	0.74	0.86	0.63	0.97	0.83	0.91
	H_e	0.82	0.84	0.82	0.82	0.81	0.87
	F	0.09	-0.02	0.23	-0.19	-0.02	-0.05
	HWE Sig	NS	NS	***	NS	NS	*
Big Summit HMA (N=36)	N_a	6	9	11	8	7	7
	N_e	3.01	5.27	6.93	4.87	4.20	5.04
	H_o	0.69	0.92	0.75	0.67	0.69	0.75
	H_e	0.67	0.81	0.86	0.80	0.75	0.80
	F	-0.04	-0.13	0.12	0.16	0.08	0.06
	HWE Sig	NS	**	**	**	NS	NS
Other HMAs (N=37)	N_a	9	10	9	9	10	9
	N_e	5.86	7.26	4.84	6.61	6.73	6.49
	H_o	0.70	0.89	0.41	0.84	0.78	0.78
	H_e	0.83	0.86	0.79	0.85	0.85	0.85
	F	0.15	-0.03	0.49	0.01	0.08	0.07
	HWE Sig	NS	NS	***	NS	NS	NS

N= Sample Size, Na= Number of Alleles, Ne= Number of Effective Alleles, H_{Obs}= Observed Heterozygosity, H_{Exp}= Expected Heterozygosity, and F=Fixation Index, HWE Sig= Hardy Weinberg Equilibrium Significance, NS= not significant, * P<0.05, ** P<0.01, *** P<0.001

Table 15. Private alleles for each locus by population and breed within populations.

Population	Locus	Breed	Allele	Frequency
Domestic	COR 112	Appaloosa	220	0.032
	COR 112	Quarter Horse	244	0.021
	COR 114	Appaloosa	213	0.011
	COR 114	Quarter Horse	215	0.011
	COR 114	Appaloosa	217	0.011
	UM011	Quarter Horse	164	0.043
	UM011	Quarter Horse	166	0.011
	UM011	Appaloosa	192	0.011
Kathiawari	UM011	-	172	0.186
Big Summit HMA	TAMU_305_93	-	229	0.014
	COR 113	-	215	0.014
Other HMA	COR 112	-	222	0.068
	COR 114	-	239	0.027
	COR 114	-	241	0.054

Table 16. Pairwise population F_{ST} values. F_{ST} value is a measure of genetic distance between the populations where 0 indicates no differentiation (same population). The greater the value of F_{ST}, greater the differentiation between the two populations

	Domestic	Kathiawari	Big Summit HMA	Other HMAs
Domestic	0.000	--	--	--
Kathiawari	0.325	0.000	--	--
Big Summit HMA	0.219	0.359	0.000	--
Other HMA	0.219	0.284	0.258	0.000

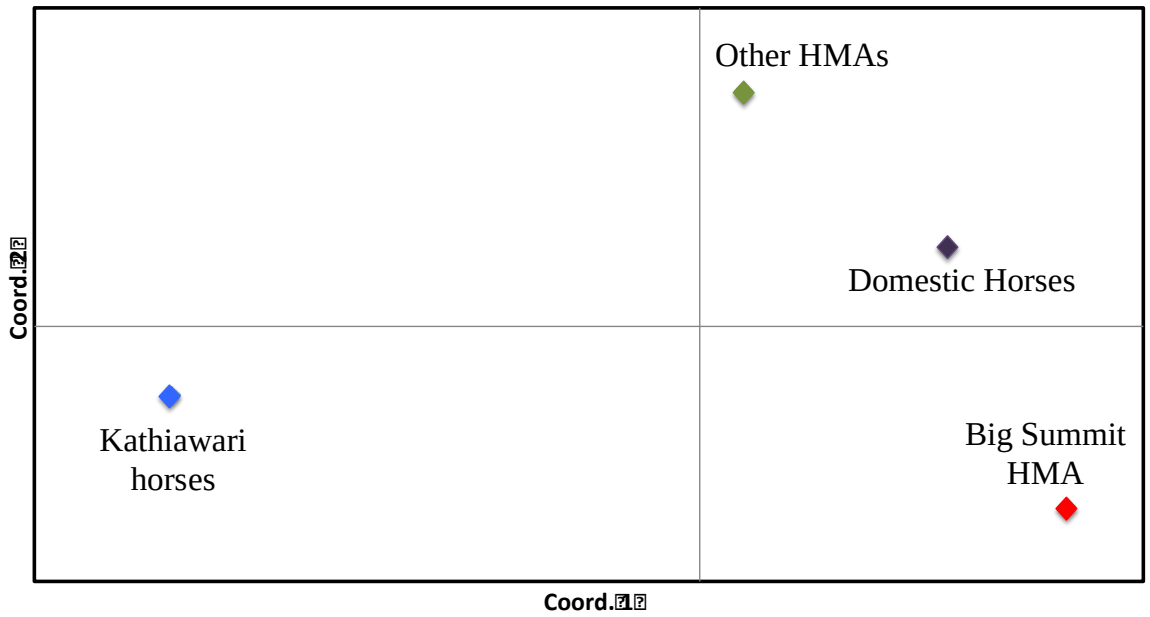

Figure 10. Principal Coordinate Analysis using pairwise F_{ST} genetic distance. PCoA shows the cluster of American horses, feral or domestic, distant from Kathiawari horses of the Indian subcontinent.

4.4. Discussion

Several studies have been carried out to identify the different ELA haplotypes among horses where serological as well parentage information was known (Brinkmeyer-Langford et al., 2013; Tseng et al., 2010). However, no studies have looked at the genetic diversity at MHC genes of different horse populations without prior knowledge of parentage or blood samples for serological ELA haplotypes. Comparing genetic variation for MHC loci in the same species separated by local and continental geographic distance provided an opportunity to understand the neutral and selective forces that may influence MHC variation.

In domesticated animals it is predicted that MHC diversity would be lower because the genes would be impacted by anthropogenic selection for certain traits.

Selection of breed specific traits can result in the reduction in allelic richness for the entire genome, including the MHC genes, and could have drastic consequences for survival in natural, less managed environment (Wynne et al., 2007). Additionally, natural selection for MHC genes associated with parasite-driven or immune-driven selection may not occur because most domesticated horses are vaccinated and de-wormed if managed under good animal husbandry practices (Neff et al., 2008). Therefore, there remains a concern that deliberate selection for breed desirable traits by selective breeding of domestic horses could completely override natural selection and lead to a loss of MHC diversity.

In the current study, relatively high allelic diversity was observed in the selectively bred domestic breeds. High allelic diversity was influenced by the private alleles that were observed exclusively in Quarter Horse and Appaloosa horses included in the study. The data support, in spite of artificial selection pressures, genetic variation is being maintained at the MHC genes. This is also seen in the feral horses and even in the small populations of isolated feral horses (Big Summit). A number of studies have found that genetic drift may overpower selection to determine MHC variation or that selection may directly influence population structure (Oliver & Piertney, 2012). Higher heterozygosity than expected was observed for most MHC loci in the horse populations. Conversely, the results of the study on Big Summit feral horses using mtDNA and neutral markers showed only two mtDNA haplotypes and loss of allelic richness compared to other feral horse populations. Studies on the Kathiawari breed based upon neutral microsatellites have shown higher genetic variability (Koringa et al., 2008) and

significant genetic differentiation from the English Thoroughbred horses and other populations. Additionally, evidence suggests that the Kathiawari breed has gone through a bottleneck event sometime in the past (Chauhan et al., 2011). In this study the detected heterozygosity in the Kathiawari horses was close to the expected heterozygosity. Similar to the findings of this study, other island populations also show high levels of variation at the MHC relative to background levels of variation at neutral loci (Aguilar et al., 2004; Seddon & Baverstock, 1999).

It is believed that events such as bottlenecks and inbreeding are expected to reduce genetic diversity (Slatkin, 1987). Other researchers argue that these conditions can also influence the population to create greater diversity through mutations and genetic mechanisms having a larger affect in a smaller population (Varvio et al., 1986). It is suggested that under balancing selection, the level of heterozygosity may be related to the population size, mutation rate of loci in question, and the selection pressures (Maruyama & Nei, 1981). Consequently, comparing the low heterozygosity at neutral markers and high heterozygosity at MHC loci in feral horses suggests that balancing selection at the MHC must have been intense in the feral horses (Chapter 3). These findings are similar to studies on isolated, inbred populations with small population sizes that have shown an excess of heterozygotes for MHC loci. Studies indicate that population bottlenecks and isolation have a larger influence on patterns of MHC variation than selection (Boyce et al., 1997). Similar increases in heterozygosity have been observed in San Nicolas Island foxes that maintain high levels of variation, given the small effective population size (Aguilar et al., 2004). For the feral horses, the observed

values of heterozygosity at MHC loci are similar to those seen in larger populations such as the island foxes where intense balancing selection was responsible for maintaining genetic diversity at the MHC loci. The findings of this study indicate that diversity within Kathiawari breed and HMA populations has been generated or persists despite the relatively small population sizes, management strategies and bottleneck events.

Furthermore examination of genetic and local geographic distances between the two HMA populations clearly delineated similarities and differences in the allelic patterns of MHC variability. Although the neutral markers showed low genetic variation, the Big Summit feral horses were distant from the other feral horses based on their F_{ST} values (Chapter 3). The smallest genetic distances for the six MHC loci were between American domestic population and the feral horse populations of Oregon (0.219) (Table 4, Figure 10). This low genetic differentiation can be attributed to the recent historical relationship between feral and modern domestic breeds. Moreover allelic sharing has been observed in studies on other domestic horse breeds such as Thoroughbred, Standardbred, and other common domestic breeds. Although private alleles were observed for these breeds, a large degree of overlap in the MHC microsatellite fragment lengths for most loci were found to be common amongst the same breeds (Brinkmeyer-Langford et al., 2013; Meyer et al., 1997; Tseng et al., 2010). Despite the allele sharing, diversity at MHC was apparent when haplotypes are compared among breeds showing variation occurs due to different arrangements of the common alleles. Allele sharing has been observed in the great crested newt from three geographically distant populations inhabiting postglacial expansion areas. However, it was suggested that these alleles could

have been maintained by balancing selection for several thousands of years since the expansion (Babik et al., 2008). Another reason for degree of overlap and allele sharing could be that horses have recently been domesticated compared to sheep or cattle that have been domesticated around 10000 ±12000 years ago (Buchanan et al., 1994; MacHugh et al., 1998). The observed similar allelic patterns across horse population as well low levels of demarcations of genetic distances within populations that are geographically separated could be due to this recent domestication of horses (Bjørnstad & Røed, 2001; Canon et al., 2000).

Comparing genetic variation for MHC loci in the same species provided an opportunity to evaluate the factors that may influence differentiation of populations. Most studies show that some degree of balancing selection is maintaining the variation at MHC genes irrespective of population size. If balancing selection is acting upon the horse populations in this study, the effect can be observed by the maintenance of allelic variability and subsequent heterozygosity in populations. However, MHC diversity must be compared to distribution of neutral markers so as to determine whether neutral demographic processes affect all loci or selection is shaping only functional genes (Boyce et al., 1997). It would be apt to look at both neutral markers and MHC genes to better understand the genetic diversity in populations.

4.5. References

Aberle, K. S., Hamann, H., Drögemüller, 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(4), 270-277.

Aguilar, A., Roemer, G., Debenham, S., Binns, M., Garcelon, D., & Wayne, R. K. (2004). High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. Proceedings of the National Academy of Sciences of the United States of America, 101(10), 3490-3494.

Albright, D., Bailey, E., & Woodward, J. G. (1991). Nucleotide sequence of a cDNA clone of the horse (*Equus caballus*) DRA gene. Immunogenetics, 34(2), 136-138.

Albright-Fraser, D. G., Reid, R., Gerber, V., & Bailey, E. (1996). Polymorphism of DRA among equids. Immunogenetics, 43(5), 315-317.

Ansari, H. A., Hediger, R., Fries, R., & Stranzinger, G. (1988). Chromosomal localization of the major histocompatibility complex of the horse (ELA) by *in situ* hybridization. Immunogenetics, 28(5), 362-364.

Babik, W., Pabijan, M., & Radwan, J. (2008). Contrasting patterns of variation in MHC loci in the alpine newt. Molecular Ecology, 17(10), 2339-2355.

Bailey, E., Marti, E., Fraser, D., Antczak, D., & Lazary, S. (2000). Immunogenetics of the horse. The Genetics of the Horse, 123-155.

Bjørnstad, G., & Røed, K. (2001). Breed demarcation and potential for breed allocation of horses assessed by microsatellite markers. Animal Genetics, 32(2), 59-65.

Boyce, W. M., Hedrick, P. W., Muggli-Cockett, N. E., Kalinowski, S., Penedo, M. C., & Ramey, R. R., 2nd. (1997). Genetic variation of major histocompatibility complex and microsatellite loci: A comparison in bighorn sheep. Genetics, 145(2), 421-433.

Brinkmeyer-Langford, C. L., Cai, J. J., Gill, C. A., & Skow, L. C. (2013). Microsatellite variation in the equine MHC. Animal Genetics, 44(3), 267-275.

Brown, J. L., & Eklund, A. (1994). Kin recognition and the major histocompatibility complex: An integrative review. American Naturalist,143(3), 435-461.

Brown, J., Thomson, W., Clegg, P., Eyre, S., Kennedy, L., Matthews, J., . . . Ollier, W. (2004). Polymorphisms of the equine major histocompatibility complex class II DRA locus. Tissue Antigens, 64(2), 173-179.

Buchanan, F. C., Adams, L. J., Littlejohn, R. P., Maddox, J. F., & Crawford, A. M. (1994). Determination of evolutionary relationships among sheep breeds using microsatellites. Genomics, 22(2), 397-403.

Canon, J., Checa, M., Carleos, C., Vega-Pla, J., Vallejo, M., & Dunner, S. (2000). The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data. Animal Genetics, 31(1), 39-48.

Chauhan, M., Gupta, A. K., & Dhillon, S. (2011). Genetic diversity and population structure of three Indian horse breeds. Molecular Biology Reports, 38(5), 3505-3511.

Edwards, S. V., & Hedrick, P. W. (1998). Evolution and ecology of MHC molecules: From genomics to sexual selection. Trends in Ecology & Evolution, 13(8), 305-311.

Fraser, D. G., & Bailey, E. (1998). Polymorphism and multiple loci for the horse DQA gene. Immunogenetics, 47(6), 487-490.

Goodwin, D. (2007). Horse behaviour: Evolution, domestication and feralisation. The welfare of horses (pp. 1-18) Springer.

Gupta, A., Chauhan, M., Bhardwaj, A., Gupta, N., Gupta, S., Pal, Y., . . . Vijh, R. (2014). Comparative genetic diversity analysis among six Indian breeds and English thoroughbred horses. Livestock Science, 163, 1-11.

Gustafson, A. L., Tallmadge, R. L., Ramlachan, N., Miller, D., Bird, H., Antczak, D. F., . . . Skow, L. C. (2003). An ordered BAC contig map of the equine major histocompatibility complex. Cytogenetic and Genome Research, 102(1-4), 189-195.

Hedrick, P. W. (1994). Evolutionary genetics of the major histocompatibility complex. American Naturalist, 146(6), 945-964.

Hedrick, P. W. (1998). Balancing selection and MHC. Genetica, 104(3), 207-214.

Hedrick, P. W., & Thomson, G. (1983). Evidence for balancing selection at HLA. Genetics, 104(3), 449-456.

Jansen, T., Forster, P., Levine, M. A., Oelke, H., Hurles, M., Renfrew, C., . . . Olek, K. (2002). Mitochondrial DNA and the origins of the domestic horse. Proceedings of

the National Academy of Sciences of the United States of America, 99(16), 10905-10910.

Kamath, P. L., & Getz, W. M. (2011). Adaptive molecular evolution of the major histocompatibility complex genes, DRA and DQA, in the genus *Equus*. BMC Evolutionary Biology, 11, 128-2148-11-128.

Kaura, R. (1961). Chapter 3 Indian breeds of livestock (including Pakistan breeds). Breeds of horses and donkeys. (pp. 48-57). Golaganj, India. Lucknow: Prem Publisher.

Kelley, J., Walter, L., & Trowsdale, J. (2005). Comparative genomics of major histocompatibility complexes. Immunogenetics, 56(10), 683-695.

Klein, J. (1986). Natural history of the major histocompatibility complex Wiley.

Knapp, L. A., Robson, J., & Waterhouse, J. S. (2006). Olfactory signals and the MHC: A review and a case study in *Lemur catta*. American Journal of Primatology, 68(6), 568-584.

Koringa, P., Joshi, C., Solanki, J., & Rank, D. (2008). Genetic characterization and bottleneck studies in Kathiawari horse breed of India. Haryana Veterinarian, 47, 77-83.

Luís, C., Cothran, E., Oom, M., & Bailey, E. (2005). Major histocompatibility complex locus DRA polymorphism in the endangered Sorraia horse and related breeds. Journal of Animal Breeding and Genetics, 122(1), 69-72.

Luis, C., Cothran, E. G., & Oom Mdo, M. (2007). Inbreeding and genetic structure in the endangered Sorraia horse breed: Implications for its conservation and management. The Journal of Heredity, 98(3), 232-237.

MacHugh, D., Loftus, R., Cunningham, P., & Bradley, D. (1998). Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. Animal Genetics, 29(5), 333-340.

Madden, D. R. (1995). The three-dimensional structure of peptide-MHC complexes. Annual Review of Immunology, 13(1), 587-622.

Makinen, A., Chowdhary, B., Mahdy, E., Andersson, L., & Gustavsson, I. (1989). Localization of the equine major histocompatibility complex (ELA) to chromosome 20 by *in situ* hybridization. Hereditas, 110(1), 93-96.

Marti, E., Szalai, G., Antczak, D. F., Bailey, E., Gerber, H., & Lazary, S. (1996). The equine major histocompatibility complex (pp. 245-267) CRC Press, Inc., Boca Raton, Florida.

Maruyama, T., & Nei, M. (1981). Genetic variability maintained by mutation and overdominant selection in finite populations. Genetics, 98(2), 441-459.

McCue, M. E., Bannasch, D. L., Petersen, J. L., Gurr, J., Bailey, E., Binns, M. M., . . . Hill, E. W. (2012). A high density SNP array for the domestic horse and extant Perissodactyla: Utility for association mapping, genetic diversity, and phylogeny studies. PLoS Genetics, 8(1), e1002451.

Meyer, A. H., Valberg, S. J., Hillers, K. R., Schweitzer, J. K., & Mickelson, J. R. (1997). Sixteen new polymorphic equine microsatellites. Animal Genetics, 28(1), 69-70.

Mittmann, E., Wrede, J., Pook, J., & Distl, O. (2010). Identification of 21,781 equine microsatellites on the horse genome assembly 2.0. Animal Genetics, 41(2), 222.

Neff, B., Garner, S., Heath, J., & Heath, D. (2008). The MHC and non-random mating in a captive population of Chinook salmon. Heredity, 101(2), 175-185.

Oliver, M. K., & Piertney, S. B. (2012). Selection maintains MHC diversity through a natural population bottleneck. Molecular Biology and Evolution, 29(7), 1713-1720.

Parham, P., Adams, E. J., & Arnett, K. L. (1995). The origins of HLA − A, B, C polymorphism. Immunological Reviews, 143(1), 141-180.

Peakall, R., & Smouse, P. E. (2006). GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes, 6(1), 288-295.

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics, 28, 2537-2539.

Pope, T. R. (1992). The influence of dispersal patterns and mating systems on genetic differentiation within and between populations of the red howler monkey (*Alouatta seniculus*). Evolution, 46(4), 1112-1128.

Pusey, A., & Wolf, M. (1996). Inbreeding avoidance in animals. Trends in Ecology & Evolution, 11(5), 201-206.

Ryden, H. (2005). America's last wild horses (Revised ed.) New York: Lyons Press, USA.

Seddon, J., & Baverstock, P. (1999). Variation on islands: Major histocompatibility complex (MHC) polymorphism in populations of the Australian bush rat. Molecular Ecology, 8(12), 2071-2079.

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. Science (New York, N.Y.), 236(4803), 787-792.

Takezaki, N., Nei, M., & Tamura, K. (2010). POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. Molecular Biology and Evolution, 27(4), 747-752.

Tozaki, T., Kakoi, H., Mashima, S., Hirota, K., Hasegawa, T., Ishida, 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.

Trowsdale, J. (1995). "Both man & bird & beast": Comparative organization of MHC genes. Immunogenetics, 41(1), 1-17.

Tseng, C., Miller, D., Cassano, J., Bailey, E., & Antczak, D. (2010). Identification of equine major histocompatibility complex haplotypes using polymorphic microsatellites. Animal Genetics, 41(s2), 150-153.

Valera, M., Molina, A., Gutiérrez, J. P., Gómez, J., & Goyache, F. (2005). Pedigree analysis in the Andalusian horse: Population structure, genetic variability and influence of the Carthusian strain. Livestock Production Science, 95(1), 57-66.

Varvio, S., Chakraborty, R., & Nei, M. (1986). Genetic variation in subdivided populations and conservation genetics. Heredity, 57(2), 189-198.

Vila, C., Leonard, J. A., Gotherstrom, A., Marklund, S., Sandberg, K., Liden, K., . . . Ellegren, H. (2001). Widespread origins of domestic horse lineages. Science, 291(5503), 474-477.

Warmuth, V., Eriksson, A., Bower, M. A., Barker, G., Barrett, E., Hanks, B. K., . . . Manica, A. (2012). Reconstructing the origin and spread of horse domestication in the Eurasian steppe. Proceedings of the National Academy of Sciences of the United States of America, 109(21), 8202-8206.

Wynne, J. W., Cook, M. T., Nowak, B. F., & Elliott, N. G. (2007). Major histocompatibility polymorphism associated with resistance towards amoebic gill disease in Atlantic salmon (*Salmo salar L.*). Fish & Shellfish Immunology, 22(6), 707-717.

Yeager, M., & Hughes, A. L. (1999). Evolution of the mammalian MHC: Natural selection, recombination, and convergent evolution. Immunological Reviews, 167(1), 45-58.

Zechner, P., Sölkner, J., Bodo, I., Druml, T., Baumung, R., Achmann, R., . . . Brem, G. (2002). Analysis of diversity and population structure in the Lipizzan horse breed based on pedigree information. Livestock Production Science, 77(2), 137-146.

CHAPTER 5

5. THE EQUINE VOLATILOME IS UNIQUE BUT CAN ALSO REFLECT KINSHIP: A POSSIBLE MECHANISM FOR INBREEDING AVOIDANCE

This manuscript has been submitted to: Journal of Chemical Ecology

5.1. Introduction

Olfactory cues function in the animal kingdom to distinguish between kin and predators, and to detect food sources and environmental toxins. The core body odors (volatilome) are emitted volatile organic compounds (VOCs) that are the end products of metabolism (Wedekind & Penn, 2000). Other factors such as genetics, diet, environment, bacteria present on the body and exogenous materials that modulate the VOCs can also contribute to an individual's odor profile. Moreover, disease processes, such as infection and endogenous metabolic disorders, can influence an individual's odor profile by producing different volatiles or by changing the proportions of VOCs that are normally produced (Shirasu & Touhara, 2011) and subsequently convey information about an individual's metabolic or psychological status.

In recent years, a number of studies have been conducted to elucidate the genetic link and biological pathway for body odor production. It has been postulated that body odors are linked to the polymorphic genes within the major histocompatibility complex (MHC) (Thomas & Parker, 1974). Since the identification of MHC-related odor cues in urine and MHC-dependent mating preferences in rodents (Beauchamp, et al., 1988;

Eklund, 1997; Yamazaki, et al., 1976; Yamazaki, et al., 1988), several animal studies have further presented evidence of the MHC genotypes influencing mate choice associated with body odor (Beauchamp, et al., 1994; Lanyon, et al., 2007; Novotny, et al., 2007; Singer, et al., 1997; Singh, et al., 1987). The olfactory differentiation for mating preferences and conspecific identification or kin recognition has been studied across various taxonomic groups including fish (Aeschlimann, et al., 2003; Milinski, et al., 2005; Olsén, et al., 2002; Rajakaruna, et al., 2006), Swedish sand lizards (Olsson et al., 2003), birds (Ekblom, et al., 2004; Strandh, et al., 2012; Zelano & Edwards, 2002), Tuco-tucos (Cutrera, et al., 2012), rodents (Busquet & Baudoin, 2005; Schwagmeyer, 1988; Todrank, et al., 1998), otters (Kean et al., 2015), lemurs (Boulet, et al., 2009; Charpentier, et al., 2008; Knapp, et al., 2006) as well as in old and new world monkeys (Setchell, et al., 2011; Smith, et al., 1997). Although most of these studies do not focus on individual components of the odor, they led to the conclusion that individuals have a distinct body-odor type, which is determined or influenced by their inherited MHC alleles, playing a pivotal role in kin recognition, mate selection and identification of dissimilar or similar individuals (Wyatt, 2003).

Animals rely on subtle signals perceived between individuals conveying information including sex, reproductive status, individual identity, ownership, competitive ability as well as health status. These cues have important influences on a variety of behaviors that are vital for reproductive success, such as parent–offspring attachment, recognition of relatedness, mate choice and territorial marking (Wyatt, 2003). Consequently, the ability to recognize individuals or their genetic relatedness plays an

important role in animal social behavior. Relatively few cases of olfactory discrimination from different conspecifics have been explored in domestic animals. Studies on pigs indicate that young pigs could use urinary cues to discriminate other individuals, not just between group members and non-group members, but also between different non-group members that facilitated the formation and maintenance of stable social groupings (McLeman, et al., 2005; Mendl, et al., 2002). Similar findings have been observed in domestic goats, calves (Baldwin, 1977) and sheep. Furthermore, in humans, the existing "individual odor hypothesis" suggests that each individual possesses a unique scent, which acts as a characteristic or odor fingerprint (Penn, et al., 2007). Thus, a combination of the presence and abundance of VOCs produces a chemical profile that is distinctive to an individual and can be seen as a biometric measurement (Curran, et al., 2010). The chemical constituents of human odor have been shown to be qualitatively similar among individuals; however, the abundance of each compound produced makes the scent specific to the individual (Curran, et al., 2005; Curran, et al., 2007; Curran, et al., 2010). Such qualitative and quantitative patterns have not been explored in domestic or wild animals and there is a need to understand the influence of odor on their behavior and interactions.

Horses, like most mammals, can recognize and discriminate chemical signals, which provide essential information for individual and herd survival and greatly influence their social behavior (Saslow, 2002). The large olfactory bulbs in a horse's brain exhibit a convoluted surface, while the large size of the olfactory epithelium suggests that olfactory information is vital to horses. Additionally, horses exhibit a well-

developed vomeronasal organ that is receptive to nonvolatile, large, species-specific molecules found in body secretions (Saslow, 2002); hence, they have a highly developed olfactory capacity. Observational studies of domestic and wild/feral horses have described how horses recognize each other on the basis of body odors (Kiley-Worthington, 1997). Recognition at the individual level guides the horse's response based on previous experience and determines the outcome of the interactions. Close proximity, and mutual sniffing have been seen during horse greetings and sexual advances that include blowing air on the face, standing parallel and sniffing the neck and under another's bellies (Stahlbaum & Houpt, 1989). Horses often sniff excrement (dung piles) (Kimura, 2001) that allows horses to recognize other individuals (Ainslie & Ledbetter, 1980; Krueger & Flauger, 2011) and to differentiate the sex of the horse by its feces (Stahlbaum & Houpt, 1989). In fact, stallions create fecal piles known as stud piles and repeatedly return to them to defecate as a territorial marking behavior (Feist & McCullough, 1976; Rubenstein & Hack, 1992). Similar practices are seen when the harem stallion covers an area with his urine where the harem female horses had previously urinated or defecated (McDonnell, 2003; Ransom & Cade, 2009). Such greetings and scent marking displayed by feral and domestic horses, suggest that odor is crucial in social encounters and is used to gather useful information from the chemical cues (Ainslie & Ledbetter, 1980)

Although some research has been carried out on olfactory communication in horses, it has been limited to behavioral observations of related and unrelated horses or mating preferences based on the identification of another individual by sniffing urine and

feces. Mozūraitis et al., identified volatiles such as diethylphthalate, m- and p-cresols in mare urine for which absolute concentrations showed a temporal patterning (Mozūraitis, et al., 2012). Similar studies on mare urine have shown significant changes in the number and abundance of volatiles during different stages of the mare's reproductive cycle (Ma & Klemm, 1997). A habituation-discrimination experiment using social cues of urine, feces and body odor samples of unfamiliar horses showed that the tested horses were able to memorize the scent of other horses' urine, feces, and body odors at only the second presentation (Hothersall, et al., 2010). These studies did not, however, aim to individualize horses based on the components of their body odor. Conversely, no chemical investigations have been performed on hair or body areas that horses sniff while greeting or grooming one another.

Though there is evidence for MHC-dependent mating preferences in these animals, the underlying mechanisms remain elusive. Moreover, it is still unclear whether a unique suite of odor compounds actually drives individual recognition. Based on the "individual odor hypothesis" put forth by Penn et al, that suggests each individual possesses a unique VOC fingerprint (Penn, et al., 2007), the objective of this present study was to investigate odor profiles from domestic horses, *Equus caballus*. VOCs were chemically analyzed using solid-phase microextraction gas chromatography–mass spectrometry (SPME GC-MS). Evidence for equine body volatiles that are presumably used for individual discrimination is lacking; therefore, this study examined the VOC profiles of domestic horses to determine the components of individual odor and whether these profiles can reliably indicate a degree kinship.

5.2. Methodology

5.2.1 Sample Collection

Triplicate hair samples were donated by the horse owners and plucked from the manes of 23 domestic horses, 11 horses with known kinships and 12 unrelated domestic horses (Appendix 2). The horse belonged to two breeds (6 Appaloosa and 17 Quarter Horses). Mane hair samples (each ≈10 mg) were collected from each horse using sterile disposable gloves. The sample were immediately placed into 10-mL glass clear, screw top vials with PTFE/Silicone septa (Supelco, Bellefonte, PA) and sealed. Samples were allowed to equilibrate for 24 h prior to SPME extraction. The usages of specific horse fly sprays, feed, and bathing routines were noted for all samples. Horses were not bathed before sampling and use of fly spray was avoided with the exception of Horse H5 who suffered from insect bite hypersensitivity and was under treatment. His body was sprayed with water before sampling since fly spray had been used one hour prior to hair collection. Empty 10-mL vials (Supelco) were exposed to the stable air for background VOC collection for each sampling time and those compound were subtracted from each VOC profile before analyses.

5.2.2 Solid-phase microextraction (SPME)-GC-MS Procedures

Divinylbenzene/Carboxen on Polydimethylsiloxane (CAR/DVB on PDMS) 50/30 μm fibers (Supelco) were used to extract the VOCs from the headspace of the vials containing the hair samples. Fiber exposure was conducted at room temperature for 12 hours. The samples were separated and analyzed by GC-MS using an Agilent 6970 GC with a 5973 MS. A HP5-MS column (0.25 mm × 30 m × 0.25 μm thickness), with

helium as the carrier gas at flow rate of 1.0 mL/min for the separation of the analytes was used. The extracted VOCs were desorbed in the injection port at 250°C for 10 min in splitless mode. The temperature program was: an initial oven temperature of 40°C for 5 min, 10°C per minute ramp to a final temperature of 250°C, followed by a final hold for 2 min for a total run time of 32 min. The mass spectrometer used was an HP 5973 MSD with a quadrupole analyzer in full scan mode (mass range: 50–550). The compounds were identified using the NIST 98 mass spectral library. The criterion for the identification of compounds was based on the quality of the detected peak, which was set at greater than or equal to 40%. Environmental blanks were assessed following the same GC-MS parameters employed for the hair samples.

5.2.3. Statistical Analysis

Prior to statistical analyses, data were reduced by removing compounds present in the environmental blanks and in Pyranha fly spray based on the MSDS description. Further, compound names were searched on ChemSpider (www.chemspider.com) (Pence & Williams, 2010) for their equine husbandry use or presence in nature. Compounds commonly associated with ointments and plant/feed related compounds were removed from the analysis. The relative ratio of each VOC's abundance was calculated for all VOCs and, subsequently transformed using square-root transformation.

Spearman rank correlation

Each horse's VOC profile was statistically evaluated to determine the similarity between each replicate as well as the differences for each of the 23 individuals sampled. The replicate profiles for each horse were then averaged to produce a single

representative VOC profile. Each individual VOC profile was correlated, in a pair wise manner, to the rest of the horses using Spearman rank correlations. The association between horses was represented by a correlation coefficient that ranged in value from -1 to +1, showing either a negative correlation or a positive correlation, respectively.

Multivariate analysis on volatiles

A non-metric multidimensional scaling (nMDS) analysis was performed in order to identify similarity/differences of (i) replicates (three) from each horse (ii) horse breeds (Appaloosa and Quarter Horse) and (iii) kinships. The analyses were performed using Primer-E ver 7 software (Clarke & Gorley, 2015) to create a similarity matrix among variables using the Bray-Curtis similarity coefficient. The overall goodness of fit of the nMDS plot models was measured by the stress statistic, which is the correlation between fitted values and ordination distances.

Differences in composition of compounds within groups were tested using an Analysis of Similarities (ANOSIM) (Clarke, 1993). ANOSIM Global R values range from 0 to 1, with an associated significance value with each Global R. A zero value indicates similarities among and within groups do not differ while a value of 1 indicates that samples within each group are more similar to each other than to those from other groups. The relative ratios were used for hierarchical cluster analysis (HCA) using the complete linkage method that takes into account the similarity of two clusters and compares it to the similarity of their most dissimilar members. In order to observe percent similarity, the cluster analysis output was overlaid on the nMDS ordination plots, indicated by the circles grouping the data.

Similarity Percentage analysis (SIMPER) analysis was performed to determine which compounds influenced the discrimination observed between the individuals, breeds and kin. The impact of compounds in groups was calculated in Primer-E ver 7 software (Clarke & Gorley, 2015) by summing the relative ratio contribution of each compound over the average dissimilarity of all the compounds (Clarke, 1993).

Heatmaps were generated in Primer-E ver 7 software (Clarke & Gorley, 2015) to visualize the contribution of top 47 compounds obtained in SIMPER analyses. The heatmap is based on relative ratio of compounds where the regions of white represent absence of compound, blue represents low relative abundance while dark red indicates higher relative abundance with highest value of 0.52.

5.3. Results

The hair samples collected from the 23 horses revealed compounds from various functional groups: alkanes, alcohols, hydrocarbons, aldehydes, alkenes, amines/amides, ketones, esters, and indols (n=187) (Figure 11). Three VOC with varying abundances were detected in 100% of horses sampled; nonanal, fluoren-9-ol, 3,6-dimethoxy-9-(2-phenylethynyl)-, and 1,2-dibromo-2-methyl- undecane. Furthermore, compounds specific to the Appaloosa breed were found. These included 3,5-bis (1,1-dimethylethyl)-1,2-benzenediol, 9-octadecen-1-ol, 1-methyl-2-phenyl-1H-indole and (1-butylnonyl)-benzene. Ninety-eight compounds were specific to Quarter Horses; however, no one compound was common across all 17 Quarter Horses.

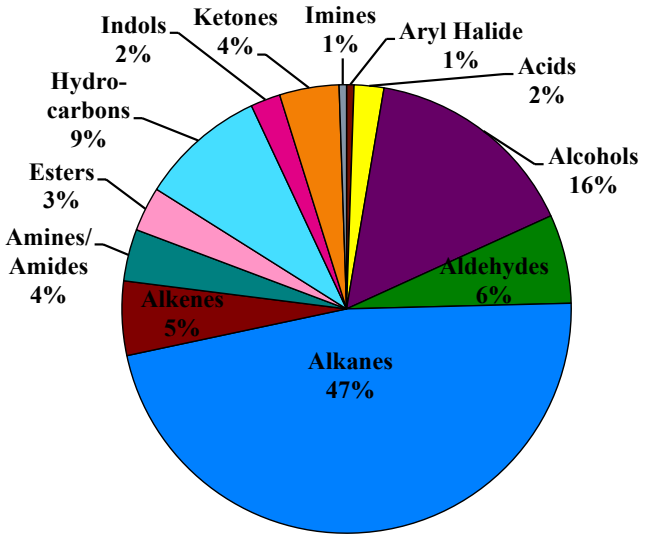

Figure 11. Functional VOC group distribution from hair samples of 23 horses. Alkanes are the most abundant in the VOCs of both breeds while Aryl halide was found only in Quarter Horses.

5.3.1. Spearman rank correlations

Correlation matrices showed the triplicate VOC profiles had a high correlation coefficient ≥ 0.8. Among horses of known relationships (parent-offspring, full sibling sharing dam and sire, half sibling sharing dam or sire), correlation coefficient of ≥ 0.7 with the exception of H15 to H16 was observed (Appendix 3). These values demonstrate that, though there are qualitative similarities within replicates from the same horse, the difference in relative abundance ratios allowed for a higher percentage of discrimination between related and unrelated horses.

5.3.2. Multivariate analysis

The nMDS plot (Figure 12) with a stress value of 0.15 and 80% similarity shows the VOC replicates for each horse. This indicated that hair sampling for replicates was

consistent and reproducible. In support of the 'individual odor hypothesis', there was a significant difference between each horse profile (R = 0.9; *p<0.001*). The nMDS plot (Figure 13) using the VOC averages with a stress value of 0.15 also showed a clear separation of Appaloosa and Quarter Horse breeds (R=0.814; *p<0.001*). Based on the cluster analysis (Figure 14) the 40% similarity grouping was observed for related horses while 20% similarity was seen to distinguish horse breeds. The nMDS plot (Figure 15) with a stress value of 0.15 indicated that the sampling site (stables) had minimal effect on grouping of the horses when related individuals are included in the study. This suggests that animal husbandry practices did not have an influence on the VOC profiles. Overall, plots showed that VOC profiles are different for each horse and subsequently can individualize them. Although the profiles are distinct, VOCs between related horses as well as some breed specific patterns were detected.

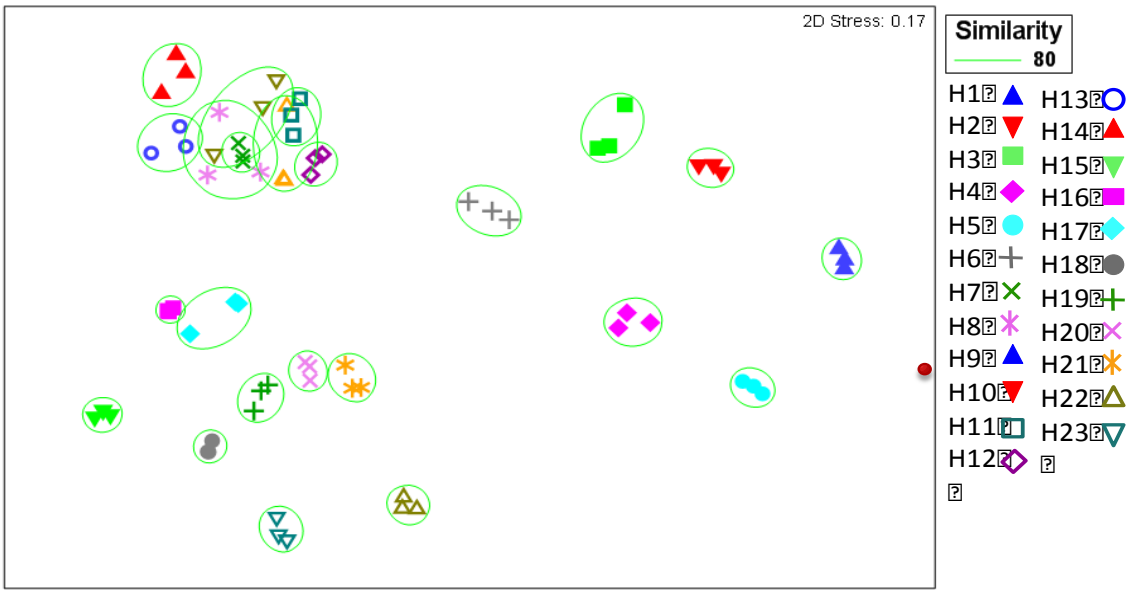

Figure 12. The nMDS plot with similarity analysis overlay showed a tight grouping of replicate samples from each horse.

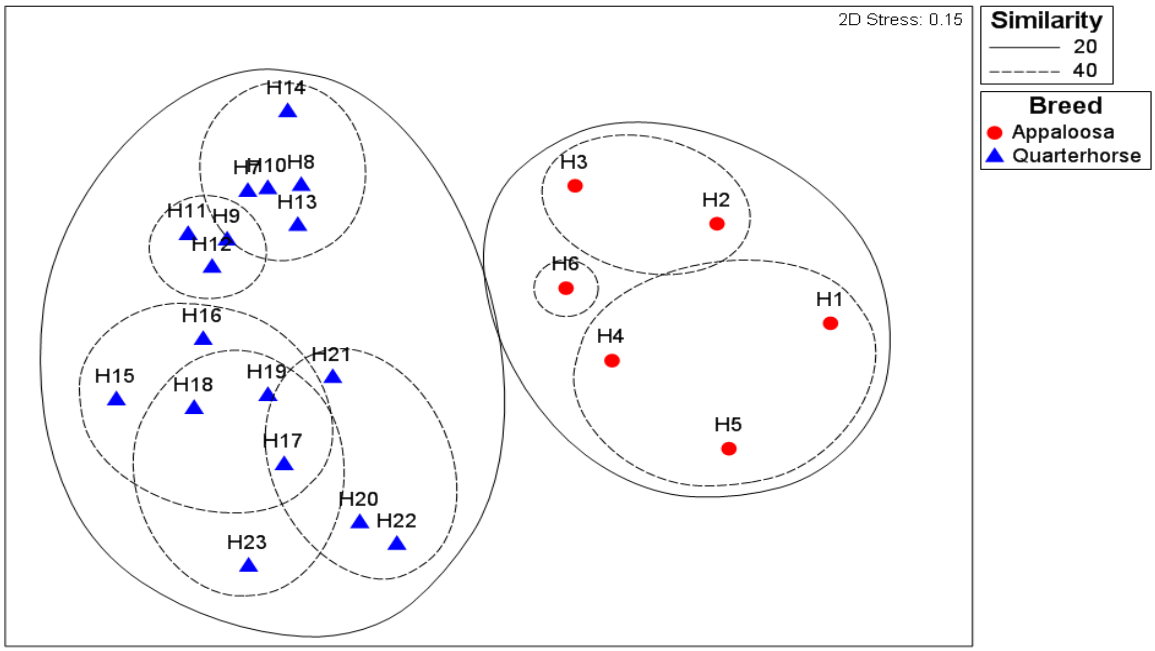

Figure 13. The nMDS plot with similarity analysis overlay grouped the horses according to the two breeds included in study.

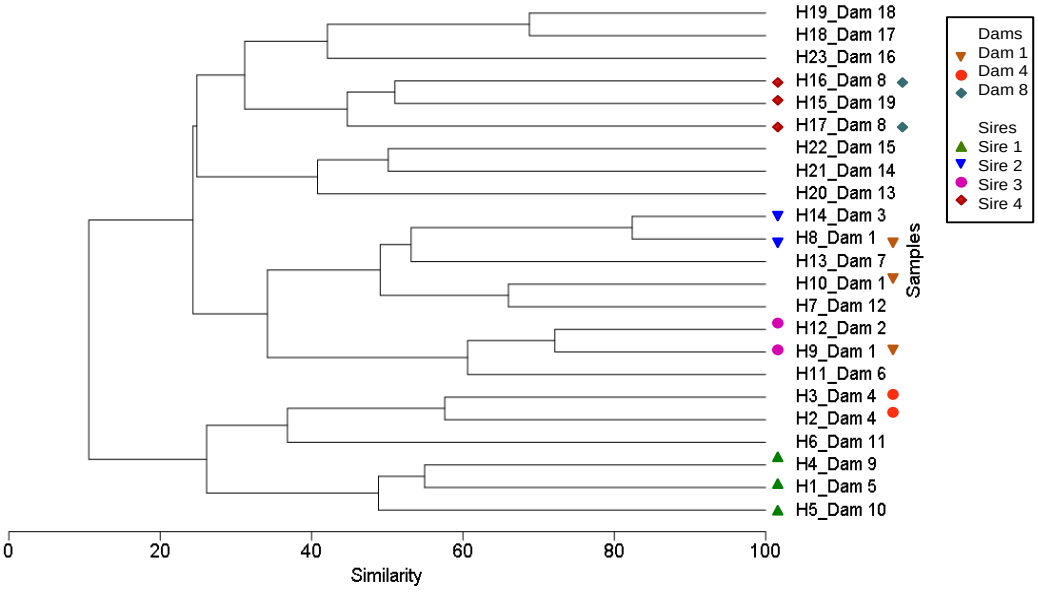

Figure 14. Dendrogram from Bray-Curtis similarity matrix using a hierarchical cluster analysis. Cluster analysis showed closer grouping of related horses and differences between breeds.

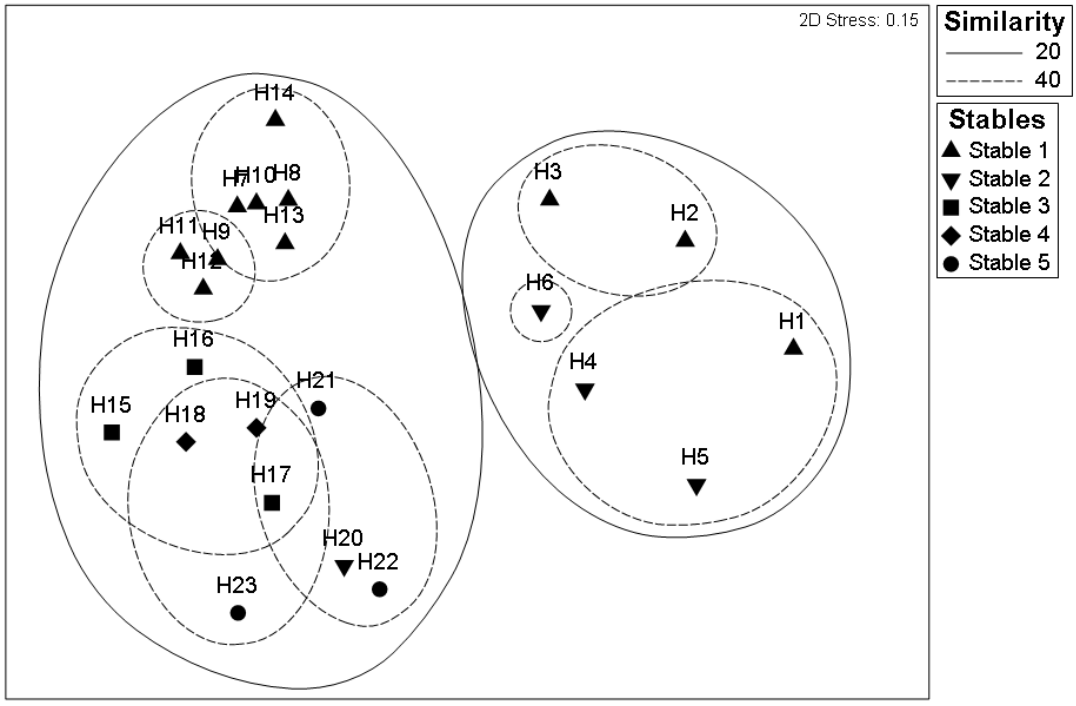

Figure 15. The nMDS plot with similarity analysis overlay indicated that the stable in which the horses had little influence on the VOC profiles for each horse. The hierarchical clustering clearly grouped the Appaloosa horses separate from the Quarter Horses. Although clustering of related horse is quite evident, clustering of distant relationships was also observed for horses H15 – H16 – H17, were H15 is the uncle to H16 and H17 from the dam's lineage. Similarly although H1-H4-H5 are related directly through the sire, H1 is also distantly related to H2, as his dam's sire is the same as H2's sire.

SIMPER analysis of the dominant 47 compounds was able to discriminate the Appaloosa and Quarter Horse breeds (Table 17). The top 10 compounds were 9-octadecen-1-o, 2-decen-1-ol, quinolone, 2-chloro-6-methoxy-4-methyl-, Undecane, 1,2-dibromo-2-methyl-, Tridecane, 4-cyanocyclohexene, 2-nonen-1-ol, (E)-, Nonanal, 4-nitro-4'-chlorodiphenylsulphoxide and Undecanal. The average dissimilarity between the breeds was 77.03%.

Table 17. List of compounds influencing variation between breeds.

Functional Group	Compound	Functional Group
Acid	C053	3,3-dimethyl-Butanoic acid
Alcohol	C011	3,5-bis(1,1-dimethylethyl)-1,2-Benzenediol
	C012	(Z)-2-Hexen-1-ol
	C014	2-Decen-1-ol
	C016	9-Octadecen-1-ol
	C017	.alpha.,.alpha.-dimethyl-Benzenemethanol
	C018	(E)-2-Nonen-1-ol
	C040	2-phenoxy-Ethanol
	C042	3,6-dimethoxy-9-(2-phenylethynyl)-Fluoren-9-ol
	C151	2,6-Bis(1,1-dimethylethyl)-4-(1-oxopropyl)phenol
	C179	Phenol, 4,4'-(1-methylethylidene)bis-
Aldehyde	C021	Nonanal
	C023	(E)-2-Nonenal
	C027	Dodecanal
	C028	Undecanal
	C039	Decanal
Alkane	C036	Dodecane
	C050	2,6-dimethyl-Undecane
	C052	2,3,5-trimethyl-Hexane
	C054	4-methyl-Dodecane
	C058	1,2-dibromo-2-methyl-Undecane
	C059	2-methyl-6-propyl-Dodecane
	C070	Tridecane
	C092	9-methyl-Nonadecane
	C093	5-methyl-Tridecane
	C104	2,6,10,14-tetramethyl-Hexadecane
	C107	Tetradecane
	C125	Cyclododecane
	C133	Pentadecane
	C135	Cyclopentadecane
	C180	Eicosane
Alkene	C008	4-Cyanocyclohexene
Amines Amides	C002	2-chloro-6-methoxy-4-methyl-Quinoline
	C033	6-Methyl-2-phenyl-7-(2,4,5-trimethylphenylmethyl) indolizine
	C043	4-(4-Chlorophenyl)-2,6-diphenylpyridine
	C069	2-Chloro-4-(4-methoxyphenyl)-6-(4-

		nitrophenyl)pyrimidine
Ester	C186	2-Amino-2-oxo-acetic acid, N-[3,4-dimethylphenyl]-, ethyl ester
Hydrocarbons	C006	Benzene, 1-phenyl-4-(2-cyano-2-phenylethenyl)
	C019	(1-methoxyethyl)-Benzene
	C123	4-Nitro-4'-chlorodiphenylsulphoxide
	C152	(1-butylheptyl)-Benzene
	C153	(1-propyloctyl)-Benzene
	C162	(1-butyloctyl)-Benzene
	C173	(1-butylnonyl)-Benzene
Imines	C001	Oxime-, methoxy-phenyl-_
Indole	C155	1-methyl-2-phenyl-1H-Indole
Ketone	C073	1-(2,2-dimethylcyclopentyl)-Ethanone

Using these 47 discriminatory compounds, the heatmap (Figure 16) indicated the relative ratio differences that provide individual uniqueness and the clustering of the two breeds. Related horses display the presence of similar compounds; however the relative ratios vary between the horses. These differences individualize the horses irrespective of breed. The heat map revealed that the related and unrelated horses contain different combinations of the 47 volatile compounds. The six Appaloosa horses (H1-H6) had higher levels of volatiles alcohols (C42, C5, C7, C31, C17, C24, C42, C5, C7, C31, C17, C24), alkene (C8, C13), amine (C3), alkane (C26), amine (C33), and hydrocarbon (C32) whereas H7 - H23 had lower levels of these compounds (Figure 16). The concentration or relative ratio of compounds differed across H1-H23 rather than just a simple presence or absence of compounds.

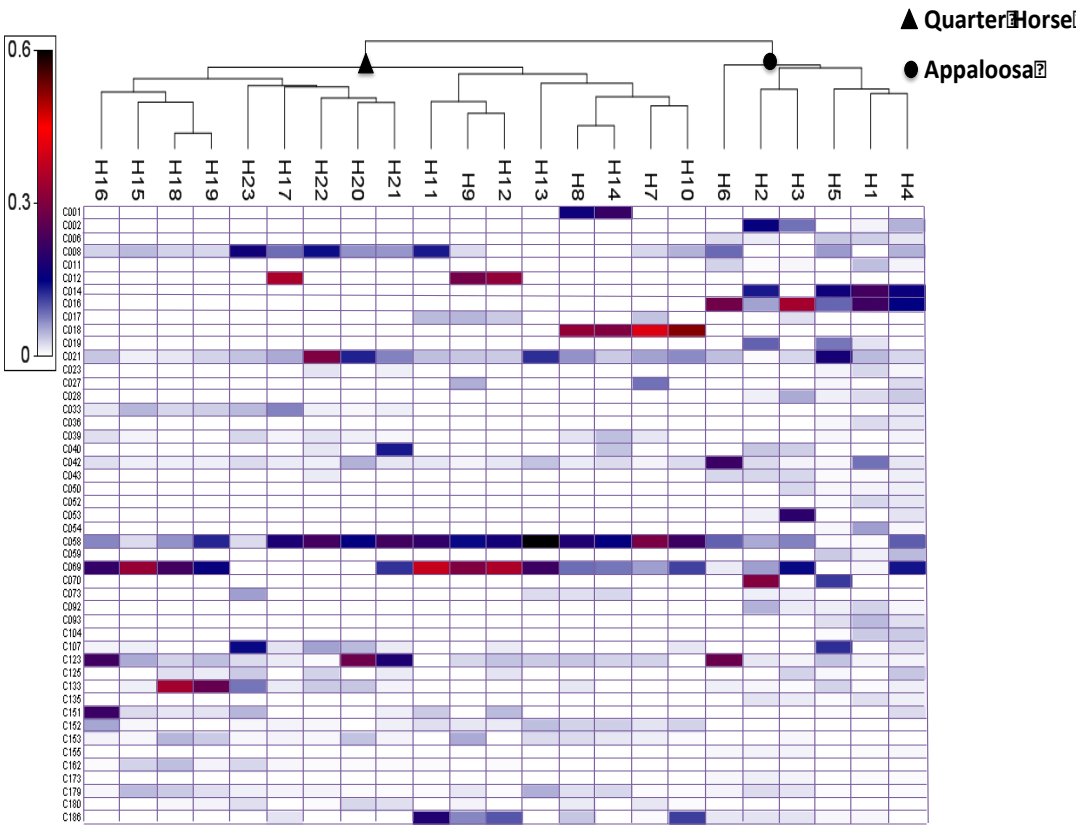

Figure 16. A hierarchical cluster analyses and heatmap of the volatiles that differentiated Appaloosa and Quarter Horse breeds as well as each individual.
The X-axis represents a grouping of related and unrelated horses, while Y-axis are the 47 compounds contributing to differences in breeds. White represents no difference in relative abundance and dark red represents maximum relative abundance.

5.4. Discussion

The potential role of olfaction in horses is known by a relatively small number of published studies concentrated primarily on the role of odor in horse mating behavior and sex identification. Such studies have previously focused on volatile chemical markers from feces and urine samples that fluctuate with ovulation or the estrus cycle in mares (Ma & Klemm, 1997; Mozūraitis, et al., 2012) and further elicit a response from stallions. The current study targeted odor profiles from hair samples in order to learn whether

particular individuals have a specific profile. Cumulative VOC profiles considering presence/absence of compounds along with relative ratios can successfully individualize horses supporting the "individual odor hypothesis" (Penn, et al., 2007). As has been seen in vertebrate species like lemurs (Palagi & Dapporto, 2006), mandrills (Setchell, et al., 2010), deer (Gassett, et al., 1997), and odor profiles in these horses, a high percentage of chemical compounds were shared among profiles of all horses. However, the significant differences between each individual suggested that the uniqueness in the chemical profile may depend more on the concentration of compounds and on complex interactions between compounds, than on the simple presence or absence of specific chemicals compounds (Singer, et al., 1997). Therefore, the variation in volatiles between individuals was indicative of an individual-specific odor that could aid in individual identification, which may influence the behavioral responses (Celerier, et al., 2010). Although only six Appaloosa horses were sampled compared to 17 Quarter Horses, it is interesting to observe that volatiles patterns differentiated between breeds. This signal of breed identity could be attributed to related groups (H1-H4-H5 and H2-H3) within the Appaloosa horses forcing them to group together away from the related and unrelated Quarter Horses. Nevertheless, the ability to utilize the individual odor profiles must rely on phenotypic comparisons which includes similarity in odor production by related individuals, the perception of such similarity, and eventually the behavioral reaction to this perception (Celerier, et al., 2010)

From an evolutionary standpoint, kin recognition aids in parental care, kin altruism, inbreeding avoidance and maintenance of optimal outbreeding. Given a choice,

most animals mate with unrelated or distant relatives. This innate selection is thought to be evolutionarily favored, as it should improve their inclusive fitness and decrease the effects of inbreeding on the population (Boulet, et al., 200)). This manner of discrimination can be treated as a communication mechanism where prior to identification, the animal must be able to receive relevant signals from other individuals and distinguish appropriately between them (Tang-Martinez, 2001). This can occur provided the individual odor cues vary with relatedness. Therefore, studying body volatiles may enable one to better understand how animals use kin-biased behavior in their interactions within breeding populations (Beecher, 1989).

When considering relationships in horses, overall lineages need to be queried rather than direct relationships of parent-offspring or siblings. Such lineages could group horses that might not share a direct relationship to sire or dam. Although all sire and dam pairs were not known in the study, VOC profiles were able to demonstrate relatedness between horses with known lineages. Although the closely related horses had similar chemical components (Figure 15), it cannot be said that horses from the same kin group share the exact odor phenotype. The greater similarity detected between odor profiles of half siblings compared to distant kin indicated a graded or continuous relationship of odors within kin groups when using relative abundances rather than a discrete one of presence/absence (Busquet & Baudoin, 2005). In this study on rodents, males recognized similarities in the odors of brothers when compared to an unrelated male. While similarities were observed between the odors of double cousins and cousins demonstrating that kinship from siblings to cousins is reflected in odor similarity

(Busquet & Baudoin, 2005). It can also be postulated that similar to wild birds (Bonadonna & Sanz-Aguilar, 2012) and primates (Boulet et al., 2009), horses may be recognizing similar odor profiles and respond to individuals based on the degrees of genetic relatedness and to identify kin/non-kin.

Horses that were housed at the same location but not related demonstrated that housing and management appeared to have little influence on their volatile profiles. This is contrary to otters housed in the same location where the similarity was attributed to common diet as a potential explanation (Davies, 2008). In this study, all the horses were fed similar feed and managed in a similar fashion. Nevertheless, grouping of horses with unknown relationships may be due to variables that were not investigated in this study, for example, age, current reproductive status or health status of the individuals.

It has been suggested that individual identification is one of the most important cues used in vertebrate chemical communication (Wilson, 1970). The evaluation of relatedness may be crucial in socio-sexual behavior of harem/herd animals like horses, especially if left to random mating instead of artificial selection now used in domestic horse breeding practices. Odor profiles can provide an array of information about the horse's social, reproductive, or health status and as was seen in this study, the animal's identity, breed, and kinship. It has been established that VOCs may play an important part in identifying individuals, establishing dominance (Marty, et al., 2009), signaling sexual readiness (Charpentier, et al., 2008), facilitating mate choice for genetically dissimilar individuals (Setchell, et al., 2009) and inbreeding avoidance (Charpentier, et al., 2005). The results in this study indicate that individual odor profiles could play a key

role in signaling individual characteristics, relatedness and that these volatile cues may possibly be used for kin recognition in horses.

5.5. References

Aeschlimann, P., Häberli, M., Reusch, T., Boehm, T., & Milinski, M. (2003). Female sticklebacks gasterosteus aculeatus use self-reference to optimize MHC allele number during mate selection. Behavioral Ecology and Sociobiology, 54(2), 119-126. doi:10.1007/s00265-003-0611-6

Ainslie, T., & Ledbetter, B. (1980). The body language of horses. William Morrow, New York.

Baldwin, B. (1977). Ability of goats and calves to distinguish between conspecific urine samples using olfaction. Applied Animal Ethology, 3(2), 145-150. doi:10.1016/0304-3762(77)90023-2

Beauchamp, G. K., Yamazaki, K., Bard, J., & Boyse, E. A. (1988). Preweaning experience in the control of mating preferences by genes in the major histocompatibility complex of the mouse. Behavior Genetics, 18(4), 537-547. doi:10.1007/BF01065520

Beauchamp, G. K., Yamazaki, K., Curran, M., Bard, J., & Boyse, E. A. (1994). Fetal H-2 odortypes are evident in the urine of pregnant female mice. Immunogenetics, 39(2), 109-113. 109. doi:10.1007/BF00188613

Beecher, M. D. (1989). Signalling systems for individual recognition: An information theory approach. Animal Behaviour, 38(2), 248-261. doi:10.1016/S0003-3472(89)80087-9

Bonadonna, F., & Sanz-Aguilar, A. (2012). Kin recognition and inbreeding avoidance in wild birds: The first evidence for individual kin-related odour recognition. Animal Behaviour, 84(3), 509-513. doi:10.1016/j.anbehav.2012.06.014

Boulet, M., Charpentier, M. J., & Drea, C. M. (2009). Decoding an olfactory mechanism of kin recognition and inbreeding avoidance in a primate. BMC Evolutionary Biology, 9, 281-2148-9-281. doi:10.1186/1471-2148-9-281

Busquet, N., & Baudoin, C. (2005). Odour similarities as a basis for discriminating degrees of kinship in rodents: Evidence from *Mus spicilegus*. Animal Behaviour, 70(5), 997-1002. doi:10.1016/j.anbehav.2004.12.023

Celerier, A., Huchard, E., Alvergne, A., Fejan, D., Plard, F., Cowlishaw, G., . . . Bonadonna, F. (2010). Detective mice assess relatedness in baboons using olfactory cues. The Journal of Experimental Biology, 213(Pt 9), 1399-1405. doi:10.1242/jeb.038349

Charpentier, M. J., Boulet, M., & Drea, C. M. (2008). Smelling right: The scent of male lemurs advertises genetic quality and relatedness. Molecular Ecology, 17(14), 3225-3233. doi:10.1111/j.1365-294x.2008.03831.x

Charpentier, M., Peignot, P., Hossaert-McKey, M., Gimenez, O., Setchell, J. M., & Wickings, E. J. (2005). Constraints on control: Factors influencing reproductive success in male mandrills (*Mandrillus sphinx)*. Behavioral Ecology, 16(3), 614-623. doi:10.1093/beheco/ari034

Clarke, K. R., & Gorley, R. N. (2015). PRIMER v7: User manual/tutorial PRIMER-E, Plymouth, 296pp.

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. Australian Journal of Ecology, 18(1), 117-143. doi:10.1111/j.1442-9993.1993.tb00438.x

Curran, A. M., Prada, P. A., & Furton, K. G. (2010). The differentiation of the volatile organic signatures of individuals through SPME-GC/MS of characteristic human scent compounds. Journal of Forensic Sciences, 55(1), 50-57. doi:10.1111/j.1556-4029.2009.01236.x

Curran, A. M., Rabin, S. I., Prada, P. A., & Furton, K. G. (2005). Comparison of the volatile organic compounds present in human odor using SPME-GC/MS. Journal of Chemical Ecology, 31(7), 1607-1619. doi:10.1007/s10886-005-5801-4

Curran, A. M., Ramirez, C. F., Schoon, A. A., & Furton, K. G. (2007). The frequency of occurrence and discriminatory power of compounds found in human scent across a population determined by SPME-GC/MS. Journal of Chromatography B, 846(1), 86-97. doi:10.1016/j.jchromb.2006.08.039

Cutrera, A. P., Fanjul, M. S., & Zenuto, R. R. (2012). Females prefer good genes: MHC-associated mate choice in wild and captive Tuco-tucos. Animal Behaviour, 83(3), 847-856. doi:10.1016/j.anbehav.2012.01.006

Davies, M. J. (2008). Chemical communication in the European otter, *Lutra lutra*. Unpublished University of Hull

Ekblom, R., Saether, S. A., Grahn, M., Fiske, P., Kålås, J. A., & Höglund, J. (2004). Major histocompatibility complex variation and mate choice in a lekking bird, the great snipe (*Gallinago media)*. Molecular Ecology, 13(12), 3821-3828. doi:10.1111/j.1365-294x.2004.02361.x

Eklund, A. (1997). The major histocompatibility complex and mating preferences in wild house mice (*Mus domesticus*). Behavioral Ecology, 8(6), 630-634. doi:10.1093/beheco/8.6.630

Feist, J. D., & McCullough, D. R. (1976). Behavior patterns and communication in feral horses. Zeitschrift Für Tierpsychologie, 41(4), 337-371. doi:10.1111/j.1439-0310.1976.tb00947.x

Gassett, J., Wiesler, D., Baker, A., Osborn, D., Miller, K., Marchinton, R., & Novotny, M. (1997). Volatile compounds from the forehead region of male white-tailed deer (*Odocoileus virginianus*). Journal of Chemical Ecology, 23(3), 569-578. doi:10.1023/b:joec.0000006397.63647.5b

Hothersall, B., Harris, P., Sörtoft, L., & Nicol, C. J. (2010). Discrimination between conspecific odour samples in the horse (*Equus caballus*). Applied Animal Behaviour Science, 126(1), 37-44. doi:10.1016/j.applanim.2010.05.002

Kean, E. F., Chadwick, E. A., & Müller, C. T. (2015). Scent signals individual identity and country of origin in otters. Mammalian Biology-Zeitschrift Für Säugetierkunde, 80(2), 99-105. doi:10.1016/j.mambio.2014.12.004

Kiley-Worthington, M. (1997). The behaviour of horses: In relation to management and training. Applied Animal Ethology, 9(1), 99. doi:10.1016/0304-3762(82)90184-5.

Kimura, R. (2001). Volatile substances in feces, urine and urine-marked feces of feral horses. Canadian Journal of Animal Science, 81(3), 411-420. doi:10.4141/a00-068

Knapp, L. A., Robson, J., & Waterhouse, J. S. (2006). Olfactory signals and the MHC: A review and a case study in *Lemur catta*. American Journal of Primatology, 68(6), 568-584. doi: 10.1002/ajp.20253

Krueger, K., & Flauger, B. (2011). Olfactory recognition of individual competitors by means of faeces in horse (*Equus caballus)*. Animal Cognition, 14(2), 245-257. doi:10.1007/s10071-010-0358-1

Lanyon, C. V., Rushton, S. P., O'donnell, A. G., Goodfellow, M., Ward, A. C., Petrie, M., . . . Penn, D. J. (2007). Murine scent mark microbial communities are genetically determined. FEMS Microbiology Ecology, 59(3), 576-583. doi: /10.1111/j.1574-6941.2006.00252.x

Ma, W., & Klemm, W. (1997). Variations of equine urinary volatile compounds during the oestrous cycle. Veterinary Research Communications, 21(6), 437-446. doi:10.1023/A:1005859419574

Marty, J. S., Higham, J. P., Gadsby, E. L., & Ross, C. (2009). Dominance, coloration, and social and sexual behavior in male drills, *Mandrillus leucophaeus*. International Journal of Primatology, 30(6), 807-823. doi:10.1007/s10764-009-9382-x

McDonnell, S. (2003). Practical field guide to horse behavior: The equid ethogram. Journal of Equine Veterinary Science, 23(1), A1. doi:10.1016/s0737-0806(03)70087-2

McLeman, M. A., Mendl, M., Jones, R. B., White, R., & Wathes, C. M. (2005). Discrimination of conspecifics by juvenile domestic pigs, *Sus scrofa*. Animal Behaviour, 70(2), 451-461. doi.org/10.1016/j.anbehav.2004.11.013

Mendl, M., Randle, K., & Pope, S. (2002). Young female pigs can discriminate individual differences in odours from conspecific urine. Animal Behaviour, 64(1), 97-101. doi: 10.1006/anbe.2002.3040

Milinski, M., Griffiths, S., Wegner, K. M., Reusch, T. B., Haas-Assenbaum, A., & Boehm, T. (2005). Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. Proceedings of the National Academy of Sciences of the United States of America, 102(12), 4414-4418. doi: 10.1073/pnas.0408264102

Mozūraitis, R., Būda, V., Kutra, J., & Borg-Karlson, A. (2012). p-and m-cresols emitted from estrous urine are reliable volatile chemical markers of ovulation in mares. Animal Reproduction Science, 130(1), 51-56. doi:10.1016/j.anireprosci.2011.12.008

Novotny, M. V., Soini, H. A., Koyama, S., Wiesler, D., Bruce, K. E., & Penn, D. J. (2007). Chemical identification of MHC-influenced volatile compounds in mouse

urine. I: Quantitative proportions of major chemosignals. Journal of Chemical Ecology, 33(2), 417-434. doi:10.1007/s10886-006-9230-9

Olsén, K. H., Grahn, M., & Lohm, J. (2002). Influence of MHC on sibling discrimination in arctic char, *Salvelinus alpinus* (L.). Journal of Chemical Ecology, 28(4), 783-795. doi:10.1023/A:1015240810676

Olsson, M., Madsen, T., Nordby, J., Wapstra, E., Ujvari, B., & Wittsell, H. (2003). Major histocompatibility complex and mate choice in sand lizards. Proceedings. Biological Sciences / the Royal Society, 270 Suppl 2, S254-6. doi:10.1098/rsbl.2003.0079

Palagi, E., & Dapporto, L. (2006). Beyond odor discrimination: Demonstrating individual recognition by scent in *Lemur catta*. Chemical Senses, 31(5), 437-443. doi:10.1093/chemse/bjj048

Pence, H. E., & Williams, A. (2010). ChemSpider: An online chemical information resource. Journal of Chemical Education, 87(11), 1123-1124. doi:10.1021/ed100697w

Penn, D. J., Oberzaucher, E., Grammer, K., Fischer, G., Soini, H. A., Wiesler, D., . . . Brereton, R. G. (2007). Individual and gender fingerprints in human body odour. Journal of the Royal Society, Interface, 4(13), 331-340. doi:10.1098/rsif.2006.0182

Rajakaruna, R., Brown, J., Kaukinen, K., & Miller, K. (2006). Major histocompatibility complex and kin discrimination in Atlantic salmon and brook trout. Molecular Ecology, 15(14), 4569-4575. doi:10.1111/j.1365-294x.2006.03113.x

Ransom, J. I., & Cade, B. S. (2009). Quantifying equid behavior--A research ethogram for free-roaming feral horses. Publications of the US Geological Survey, p 26.

Rubenstein, D. I., & Hack, M. A. (1992). Horse signals: The sounds and scents of fury. Evolutionary Ecology, 6(3), 254-260. doi:10.1007/bf02214165

Saslow, C. A. (2002). Understanding the perceptual world of horses. Applied Animal Behaviour Science, 78(2), 209-224. doi:10.1016/s0168-1591(02)00092-8

Schwagmeyer, P. (1988). Scramble-competition polygyny in an asocial mammal: Male mobility and mating success. The American Naturalist, 131(6), 885-892. doi:10.1086/284828

Setchell, J. M., Vaglio, S., Abbott, K. M., Moggi-Cecchi, J., Boscaro, F., Pieraccini, G., & Knapp, L. A. (2011). Odour signals major histocompatibility complex genotype in an old world monkey. Proceedings. Biological Sciences / the Royal Society, 278(1703), 274-280. doi:10.1098/rspb.2010.0571

Setchell, J. M., Vaglio, S., Moggi-Cecchi, J., Boscaro, F., Calamai, L., & Knapp, L. A. (2010). Chemical composition of scent-gland secretions in an old world monkey (*Mandrillus sphinx*): Influence of sex, male status, and individual identity. Chemical Senses, 35(3), 205-220. doi:10.1093/chemse/bjp105

Shirasu, M., & Touhara, K. (2011). The scent of disease: Volatile organic compounds of the human body related to disease and disorder. Journal of Biochemistry, 150(3), 257-266. doi:10.1093/jb/mvr090

Singer, A. G., Beauchamp, G. K., & Yamazaki, K. (1997). Volatile signals of the major histocompatibility complex in male mouse urine. Proceedings of the National Academy of Sciences of the United States of America, 94(6), 2210-2214. doi:10.1073/pnas.94.6.2210

Singh, P. B., Brown, R. E., & Roser, B. (1987). MHC antigens in urine as olfactory recognition cues. Nature, 327(6118), 161–164. doi:10.1038/327161a0

Smith, T. E., Abbott, D. H., Tomlinson, A. J., & Mlotkiewicz, J. A. (1997). Differential display of investigative behavior permits discrimination of scent signatures from familiar and unfamiliar socially dominant female marmoset monkeys (*Callithrix jacchus*). Journal of Chemical Ecology, 23(11), 2523-2546. doi:10.1023/b:joec.0000006664.38169.e5

Stahlbaum, C. C., & Houpt, K. A. (1989). The role of the flehmen response in the behavioral repertoire of the stallion. Physiology & Behavior, 45(6), 1207-1214. doi:10.1016/0031-9384(89)90111-x

Strandh, M., Westerdahl, H., Pontarp, M., Canback, B., Dubois, M. P., Miquel, C., . . . Bonadonna, F. (2012). Major histocompatibility complex class II compatibility, but not class I, predicts mate choice in a bird with highly developed olfaction. Proceedings. Biological Sciences / the Royal Society, 279(1746), 4457-4463. doi:10.1098/rspb.2012.1562

Tang-Martinez, Z. (2001). The mechanisms of kin discrimination and the evolution of kin recognition in vertebrates: A critical re-evaluation. Behavioural Processes, 53(1-2), 21-40. doi:10.1016/s0376-6357(00)00148-0

Thomas, L., & Parker, T. (1974). The lives of a cell. Viking Books New York.

Todrank, J., Heth, G., & Johnston, R. E. (1998). Kin recognition in golden hamsters: Evidence for kinship odours. Animal Behaviour, 55(2), 377-386. doi:10.1006/anbe.1997.0611

Wedekind, C., & Penn, D. (2000). MHC genes, body odours, and odour preferences. Nephrology Dialysis Transplantation, 15(9), 1269-1271. doi:10.1093/ndt/15.9.1269

Wilson, E. O. (1970). Chemical communication within animal species. Chemical Ecology, 9, 133-155. doi:10.1016/b978-0-12-654750-4.50013-x

Wyatt, T. D. (2003). Pheromones and animal behaviour: Communication by smell and taste. Cambridge University Press. doi:10.1017/cbo9780511615061

Yamazaki, K., Beauchamp, G. K., Kupniewski, D., Bard, J., Thomas, L., & Boyse, E. A. (1988). Familial imprinting determines H-2 selective mating preferences. Science (New York, N.Y.), 240(4857), 1331-1332. doi:10.1126/science.3375818

Yamazaki, K., Boyse, E. A., Mike, V., Thaler, H. T., Mathieson, B. J., Abbott, J., . . . Thomas, L. (1976). Control of mating preferences in mice by genes in the major histocompatibility complex. The Journal of Experimental Medicine, 144(5), 1324-1335. doi:10.1084/jem.144.5.1324

Zelano, B., & Edwards, S. V. (2002). An MHC component to kin recognition and mate choice in birds: Predictions, progress, and prospects. The American Naturalist, 160(S6), S225-S237. doi:10.1086/342897

CHAPTER 6

6. THE CORRELATION BETWEEN GENETICS, BODY VOLATILES AND RELATEDNESS IN *EQUUS CABALLUS*

This manuscript has been submitted to: Chemical Senses

6.1 Introduction

The evolution of MHC polymorphisms is a result of two selective pressures, pathogen-driven and reproductive selection (Potts & Wakeland, 1993). MHC polymorphisms produced by selection pressures drive the evolution of MHC-dependent disassortative mating preferences producing progeny that are heterozygous at the MHC and thus possess a greater fitness. MHC are ideal candidate gene complexes for understanding individual odor production as these extremely polymorphic genes produce and directly convey odor information on relatedness, compatibility or individuality. This, in turn, influences social communication, kin discrimination, parent –progeny recognition as well as mate preference (Ruff, et al. 2012). Individual recognition using olfaction is an important component of social behavior. Studies using rodents (Yamazaki, et al., 1976), sticklebacks and sand lizards (Olsson, et al., 2003; Reusch, et al. 2001), birds (Bonadonna & Nevitt, 2004), *Lemur catta* (Knapp, et al., 2006; Palagi & Dapporto, 2006), and old world monkeys (Setchell, et al., 2011) have demonstrated the role of major histocompatibility complex (MHC) in modulating odor profiles used in social recognition. MHC-mediated body odors not only deter inbreeding between closely

related individuals but also provide information about the genetic compatibility between mates.

MHC-dependent mate choice may be based upon three processes involving different aspects of MHC genes: (a) advantage of particular alleles, (b) diversity or heterozygosity of MHC alleles, and (c) genetic compatibility (similarity/dissimilarity) between partners (Piertney & Oliver, 2006). Although a detailed pathway of MHC-associated odor production is still unclear, different hypotheses have been suggested to explain how MHC genes influence an individual's 'body odor' or odor profile. The MHC molecule hypothesis suggests fragments of MHC molecules in biological fluids provide the Volatile Organic Compounds (VOCs) (Ferstl, et al., 1998; Singh, et al., 1987). The peptide hypothesis proposes MHC molecules may alter the peptides found in urine and those metabolites provide the VOCs (Singer, et al., 1997; Yamaguchi, et al., 1981). The microflora hypothesis suggests MHC genes may alter odor by shaping populations of commensal microbes (Singh, et al., 1990). The carrier hypothesis proposes MHC molecules could be altered to carry volatile aromatics (Pearse-Pratt, et al., 1992). Lastly, the widely advocated peptide-microflora hypothesis suggests that MHC molecules alter odor by changing the peptides that are available to commensal microbes (Penn & Potts, 1998; Wedekind & Penn, 2000) and they produce the VOCs as part of their metabolism.

In order to establish a link between MHC and body odor composition, captive populations of lemurs (Setchell, et al., 2011), black-legged kittiwakes (Leclaire, et al., 2012) and Antarctic fur seals (*Arctocephalus gazella*)(Stoffel, et al., 2015) have been

studied. Leclaire et al., were the first to demonstrate a link between odor and genetics in birds and the existence of odor-based mechanisms of mate choice (Leclaire, et al., 2012). Expanding on this evidence, a study on Antarctic fur seal established factors such as colony membership, mother–offspring relation, an individual's multi-locus heterozygosity, and genetic relatedness were all chemically encoded.

To date no study has investigated MHC genotypes and relatedness using odor profiles in horses. Horses are social equids where the social and breeding unit is the band, a stable association of mares, their offspring and one or more stallions that defend the mare group (Klingel, 1975). In domestic and feral horses, scent marking is a prominent feature of horse behavior. Olfaction is also observed in association with greeting between horses (Kiley-Worthington, 1997). Horses periodically test the air by sniffing as they approach one another. During mutual grooming horses approach one another and touch nostrils, and smell each other. The most extensive use of olfaction is observed among stallions in harem groups where olfaction is used periodically to sniff the air in an attempt to locate other harems as well as sniff out mares in estrous (Kimura, 2001). Furthermore, mares and foals immediately learn each other's scents and can find each other quickly in a herd. The frequent sniffing of other horses and excretions occurs with such regularity as to suggest a dependence on olfaction for individual recognition and subsequently, the need of a highly developed olfactory system (Kiley-Worthington, 1997). Studies have not demonstrated association between relationships observed based on VOC profiles and MHC genotypes. This study aimed to integrate genetic and chemical data to investigate the relationship between MHC genotype, relatedness and odor profiles in domestic horses

within two breeds. The study examined the hypotheses that MHC genotype do not influence odor signatures, and that odor similarity does not reflect genetic similarity in horses.

6.2. Methodology

6.2.1. Sample collection

Replicate hair samples (n = 3/horse) were donated by the horse owners by plucking the manes of 23 domestic horses, 11 horses with known kinship and 12 unrelated domestic horses (Table 18). The horses belonged to two breeds, Appaloosa (n = 6) and Quarter Horses (n = 17), with the Appaloosa breed being selectively outcrossed, historically, with Quarter Horses to produce the Appaloosa phenotype. Mane hair samples (each ≈10 mg) were collected from each horse using sterile disposable gloves.

Table 18. Information on horses included in the study. Relations are stated for individuals where the dam and sire was known.

Individual	Dam	Sire	Related horses	Breed
H1	Dam 5	Sire 1	H4, H5	Appaloosa
H2	Dam 4	Sire 6	H3	
H3	Dam 4	Sire 7	H2	
H4	Dam 9	Sire 1	H1, H5	
H5	Dam 10	Sire 1	H1, H4	
H6	Dam 11	Sire 11	Unknown	
H7	Dam 12	Sire 8	H8, H9, H10	Quarter Horse
H8	Dam 1	Sire 2	H7, H9, H10	
H9	Dam 1	Sire 3	H7, H8, H10, H12	
H10	Dam 1	Sire 4	H7, H8, H9	
H11	Dam 6	Sire 9	H12	
H12	Dam 2	Sire 3	H11	
H13	Dam 7	Sire 10	Unknown	
H14	Dam 3	Sire 2	Unknown	
H15	Dam 19	Sire 18	H16, H17	
H16	Dam 8	Sire 5	H15, H17	
H17	Dam 8	Sire 5	H15, H16	
H18	Dam 17	Sire 16	Unknown	
H19	Dam 18	Sire 17	Unknown	
H20	Dam 13	Sire 12	Unknown	
H21	Dam 14	Sire 13	Unknown	
H22	Dam 15	Sire 14	Unknown	
H23	Dam 16	Sire 15	Unknown	

6.2.2. VOC Analysis using Solid-phase microextraction (SPME)-GC-MS

The hair samples were immediately placed into 10-mL glass clear vials with septa (Supelco, Bellefonte, PA) and sealed. Prior to Solid-phase microextraction (SPME) extraction, a fiber was inserted in to the vial and samples were equilibrated for 24 h. The usages of specific horse fly sprays, feed, bathing routines were recorded for all samples. Horses were not bathed before sampling and use of fly spray was avoided with the exception of Horse H5 who suffered from insect bite hypersensitivity and was under treatment. His body was sprayed with water before sampling since fly spray had been used one hour prior to hair collection. Collection of background VOC for each sampling

time was carried out by exposure of empty 10-mL vials (Supelco) to the ambient stable air and those compound were subtracted from each VOC profile before analyses.

Divinylbenzene/Carboxen on Polydimethylsiloxane (CAR/DVB on PDMS) 50/30 μm fibers (Supelco) were used to extract the VOCs from the headspace of the vials containing the hair samples. Fiber exposure was conducted at room temperature for 12 hours. The samples were separated and analyzed by GC-MS using an Agilent 6890 GC with a 5973 MS. A HP5-MS column (0.25 mm × 30 m × 0.25 μm thickness), with helium as the carrier gas at flow rate of 1.0 mL/min for the separation of the analytes was used. The extracted VOCs were desorbed in the injection port at 250°C for 10 min in splitless mode. The temperature program was: an initial oven temperature of 40°C for 5 min, 10°C per minute ramp to a final temperature of 250°C, followed by a final hold for 2 min for a total run time of 32 min. The mass spectrometer used was an HP 5973 MSD with a quadrupole analyzer in full scan mode (mass range: 50–550). The compounds were identified using the NIST 98 mass spectral library. The criterion for the identification of compounds was based on the quality of the detected peak, which was set at greater than or equal to 40% and the presence in all three replicates. Environmental blanks were assessed following the same GC-MS parameters employed for the hair samples.

6.2.3. DNA Extraction

Genomic DNA was extracted from hair samples using the hair protocol provided by QIAamp® DNA Mini Kit (Qiagen, Germantown, MD). Five to twelve hair strands with bulbs were used. DNA was quantified using the Qubit® dsDNA HS assay kit on the

Qubit® 2.0 Fluorometer (Life Technologies, Grand Island, NY) and the average DNA yield was 10-30 ng.

6.2.4. Microsatellite amplification and fragment analysis

DNA was amplified using four MHC Class II microsatellites as Class II loci are thought to be most closely associated with odor production. These included COR113, COR112, COR114, and UM011 (Brinkmeyer-Langford, et al., 2013; Tseng, et al., 2010). Fragment separation and analysis were achieved using capillary electrophoresis on an ABI Prism 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA) and GeneMapper ® Software Version 4.0 (Applied Biosystems, Foster City, CA).

6.2.5. Statistical Analysis

Identification of Compounds - Similarity Percentage analysis (SIMPER) analysis in Primer-E ver 7software (Clarke & Gorley, 2015) was used to reduce the 187 total VOC identified to a smaller number where a subset of 47 compounds were identified and used for analysis. Multivariate analysis - Overall patterns of the chemical similarity using the selected 47 compounds corresponding to the variance between the 23 horses were visualized using nonmetric multidimensional scaling (nMDS) based on pairwise Bray–Curtis similarity. Relative ratios of these 47 VOCs were log(x+1) transformed prior to analysis.

Analysis of MHC genotype and chemical compounds

1. A Spearman correlation was carried out in R statistical software (R Development Core Team, 2013) to check for correlation between individual MHC loci and individual VOCs.

Correlation between loci and compounds with *p values <0.05* were selected with corresponding negative or positive correlation r-values (Table 19).

2. Mantel's r test with a two-tailed *P value* and 1000 permutations were carried out to determine association between genetic distance between individuals based on four MHC II loci and (i) odor profile similarity based on 187 VOCs and (ii) odor profile similarity based on a subset of 47 VOCs identified through SIMPER analysis (Table 20). For the Mantel test, a relatedness matrix using Queller and Goodnight estimator (Queller & Goodnight, 1989) was generated in GenAlEx (Peakall & Smouse, 2006; Peakall & Smouse, 2012) based on genotypes to observed VOCs patterns in relationships. A F_{ST} genetic distance matrix was generated in GENEPOP (Raymond & Rousset, 1995) to observe genetic differentiation between individuals based on the MHC loci and was carried out for both relatedness and F_{ST} distance matrices. A Bray Curtis similarity matrix using VOC relative ratios was generated using vegan package in R statistical software (Oksanen et al., 2007; (R Development Core Team, 2013).

3. In order to visualize patterns of variation in the VOC matrix that can be explained by the distance matrix based on genotypes, a Constrained Analysis of Principal Coordinates (CAP) was carried out in R statistical software. Since each locus is diploid, two columns represented each locus were therefore treated as eight vectors. CAP allows the use of non Euclidean dissimilarity indices like Bray Curtis to be used for comparison.

6.3. Results

6.3.1. Multivariate analysis

The nMDS plot (Figure 17) based on 47 VOCs indicates that VOC profiles are distinct for each individual horse and can therefore be used to individualize them. Although the profiles are distinct, VOCs between related horses as well as some breed specific patterns were detected. Grouping was evident with the percentage similarity analysis overlay on the nMDS. The 40% similarity grouping was observed for related horses while only 20% similarity was seen between horse breeds.

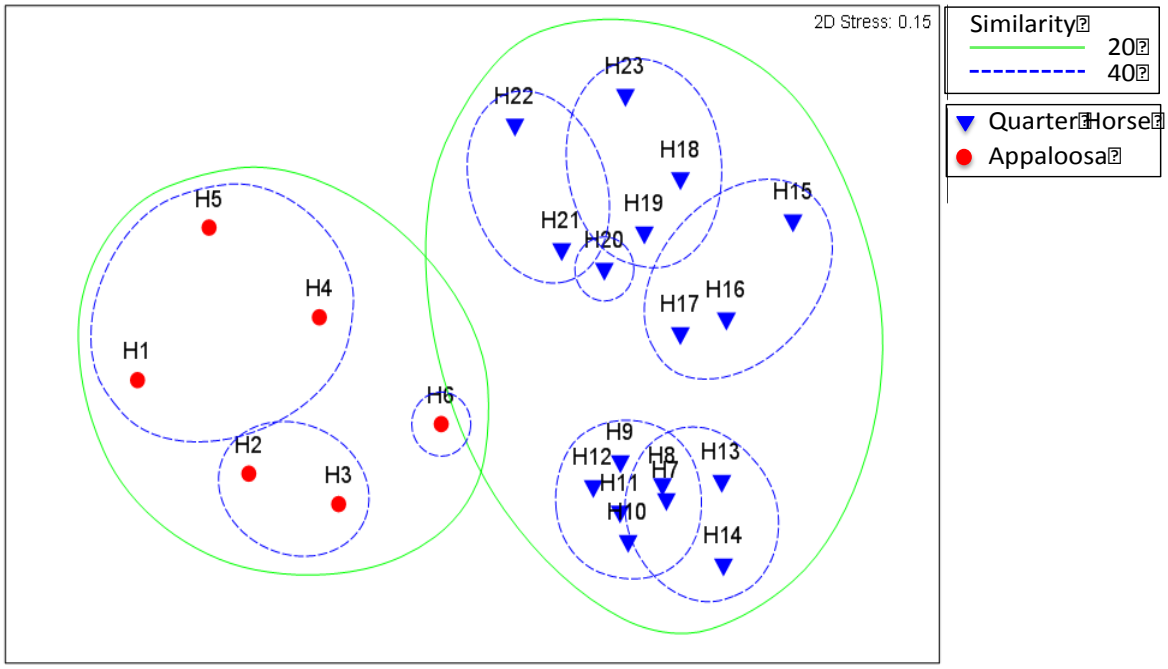

Figure 17. The nMDS plot based on subset of 47 VOC, with similarity analysis overlaid. The plot showed individual horses grouped according to relationships and as well as with horses of the same breed. Blue line representing 40% similarity showed group of related horses. Green line representing 20% similarity showed clustering of horses belonging to the same breeds where blue inverted triangle are Quarter Horses and red circles are Appaloosa horses. Known relations between the horses are given in Table 1.

6.3.2. Combined analysis of genotype and chemical compounds

The Spearman correlation showed significant correlation ($p<0.05$) for another subset of 48 compounds with the four loci (Table 19). The most positive correlations were observed at UM011 while COR 112 had the least compounds with significant correlations. No one compound showed a correlation with all the four loci. However, 3-methyl-1-pentanol was the only compound that showed a positive correlation (r=0.5, $p<0.05$) to COR113, COR114 and UM011.

Table 19. Number of compounds showing significant positive and negative correlation with MHC loci (p<0.05).

Locus	Number of compounds with significant correlations $p<0.05$	Positive correlation	Negative correlations
COR 112	7	4	3
COR 113	12	5	7
COR 114	19	15	4
UM011	29	24	5

Correlations between 187 compounds and the selected subset of 47 compounds with F_{ST} genetic distance and relatedness were observed using the Mantel's r test (Table 20). Chemical similarity based on the total 187 compounds showed a positive non-significantly correlation with genetic relatedness and genetic distance between individuals. However, when the number of compounds was reduced to a subset of 47 compounds, a significant positive correlation was observed at both relatedness and genetic distance.

Table 20. Mantel's r for relatedness and F_{ST} genetic distance for total of 187 compounds as well as subsets identified using SIMPER.

	Relatedness matrix			F_{ST} genetic distance		
Number of Compounds	R value	*P value*	Correlation	R value	*P value*	Correlation
187 (Total)	0.17	0.06	Not significant	0.08	0.61	Not significant
47 (subset)	0.26	*0.02*	Significant	0.34	*0.04*	Significant

The CAP ordination plot in Figure 18 shows the variation in VOCs that can be explained by the four loci. The p value was not significant (*p=0.10*). However, the plot is used as a visualization tool of the variance seen in the VOC described by the genotypes and their influence in discriminating breeds. The colored points are each individual (VOC profiles) displayed in CAP space and the blue vectors show how loci fall along that CAP space. Since two vectors represent each locus, influence of single locus can also be observed. The longest vectors along each CAP axis are most important in explaining variation in VOC profiles along that axis.

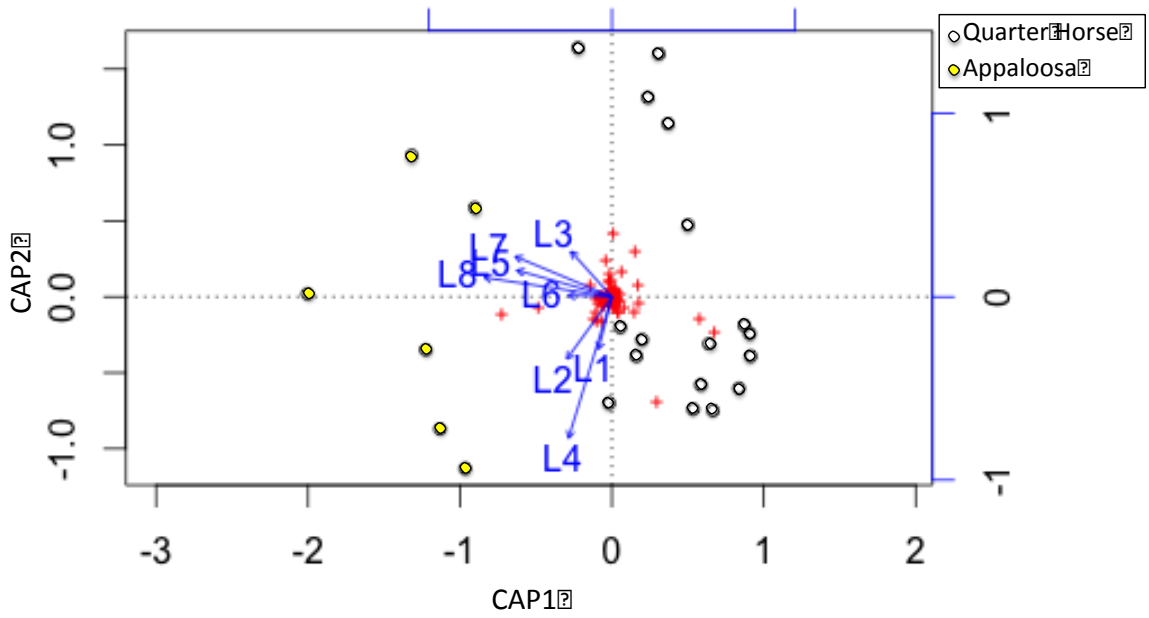

Figure 18. CAP ordination plot showing separation of horses based on breed.
Plot shows VOC variance that is influenced by four loci. Each locus is represented as two vectors considering these are diploid loci. The loci are interpreted as L1 and L2, (COR113), L3 and L4 (COR112), L5 and L6 (COR114), and L7 and L8 (UM011). The plot demonstrates the influence of genotypes on the variation between breeds.

6.4. Discussion

Olfactory cues play a fundamental role in mammalian social interactions ranging from rodents (Cheetham, et al., 2007; Gheusi, et al., 1997) to primates (Setchell, et al., 2010; Smith et al., 2001). These social interactions are perceived to be based upon the similarity in odors among related individuals, the perception of such similarity, and the behavioral response to this perception. Numerous studies have attempted to determine genetic markers influencing odor. The current study was able to establish a correlation between four widely studied equine MHC Class II loci and individual VOC profiles in domestic horses of two breeds.

A total of 187 VOCs from the domestic horses were identified. Many of the compounds recognized were volatile hydrocarbons, which have also been identified in odor profiles for primates (Knapp, et al., 2006; Setchell, et al., 2010; Smith et al., 2001). Although "chemical fingerprints" are widely thought to be predetermined by genotype, only a few studies have reported any association to relatedness (Boulet, et al., 2009), breed similarity (Chapter 5) and individual heterozygosity (Charpentier, et al., 2008). Congenic strains of mice and rats that differ only in alleles at the MHC have been used in several different studies to examine whether MHC genotypes are associated with specific odor signatures (Yamazaki, et al., 1994; Yamazaki, et al., 1990). These studies demonstrated that mice trained to distinguish the urine of MHC congenic mice, could as readily distinguish the urine of germ free mice, suggesting that the MHC genotype was a major factor in individual discrimination and these genotypes were hence, associated with specific odor cues (Yamazaki, et al., 1994; Yamazaki, et al., 1990). Likewise, correlations seen in this study provide additional evidence that an individual's odor profile could be influenced to some extent by the variation in their MHC genotype. However, analyses based on the all the chemical compounds identified in a study may mask factors that could influence specific odor compounds (Stoffel, et al., 2015). Factors such as sex (Doty & Cameron, 2009; Setchell, et al., 2010; Smith, et al., 2001), the degree of heterozygosity across loci (Charpentier, et al., 2010; Hedrick, 1992), kinship (Boulet, et al., 2009; Hurst & Beynon, 2010), age (Caspers, et al., 2011) and breeding colonies (Stoffel, et al., 2015) are often considered.

The Stoffel et al study on Antarctic fur seals focused primarily on relatedness with an inter-dependence of both chemical and genetic data on mother-offspring

relations. When all of the animals were analyzed without controlling for this factor they obtained a highly significant relation between VOCs and the relatedness matrix (Mantel's r = 0.07 n = 82, $p = 0.005$) and suggested that the VOCs in odor profiles were crucial in mother–offspring communication, kin recognition, and mate choice in Antarctic fur seals (Stoffel, et al., 2015). Another study on old world monkeys demonstrated that the mandrill odor profiles indicate MHC genotype as well as information concerning sex and male rankings (Setchell, et al., 2011). They were, however, unable to show possible discrimination of individuals possessing particular MHC genotypes based on their odor profiles. However, similar to the study on Antarctic seals, they found odor similarity in dyads and in some cases also odor similarity in pedigree relationships. The observed positive relationship between odor dissimilarity and MHC dissimilarity in dyads suggested a mechanism underlying MHC-disassortative reproduction patterns in mandrills (Setchell, et al., 2010).

In the current study, correlation between the four loci and total VOCs showed a number of compounds to be highly associated with a single locus. This implied that, apart from the combination of genotypes affecting odor, alleles at a single locus may also influence the presence/ absence or the abundance of a compound in an individual's odor. Consequently, it can be implied that a combination of compounds influenced by different MHC loci results in complex phenotypic odor profiles that may, in turn, signify MHC genotype to the animals. Interestingly, the positive significant correlation of only a subset of compounds and MHC loci could also indicate that individuals may not utilize all VOCs to identify unfamiliar or related individual but use only a subset of compounds for discrimination. Similar findings were found in Antarctic fur seal populations where

evidence of relatedness between mother-offspring dyads was encoded by a small subset of chemicals. They found that the mother-offspring similarity signal was undetected when the total chemical profile was analyzed (Stoffel, et al., 2015).

The ordination plot based on VOCs clearly grouped related individuals based on their odor profiles suggesting the role of odor in kin recognition. This analysis of domestic horse VOCs and genetic MHC profiles suggest that the quantitative differences in the array of volatiles between closely related horses result in distinct kinship patterns. A similar role in kin recognition has been seen in female mice that choose nest mates that are similar at the MHC loci, which indicated that odor recognition of kin is a surrogate for recognizing the MHC genotypes (Manning, et al., 1992). Although the analysis in the current study is not exhaustive, with inclusion of only four MHC Class II loci, the results suggest that chemical similarity within related individuals appears to be influenced by the MHC genotype pattern. These findings support other studies (Beecher, 1989; Cheetham, et al., 2007; Gheusi, et al., 1997; Kean, et al., 2015; Sherborne, et al., 2007) where odor profiles provided both a signal of individual genetic variability and aided the animals to innately assess genetic similarity for the purpose of inbreeding avoidance.

Chemical communication among individuals of the same species is a versatile and widely used means of social interaction. In horses, it is remarkable that associations beyond immediate relatives could be a major factor affecting harem affiliations (Oom & Cothran, 1994; Sigurjónsdóttir, et al., 2003). It has been observed that kinship affects interactions among mares differently than relationships between the stallion and the mares. Where mares prefer related females for bonding within the harem, stallions choose unrelated mares for their harems, thus avoiding inbreeding. If individuals, with

more similar MHC genotypes, possess similar odor phenotypes, then the horses may be identifying individuals by comparing their odor with self or known relatives (Widdig, 2007). The findings of this study support the hypothesis that MHC genotypes do indeed influence odor profiles and inversely, odor profiles are signatures of MHC similarity or dissimilarity that are identified by horses. In summary, this study indicated significant correlations between genetics and volatile scent signatures of horses, suggesting individual discrimination and kin identification can play a key role in shaping social behavior and communication in equids.

6.5 References

Beecher, M. D. (1989). Signalling systems for individual recognition: An information theory approach. Animal Behaviour, 38(2), 248-261.

Bonadonna, F., & Nevitt, G. A. (2004). Partner-specific odor recognition in an Antarctic seabird. Science (New York, N.Y.), 306(5697), 835.

Boulet, M., Charpentier, M. J., & Drea, C. M. (2009). Decoding an olfactory mechanism of kin recognition and inbreeding avoidance in a primate. BMC Evolutionary Biology, 9, 281-2148-9-281.

Brinkmeyer-Langford, C. L., Cai, J. J., Gill, C. A., & Skow, L. C. (2013). Microsatellite variation in the equine MHC. Animal Genetics, 44(3), 267-275.

Caspers, B. A., Schroeder, F. C., Franke, S., & Voigt, C. C. (2011). Scents of adolescence: The maturation of the olfactory phenotype in a free-ranging mammal. PloS One, 6(6), e21162.

Charpentier, M. J., Boulet, M., & Drea, C. M. (2008). Smelling right: The scent of male lemurs advertises genetic quality and relatedness. Molecular Ecology, 17(14), 3225-3233.

Charpentier, M. J., Crawford, J. C., Boulet, M., & Drea, C. M. (2010). Message 'scent': Lemurs detect the genetic relatedness and quality of conspecifics via olfactory cues. Animal Behaviour, 80(1), 101-108.

Cheetham, S. A., Thom, M. D., Jury, F., Ollier, W. E., Beynon, R. J., & Hurst, J. L. (2007). The genetic basis of individual-recognition signals in the mouse. Current Biology, 17(20), 1771-1777.

Clarke, K. R., & Gorley, R. N. (2015). PRIMER v7: User manual/tutorial PRIMER-E, Plymouth, 296pp.

Doty, R. L., & Cameron, E. L. (2009). Sex differences and reproductive hormone influences on human odor perception. Physiology & Behavior, 97(2), 213-228.

Ferstl, R., Eggert, F., & Muller-Ruchholtz, W. (1998). Major histocompatibility complex-associated odours. Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association, 13(5), 1117-1119.

Gheusi, G., Goodall, G., & Dantzer, R. (1997). Individually distinctive odours represent individual conspecifics in rats. Animal Behaviour, 53(5), 935-944.

Hedrick, P. W. (1992). Female choice and variation in the major histocompatibility complex. Genetics, 132(2), 575-581.

Hurst, J. L., & Beynon, R. J. (2010). Making progress in genetic kin recognition among vertebrates. Journal of Biology, 9(2), 1.

Kean, E. F., Chadwick, E. A., & Müller, C. T. (2015). Scent signals individual identity and country of origin in otters. Mammalian Biology-Zeitschrift Für Säugetierkunde, 80(2), 99-105.

Kiley-Worthington, M. (1997). The behaviour of horses: In relation to management and training. Applied Animal Ethology, 9(1), 99. doi:10.1016/0304-3762(82)90184-5.

Kimura, R. (2001). Volatile substances in feces, urine and urine-marked feces of feral horses. Canadian Journal of Animal Science, 81(3), 411-420.

Klingel, H. (1975). Social organization and reproduction in equids. Journal of Reproduction and Fertility. Supplement, (23), 7-11.

Knapp, L. A., Robson, J., & Waterhouse, J. S. (2006). Olfactory signals and the MHC: A review and a case study in *Lemur catta.* American Journal of Primatology, 68(6), 568-584.

Leclaire, S., Merkling, T., Raynaud, C., Mulard, H., Bessiere, J. M., Lhuillier, E., . . . Danchin, E. (2012). Semiochemical compounds of preen secretion reflect genetic make-up in a seabird species. Proceedings. Biological Sciences / the Royal Society, 279(1731), 1185-1193.

Manning, C. J., Wakeland, E. K., & Potts, W. K. (1992). Communal nesting patterns in mice implicate MHC genes in kin recognition. Nature, 360, 581-583.

Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggests, M. (2007). The vegan package. Community Ecology Package, 10

Olsson, M., Madsen, T., Nordby, J., Wapstra, E., Ujvari, B., & Wittsell, H. (2003). Major histocompatibility complex and mate choice in sand lizards. Proceedings.Biological Sciences / the Royal Society, 270 Suppl 2, S254-6.

Oom, M., & Cothran, E. (1994). The genetic variation of an endangered breed: The Sorraia horse. Animal Genetics, 27(Suppl 2), 35.

Palagi, E., & Dapporto, L. (2006). Beyond odor discrimination: Demonstrating individual recognition by scent in *Lemur catta.* Chemical Senses, 31(5), 437-443.

Peakall, R., & Smouse, P. E. (2006). GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes, 6(1), 288-295.

Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics, 28, 2537-2539.

Pearse-Pratt, R., Schellinck, H., Brown, R., & Roser, B. (1992). Evolutionary and immunological implications of the role of the MHC in olfactory signaling. Chemical Signals in Vertebrates 6 (pp. 167-173) Springer.

Penn, D., & Potts, W. (1998). How do major histocompatibility complex genes influence odor and mating preferences? Advances in Immunology, 69, 411-436.

Piertney, S., & Oliver, M. (2006). The evolutionary ecology of the major histocompatibility complex. Heredity, 96(1), 7-21.

Potts, W. K., & Wakeland, E. K. (1993). Evolution of MHC genetic diversity: A tale of incest, pestilence and sexual preference. Trends in Genetics, 9(12), 408-412.

Queller, D. C., & Goodnight, K. F. (1989). Estimating relatedness using genetic markers. Evolution, 43(2)258-275.

R Development Core Team. (2013). R: A language and environment for statistical computing.

Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. Journal of Heredity, 86(3), 248-249.

Reusch, T. B., HaÈberli, M. A., Aeschlimann, P. B., & Milinski, M. (2001). Female sticklebacks count alleles in a strategy of sexual selection explaining MHC polymorphism. Nature, 414(, 300-302.

Ruff, J. S., Nelson, A. C., Kubinak, J. L., & Potts, W. K. (2012). MHC signaling during social communication. Self and nonself (pp. 290-313) Springer.

Setchell, J., Charpentier, M., Abbott, K., Wickings, E., & Knapp, L. (2010). Opposites attract: MHC-associated mate choice in a polygynous primate. Journal of Evolutionary Biology, 23(1), 136-148.

Setchell, J. M., Vaglio, S., Abbott, K. M., Moggi-Cecchi, J., Boscaro, F., Pieraccini, G., & Knapp, L. A. (2011). Odour signals major histocompatibility complex genotype in an old world monkey. Proceedings. Biological Sciences / the Royal Society, 278(1703), 274-280.

Setchell, J. M., Vaglio, S., Moggi-Cecchi, J., Boscaro, F., Calamai, L., & Knapp, L. A. (2010). Chemical composition of scent-gland secretions in an old world monkey (*Mandrillus sphinx*): Influence of sex, male status, and individual identity. Chemical Senses, 35(3), 205-220.

Sherborne, A. L., Thom, M. D., Paterson, S., Jury, F., Ollier, W. E., Stockley, P., . . . Hurst, J. L. (2007). The genetic basis of inbreeding avoidance in house mice. Current Biology, 17(23), 2061-2066.

Sigurjónsdóttir, H., Van Dierendonck, M. C., Snorrason, S., & Thórhallsdóttir, A. G. (2003). Social relationships in a group of horses without a mature stallion. Behaviour, 140(6), 783-804.

Singer, A. G., Beauchamp, G. K., & Yamazaki, K. (1997). Volatile signals of the major histocompatibility complex in male mouse urine. Proceedings of the National Academy of Sciences of the United States of America, 94(6), 2210-2214.

Singh, P. B., Brown, R. E., & Roser, B. (1987). MHC antigens in urine as olfactory recognition cues.

Singh, P. B., Herbert, J., Roser, B., Arnott, L., Tucker, D. K., & Brown, R. E. (1990). Rearing rats in a germ-free environment eliminates their odors of individuality. Journal of Chemical Ecology, 16(5), 1667-1682.

Smith, T. E., Tomlinson, A. J., Mlotkiewicz, J. A., & Abbott, D. H. (2001). Female marmoset monkeys (*Callithrix jacchus*) can be identified from the chemical composition of their scent marks. Chemical Senses, 26(5), 449-458.

Stoffel, M. A., Caspers, B. A., Forcada, J., Giannakara, A., Baier, M., Eberhart-Phillips, L., . . . Hoffman, J. I. (2015). Chemical fingerprints encode mother-offspring similarity, colony membership, relatedness, and genetic quality in fur seals. Proceedings of the National Academy of Sciences of the United States of America, 112(36), E5005-12.

Tseng, C., Miller, D., Cassano, J., Bailey, E., & Antczak, D. (2010). Identification of equine major histocompatibility complex haplotypes using polymorphic microsatellites. Animal Genetics, 41(s2), 150-153.

Wedekind, C., & Penn, D. (2000). MHC genes, body odours, and odour preferences. Nephrology, Dialysis, Transplantation: Official Publication of the European Dialysis and Transplant Association - European Renal Association, 15(9), 1269-1271.

Widdig, A. (2007). Paternal kin discrimination: The evidence and likely mechanisms. Biological Reviews, 82(2), 319-334.

Yamaguchi, M., Yamazaki, K., Beauchamp, G. K., Bard, J., Thomas, L., & Boyse, E. A. (1981). Distinctive urinary odors governed by the major histocompatibility locus of the mouse. Proceedings of the National Academy of Sciences of the United States of America, 78(9), 5817-5820.

Yamazaki, K., Beauchamp, G. K., Imai, Y., Curran, M., Bard, J., & Boyse, E. A. (1994). Odor types determined by the major histocompatibility complex in mice. Olfaction and Taste XI (p. 233-236) Springer.

Yamazaki, K., Beauchamp, G. K., Imai, Y., Bard, J., Phelan, S. P., Thomas, L., & Boyse, E. A. (1990). Odor types determined by the major histocompatibility complex in germfree mice. Proceedings of the National Academy of Sciences of the United States of America, 87(21), 8413-8416.

Yamazaki, K., Boyse, E. A., Mike, V., Thaler, H. T., Mathieson, B. J., Abbott, J., . . . Thomas, L. (1976). Control of mating preferences in mice by genes in the major histocompatibility complex. The Journal of Experimental Medicine, 144(5), 1324-1335.

CHAPTER 7

7. SUMMARY

Various phenomena are responsible for maintenance of genetic variants being distributed throughout a population or within specific subpopulations. Genetic drift, habitat destruction, and change in population size are few factors that are related to social, behavioral and environmental variables affecting populations. Molecular markers are often used to examine the variables influencing the changing structure within small populations. Often, however, it is nearly impossible to obtain biological samples for performing molecular analysis on wild animals that are endangered or elusive. Non-invasive sampling (i.e., collection of feces, hair) allows the extraction of DNA without harming or causing distress to the animal. Nevertheless, owing to the drawbacks of such sampling method it is necessary to develop techniques that improve the quality and quantity of extracted DNA that can be used for downstream processes. The first aim of this study was to develop a method that would improve the DNA yield from fresh and aged fecal samples. The study was able to demonstrate that pressure cycling technology enhanced DNA yield from fecal samples and extracted DNA could individualize contributor and subsequently be used for kinship analysis. This was a proof of concept study and it would be interesting to use this method for samples in the wild where fecal samples are older than six days since this would be a more likelihood scenario in the wild.

Assessing the genetic diversity using non-invasive sampling methods in the feral horses would shed light on the effect of management strategies on such populations.

Using 14 neutral microsatellites this study showed that the small herd and isolated Big Summit horses have reduced heterozygosity and are trending towards being inbred. Though this study aimed to assess the genetic diversity primarily within the Big Summit herd, obtaining more samples from other HMAs would be a better measure of diversity within subgroups of feral horses. This would also aid in understanding how selection pressures are varied and can impact genetic diversity and subsequent social behavior among subpopulations managed in similar way. Observing the variation in the adaptive MHC loci in geographically separated horse breeds showed low overall differentiation. The information on maintenance of higher MHC diversity but low genetic distance between subpopulations of the same breed could be used in management of breeds irrespective of location. Subsequently, these finding also demonstrate that breed similarity rather than distance may influence how selection pressures effects subpopulations.

Apart from using neutral and adaptive markers to profile individuals and determine relationship in a population this study was successful in showing that odor profiles could also be used to as a discrimination tool. This study demonstrated that VOC obtained from hair of domestic horses were able to individualize horses as well as differentiate between horse breeds and display kinship. Future studies should be carried out with a larger sample size and assessing discrimination where relationships are not known. Furthermore this study was able to show a significant correlation between MHC genotypes and VOC odor profiles in domestic horses with known relationships. Although correlation was observed using only four MHC Class II loci, it would be interesting to incorporate more gene specific loci in future studies. Understanding the relationship

between MHC gene and odor using domestic animals would provide information that can be used in understanding social structure and behavior in wild populations.

APPENDICES

Appendix 1

mtDNA sequence variation with nucleotide substitutions relative to the horse reference sequence GenBank JN398377 and haplogroup assignments (Achilli et al. 2012). A '.' represents the same nucleotide base as the reference sequence, A '—' represents a deletion in the sequence.

Haplotypes	Position of nucleotide substitution																																					Haplogroups	# of Individuals
	15490	15494	15495	15496	15532	15533	15534	15536	15538	15542	15585	15597	16000	16001	16002	16003	16004	16035	16049	16050	16066	16103	16209	16223	16226	16240	16271	16276	16277	16806	16807	16811	16826	16827	16838		
JN 39837	C	T	T	A	C	A	C	T	A	C	G	A	G	T	C	T	G	C	A	A	G	T	C	G	C	G	A	C	T	A	C	C	C	A	A	C	A1	
1	.	.	C	.	.	.	.	.	.	T	A	G	.	.	T	.	.	T	.	G	A	C	.	A	.	.	.	.	.	.	.	.	.	.	.	.	G1	1
2	.	.	C	.	.	.	.	.	.	T	.	G	.	.	T	.	.	T	.	G	A	C	.	A	.	.	.	.	.	.	.	.	.	.	.	.	G1	1
3	.	.	C	.	.	—	.	C	.	.	A	.	A	.	T	.	A	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	A	4
4	T	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	A	.	.	.	.	.	.	.	T	.	G	.	.	B	3
5	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	A	.	.	A	.	.	.	.	.	.	.	.	.	G	.	.	B	2
6	.	.	C	.	.	.	.	.	G	.	A	.	.	.	T	.	.	.	.	G	.	.	T	A	.	.	.	T	.	.	.	.	.	G	.	.	I	14
7	.	.	.	.	.	.	.	.	.	.	A	.	.	C	T	.	.	.	.	.	.	.	.	A	.	.	.	T	.	.	T	.	.	.	G	T	N	3
8	.	.	C	.	.	.	.	.	.	.	.	.	.	C	T	.	.	.	.	.	.	.	.	A	.	.	.	T	.	.	T	.	.	.	G	T	N	10
9	.	C	C	G	.	.	T	.	.	.	.	.	.	.	T	C	.	.	G	.	.	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	L	12
10	.	C	C	G	.	.	T	.	.	.	.	.	.	.	T	C	.	.	G	.	.	.	.	A	.	.	.	T	.	.	.	.	.	.	G	.	L1	1
11	.	C	C	G	.	.	T	.	.	.	A	.	.	.	.	C	.	.	G	.	.	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	L2	3
12	.	C	C	G	—	.	T	.	.	.	A	.	.	.	T	C	A	.	G	.	.	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	L3	2
13	.	C	C	G	—	.	T	.	.	.	A	.	.	.	T	C	A	.	G	.	.	.	.	A	.	.	.	T	C	.	.	.	.	.	.	.	L3	1
14	.	C	C	G	.	.	T	.	.	.	.	.	.	.	T	C	A	.	G	.	.	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	L1	1
15	.	.	C	.	.	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	C	.	A	.	.	G	T	.	G	.	.	T	.	.	.	Q3	2
16	.	.	C	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	A	T	A	G	T	.	.	.	.	T	.	.	.	Q1	1

Appendix 2

Information regarding relatedness and location on horses used in study

Identifier	Sex	Known Dam	Dam	Sire	Related horses	Stable	Breed
H1	Male		Dam 5	Sire 1	H4, H5	Stable 1	Appaloosa
H2	Male		Dam 4	Sire 6	H3	Stable 1	
H3	Female		Dam 4	Sire 7	H2	Stable 1	
H4	Female		Dam 9	Sire 1	H1, H5	Stable 2	
H5	Male		Dam 10	Sire 1	H1, H4	Stable 2	
H6	Male		Dam 11	Sire 11	Unknown	Stable 2	
H7	Female	Dam 1	Dam 12	Sire 8	H8, H9, H10	Stable 1	Quarter Horse
H8	Male		Dam 1	Sire 2	H7, H9, H10	Stable 1	
H9	Female		Dam 1	Sire 3	H7, H8, H10, H12	Stable 1	
H10	Female		Dam 1	Sire 4	H7, H8, H9	Stable 1	
H11	Female	Dam 2	Dam 6	Sire 9	H12	Stable 1	
H12	Male		Dam 2	Sire 3	H11	Stable 1	
H13	Female	Dam 3	Dam 7	Sire 10	Unknown	Stable 1	
H14	Male		Dam 3	Sire 2	Unknown	Stable 1	
H15	Male		Dam 19	Sire 18	H16, H17	Stable 3	
H16	Female		Dam 8	Sire 5	H15, H17	Stable 3	
H17	Female		Dam 8	Sire 5	H15, H16	Stable 3	
H18	Female		Dam 17	Sire 16	Unknown	Stable 4	
H19	Female		Dam 18	Sire 17	Unknown	Stable 4	
H20	Male		Dam 13	Sire 12	Unknown	Stable 2	
H21	Female		Dam 14	Sire 13	Unknown	Stable 5	
H22	Male		Dam 15	Sire 14	Unknown	Stable 5	
H23	Female		Dam 16	Sire 15	Unknown	Stable 5	

Appendix 3

Spearman rank correlation matrix for 23 horses using VOCs

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23
H1																							
H2	0.5																						
H3	0.4	0.7																					
H4	0.7	0.4	0.4																				
H5	0.8	0.4	0.3	0.6																			
H6	0.4	0.5	0.5	0.4	0.4																		
H7	0.1	0.2	0.3	0.2	0.2	0.4																	
H8	0.1	0.3	0.3	0.1	0.1	0.3	0.7																
H9	0.2	0.2	0.3	0.2	0.2	0.4	0.7	0.5															
H10	0.2	0.2	0.3	0.2	0.2	0.4	0.8	0.7	0.7														
H11	0.1	0.2	0.3	0.1	0.1	0.4	0.6	0.4	0.6	0.6													
H12	0.1	0.2	0.4	0.2	0.2	0.4	0.5	0.4	0.8	0.5	0.7												
H13	0.2	0.4	0.4	0.2	0.2	0.3	0.6	0.7	0.6	0.6	0.4	0.5											
H14	0.1	0.2	0.3	0.1	0.1	0.3	0.6	0.8	0.4	0.5	0.3	0.3	0.8										
H15	0.0	0.1	0.2	0.2	0.1	0.3	0.4	0.3	0.3	0.4	0.5	0.4	0.4	0.3									
H16	0.1	0.1	0.2	0.2	0.1	0.3	0.4	0.3	0.4	0.4	0.4	0.5	0.5	0.4	0.7								
H17	0.1	0.2	0.2	0.2	0.2	0.4	0.4	0.4	0.5	0.5	0.4	0.4	0.4	0.3	0.7	0.9							
H18	0.1	0.2	0.3	0.3	0.1	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.3	0.4	0.4	0.4						
H19	0.1	0.1	0.2	0.2	0.1	0.3	0.4	0.4	0.4	0.4	0.3	0.4	0.4	0.3	0.4	0.4	0.4	0.8					
H20	0.0	0.1	0.1	0.1	0.1	0.2	0.4	0.3	0.3	0.4	0.2	0.3	0.3	0.2	0.4	0.4	0.5	0.4	0.4				
H21	0.0	0.1	0.1	0.1	0.1	0.2	0.4	0.3	0.3	0.4	0.3	0.4	0.4	0.3	0.3	0.3	0.4	0.4	0.4	0.4			
H22	0.0	0.0	0.1	0.1	0.0	0.2	0.3	0.2	0.3	0.3	0.2	0.2	0.2	0.2	0.3	0.2	0.3	0.3	0.4	0.3	0.5		
H23	0.0	0.1	0.2	0.2	0.1	0.2	0.3	0.4	0.3	0.3	0.3	0.4	0.4	0.3	0.4	0.4	0.4	0.5	0.5	0.3	0.4	0.4	

VITA

KETAKI DESHPANDE

Born, Rajasthan, India

2010 – 2016	PhD in Biology with Forensic track Florida International University Miami, Florida
2010 – 2013	MS Forensic Science Florida International University Miami, FL
2008	Certificate course in Clinical Research Symbiosis Centre of Health Care Maharashtra, India
2005 - 2007	MS Health Sciences (Genetics) Interdisciplinary School of Health Sciences (University of Pune) Maharashtra, India
2002 - 2005	Bachelor of Science (Zoology) Fergusson College, University of Pune Maharashtra, India

PUBLICATIONS & PRESENTATIONS

Deshpande, K., Villarreal, M., & Mills, D. K. (2016). Improved DNA profiles from aged horse feces using pressure cycling technology. Conservation Genetics Resources, 1-9.

Deshpande K., Perez E., Leyva N., Suarez M., Mills D., (2016). Genetic structure within the wild horse populations inhabiting Oregon's Herd Management Areas (Submitted to Journal of Heredity)

Deshpande K., Furton K., Mills D., (2016). The equine volatilome is unique but can also reflect kinship: a possible mechanism for inbreeding avoidance (Submitted to Behavioral Ecology)

Deshpande K., Mills D., (2016). The correlation between genetics, body volatiles and relatedness in *Equus caballus* (Submitted to Chemical Senses)

Deshpande K., Mills D., (2016). Genetic variation at seven Major Histocompatibility Complex microsatellites in three horse populations (In prep Animal Genetics)

Deshpande K., Oswald M., Mills D., 2016 “Application of a 6-Plex Microsatellite Kit in the Analysis of Aged Fecal DNA Samples: Prospective Use in Equine Slaughter Forensic Cases” (Poster Presentation) American Academy of Forensic Sciences (AAFS) 68th annual scientific meeting, Las Vegas, Nevada

Deshpande K., Mills D., 2015 “Major Histocompatibility Complex As Additional Markers In Wildlife Conservation and Forensics” (Oral presentation) 8th International Veterinary Forensic Science association (IVFSA), Orlando, FL.

Deshpande K., Mills D., 2015 “The correlation between genetics, body volatiles and relatedness” (Oral presentation) Biology Symposium, Miami, FL.

Deshpande K., Mills D., 2014 “Breed designation for unknown equine case samples: a forensic case study” (Poster presentation) Forensic Science Symposium, Miami, FL.

Deshpande K., Mills D., 2014 “Genetic analysis of the “wild “ horses inhabiting Ochoco National Forest, Oregon (Oral Presentation) Scholarly Forum, Miami, FL.

Deshpande K., Mills D., 2014 “Breed designation for unknown equine case samples: a forensic case study” (Poster Presentation) American Academy of Forensic Sciences (AAFS) 66th annual scientific meeting, Seattle, WA.

Deshpande K., Mills D., 2014 “Breed designation for unknown equine case samples: a forensic case study” (Poster presentation) Annual Biomedical and Comparative Immunology Symposium 16th annual scientific meeting (First prize).

Deshpande K., Mills D., 2013 “Genetic analysis of the “wild “ horses inhabiting Ochoco National Forest, Oregon (Oral Presentation) International Equine Conference, Lexington, Ky.

Deshpande K., Mills D., 2013 “Breed designation for unknown equine case samples: a forensic case study” (Oral Presentation) Florida Annual Meeting and Exposition, Tampa, FL.