GENOMIC STUDIES IN THE DROMEDARY (*CAMELUS DROMEDARIUS*)

A Dissertation

by

FAHAD ABDULLAH S. ALSHANBARI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Terje Raudsepp
Committee Members,	Rodolfo Aramayo
	Clare Gill
	E. Gus Cothran
	Brian W. Davis
	James N. Derr
Head of Department,	Jane Welsh

December 2019

Major Subject: Biomedical Sciences

ABSTRACT

Camelids are an important livestock species in many regions of the world, often used for transportation and the production of wool/fiber, meat, and milk. Furthermore, they are used as exotic animals in zoos and shows, and they are used for riding or racing in many countries around the world. All camelid species are adapted to extreme environments. Dromedary and Bactrian camels are adapted to arid conditions, whereas alpacas and llamas are adapted to high altitude. This makes them unique organisms to study how their genetic make-up is related to their physiological features. Even though there has been an increased scientific interest in camelid genetics and genome analysis in the past decade, genetic studies on different traits and disorders are still limited.

There are only a few reports about the genetics of camelid coat color and production traits and no genetic studies comparing these traits between dromedary populations. As selection has taken place since domestication, signatures of selection for different phenotypes can be traced using whole genome sequencing and bioinformatic tools. This will lead to the identification of genes that are associated with selected traits.

We collected blood samples from 200 camels for DNA isolation and tissues for RNA isolation, cell suspensions and live cell cultures. DNA samples and cell suspensions were used for Sanger sequencing and fluorescent in situ hybridization to identify variants associated with coat color phenotypes and major coat color genes as well as locate these genes in camelid genomes. We preformed a whole genome sequencing of pooled dromedary samples to investigate genetic differentiation in populations between different dromedary breeds.

We identified genetic variants in the dromedary *MC1R* and *ASIP* genes that are associated with white and black coat colors, respectively. We identified a subset of genes under selection in dromedary populations and are associated with different traits, such as coat color, milk, and meat production. The overall objective of this project was to generate genetic and genomic resources for camelids (the dromedary in particular) and to identify genes that are associated with different coat color phenotypes and genes that are under selection in different dromedary populations.

ACKNOWLEDGEMENTS

I would sincerely thank my chair, Dr. Terje Raudsepp, for her patience, guidance, wisdom, and allowing me not only to study my favorite species but also for her strong encouragement to learn and obtain my PhD degree. I would also thank my committee members Dr. Rodolfo Aramayo, Dr. Clare Gill, Dr. E. Gus Cothran, Dr. Brian Davis, and Dr. James Derr for their knowledgeable advice throughout my studies. I would also like to thank my Saudi Arabian Culture mission and Qassim University for their sponsorship to obtain my PhD.

I would like to thank my family for their support and for standing by my side all these years. Thanks to my dad for connecting me with breeders allowing me to sample their camels; this research could not have been done without you. Sorror, thanks for helping to get all materials I need for sampling. Sharaf, thank you so much for helping me sample camels and being patient in the hot, sandy, stormy desert. Fesal, thank you so much for completing the required paperwork and sending all samples to the US. This was a team effort, and without you all, I could not have accomplished this study.

I would also thank all breeders who allowed me to sample their camels. Hadaff, it was such a journey traveling with you to obtain samples. Doug Baum, what a pleasure knowing you, and so many thanks for introducing me to the US camel society and assisting me with sampling your and other's camels. Finally, I would like to thank my lab mates, friends, and Texas A&M and Qassim Universities' staff for sharing this great experience with me. Thank you all for your advice and companionship.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Dr. Terje Raudsepp [chair] and Dr. Rodolfo Aramayo of the Department of Biology, Dr. Clare Gill of the Department of Animal Science, Dr. E. Gus Cothran of the Department of Veterinary Integrative Biosciences, Dr. Brian Davis of the Department of Veterinary Integrative Biosciences and Dr. James Derr of the Department of Veterinary Pathobiology.

Dr. Raudsepp and I initiated and designed the experiments of chapters 2, 3, and 4 at the Department of Veterinary Integrative Biosciences; Mayra N. Mendoza helped with the experiments described in chapters 2 and 3; Caitlin Castaneda conducted TaqMan assays and helped with the statistical analysis described in chapter 3; Prof. Leif Andersson suggested the idea of chapter 4 ; Dr. Susanne Bornelov (Uppsala University and University of Cambridge) and Dr. Brian Davis helped with the analysis described in chapter 4. Chapters 2 and 3 were published in 2019. Dr. Susanne Bornelov (Uppsala University and University of Cambridge) wrote R scripts stated in Appendix B. The rest of the work was conducted independently by the student.

Graduate study was supported by a sponsorship from Qassim University. The research was supported by the Molecular Cytogenetics Laboratory at Texas A&M University.

NOMENCLATURE

BAC	Bacterial artificial chromosome
CNV	Copy number variation
DNA	Deoxyribonucleic acid
SA	Saudi Arabia
SNP	Single nucleotide polymorphism
US	United States
VCF	Variant Call Format

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

The family Camelidae

Evolution and phylogenetics

Camelids belong to the mammalian order Cetartiodactyla. The order consists of large to medium sized mammals (including whales and even-toed ungulates), and is one of the most diverse mammalian groups (Hassanin et al., 2012; Proskuryakova et al., 2017; Zhou, Xu, Yang, Zhou, & Yang, 2011). The order is comprised of four diverse suborders: Tylopoda (camelids), Suina (pigs and peccaries), Ruminantia (cattle, bison, sheep, goat and deer) and Cetacea (hippopotami, dolphins, whales and porpoises) (Gatesy, 2002; Hassanin et al., 2012; Murphy et al., 2005; Zhou et al., 2011; Zurano et al., 2019). Most phylogenetic and phylogenomic studies place Camelidae as one of the most basal families within the order of Cetartiodactyla (Figure 1A) (Agnarsson & May-Collado, 2008; Ayoub, McGowen, Clark, Springer, & Gatesy, 2009; Malcolm A. Ferguson-Smith & Vladimir Trifonov, 2007; Gatesy, 2002; Murphy et al., 2005; Zhou et al., 2011). However, a few recent studies have placed Camelids as the second most basal family after Suidae (Proskuryakova et al., 2017; Zurano et al., 2019). Thus, classification within this order remains elusive (Hassanin et al., 2012; Zhou et al., 2011).

The family Camelidae originated in North America during the Eocene about 40-45 million years ago (MYA) (Kozhamkulova, 1986). The last common ancestor of all extant camelid species is dated to 17.5 MYA (Figure 1B) (Prothero & Foss, 2007). Then, its descendants split into lineages that gave rise to the two modern tribes: around 15.5 MYA, *Procamelus* evolved into the tribe Camelini (the Old World camelids), and approximately 13.5 MYA, *Pleiolama* gave rise to the tribe Lamini (the New World camelids) (Figure 1B) (Burger,

2016; Prothero & Foss, 2007; S David Webb & Meachen, 2004). Approximately 3.3 MYA, Camelini species migrated to Eurasia via the Bering Strait, whereas Lamini species migrated to South America. Camelids became extinct in North America around 10,000 years ago along with most of its megafauna (Franklin, 1982; Ji et al., 2009; Stanley, Kadwell, & Wheeler, 1994; S.D. 1974 Webb, 1974). The one-hump dromedary diverged from the two-hump Bactrian camel approximately 5-8 MYA (Burger, 2016; Wu et al., 2014).

Camelidae family comprises seven extant species: The Old World camels (tribe Camelini) include the one-humped dromedary camel (*Camelus dromedarius*) and both the domestic and wild two-humped Bactrian camels (*Camelus bactrianus* and *Camelus ferus*); the New World camelids (tribe Lamini) include the domesticated alpaca (*Lama pacos*) and llama (*Lama glama*) and wild vicugna (*Vicugna vicugna*) and guanaco (*Lama guanicoe*) (Figure 1 B.) (Burger, 2016; Stanley et al., 1994). It is thought that the wild two-humped Bactrian camel is the most direct descendant of the common ancestor of the tribe Camelini (Peters & Driesch, 1997).

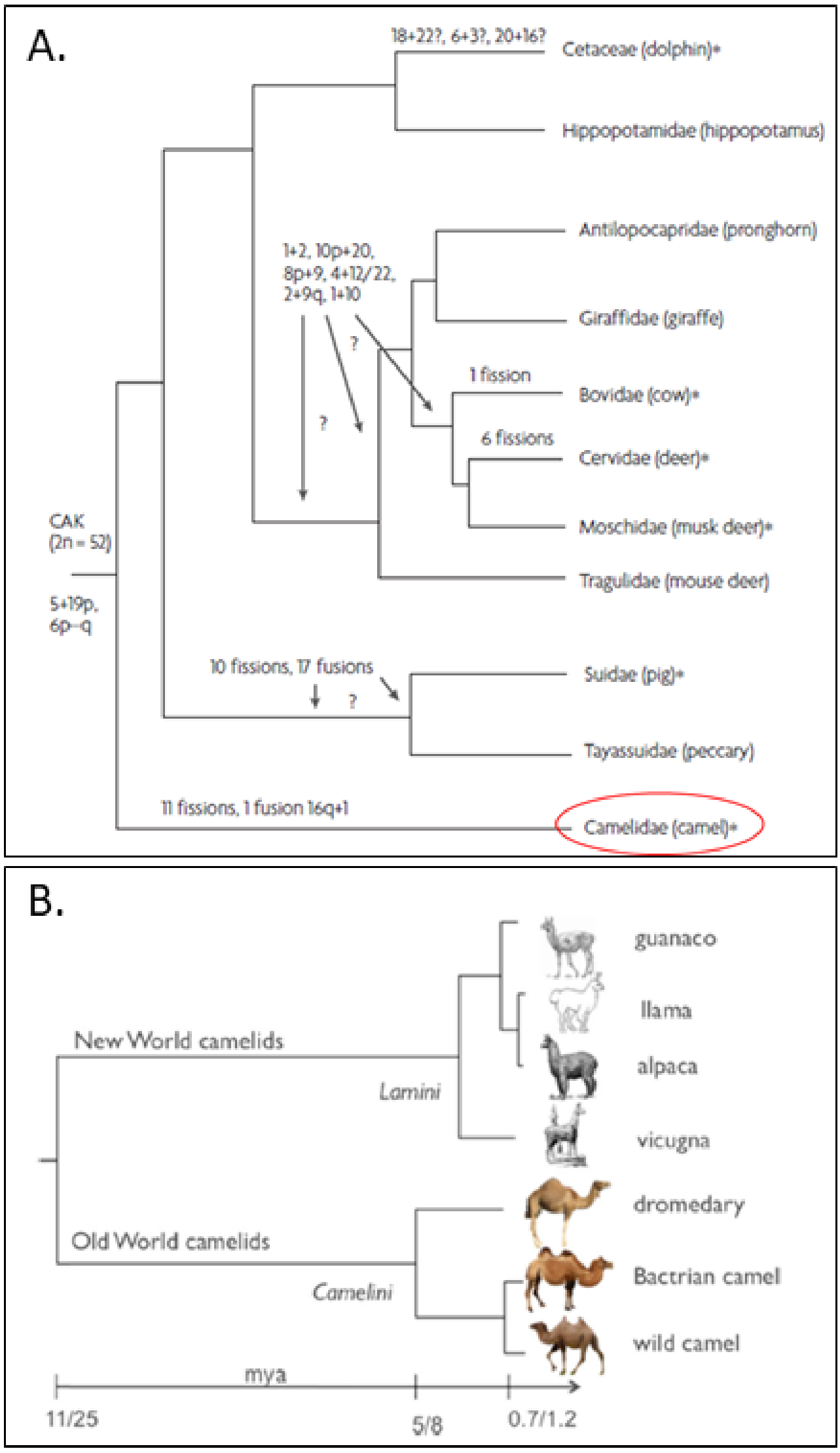

Figure 1. Phylogenetic tree and phylogenetic relationships within the family Camelidae. A. Phylogenetic tree of 11 Cetartiodactyla families is based on chromosomal rearrangements where 6 species indicated by asterisks were studied using human chromosome specific probes. This figure shows that Camelidae (red circle) is the most basal family of Cetartiodactyla (Malcom A. Ferguson-Smith & Vladimir Trifonov, 2007). B. Phylogenetic relationships within the family Camelidae showing divergence times (mya) of the species (Burger, 2016).

The demographic history of Camelini was evaluated by whole genome sequencing of dromedary (n=9), wild Bactrian camel (n=9) and domesticated Bactrian camel (n=7) (Burger, 2016; Wu et al., 2014). The Pairwise Sequentially Markovian Coalescent model (PSMC) (Heng Li & Durbin, 2011) was used to determine the demographic history of the species (Wu et al., 2014). The study showed that the effective population size declined by 70% between 100,000-20,000 years ago (ya). Furthermore, it declined again 4000-5000 ya, perhaps due to hunting by humans followed by domestication, while the wild Bactrian camel is facing extinction due to hunting and the decline of its habitat. Based on zooarchaeological evidence, the wild dromedary got extinct already around 2000 years ago (Grigson, 2014; H.-P. Uerpmann & Uerpmann, 2002; M. Uerpmann & Uerpmann, 2012; Von den Driesch & Obermaier, 2007).

Adaptation and biology

All camelids are uniquely adapted to extreme environments. The dromedaries are adapted to the hot, arid environments in Asia and North Africa, Bactrian camels are adapted to the cold desert conditions of central Asia, and the New World camelids are adapted to high altitudes in the high Andes (Altipano) of South America (Bernard Faye, 2015; Wu et al., 2014). In addition to extreme adaptations, camelids have other unique biological features. For example, they are the only mammals with small and functionally efficient heavy chain-only antibodies (CHAbs) (Cohen, 2018; Flajnik, Deschacht, & Muyldermans, 2011; Griffin et al., 2014). The first report described the CHAbs, also known and named as VHH in camelids, back in 1993 (Hamers-Casterman et al., 1993). The study characterized their size, structure and function, particularly their substantial antigen-binding repertoire (Daley et al., 2010; Hamers-Casterman et al., 1993). Due to this, camels show high resistance to infections that threaten other livestock species in the

same areas, e.g., contagious pleuro-pneumonia, trypanosomiasis, and foot-and-mouth disease (Hemida et al., 2014; Wernery & Kaaden, 2004).

VHH does not only benefit the species by strengthening immunity, but also has important applications for biomedical research. The first step was made by constructing naïve VHH phage (Monegal et al., 2009) and synthetic libraries (Goldman et al., 2006; Yan, Li, Hu, Ou, & Wan, 2014). This allowed researchers to use VHH in clinical trials as a treatment against various diseases (Könning et al., 2017). VHH has been used to successfully target infected HIV stem-cells and block the virus from infecting other cells (Jähnichen et al., 2010; McCoy et al., 2012; Strokappe et al., 2012). Maussang and colleagues contracted VHH against the chemokine receptor CXCR7 from llamas and showed that VHH reduces tumor growth in mice (Maussang et al., 2013). Further, VHH has been used in clinical trials supported by the company Ablynx (www.ablynx.com), producing products with the initial "ALX". For example, ALX-0681 (Caplacizumab) is added to Willebrand Factor-targeting (ulvWF) which is used to assess it suitability for the treatment of acquired thrombotic thrombocytopenic purpura (TTP) (Holz, 2012; Peyvandi et al., 2016). ALX-0171 targets respiratory syncytial virus (RSV) and blocks its replication in *in vitro* experiments (Detalle et al., 2016). ALX-0141 is used for bone-loss related disorders (Chiricozzi, De Simone, Fossati, & Peris, 2019). ALX-0061 increases the affinity of pleiotropic cytokine interleukin-6 receptor (IL-6R) which will target human albumin in rheumatoid arthritis (Van Roy et al., 2015). Thus, VHH has been extensively evaluated for the use in biomedical therapy (Könning et al., 2017).

Camelids, unlike other mammals, have oval-shaped erythrocytes that are small and circulate in much greater numbers. These erythrocytes are hemoglobin-rich and capable of

holding water bonds to amino acids, which led to the proposition that camelid erythrocytes evolved as a natural selection aimed to protect the species against arid conditions (Bogner, Csutora, Cameron, Wheatley, & Miseta, 1998). This morphological shape allows erythrocytes to circulate in smaller veins, which combats the extreme dehydration which both Old-World and New World Camelids face in their respective environments. In addition, the amino acid sequences of camelid erythrocytes are different from those of other mammals, and allows them to resist osmotic lysis, which in turn allows them to drink large amounts of water after extreme dehydration (Bogner et al., 1998). Finally, camelid erythrocytes can expand 240% upon re-hydration, compared to the only 150% in other mammals (C. Cebra, Anderson, Tibary, Van Saun, & Johnson, 2014).

All camelid species have high blood glucose levels, two folds higher than other mammals, as well as lower concentration of blood ketones (Al-Ali, Husayni, & Power, 1988; C. K. Cebra, Tornquist, Jester, & Stelletta, 2004; C. K. Cebra, Tornquist, Van Saun, & Smith, 2001). Due to a weak insulin response, camelids process glucose much slower than other mammals, which suggests that glucose will remain in the blood for a longer period of time. This gives the species a distinct advantage: the ability to cross longer distances with lower food intake in a shorter amount of time.

Camelids have stomach with 3 compartments compared to other ruminants which have 4 compartments (Lechner-Doll et al., 1995). The classical ruminant stomach consists of the rumen, the reticulum, the omasum and abomasum. In camelids, the stomach comprises of a large compartment 1 (C1) which is divided into cranial and caudal by a strong linear muscle. The compartment 2 (C2) is relatively small in size and not separated from C1. Compartment 3(C3) is

originated from C2 and situated to the right side of C1 where HCl production takes a place by the end of C3 called hindstomach.

There are different factors which allow camelids to save water in their bodies during a dehydration state. All camelids share lower food intake compared to other mammals, which explains their adaptation to arid environments with limited food resources (Bouaouda et al., 2014; Dittmann et al., 2014). Furthermore, Old World Camelids have the ability to regulate their body temperatures, which can vary from 34°C to 41°C depending on the time of day. This is advantageous for the water balance during a dehydration state, as the desert temperature varies significantly from early morning to late evening (Schmidt-Nielsen, Crawford, Newsome, Rawson, & Hammel, 1967; Schmidt-Nielsen, Schmidt-Nielsen, Jarnum, & Houpt, 1957). Therefore, they can survive the loss of more than 25% of their total body weight, while a 12% loss of body weight in other mammals is lethal due to blood circulation failure (McKinley, McBurnie, & Mathai, 2001). Osmolality is controlled by two systems: the first controls the intracellular fluid in tissues, and the other controls the extracellular fluid in blood (B. Andersson, Olsson, & Rundgren, 1980). Besides these water controlling features, the Old-World Camelids excrete only one time a day, and their kidneys absorb most of the water, producing very concentrated urine. Additionally, they do not sweat and their wool works as a cooling system. Therefore, shorn camels sweat 60% more than unshorn camels. Finally, the large intestine absorbs most of the water content before evacuating stool (Ali, Baby, & Vijayan, 2019; Davidson, Jaine, & Vannithone, 2014; Schmidt-Nielsen, 1959).

Another feature the Old World camelids have, is storing 45% of the fat in the their humps, whereas 8% of their carcass contains fat (B Faye, Bengoumi, Messad, & Chilliard,

2002). Fat deposition to the hump starts after two months of age (Bernard Faye, Bengoumi, Cleradin, Tabarani, & Chilliard, 2001). One of the most important hormones that plays a role in regulating adipose tissue is leptin (Chilliard, Delavaud, & Bonnet, 2005). Leptin regulates orexigenic peptides such as neuropeptide Y (NPY), galanin, melanin-concentrating hormone (MCH) and orexins, as well as anorexigenic peptides, e. g., thyrotropin-releasing hormone (TRH) and corticotrophin-releasing hormone (CRF) (Ahima & Hileman, 2000; Ingvartsen & Boisclair, 2001). These peptides do not only regulate energy metabolism but also regulate reproductive processes and the immune system (Matarese, 2000; Parent, Lebrethon, Gerard, Vandersmissen, & Bourguignon, 2000). Comparative analysis between dromedary, cattle and water buffalo shows that leptin cDNA is very similar in these species (Bartha, Sayed-Ahmed, & Rudas, 2005). C. Delavaud *et al.* compared feeding and dehydration in the dromedary in two different experiments measuring leptin and fatty acids in the plasma as well as hump sizes (Delavaud, Bengoumi, Faye, Levieux, & Chilliard, 2013). In one experiment, they showed that plasma leptin decreased by 28%, whereas fatty acids significantly increased (~300%) when camels were underfed. However, in another experiment with underfed camels, both leptins (28%) and fatty acids (419%) were significantly increased. They also reported that hump sizes were significantly decreased during a dehydration state (45%), and there was a reduction of body weight (25%). These results show that camel humps are used when there is both low food and water resources.

Reproduction

Another difference that camelids have from other mammals is the physiology of their reproductive system. Females do not have regular estrous cycles, and their ovulation occurs after

mating; however, spontaneous ovulation has been reported in a small number of females (C. Cebra et al., 2014; Eiwishy, 1987; Skidmore, Billah, Binns, Short, & Allen, 1999). The uterus is bicornuate and its left horn is bigger than the right, due to which implantation is more likely to occur in the left horn (elWishy, 1988). Furthermore, they have an endotheliochorial placenta, where the chorion and the epithelium of the uterus are diffused, non-invasive, and microcotylendonary. This means that they have the advantage of combining the structural features of the cow (diffused) and the horse (microcotylendonary) placentas (Aba, 2013; Olivera, Zago, Jones, & Bevilacqua, 2003; Skidmore, Billah, & Allen, 1996).

Despite the significant difference in their body sizes (the Old-World Camelids are much bigger than the New-World Camelids), camelids can crossbreed naturally (within genus) or by artificial insemination (between genera) and produce live hybrid offspring (Fowler & Bravo, 2010; Skidmore et al., 1999). Hybrids between dromedary and Bactrian camels or alpaca and llama commonly produce fertile offspring. On the other hand, hybrids between genera (e. g. dromedary with alpaca or llama) are rare, their fertility is uncertain, and results are difficult to achieve (Fowler & Bravo, 2010; C. Jones, Abd-Elnaeim, Bevilacqua, Oliveira, & Leiser, 2002; C. J. Jones, Skidmore, & Aplin, 2008). Despite of several anecdotal stories, there are no published reports about hybrids between Bactrian camels and alpacas or llamas. J. A. Skidmore and colleagues reported that hybridization between dromedary and guanaco has very low success rate. Out of 50 attempts, they produced only 6 zygotes, of which 4 aborted during embryonic development, and one female was born but lived for only one day, and one lived born who succeeded (Skidmore et al., 1999). Although the diploid chromosome number of all camelids is the same (2n=74), this failure can be explained by genetic differences between the two genera

(Skidmore et al., 1999). Figure 2 shows pictures of all camelid species and their various hybrids. It is worth mentioning that camelid hybrids are larger in size, stronger, and frequently used for wrestling competitions for entertainment in eastern Asia.

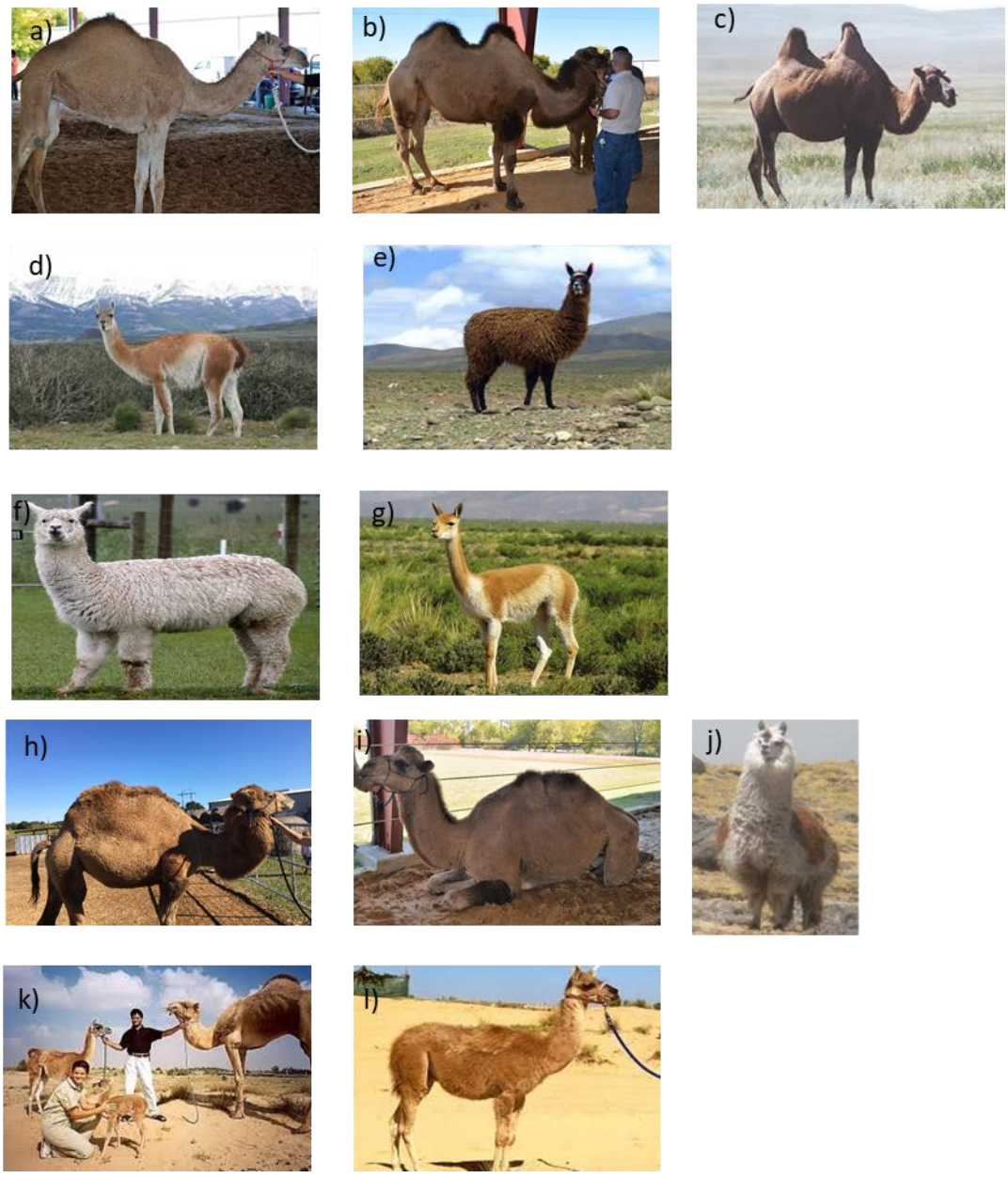

Figure 2. Camelids species: a) Dromedary; b) Domestic Bactrian camel; c) Wild Bactrian camel (Hare, J. 2008); d) Guanaco. (Root-Bernstein, 2014); e) Llama (Stein, V. 2018); f) Alpaca (Hisgett, T. 2016); g) Vicugna. (Arzamendia, Y., Baldo, J. L., & Vilá, B. 2012); h) F1 hybrid of dromedary and Bactrian camel, i) F2 hybrid of dromedary and Bactrian camel (F1 x Bactrian camel), j) Paco - Vicuna (Alpaca x Vicuna hybrid) (Wheeler, J. C. 2012); k) *Rama the Cama* at birth (Guanaco x Dromedary hybrid) (Farag, T. 2003) and l) Adult *Rama the Cama*. (Farag, T. 2003).

Cytogenetics

Cytogenetics is a study of chromosomes, their structure, function, normal chromosome number of a species, karyotypes, chromosomal location of genes, but also chromosomal abnormalities associated with reduced fertility and cancers (Gersen, Keagle, Gersen, & Keagle, 2005; Raudsepp & Chowdhary, 2016). The correct chromosome number of humans (2n=46) was published in 1956 (Ferguson-Smith, 1960; Tjio & Levan, 1956; Tjio & Puck, 1958), while banding protocols were developed late in the 1960s and 1970s, in which clinical cytogenetic studies emerged. The basic principle of cytogenetics relies on understanding the cell cycle in general and cell division during mitosis and miosis in particular. Cytogenetic studies have been extensively reviewed in many domestic species including cattle (Larkin & Farré Belmonte, 2014), pigs (Ducos et al., 2007; Raudsepp & Chowdhary, 2011), horses (Raudsepp et al., 2008), and sheep (Goldammer et al., 2009). On the other hand, fewer reports are available for goats (Schibler, Di Meo, Cribiu, & Iannuzzi, 2009), dogs (Poth, Breuer, Walter, Hecht, & Hermanns, 2010), and cats. Several studies have also investigated camelid cytogenetics.

The first two reports stated that camelids have a diploid karyotype number of 2n = 72 (Capanna, 1965; Hungerford & Snyder, 1966), but it was shortly corrected to 2n=74 (L. Koulischer, Tijskens, & Mortelmans, 1971; Samman, Al-Saleh, & Sheth, 1992; Schmidt-Nielsen et al., 1967; K. Taylor, Hungerford, Snyder, & Ulmer Jr, 1968). Since then, several studies have been conducted studying camelid chromosomes to compare karyotypes of different species and to identify chromosomal aberrations that are associated with genetic disorders (Felipe Avila et al., 2015; T. D. Bunch, 1985; Drew, Meyers-Wallen, Acland, Guyer, & Steinheimer, 1999; Hinrichs, Horin, Buoen, Zhang, & Ruth, 1997; L Koulischer, 1971; Raudsepp, 2014; Raudsepp, Avila, P. Baily, Merriwether, & A. Kutzler, 2015; Raudsepp & Chowdhary, 2016). This is a

difficult undertaking due to the high number of chromosomes (2n=74) and very similar karyotypes among all camelids. More recently, molecular cytogenetic tools were developed based on dromedary chromosome painting probes that were used for Zoo-FISH studies to identify chromosome homologies between dromedary, human, cattle, and pig (Balmus et al., 2007; Kulemzina et al., 2014). Soon after that, clones from the alpaca genome Bacterial Artificial Chromosome (BAC) library CHORI-246 were used for gene mapping and molecular cytogenetic studies across camelids (Alshanbari et al., 2019; Felipe Avila et al., 2015; F. Avila, Baily, et al., 2014; F. Avila, Das, et al., 2014; Mendoza, Raudsepp, Alshanbari, Gutiérrez, & Ponce de León, 2019).

These molecular cytogenetic tools have been used to improve the reference genome assembly for alpacas and dromedary camels by assigning genes and DNA sequences to chromosomes by florescence *in situ* hybridization (FISH) and other genetic methodologies as described below (Alshanbari et al., 2019; Felipe Avila et al., 2015; F. Avila, Baily, et al., 2014; Mendoza et al., 2019; Richardson et al., 2019). It is important to mention that because of evolutionary conservation of camelid genomes, clones from the alpaca BAC library can be mapped to all camelid species (F. Avila, Das, et al., 2014; Martin Plasil et al., 2016).

Domestication

Even though camelids inhabit different geographic locations, they were domesticated for the same purposes: for food (meat and milk), wool production, and as beasts of burden (A. M. Al-Swailem et al., 2010; Kadwell et al., 2001; Skidmore et al., 1999). They have also been used as companion animals, for sports, and dromedaries even for beauty contests. The first domestication of the dromedary was 4000-6000 ya in the Southeast Arabian Peninsula, the

Bactrian camel was domesticated in the central Asian desert 5000-6000 ya, and the domestication of alpacas and llamas took place in the high Andes 6000-7000 ya (Almathen et al., 2016; Ji et al., 2009; Kadwell et al., 2001). Therefore, humans have been using camelids as long as cattle, horses, or dogs (N. Chen et al., 2018; Fages et al., 2019; Librado et al., 2017; Pitt et al., 2018; Wheeler, 1995).

Camelids are well-adapted to harsh conditions and have had remarkable impact on many cultures since domestication. For example, it is thought that the agricultural success of the ancient Inca Empire largely owes to llama dung as the main fertilizer in the poor environment of high Andes (The Guardian 2011).

Alpacas and llamas have been extensively used for the production of fiber. While alpacas have been selected for a variety of color phenotypes, white coat color is of particular interest due to its attractiveness for the textile industry. On the other hand, dromedaries and Bactrian camels have been used for transportation and the production of meat and milk. In the 1800s, dromedaries were transported to Australia to help the exploration of the “red center” of the continent, and later for building the railroad in through the vast central Australian desert. Once the railroad was completed, dromedaries were left in the wild, where the current half a million size feral population of dromedaries is causing ecological issues in the country till now (Crowley, 2014). Dromedaries were also introduced to the United States to be used for transporting military materials in the western US deserts during the Seminole War starting in 1851 (Young, 1982).

A comparative analysis of the extinct wild dromedary, early dromedary domesticates, and contemporary dromedary DNA samples shows that there has been extensive gene flow between

dromedary populations from Southern Asia to Central Africa, whereas the West African populations are different due to their isolation from other dromedary populations (Almathen et al., 2016; M. Plasil et al., 2019). Similar conclusions about low genetic diversity in the dromedary were reached by another study that involved more ancient wild and early domestic dromedary samples, and improved ancient sample sequence coverage. The study concluded that genetic diversity in the dromedary is lower than in the Bactrian camel and other domestic species (Mohandesan et al., 2017).

Alpaca reference genome reveals adaptation to high-altitude

Nyala's Accoyo Empress Carlotta, a Huacaya female alpaca born in 2005, was the donor for all reference genome assemblies as well as the BAC library CHORI-246. The first alpaca assembly was released in 2008 and produced 1.9 Gb (~65% of the genome) in almost 300 thousand scaffolds. The second genome assembly was released in 2011, which contains 2.1 Gb (72.4%). Recently, we generated an improved, high-quality, chromosome-assigned reference genome assembly (Richardson et al., 2019). In this work, 190X genome coverage of paired-end and meta-pair short-reads, 8X PacBio long reads, and 60X genome coverage Dovetail Chicago chromatin interaction scaffolding data were generated from Carlotta (the donor animal stated above). The short reads as well as the first two genomes were *de novo* assembled. Then, PacBio long reads were incorporated to the assembly. Finally, in order to reduce the rate of error, the assembly iteration was scaffolded with the short reads data and Dovetail Chicago data. Sequence information from the alpaca cytogenetic map (F. Avila, Baily, et al., 2014) along with available information from comparative chromosome painting between camel, human, cattle and pig

(Balmus et al., 2007) were used to anchor sequencing from the assembly to their corresponding chromosomes.

In this study, transcriptome data was generated from testes and skin samples. Out of these, 22,462 coding genes were predicted which in general, is similar to that in other mammalian species. In addition, a total of 19,927 genes were predicted that have no similarities to other mammalian peptides and likely represent genes encoding for different regulatory RNA species. The new assembly covers 90% of alpaca genome.

The major histocompatibility complex (MHC) is one of the most complex regions in mammalian genomes due to high polymorphisms and the complexity of structures by copy number variations (CNVs) and segmental duplications. MHC is located in chromosome 20 in all camelid species and consists of three classes: MHC I, MHC II and MHC III. MHC class I shows structural similarities between alpacas and Bactrian camels, whereas the dromedary was much different. On the other hand, MHC class II is more similar between dromedaries and Bactrian camels, but different in alpacas (Richardson et al., 2019).

Camelid chromosome 36 was the only one chromosome not revealed by Zoo-FISH (Balmus et al., 2007). It was suggested that this chromosome is very heterochromatic and gene-poor, if containing any genes at all. However, two genes (*ZPBP* and *VWC2*) were mapped to chromosome 36 showing its homology to human chromosome 7 (Felipe Avila et al., 2015; F. Avila, Baily, et al., 2014). More importantly, alpaca chromosome 36 is associated with *minute chromosome syndrome* (*MCS*) (Felipe Avila et al., 2015). The syndrome is associated with infertility and abnormal sexual development in female alpacas and llamas (Felipe Avila et al., 2015; Raudsepp & Chowdhary, 2016). *MCS* is characterized by different sizes of chr36

homologs, which the larger chromosome contains a massive nucleolus organizer (NOR) that is not found in normal individuals (Felipe Avila et al., 2015). A total of 64 genes were mapped to alpaca chr36 of which 24 genes are orthologs to human chr7.

The report investigated genes that are associated with high altitude adaptation by analyzing signatures of selection in 20 genes that are associated with adaptation in other species. Out of the 20 high-altitude adaptation genes, 9 genes were under negative (purifying) selection in alpacas compared to other camelids. This suggests that selection favored to remove deleterious mutations that alter these genes function. On the other hand, two genes show positive selection as allele frequencies are significantly increased in the alpaca.

The dromedary

Due to their adaptation to arid environment conditions, dromedaries (also known as the Arabian camel) were used as the main source of transportation in Asia, the Middle East, and Africa. They have many useful traits: they can carry approximately 400 kg (Nelson, Bwala, & Nuhu, 2015), they can live without water for more than 10 days, and they can adjust their body temperature (Wu et al., 2014). Furthermore, they are an important source of meat, milk, and wool in desert societies (Kadim, Mahgoub, & Purchas, 2008; Konuspayeva, Faye, & Loiseau, 2009; Wu et al., 2014). There is an increasing interest for camel milk in North America and Europe, and a few companies around the world specialize in dromedary dairy production.

Despite being a species of cultural and economic importance and broad biological interest, analysis of the dromedary genome had a late start and lags behind the genome studies of other domestic species. This may be partially attributed to limited geographical distribution and

lack of funding, but may also be caused by features of dromedary biology, such as the long, 13 month-pregnancy and achievement of sexual maturity at 4-5 years of age (Bernard Faye, 2015). Furthermore, traditional cultural practices, which are based on "breeders' knowledge" and oral communication and a lack of written stud books, make it difficult to adopt contemporary breeding strategies and systematically collect resources such as DNA samples, pedigrees, and phenotypes for genetic studies.

Dromedary genome analysis

The dromedary reference genome

The first effort towards sequencing the whole dromedary genome was sequencing expressed sequence tags (ESTs) funded by King Abdulaziz City for Science and Technology (A. M. Al-Swailem et al., 2010). The authors obtained a total of 23,602 putative gene sequences of which 4,500 were novel, species-specific or fast-evolving. The sequences were generated by sequencing RNA samples isolated from 11 different tissues collected from various dromedary breeds from Saudi Arabia. This effort provided the basic sequence information needed to annotate the complete genome sequence.

In 2014, the first assembled annotated dromedary reference genome was published using a male dromedary sample from Saudi Arabia (PRJNA234474_Ca_dromedarius_V1.0) (Table 1) (Wu et al., 2014). The first genome assembly contained 88.91% of the genome. This report investigated the demographic history of dromedaries, Bactrian camels and alpacas, as well as adaptation of camels to arid environments. It was shown that the dromedary genome shared high synteny with the Bactrian camel and alpaca genomes (97.84% and 97.31%, respectively).

Furthermore, the authors reported that dromedary genome has 28.4% repetitive elements which is considered to be lower than other species; for example, repeat content of the Bactrian camel is 30%, alpacas 32%, humans 66% (de Koning, Gu, Castoe, Batzer, & Pollock, 2011), mouse 38%, (Bult et al., 2008), cattle 39% (Bovine Genome et al., 2009) and sheep 47% (Y. Yang et al., 2017).

Thanks to the improvement of bioinformatic tools and cost reduction to generate whole genome sequences (WGS), additional dromedary genomes have recently been assembled (Table 1). A second genome was published in 2016 (Fitak, Mohandesan, Corander, & Burger, 2016) using a female dromedary whose ancestors originate from an isolated population (the spotted camel) from the Canary Islands. The work was funded by Academy of Finland to COIN Centre of Excellence. Around 1.4 million single nucleotide polymorphisms (SNP) were reported when the assembly was compared to the dromedary reference genome v1. This assembly was then improved to be used as the dromedary second reference genome (Elbers et al., 2019), funded by Russian Science Foundation and Russian Foundation for Basic Research (RUBR). This work generated Dovetail Chicago scaffolds, Dovetail Hi-C reads, and Pacific Bioscience (PacBio) long reads from the same individual and resulted in a high-quality chromosome-level reference genome. The previous assembly, which was based on short Illumina reads, was used to generate a *de novo* assembly. The Chicago scaffolds, along with Dovetail Hi-C libraries, were used to map scaffolds from the previous step. PacBio long reads were used to fill in the gaps followed by filtering PacBio reads with Illumina short reads for error correction. As a result, the scaffold length increased 12-fold from 9 Mbp to 124 Mbp. The RNA reads were generated from an adult male Algerian dromedary in a separate project (Alim et al., 2019) and mapped to the assembly.

Chromosomal mapping was done with the help of 4,891 markers from the alpaca radiation hybrid map previously assigned to alpaca chromosomes (W. E. Johnson unpublished data; Avila *et al.* 2014a). This assembly is much better than the first assembly and contains 95.13% of the dromedary genome.

Another study was recently done to generate a chromosome-level assembly of a male dromedary from North Africa (Table 1) (Ruvinskiy, Larkin, & Farre, 2019). The work is still in progress and has taken three steps so far: first, it used the Reference-Assisted Chromosome Assembly (RACA) to generate predicted chromosome fragments (PCFs) to identify chimeric scaffolds (Damas et al., 2017); second, it checked chimeric scaffolds by PCR and removed falsely assembled scaffolds; third, it took the assembled scaffolds and assigned them to chromosomes using both the dromedary and the alpaca physical maps (F. Avila, Baily, et al., 2014; Balmus et al., 2007). The work is funded by the Biotechnology and Biological Sciences Research Council Grant and Russian Foundation for Basic Research (RUBR).

Another two dromedary genomes were assembled using two different individuals from Iran. The aim was to study their transposable and repetitive elements (SINES, LINES, and microsatellites) (Table 1) (Khalkhali-Evrigh, Hedayat-Evrigh, Hafezian, Farhadi, & Bakhtiarizadeh, 2019). This study provides valuable information regarding the structure of the dromedary genome together with a list of transposable and repetitive elements.

Table 1: A summary of published dromedary genome assemblies: Data presented in this table is adapted from NCBI (https://ncbi.nlm.nih.gov) and/or published resources. * indicates that available information was averaged between the two Iranian dromedary assemblies

GenBank accession #	Ca_dromedarius_V 1.0	CamDrom2	ASM164081v1	Ave.* Iranian Ass.
Sex	Male	Female	Male	NA
Breed	Arabian	African	Targui	Iranian
Published Date	10/21/2014	5/17/2019	2/16/2019	8/14/2019
Total sequence length	2.01 Gb	2.15 Gb	1.88 Gb	1.94
Total ungapped length	1.98 Mb	2.13 Gb	NA	NA
# of scaffolds	32,573	23,439	1,797	
Scaffold N50	4.188 Mb	75,021 Mb	54,360 Mb	54.33 Mb
Scaffold L50	132	11	NA	NA
# of contigs	105,348	59,913	207,871	145,772
Contig N50	69 Kb	204 Kb	31 Kb	NA
Contig L50	7,850	3,014	19,666	NA
# of annotated genes	20,714	24,475	NA	NA
Genome Coverage	65x	65x	46.43x	~40x
Chromosomal assignment	none	yes	none	none

Generation of Genome Analysis Tools - Whole Genome Radiation Hybrid Panels

Recently, whole genome radiation hybrid (RH) panels were constructed for the dromedary as an additional tool to study the dromedary genome (Perelman et al., 2018). The first panel was 5000rad with a retention frequency of 47.7%, constructed from a female dromedary and comprised of 93 clones corresponding to autosomes and the X chromosome. The second panel was 15,000rad with an average retention frequency of 39.9%, constructed from a male and comprised of 90 clones. The 15,000rad panel is expected to provide the needed resolution for

mapping complex genomic regions such as the Major Histocompatibility Complex (*MHC*) and the Y chromosome. Even though the number of RH clones in the panels is not very high, application of genotyping by clone sequencing will eliminate the need of clone expansion. Furthermore, as stated above, the alpaca radiation hybrid map was successfully used to bring the dromedary assembly to the chromosomal level (Elbers et al., 2019).

Justification

There are several good reasons to study the dromedary genome and the underlying genetics of traits and biological features. The dromedary camel is a culturally and economically important species in many countries and it is important to generate basic resources and tools for genomic and genetic studies and for improved breeding systems. Limited available genomics tools, from one side, and the lack of national reporting system and stud book, from another, hinder the use of advanced breeding strategies, such as marker assisted-selection (Hayes, Lewin, & Goddard, 2013), and are among the major limitations to improve our knowledge about the species. Even though, there has been an increased scientific interest in camelid genetics/genome analysis in the past decade, genetic studies based on different traits and disorders are still limited. Therefore, it is essential to expand on the study of genetics of different dromedary traits and disorders to improve breeding system and welfare of animals. With the availability of dromedary reference genome, it is now possible to investigate genetic loci associated with different phenotypes. For example, the genes associated with coat color patterns. Furthermore, compared to other ruminants (camelids are pseudoruminants), there are no studies concerning genes associated with meat and milk production in the dromedary. As selection took place since domestication, signatures of selection for different phenotypes can be traced. With the cost

reduction of whole genome sequencing (WGS) and the improved bioinformatics tools, it is possible to identify selective sweeps showing differentiation between dromedary populations based on certain phenotypes and using nucleotide diversity and various strategies that are well established in other species. This will lead to the identification of genes that are associated with different traits and the discovery of genetic variants that are associated with traits and genetic disorders that are common in one population but rare in other populations. This will generate a panel of genetic loci that may explain some phenotypic differences between dromedary populations**.**

The overall goal of this research was to generate genetic resources for dromedary genomics and to improve knowledge of the dromedary genome. This was achieved through four specific aims:

Specific Aim 1: Resource collection and generation of tools for dromedary genomics;

Specific Aim 2: Identify gene variants underlying the basic coat colors in the dromedary and develop genetic assays for testing;

Specific Aim 3: Comparative cytogenetic mapping of basic coat color genes in camelids;

Specific Aim 4: Population differentiation between white, black and red dromedaries from Saudi Arabia.

CHAPTER II

GENERATION OF RESOURCES AND TOOLS FOR DROMEDARY GENOMICS

Introduction

Compared to other livestock species, such as cattle, sheep, horses and pigs, where gene mapping and genome analysis has been ongoing for decades, genome analysis in the dromedary had a late start. The main progress has been made just in the past 5 years (Almathen et al., 2016; Burger, 2016; Elbers et al., 2019; Fitak et al., 2016; Wu et al., 2014). This also has delayed internationally organized generation of genomic resources and tools for the species. The goal of this study was to contribute to the generation of such resources and tools and their application for genomics research in the dromedary and other camelids.

Methods

1) DNA and phenotype collection (dromedary and Bactrian). **Blood samples** for DNA isolation were collected from 194 dromedaries, 6 Bactrian camels, and one F1 and one F2 dromedary x Bactrian hybrids together with as comprehensive phenotype descriptions as possible. The samples were collected from Saudi Arabia (n=167 dromedaries) and the US (n=35 total). All collected samples have a written phenotypic description including, if available, the color, breed, type (production/riding/racing), a photo, information about parents and origin, sample date and owner. Due to different spelling of breed names, we adopted the spelling of breed names from Porter *et al*. (2016) (Porter, Alderson, Hall, & Sponenberg, 2016). Genomic DNA was isolated from peripheral blood lymphocytes using Gentra Puregene Blood Kit (Qiagen) following the manufacturer's protocol, or by standard phenol-chloroform method (J.

Sambrook, E. F. Fritsch and T. Maniatis 1989). DNA stock solutions are stored at -80^{o}C until needed.

2) Tissue collection for transcriptome analysis. Testis tissue was obtained from 11 male dromedaries during castration. The tissue was cut into small, approximately 1cm^3-size pieces and each stored in 4 mL RNAlater solution (Ambion) at -80^{o}C until needed.

3) Establishment and cryopreservation of primary fibroblast cultures. Primary fibroblast cultures were established from 3 dromedaries: a female from San Diego zoo, a male established from placenta, and a male established from skin biopsy. Initiation of the cultures, growing the fibroblasts and cryopreservation followed standard procedures described elsewhere (Siengdee, Klinhom, Thitaram, & Nganvongpanit, 2018). Briefly, pieces of tissue were cut in sterile conditions into small 0.5 cm^2-size pieces, washed at least 10 times in Hanks Balanced Salt Solution (HBSS; Gibco) supplemented with 2X antibiotic-antimycotic solution (100X stock; Invitrogen). The pieces were transferred into T25 cell culture flasks, approximately 4 to 6 pieces per flask and covered with a few drops of culture medium (MEM alpha with Glutamax and nucleosides; Gibco) supplemented with 20% Fetal Bovine Serum (Atlanta Biologicals) and 1X antibiotic-antimycotic solution. The medium was changed every day and the pieces were monitored for cell outgrowth using an inverted microscope Zeiss Primo Vert and 4X and 10X phase contrast objectives. Once all pieces had given outgrowth of approximately 100 cells, the pieces were removed and frozen in freezing medium (MEM alpha with 20% FBS, 10% DMSO and 1X antibiotics-antimycotics) in liquid nitrogen (LN2), and designated as P0. The cells in the flasks were detached with 0.25% Trypsin-EDTA, transferred into a fresh T25 flask for even distribution and cultured with 5 mL medium and 5% CO^2 at 37^{o}C until confluency. Culture

medium was replaced every 2-3 days. Confluent cultures with approximately 1 million cells in each, were discontinued. The cells were trypsinized as described above, mixed with 1 mL freezing medium in cryotubes and frozen in LN2 until needed.

4) Preparation of fixed cell suspensions for chromosome analysis and gene mapping. To generate this resource, blood was collected in Na-heparin vacutainers™ (Becton Dickinson) from 24 male and female dromedaries, 5 Bactrian camels and 2 interspecific hybrids. Short-term lymphocyte cultures were established and harvested for metaphase chromosome preparations following standard procedures described elsewhere (Raudsepp & Chowdhary, 2008). Briefly, approximately 1 mL whole blood was added into 9 mL sterile medium containing RPMI Medium 1640 with Glutamax™ and 25 m*M* HEPES Buffer (Gibco) supplemented with 30% FBS, 1X antibiotic-antimycotic solution (100X stock; Gibco), and 1% pokeweed mitogen (lectin from *Phytolacca Americana;* Sigma Aldrich). The cells were cultured for 72 h at 37°C and harvested by adding 100 µL of demecolcine solution (10 mg/mL in HBSS; Sigma-Aldrich) to stop the movement of cell spindle and capture chromosomes at metaphase. The cells were treated with hypotonic solution (Optimal Hypotonic Solution; Genial Helix) to burst the cytoplasm.

Finally, the cells were fixed with methanol/glacial acetic acid in a ratio of 3:1 and used to make chromosome preparations on clean, wet microscope slides. The remaining of the cell suspension was frozen at -20^{o}C until needed. Chromosomes were stained with Giemsa or by G-banding (Seabright 1971). For chromosome analysis and karyotyping, 10-20 metaphase spreads per individual were captured, analyzed and karyotyped using Zeiss semiautomated Axioplan2 microscope and IKAROS image analysis software (MetaSystems GmbH). Dromedary and

Bactrian chromosomes were arranged into karyotypes following the nomenclature proposed by Balmus and colleagues (Balmus et al., 2007).

5) Generate a collection of alpaca BAC clones for comparative genomics in the dromedary. We used the alpaca genomic BAC library CHORI-246 (https://bacpacresources.org/), which is the only BAC library available for camelids. In collaboration with researchers at National Agrarian University La Molina, Peru, we identified 11 candidate genes for coat color and hair characteristics (*COL1A1, CTNNB1, DAB2IP, KRT15, KRTAP13-1, TNFSF12, ALX3, NCOA6, SOX9, ZIC1* and *ZIC5*). Genomic sequences for these genes were retrieved from the Ensembl Genome Browser (http://useast.ensembl.org/index.html), masked for repeats (RepeatMasker: http://www.repeatmasker.org/) and used for the design of polymerase chain reaction (PCR) primers in Primer3 software (http://frodo.wi.mit.edu/primer3/), as well as overgo primers in or around the PCR amplicons (Gustafson et al., 2003) (Table 2). Overgo primers were radioactively labeled with [^{32}P] 2'-deoxyadenosine triphosphate (dATP) and [^{32}P] deoxycytidine triphosphate (dCTP; Amersham Biosciences, USA) as previously described (Gustafson et al., 2003). Equal amounts of overgo probes were pooled and hybridized to high-density filters of the CHORI-246 alpaca bacterial artificial chromosome (BAC) library (http://bacpac.chori.org/library.php?id=448). The hybridization solution, containing the labeled probes, 20X SSPE, 10% sodium dodecyl sulfate, 5% dry milk, 100X Denhardt's solution, and 50% formamide, was denatured by boiling for 10 min, chilled, and hybridized to library filters at 42°C for 16h. The filters were washed 3 times in 2X SSPE at 55°C for 15 min, exposed to autoradiography films over intensifying screens for 2–3 days at –80°C, and the autoradiograms were developed. Positive BAC clones were identified and picked from the library. The BAC

clones corresponding to individual genes were identified by PCR using gene-specific primers and BAC cell lysates as templates. Isolation of DNA from individual BACs was carried out with the Plasmid Midi Kit (Qiagen) according to the manufacturer's protocol. The quality and quantity of BAC DNA was evaluated by gel electrophoresis and nanodrop spectrophotometry. The physical location of the genes was determined by fluorescence *in situ* hybridization (FISH) to alpaca metaphase chromosomes according to our protocols (Raudsepp & Chowdhary, 2008). Briefly, DNA from individual BAC clones was labeled with biotin-16-deoxyuridine, 5'-triphosphate (dUTP) or digoxigenin (DIG)-11-dUTP, using Biotin- or DIG-Nick Translation Mix (Roche), respectively. Differently labeled probes were hybridized in pairs to metaphase/interphase chromosomes. Biotin and DIG signals were detected with avidin-fluorescein isothiocyanate and anti-DIG-Rhodamine, respectively. Images for a minimum of 10 metaphase spreads and 10 interphase cells were captured for each experiment and analyzed with a Zeiss Axioplan2 fluorescence microscope equipped with Isis Version 5.2 (MetaSystems GmbH) software. The chromosomes were counterstained with 4'-6-diamidino-2-phenylindole (DAPI) and identified according to the nomenclature proposed by Balmus and colleagues (2007). We have stored the remaining unlabeled BAC DNA and labeled probes at -80oC for future comparative mapping and molecular cytogenetics in camelids.

Table 2: Gene specific and overgo primers. (Mendoza et al., 2019)

Gene symbol	Identified BAC clones	PCR primer 50 -30	PCR product size (bp)	Overgo primer sequence 50 -30
ALX3	115I10	F: TATGTCTCCGTACTCCCACTCTC	161	F: GCTCTAGGGGGCCACAGCTTTGAG
		R: GGAGACTTATAGTCGTCATCTGG		R: CGTCATCTGGGGAGGGCTCAAAGC
COL1A1	198E13	F: CCATTGGTAGTGTTGGTGCT	365	F: GCCCTGTTGGCAAAGAAGGCAGCA
	204B18	R: AGGGAAGCCTCTTTCTCCTC		R: TCACCACGAGGACCTTTGCTGCCT
	264O11			
	271L20			
	295O17			
CTNNB1	129B09	F: ATCCCAGCTATCGTTCTTTTCA	300	F: CACTCCGGTGGATACGGACAGGAT
	150A21	R: CCTACCAACCCAAGTCTTTCTG		R: GGTCCATACCCAAGGCATCCTGTC
DAB2IP	101B06	F: TACTGAGAACGGCGAGTTCA	107	F: GAACGGCGAGTTCAGAAACAGCAGCAA
		R: AAAGCTCAGCCTCTCTCTCG		R: CGTGCCTGGGACACTTGAATTGCTGCT
KRT15	263E22	F: GGCAAAGTCCGCATCAATGTT	218	F: TGGCCAGAGGGGCCAGAAGGGCAAA
	268A9	R: ATGCCAAGCAGCCAACTAGG		R: CCCCTCTGGGTCTAGAGTTTGCCCT
	274A22			
NCOA6	34F15	F: CCCAAGATTTTCTAAAGACAGGAA	109	F: CAGCTGTGTTTACAACTCCTCCAGCCAAG
	46J23	R: CTGGTCAGTATGGGCTTATCTCTT		R: CTGGTCAGTATGGGCTTATCTCTTGGCTG
	59N23			
	86O24			
SOX9	13O23	F: AAATGCTCTTATTTTTCCAACAGC	220	F: GTGTTATGGGATCAGTTTGGGGGGTTA
	30B6	R: AATCACAAAGCCTGAGGAATTAAG		R: CTGAGGAATTAAGCAAAGCTAACCCCC
	32I21			
	58P4			

Table 2 continued

Gene symbol	Identified BAC clones	PCR primer 50 -30	PCR product size (bp)	Overgo primer sequence 50 -30
	68P18			
	115A15			
	122H18			
	169O5			
	172F10			
	186L14			
	202F10			
	231H17			
	249C14			
	279H10			
	297J24			
	306O5			
TNFSF12	133N9	F: GACCTGAATCCCCAGACAGA	94	F: AGCCAGGACACCGTGTCTTTCCTG
	169O5	R: GTGGTTTCCGGCCTTTAGGT		R: GAGGCCGAACCAGTTTCAGGAAAG
	172F10			
ZIC1	127I17	F: AGTCCGCGTTCAGAGCACTAT	192	F: GCGCCGGCGCTTTCTTCCGCTACATG
	135I16	R: GAAAGTTTTGTTGCACGACTTTTT		R: CTGTTTGATGGGCTGGCGCATGTAGC
ZIC5	211H22	F: GCAAACTTTCTGCAAGTGCAAC	199	F: AGGGGGCACGAAGCGAAAGCGAAG
	224A3	R: GGAAGCCTGTCATATTCTGAAAC		R: CTGTGCTCACTGACGCCTTCGCTT

Results

1) DNA and phenotype collection. The DNA collection includes dromedaries of Saudi Arabia and US origin and incorporates male and female animals of multiple breeds, color phenotypes and types (Table 3). For comparative purposes, DNA samples were also obtained

from 6 Bactrian camels and two camel hybrids. In the present dissertation, this resource was used for the studies described in Chapters III and IV.

Table 3: Summary information about the genomic DNA samples collected from the dromedary

Number of individuals	Color	Breed	Geographic location	Type (purpose)
51	White	Waddah	South-west SA	Production
38	Black	Majaheem	South-west SA	Production
39	Brown-red	Saheli	West SA	Production
7	Red	Homor	North SA	Multipurpose
13	Varies	Muhajanat	West SA	Racing
4	Dark brown	Sofor	North SA	Production
4	Light beige	Shageh	North SA	Production
9	Brown	Shaele	North SA	Multipurpose
2	Light beige	Omani	North SA	Multipurpose
2	Brown	Sudanese	Sudan	Multipurpose
2	Spotted	NA	US	Multipurpose
25	Varies	NA	US	Multipurpose

2) Testis tissue collection. Testis tissue for transcriptome/genome analysis was archived from 11 individual dromedary males. One of the samples has already been used (not part of this dissertation) in another project to synthesize cDNA for cDNA selection experiments with flow-sorted dromedary Y chromosome, and also for testis RNAseq to generate testis transcriptome data for Y chromosome annotation. Testis tissue collection will have future use for functional annotation of the dromedary genome assembly and for the discovery of male fertility genes.

3) Primary fibroblast cultures. Primary fibroblast cultures were successfully established from all three individuals (one female, 2 male dromedaries) and the cells have been cryopreserved for future applications. The cell line that was established from a skin biopsy of a male dromedary CJ (lab ID: CDR83; Figure 2) from Franklin Safari, Texas has already had multiple use in different research projects (not part of this dissertation):

a) To construct a 15,000rad radiation hybrid (RH) panel (Perelman et al., 2018).

b) To flow sort 20,000 Y chromosomes (collaboration with Dr. Ferguson-Smith, Cambridge, UK); the flow-sorted Y chromosomes were sequenced by PacBio and Illumina MySeq and used for cDNA selection with testis cDNA. The data is currently used to assemble and annotate the dromedary Y chromosome (another PhD project in Dr. Raudsepp's lab).

c) To improve the genome assembly of the dromedary. The whole genome of CJ was sequenced at the University of Florida (Dr. Samantha Brooks) and the data has been accepted by the Vertebrate Genome Project (VGP: https://genome10k.soe.ucsc.edu/about/).

d) In addition, blood and DNA of CJ was used to construct whole genome BAC library at the Children's Hospital Oakland Research Institute (CHORI) by Dr. Pieter de Jong. Fresh blood

samples were collected and sent to CHORI twice. Unfortunately, the generation of dromedary BAC library failed due to using the wrong restriction enzyme.

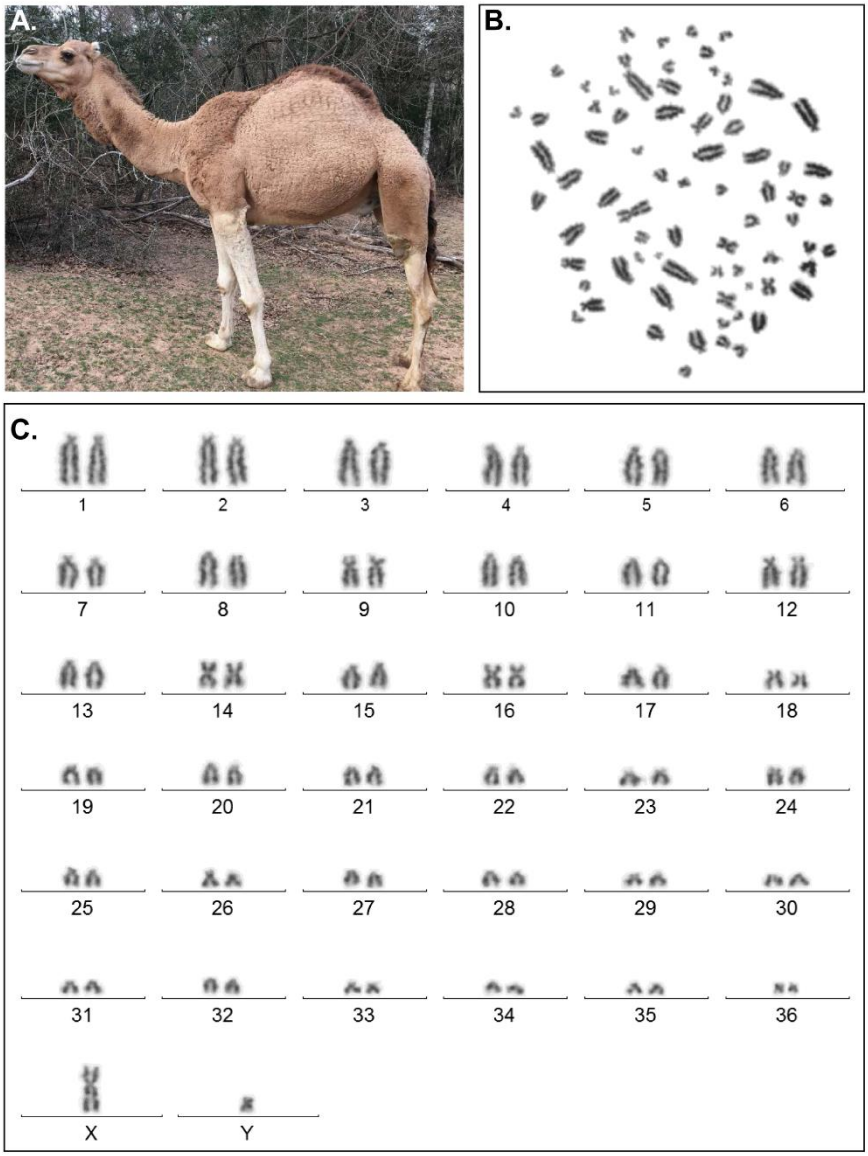

Figure 3. Dromedary CDR83, named *CJ* from Franklin Safari, Texas. A. Photo of CDR83; **B.** CDR83 metaphase spread; **C.** CDR83 karyotype showing normal male karyotype 74,XY. ***Note:*** This animal was the fibroblast donor for 15,000rad RH panel and for flow-sorting dromedary Y chromosome, as well as DNA donor for WG sequencing (U of Florida; Dr. Brooks group), and for the attempts to construct a genomic BAC library at CHORI (by Dr. Pieter de Jong).

4) Fixed cell suspensions and chromosome preparations. Chromosome preparations and karyotypes were obtained for 24 dromedaries, 5 Bactrian camels and 2 interspecific hybrids. All individuals were chromosomally normal with 74,XY in males (Figure 2) or 74,XX in females. There were no obvious differences between dromedary and Bactrian camel karyotypes except for the Y chromosome, which is different between the two species. The cell suspensions and chromosome preparations are important tools for gene mapping and comparative gene mapping, and for the study of normal and aberrant chromosomes. Even though no genetic disorders were recorded in this collection, it is essential to know the normal karyotype of the species. The chromosome preparations were successfully used for gene mapping as detailed in Chapter III.

5) Collection of alpaca BAC clones for comparative genomics in the dromedary and other camelids. As mentioned above, all camelids have the same number of chromosomes (2n=74), and their karyotypes are very similar. Due to these similarities, alpaca probes can be hybridized to other camelid species. We successfully identified and isolated alpaca BAC clones containing 11 candidate genes for coat color and hair characteristics (*COL1A1, CTNNB1, DAB2IP, KRT15, KRTAP13-1, TNFSF12, ALX3, NCOA6, SOX9, ZIC1* and *ZIC5*). The genes were FISH mapped to alpaca chromosomes (Mendoza et al., 2019), and will also be mapped to the dromedary and Bactrian chromosomes to facilitate the development of the camelid cytogenetic map and for chromosomal anchoring of the reference genomes. In addition, we isolated BACs for another 32 markers that were mapped to alpaca chromosomes. These markers are associated with polymorphic regions and will assist with the generation of alpaca whole-

genome SNP chip as an important tool for genome-wide association studies, mutation discovery and population analysis.

Discussion

Resources that were generated for the dromedary have a broad application in dromedary and comparative genomics. Some of these resources, such as the phenotypic collection of DNA samples, chromosome preparations and alpaca BAC clones for coat color genes, were used for the research in this PhD project. DNA samples, fibroblast cell lines, chromosome preparations and alpaca BAC clones were also used in collaborative projects (Mendoza et al., 2019; Perelman et al., 2018). However, these unique resources will have more applications in the future. For example, frozen live fibroblasts have a potential use for regular and single cell RNAseq and chromatin interaction mapping by Hi-C (Belton et al., 2012). The frozen fibroblasts are also an excellent source of extra-high molecular weight DNA for such cutting-edge genome analysis platforms like BioNano. The genomic DNA collection from almost 200 dromedaries will have multiple applications in populations studies. Finally, the genomic DNA obtained from an F1 dromedary-Bactrian camel hybrid will be an excellent source for generating high quality sequences for both parent genomes (Langdon, Peris, Kyle, & Hittinger, 2018).

CHAPTER III

COAT COLOR GENETICS

Mammalian Coat Color Genetics

Overview

Mammalian coat color is a phenotypic trait that serves for camouflage and communication in the wild, while in domestic species, selection for color phenotypes have mainly commercial, historical, or aesthetic reasons. For example, white coat color in pigs (large white) is associated with higher commercial value of pork (Giuffra et al., 2002); certain horse breeds such as Paint or Friesians, are restricted to certain color patterns; a lighter coat color in alpacas has been associated with supreme fiber quality (Gutiérrez et al., 2011). Overall, color phenotypes have played an important role in domestication and selection of domestic animal species (L. Andersson & Georges, 2004).

Reports on coat color inheritance of domestic animals go back to the mid-1850s; in fact, Charles Darwin published the first book on Variation Under Domestication back in 1868 (Darwin, 1868). Since then, heredity of color variation was further investigated, in which rodent species were of particular interest. This is perhaps due to their short life spans, which allow for quick results to be shown of offspring crossing (Castle, 1905; Dunn, 1921; Punnett & Punnett, 1912; Sturtevant, 1913). Little reviewed published reports of rodents and carnivores summarizing 41 and 30 alleles are associated with coat colors (Little, 1958). In addition, he summarized lethal and pleiotropic effects of some alleles that are associated with blindness in cats. However, the following molecular genetics studies showed that there are more genes and gene interactions that underlie coat color and associated traits.

Mammalian coat color is a genetic trait governed by a few major pigmentation genes, and further modified by a larger number (over 350) of dilution and spotting genes (Oetting, Austin, & Bennett, 2011). Importantly, since the pigment cells are derived from the embryonic neural crest, mutations in pigmentation genes frequently cause pleiotropic effects involving sight, hearing, and neurologic functioning (Monika Reissmann & Arne Ludwig, 2013). For example, certain hypopigmentation phenotypes in alpacas, cats, and dogs are associated with congenital deafness (Jackling, Johnson, & Appleton, 2014; Kaas, 2005), while certain spotting patterns in horses (appaloosa) are associated with congenital night blindness (Fritz et al., 2014).

Mechanisms of major coat color genes

Key genes are expressed during embryonic development to ensure that melanocytes are transported from the neural crest to the skin (Michael Cieslak, Monika Reissmann, Michael Hofreiter, & Arne Ludwig, 2011). Melanocytes are differentiated from melanoblasts that are derived from the neural crest that also differentiate into the peripheral nervous system (Barsh, 2001). Reported genes that are associated with the transportation are *v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog* (*KIT*)/*KIT ligand* (*KITL*), *endothelin receptor type B* (*EDNRB*), *endothelin 3* (*EDN3*), *microphthalmia-associated transcription factor* (*MITF*), *sex-determining region Y (SRY-) box 10* (*SOX10*), and *paired box 3* (*PAX3*). If these cells partially or completely fail to migrate, white spotting or complete depigmentation occurs, respectively. The most remote regions of the melanoblasts are the legs, the forehead and the belly areas, which explains why these regions are often depigmented in many mammalian species.

Melanosomes are the organelles where melanogenesis takes place. Melanosomes require different components, such as structural components and membrane transport channels (Michael Cieslak et al., 2011). An example of structural components is *silver homolog* (*SILV*) also known as (*PMEL17*), whereas examples of membrane transport channels are *oculocutaneous albinism II* (*OCA2*), *solute carrier family 45 member 2* (*SLC45A2*), *solute carrier family 36 member 1* (*SLC36A1*), *solute carrier family 7 member 11* (*SLC7A11*), and *solute carrier family 24 member 5* (*SLC24A5*), in which these genes are associated with forming melanocytes.

While mammalian color phenotypes are influenced by many genes, the type, amount, and distribution of the main pigments are regulated by just a few, specific genes. Among these, the two key genes are *melanocortin 1 receptor (MC1R)* and *agouti signaling protein* (*ASIP)* (Hitoshi Suzuki, 2013). These two genes play the major role in regulating the production of black/brown pigment (eumelanin) and red/yellow pigment (pheomelanin) (Klungland, Vage, Gomez-Raya, Adalsteinsson, & Lien, 1995; Hitoshi Suzuki, 2013). It has been shown that SNPs in these two genes are associated with different coat colors in mammals (Hitoshi Suzuki, 2013).

These pigments, eumelanin and pheomelanin, are produced by melanocytes (Hitoshi Suzuki, 2013). The *MC1R* gene is tissue-specific and expressed primarily in the melanocyte membrane. This gene plays a major role in determining the type of pigmentation by responding to two signaling molecules: α-melanocyte stimulating hormone (αMSH) and ASIP (Hitoshi Suzuki, 2013). When αMSH is present, it binds to its receptor, MC1R, and activates the pathway of eumelanin production (Hitoshi Suzuki, 2013). On the other hand, ASIP blocks MC1R and stops the eumelanin pathway, leading to the production of pheomelanin (Michael Cieslak et al., 2011). Briefly, the expression ratio of both melanin products is determined by the balance

(amount) between the agonist αMSH acting on MC1R or the antagonist ASIP blocking the receptor (Michael Cieslak et al., 2011).

MC1R and *ASIP* play important roles in evolutionary coat color changes in mammals (Hitoshi Suzuki, 2013). This is because mutations in these two genes are less harmful to survival than mutations in several other coat color genes. For example, mutations in *the receptor tyrosine kinase* (*c-KIT*) or *endothelin receptor type B* (*EDNRB*) cause various types of depigmentation that can be destructive or lethal (Hitoshi Suzuki, 2013; A. A. Webb & Cullen, 2010). Therefore, comparative sequence analysis of *MC1R* and *ASIP* among mammals is of interest for their role in evolution.

An important step of eumelanin and pheomelanin production is the catalyst of tyrosine by the tyrosinase-related protein family (Murisier & Beermann, 2006). Variations in *tyrosinase* (*TYR*) gene are associated with depigmentation, whereas variations in *tyrosinase related protein 1* and *2* (*TYRP1*, *TYRP2*) are associated with brown and dark gray color, respectively (Bennett, Huszar, J Laipis, Jaenisch, & J Jackson, 1990; Guyonneau, Murisier, Rossier, Moulin, & Beermann, 2004).

The phenotypic classification is problematic because different haplotypes are associated with a single gene that results in different coat colors; this can be further complicated because a combination of variations in two or more genes influence the same color (Michael Cieslak et al., 2011). For example, variations in either *TYR*, *KIT*, *EDNRB* or *STX17* genes can result in white coat color. Subsequently, phenotypic classification provides limited insights in understanding the mechanisms of color variations. Therefore, due to these large numbers of variations in major and diluted genes, the mammalian coat color is a surprisingly complex trait. Table 4 shows some

examples of major genes and their association with coat color phenotypes in different species.

Table 4: Examples of genes and their phenotypic effect and their association with coat color phenotype in different species

Gene ID	Phenotype	Pleotropic effect	species	reference
EDNRB	Frame-ovaro white spotted	Lethal homozygote	Horse	(Metallinos, Bowling, & Rine, 1998)
KIT	White/white-spoted	NA	Horse	(Brooks & Bailey, 2005)
MITF	White/white-spoted white-spoted	Microphthalmia NA	Dog Cattle	(Stritzel, Wohlke, & Distl, 2009) (Hofstetter et al., 2019)
β-defensin	Black	NA	Dog	(Candille et al., 2007)
SILV	Silver coat color	NA	Horse	(Brunberg et al., 2006)
SLC45A2	Cream white	NA	Horse Tiger	(Cook, Brooks, Bellone, & Bailey, 2008) (X. Xu et al., 2013)
SLC36A1	Champagne	NA	Horse	(Mariat, Taourit, & Guerin, 2003)
TYRP1	Brown	NA	Cattle	(Berryere, Schmutz, Schimpf, Cowan, & Potter, 2003)
TYR	Hypopigmentation	Albinism	Cattle	(Schmutz et al., 2004)

Comparative FISH-mapping of *MC1R, ASIP* and *TYRP1* in New and Old World camelids and association analysis with coat color phenotypes in the dromedary (*Camelus dromedarius*)*

Introduction

Mammalian coat color is a phenotypic trait that serves for camouflage and communication in the wild, and has been a target for selection by humans in farm and companion species since their domestication (L. Andersson, 2001; Michael Cieslak et al., 2011). As a result, domestic animals display a perplexing variety of colors, patterns and markings, which reflect the genetic diversity of a breed or species, as well as historic and aesthetic preferences or commercial needs of humans.

Many genes regulate coat color. This was already noted by J.B.S. Haldane over 90 years ago when he studied color genetics in rodents and carnivores and suggested that there are at least 20 different color genes in mammals (Haldane, 1927). Since then, approximately 150 coat-color associated genes have been described in mice, humans and domestic animals (Bellone, 2010; Michael Cieslak et al., 2011; M. Reissmann & A. Ludwig, 2013), whereas Color Genes database lists 378 mouse loci with their human and zebrafish homologues that are associated with various pigmentation phenotypes.

---

* Reprinted from Alshanbari, F., Castaneda, C., Juras, R., Hillhouse, A., Mendoza, M. N., Gutiérrez, G. A., . . . Raudsepp, T. (2019). **Comparative FISH-Mapping of MC1R, ASIP, and TYRP1 in New and Old World Camelids and Association Analysis With Coat Color Phenotypes in the Dromedary (*Camelus dromedarius*)**. *Frontiers in Genetics*, 10, 340, Copyright 2019 by Frontiers Media SA.

Despite the large number of genes involved, the production, amount and distribution of main pigments, the brown/black eumelanin and the red/yellow pheomelanin, are controlled by just a few major pigmentation genes (Bellone, 2010; Rees, 2003; Hitoshi Suzuki, 2013). These include *melanocortin 1 receptor* (*MC1R*), *agouti signaling protein* (*ASIP*) and *tyrosinase related protein 1* (*TYRP1*). Melanocortin 1 receptor is the key switch between the synthesis of eumelanin or pheomelanin; ASIP is an antagonist ligand that regulates MC1R signaling by inhibiting the MC1R receptor, and TYRP1 is a melanogenic enzyme that influences the quantity and quality of melanins (Bellone, 2010; Pielberg, 2004; Sturm & Duffy, 2012; H. Suzuki, 2013). Associations between basic coat colors and DNA sequence polymorphisms in *MC1R*, *ASIP*, *TYRP1*, are known for most domestic species (Bellone, 2010; M. Cieslak, M. Reissmann, M. Hofreiter, & A. Ludwig, 2011; Pielberg, 2004; Rieder, Taourit, Mariat, Langlois, & Guerin, 2001; Schmutz & Berryere, 2007), and are routinely used for genetic testing.

In contrast to other domestic species, coat color genomics in camelids had a late start, even though fiber color is an important trait for the alpaca industry (Morante et al., 2009) and there is an interest for breeding white or black dromedaries in some Arabian countries (Almathen, Elbir, Bahbahani, Mwacharo, & Hanotte, 2018). A few studies in alpacas have associated mutations in *ASIP* with the black color (Chandramohan, Renieri, La Manna, & La Terza, 2013; Feeley, Bottomley, & Munyard, 2011) and identified *MC1R* mutations that may determine light phenotypes, though the findings about the alpaca *MC1R* remain inconclusive (Chandramohan et al., 2013; Feeley & Munyard, 2009; Guridi, Soret, Alfonso, & Arana, 2011).

Research on dromedary color genes is even more recent with just two publications. The first study revealed that a frameshift mutation in the *KIT* gene explains some, though not all

forms of white-spotting phenotypes in the dromedary (Holl et al., 2017). The most recent study identified a missense mutation in *MC1R* that is associated with the white color, and a deletion and a single nucleotide polymorphism (SNP) in *ASIP* exon 2 that are associated with the black/dark brown color in dromedaries (Almathen et al., 2018). Current reference genomes for the alpaca and the dromedary are in scaffolds and not assigned to chromosomes. Because of this, chromosomal location is known only for the few coat color genes that were included in the alpaca whole genome cytogenetic map (F. Avila, Baily, et al., 2014). Among the main pigmentation genes, *ASIP* and *TYRP1* have been mapped in the alpaca but not in other camelids, whereas *MC1R* is not mapped in any camelid species.

The aim of this study is to confirm and refine the recently reported *MC1R* and *ASIP* mutations for white and black coat color in dromedaries, and search for novel color-related variants in *TYRP1*. We compare the accuracy of genotyping the white and black mutations in large dromedary populations by direct sequencing and with a TaqMan™ assay. Finally, we cytogenetically map *MC1R, ASIP* and *TYRP1* in three camelid species.

Materials and Methods

Ethics statement. Procurement of peripheral blood was performed according to the United States Government Principles for the Utilization and Care of Vertebrate Animals Used in Testing, Research and Training. These protocols were approved by Animal Use Protocol AUP #2011-96, # 2018-0342 CA and CRRC #09-47 at Texas A&M University.

Animals and phenotypes. We sampled 188 dromedaries originating from Saudi Arabia (SA; n=171) and from the United States (US; n=17). Coat color phenotypes were determined by visual inspection, recorded in written notes and/or photos, and were as follows: white/cream

(n=53), black/dark brown (n=38), and brown/beige (n=97) (Fig. 4). Two brown dromedaries had white markings and blue eyes. We use 'brown' as a generic term to denote animals with wild-type coat color, which can range from light beige to darker reddish-brown with either matching or darker tail and hump. Blood was collected by jugular venipuncture into EDTA-containing Vacutainers (Becton Dickinson).

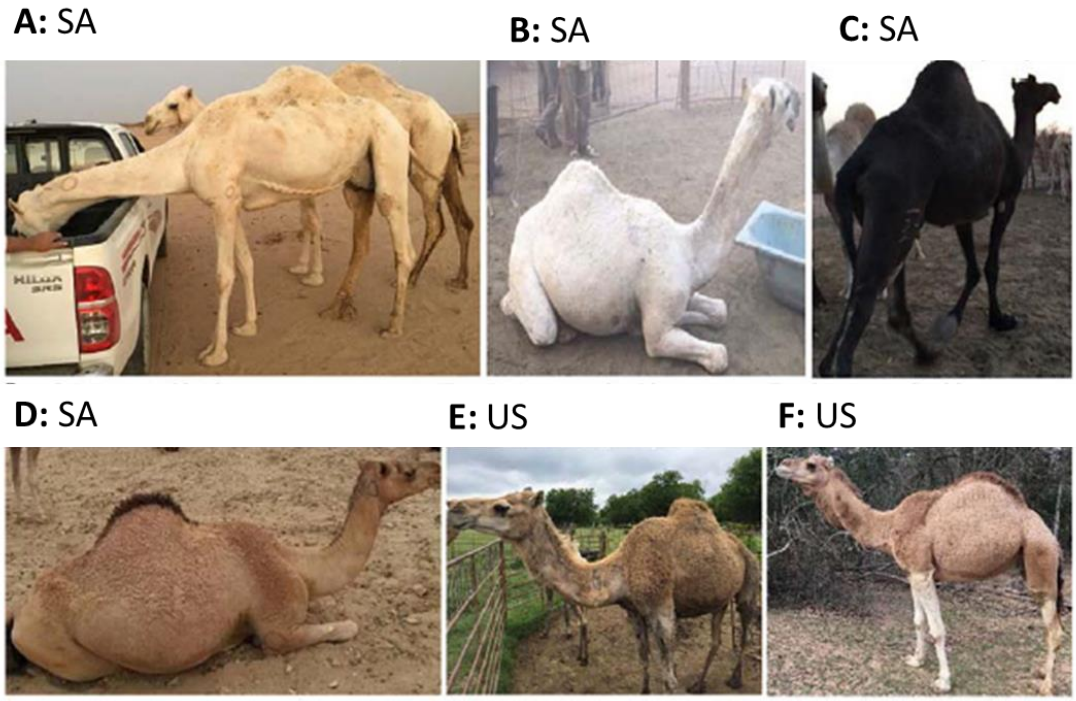

Figure 4. Examples of animals and coat colors used for this study. A. White/cream; **B.** White; **C.** Black; **D.** Reddish brown with dark hump and tail; **E.** Medium brown, **F.** Light brown; SA - Saudi Arabia; US - United States

DNA isolation. Genomic DNA was isolated from peripheral blood lymphocytes using Gentra Puregene Blood Kit (Qiagen) following the manufacturer's protocol, or by standard phenol-chloroform method (J. Sambrook, E. F. Fritsch and T. Maniatis 1989). We evaluated DNA quality and quantity by NanoDrop 2000 spectrophotometer (Thermo Scientific) and by 1% agarose gel electrophoresis.

Primers, PCR and sequencing. We used the available sequence information for the dromedary and alpaca *MC1R*, *ASIP* and *TYRP1* in NCBI , UCSC , and Ensembl genome browsers, or sequences of the Bactrian camel (Wu et al., 2014) and Primer3 software (Untergasser et al., 2012) to design primers. For *ASIP* and *TYRP1*, primers were designed to amplify all exons and exon-intron boundaries. For *MC1R*, primers were designed to amplify overlapping fragments covering the single exon and the 5' UTR. Primer details are presented in Table 5. PCR was conducted in 10 µL reactions containing 50 ng dromedary genomic DNA and 0.5 unit of JumpStart Taq ReadyMix (Sigma Aldrich). For MC1R, primers 5' UTR.1 F and 5' UTR.2 R (Table 5) were combined to amplify the entire 2 kb of the 5'UTR. The PCR products were cleaned using ExoSAP (Affymetrix) and sequenced using BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) and the manufacturer's protocol. Sequencing reactions were cleaned in Spin-50 mini columns (BioMax, Inc) and resolved on 3100 automated sequencer (Applied Biosystems).

Table 5: Primers used for PCR and sequencing of *MC1R*, *ASIP* and *TYRP1*; ***** denotes overgo primers for screening *MC1R* from CHORI-246 BAC library

Gene symbol and primer ID	Region	Forward 5' - 3'	Reverse 5' - 3'	Product size, bp
***MC1R* 5' UTR 1**	5' UTR	TCTCACGCCCTTTGAAGTCT	GCTGACGACAAACCCTTCTC	2007
***MC1R* 5' UTR 2**	5' UTR	ACACACCTTAACGGGACACC	GCAGGAAAGGGTCTTCACTCT	1101
***MC1R* 1**	Exon 1	TCCTCCTCTGTCTCGTCAGC	GTTGGCTGGCACTGTCTCC	918
***MC1R* 2**	Exon 1	GGCGCTGTCTCTTGTGGAG	GACCAGAAGAGACGCAGGAG	912
***MC1R* 3**	Exon 1	CTCCCTGGCAGGACGATG	AGTCCGAGGTGGGTGGTG	92
MC1R*_OV	Exon 1	CATGGTGTCCAGCCTCTGCTCTCT	CACGGCGATAGCACCCAGAGAGCA	n/a
***ASIP* Prom**	Promoter	GGATTTGGGGTCAGTCTGTA	CCCATCCCTTTAGCCTCCTA	2608
***ASIP* Ex1**	Exon 1	GTGTGAGTCAGTGGCAGGAA	AAATTCTGGGTGGGCTAAGG	1130
***ASIP* Ex1b**	Exon 1	ACTTAAGGCAGGCTGGACCT	ATGTGCCCATCCCTTTAGC	1502
***ASIP*ex1seq**	Exon 1	AGACCCTGCATTAAGCTGCTC	Sequencing primer	n/a
***ASIP*ex1seqb**	Exon 1	GCTTTTCTGATAATGAAATA	Sequencing primer	n/a
***ASIP* Ex 2**	Exon 2	CTTCAGTCTCCCTCCCTTCC	GCCAGGTATTTTTCCCTGAG	828
***ASIP* Ex 3**	Exon 3	TCCAGGGCCTTATTGGACTT	CTGGAAAGGCTCAGTTTGCT	953
***ASIP* Ex 4**	Exon 4	ACTGTAAGAGGGCCAGAGCA	TAAAGTAGGGGGCAGCATTG	511
***TYRP1* Ex 1**	Exon 1	AGCACTTTGAAGGTGGGTTG	AGTCAGAAGACTGGAGCATCAA	698
***TYRP1* Ex 2**	Exon 2	AAGAGAGGGAGTGGAAGGGAGA	ATGTGAAATTGCTTGGTCAGTG	650
***TYRP1* Ex 3**	Exon 3	TGAGTTGGGTTTCATTCCTT	CACTTTCTTTTTCCCCTGGA	629
***TYRP1* Ex 4**	Exon 4	TGGACATGGTAACTTGGGTTT	GGCCAGCAACCTAACTTTGA	722
***TYRP1* Ex 5**	Exon 5	GGCCACCAACCATAGGTACA	GACTTCCTGTCTGCCTTTTCA	525
***TYRP1* Ex 6**	Exon 6	CCTGGGCTGCTGTAGTGAA	CTGGGGGCTCTCAACAAACT	500
***TYRP1* Ex 7**	Exon 7	GGAATTAGGAAGTGCCCTGA	AACATGCCCCAAATCTTCAC	595

Sequence analysis and mutation discovery. For initial mutation discovery, we sequenced PCR products of *MC1R, ASIP* and *TYRP1* in 4 white, 4 black and 4 brown dromedaries. Sequences were analyzed for mutations using Sequencher v 5.3 software (Gene Codes Corp.). Effects of single nucleotide changes and indels on protein structure and function were evaluated with Protein Variation Effect Analyzer (PROVEAN) toolkit (Choi & Chan, 2015; Choi, Sims, Murphy, Miller, & Chan, 2012). Amino acid sequences of different species were retrieved from NCBI and Ensembl. Comparative analysis of the MC1R protein across species was performed by aligning amino acid sequences in ClustalW (Thompson, Higgins, & Gibson, 1994). We used Transmembrane Protein Topology with a Hidden Markov Model (Moller, Croning, & Apweiler, 2001) to determine MC1R transmembrane domains and evaluate the effect of SNPs; GeneCluster 2.0 (Reich, Ohm, Angelo, Tamayo, & Mesirov, 2004) for comparative analysis of MC1R across species, and ExPASy webtools (Gasteiger et al., 2003) to translate genomic sequence into protein.

Large cohort genotyping and association analysis. Putative causative mutations in *MC1R* and *ASIP* were further analyzed for genotype-phenotype association by Sanger sequencing the regions in 69 dromedaries (29 white, 17 black, 23 brown). Custom TaqMan™ SNP genotyping assays were designed for *MC1R* and *ASIP* mutations according to manufacturer specification (Applied Biosystems) (Table 6), and used for genotyping all 188 dromedaries. We used CFX-96 Real Time-PCR machine (Bio Rad) and corresponding software for PCR amplifications, genotyping and allelic discrimination. The thermal conditions were: priming at 60℃ for 1 min, initial denaturation at 95℃ for 10 min, 40 cycles of 92℃ for 15 sec, annealing at primer-specific t℃, extension for 1 minute at 60℃, followed by a final extension at 65℃. The 8

µL reactions contained 0.208 µL of TaqMan™ assay, 30 ng template DNA and 4.2 µL of ABI TaqMan Universal Master mix, no UNG (Applied Biosystems).

Table 6: TaqMan assays for genotyping *MC1R* g901C>T and *ASIP* g.174495T>Del

Primer/probe	**5'-3'**
***MC1R*-forward**	CTCATCATCTGCAACTCCATCGT
***MC1R*-reverse**	CAGCACCTCTTGGAGTGTCTTC
***MC1R*_VICprobe**	ATGCCTTCCGCAGCCA
***MC1R*_FAMprobe**	CTATGCCTTCTGCAGCCA
***ASIP*-forward**	CCACTCAGATATCCCAGGATGGA
***ASIP*-reverse**	GCTGTAGGCATTGAGGAAGCA
***ASIP*_VICprobe**	CCTCTTCCTAGCTACCC
***ASIP*_FAMprobe**	CCTCTTCCAACTACCC

Statistical Analysis. We conducted contingency analysis with JMP Program v12 (JMP®, Version 13. SAS Institute Inc., Cary, NC, 1989-2007) to examine the relationship between color phenotypes and genotypes at each variable site. Genotype-phenotype associations were determined by the standard Chi-square test.

Chromosome preparations. Alpaca, dromedary and Bactrian camel chromosome slides were prepared from methanol : acetic acid (3:1)-fixed cell suspensions available in the depository of the Molecular Cytogenetics laboratory at Texas A&M University. All cell suspensions originated from normal individuals with normal karyotypes.

Fluorescence in situ hybridization (FISH). We used alpaca CHORI-246 genomic Bacterial Artificial Chromosome (BAC) library to obtain probes for FISH. BAC clones containing *ASIP* and *TYRP1* were previously identified and mapped in the alpaca (F. Avila, Baily, et al., 2014). To obtain BACs for MC1R, we screened CHORI-246 filters with *MC1R*-specific radioactively labeled ([32P] dATP/dCTP) overgo primers (Table 6) as described by Avila et al. (2014b) and Mendoza et al. (Mendoza et al., 2019). The final BACs containing *MC1R* were further verified by PCR with *MC1R* exon primers (Table 5). BAC DNA was isolated with Plasmid Mini Kit (Qiagen) according to the manufacturer's protocol. Probe labeling, hybridization and signal detection were conducted according to standard protocols (Raudsepp & Chowdhary, 2008). Because of difficulties to unambiguously identify camelid chromosomes by conventional cytogenetic methods (F. Avila, Das, et al., 2014), BACs containing the three genes were co-hybridized with a differently labeled reference gene from the alpaca cytogenetic map (F. Avila, Baily, et al., 2014). Composite information about the BACs used for comparative FISH mapping is presented in Table 7. Images for at least 10 metaphases for each experiment were captured and

analyzed using a Zeiss Axioplan 2 fluorescence microscope, equipped with the Isis Version 5.2 (MetaSystems GmbH) software.

Table 7: Comparative cytogenetic mapping of *ASIP, MC1R* and *TYRP1*. Details about alpaca BAC clones, corresponding genes and cytogenetic locations; * denotes reference BACs/genes for chromosome identification

CHORI-246 BAC	**Gene symbol**	**Camelid chr.**	**Alpaca**	**Dromedary**	**Bactrian**	**Human chr.**
018C13	*ASIP*	19q12	Avila et al. 2014a	This study	This study	20q11.2-q12
125P19*	*EDN3**	19q23	Avila et al. 2014a	This study	This study	20q13.2-q13.3
166N17	*MC1R*	21q15	This study	This study	This study	16q23
128F16*	*MYOC**	21q13	Avila et al. 2014a	This study	This study	1q23-q24
129N17	*TYRP1*	4q21dist-q22	Avila et al. 2014a	This study	This study	9p23
135B22*	*MRPL41**	4q36	Avila et al. 2014a	This study	This study	9q34.3

Results

Mutation discovery and association analysis of *MC1R*

The initial sequence analysis of 12 individuals (4 white, 4 black and 4 brown) identified 7 sequence variants inside and around *MC1R* (Table 8). All variants were SNPs including the previously reported c.901C>T (p.Arg301Cys) missense mutation in the *MC1R* coding region (Almathen et al., 2018). Three SNPs were in the promoter region, two in 5'UTR and one in 3'UTR. The c.901C>T missense mutation was genotyped in large cohorts by Sanger sequencing (n=68) and by TaqMan™ genotyping (n=188) showing that this mutation is significantly associated ($P<0.0001$) with the white color (Table 9), thus confirming the findings of Almathen et al. (2018). To evaluate the possible effect of the p.301R>C mutation on MC1R function, we constructed transmembrane protein topology and showed that the amino acid change affects the last of the 7 transmembrane domains (Fig. 5). We also aligned the amino acid sequences of the MC1R last transmembrane domain in diverse mammalian and vertebrate species and showed that arginine at this position is highly conserved across species (Fig. 6), suggesting its importance for MC1R normal function.

Table 8: Sequence polymorphisms in *ASIP*, *MC1R* and *TYRP1*. Sequence variants were discovered by sequencing the three genes in 4 white, 4 black and 4 brown dromedaries. Sequence positions correspond to dromedary whole genome assembly: GCA_000767585.1 PRJNA234474_Ca_dromedarius_V1.0.; *MC1R* scaffold ID: NW_011592664.1, *ASIP* scaffold ID: NW_011591043.1 and *TYRP1* scaffold ID: NW_011591511.1; Numbers in columns White, Black, Brown denote the number of animals with the corresponding genotype; ORF- open reading frame; D – deletion; ***P-value*** **for genotype-phenotype association was determined by contingency analysis in JMP**

Gene symbol	**Location in the gene**	**Variant**	**Reference**	**Effect on protein**	**Phenotype-genotype**			***P-value***
					White	**Black**	**Brown**	
MC1R	Promoter (1802 bp from ORF	g.535236C>T	This study	noncoding	4CC	4CT	4CT	0.0005
MC1R	Promoter (557 bp from ORF)	g.536482G>A	This study	noncoding	2GG, 2GA	1GG, 2GA, 1AA	2GG, 2GA	0.6208
MC1R	Promoter (420 bp from ORF)	g.536623G>A	This study	noncoding	4GG	3GG, 1GA	3GG, 1GA	0.4033
MC1R	5'UTR	g.537027G>A	This study	noncoding	4GG	3GG, 1GA	3GG, 1GA	0.4033
MC1R	5'UTR	g.537028A>T	This study	noncoding	4AA	3AA, 1AT	3AA, 1AT	0.4033
MC1R	Exon	g.537961C>T; c.901C>T	Almathen et al. 2018	Missense; p.Arg301Cys	3CT, 1TT	4CC	4CC	0.0042
MC1R	3'UTR	g.538058G>A	This study	noncoding	4GG	2GG, 2GA	4GG	0.0718
ASIP	Exon 2	g.174495T_del; c.23T_del	Almathen et al. 2018	frameshift; p.24X	4TD	4DD	2TT,2TD	0.0009

Table 9: Genotype frequencies of *MC1R* c.901C>T missense mutation in a large study cohort (n=188). The mutation is significantly (*P*<0.0001) associated with white coat color

Genotypes	Frequency (count) in dromedary color groups		
	White (n=53)	**Black (n=38)**	**Brown (n=97)**
CC	0.037 (7)	0.20 (38)	0.41 (78)
CT	0.13 (25)	0	0.085 (16)
TT	0.11 (21)	0	0.016 (3)

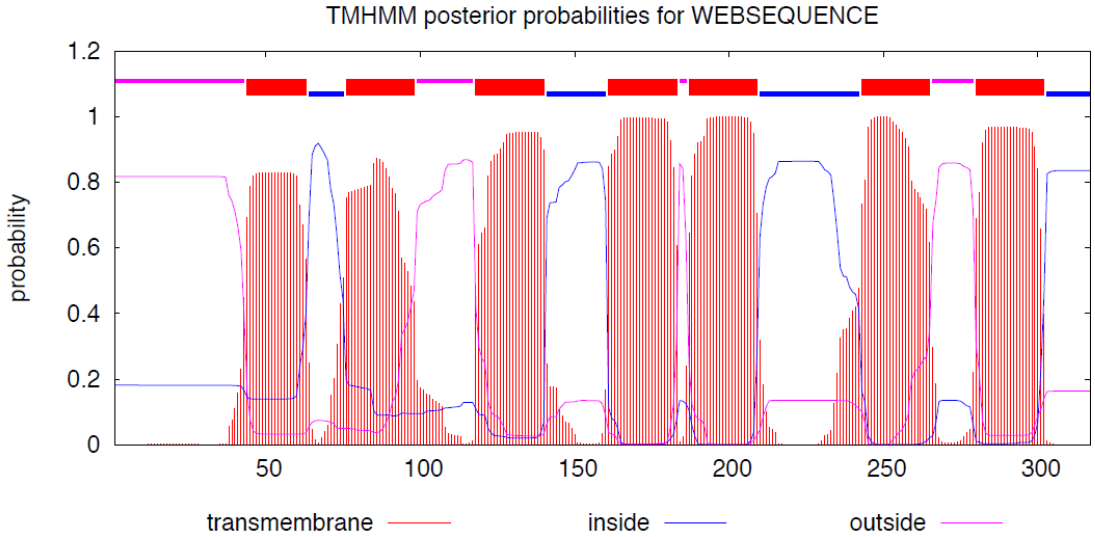

Figure 5. MC1R protein functional domains. The seven MC1R transmembrane domains and the position of p.301R>C mutation (arrow); y-axis: the probability of the amino acid sequences to be cytoplasmic (blue), extracellular (magenta), or part of the transmembrane helix (orange); x-axis: amino acid sequence. We used Transmembrane Protein Topology with a Hidden Markov Model (Moller et al., 2001)

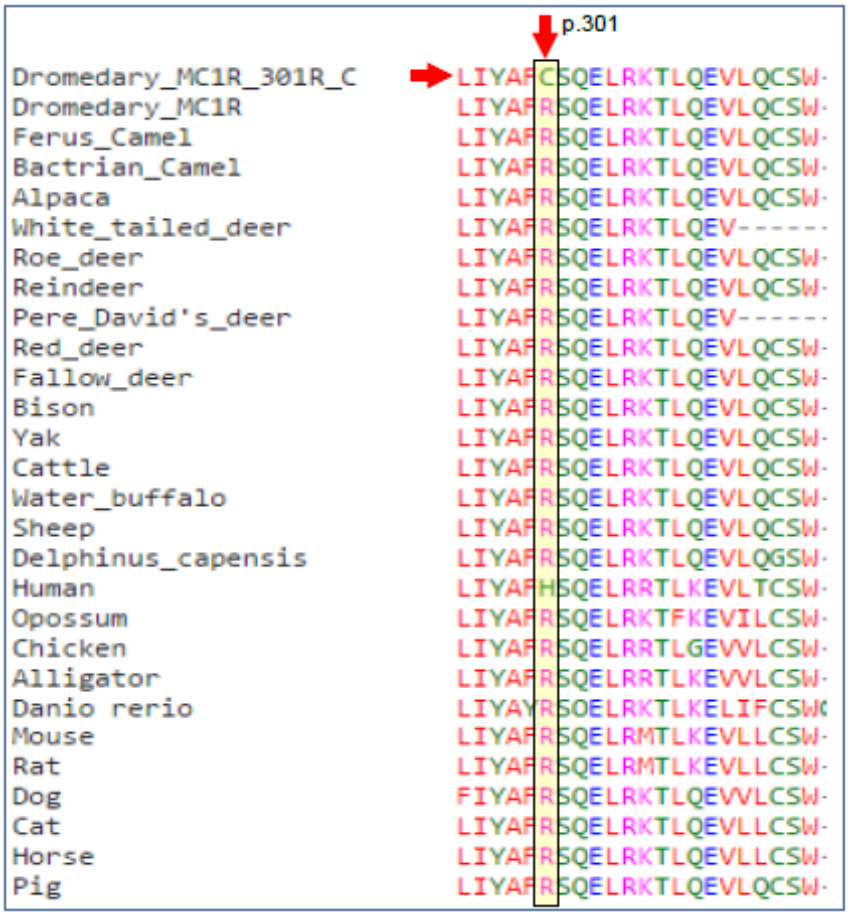

Figure 6: MC1R comparative. Comparative alignment of MC1R transmembrane domain 7 amino acid sequences in diverse mammalian and vertebrate species. The p.301 position is indicated by a vertical arrow and highlighted; horizontal arrow shows the p.301R>C mutation in the dromedary. Note that Arginine is highly conserved across species, except in humans who have Histidine, which is another positively charged amino acid.

The initial analysis also indicated that the SNP g.538058G>A in *MC1R* 3'UTR may be associated with color phenotype because genotype GA was present only in black dromedaries (Table 8). Large cohort (n=68) genotyping by sequencing confirmed this and showed that GA genotype was more frequent ($P<0.0004$; Table 10) in black animals. Notably, we did not find dromedaries homozygous for the A-allele (AA) at this site. Furthermore, the brown dromedaries (n=31) with dark to black wool on the hump

and black hair on the tails did not have A-allele and were all homozygous for the G-allele (GG) (Table 10)

Table 10: Genotype frequencies of *MC1R* 3'UTR variant g.538058G>A in a large study cohort (n=68). The SNP is significantly (*P*<0.0004) associated with black coat color

Genotypes	**Frequency (count) in dromedary color groups**		
	White (n=22)	**Black (n=15)**	**Brown (n=31)**
GG	0.29 (20)	0.13 (9)	0.46 (31)
GA	0.029 (2)	0.088 (6)	0
AA	0	0	0

Mutation discovery and association analysis of *ASIP*

We identified three sequence variants in the four exons of the dromedary *ASIP* gene: – two in exon 2 and one in exon 4 (Table 8). A single nucleotide deletion in exon 2 (g.174495T_del; c.23T_del) combined with a SNP two base-pairs later (g.174497A>G; c.25A>G; Table 8) caused a shift in the reading frame, an insertion of a premature stop at codon 24, and truncated protein (Fig. 7). All black dromedaries in the discovery cohort were homozygous for the frameshift deletion. Therefore, we genotyped the frameshift mutation in large dromedary cohorts by sequencing (n=68) and TaqMan™ assay (n=188). The results showed that the frameshift mutation in exon 2 is significantly associated (P<0.0001) with black coat color (Table 11), consistent with the previous

findings (Almathen et al., 2018). However, in the large study cohort, one black animal did not have the frameshift deletion and three black dromedaries were heterozygous for it (Table 11). In these four animals, we analyzed ASIP sequence further and discovered that two dromedaries were heterozygous for another frameshift mutation in exon 4 at g.178388C_del (Table 11). The mutation shifted normal stop at codon 133 to codon 254, resulting in 120 amino acids longer polypeptide (Fig. 7). However, the other two animals did not have this deletion and, overall, we were not able to associate exon 4 mutation with black color in our study cohort.

Table 11: Genotype frequencies of *ASIP* exon 2 g.174495T_del (D) nonsense mutation in a large study cohort (n=188). The mutation is significantly (P<0.0001) associated with black coat color

Genotypes	Frequency (count) in dromedary color groups		
	White (n=53)	Black (n=38)	Brown (n=97)
TT	0.12 (23)	0.005 (1)	0.20 (37)
TD	0.13 (25)	0.016 (3)	0.18 (33)
DD	0.03 (5)	0.18 (34)	0.14 (27)

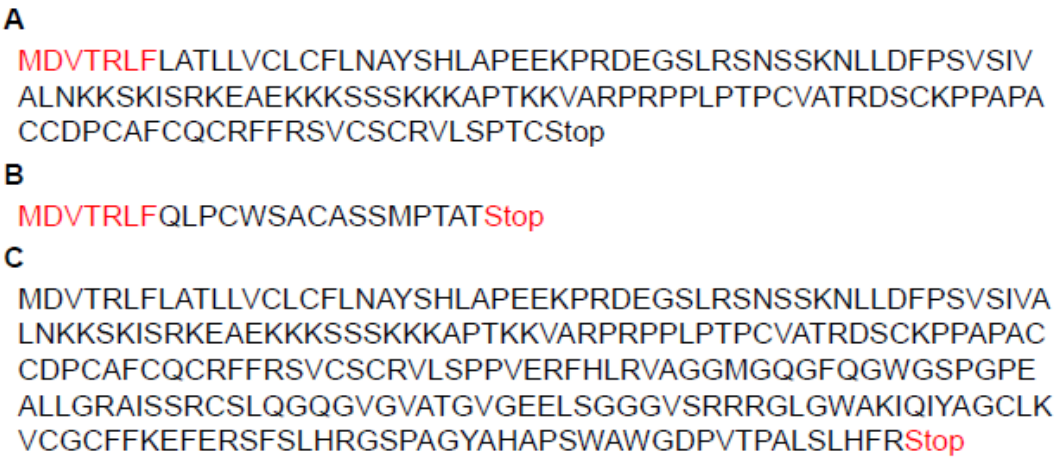

Figure 7. The effect of frameshift mutations on ASIP polypeptide. A. Normal ASIP polypeptide with 133 amino acids and stop at codon 134; **B.** Truncated ASIP protein with 24 amino acids and stop at codon 25 due to frameshift mutation in exon2; **C.** Abnormally long polypeptide with 253 amino acids due to a frameshift mutation in exon 4. Amino acids in red font in A. and B. are before frameshift and, thus shared between the normal and truncated ASIP.

As noted above, we conducted large cohort genotyping for *MC1R* c.901C>T and ASIP c.23T_del mutations in 68 dromedaries both by Sanger sequencing and TaqMan™ assays. The results between the two methods were in full agreement, suggesting that the high throughput, faster and cheaper TaqMan™ assay should be the method of choice for any further genotyping of large numbers of additional animals.

Mutation discovery in *TYRP1*

Sequence analysis of the 7 exons and exon-intron boundaries of the *TYRP1* gene in the discovery cohort of 12 dromedaries, identified 5 sequence variants: 4 SNPs and one insertion (Table 8). However, all variants were in non-coding regions (introns and 3'UTR) and not associated with dromedary color phenotypes. Therefore, we did not conduct any large cohort genotyping for *TYRP1*.

Comparative FISH mapping

We mapped *TYRP1*, *ASIP*, and *MC1R* by FISH to metaphase chromosomes in the alpaca, dromedary and Bactrian camel (Fig. 8). For unambiguous chromosome identification, we used chromosome-specific reference markers from the alpaca cytogenetic map (F. Avila, Baily, et al., 2014). The *TYRP1* and *ASIP* genes were previously FISH mapped to alpaca chromosomes 4 and 19, respectively (F. Avila, Baily, et al., 2014). Here we mapped *TYRP1* to chr4 and *ASIP* to chr19 in both camel species (Fig. 8A, B). The results are in agreement with karyotype conservation across Old and New World camelids (Bianchi, Larramendy, Bianchi, & Cortes, 1986; T.D. Bunch, Foote, & Maciulis, 1985) and consistent with human-dromedary Zoo-FISH data on conserved synteny segments between these species (Balmus et al., 2007). The *MC1R* gene has not been chromosomally assigned in any camelid genome. Here we mapped *MC1R* to the very terminal region in chr21q in the alpaca, dromedary and Bactrian camel (Fig. 8C) and revealed a hitherto unknown conserved synteny block between camelid chr21 and HSA16q.

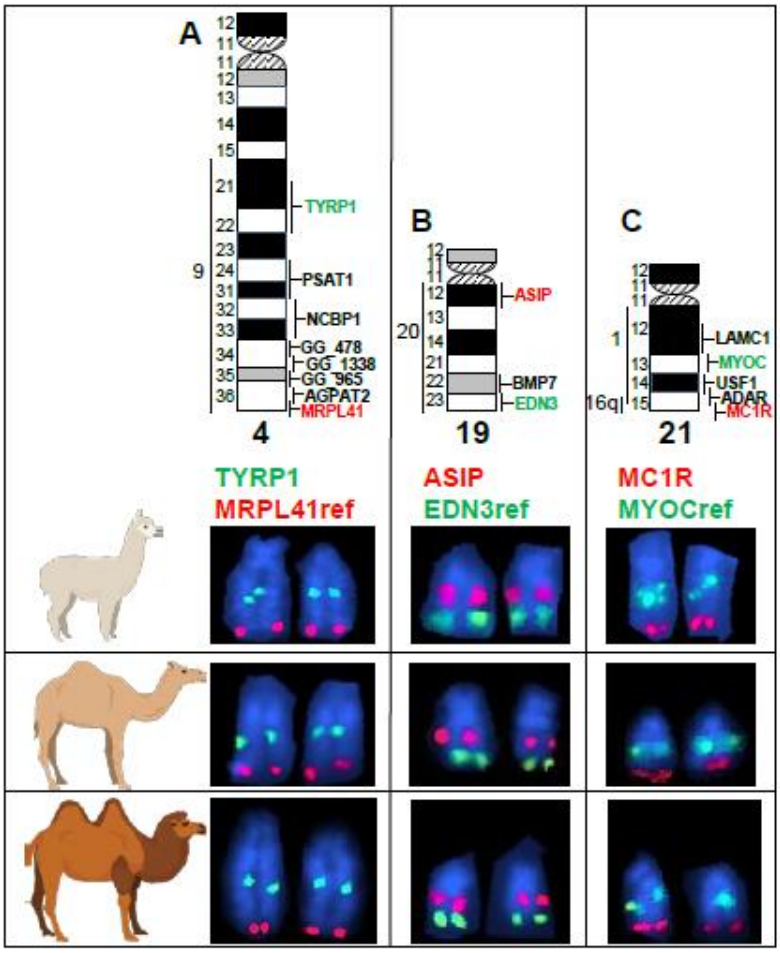

Figure 8. Comparative FISH mapping. Comparative mapping of *TYRP1* (**A**), *ASIP* (**B**), and *MC1R* (**C**) in alpaca, dromedary and Bactrian camel chr4, chr19 and chr21, respectively. Chromosome ideograms with all mapped markers (Avila et al. 2014a) are shown at the top. Vertical lines with numbers to the left of chromosome ideograms indicate homology segments to human chromosomes. Ref – reference gene for chromosome identification. Green and red font colors for coat color genes and reference genes correspond to the green and red FISH signals in partial microscope images below ideograms.

Discussion

Here we validated the recently published mutations for white and black/dark brown coat color in dromedaries (Almathen et al., 2018) using independent dromedary

populations of US and Saudi Arabian origin. In addition, we designed for both mutations TaqMan assays and confirmed their accuracy and efficiency for large cohort genotyping. Overall, our results are consistent with the recently published data, but also refine and expand it.

The likely causative mutation for the white color in *MC1R* c.901C>T in dromedaries was, at the first sight intriguing because, according to published references (Almathen et al., 2018; Rieder et al., 2001), a mutation at the same position (c.901C>T) is responsible for recessive chestnut coat color in horses. This poses a question why the same missense mutation results in a depigmentation phenotype in the dromedary, but a pheomelanic phenotype in horses. However, closer inspection of the original publication for the horse chestnut mutation (Marklund, Moller, Sandberg, & Andersson, 1996); reviewed by (L. Andersson, 2003) reveals that the horse chestnut and dromedary white mutations are different. The horse chestnut is due to p.Ser83Phe, which affects MC1R second transmembrane domain (Marklund et al., 1996), while the dromedary mutation p.Arg301Cys is in the last (seventh) transmembrane domain (Fig. 5).

The dromedary mutation, however, shares functional and phenotypic similarity to recently reported *MC1R* sequence variants in Australian cattle dogs and Alaskan and Siberian huskies (Durig et al., 2018). Cream color in Australian cattle dogs is associated with a combination of c.916C>T (p.Arg306Ter) and a promoter variant affecting MITF binding site. White huskies, on the other hand, are homozygous for a deletion c.816-delCT. Even though causative sequence variants are different in Australian cattle dogs, huskies and dromedaries, they share essential similarities: all occur in the last

transmembrane domain, negatively affect MC1R function and result in depigmentation phenotypes. The mutations in cream-colored Australian cattle dogs cause downregulation of MC1R transcription, while MC1R in white huskies has lost the last transmembrane domain and the cytoplasmic C-terminal tail (Durig et al., 2018). Though no functional data are available for white dromedaries, we theorize based on the predicted effect of p.Arg301Cys on the last transmembrane domain (Fig. 5) and the resulting white phenotype that the mutation is loss-of-function. Functional importance of this portion of the MC1R protein is illustrated by highly conserved sequence of 17 amino acids (p.296-312) across diverse mammalian species (Fig. 6). Notably, the dromedary differs from other mammals at p.301 because a white and not a wild-type animal was used for the reference sequence.

In contrast to huskies where white color is a recessive trait (Durig et al., 2018), the dromedary white mutation is dominant because heterozygosity for the T-allele at c901C>T is sufficient for the white phenotype (Table 9). Therefore, we suggest that the *MC1R* mutation in white dromedaries has dominant negative effect, i.e., it alters the function of the wild type C-allele and has dominant or semi-dominant phenotype. Similar dominant negative effect on wild-type MC1R receptor cell surface expression or wild-type MC1R cAMP signaling has been described for several *MC1R* sequence variants in humans (Beaumont et al., 2007).

Another observation about dromedary *MC1R*, as also noted by Almathen and colleagues (2018), is the low level of sequence variation (just c901C>T) in the coding region, contrasting the 21 sequence variants found in the alpaca *MC1R* (Feeley &

Munyard, 2009). However, a recent study of *MC1R* sequence variants across all four South American camelids (vicugna, guanaco, llama and alpaca) suggests that variation in alpacas is the result of human selection for a variety of fiber colors, whereas in wild South American camelids (guanacos and free living vicugnas), there is a selection against non-synonymous substitutions in *MC1R* (Marin et al., 2018). Likewise, there is low sequence variation of *MC1R* in wild pigs, but many more variants in domestic pig breeds as a result of human selection (L. Andersson, 2003). Thus, we suggest that low sequence variation of *MC1R* in dromedaries is because human selection for coat color in this species is a more recent event in course of domestication.

On the other hand, sequence variants are present immediately outside the dromedary *MC1R* coding region, in 5'- and 3'-UTRs and in the promoter (Table 8). Whether any of these have regulatory roles in shaping pigmentation phenotypes, is a subject of future studies. This also applies to the 3'UTR variant g.538058G>A (Table 8), which showed association ($P<0.0004$) with black coat color (Table 10). Though, it is also possible that the statistical significance may be influenced by relatedness between black animals.

The causative mutation for black coat color, as reported earlier (Almathen et al., 2018) and confirmed in this study (Table 11), is a frameshift deletion in *ASIP* exon 2, resulting in premature stop codon and truncated protein (Fig. 7). Like in previous study (Almathen et al., 2018), we also observed a synonymous SNP 2 bp after the frameshift deletion (Table 8), but did not conduct association analysis because it was irrelevant for the premature stop codon. Similar, though not identical, loss-of-function mutations in

ASIP underlie recessive black color in several domestic and wild species. For example, in alpacas (Feeley et al., 2011), sheep (Norris & Whan, 2008; Royo et al., 2008), Iranian Markhoz goats (Nazari-Ghadikolaei et al., 2018), donkeys (Abitbol, Legrand, & Tiret, 2015), horses (Rieder et al., 2001), dogs (Kerns et al., 2004), cats (Eizirik et al., 2003) and impala antelope (Miller, Guthrie, & Harper, 2016). Like in these species, we are confident that the black color in the dromedary is a recessive trait because the majority (34/38) of black dromedaries in this study were homozygous for the deletion (Table 11). However, 4 black animals in our study cohort did not follow this pattern (Table 11). Two of these carried another frameshift deletion in *ASIP* exon 4, resulting in abnormally long and likely non-functional ASIP protein (Fig. 7). We suggest that the second frameshift deletion may be causative for black color in the absence of the first deletion, though it was not possible to conduct association analysis with just 2 individuals. Of the remaining two black dromedaries, one was heterozygous for the exon 2 deletion and the other had no mutations in *ASIP*. This is similar to observations in alpacas where homozygous recessive loss-of-function mutations in *ASIP* explain the majority but not all cases of the black phenotype (Feeley et al., 2011). Thus, like in alpacas, black coat color in dromedaries may be influenced by additional regulatory mutations and MC1R interactions with ASIP and α-melanocyte stimulating hormone (α-MSH). One should also consider possible errors in phenotyping.

We investigated the *TYRP1* gene as a possible contributor to various shades of brown coat color in the dromedary. The gene encodes for an important enzyme for the synthesis of eumelanin (del Marmol & Beermann, 1996) and *TYRP1* mutations are

associated with brown or chocolate coat color on black genetic background in many mammals and other vertebrates (J. Li et al., 2018). However, all *TYRP1* variants found in this study, were in noncoding regions (Table 8) and we did not detect the two SNPs in dromedary *TYRP1* exon1 as reported by a prior study (Almathen et al., 2018). Nevertheless, both the noncoding SNPs and the exon 1 SNPs were not associated with any color phenotypes. Likewise, no candidate coat color mutations have been detected in alpaca *TYRP1* (Cransberg & Munyard, 2011). Despite these findings, *TYRP1* remains an important candidate gene for color phenotypes in camelids and should be included in future studies.

Finally, we comparatively FISH mapped the three coat color genes in three camelid species – the alpaca, the dromedary and the Bactrian camel. In agreement with the known conservation of camelid karyotypes (T.D. Bunch et al., 1985; K. M. Taylor, Hungerford, Snyder, & Ulmer, 1968) and prior mapping of *TYRP1* and *ASIP* in alpacas (F. Avila, Baily, et al., 2014), the genes mapped to the same cytogenetic location in the same chromosomes in all species: *TYRP1* to chr4q21-q22, *ASIP* to chr19q12, and *MC1R* to chr21qter (Fig. 8). While the locations of *TYRP1* and *ASIP* were in good agreement with human-dromedary Zoo-FISH (Balmus et al., 2007), mapping *MC1R* to chr21 came as a surprise. This is because camelid chr21 shares known conserved synteny with part of HSA1q only (F. Avila, Baily, et al., 2014; Balmus et al., 2007). Since human *MC1R* is located very terminal in the long arm of chr16 (HSA16q24.3; 89.9 Mb), we anticipated mapping *MC1R* to camelid chr9, which is homologous to HSA16q (F. Avila, Baily, et al., 2014; Balmus et al., 2007). Furthermore, camelid chr9 shares also homology with

HSA19q, and HSA16q/HSA19q correspond to an ancestral eutherian synteny combination, which has been conserved in many eutherian karyotypes (Chowdhary, Raudsepp, Fronicke, & Scherthan, 1998; M. A. Ferguson-Smith & V. Trifonov, 2007). Our findings indicate that this ancestral synteny combination has undergone rearrangements during camelid karyotype evolution, so that a segment homologous to HSA16q containing *MC1R* has become of part of camelid chr21 and shares synteny with sequences corresponding to HSA1q. Inspection of the current dromedary genome assembly PRJNA234474_Ca_dromedarius_V1.0 scaffolds confirmed FISH results for *MC1R* and showed that sequences corresponding to HSA1q: 145-147 Mb and HSA16q: 85-90 Mb are together in dromedary scaffold NW_011592664.1 (https://www.ncbi.nlm.nih.gov/gene/105104349). Therefore, cytogenetic mapping of *MC1R* in camelids revealed a novel human-camelid synteny segment, confirmed sequence assembly of scaffold NW_011591415.1, and anchored alpaca, dromedary and Bactrian camel scaffolds containing *TYRP1*, *ASIP* and *MC1R* to chromosomes.

CHAPTER IV

POPULATION DIFFERENTIATION IN SAUDI ARABIAN DROMEDARIES

Introduction

Since domestication, the genetic diversity in domestic species has been extensively shaped by human selection, which in some instances, has led to increased distances between closely-related populations (Boichard, Ducrocq, Croiseau, & Fritz, 2016; Groeneveld et al., 2010). As a result, this has led to increased genetic variation between subpopulations/breeds, which can be traced by estimating nucleotide diversity across the genome to identify regions under selection potentially harboring genes associated with traits under selection (L. Andersson, 2001; L. Andersson, 2012). This methodology has been used in many species, including horses, sheep, and goats to identify regions with potential in-process or complete selective sweeps (Guo et al., 2018; Ruiz-Larranaga et al., 2018; C. Zhang et al., 2018). In these studies, association between phenotypes under selection and genetic loci influencing these phenotypes were identified. To date, no similar studies have been conducted for comparing dromedary subpopulations.

Thanks to the decreasing cost of whole genome sequencing (WGS), the improvement of bioinformatics tools, and the availability of the dromedary reference genome (Wu et al., 2014), it is now possible to compare the genomes of dromedary subpopulations based on traits that have been of human interest and potentially under selection. In order to achieve accurate results, it is important to consider that the

comparison between traits could be moderate to extreme in their phenotypic differences and influenced by different geographic locations of the subpopulations.

Selective sweeps are the reduction of genetic diversity and increasing the homozygosity in a population as a result of positive selection (Messer & Neher, 2012; Smith & Haigh, 1974). This will lead to change allele frequencies at closely linked loci (Smith & Haigh, 1974). Therefore, phenotypes that improve ability to survive and reproduce become more common, whereas the poor phenotype for survival and reproduction will be reduced and may disappear (Palladino, Spencer, Cummings, & Klug, 2015).

Selection in dromedaries and classification of breeds has not been systematic and varies between countries. Therefore, breeds can be defined by color or human tribe names. It was reported that there are 48 different dromedary breeds in the world, which can be classified into 8 subgroups. Saudi Arabia has 12 breeds that were classified based mainly on coat color and body measurements (Abdallah & Faye, 2012; Bernard Faye, 2015). Saudi Arabian breeds are distributed in certain locations, however, these populations are interbred and not isolated (Bernard Faye, 2015). Even though 12 breeds have been documented, only three subtypes were identified using the rapid amplified polymorphic DNA (RAPD) technique; demonstrating that white dromedary (Waddah) is the least similar to the other two subtypes (black and brown) (Abdulaziz M Al-Swailem, Al-Busadah, Shehata, Al-Anazi, & Askari, 2007). Further, it was documented that there is gene flow between dromedary breeds in Saudi Arabia based on microsatellite analysis (Mahmoud, Alshaikh, Aljumaah, & Mohammed, 2012). A genetic parentage testing

using 17 polymorphic microsatellite markers was reported for racing dromedary (Spencer, Wilson, & Tinson, 2010). As mentioned above, there are 12 reported dromedary breeds/subpopulations in Saudi Arabia, and currently there is no information about the degree of genetic differentiation between them.

In this this study, we investigate genetic differentiation between three Saudi Arabia dromedary populations. The populations are phenotypically defined by color: white (the Waddah breed), black (the Majaheem breed) and brown (the Saheli breed) (Figure 9). For clarity, from here the breeds are named by coat colors instead of the breed name. Besides the coat color, the three breeds have also other phenotypic differences. For example, the white dromedary is a good dairy animal with medium to large body size, and many Saudis consider it prestigious to own a white camel (Abdallah & Faye, 2012). On the other hand, the black breed is valued for its large body size and the greatest meat and milk production. The brown breed is small to medium in size, known for good quality milk production, but is overall a multipurpose animal (Abdallah & Faye, 2012). Here, we conducted WGS of pooled samples of 27 white, 17 black, and 31 brown dromedaries to identify the signatures of selection leading to identify genetic regions that under selection associated with traits such as performance and/or production.

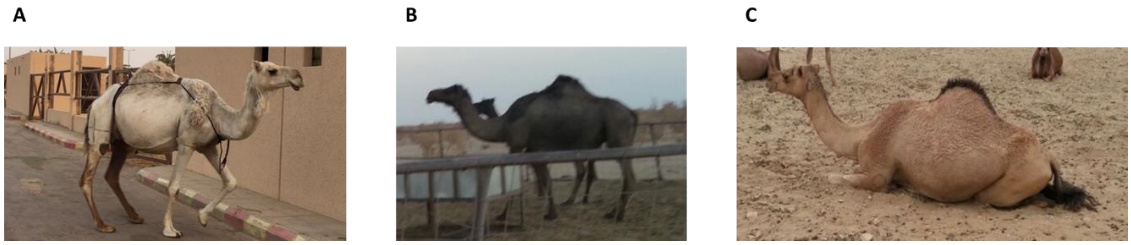

Figure 9. Examples of animals and coat colors used for this study. A. White/cream**; B.** Black**; C.** Brown

Materials and Methods

Animals and phenotypes. Blood was collected by jugular venipuncture from 75 dromedaries from Saudi Arabia representing three different breeds and color phenotypes: white/cream (n=27) from the middle (Waddah); black (n=17) from the south (Majaheem); and brown coat color (n=31) from the west coast (Saheli). Coat color phenotypes were determined by visual inspection and recorded in written notes and/or photos.

DNA isolation. Genomic DNA was extracted from peripheral blood leukocytes using Gentra Puregene Blood Kit (Qiagen) following the manufacturer's protocol or by standard phenol-chloroform method (J. Sambrook, Fritsch, & Maniatis, 1989). DNA quality and quantity were evaluated by NanoDrop 2000 (Thermo Scientific) spectrometry and agarose gel electrophoresis.

Genome Sequencing, alignment and SNP calling. We pooled DNA samples based on color phenotypes. Pooled samples were then prepared for Illumina TruSeq PCR-free library. These libraries were sequenced 100 bp paired-end on the HiSeq2500 platform, generating 15X average genome coverage per library. After quality trimming, reads were mapped to the dromedary reference genome Ca_dromedarius_V1.0 (GenBank assembly accession no. GCA_000767585) using the Burrows-Wheeler alignment tools (BWA) MEM with default parameters as previously described (Heng Li & Durbin, 2009). We used Samtools (version0.1.19-44428cd) to preform SNP calling by creating mpileup files for each breed and across all breeds (H. Li, 2011). The mpileup files were then used to create synchronized files that contain allele frequencies for each position in every pool using Popoolation and Popoolation2 (Kofler, Orozco-terWengel, et al., 2011; Kofler, Pandey, & Schlotterer, 2011). Allele frequencies in the synchronized files were called after filtering for base quality was applied as previously described (Kofler, Orozco-terWengel, et al., 2011; Kofler, Pandey, et al., 2011). Figure 10 summarizes strategies used in this study.

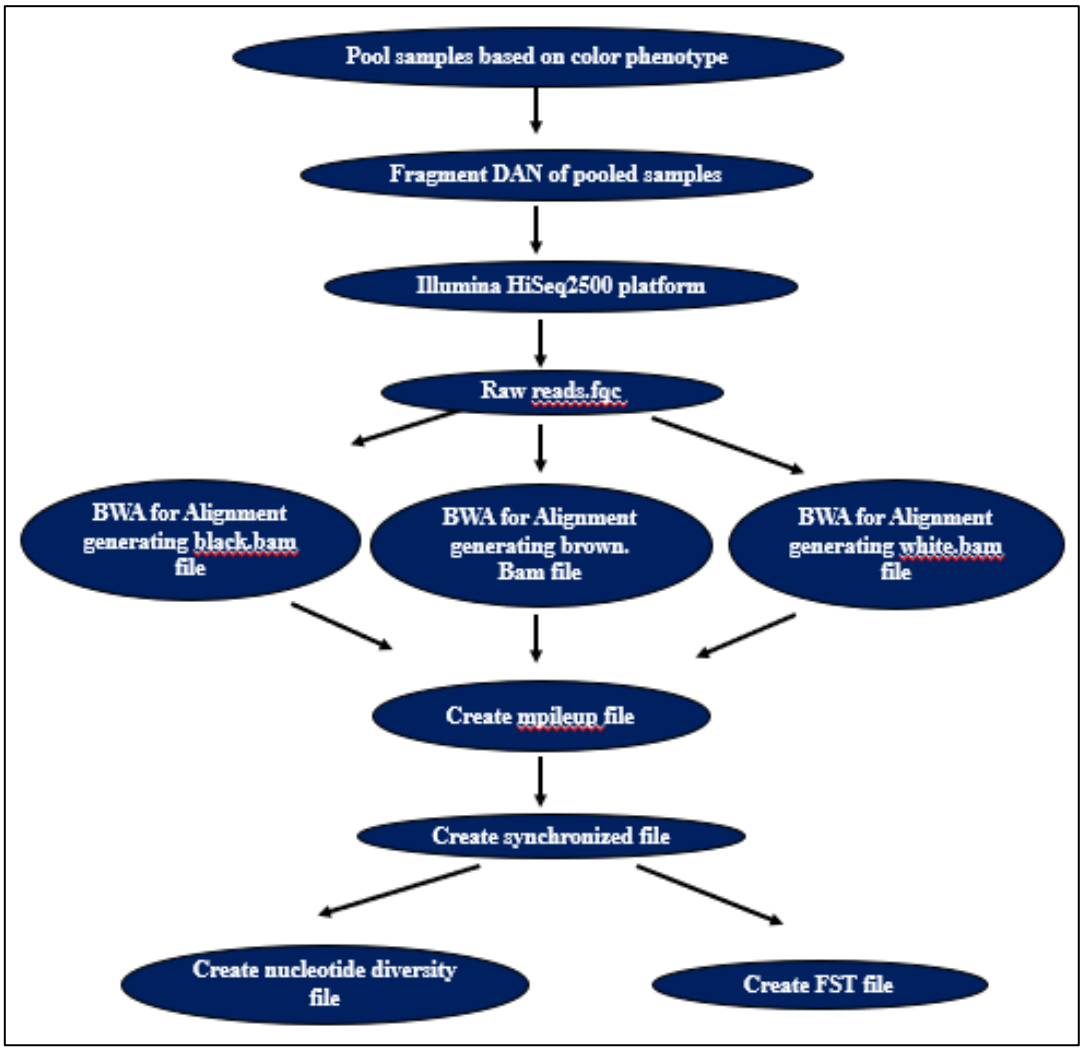

Figure 10: Data analysis strategies in this study.

Identification of directional selection regions

FST and divergence. In order to identify selective sweep regions, which are genomic regions with variations that are associated with selective breeding between breeds, we preformed fixation index (F_{ST}) analysis for the three breeds. F_{ST} is used to measure population differentiation between subpopulations by comparing the total number of variants found in a subpopulation relative to the total variants in the entire

population. The synchronized file across breeds was used to estimate F_{ST} using a sliding window approach (window size 25 kb, step 1 kb) using Popoolation2 tools (Kofler, Pandey, et al., 2011). The F_{ST} file contains columns include scaffold ID, position, quality and F_{ST} values. The F_{ST} values, which range from 0 to 1 where high F_{ST} values indicate significant differentiation, were z transformed according to the formula ($zF_{ST} = F_{ST} – mF_{ST} / s$), where mF_{ST} is the mean of all F_{ST} values and s is the standard deviation of F_{ST} values. The transformation was done with Matlab software v13 MATLAB and Statistics Toolbox Release 2017b, The MathWorks, Inc., Natick, Massachusetts, United States. R i386 v3.5.1; a custom-made script was used to create plots of zF_{ST} values for all pairwise comparisons. Genes, within windows that showed signatures of selection, were identified using zF_{ST} score of 8 or greater and displayed using R custom-made script.

Heterozygosity and nucleotide diversity estimates from the pooled sequence data. Popoolation2 (Kofler, Pandey, et al., 2011) was also used to estimate nucleotide diversity across breeds, creating two files: the first file contained allele frequency for the major and minor alleles for every SNP; the second file comprised allele frequency differences for every pairwise comparison that was included in the synchronized file (3 population total). The second file was used to determine heterozygosity and to confirm differences between breeds in previously-identified selected windows from the F_{ST} file.

Intrapopulation analysis. Further investigation was applied to confirm the effectiveness of our analysis. Tajima's Pi was calculated using Popoolation tools (Kofler, Orozco-terWengel, et al., 2011) for each pool separately using created synchronized file for each pool. Delta Pi was calculated as described above. Tajima's Pi

detect selection signature by comparing the mean of pairwise contrasting and the number of segregating sites.

Identification of candidate gene variants differing between subpopulations. The regions identified based on the above-described criteria were further analyzed to detect variants that differentiate one breed from the other two. We included 25 kb to each side of every selected window to reduce the risk of excluding the outer portions of selected haplotypes and to include loci in regulatory sites. CRISP (Bansal, 2010) was used to generate a variant call format (VCF) file from the aligned BAM files to identify call variants in each pool. The VCF file was converted using CRISP (Bansal, 2010) and annotated using snpEff (Cingolani et al., 2012). Genotypes for regions of interest were extracted out from the converted annotated VCF file using bedtools (Quinlan & Hall, 2010). SNPs with high impact and/or in the coding regions were analyzed.

Validation of Identified Variants

Sanger Sequencing: Primers, PCR and sequencing. Because the samples in the pools that were sequenced were not individually barcoded, we generated additional Sanger sequencing data for select genes that showed highly significant signatures of selection. Sanger sequencing was done in 4 dromedaries from each breed/color pool to confirm that these regions were truly polymorphic. Sequence information for LCORL 6 exons, ABCD2 10 exons, CNDP1 exons 6, 11 and 12, and MRPL15 exons 3, 4 and 5 was retrieved form NCBI genome browser and the primers were designed with Primer3 software (Untergasser et al., 2012). Primer details are presented in Table 12. Primers were designed to amplify all exons and exon-intron boundaries. Amplification of the select regions was done by PCR in 10 µL reactions containing 50 ng dromedary genomic DNA (25 ng/ µL) and 0.5 unit of JumpStart Taq ReadyMix (Sigma Aldrich). The PCR products were cleaned with ExoSAP (Affymetrix) and sequenced using BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) and the manufacturer's protocol. Sequencing reactions were cleaned in Spin-50 mini columns (BioMax, Inc) and resolved on 3100 automated sequencer (Applied Biosystems).

TABLE 12: Primers used for PCR and sequencing of *LCORL*, ABCD2, *CNDP1* and *MRPL15*

Gene symbol and primer ID	Region	Forward 5' - 3'	Reverse 5' - 3'	Product size, bp
LCORL	Exon 1	ACCTGGGGATGTTTCTAAATGA	TTTTACCACCAAGCCAAAAA	600
LCORL	Exon 2	TTTTTAGGGGGACAGCATTC	CCTCCAACCACCAATTCCTA	872
LCORL	Exon 3	AATGCTAATTTGGCCCTCTG	GGTCTACAGTGACTGCAAATCCT	511
LCORL	Exon 4	ACATTTGAATTGAAACTCTGAAAA	CTGCACAGCAAGTGCTTCAT	922
LCORL	Exon 5	TTCCTACTTTCCCAAAGCATTA	TTTTCCTTAAAATTGCCTCCT	801
LCORL	Exon 6	CCAGGACTGTAGCGTTTGCT	GAAAACCTGAATCCTTTTCACC	1512
ABCD2	Exon 1	TCTGGGTAGCTGAGGCTAGG	CCACCCCAACTTGTCACTTT	1909
ABCD2	Exon 2	TTTGAAAACACAATTGTAAATCACAT	TGCTCAAGATCAGCTTCAAAAA	713
ABCD2	Exon 3	CCCAACTGTAACAAAGGCAGA	GGAAGAAGGTGGATCACCAA	967
ABCD2	Exon 4	CATTTTGAACCCTGACTTGAAC	ACCTTGTATGGGCTCTCAGG	1512
ABCD2	Exon 5	TGAGCCTGCTATTCCTTGAGA	GGAAAGTCAGTGCTCATTCCA	66
ABCD2	Exon 6	TGTGGTGCCTGAAAAGAAAA	TGCTGGATCAAAACCTCAAA	907
ABCD2	Exon 7	CTGAATACCTACCTACTGTGTTCACT	TCTCCTACCACCCTCCAAAA	800
ABCD2	Exon 8	TCCACCTACTGGGAGATCAAA	TGCCCATCTTTCCAATTCTT	701
ABCD2	Exon 9	ACACCTGGTTTCAACCCAAA	GCATGTCTTCCTTTGCACTTC	857
CNDP1	Exon 6	TTGATGGGTCACTGCCATTA	CGTCACAAACCCCTCTGAAT	612
CNDP1	Exon 11-12	GCAACCATGCTTTTCTTCTTG	GGGAAACCAAGTCGTTGAAG	836
MRPL15	Exon 3	TGCTTCCAGAGGCAAATACTG	AGCAAGCGGAACAGACACAT	573
MRPL15	Exon 4	TGTAGCCACTTTGGGGAGAA	ACCGCAAGCAGAACATCCTA	362
MRPL15	Exon 5	ATGCACGTAGCCAATGTTTG	AAACTTAGGTCAGCAATGTATGAGG	1323

Sequence analysis and mutation discovery. W first sequenced PCR products of *LCORL* 6 exons, *ABCD2* 10 exons, *CNDP1* exons 6, 11 and 12, and *MRPL15* exons 3, 4 and 5 in 4 white, 4 black, and 4 brown dromedaries. Sequences were analyzed for mutations using Sequencher v 5.3 software (Gene Codes Corp.).

Copy number variations (CNVs) analysis: As no significant variants were identified from Sanger sequencing results, we analyzed the WGS data for copy number variants (CNVs) using CNVkit (Talevich, Shain, Botton, & Bastian, 2016) to compare white dromedaries with the black and brown. The tools in CNVkit are designed for the comparison of BAM files of cancer and control samples, as well as of a reference genome. In our analysis, we used white as a 'cancer' sample against the black and brown as 'normal' samples. The tools also require a *bed*-file that contains specific regions of interest in order to scan for identical, specific sequences.

Results

Selective sweeps, genes and genetic variations

Genomic variation: Overall, 94%, 97%, and 98% and 35 Gbp, 29 Gbp, and 32 Gbp for the white, black, and brown pools, respectively, were aligned to the Ca_dromedarius_V1.0 reference with average coverage of 15x (14.9x black pool, 11x red pool, 20.4x white pool). Total of 1,971,138 high quality SNPs were included in the analysis for the Fst sliding window approach.

Genetic differentiation between white, black and brown dromedary breeds: Based on F_{ST} analysis, we have identified 22 selective sweeps for the white against the black and brown breeds (Figure 11 a and b) ($zF_{ST}>8$). The 22 selective sweeps were

identical between white-black and white-brown, suggesting that the white breed is much different than the other two. In contrast, we identified 13 selective sweeps regions between black and brown (Figure 11 c) (zF_{ST}>8).

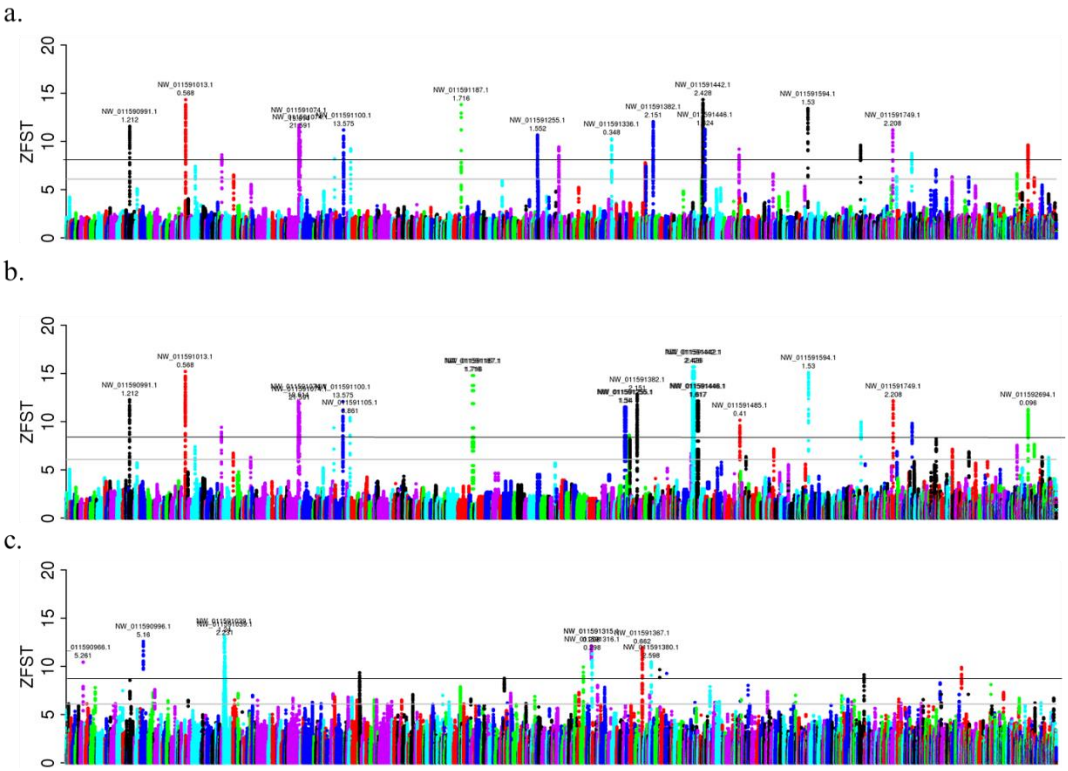

Figure 11. Regions of selective sweeps between the three dromedary breeds. Signals of dromedary directional selection between the three dromedary populations based on Z transformed values (zF_{ST}) of the fixation index. The zFST values are plotted along all scaffolds of the dromedary reference genome comparing: a. the white breed against the black breed, b. the white against the brown and c. the brown against the black. The x-axis represents the dromedary scaffolds where every color is a scaffold. The y-axis shows zFST values and the black horizontal line denotes threshold value zFST=8, whereas gray horizontal line denotes threshold value zFST=6. Scaffolds IDs are printed given for scaffolds with zFST >10.

Identification of genes in selective sweep regions: We used a custom R script to plot regions under selective sweeps along with gene IDs and gene annotations, if any

(Figure 12). Only 13 out of 35 regions contained genes (Table 13). No plots were made for regions with no annotated genes. The NCBI database was used to confirm gene IDs and positions of these regions. Files with scaffold ID, start, stop and zFST values for each region with zFST>8 were generated using MATLAB.

Table 13: A summary of identified genes in selective sweep regions along with zFST values and CNVs estimate from CNVkit

Scaffld ID	**Start**	**Stop**	**Genes**	**zF_{ST}**	**CNVs**
NW_011590991.1	1169500	1252500	*LXN*	12.27	11
NW_011591013.1	569819	594456	*CNDP1*	15.2	3
NW_011591038.1	1547500	1620500	*CTSV*	9.41	9
NW_011591074.1	21334107	21335085	*PPM1L*	12.12	1
NW_011591098.1	851500	983500	*LOC105086721*	13.47	17
NW_011591100.1	13485500	13524923	*USP25*	13.56	5
NW_011591100.1	13524923	13580115	*USP25*	13.56	7
NW_011591100.1	13580115	13690500	*USP25*	13.56	14
NW_011591187.1	1685487	2062276	*ABCD2, SLC2A13*	18.88	41
NW_011591255.1	1450431	1587003	*VPS36, THSD1, SLC25A15*	11.77	7
NW_011591275.1	5593633	5641177	*LGALS3, DLGAP5*	9.72	6
NW_011591382.1	1787610	2189516	*PDE5A, MAD2L1*	12.83	17
NW_011591442.1	2081500	2474500	*HS3ST4*	15.69	51
NW_011591485.1	336500	449500	*ATRX*	10.51	15
NW_011591594.1	1520950	1558703	*MRPL15, LYPLA1*	15.09	4
NW_011591685.1	2465604	2550967	*LCORL*	9.98	11
NW_011591749.1	2172500	2269500	*TRMT1L*	13.28	13
NW_011591789.1	1039668	1102377	*TTC1, PWWP2A*	9.82	7
NW_011591890.1	1014255	1091513	*MPRIP*	8.23	10

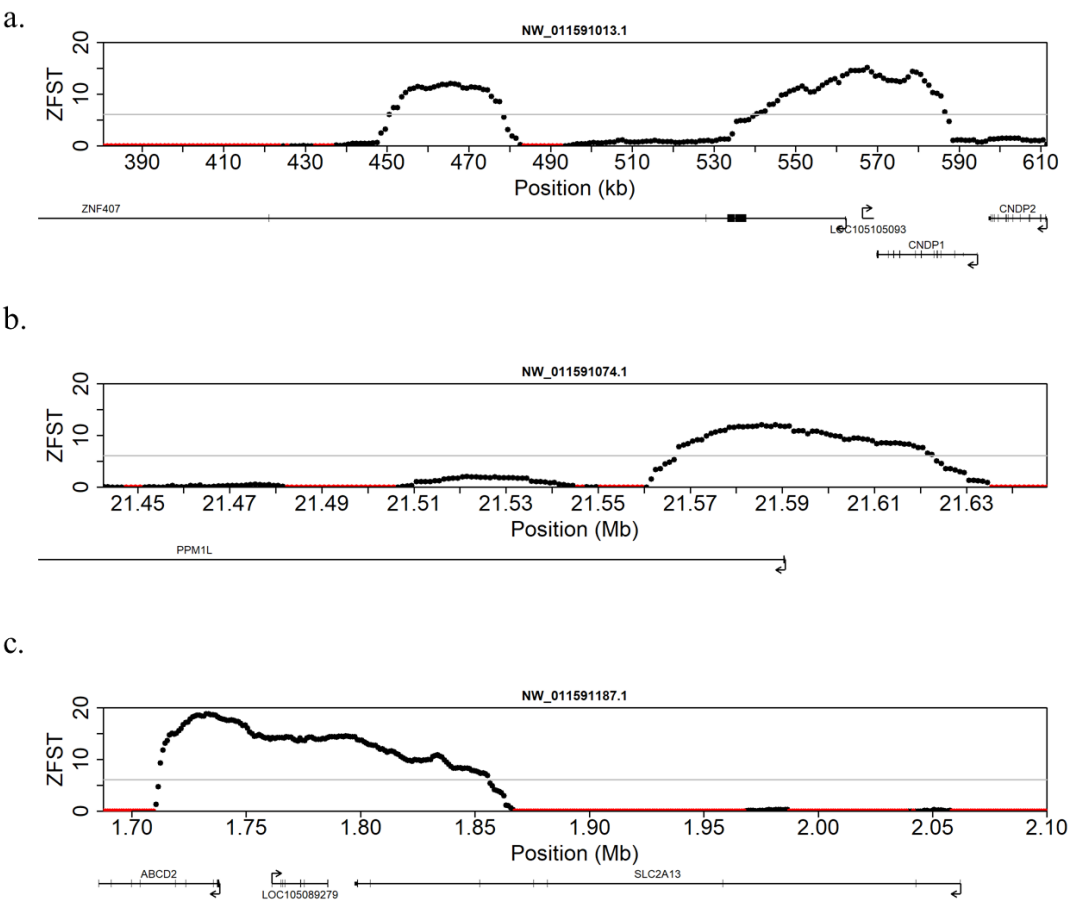

Figure 12. An example of R script output for genes in regions of selective sweeps. windows of selective sweep regions that contain a. *CNDP1*, b. *PPM1L* and c. *ABCD2* and *SLC2A13*. x-axis: position in scaffold (kb or Mbp)

Identification of high impact variants under selective sweeps: We used bed tools to extract exonic regions from the VCF-converted annotated file and analyzed missense, non-sense, frameshift, and high impact variants. A total of 99 mutations were identified in 15 genes, and allele frequencies were calculated for all variants (Table 14).

Table 14: Detailed list of variants found in exonic regions of genes from the VCF converted annotated file

Gene ID	Scaffold ID	Possition	Exon #	Variant Type	Black Ref	Alternative	Red Ref	Alternative	White ref	Alternative
CNDP1	NW_011591013.1	569893	12	missense mutation	1	0	1	0	0.52	0.48
CNDP1		569921	11	missense mutation	1	0	1	0	0.52	0.48
CNDP1		570136	11	missense mutation	1	0	1	0	0.11	0.89
CNDP1		572617	10	missense mutation	1	0	1	0	0.09	0.91
CNDP1		575445	8	missense mutation	1	0	0.89	0.11	0.14	0.86
CNDP1		580810	6	missense mutation	1	0	1	0	0.83	0.17
PPM1L	NW_011591074.1	21590471	1	missense mutation	1	0	1	0	0.72	0.28
ABCD2	NW_011591187.1	1722781	5	missense mutation	1	0	1	0	0.92	0.08
ABCD2		1722790	5	missense mutation	1	0	1	0	0.61	0.39
ABCD2		1722791	5	missense mutation	1	0	1	0	0.55	0.45
ABCD2		1722839	5	missense mutation	1	0	1	0	0.46	0.54
ABCD2		1723669	4	missense mutation	1	0	1	0	0.52	0.48
ABCD2		1723681	4	missense mutation	1	0	1	0	0.42	0.58
ABCD2		1735654	2	missense mutation	1	0	1	0	0.88	0.12
ABCD2		1735670	2	missense mutation	1	0	1	0	0.87	0.13

Table 14 continued

Gene ID	Scaffold ID	Possition	Exon #	Variant Type	Black Ref	Alternative	Red Ref	Alternative	White ref	Alternative
ABCD2		1735687	2	missense mutation	1	0	1	0	0.45	0.55
ABCD2		1737303	1	missense mutation	1	0	1	0	0.67	0.33
ABCD2		1737687	1	missense mutation	1	0	1	0	0.46	0.54
ABCD2		1737733	1	missense mutation	1	0	1	0	0.41	0.59
ABCD2		1737796	1	missense mutation	1	0	0.91	0.09	0.47	0.53
ABCD2		1737801	1	missense mutation	1	0	1	0	0.75	0.25
ABCD2		1737814	1	missense mutation	1	0	1	0	0.53	0.47
ABCD2		1737850	1	missense mutation	1	0	1	0	0.75	0.25
ABCD2		1738012	1	missense mutation	1	0	1	0	0.62	0.38
ABCD2		1738012	1	missense mutation	1	0	1	0	0.62	0.38
ABCD2		1738027	1	missense mutation	1	0	1	0	0.67	0.33
ABCD2		1738063	1	missense mutation	1	0	1	0	0.77	0.23
CKAP2	NW_011591255.1	1436982	2	missense mutation	0.79	0.21	0.92	0.08	0.91	0.09
VPS36		1465135	12	missense mutation	1	0	1	0	0.74	0.26
THSD1		1474646	1	missense mutation	1	0	1	0	0.71	0.29

Table 14 continued

Gene ID	Scaffold ID	Possition	Exon #	Variant Type	Black Ref	Alternative	Red Ref	Alternative	White ref	Alternative
THSD1		1478745	2	missense mutation	0.83	0.17	0.6	0.4	0.83	0.17
THSD1		1488916	3	missense mutation	1	0	1	0	0.69	0.31
THSD1		1493660	5	missense mutation	1	0	1	0	0.77	0.23
THSD1		1493665	5	missense mutation	1	0	1	0	0.075	0.25
THSD1		1493773	5	missense mutation	1	0	1	0	0.42	0.058
THSD1		1493794	5	missense mutation	1	0	1	0	0.67	0.33
THSD1		1493989	6	missense mutation	1	0	1	0	0.38	0.62
THSD1		1493995	6	missense mutation	1	0	1	0	0.36	0.64
SLC25A15		1572743	1	missense mutation	1	0	1	0	0.75	0.25
LGALS3	NW_011591275.1	5597579	3	missense mutation	1	0	1	0	0.73	0.27
LGALS3		5597580	3	missense mutation	1	0	1	0	0.68	0.32
DLGAP5		5616722	13	missense mutation	1	0	1	0	0.8	0.2
DLGAP5		5621529	12	missense mutation	1	0	1	0	0.79	0.21
DLGAP5		5622385	10	missense mutation	1	0	1	0	0.56	0.44
DLGAP5		5622436	10	missense mutation	1	0	1	0	0.66	0.34

Table 14 continued

Gene ID	Scaffold ID	Possition	Exon #	Variant Type	Black Ref	Alternative	Red Ref	Alternative	White ref	Alternative
DLGAP5		5630115	4	splice acceptor variant (High)	1	0	1	0	0.68	0.32
DLGAP5		5632800	3	missense mutation	1	0	1	0	0.81	0.19
DLGAP5		5632813	3	missense mutation	1	0	1	0	0.8	0.2
DLGAP5		5639097	2	missense mutation	1	0	1	0	0.47	0.53
DLGAP5		5639101	2	missense mutation	1	0	1	0	0.65	0.35
PDE5A	NW_011591382.1	1908922	1	missense mutation	1	0	1	0	0.91	0.09
PDE5A		2184296	5	missense mutation	1	0	1	0	0.85	0.15
MRPL15	NW_011591594.1	1524944	3	missense mutation	1	0	1	0	0.77	0.23
MRPL15		1525064	3	missense mutation	1	0	1	0	0.57	0.43
MRPL15		1525698	2	missense mutation	1	0	1	0	0.69	0.31
LCORL		2468830	6	missense mutation	1	0	1	0	0.14	0.86
LCORL		2468833	6	nonsense mutation	1	0	1	0	0.9	0.1
LCORL		2468844	6	missense mutation	1	0	1	0	0.29	0.71
LCORL		2468845	6	missense mutation	1	0	1	0	0.029	0.71

Table 14 continued

Gene ID	Scaffold ID	Possition	Exon #	Variant Type	Black Ref	Alternative	Red Ref	Alternative	White ref	Alternative
LCORL		2468899	6	missense mutation	1	0	1	0	0.13	0.87
LCORL		2468934	6	missense mutation	1	0	1	0	0.86	0.14
LCORL		2469123	6	missense mutation	1	0	1	0	0.18	0.82
LCORL		2469179	6	missense mutation	1	0	1	0	0.14	0.86
LCORL		2469180	6	missense mutation	1	0	1	0	0.14	0.94
LCORL		2469268	6	missense mutation	1	0	1	0	0.06	0.94
LCORL		2469345	6	missense mutation	1	0	1	0	0.13	0.87
LCORL		2469463	6	missense mutation	1	0	1	0	0.83	0.17
LCORL		2469524	6	missense mutation	1	0	1	0	0.1	0.9
LCORL		2469573	6	missense mutation	1	0	1	0	0.11	0.89
LCORL		2469586	6	missense mutation	1	0	1	0	0.14	0.86
LCORL		2469684	6	missense mutation	1	0	1	0	0.14	0.86
LCORL		2469779	6	missense mutation	1	0	1	0	0.18	0.82
LCORL		2471166	5	missense mutation	0.94	0.06	1	0	0.17	0.83

Table 14 continued

Gene ID	Scaffold ID	Possition	Exon #	Variant Type	Black Ref	Alternative	Red Ref	Alternative	White ref	Alternative
LCORL		2489000	4	missense mutation	1	0	0.9.0.1	0	0.05	0.95
PWWP2A	NW_011591789.1	1101220	1	missense mutation	1	0	1	0	0.54	0.46
PWWP2A		1101280	1	missense mutation	1	0	1	0	0.26	0.74
PWWP2A		1101300	1	missense mutation	1	0	1	0	0.82	0.18
PWWP2A		1101307	1	missense mutation	1	0	1	0	0.34	0.66
PWWP2A		1101319	1	missense mutation	1	0	1	0	0.9	0.1
PWWP2A		1101406	1	missense mutation	1	0	1	0	0.47	0.53
PWWP2A		1101410	1	nonsense mutation	1	0	1	0	0.9	0.1
PWWP2A		1101414	1	missense mutation	1	0	1	0	0.37	0.63
PWWP2A		1101425	1	missense mutation	1	0	1	0	0.36	0.64
PWWP2A		1101433	1	frameshift deletion	0.54	0.46	0	1	0.64	0.36
CTTN	NW_0115926941.1	52856	13	missense mutation	1	0	0.83	0.17	0.29	0.71
CTTN		52857	13	missense mutation	1	0	0.83	0.17	0.29	0.71
CTTN		52869	13	missense mutation	1	0	1	0	0.56	0.44

Table 14 continued

Gene ID	Scaffold ID	Possition	Exon #	Variant Type	Black Ref	Alternative	Red Ref	Alternative	White ref	Alternative
CTTN		57914	11	missense mutation	1	0	1	0	0.61	0.39
CTTN		57959	11	missense mutation	1	0	1	0	0.71	0.29
CTTN		59349	9	missense mutation	1	0	1	0	0.25	0.75
PPFIA1		91222	26	missense mutation	1	0	1	0	0.77	0.23
PPFIA1		91345	26	missense mutation	1	0	1	0	0.84	0.16
PPFIA1		91373	26	missense mutation	1	0	1	0	0.37	0.63
PPFIA1		95145	22	missense mutation	1	0	1	0	0.74	0.26
PPFIA1		95233	22	missense mutation	1	0	1	0	0.73	0.27
PPFIA1		108688	18	missense mutation	1	0	1	0	0.6	0.4
PPFIA1		118924	13	missense mutation	0.82	0.18	1	0	1	0
PPFIA1		119391	12	missense mutation	0.67	0.33	0.75	0.25	0.06	0.94
PPFIA1		126658	7	missense mutation	1	0	1	0	0.67	0.33

Further analysis to identify what drove the signals

Sanger sequencing: due to many regions were identified, we selected 20 regions of 4 genes to identified variants because of high number of missense variants in these genes. We sequenced 4 individuals from each group (4 black, 4 red and 4 white). However, no variations were identified in these regions.

Identification of CNVs: After further review of the VCF and confirming with Integrative Genomics Viewer (IGV), data shows that most regions with high zF_{ST} have greater than 2 times the average genome-wide read depth. We used CNVkit to identify CNVs in these regions. All regions show 2 or more copies except PPM1L scaffold NW_011591074.1, which has only one copy. This analysis was done for only the white breed versus the other two groups (table 13).

The F_{ST} analysis identified 13 regions of significant divergencebetween black vs brown subpopulations (Figure 11 c) ($zF_{ST}>8$). Arginine and serine rich coiled-coil 1 (*RSRC1*) the short stature homebox (*SHOX*) and Fatty acid elongase7 (*ELOVL7*) were selected for Sanger sequencing because these genes are associated with adaptation and QTL traits in other species (Binder, 2011; B. Yang et al., 2013; B. Zhang et al., 2018). However, additional sequence analysis did not detect any variants in the coding regions of these genes in the individuals studied.

Discussion

In this study, we compared 3 dromedary breeds, defined by coat color, from Saudi Arabia and showed that these breeds are genetically different, showing distinct regions with signatures of selection. We further showed that the white dromedary breed

is the most different from the black and brown breeds, which is in agreement with previous findings (Abdulaziz M Al-Swailem et al., 2007). We identified 35 genes that appeared to be under selection in these three populations (Table 12), some of which may have important functions in dromedary biology.

The only single-copy gene in the white breed that was found in a region with significant zFST when compared to black and brown dromedaries, was protein phosphatase, Mg2+/Mn2 dependent 1L (*PPM1L*). The gene has been associated with metabolic syndrome, obesity, and growth in mice (Y. Chen et al., 2008). Also, *PPM1L* has been associated with food intake and growth traits in US cattle (Seabury et al., 2017). More recently, *PPM1L* was found to be one of many candidate genes under directional selection for production traits in Iraqi cattle (Alshawi, Essa, Al-Bayatti, & Hanotte, 2019). These findings make *PPM1L* an attractive candidate gene for growth and production also in the dromedary.

In this study, ATP binding cassette subfamily D member 2 (*ABCD2*) and solute carrier family 2 member 13 (*SLC2A13*) were found under the highest zF_{ST} peak (zF_{ST}>18), most likely due to CNVs (CNVs=41). Both genes function as extra- and intra-cellular transporters and both have been associated with obesity in mice (J. Liu et al., 2012; Stewart, Kim, Saxton, & Kim, 2010). The *ABCD2* gene has repeatedly been associated with growth and body gain in pigs (Fernandez et al., 2012), growth and carcass traits in cattle (Kim, Kim, Raney, & Ernst, 2012), fatty acid accumulation in white pigs (B. Yang et al., 2013), and body weight gain in chickens (Yuan et al., 2018). The *SLC2A13* gene, on the other hand, has been associated with resistance against clinical mastitis in dairy

cattle in a QTL analysis (Cai, Guldbrandtsen, Lund, & Sahana, 2018). Thus, both genes may also be of interest regarding growth in the dromedary.

The region with the second highest zF_{ST} values (zF_{ST}>15) contained four genes. Of these, the heparan sulfate-glucosamine3-sulfotransferase 4 (*HS3ST4*) gene, with known function in heparan sulfate modification (Antoine, Yakoub, Maus, Shukla, & Tiwari, 2014), showed 51 CNVs (Table 12). The Carnosine dipeptidase 1 (*CNDP1*) gene had 3 CNVs and has been associated with food intake and body gain in pigs (Shen et al., 2015). Mitochondrial ribosomal protein L15 (*MRPL15*) (CNV=4) has been associated with weaning and body gain (Cardoso et al., 2018; Saatchi, Schnabel, Taylor, & Garrick, 2014), while lysophospholipase 1 (*LYPLA1*) (CNVs=4) has been associated with QTL traits such as food intake, body gain weight, and dairy production in both cattle and pigs (Cheruiyot et al., 2018; Do et al., 2014; Lindholm-Perry et al., 2012; Magalhaes et al., 2016).

Further, ubiquitin specific peptide 25 (*USP25*) with 14 CNVs in this study, has been associated with skatole levels in boars (Ramos et al., 2011); tRNA methyltransferase 1 like (*TRMT1L*) (CNVs=13), also is a known CNV gene in humans (Rodrigues-Peres et al., 2019), and ligand of number protein X (*LXN*) (CNVs=11) is involved in signal transduction and protein interaction (REF)(Boschiero et al., 2018).

A few genes identified in this study, have been associated with various reproductive phenotypes in other species. These include putative spermatogenesis-associated protein 31D4 (*LOC105086721*) (CNVs=17); phosphodiesterase 5A (*PDE5A*) (CNVa=17) associated with both scrotal circumference (Utsunomiya et al., 2014), and

female fertility in cattle (Kiser et al., 2019), as well as blood pressure in these species (Warren et al., 2017).

Some genes identified in this study could be of potential interest regarding several other phenotypic traits. For example, mitotic arrest deficient 2 like 1 (*MAD2L1*) (CNVs=17) has been associated with coat color and/or morphological characteristics in cattle (Ramey et al., 2013), Vacuolar protein sorting 36 homolog (*VPS36*) (CNVs=7) has been associated with wool production in Chinese sheep (S. Liu et al., 2017), Thrombospondin type 1 domain containing 1 (*THSD1*) (CNVs=7) is associated with fat deposition in white-tailed deer (B. Li et al., 2018), and *SLC25A3* (CNVs=7) is associated with low food intake and higher fat accumulation in cattle (Mukiibi et al., 2018; Olivieri et al., 2016). In addition, two genes the *cortactin* (*CTTN*) and interacting protein alpha 1 (*PPFIA1*) (CNVs=14) were in the same region and function as regulators; Cathepsin V (*CTSV*) is also a regulator, and the ATRX chromatin remodeler (*ATRX*) (CNVs=15) is involved in DNA methylation and X-linked syndromes (Schutt et al., 2003).

Perhaps a gene of particular interest was ligand dependent nuclear receptor corepressor like (*LCORL*). It is a well-studied gene and associated with body size in many species including cattle (Bouwman et al., 2018; L. Xu et al., 2019), sheep (Al-Mamun et al., 2015), horses (Gurgul et al., 2019), pig (Rubin et al., 2012), and dogs (Schoenebeck & Ostrander, 2014). Our analysis detected 11 CNVs in the *LCORL* gene, suggesting its possible role in body size regulation also in the dromedary.

Finally, two genes in significant zFST regions have known functions during the development: Tetratricopeptide domain 1 (*TTC1*) is associated with cattle growth and

development (Mesbah-Uddin et al., 2017) and myosin phosphatase Rho interacting protein (*MPRIP)* is associated with congenital and bone development in newborn red dairy cattle(Agerholm, McEvoy, Menzi, Jagannathan, & Drogemuller, 2016). No known functions could be found for some genes, such as PWWP domain containing 2 A (*PWWP2A*).

Taken together, it is noteworthy that the majority of the genes from the regions under selection in the three dromedary populations, are predominantly involved in biological processes associated with growth, development and reproduction. On the other hand, this is probably not unexpected because during the course of domestication, these have been the traits of main interest for humans. Our findings show that separation of dromedary populations by color is accompanied by signatures of selection for other traits of human interest.

This is among the first and quite a limited analysis of Saudi Arabian dromedary populations. The findings are preliminary, should be taken with caution and require further verification. However, these preliminary data certainly warrant further and more detailed genomic research in the dromedary. Since the species does not have advanced genomics tools such as WG SNP arrays, the further progress encompasses generation of WGS of more individuals.

CHAPTER V

CONCLUSIONS

This study represents a small but focused contribution to the genome analysis, trait genetics and population studies in the dromedary and other camelids. Compared to the recent unparalleled progress in the genomics of other livestock species (Ghosh et al., 2018) (Raudsepp et al. 2019, submitted), camelid genomics lags behind at almost all accounts. There has been no internationally organized and coordinated collection of genomics resources (DNA, tissue, cells, phenotypes, pedigrees); there is only one genomic BAC library (CHORI246 from a female alpaca) and 2 radiation hybrid panels (Perelman et al., 2018), but no radiation hybrid maps. Despite of the available reference sequence assemblies for the alpaca (Richardson et al., 2019) and the camels (Wu et al., 2014), no SNP genotyping arrays are available for any of the seven extant species. As a consequence, current knowledge about the genes and mutations underlying camelid traits, diseases, production, reproduction and performance are very limited.

In this research, we generated a collection of over 200 DNA samples and phenotypes for dromedaries from Saudi Arabia and US. Besides being used for coat color genetics and population studies in this research, the DNA collection will have a prospective use for generating WGS for individual animals, targeting complex regions in the dromedary genome and adding variants to SNP and CNV collection for the species. This, in turn, will facilitate the discovery of genes underlying Mendelian and complex traits.

Collection and preservation of dromedary testis tissue for transcriptome studies is a step towards functional annotation of the dromedary genome. The resource has already been used for RNAseq and Y chromosome annotation, and it has further applications for genome sequence annotation and the discovery of male fertility genes.

The collection of live cryopreserved cells has already had several applications in collaborative projects, such as generation of a 15,000rad RH panel (Perelman et al., 2018) and flow sorting of the Y chromosome. However, live cells are critically needed for several novel cutting-edge technologies. For example, to generate Hi-C libraries for fine mapping of chromatin contacts, which approach is widely used for scaffolding and increasing the contiguity of genome assemblies (PMID 30456315; 24185095). Very recently, this was applied for improving the Illumina assembly of North African dromedary (Elbers et al., 2019). Thus, the resource has an important application for chromosome-scale scaffolding of the dromedary genome. Likewise, live cells are needed for transcriptomics, particularly for transcriptomics at a single cell level. Live cells are also the excellent source of long DNA molecules that are needed for BioNano genomics (https://bionanogenomics.com/technology/), which is a novel platform for chromosome-level genome assemblies and for the detection of all kinds of structural variants (SNPs, CNVs, translocations, inversions) in the genome. Even though the immediate application of these platforms for the dromedary is not very likely, without the cryopreserved cells, it cannot happen in the future either.

We also prepared fixed cell suspensions and chromosome preparations for several dromedaries and Bactrian camels in this study. This is a resource for comparative

and clinical cytogenetics, as well as for gene mapping and comparative gene mapping in camelids. In Chapter III, we describe the use of this resource for comparative gene mapping of coat color genes. However, cytogenetic mapping by FISH requires suitable probes. Therefore, another important resource generated in this study, was the collection of alpaca BAC clones for FISH. Some were used for mapping coat color genes in this study (Chapter III), others contributed to chromosomal assignment of 11 candidate genes for coat color and hair characteristics in the alpaca (Mendoza et al., 2019), while an additional 32 markers are ready to use for chromosomal assignment of selected markers to assist with the generation of alpaca SNP chip. The latter is an awaited tool in camelid genomics, for genome-wide association studies, mutation discovery and population analysis.

Coat color is among the traits, humans have been selecting for in almost all domestic animals, and coat color genes for most livestock species are well documented (Baxter, Watkins-Chow, Pavan, & Loftus, 2019; M. Reissmann & A. Ludwig, 2013; H. Suzuki, 2013) Here again, the dromedary coat color genetics lags far behind. Even though many genes are involved in determining the coat color, in this study we focused only on the major genes. We identified variants in *MC1R* and *ASIP* and showed their association with white and black coat color, respectively. The results are not surprising, because these genes regulate pigmentation in many other species. However, findings contribute to comparative coat color genetics and the development of molecular tests in camelids, whereas cytogenetic mapping of these genes will assist the assignment of sequence scaffolds to chromosomes in camelid genome reference assemblies.

Comparisons between Saudi Arabian dromedary breeds showed that these breeds are not only phenotypically but also genetically different. Strikingly, white dromedaries, the ones which in some parts of Saudi Arabia associate with the social status of the owner and are used for ‘beauty contests’, were genetically most different. This preliminary genome scan led the identification of selective sweeps and genes in these regions. Future studies are needed to reveal which of these genes and genomic regions under selection carry variants governing economically important traits and features beneficial for dromedary biology.

REFERENCES

Aba, M. (2013). Anatomy and Physiology of Reproduction in the Female Llama and Alpaca. *Llama and Alpaca Care: (pp. 140-150). WB Saunders.*

Abdallah, H. R., & Faye, B. (2012). Phenotypic classification of Saudi Arabian camel (*Camelus dromedarius*) by their body measurements. *Emirates Journal of Food and Agriculture, 24*(3), 272-280.

Abitbol, M., Legrand, R., & Tiret, L. (2015). A missense mutation in the agouti signaling protein gene (*ASIP*) is associated with the no light points coat phenotype in donkeys. *Genet Sel Evol, 47*, 28. doi:10.1186/s12711-015-0112-x

Agerholm, J. S., McEvoy, F. J., Menzi, F., Jagannathan, V., & Drogemuller, C. (2016). A CHRNB1 frameshift mutation is associated with familial arthrogryposis multiplex congenita in Red dairy cattle. *BMC Genomics, 17*, 479. doi:10.1186/s12864-016-2832-x

Agnarsson, I., & May-Collado, L. J. (2008). The phylogeny of Cetartiodactyla: the importance of dense taxon sampling, missing data, and the remarkable promise of cytochrome b to provide reliable species-level phylogenies. *Mol Phylogenet Evol, 48*(3), 964-985. doi:10.1016/j.ympev.2008.05.046

Ahima, R. S., & Hileman, S. M. (2000). Postnatal regulation of hypothalamic neuropeptide expression by leptin: implications for energy balance and body weight regulation. *Regulatory Peptides, 92*(1-3), 1-7.

Al-Ali, A. K., Husayni, H. A., & Power, D. M. (1988). A comprehensive biochemical analysis of the blood of the camel (*Camelus dromedarius*). *Comp Biochem Physiol B, 89*(1), 35-37. doi:10.1016/0305-0491(88)90257-x

Al-Mamun, H. A., Kwan, P., Clark, S. A., Ferdosi, M. H., Tellam, R., & Gondro, C. (2015). Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. *Genet Sel Evol, 47*, 66. doi:10.1186/s12711-015-0142-4

Al-Swailem, A. M., Al-Busadah, K. A., Shehata, M. M., Al-Anazi, I. O., & Askari, E. (2007). Classification of Saudi Arabian camel (*Camelus dromedarius*) subtypes based on RAPD technique. *Journal of Food Agriculture and Environment, 5*(1), 143.

Al-Swailem, A. M., Shehata, M. M., Abu-Duhier, F. M., Al-Yamani, E. J., Al-Busadah, K. A., Al-Arawi, M. S., . . . Otu, H. H. (2010). Sequencing, analysis, and annotation of expressed sequence tags for *Camelus dromedarius*. *PLoS One, 5*(5), e10720. doi:10.1371/journal.pone.0010720

Ali, A., Baby, B., & Vijayan, R. (2019). From Desert to Medicine: A Review of Camel Genomics and Therapeutic Products. *Front Genet, 10*, 17. doi:10.3389/fgene.2019.00017

Alim, F. Z. D., Romanova, E. V., Tay, Y.-L., Rahman, A. Y. B. A., Chan, K.-G., Hong, K.-W., . . . Hindmarch, C. C. T. (2019). Seasonal adaptations of the hypothalamo-neurohypophyseal system of the dromedary camel. *PLoS One, 14*(6), e0216679-e0216679. doi:10.1371/journal.pone.0216679

Almathen, F., Charruau, P., Mohandesan, E., Mwacharo, J. M., Orozco-terWengel, P., Pitt, D., . . . Burger, P. A. (2016). Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proc Natl Acad Sci U S A, 113*(24), 6707-6712. doi:10.1073/pnas.1519508113

Almathen, F., Elbir, H., Bahbahani, H., Mwacharo, J., & Hanotte, O. (2018). Polymorphisms in *MC1R* and *ASIP* Genes are Associated with Coat Color Variation in the Arabian Camel. *J Hered, 109*(6), 700-706. doi:10.1093/jhered/esy024

Alshanbari, F., Castaneda, C., Juras, R., Hillhouse, A., Mendoza, M. N., Gutiérrez, G. A., . . . Raudsepp, T. (2019). Comparative FISH-Mapping of *MC1R*, *ASIP*, and *TYRP1* in New and Old World Camelids and Association Analysis With Coat Color Phenotypes in the Dromedary (*Camelus dromedarius*). *Front Genet, 10*, 340-340. doi:10.3389/fgene.2019.00340

Alshawi, A., Essa, A., Al-Bayatti, S., & Hanotte, O. (2019). Genome Analysis Reveals Genetic Admixture and Signature of Selection for Productivity and Environmental Traits in Iraqi Cattle. *Frontiers in Genetics*, 10, 340.

Andersson, B., Olsson, K., & Rundgren, M. (1980). ADH in regulation of blood osmolality and extracellular fluid volume. *JPEN J Parenter Enteral Nutr, 4*(2), 88-96. doi:10.1177/014860718000400207

Andersson, L. (2001). Genetic dissection of phenotypic diversity in farm animals. *Nat Rev Genet, 2*(2), 130-138. doi:10.1038/35052563

Andersson, L. (2003). Melanocortin receptor variants with phenotypic effects in horse, pig, and chicken. *Ann N Y Acad Sci, 994*, 313-318.

Andersson, L. (2012). How selective sweeps in domestic animals provide new insight into biological mechanisms. *Journal of Internal Medicine, 271*(1), 1-14.

Andersson, L., & Georges, M. (2004). Domestic-animal genomics: deciphering the genetics of complex traits. *Nature Reviews Genetics, 5*(3), 202.

Antoine, T. E., Yakoub, A., Maus, E., Shukla, D., & Tiwari, V. (2014). Zebrafish 3-O-sulfotransferase-4 generated heparan sulfate mediates HSV-1 entry and spread. *PLoS One, 9*(2), e87302. doi:10.1371/journal.pone.0087302

Arzamendia, Y., Baldo, J. L., & Vilá, B. (2012). Lineamientos para un plan de conservación y uso sustentable de vicuñas en Jujuy, Argentina. *VICAM, EDIUNJU.* Jujuy, Argentina.

Avila, F., Baily, M. P., Merriwether, D. A., Trifonov, V. A., Rubes, J., Kutzler, M. A., . . . Raudsepp, T. (2015). A cytogenetic and comparative map of camelid chromosome 36 and the minute in alpacas. *Chromosome Research, 23*(2), 237-251.

Avila, F., Baily, M. P., Perelman, P., Das, P. J., Pontius, J., Chowdhary, R., . . . Raudsepp, T. (2014). A comprehensive whole-genome integrated cytogenetic map for the alpaca (*Lama pacos*). *Cytogenet Genome Res, 144*(3), 196-207. doi:10.1159/000370329

Avila, F., Das, P. J., Kutzler, M., Owens, E., Perelman, P., Rubes, J., . . . Raudsepp, T. (2014). Development and application of camelid molecular cytogenetic tools. *J Hered, 105*(6), 858-869. doi:10.1093/jhered/ess067

Ayoub, N. A., McGowen, M. R., Clark, C., Springer, M. S., & Gatesy, J. (2009). Evolution and phylogenetic utility of the melanocortin-1 receptor gene (*MC1R*) in Cetartiodactyla. *Mol Phylogenet Evol, 52*(2), 550-557. doi:10.1016/j.ympev.2009.03.008

Balmus, G., Trifonov, V. A., Biltueva, L. S., O'Brien, P. C., Alkalaeva, E. S., Fu, B., . . . Ferguson-Smith, M. A. (2007). Cross-species chromosome painting among camel, cattle, pig and human: further insights into the putative Cetartiodactyla ancestral karyotype. *Chromosome Res, 15*(4), 499-515. doi:10.1007/s10577-007-1154-x

Bansal, V. (2010). A statistical method for the detection of variants from next-generation resequencing of DNA pools. *Bioinformatics, 26*(12), i318-i324.

Barsh, G. (2001). Coat color mutations, animals. *Encyclopedia of Genetics (Brenner S & Miller JH, eds.) Academic*, 397-401.

Bartha, T., Sayed-Ahmed, A., & Rudas, P. (2005). Expression of leptin and its receptors in various tissues of ruminants. *Domestic Animal Endocrinology, 29*(1), 193-202.

Baxter, L. L., Watkins-Chow, D. E., Pavan, W. J., & Loftus, S. K. (2019). A curated gene list for expanding the horizons of pigmentation biology. *Pigment Cell Melanoma Res, 32*(3), 348-358. doi:10.1111/pcmr.12743

Beaumont, K. A., Shekar, S. N., Newton, R. A., James, M. R., Stow, J. L., Duffy, D. L., & Sturm, R. A. (2007). Receptor function, dominant negative activity and phenotype correlations for *MC1R* variant alleles. *Human Molecular Genetics,* 16(18), 2249-2260.

Bellone, R. R. (2010). Pleiotropic effects of pigmentation genes in horses. *Anim Genet, 41 Suppl 2*, 100-110. doi:10.1111/j.1365-2052.2010.02116.x

Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., & Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods, 58*(3), 268-276. doi:10.1016/j.ymeth.2012.05.001

Bennett, D., Huszar, D., J Laipis, P., Jaenisch, R., & J Jackson, I. (1990). Phenotypic rescue of brown melanocytes by a retrovirus bearing a wildtype tyrosinase-related protein gene. *Development,* Cambridge, England, 110, 471-475.

Berryere, T. G., Schmutz, S. M., Schimpf, R. J., Cowan, C. M., & Potter, J. (2003). *TYRP1* is associated with dun coat colour in Dexter cattle or how now brown cow? *Anim Genet, 34*(3), 169-175.

Bianchi, N. O., Larramendy, M. L., Bianchi, M. S., & Cortes, L. (1986). Karyological conservation in South American camelids. *Experientia, 42,* 622-624.

Binder, G. (2011). Short stature due to *SHOX* deficiency: genotype, phenotype, and therapy. *Horm Res Paediatr, 75*(2), 81-89. doi:10.1159/000324105

Bogner, P., Csutora, P., Cameron, I. L., Wheatley, D. N., & Miseta, A. (1998). Augmented water binding and low cellular water content in erythrocytes of camel and camelids. *Biophysical Journal, 75*(6), 3085-3091.

Boichard, D., Ducrocq, V., Croiseau, P., & Fritz, S. (2016). Genomic selection in domestic animals: Principles, applications and perspectives. *Comptes Rendus Biologies,* 339(7-8), 274-277.

Boschiero, C., Moreira, G. C. M., Gheyas, A. A., Godoy, T. F., Gasparin, G., Mariani, P. D. S. C., . . . Coutinho, L. L. (2018). Genome-wide characterization of genetic variants and putative regions under selection in meat and egg-type chicken lines. *BMC Genomics, 19*(1), 83-83. doi:10.1186/s12864-018-4444-0

Bouaouda, H., Achaaban, M. R., Ouassat, M., Oukassou, M., Piro, M., Challet, E., . . . Pevet, P. (2014). Daily regulation of body temperature rhythm in the camel (*Camelus dromedarius*) exposed to experimental desert conditions. *Physiol Rep, 2*(9). doi:10.14814/phy2.12151

Bouwman, A. C., Daetwyler, H. D., Chamberlain, A. J., Ponce, C. H., Sargolzaei, M., Schenkel, F. S., . . . Hayes, B. J. (2018). Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat Genet, 50*(3), 362-367. doi:10.1038/s41588-018-0056-5

Bovine Genome, S., Analysis, C., Elsik, C. G., Tellam, R. L., Worley, K. C., Gibbs, R. A., . . . Zhao, F.-Q. (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science, 324*(5926), 522-528. doi:10.1126/science.1169588

Brooks, S. A., & Bailey, E. (2005). Exon skipping in the *KIT* gene causes a Sabino spotting pattern in horses. *Mamm Genome, 16*(11), 893-902. doi:10.1007/s00335-005-2472-y

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

Bult, C. J., Eppig, J. T., Kadin, J. A., Richardson, J. E., Blake, J. A., & Group, t. M. G. D. (2008). The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Research, 36*(suppl_1), D724-D728. doi:10.1093/nar/gkm961

Bunch, T. D. (1985). Chromosome banding pattern homologies and NORs for the Bactrian camel, guanaco and llama. *Journal of Heredity, 76*(2), 115-118.

Bunch, T. D., Foote, W. C., & Maciulis, A. (1985). Chromosome banding pattern homologies and NORs for the Bactrian camel, guanaco, and llama. *J Hered, 76*, 115-118.

Burger, P. A. (2016). The history of Old World camelids in the light of molecular genetics. *Trop Anim Health Prod, 48*(5), 905-913. doi:10.1007/s11250-016-1032-7

Cai, Z., Guldbrandtsen, B., Lund, M. S., & Sahana, G. (2018). Prioritizing candidate genes post-GWAS using multiple sources of data for mastitis resistance in dairy cattle. *BMC Genomics, 19*(1), 656. doi:10.1186/s12864-018-5050-x

Candille, S. I., Kaelin, C. B., Cattanach, B. M., Yu, B., Thompson, D. A., Nix, M. A., . . . Barsh, G. S. (2007). A -defensin mutation causes black coat color in domestic dogs. *Science, 318*(5855), 1418-1423. doi:10.1126/science.1147880

Capanna, E. (1965). The chromosomes of the three species of neotropical Camelidae. Mammal. *Chrom. Newsl., 17*, 75-79.

Cardoso, D. F., de Albuquerque, L. G., Reimer, C., Qanbari, S., Erbe, M., do Nascimento, A. V., . . . Tonhati, H. (2018). Genome-wide scan reveals population stratification and footprints of recent selection in Nelore cattle. *Genet Sel Evol, 50*(1), 22. doi:10.1186/s12711-018-0381-2

Castle, W. E. (1905). *Heredity of Coat Characters in Guinea-Pigs and Rabbits* (Vol. 1): Carnegie Institution.

Cebra, C., Anderson, D. E., Tibary, A., Van Saun, R. J., & Johnson, L. W. (2014). *Llama and Alpaca Care-E-Book: Medicine, Surgery, Reproduction, Nutrition, and Herd Health*: Elsevier Health Sciences.

Cebra, C. K., Tornquist, S. J., Jester, R. M., & Stelletta, C. (2004). Assessment of the effects of feed restriction and amino acid supplementation on glucose tolerance in llamas. *Am J Vet Res, 65*(7), 996-1001.

Cebra, C. K., Tornquist, S. J., Van Saun, R. J., & Smith, B. B. (2001). Glucose tolerance testing in llamas and alpacas. *Am J Vet Res, 62*(5), 682-686.

Chandramohan, B., Renieri, C., La Manna, V., & La Terza, A. (2013). The alpaca agouti gene: genomic locus, transcripts and causative mutations of eumelanic and pheomelanic coat color. *Gene, 521*(2), 303-310. doi:10.1016/j.gene.2013.03.060

Chen, N., Cai, Y., Chen, Q., Li, R., Wang, K., Huang, Y., . . . Lei, C. (2018). Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat Commun, 9*(1), 2337-2337. doi:10.1038/s41467-018-04737-0

Chen, Y., Zhu, J., Lum, P. Y., Yang, X., Pinto, S., MacNeil, D. J., . . . Schadt, E. E. (2008). Variations in DNA elucidate molecular networks that cause disease. *Nature, 452*(7186), 429-435. doi:10.1038/nature06757

Cheruiyot, E. K., Bett, R. C., Amimo, J. O., Zhang, Y., Mrode, R., & Mujibi, F. D. N. (2018). Signatures of Selection in Admixed Dairy Cattle in Tanzania. *Front Genet, 9*, 607. doi:10.3389/fgene.2018.00607

Chilliard, Y., Delavaud, C., & Bonnet, M. (2005). Leptin expression in ruminants: Nutritional and physiological regulations in relation with energy metabolism. *Domestic Animal Endocrinology, 29*(1), 3-22. doi:https://doi.org/10.1016/j.domaniend.2005.02.026

Chiricozzi, A., De Simone, C., Fossati, B., & Peris, K. (2019). Emerging treatment options for the treatment of moderate to severe plaque psoriasis and psoriatic arthritis: evaluating bimekizumab and its therapeutic potential. *Psoriasis: Targets and Therapy, 9*, 29.

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics, 31*(16), 2745-2747. doi:10.1093/bioinformatics/btv195

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS One, 7*(10), e46688. doi:10.1371/journal.pone.0046688

Chowdhary, B. P., Raudsepp, T., Fronicke, L., & Scherthan, H. (1998). Emerging patterns of comparative genome organization in some mammalian species as revealed by Zoo-FISH. *Genome Res, 8*(6), 577-589.

Cieslak, M., Reissmann, M., Hofreiter, M., & Ludwig, A. (2011). Colours of domestication. *Biological Reviews, 86*(4), 885-899.

Cieslak, M., Reissmann, M., Hofreiter, M., & Ludwig, A. (2011). Colours of domestication. *Biol Rev Camb Philos Soc, 86*(4), 885-899. doi:10.1111/j.1469-185X.2011.00177.x

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., . . . Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly, 6*(2), 80-92.

Cohen, J. (2018). Llama antibodies inspire gene spray to prevent all flus. *American Association for the Advancement of Science.* 511-511.

Cook, D., Brooks, S., Bellone, R., & Bailey, E. (2008). Missense mutation in exon 2 of *SLC36A1* responsible for champagne dilution in horses. *PLoS Genetics, 4*(9), e1000195. doi:10.1371/journal.pgen.1000195

Cransberg, R., & Munyard, K. A. (2011). Polymorphisms detected in the tyrosinase and *matp* (*slc45a2*) genes did not explain coat colour dilution in a sample of Alpaca (*Viguna pacos*). *Small Ruminant Research, 95*, 92-96.

Crowley, S. L. (2014). Camels out of place and time: the dromedary (*Camelus dromedarius*) in Australia. *Anthrozoös, 27*(2), 191-203.

Daley, L., Kutzler, M., Bennett, B., Smith, M., Glaser, A., & Appleton, J. (2010). Effector functions of camelid heavy-chain antibodies in immunity to West Nile virus. *Clin. Vaccine Immunol., 17*(2), 239-246.

Damas, J., O'Connor, R., Farre, M., Lenis, V. P. E., Martell, H. J., Mandawala, A., . . . Larkin, D. M. (2017). Upgrading short-read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. *Genome Res, 27*(5), 875-884. doi:10.1101/gr.213660.116

Darwin, C. (1868). The Variation of Animals and Plants under Domestication. *The British and Foreign Medico-Chirurgical Review, 42*(83), 143-166.

Davidson, A., Jaine, T., & Vannithone, S. (2014). *The Oxford Companion to Food* (3rd edition. ed.). Oxford: Oxford University Press.

de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A., & Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS genetics, 7*(12), e1002384-e1002384. doi:10.1371/journal.pgen.1002384

del Marmol, V., & Beermann, F. (1996). Tyrosinase and related proteins in mammalian pigmentation. *FEBS Lett, 381*(3), 165-168.

Delavaud, C., Bengoumi, M., Faye, B., Levieux, D., & Chilliard, Y. (2013). Plasma leptin, glucose and non-esterified fatty acid variations in dromedary camels exposed to prolonged periods of underfeeding or dehydration. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 166*(1), 177-185. doi:https://doi.org/10.1016/j.cbpa.2013.05.026

Detalle, L., Stohr, T., Palomo, C., Piedra, P. A., Gilbert, B. E., Mas, V., . . . Allosery, K. (2016). Generation and characterization of ALX-0171, a potent novel therapeutic nanobody for the treatment of respiratory syncytial virus infection. *Antimicrobial Agents and Chemotherapy, 60*(1), 6-13.

Dittmann, M. T., Hummel, J., Runge, U., Galeffi, C., Kreuzer, M., & Clauss, M. (2014). Characterising an artiodactyl family inhabiting arid habitats by its metabolism: Low metabolism and maintenance requirements in camelids. *Journal of Arid Environments, 107*, 41-48. doi:https://doi.org/10.1016/j.jaridenv.2014.04.005

Do, D. N., Ostersen, T., Strathe, A. B., Mark, T., Jensen, J., & Kadarmideen, H. N. (2014). Genome-wide association and systems genetic analyses of residual feed intake, daily feed consumption, backfat and weight gain in pigs. *BMC Genet, 15*, 27. doi:10.1186/1471-2156-15-27

Drew, M. L., Meyers-Wallen, V., Acland, G., Guyer, C., & Steinheimer, D. (1999). Presumptive Sry-negative XX sex reversal in a llama with multiple congenital anomalies. *Journal of the American Veterinary Medical Association, 215*(8), 1134-1139.

Ducos, A., Berland, H. M., Bonnet, N., Calgaro, A., Billoux, S., Mary, N., . . . Pinton, A. (2007). Chromosomal control of pig populations in France: 2002-2006 survey. *Genet Sel Evol, 39*(5), 583-597. doi:10.1051/gse:2007023

Dunn, L. (1921). Unit character variation in rodents. *J. Mamm, 2*(3), 12.

Durig, N., Letko, A., Lepori, V., Hadji Rasouliha, S., Loechel, R., Kehl, A., . . . Leeb, T. (2018). Two MC1R loss-of-function alleles in cream-coloured Australian Cattle Dogs and white Huskies. *Anim Genet, 49*(4), 284-290. doi:10.1111/age.12660

Eiwishy, A. B. (1987). Reproduction in the female dromedary (*Camelus dromedarius*): A review. *Animal Reproduction Science, 15*(3), 273-297. doi:https://doi.org/10.1016/0378-4320(87)90049-2

Eizirik, E., Yuhki, N., Johnson, W. E., Menotti-Raymond, M., Hannah, S. S., & O'Brien, S. J. (2003). Molecular genetics and evolution of melanism in the cat family. *Curr Biol, 13*(5), 448-453.

Elbers, J. P., Rogers, M. F., Perelman, P. L., Proskuryakova, A. A., Serdyukova, N. A., Johnson, W. E., . . . Burger, P. A. (2019). Improving Illumina assemblies with Hi-C and long reads: An example with the North African dromedary. *Molecular Ecology Resources, 19*(4), 1015-1026. doi:10.1111/1755-0998.13020

elWishy, A. B. (1988). A study of the genital organs of the female dromedary (*Camelus dromedarius*). *J Reprod Fertil, 82*(2), 587-593. doi:10.1530/jrf.0.0820587

Fages, A., Hanghøj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., . . . Orlando, L. (2019). Tracking Five Millennia of Horse Management with Extensive Ancient Genome Time Series. *Cell, 177*(6), 1419-1435.e1431. doi:10.1016/j.cell.2019.03.049

Farag, T, 2003. Emirates home to new hybrid species: Cama. Retrieved from http://ambassadors.net/archives/issue12/features.htm

Faye, B. (2015). Role, distribution and perspective of camel breeding in the third millennium economies. *Emirates Journal of Food and Agriculture*, Vol. 27, 318-327.

Faye, B., Bengoumi, M., Cleradin, A., Tabarani, A., & Chilliard, Y. (2001). Body condition score in dromedary camel: A tool for management of reproduction. *Emirates Journal of Food and Agriculture*, 1-6.

Faye, B., Bengoumi, M., Messad, S., & Chilliard, Y. (2002). Assessment of Body Reserves in Camels. *Revue D Elevage Et De Medicine Veterinaire Des Pays Tropicaux, 55*(1), 69-78.

Feeley, N. L., Bottomley, S., & Munyard, K. A. (2011). Three novel mutations in *ASIP* associated with black bibre in alpacas (*Vicugna pacos*). *Journal of Agricultural Science, 149*, 529-538.

Feeley, N. L., & Munyard, K. A. (2009). Characterisation of the melanocortin-1 receptor gene in alpaca and identification of possible markers associated with phenotypic variations in colour. *Animal Production Science, 49*, 675-681.

Ferguson-Smith, M. (1960). Cytogenetics in man. *AMA Archives of Internal Medicine, 105*(4), 627-639.

Ferguson-Smith, M. A., & Trifonov, V. (2007). Mammalian karyotype evolution. *Nature Reviews Genetics, 8*, 950. doi:10.1038/nrg2199

Ferguson-Smith, M. A., & Trifonov, V. (2007). Mammalian karyotype evolution. *Nat Rev Genet, 8*(12), 950-962. doi:10.1038/nrg2199

Fernandez, A. I., Perez-Montarelo, D., Barragan, C., Ramayo-Caldas, Y., Ibanez-Escriche, N., Castello, A., . . . Rodriguez, M. C. (2012). Genome-wide linkage analysis of QTL for growth and body composition employing the PorcineSNP60 BeadChip. *BMC Genet, 13*, 41. doi:10.1186/1471-2156-13-41

Fitak, R. R., Mohandesan, E., Corander, J., & Burger, P. A. (2016). The de novo genome assembly and annotation of a female domestic dromedary of North African origin. *Mol Ecol Resour, 16*(1), 314-324. doi:10.1111/1755-0998.12443

Flajnik, M. F., Deschacht, N., & Muyldermans, S. (2011). A case of convergence: why did a simple alternative to canonical antibodies arise in sharks and camels? *PLoS Biology, 9*(8), e1001120.

Fowler, M. E., & Bravo, P. W. (2010). Congenital/hereditary conditions. *Medicine and Surgery of Camelids*, 525-558.

Franklin, W. L. (1982). Biology, ecology, and relationship to man of the South American camelids. *Mammalian Biology in South America, 6*, 457-489.

Fritz, K., Kaese, H., Valberg, S., Hendrickson, J. A., Rendahl, A., Bellone, R., . . . Cuomo, F. (2014). Genetic risk factors for insidious equine recurrent uveitis in A ppaloosa horses. *Animal Genetics, 45*(3), 392-399.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res, 31*(13), 3784-3788.

Gatesy, J. (2002). Relative quality of different systematic datasets for cetartiodactyl mammals: assessments within a combined analysis framework. *Exs*(92), 45-67.

Gersen, S. L., Keagle, M. B., Gersen, S., & Keagle, M. (2005). *The Principles of Clinical Cytogenetics.* Springer Publishers, New York.

Ghosh, M., Sharma, N., Singh, A. K., Gera, M., Pulicherla, K. K., & Jeong, D. K. (2018). Transformation of animal genomics by next-generation sequencing technologies: a decade of challenges and their impact on genetic architecture. *Crit Rev Biotechnol, 38*(8), 1157-1175. doi:10.1080/07388551.2018.1451819

Giuffra, E., Törnsten, A., Marklund, S., Bongcam-Rudloff, E., Chardon, P., Kijas, J. M., . . . Andersson, L. (2002). A large duplication associated with dominant white color in pigs originated by homologous recombination between LINE elements flanking *KIT*. *Mammalian Genome, 13*(10), 569-577.

Goldammer, T., Di Meo, G. P., Luhken, G., Drogemuller, C., Wu, C. H., Kijas, J., . . . Cockett, N. E. (2009). Molecular cytogenetics and gene mapping in sheep (*Ovis aries*, 2n = 54). *Cytogenet Genome Res, 126*(1-2), 63-76. doi:10.1159/000245907

Goldman, E. R., Anderson, G. P., Liu, J. L., Delehanty, J. B., Sherwood, L. J., Osborn, L. E., . . . Hayhurst, A. (2006). Facile generation of heat-stable antiviral and antitoxin single domain antibodies from a semisynthetic llama library. *Analytical chemistry, 78*(24), 8245-8255.

Griffin, L. M., Snowden, J. R., Lawson, A. D., Wernery, U., Kinne, J., & Baker, T. S. (2014). Analysis of heavy and light chain sequences of conventional camelid antibodies from *Camelus dromedarius* and *Camelus bactrianus* species. *Journal of Immunological Methods, 405*, 35-46.

Grigson, C. (2014). The history of the camel bone dating project. *Anthropozoologica, 49*(2), 225-236.

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., . . . Weigend, S. (2010). Genetic diversity in farm animals--a review. *Anim Genet, 41 Suppl 1*, 6-31. doi:10.1111/j.1365-2052.2010.02038.x

Guo, J., Tao, H., Li, P., Li, L., Zhong, T., Wang, L., . . . Zhang, H. (2018). Whole-genome sequencing reveals selection signatures associated with important traits in six goat breeds. *Sci Rep, 8*(1), 10405.

Gurgul, A., Jasielczuk, I., Semik-Gurgul, E., Pawlina-Tyszko, K., Stefaniuk-Szmukier, M., Szmatola, T., . . . Bugno-Poniewierska, M. (2019). A genome-wide scan for diversifying selection signatures in selected horse breeds. *PLoS One, 14*(1), e0210751. doi:10.1371/journal.pone.0210751

Guridi, M., Soret, B., Alfonso, L., & Arana, A. (2011). Single nucleotide polymorphisms in the Melanocortin 1 Receptor gene are linked with lightness of fibre colour in Peruvian Alpaca (*Vicugna pacos*). *Anim Genet, 42*(6), 679-682. doi:10.1111/j.1365-2052.2011.02205.x

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. *Cytogenet Genome Res, 102*(1-4), 189-195. doi:10.1159/000075747

Gutiérrez, J., Varona, L., Pun, A., Morante, R., Burgos, A., Cervantes, I., & Pérez-Cabal, M. (2011). Genetic parameters for growth of fiber diameter in alpacas. *Journal of Animal Science, 89*(8), 2310-2315.

Guyonneau, L., Murisier, F., Rossier, A., Moulin, A., & Beermann, F. (2004). Melanocytes and pigmentation are affected in dopachrome tautomerase knockout mice. *Mol Cell Biol, 24*(8), 3396-3403. doi:10.1128/mcb.24.8.3396-3403.2004

Haldane, J. B. S. (1927). The comparative genetics of colour in rodents and carnivora. *Biological Reviews, 2*(3), 199-212.

Hamers-Casterman, C., Atarhouch, T., Muyldermans, S., Robinson, G., Hammers, C., Songa, E. B., . . . Hammers, R. (1993). Naturally occurring antibodies devoid of light chains. *Nature, 363*(6428), 446.

Hare, J. 2008. *Camelus ferus*. *The IUCN Red List of Threatened Species* 2008. Retrieved from http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T63543A12689285.en.

Hassanin, A., Delsuc, F., Ropiquet, A., Hammer, C., Jansen van Vuuren, B., Matthee, C., . . . Couloux, A. (2012). Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C R Biol, 335*(1), 32-50. doi:10.1016/j.crvi.2011.11.002

Hayes, B. J., Lewin, H. A., & Goddard, M. E. (2013). The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends in Genetics, 29*(4), 206-214.

Hemida, M. G., Chu, D. K., Poon, L. L., Perera, R. A., Alhammadi, M. A., Ng, H. Y., . . . Peiris, M. (2014). MERS coronavirus in dromedary camel herd, Saudi Arabia. *Emerg Infect Dis, 20*(7), 1231-1234. doi:10.3201/eid2007.140571

Hinrichs, K., Horin, S. E., Buoen, L. C., Zhang, T. Q., & Ruth, G. R. (1997). X-chromosome monosomy in an infertile female llama. *J Am Vet Med Assoc, 210*(10), 1503-1504.

Hisgett, T. 2016 flickr. Retrieved from https://www.flickr.com/photos/hisgett/31562329701

Hofstetter, S., Seefried, F., Hafliger, I. M., Jagannathan, V., Leeb, T., & Drogemuller, C. (2019). A non-coding regulatory variant in the 5'-region of the *MITF* gene is associated with white-spotted coat in Brown Swiss cattle. *Anim Genet, 50*(1), 27-32. doi:10.1111/age.12751

Holl, H., Isaza, R., Mohamoud, Y., Ahmed, A., Almathen, F., Youcef, C., . . . Brooks, S. (2017). A Frameshift Mutation in *KIT* is Associated with White Spotting in the Arabian Camel. *Genes (Basel), 8*(3). doi:10.3390/genes8030102

Holz, J.-B. (2012). The TITAN trial–assessing the efficacy and safety of an anti-von Willebrand factor Nanobody in patients with acquired thrombotic thrombocytopenic purpura. *Transfusion and Apheresis Science, 46*(3), 343-346.

Hungerford, D., & Snyder, R. (1966). Chromosomes of European wolf (*Canis lupus*) and of a Bactrian camel (*Camelus bactrianus*). *Mamm Chrom Newsl, 20*, 72.

Ingvartsen, K. L., & Boisclair, Y. (2001). Leptin and the regulation of food intake, energy homeostasis and immunity with special focus on periparturient ruminants. *Domestic Animal Endocrinology, 21*(4), 215-250.

Jackling, F. C., Johnson, W. E., & Appleton, B. R. (2014). The genetic inheritance of the blue-eyed white phenotype in alpacas (*Vicugna pacos*). *J Hered, 105*(6), 847-857. doi:10.1093/jhered/ess093

Jähnichen, S., Blanchetot, C., Maussang, D., Gonzalez-Pajuelo, M., Chow, K. Y., Bosch, L., . . . Vandevelde, W. (2010). CXCR4 nanobodies (VHH-based single variable domains) potently inhibit chemotaxis and HIV-1 replication and mobilize stem cells. *Proceedings of the National Academy of Sciences, 107*(47), 20565-20570.

Ji, R., Cui, P., Ding, F., Geng, J., Gao, H., Zhang, H., . . . Meng, H. (2009). Monophyletic origin of domestic bactrian camel (*Camelus bactrianus*) and its evolutionary relationship with the extant wild camel (*Camelus bactrianus ferus*). *Animal Genetics, 40*(4), 377-382.

Jones, C., Abd-Elnaeim, M., Bevilacqua, E., Oliveira, L., & Leiser, R. (2002). Comparison of uteroplacental glycosylation in the camel (*Camelus dromedarius*) and alpaca (*Lama pacos*). *Reproduction, 123*(1), 115-126.

Jones, C. J., Skidmore, J. A., & Aplin, J. D. (2008). Placental glycosylation in a cama (camel-llama cross) and its relevance to successful hybridisation. *Mol Phylogenet Evol, 49*(3), 1030-1035. doi:10.1016/j.ympev.2008.09.024

Kaas, J. H. (2005). Serendipity and the Siamese cat: the discovery that genes for coat and eye pigment affect the brain. *ILAR journal, 46*(4), 357-363.

Kadim, I., Mahgoub, O., & Purchas, R. (2008). A review of the growth, and of the carcass and meat quality characteristics of the one-humped camel (*Camelus dromedaries*). *Meat Science, 80*(3), 555-569.

Kadwell, M., Fernandez, M., Stanley, H. F., Baldi, R., Wheeler, J. C., Rosadio, R., & Bruford, M. W. (2001). Genetic analysis reveals the wild ancestors of the llama and the alpaca. *Proceeding, Biological Sciences, 268*(1485), 2575-2584. doi:10.1098/rspb.2001.1774

Kerns, J. A., Newton, J., Berryere, T. G., Rubin, E. M., Cheng, J. F., Schmutz, S. M., & Barsh, G. S. (2004). Characterization of the dog Agouti gene and a nonagoutimutation in German Shepherd Dogs. *Mamm Genome, 15*(10), 798-808. doi:10.1007/s00335-004-2377-1

Khalkhali-Evrigh, R., Hedayat-Evrigh, N., Hafezian, S. H., Farhadi, A., & Bakhtiarizadeh, M. R. (2019). Genome-Wide Identification of Microsatellites and Transposable Elements in the Dromedary Camel Genome Using Whole-Genome Sequencing Data. *Front Genet, 10*, 692. doi:10.3389/fgene.2019.00692

Kim, K. S., Kim, S. W., Raney, N. E., & Ernst, C. W. (2012). Evaluation of BTA1 and BTA5 QTL Regions for Growth and Carcass Traits in American and Korean Cattle. *Asian-Australas J Anim Sci, 25*(11), 1521-1528. doi:10.5713/ajas.2012.12218

Kiser, J. N., Keuter, E. M., Seabury, C. M., Neupane, M., Moraes, J. G. N., Dalton, J., . . . Neibergs, H. L. (2019). Validation of 46 loci associated with female fertility traits in cattle. *BMC Genomics, 20*(1), 576. doi:10.1186/s12864-019-5935-3

Klungland, H., Vage, D., Gomez-Raya, L., Adalsteinsson, S., & Lien, S. (1995). The role of melanocyte-stimulating hormone (*MSH*) receptor in bovine coat color determination. *Mammalian Genome, 6*(9), 636-639.

Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R. V., Nolte, V., Futschik, A., . . . Schlotterer, C. (2011). PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One, 6*(1), e15925. doi:10.1371/journal.pone.0015925

Kofler, R., Pandey, R. V., & Schlotterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics, 27*(24), 3435-3436. doi:10.1093/bioinformatics/btr589

Könning, D., Zielonka, S., Grzeschik, J., Empting, M., Valldorf, B., Krah, S., . . . Kolmar, H. (2017). Camelid and shark single domain antibodies: structural features and therapeutic potential. *Current Opinion in Structural Biology, 45*, 10-16.

Konuspayeva, G., Faye, B., & Loiseau, G. (2009). The composition of camel milk: a meta-analysis of the literature data. *Journal of Food Composition and Analysis, 22*(2), 95-101.

Koulischer, L. (1971). Mammalian cytogenetics IV. The chromosomes of two male Camelidae: *Camelus bacterianus* and *Lama vicugna*. *Acta Zool Pathol Antverp, 52*, 89-92.

Kozhamkulova, B. (1986). The late Cenozoic two-humped (Bactrian) camels of Asia. *Quartärpläontolgie (Abh. Ber. Inst. Quartärpläontolgie Weimar), 6*, 93-97.

Kulemzina, A. I., Perelman, P. L., Grafodatskaya, D. A., Nguyen, T. T., Thompson, M., Roelke-Parker, M. E., & Graphodatsky, A. S. (2014). Comparative chromosome painting of pronghorn (*Antilocapra americana*) and saola (*Pseudoryx nghetinhensis*) karyotypes with human and dromedary camel probes. *BMC Genetics, 15*(1), 68.

Langdon, Q. K., Peris, D., Kyle, B., & Hittinger, C. T. (2018). sppIDer: A Species Identification Tool to Investigate Hybrid Genomes with High-Throughput Sequencing. *Mol Biol Evol, 35*(11), 2835-2849. doi:10.1093/molbev/msy166

Lechner-Doll, M., Von Engelhardt, W., Abbas, H., Mousa, L., Luciano, L., & Reale, E. (1995). Particularities in forestomach anatomy, physiology and biochemistry of camelids compared to ruminants. *Elevage et Alimentation du Dromadaire–Camel Production and Nutrition Options Méditerranéennes, Serie B. Etudes et Recherches*(13), 19-32.

Li, B., Qiao, L., An, L., Wang, W., Liu, J., Ren, Y., . . . Liu, W. (2018). Transcriptome analysis of adipose tissues from two fat-tailed sheep breeds reveals key genes involved in fat deposition. *BMC genomics, 19*(1), 338. doi:10.1186/s12864-018-4747-1

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics, 27*(21), 2987-2993. doi:10.1093/bioinformatics/btr509

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25*(14), 1754-1760.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature, 475*(7357), 493.

Li, J., Bed'hom, B., Marthey, S., Valade, M., Dureux, A., Moroldo, M., . . . Tixier-Boichard, M. (2018). A missense mutation in *TYRP1* causes the chocolate plumage color in chicken and alters melanosome structure. *Pigment Cell Melanoma Res*. doi:10.1111/pcmr.12753

Librado, P., Gamba, C., Gaunitz, C., Der Sarkissian, C., Pruvost, M., Albrechtsen, A., . . . Orlando, L. (2017). Ancient genomic changes associated with domestication of the horse. *Science, 356*(6336), 442-445. doi:10.1126/science.aam5298

Lindholm-Perry, A. K., Kuehn, L. A., Smith, T. P., Ferrell, C. L., Jenkins, T. G., Freetly, H. C., & Snelling, W. M. (2012). A region on BTA14 that includes the positional candidate genes *LYPLA1*, *XKR4* and *TMEM68* is associated with feed intake and growth phenotypes in cattle(1). *Anim Genet, 43*(2), 216-219. doi:10.1111/j.1365-2052.2011.02232.x

Little, C. (1958). Coat color genes in rodents and carnivores. *The Quarterly Review of Biology, 33*(2), 103-137.

Liu, J., Liang, S., Liu, X., Brown, J. A., Newman, K. E., Sunkara, M., . . . Graf, G. A. (2012). The absence of ABCD2 sensitizes mice to disruptions in lipid metabolism by dietary erucic acid. *J Lipid Res, 53*(6), 1071-1079. doi:10.1194/jlr.M022160

Liu, S., He, S., Chen, L., Li, W., Di, J., & Liu, M. (2017). Estimates of linkage disequilibrium and effective population sizes in Chinese Merino (*Xinjiang type*) sheep by genome-wide SNPs. *Genes Genomics, 39*(7), 733-745. doi:10.1007/s13258-017-0539-2

Magalhaes, A. F., de Camargo, G. M., Fernandes, G. A. J., Gordo, D. G., Tonussi, R. L., Costa, R. B., . . . de Albuquerque, L. G. (2016). Genome-Wide Association Study of Meat Quality Traits in Nellore Cattle. *PLoS One, 11*(6), e0157845. doi:10.1371/journal.pone.0157845

Mahmoud, A., Alshaikh, M., Aljumaah, R., & Mohammed, O. (2012). Genetic variability of camel (*Camelus dromedarius*) populations in Saudi Arabia based on microsatellites analysis. *African Journal of Biotechnology, 11*(51), 11173-11180.

Mariat, D., Taourit, S., & Guerin, G. (2003). A mutation in the *MATP* gene causes the cream coat colour in the horse. *Genet Sel Evol, 35*(1), 119-133. doi:10.1051/gse:2002039

Marin, J. C., Rivera, R., Varas, V., Cortes, J., Agapito, A., Chero, A., . . . Orozco-terWengel, P. (2018). Genetic Variation in Coat Colour Genes *MC1R* and *ASIP* Provides Insights Into Domestication and Management of South American Camelids. *Front Genet, 9*, 487. doi:10.3389/fgene.2018.00487

Marklund, L., Moller, M. J., Sandberg, K., & Andersson, L. (1996). A missense mutation in the gene for melanocyte-stimulating hormone receptor (*MC1R*) is associated with the chestnut coat color in horses. *Mamm Genome, 7*(12), 895-899.

Matarese, G. (2000). Leptin and the immune system: how nutritional status influences the immune response. *European Cytokine Network, 11*(1), 7-14.

Maussang, D., Mujić-Delić, A., Descamps, F. J., Stortelers, C., Vanlandschoot, P., Stigter-van Walsum, M., . . . Gonzalez-Pajuelo, M. (2013). Llama-derived single variable domains (nanobodies) directed against chemokine receptor CXCR7 reduce head and neck cancer cell growth in vivo. *Journal of Biological Chemistry, 288*(41), 29562-29572.

McCoy, L. E., Quigley, A. F., Strokappe, N. M., Bulmer-Thomas, B., Seaman, M. S., Mortier, D., . . . Weiss, R. A. (2012). Potent and broad neutralization of HIV-1 by a llama antibody elicited by immunization. *J Exp Med, 209*(6), 1091-1103. doi:10.1084/jem.20112655

McKinley, M. J., McBurnie, M. I., & Mathai, M. L. (2001). Neural mechanisms subserving central angiotensinergic influences on plasma renin in sheep. *Hypertension, 37*(6), 1375-1381. doi:10.1161/01.hyp.37.6.1375

Mendoza, M. N., Raudsepp, T., Alshanbari, F., Gutiérrez, G., & Ponce de León, F. A. (2019). Chromosomal Localization of Candidate Genes for Fiber Growth and Color in Alpaca (*Vicugna pacos*). *Front Genet, 10*, 583-583. doi:10.3389/fgene.2019.00583

Mesbah-Uddin, M., Guldbrandtsen, B., Iso-Touru, T., Vilkki, J., De Koning, D. J., Boichard, D., . . . Sahana, G. (2017). Genome-wide mapping of large deletions and their population-genetic properties in dairy cattle. *DNA Res*. doi:10.1093/dnares/dsx037

Messer, P. W., & Neher, R. A. (2012). Estimating the strength of selective sweeps from deep population diversity data. *Genetics, 191*(2), 593-605. doi:10.1534/genetics.112.138461

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

Miller, S. M., Guthrie, A. J., & Harper, C. K. (2016). Single base-pair deletion in *ASIP* exon 3 associated with recessive black phenotype in impala (*Aepyceros melampus*). *Anim Genet, 47*(4), 511-512. doi:10.1111/age.12430

Mohandesan, E., Speller, C. F., Peters, J., Uerpmann, H.-P., Uerpmann, M., De Cupere, B., . . . Burger, P. A. (2017). Combined hybridization capture and shotgun sequencing for ancient DNA analysis of extinct wild and domestic dromedary camel. *Molecular Ecology Resources, 17*(2), 300-313. doi:10.1111/1755-0998.12551

Moller, S., Croning, M. D., & Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics, 17*(7), 646-653.

Monegal, A., Ami, D., Martinelli, C., Huang, H., Aliprandi, M., Capasso, P., . . . de Marco, A. (2009). Immunological applications of single-domain llama recombinant antibodies isolated from a naive library. *Protein Engineering, Design & Selection, 22*(4), 273-280.

Montoliu L, Oetting WS, Bennett DC. Color Genes. October 2011. European Society for Pigment Cell Research. World Wide Web, http://www.espcr.org/micemut

Morante, R., Goyache, F., Burgos, A., Cervantes, I., Pérez-Cabal, M. A., & Gutiérrez, J. P. (2009). Genetic improvement for alpaca fibre production in the Peruvian Altiplano: the Pacomarca experience. *Animal Genetic Resources Information, 45*, 37–43. doi:10.1017/S1014233909990307

Mukiibi, R., Vinsky, M., Keogh, K. A., Fitzsimmons, C., Stothard, P., Waters, S. M., & Li, C. (2018). Transcriptome analyses reveal reduced hepatic lipid synthesis and accumulation in more feed efficient beef cattle. *Sci Rep, 8*(1), 7303. doi:10.1038/s41598-018-25605-3

Murisier, F., & Beermann, F. (2006). Genetics of pigment cells: lessons from the tyrosinase gene family. *Histol Histopathol, 21*(5), 567-578. doi:10.14670/hh-21.567

Murphy, W. J., Larkin, D. M., Everts-van der Wind, A., Bourque, G., Tesler, G., Auvil, L., . . . Lewin, H. A. (2005). Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science, 309*(5734), 613-617. doi:10.1126/science.1111387

Nazari-Ghadikolaei, A., Mehrabani-Yeganeh, H., Miarei-Aashtiani, S. R., Staiger, E. A., Rashidi, A., & Huson, H. J. (2018). Genome-Wide Association Studies Identify Candidate Genes for Coat Color and Mohair Traits in the Iranian Markhoz Goat. *Front Genet, 9*, 105. doi:10.3389/fgene.2018.00105

Nelson, K., Bwala, D., & Nuhu, E. (2015). The dromedary camel; a review on the aspects of history, physical description, adaptations, behavior/lifecycle, diet,

reproduction, uses, genetics and diseases. *Nigerian Veterinary Journal, 36*(4), 1299-1317.

Norris, B. J., & Whan, V. A. (2008). A gene duplication affecting expression of the ovine *ASIP* gene is responsible for white and black sheep. *Genome Res, 18*(8), 1282-1293. doi:10.1101/gr.072090.107

Olivera, L. V., Zago, D. A., Jones, C. J., & Bevilacqua, E. (2003). Developmental changes at the materno-embryonic interface in early pregnancy of the alpaca, *Lamos pacos*. *Anat Embryol (Berl), 207*(4-5), 317-331. doi:10.1007/s00429-003-0346-1

Olivieri, B. F., Mercadante, M. E., Cyrillo, J. N., Branco, R. H., Bonilha, S. F., de Albuquerque, L. G., . . . Baldi, F. (2016). Genomic Regions Associated with Feed Efficiency Indicator Traits in an Experimental Nellore Cattle Population. *PLoS One, 11*(10), e0164390. doi:10.1371/journal.pone.0164390

Palladino, M. A., Spencer, C. A., Cummings, M. R., & Klug, W. S. (2015). *Concepts of Genetics*. 11th Edition. Pearson Higher Ed, New York.

Parent, A.-S., Lebrethon, M.-C., Gerard, A., Vandersmissen, E., & Bourguignon, J.-P. (2000). Leptin effects on pulsatile gonadotropin releasing hormone secretion from the adult rat hypothalamus and interaction with cocaine and amphetamine regulated transcript peptide and neuropeptide Y. *Regulatory Peptides, 92*(1-3), 17-24.

Perelman, P. L., Pichler, R., Gaggl, A., Larkin, D. M., Raudsepp, T., Alshanbari, F., . . . Periasamy, K. (2018). Construction of two whole genome radiation hybrid panels for dromedary (*Camelus dromedarius*): 5000RAD and 15000RAD. *Sci Rep, 8*(1), 1982. doi:10.1038/s41598-018-20223-5

Peters, J., & Driesch, A. v. d. (1997). The two-humped camel (*Camelus bactrianus*): new light on its distribution, management and medical treatment in the past. *Journal of Zoology, 242*(4), 651-679.

Peyvandi, F., Scully, M., Kremer Hovinga, J. A., Cataland, S., Knöbl, P., Wu, H., . . . Jilma, B. (2016). Caplacizumab for acquired thrombotic thrombocytopenic purpura. *New England Journal of Medicine, 374*(6), 511-522.

Pielberg, G. (2004). *Molecular Coat Color Genetics.* (Doctoral), Swedish University of Agricultural Sciences, Uppsala.

Pitt, D., Sevane, N., Nicolazzi, E. L., MacHugh, D. E., Park, S. D. E., Colli, L., . . . Orozco-terWengel, P. (2018). Domestication of cattle: Two or three events? *Evolutionary Applications, 12*(1), 123-136. doi:10.1111/eva.12674

Plasil, M., Mohandesan, E., Fitak, R. R., Musilova, P., Kubickova, S., Burger, P. A., & Horin, P. (2016). The major histocompatibility complex in Old World camelids and low polymorphism of its class II genes. *BMC Genomics, 17*, 167-167. doi:10.1186/s12864-016-2500-1

Plasil, M., Wijkmark, S., Elbers, J. P., Oppelt, J., Burger, P. A., & Horin, P. (2019). The major histocompatibility complex of Old World camelids: Class I and class I-related genes. *Hla, 93*(4), 203-215. doi:10.1111/tan.13510

Porter, V., Alderson, L., Hall, S. J. G., & Sponenberg, D. P. (2016). *Mason's World Encyclopedia of Livestock Breeds and Breeding : Volume 1*. Wallingford: CABI.

Poth, T., Breuer, W., Walter, B., Hecht, W., & Hermanns, W. (2010). Disorders of sex development in the dog-Adoption of a new nomenclature and reclassification of reported cases. *Anim Reprod Sci, 121*(3-4), 197-207. doi:10.1016/j.anireprosci.2010.04.011

Proskuryakova, A., Kulemzina, A., Perelman, P., Makunin, A., Larkin, D., Farré, M., . . . Beklemisheva, V. (2017). X Chromosome evolution in Cetartiodactyla. *Genes, 8*(9), 216.

Prothero, D. R., & Foss, S. E. (2007). *The Evolution of Artiodactyls*: JHU Press.

Punnett, R., & Punnett, R. (1912). Inheritance of coat-colour in rabbits. *Journal of Genetics, 2*(03).

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841-842.

Ramey, H. R., Decker, J. E., McKay, S. D., Rolf, M. M., Schnabel, R. D., & Taylor, J. F. (2013). Detection of selective sweeps in cattle using genome-wide SNP data. *BMC Genomics, 14*, 382. doi:10.1186/1471-2164-14-382

Ramos, A. M., Duijvesteijn, N., Knol, E. F., Merks, J. W., Bovenhuis, H., Crooijmans, R. P., . . . Harlizius, B. (2011). The distal end of porcine chromosome 6p is involved in the regulation of skatole levels in boars. *BMC Genet, 12*, 35. doi:10.1186/1471-2156-12-35

Raudsepp, T. (2014). Cytogenetics and infertility. *Llama and Alpaca Care* (pp. 243-249): Elsevier, New York.

Raudsepp, T., Avila, F., P. Baily, M., Merriwether, D., & A. Kutzler, M. (2015). Molecular Cytogenetics of the Minute Chromosome Syndrome in Alpacas. *Annual Review of Animal Biosciences*, 4, 15-43.

Raudsepp, T., & Chowdhary, B. P. (2008). FISH for mapping single copy genes. *Methods Mol Biol, 422*, 31-49. doi:10.1007/978-1-59745-581-7_3

Raudsepp, T., & Chowdhary, B. P. (2011). Cytogenetics and chromosome maps. *The Genetics of the Pig: Second Edition*, 134-178. doi:10.1079/9781845937560.0134

Raudsepp, T., & Chowdhary, B. P. (2016). Chromosome Aberrations and Fertility Disorders in Domestic Animals. *Annu Rev Anim Biosci, 4*, 15-43. doi:10.1146/annurev-animal-021815-111239

Raudsepp, T., Gustafson-Seabury, A., Durkin, K., Wagner, M. L., Goh, G., Seabury, C. M., . . . Chowdhary, B. P. (2008). A 4,103 marker integrated physical and comparative map of the horse genome. *Cytogenet Genome Res, 122*(1), 28-36. doi:10.1159/000151313

Rees, J. L. (2003). Genetics of hair and skin color. *Annu Rev Genet, 37*, 67-90. doi:10.1146/annurev.genet.37.110801.143233

Reich, M., Ohm, K., Angelo, M., Tamayo, P., & Mesirov, J. P. (2004). GeneCluster 2.0: an advanced toolset for bioarray analysis. *Bioinformatics, 20*(11), 1797-1798.

Reissmann, M., & Ludwig, A. (2013). Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. *Semin Cell Dev Biol, 24*(6-7), 576-586. doi:10.1016/j.semcdb.2013.03.014

Richardson, M. F., Munyard, K., Croft, L. J., Allnutt, T. R., Jackling, F., Alshanbari, F., . . . Raudsepp, T. (2019). Chromosome-Level Alpaca Reference Genome VicPac3.1 Improves Genomic Insight Into the Biology of New World Camelids. *Front Genet, 10*(586). doi:10.3389/fgene.2019.00586

Rieder, S., Taourit, S., Mariat, D., Langlois, B., & Guerin, G. (2001). Mutations in the agouti (*ASIP*), the extension (*MC1R*), and the brown (*TYRP1*) loci and their association to coat color phenotypes in horses (Equus caballus). *Mamm Genome, 12*(6), 450-455. doi:10.1007/s003350020017

Rodrigues-Peres, R. M., de, S. C. B., Anurag, M., Lei, J. T., Conz, L., Goncalves, R., . . . Sarian, L. O. (2019). Copy number alterations associated with clinical features in an underrepresented population with breast cancer. *Mol Genet Genomic Med, 7*(7), e00750. doi:10.1002/mgg3.750

Root-Bernstein, M, 2014. When rewilding isn't mad: guanacos can transform the espinal of Chile. Retrieved from https://theconversation.com/when-rewilding-isnt-mad-guanacos-can-transform-the-espinal-of-chile-24248

Royo, L. J., Alvarez, I., Arranz, J. J., Fernandez, I., Rodriguez, A., Perez-Pardal, L., & Goyache, F. (2008). Differences in the expression of the ASIP gene are involved in the recessive black coat colour pattern in sheep: evidence from the rare Xalda sheep breed. *Anim Genet, 39*(3), 290-293. doi:10.1111/j.1365-2052.2008.01712.x

Rubin, C. J., Megens, H. J., Martinez Barrio, A., Maqbool, K., Sayyab, S., Schwochow, D., . . . Andersson, L. (2012). Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci U S A, 109*(48), 19529-19536. doi:10.1073/pnas.1217149109

Ruiz-Larranaga, O., Langa, J., Rendo, F., Manzano, C., Iriondo, M., & Estonba, A. (2018). Genomic selection signatures in sheep from the Western Pyrenees. *Genet Sel Evol, 50*(1), 9. doi:10.1186/s12711-018-0378-x

Ruvinskiy, D., Larkin, D. M., & Farre, M. (2019). A Near Chromosome Assembly of the Dromedary Camel Genome. *Front Genet, 10*, 32. doi:10.3389/fgene.2019.00032

Saatchi, M., Schnabel, R. D., Taylor, J. F., & Garrick, D. J. (2014). Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC genomics, 15*, 442. doi:10.1186/1471-2164-15-442

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). *Molecular Cloning: A Laboratory Manual.* (Vol. 61). Cold Spring Harbor: Elsevier. New York.

Samman, M. A., Al-Saleh, A. A., & Sheth, K. (1992). Karyotyping the Arabian Camel (*Camelus dromedarius*), Using C-Banding Technique. *Cytologia, 57*(3), 383-388. doi:10.1508/cytologia.57.383

Schibler, L., Di Meo, G. P., Cribiu, E. P., & Iannuzzi, L. (2009). Molecular cytogenetics and comparative mapping in goats (*Capra hircus*, 2n = 60). *Cytogenet Genome Res, 126*(1-2), 77-85. doi:10.1159/000245908

Schmidt-Nielsen, K. (1959). The physiology of the camel. *Sci Am, 201*, 140-151.

Schmidt-Nielsen, K., Crawford, E. C., Jr., Newsome, A. E., Rawson, K. S., & Hammel, H. T. (1967). Metabolic rate of camels: effect of body temperature and dehydration. *Am J Physiol, 212*(2), 341-346. doi:10.1152/ajplegacy.1967.212.2.341

Schmidt-Nielsen, K., Schmidt-Nielsen, B., Jarnum, S. A., & Houpt, T. R. (1957). Body temperature of the camel and its relation to water economy. *Am J Physiol, 188*(1), 103-112. doi:10.1152/ajplegacy.1956.188.1.103

Schmutz, S. M., & Berryere, T. G. (2007). Genes affecting coat colour and pattern in domestic dogs: a review. *Anim Genet, 38*(6), 539-549. doi:AGE1664 [pii] 10.1111/j.1365-2052.2007.01664.x

Schmutz, S. M., Berryere, T. G., Ciobanu, D. C., Mileham, A. J., Schmidtz, B. H., & Fredholm, M. (2004). A form of albinism in cattle is caused by a tyrosinase frameshift mutation. *Mamm Genome, 15*(1), 62-67. doi:10.1007/ s00335-002-2249-5

Schoenebeck, J. J., & Ostrander, E. A. (2014). Insights into morphology and disease from the dog genome project. *Annu Rev Cell Dev Biol, 30*, 535-560. doi:10.1146/annurev-cellbio-100913-012927

Schutt, S., Florl, A. R., Shi, W., Hemberger, M., Orth, A., Otto, S., . . . Fundele, R. H. (2003). DNA methylation in placentas of interspecies mouse hybrids. *Genetics, 165*(1), 223-228.

Seabury, C. M., Oldeschulte, D. L., Saatchi, M., Beever, J. E., Decker, J. E., Halley, Y. A., . . . Taylor, J. F. (2017). Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics, 18*(1), 386. doi:10.1186/s12864-017-3754-y

Shen, L., Luo, J., Du, J., Liu, C., Wu, X., Pu, Q., . . . Zhu, L. (2015). Transcriptome Analysis of Liangshan Pig Muscle Development at the Growth Curve Inflection Point and Asymptotic Stages Using Digital Gene Expression Profiling. *PLoS One, 10*(8), e0135978. doi:10.1371/journal.pone.0135978

Siengdee, P., Klinhom, S., Thitaram, C., & Nganvongpanit, K. (2018). Isolation and culture of primary adult skin fibroblasts from the Asian elephant (Elephas maximus). *PeerJ, 6*, e4302. doi:10.7717/peerj.4302

Skidmore, J. A., Billah, M., & Allen, W. R. (1996). Patterns of hormone secretion throughout pregnancy in the one-humped camel (*Camelus dromedarius*). *Reprod Fertil Dev, 8*(5), 863-869.

Skidmore, J. A., Billah, M., Binns, M., Short, R. V., & Allen, W. R. (1999). Hybridizing Old and New World camelids: *Camelus dromedarius* x *Lama guanicoe*. *Proceedings: Biological Sciences, 266*(1420), 649-656. doi:10.1098/rspb.1999.0685

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genet Res, 23*(1), 23-35.

Spencer, P. B., Wilson, K. J., & Tinson, A. (2010). Parentage testing of racing camels (*Camelus dromedarius*) using microsatellite DNA typing. *Anim Genet, 41*(6), 662-665. doi:10.1111/j.1365-2052.2010.02044.x

Stanley, H. F., Kadwell, M., & Wheeler, J. C. (1994). Molecular evolution of the family Camelidae: a mitochondrial DNA study. *Proc. R. Soc. Lond. B, 256*(1345), 1-6.

Stein, V. 2018. How llamas could help us fight the flu, Science. Retrieved from https://www.pbs.org/newshour/science/how-llamas-could-help-us-fight-the-flu

Stewart, T. P., Kim, H. Y., Saxton, A. M., & Kim, J. H. (2010). Genetic and genomic analysis of hyperlipidemia, obesity and diabetes using (C57BL/6J x TALLYHO/JngJ) F2 mice. *BMC Genomics, 11*, 713. doi:10.1186/1471-2164-11-713

Stritzel, S., Wohlke, A., & Distl, O. (2009). A role of the microphthalmia-associated transcription factor in congenital sensorineural deafness and eye pigmentation in Dalmatian dogs. *J Anim Breed Genet, 126*(1), 59-62. doi:10.1111/j.1439-0388.2008.00761.x

Strokappe, N., Szynol, A., Aasa-Chapman, M., Gorlani, A., Forsman Quigley, A., Hulsik, D. L., . . . Verrips, T. (2012). Llama antibody fragments recognizing various epitopes of the CD4bs neutralize a broad range of HIV-1 subtypes A, B and C. *PLoS One, 7*(3), e33298. doi:10.1371/journal.pone.0033298

Sturm, R. A., & Duffy, D. L. (2012). Human pigmentation genes under environmental selection. *Genome Biol, 13*(9), 248. doi:10.1186/gb-2012-13-9-248

Sturtevant, A. (1913). The Himalayan rabbit case, with some considerations on multiple allelomorphs. *The American Naturalist, 47*(556), 234-239.

Suzuki, H. (2013). Evolutionary and phylogeographic views on *Mc1r* and *Asip* variation in mammals. *Genes & Genetic Systems, 88*(3), 155-164.

Talevich, E., Shain, A. H., Botton, T., & Bastian, B. C. (2016). CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol, 12*(4), e1004873. doi:10.1371/journal.pcbi.1004873

Taylor, K., Hungerford, D., Snyder, R., & Ulmer Jr, F. (1968). Uniformity of karyotypes in the Camelidae. *Cytogenet Genome Res, 7*(1), 8-15.

Taylor, K. M., Hungerford, D. A., Snyder, R. L., & Ulmer, F. A., Jr. (1968). Uniformity of kryotypes in the Camelidae. *Cytogenetics, 7*(1), 8-15.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res, 22*(22), 4673-4680.

Tjio, J. H., & Levan, A. (1956). The chromosome number of man. *Hereditas, 42*(1-2), 1-6.

Tjio, J. H., & Puck, T. T. (1958). Genetics of somatic mammalian cells: II. Chromosomal constitution of cells in tissue culture. *J Exp Med, 108*(2), 259.

Uerpmann, H.-P., & Uerpmann, M. (2002). The appearance of the domestic camel in south-east Arabia. *J Oman Stud, 12*, 235-260.

Uerpmann, M., & Uerpmann, H.-P. (2012). IV–Dromedaries *(Camelus dromedarius):* Archeozoology of Camels in South-Eastern Arabia. *Verlag der Österreichischen Akademie der Wissenschaften.* Vol. 451, pp. 109-122.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Res, 40*(15). doi:ARTN e115

Utsunomiya, Y. T., Carmo, A. S., Neves, H. H., Carvalheiro, R., Matos, M. C., Zavarez, L. B., . . . Garcia, J. F. (2014). Genome-wide mapping of loci explaining variance in scrotal circumference in Nellore cattle. *PLoS One, 9*(2), e88561. doi:10.1371/journal.pone.0088561

Van Roy, M., Ververken, C., Beirnaert, E., Hoefman, S., Kolkman, J., Vierboom, M., . . . Hemeryck, A. (2015). The preclinical pharmacology of the high affinity anti-IL-6R Nanobody® ALX-0061 supports its clinical development in rheumatoid arthritis. *Arthritis Research & Therapy, 17*(1), 135.

Von den Driesch, A., & Obermaier, H. (2007). The hunt for wild dromedaries during the 3rd and 2nd millennia BC on the United Arab Emirates coast. Camel bone finds

from the excavations at Al Sufouh 2 Dubai, UAE. *Skeletal Series and Their Socio-Economic Context, Documenta Archaeolbiologiae*, 133-167.

Warren, H. R., Evangelou, E., Cabrera, C. P., Gao, H., Ren, M., Mifsud, B., . . . Wain, L. V. (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet, 49*(3), 403-415. doi:10.1038/ng.3768

Webb, A. A., & Cullen, C. L. (2010). Coat color and coat color pattern-related neurologic and neuro-ophthalmic diseases. *The Canadian Veterinary Journal 51*(6), 653-657.

Webb, S. D. (1974). Pleistocene llamas of Florida, with a brief review of the Lamini. In: Pleistocene Mammals of Florida. *The University Presses of Florida*, Gainesville, Florida, USA, pp. 170-213.

Webb, S. D., & Meachen, J. (2004). On the origin of Lamine Camelidae (Mammalia: Camelidae): a re-evaluation including a new genus from the Late Miocene of the High Plains. *J Vert Paleontol., 6*(65-75).

Wernery, U., & Kaaden, O.-R. (2004). Foot-and-mouth disease in camelids: a review. *The Veterinary Journal, 168*(2), 134-142.

Wheeler, J. C. (1995). Evolution and present situation of the South American camelidae. *Biological Journal of the Linnean Society, 54*(3), 271-295. doi:https://doi.org/10.1016/0024-4066(95)90021-7

Wheeler, J. C. (2012). South American camelids: past, present and future. *Journal of Camelid Science, 5*(1), 1-24.

Wu, H., Guang, X., Al-Fageeh, M. B., Cao, J., Pan, S., Zhou, H., . . . Wang, J. (2014). Camelid genomes reveal evolution and adaptation to desert environments. *Nat Commun, 5*, 5188. doi:10.1038/ncomms6188

Xu, L., Yang, L., Zhu, B., Zhang, W., Wang, Z., Chen, Y., . . . Li, J. (2019). Genome-wide scan reveals genetic divergence and diverse adaptive selection in Chinese local cattle. *BMC Genomics, 20*(1), 494. doi:10.1186/s12864-019-5822-y

Xu, X., Dong, G. X., Hu, X. S., Miao, L., Zhang, X. L., Zhang, D. L., . . . Luo, S. J. (2013). The genetic basis of white tigers. *Curr Biol, 23*(11), 1031-1035. doi:10.1016/j.cub.2013.04.054

Yan, J., Li, G., Hu, Y., Ou, W., & Wan, Y. (2014). Construction of a synthetic phage-displayed Nanobody library with CDR3 regions randomized by trinucleotide cassettes for diagnostic applications. *Journal of Translational Medicine, 12*(1), 343.

Yang, B., Zhang, W., Zhang, Z., Fan, Y., Xie, X., Ai, H., . . . Ren, J. (2013). Genome-wide association analyses for fatty acid composition in porcine muscle and abdominal fat tissues. *PLoS One, 8*(6), e65554. doi:10.1371/journal.pone.0065554

Yang, Y., Wang, Y., Zhao, Y., Zhang, X., Li, R., Chen, L., . . . Wang, K. (2017). Draft genome of the Marco Polo Sheep (*Ovis ammon polii*). *GigaScience, 6*(12), 1-7. doi:10.1093/gigascience/gix106

Young, J. A. (1982). Camels on the western range USA, history. *Rangelands Archives, 4*(6), 248-251.

Yuan, Y., Peng, D., Gu, X., Gong, Y., Sheng, Z., & Hu, X. (2018). Polygenic Basis and Variable Genetic Architectures Contribute to the Complex Nature of Body Weight -A Genome-Wide Study in Four Chinese Indigenous Chicken Breeds. *Front Genet, 9*, 229. doi:10.3389/fgene.2018.00229

Zhang, B., Chang, L., Lan, X., Asif, N., Guan, F., Fu, D., . . . Li, S. (2018). Genome-wide definition of selective sweeps reveals molecular evidence of trait-driven domestication among elite goat (*Capra species*) breeds for the production of dairy, cashmere, and meat. *GigaScience, 7*(12). doi:10.1093/gigascience/giy105

Zhang, C., Ni, P., Ahmad, H. I., Gemingguli, M., Baizilaitibei, A., Gulibaheti, D., . . . Zhao, S. (2018). Detecting the Population Structure and Scanning for Signatures of Selection in Horses (*Equus caballus*) From Whole-Genome Sequencing Data. *Evol Bioinform Online, 14*, doi:10.1177/1176934318775106

Zhou, X., Xu, S., Yang, Y., Zhou, K., & Yang, G. (2011). Phylogenomic analyses and improved resolution of Cetartiodactyla. *Mol Phylogenet Evol, 61*(2), 255-264. doi:10.1016/j.ympev.2011.02.009

Zurano, J. P., Magalhaes, F. M., Asato, A. E., Silva, G., Bidau, C. J., Mesquita, D. O., & Costa, G. C. (2019). Cetartiodactyla: Updating a time-calibrated molecular phylogeny. *Mol Phylogenet Evol, 133*, 256-262. doi:10.1016/j.ympev.2018.12.015

APPENDIX A

SOFTWARE PACKAGES USED IN CHAPTER IV

Burrows-Wheeler Aligner (BWA) is a software package used for mapping sequence reads generated from an organism to its correspondent reference genome.

BWA_0.7.12-r1039

Indexing the reference genome:

```
$ bwa index -a reference.fa
```

Aligning generated reads to the dromedary reference genome:

```
$ #!/bin/sh
$ dir="/data/falshanbari/Camel_Coat_Color/Sample_Black_Camel/"
$ gen="/data/falshanbari/Drom_ref_gen/Ca_drom.fa"
$ bwa mem -aM -t 8 -R
"@RG\tID:black_pool\tPL:ILLUMINA\tLB:black_pool\tDS:black_pool\tPU:black_pool
\tSM:black_pool" ${gen} ${dir}Black_Camel_NoIndex_L001_R1_001.fastq.gz
${dir}Black_Camel_NoIndex_L001_R2_001.fastq.gz | samtools view -b -S -h -o
${dir}black_pool.bam –
```

Samtools is a computational package consisted of tools that are used for large data.

SAmtools.0.1.19

Sorting and indexing bam files:

```
$ #!/bin/sh
$ dir="/data/falshanbari/Camel_Coat_Color/Pool_seq_ana/"
```

$ samtools sort ${dir}black_pool.bam ${dir}black_pool.sort; samtools index ${dir}black_pool.sort.bam

Identifying the coverage:

$ #/bin/sh

$ dir="/data/falshanbari/Camel_Coat_Color/Pool_bam/"

$ outdir="/data/falshanbari/Camel_Coat_Color/Pool_bam/"

$ samtools depth ${dir}black_pool.sort.bam | awk '{sum+=$3} END {print sum/NR}' > ${outdir}black.coverage

Creating mpileup file:

S amtools mpileup /data1/CamDro/Coat_Color/Bam_files/black_pool.sort.bam > /data1/CamDro/Coat_Color/Popoolation/black.mpileup

Popoolation2 is a software has tools to compare between two populations to identify significant differences between populations. The first step is to calculate allele frequencies within population by generating synchronized files.

Popooation2_1201

$ perl /data/falshanbari/popoolations/mpileup2sync.pl --fastq-type sanger --min-qual 20 --input /data/falshanbari/Camel_Coat_Color/popoolation/black_red_white.mpileup --output /data/falshanbari/Camel_Coat_Color/popoolation/black-red-white.sync

Calculate divergent using sliding window and generate FST file:

$ perl /data/falshanbari/popoolations/popoolation2_1201/fst-sliding.pl --input /data/falshanbari/Camel_Coat_Color/popoolation/black_red_white.sync --output /data/falshanbari/Camel_Coat_Color/popoolation/b-r-w-25000-1000.fst --min-count 2 --

min-coverage 4 --max-coverage 250 --min-covered-fraction 0.6 --window-size 25000 --step-size 1000 --pool-size 34

Calculate allele frequency differences

$ perl /data1/_software/popoolation2_1201/snp-frequency-diff.pl --input black_red_white.sync --output-prefix black_red_white.snp-freq-diff --min-count 6 --min-coverage 6 --max-coverage 200

CRISP is a software used to generate variant calling format file (VCF).

$ /data1/_software/crisp/CRISP --bams /data1/ data1/CamDro /Camel_Color_bam_files.txt --ref /data1/CamDro/Drom_ref_gen/Ca_drom.fa --VCF /data1/ data1/CamDro /Coat_color/CDRISP/black_red_white_CCC_CRISP.vcf

Convert VCF file:

$ /data1/_software/crisp/scripts/convert_pooled_vcf.py /data1/CamDro/Coat_Color/crisp/region_of_interest_wrong_pool_sizes.vcf /data1/CamDro/Coat_Color/Bam_files/Camel_Color_bam_files.txt > /data1/CamDro/Coat_Color/crisp/region_of_interest_wrong_pool_sizes_converted.vcf

snpEff is a software that is used to predict the effect of variants that were identified in the VCF file. The software requires a gtf file that contains information of coding gene positions in a reference genome.

$ java -Xmx4g -jar snpEff.jar eff CamDro1 /data1/CamDro/Coat_Color/black_red_white_all.crisp.vcf > /data1/CamDro/Coat_Color/crisp/black_red_white_converted.ann.vcf

Running R scripts to generate Manhattan plots and plots with genes in specific windows:

$ source("plot_freq_selected_region.R")

plot_freq_selected_region("data/NW_011591187.1_allele_freq.txt","NW_011591187.1",1590000,1990000,"out/Freq_above_0-5.png",snp_threshold=0,allele_threshold=0.5)

plot_freq_selected_region("data/NW_011591187.1_allele_freq.txt","NW_011591187.1",1590000,1990000,"out/Freq_above_0-8.png",snp_threshold=0,allele_threshold=0.8)

Popoolation is a software used to estimate Tajima's Pi and other population parameters such as Watterson's Theta and Tajima's D using sliding window approach

$ samtools mpileup /data1/CamDro/Coat_Color/Bam_files/black_pool.sort.bam > /data1/CamDro/Coat_Color/Popoolation/black.mpileup

$ perl /data1/_software/popoolation_1.2.2/Variance-sliding.pl --measure pi --input black.mpileup --min-count 2 --min-qual 20 --min-coverage 4 --max-coverage 70 --pool-size 34 --window-size 1000 --step-size 20000 --output black_pool.pi --fastq-type sanger

CNVkit is a software used to estimate copy number variation (CNVs) in a large data such as whole genome sequencing.

$ # From baits and tumor/normal BAMs

cnvkit.py batch --method wgs

*/data1/CamDro/Coat_Color/Bam_files/white_pool.sort.bam \

--normal */data1/CamDro/Coat_Color/Bam_files/black_pool.sort.bam

APPENDIX B

R SCRIPTS USED IN CHAPTER IV

R script that was used to generate Manhattan plots from the zFST file:

```
###### Parameters ######


# Set working directory
setwd("/media/7728e2f2-0885-4fac-957c-
e1a995c7514e/data/Uppmax/Camel/vcf_180110")


# Input file name
#fname="BLACKvs-Red-1k-25k-zFST.txt"
#fname="Black-vs-White-zFST_1k-25k.txt"
fname="Red-vs-White-1k-25k-zFST.txt"


# Should we add labels showing position of top peaks (time consuming)?
mark_positions=TRUE


# Set y coordinate limits (lowest and highest value shown)
min_coordinate=0
max_coordinate=20


########################
```

```
library(zoo)

# Create output folder
folder = sub(".zFST","",sub(".txt","",fname))
system(paste("mkdir -p ",folder,sep=""))

# Read data file
dat = read.csv(fname,sep="\t",header=FALSE)

# Which column has the zFST value?
column=5

# Exclude regions with less than half or more than double the amount of variants
medianReads=median(dat[,3])
minReads=medianReads*0.5
maxReads=medianReads*2
dat = dat[dat[,3]>minReads,]
dat = dat[dat[,3]<maxReads,]

# Select only chr 1:28
#dat = dat[dat[,1] %in% 1:28,]
```

```
# 25k windows with 1k overlap means that the number of independent tests are
approximately the number of rows/25
thres = -qnorm(0.025/dim(dat)[1]/25)

# Replace negative values by 0
#dat[ dat[,5]<0 , 5 ]=0

# Chromosome names
chrs = as.character(unique(dat[,1]))

# Retrieve ZFST values
Z = dat[,column]

colors = c("#000000","#CC00FF","#00FFFF","#FF0000","#0000FF","#00FF00")
start = 1

# Make the combined plot
#pdf(paste(folder,"/all.pdf",sep=""),width=10,height=2,pointsize=6)
png(paste(folder,"/all.png",sep=""),width=10,height=2,units="in",res=300,pointsize=6)
par(
  mar      = c(0, 4, 0.5, 0)+0.5,
  xaxs     = "i",
```

```
  yaxs     = "i",
  cex.axis = 2,
  cex.lab  = 2
)
for (i in 1:length(chrs)) {
  Zchr = Z[dat[,1]==chrs[i]]
  trans="FF"
  size = 19
  if (length(Zchr)==1) {
    #trans="33"
    size=4
  }
  positions = dat[,2][dat[,1]==chrs[i]]
  x = start:(start+length(Zchr)-1)
  Zchr[Zchr<0]=0

  if (mark_positions) {
    if (length(Zchr)>1001) {
      z = zoo(Zchr)
      zmax = rollapply(c(rep(0,500),z,rep(0,500)),1001,function(x) which.max(x)==501)
    } else {
      zmax = which(Zchr==max(Zchr))
```

```
  }
 }

 if (i==1) {

plot(x,Zchr,type="p",pch=size,cex=.6,ylim=c(min_coordinate,max_coordinate),xlim=c(
0,length(Z)),col=paste(colors[i],trans,sep=""),bty="n",ylab="ZFST",xlab="",xaxt='n',xp
d=NA)
 } else {
  points(x,Zchr,typ="p",pch=size,cex=.6,col=paste(colors[(i-
1)%%6+1],trans,sep=""),xpd=NA)
 }


 if (mark_positions) {
  coord = zmax
  if (is.logical(coord)) {
   coord = which(coord)
  }
  coord = coord[Zchr[coord]>10]

  if (length(coord)) {
```

```
      y = Zchr[coord]
      y[y>max_coordinate-1]=max_coordinate-1

text(x[coord],y,paste(chrs[i],"\n",round(positions[coord]/1e6,digits=3),sep=""),pos=3,ce
x=0.8,offset=0.4)
    }
  }
  start = start+length(Zchr)
}

segments(0,thres,length(Z),thres,col="grey")
dev.off()

# Make individual plots for each chromosome
lines = dim(dat)[[1]]

from = 1
to = 0

for (i in 1:length(chrs)) {
  datChr = dat[dat[,1]==chrs[i],]
  datChr = datChr[datChr[,3]>minReads,]
```

```
  datChr = datChr[datChr[,3]<maxReads,]


  x = datChr[,2]
  y = datChr[,column]


  # Only plot contigs with signal greater than 8
  if (length(x) > 0 && max(y)>=8) {
    pdf(paste(folder,"/",chrs[i],".pdf",sep=""),width=5,height=2,pointsize=6)

#png(paste(folder,"/",chrs[i],".png",sep=""),width=5,height=2,units="in",res=300,pointsi
ze=6)
    par(
      mar      = c(5, 5, 2, 2),
      xaxs     = "i",
      yaxs     = "i",
      cex.axis = 2,
      cex.lab  = 2
    )


    from = min(from,y)
    to = max(to,y)
```

```
    y[y<0]=0


    plot(x/1000000,y,type="p",pch=19,cex=1,col=colors[(i-
1)%%6+1],ylab="ZFST",xlab="Position
(Mb)",main=paste(chrs[i],sep=""),ylim=c(min_coordinate,max_coordinate),xlim=c(min(
x),max(x))/1000000,xpd=NA)
    axis(1,at=10*(1:20))
    segments(0,thres,x[length(x)]/1000000,thres,col="grey")
    dev.off()
  }
}
```

R script that was used to print specific regions with gene IDs and annotation of exons/introns as shown in figure 12

```
###### Parameters ######

# Set working directory
setwd("/media/7728e2f2-0885-4fac-957c-
e1a995c7514e/data/Uppmax/Camel/vcf_180110")

# Input file name
#fname="BLACKvs-Red-1k-25k-zFST.txt"
#fname="Black-vs-White-zFST_1k-25k.txt"
```

```
fname="Red-vs-White-1k-25k-zFST.txt"

# Should we add labels showing position of top peaks (time consuming)?
mark_positions=TRUE

# Set y coordinate limits (lowest and highest value shown)
min_coordinate=0
max_coordinate=20

########################
library(zoo)

# Create output folder
folder = sub(".zFST","",sub(".txt","",fname))
system(paste("mkdir -p ",folder,sep=""))

# Read data file
dat = read.csv(fname,sep="\t",header=FALSE)

# Which column has the zFST value?
column=5
```

```
# Exclude regions with less than half or more than double the amount of variants
medianReads=median(dat[,3])
minReads=medianReads*0.5
maxReads=medianReads*2
dat = dat[dat[,3]>minReads,]
dat = dat[dat[,3]<maxReads,]

# Select only chr 1:28
#dat = dat[dat[,1] %in% 1:28,]

# 25k windows with 1k overlap means that the number of independent tests are
approximately the number of rows/25
thres = -qnorm(0.025/dim(dat)[1]/25)

# Replace negative values by 0
#dat[ dat[,5]<0 , 5 ]=0

# Chromosome names
chrs = as.character(unique(dat[,1]))

# Retrieve ZFST values
Z = dat[,column]
```

```
colors = c("#000000","#CC00FF","#00FFFF","#FF0000","#0000FF","#00FF00")
start = 1


# Make the combined plot
#pdf(paste(folder,"/all.pdf",sep=""),width=10,height=2,pointsize=6)
png(paste(folder,"/all.png",sep=""),width=10,height=2,units="in",res=300,pointsize=6)
par(
  mar      = c(0, 4, 0.5, 0)+0.5,
  xaxs     = "i",
  yaxs     = "i",
  cex.axis = 2,
  cex.lab  = 2
)
for (i in 1:length(chrs)) {
  Zchr = Z[dat[,1]==chrs[i]]
  trans="FF"
  size = 19
  if (length(Zchr)==1) {
    #trans="33"
    size=4
  }
```

```
 positions = dat[,2][dat[,1]==chrs[i]]
 x = start:(start+length(Zchr)-1)
 Zchr[Zchr<0]=0

 if (mark_positions) {
  if (length(Zchr)>1001) {
    z = zoo(Zchr)
    zmax = rollapply(c(rep(0,500),z,rep(0,500)),1001,function(x) which.max(x)==501)
  } else {
    zmax = which(Zchr==max(Zchr))
   }
 }

 if (i==1) {

plot(x,Zchr,type="p",pch=size,cex=.6,ylim=c(min_coordinate,max_coordinate),xlim=c(
0,length(Z)),col=paste(colors[i],trans,sep=""),bty="n",ylab="ZFST",xlab="",xaxt='n',xp
d=NA)
 } else {
   points(x,Zchr,typ="p",pch=size,cex=.6,col=paste(colors[(i-
1)%%6+1],trans,sep=""),xpd=NA)
 }
```

```
  if (mark_positions) {
    coord = zmax
    if (is.logical(coord)) {
      coord = which(coord)
    }
    coord = coord[Zchr[coord]>10]

    if (length(coord)) {
      y = Zchr[coord]
      y[y>max_coordinate-1]=max_coordinate-1

text(x[coord],y,paste(chrs[i],"\n",round(positions[coord]/1e6,digits=3),sep=""),pos=3,ce
x=0.8,offset=0.4)
    }
  }
  start = start+length(Zchr)
}

segments(0,thres,length(Z),thres,col="grey")
dev.off()
```

```
# Make individual plots for each chromosome
lines = dim(dat)[[1]]

from = 1
to = 0

for (i in 1:length(chrs)) {
  datChr = dat[dat[,1]==chrs[i],]
  datChr = datChr[datChr[,3]>minReads,]
  datChr = datChr[datChr[,3]<maxReads,]

  x = datChr[,2]
  y = datChr[,column]

  # Only plot contigs with signal greater than 8
  if (length(x) > 0 && max(y)>=8) {
    pdf(paste(folder,"/",chrs[i],".pdf",sep=""),width=5,height=2,pointsize=6)

#png(paste(folder,"/",chrs[i],".png",sep=""),width=5,height=2,units="in",res=300,pointsi
ze=6)
    par(
```

```
      mar      = c(5, 5, 2, 2),
      xaxs     = "i",
      yaxs     = "i",
      cex.axis = 2,
      cex.lab  = 2
    )

    from = min(from,y)
    to = max(to,y)

    y[y<0]=0

    plot(x/1000000,y,type="p",pch=19,cex=1,col=colors[(i-
1)%%6+1],ylab="ZFST",xlab="Position
(Mb)",main=paste(chrs[i],sep=""),ylim=c(min_coordinate,max_coordinate),xlim=c(min(
x),max(x))/1000000,xpd=NA)
    axis(1,at=10*(1:20))
    segments(0,thres,x[length(x)]/1000000,thres,col="grey")
    dev.off()
  }
}
```