COMPARATIVE MAPPING OF THE ALPACA GENOME

A Dissertation

by

FELIPE FAGUNDES DE AVILA

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
Co-Chair of Committee,	Penny K. Riggs
Committee Members,	Bhanu P. Chowdhary
	James E. Womack
	William J. Murphy
Head of Department,	Evelyn Tiffany-Castiglioni

August 2014

Major Subject: Biomedical Sciences

ABSTRACT

The development of gene maps constitutes a key feature for understanding genome architecture and comparative evolution. The genomes of some livestock species such as cattle, horses and pigs, have received considerable attention over the years due to their economic importance. In contrast, though camelids are gaining worldwide popularity as production and companion animals, cytogenetics and genome mapping in these species lag far behind those of other mammals.

One of the reasons for the scarce body of knowledge regarding the camelid genome is their particularly difficult karyotype for analysis. All six extant camelid species have a diploid number of 74 chromosomes; the gross morphological similarities shared by many of the autosomes, combined with the relatively small size of some chromosome pairs, present serious challenges for identifying individual chromosomes using conventional cytogenetic techniques. The Alpaca Genome Project includes whole genome sequencing, radiation hybrid (RH) mapping and human-camel comparative chromosome painting (Zoo-FISH). However, there is no common platform that aligns various maps and precisely assigns them to individual chromosomes.

Therefore, the goal of this research project was to construct a cytogenetic map for the alpaca genome by fluorescence *in situ* hybridization (FISH) of large insert clones from the alpaca CHORI-246 genomic BAC library. The BACs were

selected based on the available Zoo-FISH, RH and sequence map data to target evolutionarily conserved genes and to get uniform distribution of markers throughout the alpaca genome. Candidate genes for traits of interest such as various congenital and reproduction-related disorders, as well as for phenotypic traits such as fiber color and texture, were also selected for mapping. A total of 230 markers were mapped to the 36 alpaca autosomes and the sex chromosomes; moreover, comparative mapping showed exceptional conservation of both gene synteny and order between alpaca and dromedary camel chromosomes.

The cytogenetic map of the alpaca genome is a platform that effectively integrates the whole genome sequence and the radiation hybrid map with cytogenetic data, thus facilitating the discovery of genes of interest and providing tools for studying chromosome evolution and for clinical cytogenetics by means of a collection of chromosome-specific markers for camelids.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Terje Raudsepp, for her guidance, patience, words of wisdom, and for allowing me to fulfill the dream of obtaining my doctorate degree. I would also like to thank my Co-Chair, Dr. Penny K. Riggs, and my committee members, Dr. Bhanu P. Chowdhary, Dr. James E. Womack, and Dr. William J. Murphy, for their support and invaluable contributions throughout the course of this research.

I would also like to thank my family for all they have done for me during this amazing journey. Pai, Mãe, Leo, Lu, Vó, I will be forever grateful for your encouragement, for your attention, for your love, and for always being there for me, even from a distance. This accomplishment is not only mine; I could not have done it without you. Muito obrigado por tudo!

Many thanks to Malorie for everything you have done. Thanks for helping me in the laboratory, for helping me write this dissertation, for sharing the happy and the difficult days with me, for your love, but most importantly, for our girl. Julia, I hope I will make you proud.

Thanks to my former and current laboratory mates, my friends and colleagues at Texas A&M University and in Brazil, to the department faculty and staff, for making this a great experience. I will treasure all these good memories forever; good luck on your life journeys.

I would also like to extend my gratitude to all our collaborators, to alpaca breeders and funding agencies, for allowing this research to happen with ideas, data, samples, and financial support.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

THE FAMILY CAMELIDAE

Phylogenetics and Evolution

Camelidae is a mammalian family belonging to the order Cetartiodactyla. This order, which consists of the most diverse large mammals in the world (Zhou *et. al.*, 2011; Hassanin *et. al.*, 2012; Rubes *et. al.*, 2012), contains the majority of domesticated mammalian species (cattle, sheep, pigs, goats, camelids, and cervids). It comprises four major lineages (or suborders): Tylopoda (camelids), Suiformes (pigs, peccaries and hippopotamuses), Ruminantia (bovids, deer, giraffes and tragulids), and Cetacea (whales, dolphins, and porpoises) (Murphy *et. al.*, 2005; Gatesy *et. al.*, 2002; Zhou *et. al.*, 2011; Hassanin *et. al.*, 2012). Even though the root of Cetartiodactyla has proven to be elusive (Zhou *et. al.*, 2011; Hassanin *et. al.*, 2012; Rubes *et. al.*, 2012), most phylogenetic and phylogenomics studies place Camelidae as the most basal family within the order (Murphy *et. al.*, 2005; Gatesy *et. al.*, 2002; Agnarsson & May-Collado, 2008; Ayoub *et. al.*, 2009; Zhou *et. al.*, 2011) (Figure 1).

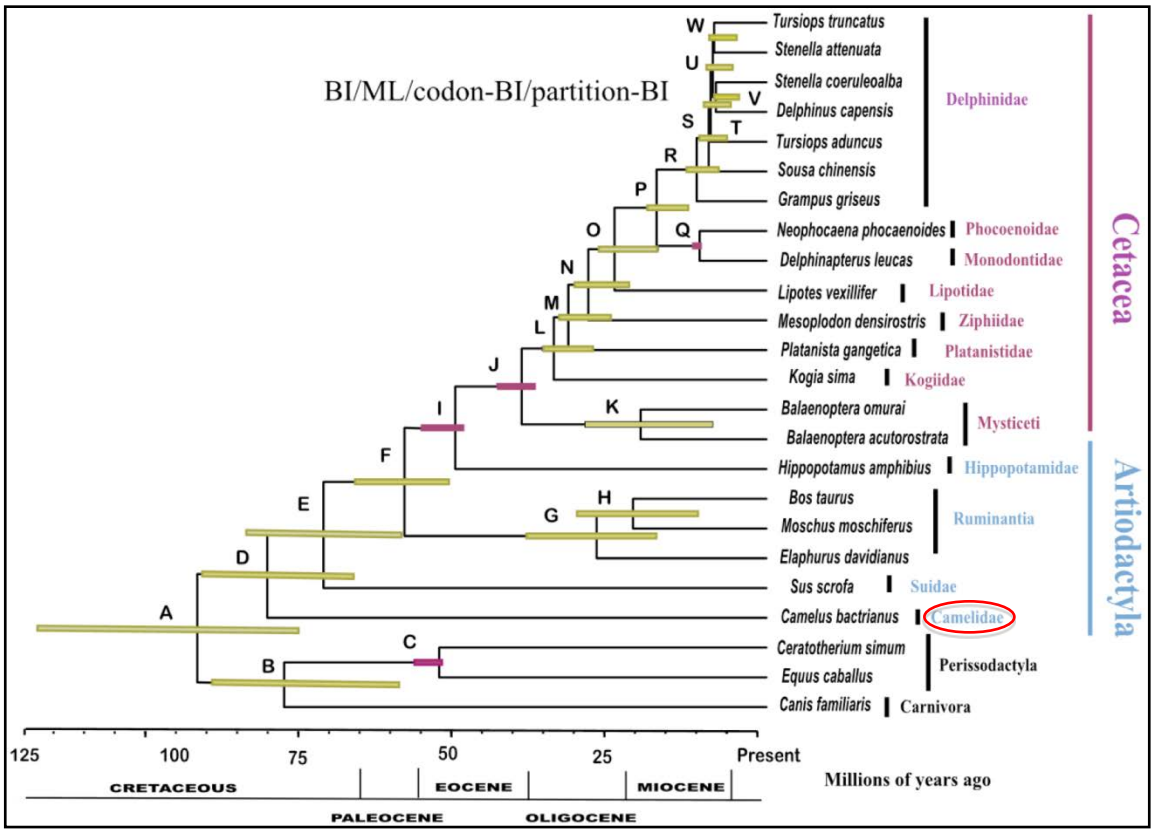

Figure 1. Phylogenetic relationships of 21 Cetartiodactyla and 3 outgroup species, inferred from 110 nucleus-encoded proteins using Bayesian inference (BI) and Maximum Likelihood (ML) methods. Using this approach, Zhou *et. al.* (2011) determined that Camelidae (red circle) is the most basal family of the order Cetartiodactyla (PP=1.0, BP=100%). Reprinted from Zhou, X.; Xu, S.; Yang, Y.; Zhou, K.; Yang, G. 2011. Phlyogenomic analyses and improved resolution of Cetartiodactyla. Mol Phylogenet Evol. 61 (2): 255-264. Copyright (2011), with permission from Elsevier.

Archaeozoological and mitochondrial DNA studies show that camelids evolved in North America during the Eocene, between 45 and 40 million years ago (MYA) (Webb, 1974; Wheeler *et. al.*, 1984; Stanley *et. al.*, 1994; Skidmore *et. al.*, 1999), whereas phylogenomics suggests that the divergence between Camelidae and the remaining Cetartiodactyla species occurred 96-66 MYA (Zhou *et. al.*, 2011). The last common ancestor of all extant camelid species,

Aepycamelus, is dated to 17.5 MYA (Honey *et. al.,* 1998); then, its descendants split into lineages that gave rise to the two modern tribes: around 15.5MYA, *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) (Honey *et. al.,* 1998; Webb & Meachen, 2004). Approximately 3.3 million years ago, during the late Tertiary of the Pliocene epoch, Camelini species migrated to Eurasia through the Bering Strait due to lowered sea levels, whereas Lamini species migrated to South America through a sub-tropical, savannah-like corridor, becoming extinct in North America around 10,000 years ago together with most of its megafauna (Webb, 1974; Franklin, 1982; Stanley *et. al.*, 1994; Ji *et. al.*, 2009).

The tribe Camelini - or the Old World camelids - is comprised of two extant species: the two-humped Bactrian camel (*Camelus bactrianus*, CBA) and the one-humped dromedary camel (*Camelus dromedarius*, CDR). Dromedary camels can be found from North Africa to southwest Asia, whereas Bactrian camels range in desert regions of Mongolia and China (Stanley *et. al.*, 1994; Ji *et. al.*, 2009) (Figure 2).

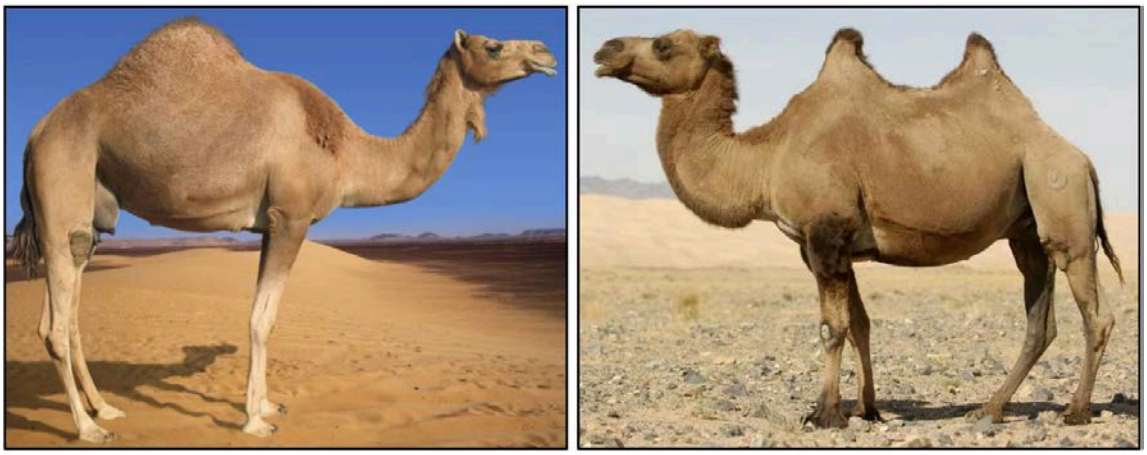

Figure 2. Old World camelids. Left: dromedary camel (Camelus dromedarius). Right: Bactrian camel (Camelus bactrianus).

The tribe Lamini - New World or South American camelids - is composed of four extant species: the wild vicugna (*Vicugna vicugna*, VVI) and guanaco (*Lama guanicoe*, LGU) and their respective putative domesticated counterparts, the alpaca (*Lama pacos*, LPA) and the llama (*Lama glama*, LGL). These animals can be found from the high altiplano of the Andes to the bush areas of South America (Stanley *et. al.*, 1994; Kadwell *et. al.*, 2001) (Figure 3).

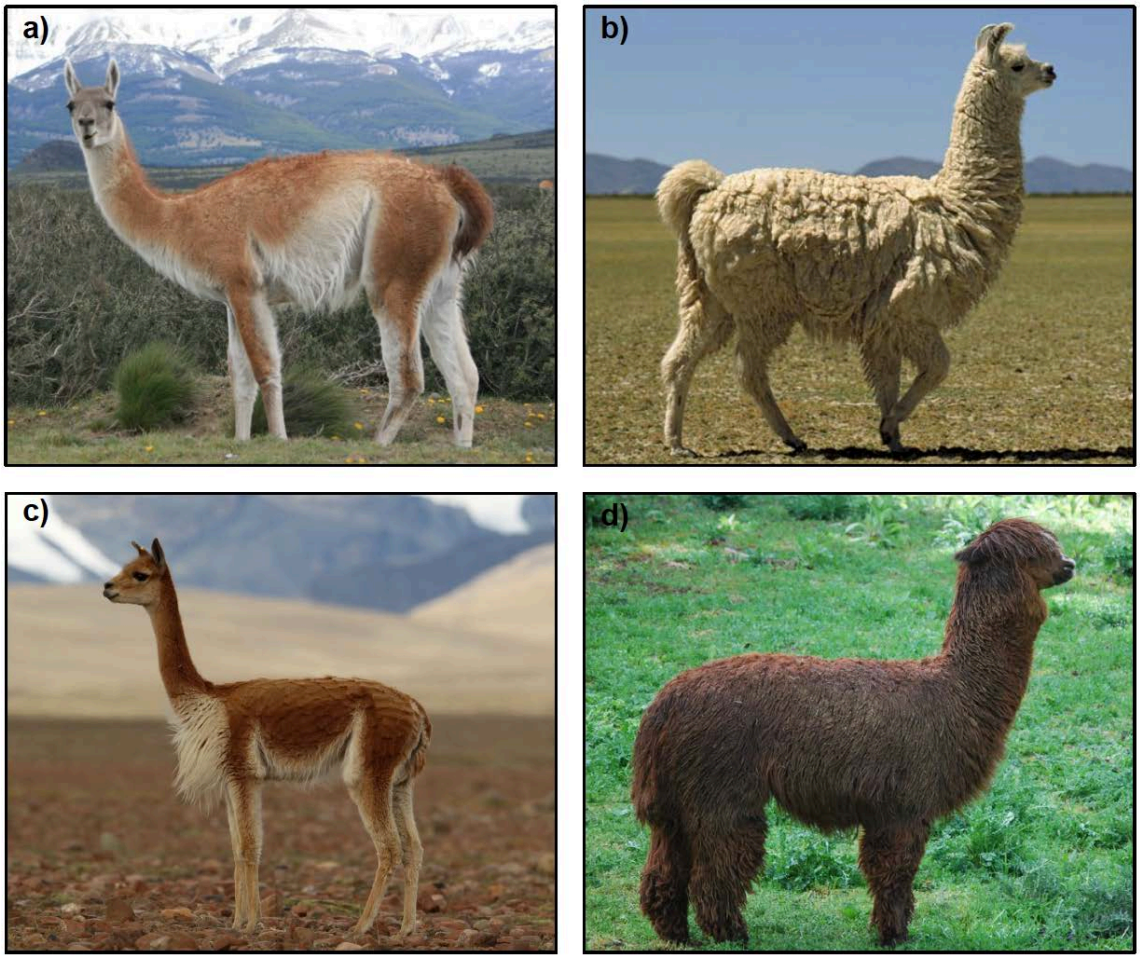

Figure 3. New World camelids. a) Guanaco (Lama guanicoe); b) Llama (Lama glama); c) Vicugna (Vicugna vicugna); d) Alpaca (Lama pacos).

Although the phylogenetic relationships between the two camelid tribes as well as within Camelini have been elucidated (Ji *et. al.*, 2009), the ancestry of the alpaca and the llama is still inconclusive and has been a subject of historical debate. Based on fossil records, it has been proposed that the llama descends from the guanaco, whereas, the vicugna gave rise to the alpaca (Wheeler, 1995; Wheeler, 2006). Alternatively, early phylogenetic studies propose that both the

llama and the alpaca descend from the guanaco, and that the vicugna has never been domesticated (Kleinschmidt *et. al*., 1986; Jurgens *et. al.,* 1988). More recently, phylogenetic studies of mitochondrial DNA (both the cytochrome b gene and the D-loop region) have established that guanacos and vicugnas form two separate monophyletic clades, whereas their relationships to llamas and alpacas could not be resolved due to extensive hybridization between the two domesticated species (Stanley *et. al.*, 1994; Kadwell *et. al.*, 2001; Di Rocco *et. al.*, 2010; Barreta *et. al.*, 2012). For this reason, taxonomists and camelid geneticists alike still debate the origin of the alpaca scientific name based on phylogenetic relationships: until 2001, the most widely used species name was *Vicugna pacos* (inferring close relationship to the vicugna, both in the genus *Vicugna*) (Wheeler, 1995); then, a phylogenetic study published in 2001 suggested the alpaca species name be changed to *Lama pacos* (Kadwell *et. al.*, 2001).

Biology and Adaptations

Old and New World camelids share remarkable anatomical and physiological similarities, but have also developed distinctive features as a result of their adaptation to extremely diverse environments. Camelidae is the only mammalian family to produce a unique class of antibodies, which lack light chains, called the Heavy-chain only antibodies (or HCAbs), along with conventional antibodies found in other mammals (Conrath *et. al.*, 2003; De

Genst *et. al.*, 2006; Flajnik *et. al.*, 2011). This class of antibodies, discovered in the early 1990s (Hamers-Casterman *et. al.*, 1993) functions alongside conventional antibodies in these animals, constituting between 45% and 75% of serum IgG across different camelid species (Daley *et. al.*, 2010), and have a remarkably high affinity to a wide array of antigens (Flajnik *et. al.*, 2011). Notably, the heavy chain of HCAbs is devoid of the first constant domain (CH1) found in other mammalian species, and contains a specialized variable domain (VHH) (De Genst *et. al.*, 2006). These antibodies represent a unique evolutionary event in mammals, and present evolutionary convergence with antibodies produced by some shark species (Conrath *et. al.*, 2003; Criscitiello, 2014).

Heavy-chain only antibodies, due to their distinctive properties and smaller size, have many potential applications in biomedical research, offering several advantages over conventional mammalian antibodies. Reports have shown that camelid antibodies have been successfully used to target and neutralize the HIV-1 virus (McCoy *et. al.*, 2012; Strokappe *et. al.*, 2012), the West Nile virus (Daley *et. al.*, 2010) and for the production of markers for early breast cancer detection (Even-Desrumeaux *et. al.*, 2012). In humans and other mammalian species, each immunoglobulin segment (*i.e.*, the kappa and lambda light chains and the heavy chain) is encoded by clusters of genes, located in different genomic loci. In camelids, the location of these genomic loci has not been reported as of yet, except for the IgH locus, that was recently mapped by

fluorescence *in situ* hybridization (FISH) close to the telomere on the long arm of alpaca chromosome 4 (LPA4) (Achour *et. al.*, 2008). In that report, the authors followed the alpaca chromosome nomenclature proposed by Di Berardino and colleagues (2006).

Further, camelids are the only mammals with elliptical red blood cells – these flat ellipsoid erythrocytes are smaller and circulate in larger numbers than those found in other mammalian species, and their unique shape allows them to be oriented with the long axis, parallel to the blood flow (Fowler, 2010). This morphological feature, derived from their existence both at high altitudes (where oxygen is scarce) and arid environments, provides resistance to osmotic lysis and allows them to travel through small capillaries even when blood viscosity increases due to dehydration. In addition, these cells can expand up to 240% upon re-hydration without rupturing, as opposed to the 150% of erythrocytes found in other mammals (Fowler, 2010; Tornquist & Cebra, 2014).

Other similarities among camelids include aspects of their reproductive physiology: females are induced ovulators, *i.e.,* have no set estrous cycle and their ovulation only occurs post-coitus, although spontaneous ovulation has been observed in a very small percentage of females (El Wishy, 1987; Skidmore *et. al.*, 1999; Pearson *et. al.*, 2014). Their uterus is bicornuate, with the left horn larger than the right and always where implantation occurs, even though ovulation can occur from both ovaries (El Wishy, 1988). Also, their epitheleochorial placenta is diffuse, microcotyledonary and non-invasive, similar

to that of the mare and the sow (Skidmore *et. al.*, 1996; Olivera *et. al.*, 2003; Aba, 2014). Different species within Camelini or Lamini can cross and produce fertile offspring (Gray, 1972), and there has been a report of a successful hybridization between a dromedary camel and a guanaco via artificial insemination (Skidmore *et. al.*, 1999). Camelids also share similarities in their digestive physiology: they are considered pseudoruminants and intermediate feeders/grazers, with a three-chambered digestive tract (their stomach is divided into C1, C2 and C3), their upper lip is split into two separate mobile parts in order to help with feeding, and they have an isolated incisor in the upper jaw (Fowler & Miller, 2008).

On the other hand, there are several differences within the family: New World camelids have a significantly smaller body size than their Old World relatives, and are covered with a dense coat of fiber in order to withstand the cold temperatures of the Andean region of South America (Skidmore *et. al.*, 1999). Old World camels, which inhabit desert regions, can fluctuate their body temperature from 34 to 41.7 degrees Celsius throughout the day, and are capable of drinking 100 liters of water in as little as ten *minute*s (Al-Swailem *et. al.*, 2010).

Camelid Domestication

Although different species of camelids were domesticated in different parts of the world at various times, humans sought out these animals for similar

purposes: for food (milk and meat) and fertilizers, for their leather and fiber, as beasts of burden (Skidmore *et. al.*, 1999; Kadwell *et. al.*, 2001; Al-Swailem *et. al.*, 2010) and, more recently, as companion and sports animals. Domestication of the two-humped camel occurred in China and Mongolia between 5,000 and 6,000 years ago, while llamas and alpacas were domesticated in the high Andes around 7,000 years ago (Wheeler, 1995; Kadwell *et. al.*, 2001). Thus, camelids have been associated with humans for almost as long as cattle (Taberlet *et. al.*, 2011), horses (Groeneveld *et. al.*, 2010) and dogs (Galibert *et. al.*, 2011).

Since their domestication, both Old and New World camelids can be considered among the most important assets and the economical mainstays of several communities throughout the world, contributing in various economic aspects. The United States alpaca industry has been steadily growing since 1984, when the first animals were brought from Peru (which has the largest alpaca herd in the world, composed of 3.5 million animals). The latest census in 2009 reported 150,000 registered animals, and it is believed that there are upwards of 300,000 alpacas in the country (Saitone & Sexton, 2012). In the U.S., alpacas are economically important companion and fiber animals, which can be exemplified by the number of Alpaca Owners Association (AOA) Certified events: more than 38 throughout the country from April to December of 2014 (https://www.alpacaowners.com/calendar.asp, accessed on April 20th, 2014).

EARLY CYTOGENETICS STUDIES IN MAMMALS

The field of cytogenetics started in the early 1900s, when technological and procedural advancements allowed for the visualization of plant and animal chromosomes, and the characterization of the first mammalian karyotypes. This was when, for the first time, the chromosome number was assessed for humans and a few domestic species, such as cattle, horses, rabbits and pigs (Krallinger, 1931). Due to the poorly developed methodology at the time, the quality of these preparations was low, resulting in overall erroneous assessments of chromosome number and morphology. Remarkably, for over 30 years the chromosome number of humans was incorrectly noted: in 1923, a study by Theophilus Painter during his tenure at the University of Texas at Austin erroneously reported the diploid number as 2n=48 (Painter, 1923), and the correct diploid number (2n=46) was established only in 1956 at the Institute of Genetics of the University of Lund, Sweden (Tjio and Levan, 1956; reviewed by Gartler, 2006). This was largely due to the advancements in cell culture methodology and the accidental discovery of the hypotonic treatment to separate chromosomes from the cytoplasm, which significantly improved the quality of preparations, allowing for the correct estimation of chromosome number and morphology (Tjio and Levan, 1956; Trask, 2002; Gartler, 2006). Animal cytogeneticists almost immediately adopted methodological advancements in human chromosome preparation and analysis, and the 1960's

mark the beginning of conventional chromosome analysis and karyotyping in most of the domestic and many wild animal species.

One of the most outstanding achievements of that time - and in the field of cytogenetics - was the publication of a 10 volume Atlas of Mammalian Chromosomes by T. C. Hsu and K. Benirschke, from 1967 to 1977. In the course of 10 years, this atlas released (as annual loose-leaf issues) Giemsa stained karyotypes of hundreds of mammalian species (Hsu & Benirschke, 1967-1977). In 2006, S. J. O'Brien, J. C. Menninger and W.G. Nash published an improved Atlas of Mammalian Chromosomes with ideograms and banded karyotypes for 850 species (O'Brien *et. al.*, 2006). Thus, together with chromosome studies published elsewhere, karyotypes have been determined for around 1,000 mammalian species. Reports show that chromosome number varies greatly in mammals, with the lowest known diploid number (2n=6/7) in the Indian muntjac deer (*Muntiacus muntjak*) (Wurster & Benirschke, 1970) and the highest (2n=102) in the red vizcacha rat (*Tympanoctomys barrerae*), a rodent found in South America (Gallardo *et. al.*, 2006).

Significant variation in chromosome number and morphology has been observed also between closely related species. For example, in extant equids – horses, asses and zebras – diploid numbers range from 2n=32 in the Hartmann's mountain zebra (*Equus zebra hartmannae*) to 2n=66 in the Przewalski's horse (*Equus przewalskii*) (Chowdhary & Raudsepp, 2000), while extensive rearrangements including inversions, fissions, fusions, translocations

and centromere repositioning distinguish between the karyotypes of the horse and the donkey – the two equids known to produce viable hybrids (Allen & Short, 1997; Raudsepp *et. al.*, 2001; Musilova *et. al.,* 2013). The same phenomena have been observed in the family Cervidae (also a member of the order Cetartiodactyla, being closely related to camelids), in which the diploid number varies greatly among species: from 2n=6/7 for the Indian muntjac (*Muntiacus muntjak*) to 2n=70 in the gray brocket deer (*Mazama gouazoubira*), and karyotypical variation has even been observed at the subpopulation level (Rubes *et. al.*, 2012). In light of such variation, it is noteworthy that the diploid number of all six extant species of the family Camelidae (correctly reported for the first time in 1967 by T. C. Hsu and K. Benirschke) is 2n=74, and chromosome morphology among camelid species is also fairly conserved (Hsu & Benirschke, 1967; Bianchi *et. al.*, 1986; Bunch *et. al.*, 1985; Samman *et. al.*, 1992; Di Berardino *et. al.*, 2006; Balmus *et. al.*, 2007; Avila *et. al.*, 2012; Raudsepp, 2014).

In the 1970's, conventional cytogenetics techniques became increasingly popular in the study of mammalian chromosomes, especially after differential chromosomal banding techniques, such as Q-, G-, C-, and R-banding were developed and/or popularized (reviewed by Schreck & Disteche, 2001). These procedures allow researchers to identify consistent structural landmarks within each individual chromosome due to differential staining of chromatin based on its state of condensation (euchromatin -or loosely packed-, versus

heterochromatin or tightly packed chromatin) and GC content. Giemsa staining is undeniably the most widely used technique for chromosome analysis, as it stains chromatin allowing for the visualization of chromosome morphology.

Various other banding techniques were named after the reagents used to produce the banding pattern of metaphase chromosomes: G-banding employs Giemsa and trypsin to stain AT-rich chromatin (or gene poor regions) darker; the reverse of G-bands is obtained by R-banding. Alternatively, Q-banding uses quinacrine mustard and DAPI-banding uses 4',6-Diamidino-2-Phenylindole, Dihydrochloride stain to produce fluorescent signals, in which bright bands correspond to gene-poor AT-rich regions and are similar to G-bands. The methods that produce DAPI-banding are widely used in fluorescence microscopy and gene mapping. Other banding methods include C-banding (highlights centromeres and constitutive heterochromatin) and NOR-banding, which uses silver nitrate to highlight functionally active Nucleolus Organizer Regions (NORs), composed of tandem copies of ribosomal RNA genes (rDNA) (Schreck & Disteche, 2001).

In the late 1980's, extensive studies of cattle, sheep, goat, river buffalo, pig, horse, dog, cat and rabbit chromosomes were carried out in laboratories worldwide, and the karyotypes of these species were characterized in detail (reviewed by Ducos *et. al.*, 2008; Lear & Bailey, 2008; Villagomez & Pinton, 2008; Villagomez *et. al.*, 2009). Concomitantly, the first worldwide, standardized chromosome nomenclature systems for domestic animals were developed,

allowing proper cross-talk between laboratories and establishing a necessary foundation for unequivocal identification of normal and aberrant chromosomes. Clinical cytogenetics also began receiving strong support from veterinarians, animal breeders and owners due to the fact that the direct connection between chromosome aberrations and developmental and reproductive abnormalities started being thoroughly reported at that time (Villagomez & Pinton, 2008; Villagomez *et. al.*, 2009; Raudsepp *et. al.*, 2010).

THE ADVENT OF MOLECULAR CYTOGENETICS

In the 1990s, the advent of molecular hybridization-based methods - such as Fluorescence *In Situ* Hybridization (FISH) - in mammalian chromosome analysis essentially blurred the boundaries between cytogenetics and molecular biology (Trask, 2002; Speicher & Carter, 2005). Molecular cytogenetics techniques rely on the principle of Watson-Crick base-pairing complementarity, and allow for the localization of DNA sequences (henceforth called probes) in their original place (*in situ*) on mitotic or meiotic chromosomes (henceforth, the targets) at different stages of the cell cycle, depending on the application (Trask, 2002; Speicher & Carter, 2005; Raudsepp & Chowdhary, 2008; Rubes *et. al.*, 2009). In the case of FISH, both the probe and the target must be denatured into single-stranded molecules for the hybridization to occur, and the probes are labeled with haptens (biotin, digoxigenin) or with fluorochromes (spectrum

orange, spectrum green, spectrum red). The latter can be directly visualized under fluorescence microscopy, whereas hapten-labeled probes need signal detection with antibodies conjugated with fluorochromes, such as fluorescein isothiocyanate (FITC) and rhodamine (Raudsepp & Chowdhary, 2008; Rubes *et. al.*, 2009).

Since its inception, FISH has become an essential tool in genomics and clinical cytogenetics. With this technique, researchers can precisely map and order molecular markers within a chromosome, identify numerical and structural chromosomal aberrations, define chromosome homologies between species, study insertions and/or deletions in chromosomal segments, and identify marker chromosomes involved in certain hereditary diseases and cancers (Rubes *et. al.*, 2009). Importantly, even though molecular cytogenetics has greatly benefited the study of mammalian chromosomes, conventional cytogenetics still plays an essential role to this day and should not be disregarded as so.

The resolution of FISH (*i.e.,* the minimum distance between two probes in which signals can be distinguished) depends on the desired application, and can vary from 1-5 Mb in metaphase chromosomes to around 100-300 kb in the case of mechanically stretched chromatin fibers (fiber-FISH) (Raudsepp & Chowdhary, 2008; Rubes *et. al.*, 2009). Probe size may also vary from short telomeric or centromeric repeats spanning a few hundred nucleotides, to large 120-200 kb long genomic inserts cloned in Bacterial or Yeast Artificial Chromosomes (BACs and YACs, respectively), or even composite probes

spanning the entire chromosome, called painting probes (Chowdhary & Raudsepp, 2001; Raudsepp & Chowdhary, 2008; Rubes *et. al.*, 2009). The latter are generated by chromosome flow-sorting or microdissection, and subsequent amplification by degenerate-oligonucleotide-primed PCR (DOP-PCR) (Telenius *et. al.* 1992; Rens, 2006) or by other whole genome amplification (WGA) methods (Sorensen *et. al.,* 2007, Arneson *et. al.,* 2012). These probes can then be used for chromosome painting or comparative chromosome painting, also known as Zoo-FISH (Scherthan *et. al.*, 1994; Chowdhary & Raudsepp, 2001). The main difference between the two techniques is that in chromosome painting the probe and the target originate from the same species, whereas in comparative painting (or Zoo-FISH) the probe and the target are generated from different species (Chowdhary & Raudsepp, 2001; Raudsepp & Chowdhary, 2008; Rubes *et. al.*, 2009).

Zoo-FISH experiments were first reported in 1994, when human chromosome-specific painting probes were successfully hybridized to chromosomes of mouse, Indian muntjac and fin whale (Scherthan *et. al.*, 1994). Remarkably, these species diverged around 80 million years ago (MYA). Since then, Zoo-FISH has been instrumental in transferring gene map information from “map-rich” species (such as humans) to “map-poor” species, and to study chromosomal homologies among close to 100 distantly and closely related mammalian species (Ferguson-Smith & Trifonov, 2007; Rubes *et. al.*, 2009; Raudsepp & Chowdhary, 2011; Raudsepp & Chowdhary, 2013). In animals,

nonetheless, the use of Zoo-FISH is limited by the availability of probes (Chowdhary & Raudsepp, 2001).

Importantly, with the use of FISH to precisely assign and order molecular markers on chromosomes, the need for high-resolution standardized chromosome nomenclatures for domestic species has become even more crucial. In order to meet the new demands, chromosome nomenclature standards were improved during the 1990s for a number of domestic species, such as dogs (*Canis familiaris*) (Breen *et. al.*, 1999), cattle (*Bos taurus*) (ISCNDB, 2000) and horses (*Equus caballus*) (ISCNH, 1997).

MOLECULAR CYTOGENETICS STUDIES IN MAMMALS

Over the course of approximately 50 years, from the inception of animal cytogenetics until the present day, chromosomes of most species of domestic animals have been studied in detail, and their evolutionary relationships established, through conventional and/or molecular cytogenetics. Reports on the use of these tools in goats (see Rubes *et. al.,* 2009; Schibler *et. al.*, 2009; Dong *et. al.*, 2013), sheep (see Iannuzzi *et. al.,* 2003; Di Meo *et. al.,* 2007; Goldammer *et. al.*, 2009; Rubes *et. al.,* 2009; Iannuzzi *et. al.*, 2014), bovids (see Frönicke & Wienberg, 2001; Iannuzzi *et. al.*, 2003; Chaves *et. al.*, 2004; Rubes *et. al.,* 2009), primates (reviewed by Stanyon *et. al.,* 2012), cattle (reviewed by Iannuzzi & Di Berardino, 2008), and many others show the usefulness of these

techniques to understand the evolution of structure and organization of mammalian genomes, as well as karyotypical relationships between different mammalian lineages.

Molecular cytogenetics tools have been extremely helpful for the study of species with high diploid number and uniform chromosome morphology. A good example is the cow (2n=60), in which the utility of conventional cytogenetics is hindered by their karyotype: all bovine autosomes are acrocentric, making the identification of individual chromosomes, even by differential banding techniques, rather difficult and unreliable. Thus, even though cytogenetics has been used to study bovine chromosomes for more than 50 years, it cannot be reliably used to characterize individual chromosomes involved in aberrations, especially in the case of smaller autosomes (reviewed by Iannuzzi & Di Berardino, 2008). Because of this, a standardized chromosomal nomenclature for cattle was only adopted in the late 1990s, after molecular markers were used to identify and characterize each bovine chromosome pair (ISCNDB, 2000). Moreover, the molecular markers and chromosome painting probes generated for cattle have been essential in the genomics era, by facilitating and improving the assembly of the bovine genome through integrated chromosome maps and evolutionary genomics studies (reviewed by Rubes *et. al.,* 2009).

The dog (2n=78) is another example of a species with a difficult karyotype. Similarly to cattle, all the canine autosomes are acrocentric. This, combined with the relatively high number of chromosomes, makes the

identification of individual chromosomes particularly difficult. The gross morphological similarities shared by the autosomes made the use of molecular cytogenetics tools necessary not only for the development of a standardized chromosome nomenclature, but also for the generation of integrated gene maps, essential in the improvement of the genome sequence assembly (reviewed by Breen, 2008; Becker *et. al.,* 2011). Molecular cytogenetic analyses played a pivotal role in anchoring the 7.6X genome assembly to physical chromosomes by assigning approximately 1,000 BAC clones and genome sequence scaffolds to their chromosomal location by FISH (Breen *et. al.,* 1999; Breen *et. al.*, 2004; Breen, 2008).

The horse (2n=64) constitutes another example of widespread and effective use of molecular cytogenetics in evolutionary and clinical studies. Before the genome sequence was made available in 2009 (Wade *et. al.*, 2009), a plethora of tools were available for researchers to study the structure and organization of the horse genome, as well as chromosomal abnormalities and equine evolution (Chowdhary & Raudsepp, 2008). Notably, the horse is one of the farm animals (together with the cow) with the most complete set of genome resources, which include a ~1,000 marker cytogenetic map, a radiation hybrid (RH) map composed of more than 4,000 markers (Chowdhary *et. al.,* 2003; Raudsepp *et. al.*, 2008), BAC contig maps spanning poorly assembled and/or regions of interest in the genome such as the PAR (Raudsepp & Chowdhary 2008) and the Y chromosome (Raudsepp *et. al.*, 2004), complete sets of whole

chromosome and chromosome-arm specific painting probes (Raudsepp & Chowdhary 1999; Yang *et. al.*, 2004; Trifonov *et. al.*, 2008), among others (reviewed by Rubes *et. al.*, 2009; Raudsepp & Chowdhary, 2013).

The pig (*Sus scrofa*, 2n=38) is another example of a domestic species whose economic importance led researchers to develop tools for chromosomal studies relatively early, mainly to investigate structural and numerical aberrations that affect the health and welfare of these animals (reviewed by Rubes *et. al.*, 2009; Raudsepp & Chowdhary, 2011).

A common feature among all the aforementioned species is that conventional and/or molecular cytogenetics studies preceded the genomics era and the availability of whole genome sequence assemblies by over a decade. Therefore, molecular tools were used to locally improve the sequence assembly and, in some cases, as an accessory to genome-wide studies, such as the use of FISH markers to validate CNVs and/or INDELs (Liu *et. al.,* 2010; Das *et. al.*, 2013; Ghosh *et. al.*, unpublished data).

More recently, the importance of a cytogenetic map to improve and further validate the existing low-coverage (2X) next-generation genome sequence assembly was shown in the tammar wallaby (*Macropus eugenii*, 2n=16). The assignment of genome sequence data to physical chromosomes by FISH started with the generation of high density physical maps for two chromosomes homologous to the human X: wallaby chromosomes X and 5 (Deakin *et. al.*, 2008). In 2011, Wang and colleagues developed the first-

generation tammar wallaby integrated genome map, by successfully combining the linkage map, a set of 492 cytogenetically mapped markers and the 2X genome sequence assembly, which provided the backbone for chromosome-wise assembly of the genome sequence. This was followed by a second-generation integrated genome map, with the addition of molecular markers to increase coverage and assist with anchoring and orientation of linkage groups. More importantly, the authors used FISH probes generated from genome sequence scaffolds to determine their precise chromosomal location, thus improving the genome sequence assembly (Wang *et. al.,* 2011). These resources were also used in a comparative study aimed at reconstructing the ancestral marsupial karyotype (Deakin *et. al.*, 2013). The overall strategy to construct an integrated map for the tammar wallaby genome, which involved the use of human homology and comparative data for the generation of molecular markers and integration of linkage, FISH and sequence maps, closely resembles the approach employed for the alpaca, which will be discussed below.

The current knowledge about the genomes of camelid species lags far behind what we know about the genomes of other domestic species (Avila *et. al.*, 2012; Raudsepp, 2014), even though the domestication of these species was contemporary to that of cattle (Taberlet *et. al.*, 2011), horses (Groeneveld *et. al.*, 2010) and dogs (Galibert *et. al.*, 2011), and despite their economic importance. This also means that the cytogenetic, molecular genetic and genomics tools to effectively identify and study genetic disorders, diseases, and

traits of interest in these valued species are limited. One of the reasons for this lack of knowledge lies in the camelid karyotype.

CYTOGENETIC STUDIES APPLIED TO CAMELIDS

Reports about the karyotypes of camelid species date back to 1960's, when first an erroneous diploid number of 2n=72 was proposed (Capanna & Civitelli, 1965; Hungerford & Snyder, 1966), but immediately corrected to 2n=74 (Hsu & Benirschke, 1967; Hsu & Benirschke, 1974; Koulischer *et. al.*, 1971; Taylor *et. al.*, 1968; Samman *et. al.*, 1993). These descriptions from almost 50 years ago have been followed only by around 20 published reports describing normal or aberrant chromosomes in these species (e.g., Drew *et. al.*, 1999; Fowler, 1990; Hinrichs *et. al.*, 1999; Hinrichs *et. al.*, 1997; Tibary, 2008; Wilker *et. al.*, 1994), and only 3 studies aimed at developing molecular cytogenetic tools for camelids (Di Berardino *et. al.*, 2006; Balmus *et. al.*, 2007; Avila *et. al.*, 2012).

One of the main complications in camelid cytogenetic analysis comes from their particularly difficult karyotype for analysis. Despite distinct anatomical and physiological differences, and specialized adaptations of the six extant species, *viz.,* the Bactrian and dromedary camels, alpaca, llama, vicugna and the guanaco, their karyotypes are extremely conserved, sharing the same diploid number of 74 chromosomes and nearly identical chromosome

morphology and banding patterns (Bianchi *et. al.*, 1986; Bunch *et. al.*, 1985; Samman *et. al.*, 1993; Di Berardino *et. al.*, 2006; Balmus *et. al.*, 2007; Avila *et. al.*, 2012; Raudsepp, 2014). The camelid karyotype arose as the result of rearrangements in the ancestral Cetartiodactyla karyotype: 12 fissions, 1 inversion and 1 fusion converted the putative Cetartiodactyla Ancestral Karyotype (CAK) (2n=52) into the camelid karyotype (Rubes *et. al.*, 2012) (Figure 4).

Morphological similarities and the relatively small size of some of the autosomes present serious challenges for identifying individual chromosomes within a species. The development of banding methods has helped resolve chromosome identification in several mammalian karyotypes, but not in camelids. Similarities in G-banding patterns between different chromosome pairs have resulted in discrepant karyotype arrangements in different studies (Bianchi *et. al.*, 1986; Bunch *et. al.*, 1985; Vidal-Rioja *et. al.*, 1989; Samman *et. al.*, 1993; Zhang *et. al.*, 2005; Di Berardino *et. al.*, 2006; Balmus *et. al.*, 2007).

Recent attempts to generate chromosome band nomenclature for the alpaca (Di Berardino *et. al.*, 2006; Avila *et. al.*, 2012) and the dromedary camel (Balmus *et. al.*, 2007) are not in agreement and provide no common platform for chromosome identification. As a result, and in contrast to most other domestic species, camelids still lack an international standard chromosome nomenclature.

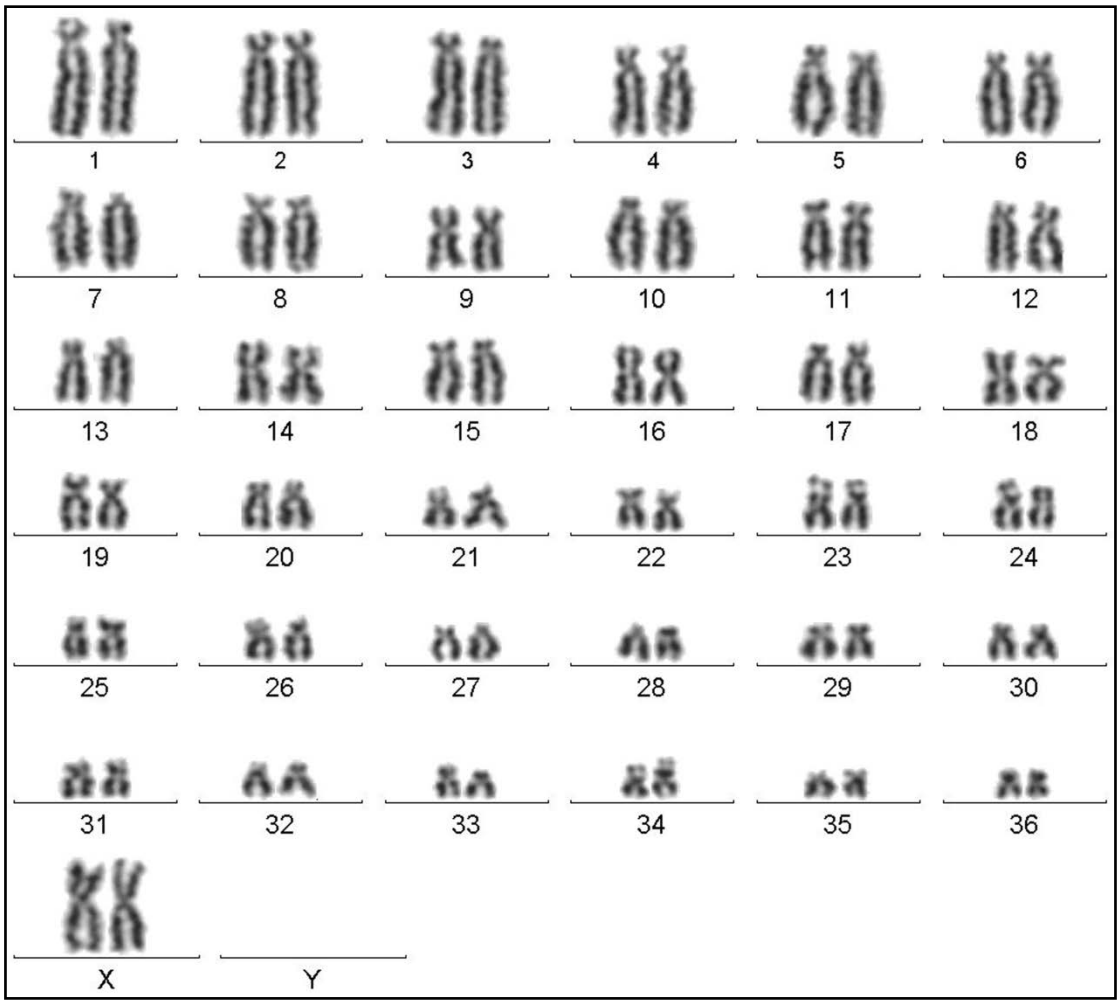

Figure 4. A Giemsa stained karyotype of a normal female alpaca.

This sets serious limitations for the advance of physical gene mapping, clinical cytogenetics, as well as for efficient cross-talk among laboratories (Avila *et. al.*, 2012; Raudsepp, 2014).

Thus, lessons from mammalian species with difficult karyotypes show that molecular and clinical cytogenetics can essentially benefit from the development of physical maps, which provide molecular markers for the identification of

individual chromosomes, chromosomal regions or bands (Avila *et. al.*, 2012). Besides, the development and mapping of molecular markers in the alpaca will benefit disease genetics and the ongoing alpaca genome sequence assembly, as well as cytogenetics and genomics in all other camelid species due to their karyotypical conservation.

THE ALPACA GENOME PROJECT

The Alpaca Genome Sequence Assembly and BAC Library

In 2002, the alpaca was nominated and funded by the National Human Genome Research Institute (NHGRI) at the National Institutes of Health (NIH) for low coverage (2X) whole genome sequencing, together with 21 other mammalian species. This was done primarily to use comparative genomic information to improve the annotation of the human genome. Also important considerations were the position of the family Camelidae at the base of the Cetartiodactyla radiation, as well as the need to increase the number of species with an available genome sequence for mammalian evolutionary studies (Johnson, 2011; Raudsepp, 2012). Information about the alpaca genome can also be applied to other camelid species, for linking genetic and phenotypic variation and for identifying genomic regions associated with genetic disorders and economic traits of interest (Raudsepp, 2012; Burger & Palmieri, 2013). The animal chosen for sequencing was a Huacaya female born in 2005, named

Nyala's Accoyo Empress Carlotta (who was also the DNA donor for the alpaca genomic BAC library CHORI-246; https://bacpac.chori.org/library.php?id=448) (Figure 5).

Figure 5. Nyala's Accoyo Empress Carlotta.

The first round of genome sequencing (*VicPac1.0*; http://www.ncbi.nlm.nih.gov/genome/905) was generated at the Broad Institute by whole genome shotgun (WGS) and Sanger technology, and released in 2008. The sequences produced 2.51X coverage of the estimated 2.9 Gb genome. Approximately 65% of the genome (1.9 Gb) was assembled into 298,413 sequence scaffolds, with a scaffold N50 of 230,521 (available at http://www.ncbi.nlm.nih.gov/assembly/

GCA_000164845.1/). In 2011, a 22X alpaca genome sequence assembly (*VicPac2.0.1*) was generated at Washington University (WashU) by combining the pre-existing WGS release with next-generation sequencing (NGS) technology (Roche 454-titanium platform) and assembled into 276,726 scaffolds, with a scaffold N50 of 7,263,804, covering 2.1Gb (72.4%) of the estimated 2.9Gb alpaca genome (http://www.ncbi.nlm.nih.gov/assembly/557168/).

Additionally, a Bacterial Artificial Chromosome (BAC) library is available for the alpaca. The CHORI-246 BAC library was developed by Dr. Pieter de Jong's laboratory at the Children's Hospital Oakland Research Institute (Osoegawa *et. al.,* 1998). The genomic DNA used for the construction of the alpaca BAC library was also obtained from Carlotta's (Figure 4) peripheral blood lymphocytes. It is composed of 202,752 clones, digested with EcoRI/EcoRI Methylase, inserted into the pTARBAC2.1 vector and transformed into DH10B (T1 resistant) electro-competent cells, with an average insert size of 200 kb. The library is also available as a set of 11 positively charged high-density nylon filters for screening by radioactive overgo hybridization. Each filter contains 36,864 clones, which represent 18,432 independent clones spotted in duplicate in a 4x4 clone array (https://bacpac.chori.org/library.php?id=448). In addition, a llama BAC library with approximately 9X genome coverage was constructed by Airmet and colleagues (2012). It is composed of approximately 196,224 clones grown into the pECBAC1 vector, with an average insert size of 138 kb. The sex of the

individual used to construct this library, however, is unknown (Airmet *et. al.*, 2012).

Despite these advancements, the utility of the alpaca genome sequence draft assembly is still limited because the sequences have not yet been properly annotated for genes and functional elements, and have not been anchored to physical chromosomes. One reason is that the entire progression of alpaca genomics has been inversed compared to the genome projects in other mammalian species. Historically, domestic animal genome sequence projects (the cow, the horse and the dog, for example) started with the construction of various maps (genetic linkage, synteny, radiation hybrid and cytogenetic maps) before heading for the ultimate map: the whole genome (WG) sequence. This was due in part to the technological limitations experienced at the time these genome sequence projects started (reviewed by Rubes *et. al.*, 2009; Chowdhary & Raudsepp, 2008). Such maps have been (and still are) instrumental for aiding and improving the genome sequence assembly in humans and most domestic species. In camelids, however, these maps are emerging only recently, after the genome has been sequenced and assembled *de novo.*

The Alpaca Whole Genome (WG) Radiation Hybrid (RH) Map

In 2007, a 5000rad alpaca x hamster RH panel consisting of 92 hybrid clones was generated and used for the construction of the first generation alpaca RH map. Among the 428 markers, systematically developed based on

human-dromedary camel chromosome homology (Balmus *et. al.*, 2007), 371 represented Type I (protein coding genes), and 57 represented Type II (microsatellite) markers, which corresponded to an average of 10 markers per chromosome. Additionally, higher resolution RH maps were generated for LPA16 and LPAX, with 50 and 60 markers, respectively (Perelman *et. al.*, 2011). Recently, taking advantage of available high-throughput genotyping technologies, a second generation RH map was constructed, harboring 4,590 markers distributed among 137 linkage groups (P. Perelman, personal communication). This high-resolution RH map, built upon the first generation map using a custom alpaca SNP chip, an Illumina GoldenGate assay and the Bovine SNP50 v2 BeadChip, represents one of the most complete and robust RH maps constructed for any domestic species, complemented by its inherent comparative value due to the strategy used for marker development (Perelman, 2011). The custom alpaca SNP chip designed for this map contained a total of 767 markers, of which 529 were used on the RH panel. Additionally, in order to obtain even chromosome-wide coverage of markers, an Illumina GoldenGate assay containing 1,560 probes was designed based on the 2X alpaca genome sequence assembly. Finally, approximately 4% of the bovine SNPs produced suitable signals for mapping on the alpaca (Perelman *et. al.*, 2011). The alpaca RH map is instrumental for supporting and guiding the genome sequence assembly, but also needs some support of its own. While radiation hybrid groups show the order and distance of genes and markers, they do not necessarily

show on which chromosome these markers are located, or whether markers from different RH groups are syntenic, nor do they show the centromere-telomere orientation of the RH linkage groups. This is where the physical chromosomes maps, also known as cytogenetic maps, come into the picture - to physically anchor the RH groups and genome sequence scaffolds to chromosomes.

The Dromedary Camel EST Library

Resources for camelid genomics also include an expressed sequence tag (EST) library for the dromedary camel, published by Al-Swailem and colleagues (2010). As a part of the Camel Genome Project efforts (available at http://camel.kacst.edu.sa), this study aimed at sequencing 70,272 ESTs from eleven tissues from three different dromedary camel cDNA libraries (from three different inbred individuals), in order to increase the efficiency of the dromedary camel genome sequence assembly (Al-Swailem *et. al.*, 2010). The authors generated 23,602 putative gene sequences, of which more than 4,500 were novel (*i.e*., are not homologous to any other mammalian genomic sequences). This effort, the first towards obtaining the whole genome sequence for the dromedary camel, also constitutes the foundation for comparative genomics studies in camelids by providing useful information on the transcriptome, gene duplication and for marker development in this species. Interestingly, the authors found an over-representation of “oxidation reduction” and “sensory perception of

smell" genes in the dromedary camel according to GO analyses (Al-Swailem *et. al.*, 2010). Also, the over 4,500 camel-specific sequences represent potentially novel or fast evolving genes in camelids, which can provide clues on unique adaptive features in these animals (Al-Swailem *et. al.*, 2010). Nonetheless, a whole genome sequence assembly is yet to be available for the dromedary camel.

The Bactrian Camel Genome

In 2012, The Bactrian Camel Genome Sequencing and Analysis Consortium published a report in which the genomes of a wild male Bactrian camel and a male domestic Bactrian camel, both from Mongolia, were sequenced using multiple NGS platforms and assembled *de novo* with *SOAPdenovo* (Bactrian Camels Genome Sequencing and Analysis Consortium, 2012). In this study, the N50 length of the scaffolds longer than 1 kb was 2 Mb, and the effective depth of coverage was 76X and 24X for the wild and domestic camel genomes, respectively. The estimated genome size for the Bactrian camel was 2.38Gb, with 34% being repetitive DNA. The 20,821 putative protein-coding genes had an average of eight exons and 1.3 kb of coding region per gene. Evolutionary studies using the *de novo* assembled genomes showed that cattle and pigs are the closest relatives of Bactrian camels, which diverged 55-60 MYA. Importantly, the authors also identified 40 IgH genes located in 16 different scaffolds, and thus were able to infer the organization of the IgH locus

in the Bactrian camel, an important step towards elucidating the molecular origins of camelid HCAbs (Bactrian Camels Genome Sequencing and Analysis Consortium, 2012).

In 2013, Burger & Palmieri sequenced the genome of a single male Bactrian camel using the Illumina Genome Analyzer IIx platform and assembled it *de novo* with CLC Assembly Cell 4.0.1beta. The authors then compared the contigs with dromedary camel ESTs (Al-Swailem *et. al.*, 2010). The 1.57 Gb assembly had an average 6.6X coverage and consisted of 781,462 contigs with an N50 of 2,814 bp. A total of 116,313 heterozygous SNPs were detected, representing only 5.5% of the 2,129,442 heterozygous SNPs reported by the Bactrian Camels Genome Sequencing and Analysis Consortium (2012) - probably due to low sequence coverage. Notably, a comparison between the Bactrian camel genome assembly and the dromedary camel EST library (Al-Swailem *et. al.*, 2010) identified an overlap of 20,014 out of the estimated 23,602 protein coding genes (84.8%) between the two species. Taken together, these 3 studies represent the foundation for subsequent genomics studies in camelids, providing the tools for genome-wide association, comparative and evolutionary studies in these species.

The Human-Dromedary Camel Comparative Map

The dromedary camel, human, cattle, and pig comparative chromosome map (also known as Zoo-FISH map) was published in 2007 by Balmus and

colleagues. This map identified human, cattle and pig counterparts for all camel chromosomes, except for chromosome 36 and the Y, and constitutes an important tool for chromosomal studies in camelids. Subsequently, as a testament of its importance, this study laid the foundation for the design of markers for the alpaca RH map (Perelman, 2011) and for the first alpaca cytogenetic map (Avila *et. al.,* 2012; see Chapter II). Balmus and colleagues (2007) used chromosome paints prepared from dromedary camel flow-sorted chromosomes (and further amplified by DOP-PCR) to characterize the karyotypes of the dromedary camel, the Bactrian camel, the guanaco, the alpaca, and a male dromedary camel x female guanaco hybrid (2n=74). Also, comparative homology maps between the dromedary camel and human, pig and cattle were generated by reciprocal Zoo-FISH, leading to a putative ancestral Cetartiodactyla karyotype with a diploid number of 2n=52 (Balmus *et. al.*, 2007). Notably, no homology was detected with human, cattle or pig using the CDR36-specific paint, possibly due to the largely heterochromatic nature of this small autosome, and because the amount of euchromatin is too small for FISH signals to be detected (Balmus *et. al.*, 2007). However, this technique does not allow the precise demarcation of evolutionary breakpoints within chromosomes, which can only be done by using selected FISH markers flanking these regions. This is where a map of cytogenetically mapped markers is needed for any species (Avila *et. al*, 2012).

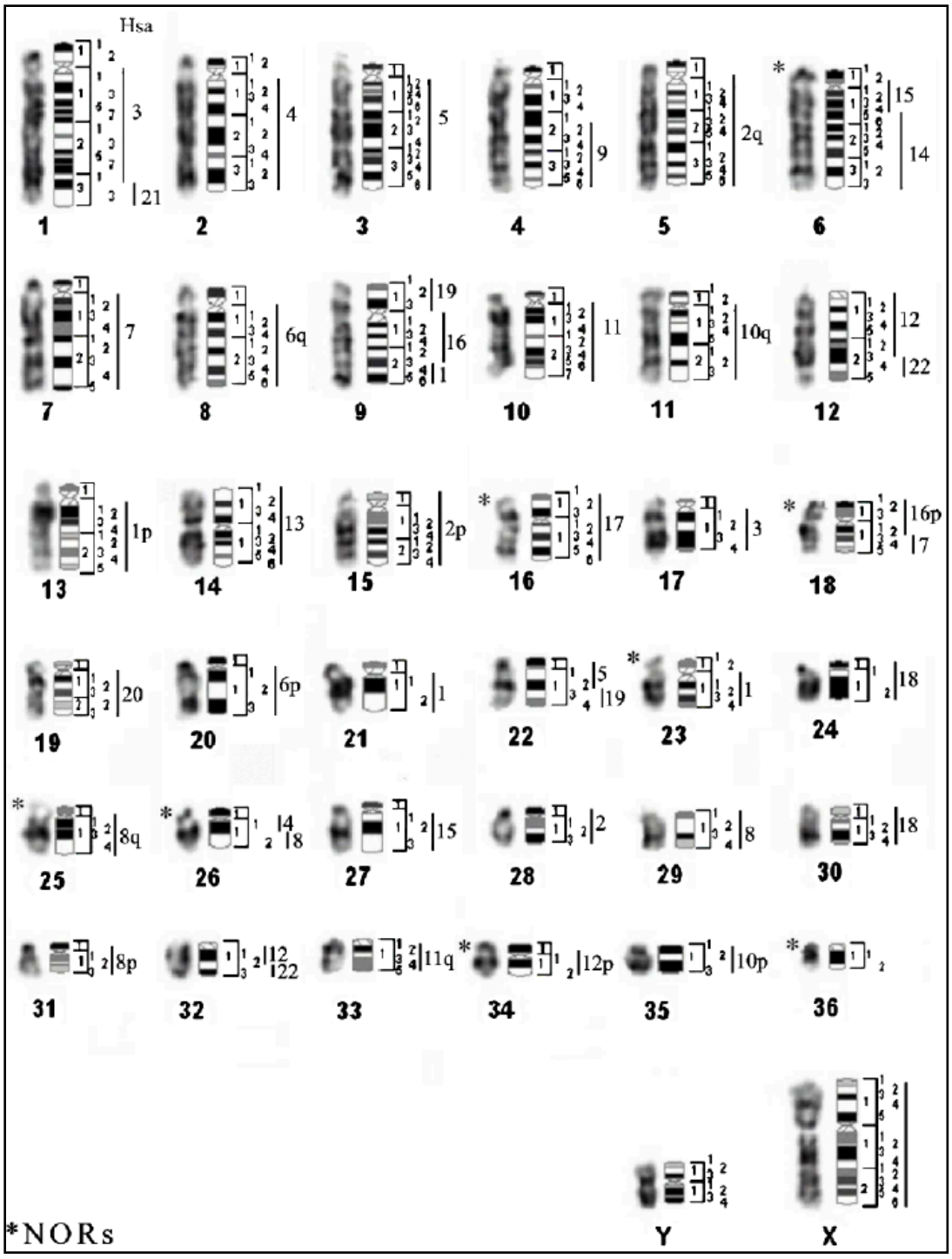

Figure 6. The human-camel Zoo-FISH map. GTG-banded chromosomes (left) and ideogram (right) of Camelus dromedarius showing chromosomal homology with human (HSA, right of each ideogram). Asterisks represent NOR chromosomes.

The authors reported no obvious differences at the chromosomal level among the different camelid species, except for size differences in heterochromatic blocks on the short arms of most chromosomes (CDR1, 2, 4-8, 11-13, 17-23, 29, 31-33, and 36) between Old and New World camelids. The results from reciprocal Zoo-FISH between camel and human were in full agreement with each other, with camel chromosome paints revealing 48 conserved segments in the human genome, except for chromosome 36, the Y chromosome and some heterochromatic blocks (Figure 6).

CAMELID CLINICAL CYTOGENETICS

Development of molecular markers for individual camelid chromosomes is also needed for the detection and analysis of chromosome rearrangements and abnormalities. Among these, of particular interest is a condition called the *Minute* Chromosome Syndrome (MCS). Cytogenetically, the *minute* is characterized as an abnormally small alpaca or llama chromosome 36. The condition has invariably been associated with female infertility due to ovarian hypoplasia or with disorders of sexual development (DSDs) (Drew *et. al.*, 1999; Avila *et. al.,* 2012). So far, all described cases have been heterozygous for the *minute*, suggesting that homozygosity might not be viable (Avila *et. al.*, 2012). Otherwise, the diploid number (2n=74) and gross morphology of other chromosomes are normal. Cytogenetic detection of the *minute* is relatively easy

due to its small size compared to other autosomes, though conventional cytogenetic methods are not capable of determining the origin of this aberration. Further, no molecular markers had been mapped to LPA36 (Avila *et. al.*, 2012) or its homologs in other camelid species. This chromosome also showed no Zoo-FISH homology to human or other non-camelid genomes (Balmus *et. al.*, 2007; Kulemzina *et. al.*, 2009), probably due to low representation of euchromatic sequences. Thus, there are no molecular tools currently available to study the mechanisms leading to the formation of the *minute* or underlying the particular phenotype of MCS. Overall, MCS is a unique cytogenetic condition found in alpacas and llamas, and its relatively frequent occurrence among infertile individuals (approximately 15% of all individuals with reproductive and/or developmental disorders subjected to cytogenetic analysis) (Avila *et. al.* 2012) justifies the launch of advanced studies to determine the molecular nature of the *minute* and to develop tools for diagnostics.

JUSTIFICATION

Alpacas and llamas have been steadily gaining popularity as fiber-producing and companion animals in the last 30 years, especially in the US. Nonetheless, their genomes remain the least studied among livestock species. This sets serious limitations for the use of advanced genomics tools in breeding programs to improve health and reproduction, and to propagate traits of interest

in these species. The Alpaca Genome Project, a starting point for camelid genomics, includes whole genome (WG) sequencing, radiation hybrid (RH) mapping and human-camel comparative chromosome painting (Zoo-FISH). However, there is no common platform that aligns various maps and precisely assigns them to individual chromosomes. Additionally, because of a high diploid chromosome number (2n=74) and similar morphology and banding patterns between different chromosome pairs, identification of individual chromosomes in alpacas and other camelids is difficult. Thus, there is an urgent need to develop a whole-genome cytogenetic map for the alpaca, in order to physically anchor the genome sequence and RH maps to chromosomes in this species, and to generate molecular tools for cytogenetic analysis in camelids. Among these tools, of special interest are those with the potential to be applied in assessing the molecular origin of the *minute* chromosome. Moreover, the alpaca whole genome cytogenetic map will serve as a resource for comparative evolutionary studies within the family Camelidae, by assessing the conservation of gene synteny and order in selected genomic regions of extant camelid species through comparative gene mapping.

OBJECTIVES

Whole Genome Cytogenetic Map for the Alpaca

The goal of this study is to generate an integrated cytogenetic map for the alpaca genome by assigning molecular markers to physical chromosomes by fluorescence *in situ* hybridization (FISH). The markers will be chosen to align the cytogenetic map with the existing radiation hybrid (RH), genome sequence and Zoo-FISH maps into an integrated, comprehensive physical map of the alpaca genome.

Cytogenetic and Molecular Analysis of the Minute Chromosome

The alpaca whole genome cytogenetic map will serve as a tool for the detection of chromosomal aberrations in camelids, as well as for the identification of candidate genes for economically and biologically important traits, such as coat color, diseases and reproduction related disorders, including the *Minute* Chromosome Syndrome. With regards to the *minute*, four more objectives will be pursued:

a) Molecular analysis of the normal chromosome 36 and the *minute*;

b) Development of molecular markers for the normal and *minute* chromosome 36 using unique sequences from this chromosome to develop DNA probes for FISH;

c) Next generation sequencing (NGS) and sequence analysis of flow sorted normal dromedary camel chromosome 36 (CDR36) and the alpaca *minute* chromosome;

d) Large cohort karyotyping of a population of alpacas with breeding records (consisting of 100 individuals) to assess the presence of the *minute* chromosome in normal individuals.

Comparative Mapping Between Alpaca and Dromedary Camel in Selected Genomic Regions

Gene synteny and order in selected genomic regions of different camelid species provide valuable insight on camelid chromosomal evolution. Therefore, comparative FISH mapping will be used to assess the architectural conservation between particular chromosomes in the alpaca and dromedary camel, such as those homologous to human chromosomes 4 and 8, which represent a mammalian ancestral synteny combination, and chromosomes 17 and X, representing evolutionarily conserved chromosomes in mammalian genomes. Moreover, the goal is to generate a medium resolution cytogenetic map for LPAX, so that it can be compared with dromedary camel, human, cattle and other mammalian X chromosomes, thus incorporating camelids in comparative evolutionary studies of mammalian sex chromosomes.

EXPECTED OUTCOMES

a) Integration of RH map and whole genome sequence data with alpaca chromosomes for a composite physical map for this species;

b) Development of molecular tools for clinical and molecular cytogenetics, especially for the identification of individual chromosomes involved in aberrations, such as the *minute* chromosome;

c) Validation and refinement of Zoo-FISH data;

d) Improvement of knowledge on comparative chromosome evolution within Camelidae;

e) Improvement of alpaca chromosome nomenclature;

f) Improvement of alpaca genome sequence assembly.

CHAPTER II

DEVELOPMENT AND APPLICATION OF CAMELID MOLECULAR CYTOGENETIC TOOLS*

INTRODUCTION

The development of cytogenetic maps for mammalian species constitutes a key feature for understanding the architecture and comparative evolution of chromosomes and karyotypes. Most domestic species have received considerable attention over the years due to their importance as production, model or companion animals. Detailed cytogenetic maps are available for individual cattle (Di Meo *et. al.*, 2011; Goldammer *et. al.*, 2009) and pig (see Raudsepp & Chowdhary, 2011) chromosomes and for the whole genome in horses (Raudsepp *et. al.*, 2008), dogs (Breen *et. al.*, 2004; Breen, 2008), cats (Davis *et. al.*, 2009), river buffalo (Di Meo *et. al.*, 2008), and sheep (Di Meo *et. al.*, 2007). These maps have been critical for anchoring genetic linkage and radiation hybrid maps, as well as genome sequence draft assemblies of these species to physical chromosomes.

---

*Reprinted from Avila, F.; Das, P.J.; Kutzler, M.; Owens, E.; Perelman, P.; Rubes, J.; Hornak, M.; Johnson, W.E.; Raudsepp, T. 2012. **Development and application of camelid molecular cytogenetic tools.** *J Hered*. Oct 29. [Epub ahead of print].

Also, cytogenetically assigned markers are important in clinical studies for precise demarcation of chromosome abnormalities and aberration breakpoints (reviewed by Ducos *et. al.,* 2008; Lear & Bailey, 2008; Rubes *et. al.,* 2009; Raudsepp & Chowdhary, 2011).

Even though the domestication of camelid species dates back to approximately 7,000 years ago (Kadwell *et. al.,* 2001), as long back as that of cattle (Taberlet *et. al.,* 2011), horses (Groeneveld *et. al.,* 2010), and dogs (Galibert *et. al.,* 2011), and considering that alpacas and llamas are gaining popularity as production and companion animals, camelid cytogenetics and physical chromosome mapping lag far behind those of other domesticated species. Reports about the karyotypes of camelid species date back to the 1960s, when first an erroneous diploid number of 2n=72 was proposed (Capanna & Civitelli, 1965; Hungerford & Snyder, 1966), which was quickly corrected to 2n=74 (Hsu & Benirschke, 1967; Taylor *et. al.,* 1968; Koulischer *et. al.,* 1971; Hsu & Benirschke, 1974). These studies from 50 years ago have been followed by only about 20 published reports describing normal or aberrant chromosomes in these species (e.g., Fowler 1990; Wilker *et. al.,* 1994; Hinrichs *et. al.,* 1997; Drew *et. al.,* 1999; Hinrichs *et. al.,* 1999; Tibary 2008), and only 1 effort has been made to develop molecular cytogenetic tools for camelids (Balmus *et. al.,* 2007).

One of the main complications in camelid cytogenetics is their particularly difficult karyotype. Despite distinct anatomical and physiological differences and

the specialized adaptations of the 6 extant species, namely, the Bactrian (*Camelus bactrianus, CBA*) and dromedary (*Camelus dromedarius, CDR*) camels, alpaca (*Lama pacos, LPA*), llama (*Lama glama, LGL*), vicugna (*Vicugna vicugna, VVI*), and guanaco (*Lama guanicoe, LGU*; Stanley *et. al.,* 1994), their karyotypes are extremely conserved, with the same diploid numbers and almost identical chromosome morphology and banding patterns (Bunch *et. al.,* 1985; Bianchi *et. al.,* 1986; Di Berardino *et. al.,* 2006; Balmus *et. al.,* 2007). Morphological similarities and the relatively small size of some of the autosomes present serious challenges for identifying individual chromosomes within a species. The development of banding methods has helped resolve chromosome identification in several mammalian karyotypes, but not in camelids. Similarities in G-banding patterns between different chromosome pairs have resulted in discrepant karyotype arrangements in different studies (Bunch *et. al.,* 1985; Bianchi *et. al.,* 1986; Vidal-Rioja *et. al.,* 1989; Zhang *et. al.,* 2005; Di Berardino *et. al.,* 2006; Balmus *et. al.,* 2007).

Likewise, the 2 recent remarkable attempts to generate chromosome band nomenclature for the alpaca (Di Berardino *et. al.,* 2006) and the dromedary camel (Balmus *et. al.,* 2007) provide no common platform for chromosome identification. As a result, and in contrast to other domestic species, camelids still lack an internationally accepted chromosome nomenclature, which sets serious limitations for the advance of physical gene mapping and clinical cytogenetics, as well as for efficient cross-talk between laboratories.

Lessons from other mammalian species with difficult karyotypes show that clinical cytogenetics can benefit from the development of physical maps that provide molecular markers for the identification of individual chromosomes, chromosome regions, or bands. An outstanding example is the domestic dog, a mammalian species with a high diploid number (2n=78) and a set of morphologically similar (acrocentric) autosomes that gradually decrease in size (Breen *et. al.,* 1999; Breen 2008). The need for unambiguous identification of individual canine chromosomes led to the generation of a collection of molecular markers for chromosome identification by fluorescence *in situ* hybridization (FISH; Breen *et. al.,* 1999; Breen *et. al.,* 2004; Breen 2008) and, subsequently, to a standardized chromosome nomenclature.

Building on these experiences, we developed a genome-wide set of molecular markers for the alpaca, assigned the markers to individual chromosomes by FISH, and applied the new tool in alpaca and llama clinical cytogenetics.

MATERIALS AND METHODS

Animals

A depository of fixed cell suspensions and chromosome slides of alpacas and llamas of the Molecular Cytogenetics and Genomics Laboratory at Texas A&M University was used for molecular cytogenetic analyses in this study. The

depository was established in 2005 and currently contains samples from 56 alpacas and 4 llamas. The samples have been cytogenetically characterized, cataloged, and stored at -20^{o}C.

Cell Cultures, Chromosome Preparations, and Karyotyping

Metaphase and interphase chromosome spreads were prepared from peripheral blood lymphocytes according to standard protocols (Raudsepp & Chowdhary, 2008a). The cells were dropped on clean, wet glass slides and checked under phase contrast microscope (x300) for quality. Chromosomes were stained with Giemsa, counted, and arranged into karyotypes using the Ikaros (MetaSystems GmbH) software. A minimum of 20 cells were analyzed per individual. Aberrant chromosomes were further analyzed by G- (Seabright 1971) and C-banding (Arrighi and Hsu 1971). The remaining cell suspensions were stored at -20^{o}C until needed.

Marker Selection and Primer Design

Human-camel Zoo-FISH data (Balmus *et. al.,* 2007) were used to select regions in the human genome that are homologous to individual alpaca chromosomes. Based on this, 24 human orthologs in segments homologous to 18 alpaca chromosomes (16 autosomes and the sex chromosomes) were identified in the National Center for Biotechnology Information (NCBI) Human Genome Map Viewer (http://www.ncbi.nlm.nih.gov/projects/genome/guide/

human/). Whenever possible, human genes were selected according to their likely involvement in reproduction or other economically important traits in alpacas. The alpaca genomic sequence for each gene was 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). Additionally, PCR and overgo primers for 22 genes, expected to map to 22 different alpaca chromosomes, were designed from alpaca cDNA sequences (generated by L. Wachter and kindly provided by Pontius J, Johnson WE, unpublished data). Details of all selected genes and the PCR and overgo primers are presented in Table 1 and Table 2, respectively.

Table 1. List of gene-specific markers and their cytogenetic locations in alpaca and human chromosomes and in human sequence map.

Gene Symbol	cDNA ID	Gene Name	Alpaca cytogenetic location	Human cytogenetic location	Human sequence map (chr:Mb)
AGPAT2	Lgnuc411	1-acylglycerol-3-phosphate O-acyltransferase 2 (lysophosphatidic acid acyltransferase, beta)	4q35-36	9q34.3	11:19.5
ARHGDIG	Lgnuc612	Rho GDP dissociation inhibitor (GDI) gamma	18q12-q13	16p13.3	16:00.3
ASIP	n/a	Agouti signaling protein	19q13-q14	20q11.2-q12	20:32.8
ATP6AP1	Lgnuc610	ATPase, H+ transporting, lysosomal accessory protein 1	Xq25	Xq28	X:153.6
BAG4	n/a	BCL2-associated athanogene 4	26q13	8p11.23	08:38.0
BRE	Lgnuc82	Brain and reproductive organ-expressed (TNFRSF1A modulator)	15q22-q23	2p23.2	02:28.1
C6orf211	Lgnuc618	Chromosome 6 open reading frame 211	8q24-q26	6q25.1	08:31.7
CAT56	n/a	MHC class I region proline-rich protein CAT56	20q13	6p21.33	06:30.5
CDC42BPB	Lgnuc584	CDC42 binding protein kinase beta (DMPK-like)	6q33	14q32.3	15:43.3
CSTF2T	n/a	Cleavage stimulation factor, 3' pre-RNA, subunit 2, 64kDa, tau variant	11q21	10q11	10:53.4
DSCC1	n/a	Defective in sister chromatid cohesion 1 homolog (*S. cerevisiae*)	25q14	8q24.12	10:00.8
DYRK1A	Lgnuc737	Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A	1q26-q31	21q22.13	21:38.7
EDN3	n/a	Endothelin 3	19q23	20q13.2-q13.3	20:57.8
FDFT1	n/a	Farnesyl-diphosphate farnesyltransferase 1	31q12-q13	8p23.1-p22	08:11.6
FGF5	n/a	Fibroblast growth factor 5	2q21-q22	4q21	05:21.1
FGFR2	n/a	Fibroblast growth factor receptor 2	11q22	10q26	12:03.2
GNB1L	Lgnuc743	Guanine nucleotide binding protein (G protein), beta polypeptide 1-like	32q13-q14	22q11.2	22:19.7
HEYL	n/a	Hairy/enhancer-of-split related with YRPW motif-like	13q22-q23	1p34.3	01:40.0
HS3ST3A1	n/a	Heparan sulfate (glucosamine) 3-O-sulfotransferase 3A1	16p13	17p12	17:13.3
HSD17B12	Lgnuc524	Hydroxysteroid (17-beta) dehydrogenase 12	33q12	11p11.2	11:43.7
KITLG	n/a	KIT ligand	12q22-q23	12q22	13:28.8
LARP4B	Lgnuc417	La ribonucleoprotein domain family, member 4B	35q13-q14	10p15.3	10:00.8

Table 1 continued

Gene Symbol	cDNA ID	Gene Name	Alpaca cytogenetic location	Human cytogenetic location	Human sequence map (chr:Mb)
LMO3	Lgnuc510	LIM domain only 3 (rhombotin-like 2)	34q12-q13	12p12.3	12:16.7
LPGAT1	Lgnuc63	Lysophosphatidylglycerol acyltransferase 1	23q14-q15	1q32	04:31.9
MITF	n/a	Microphthalmia-associated transcription factor	17q14	3p14.2-p14.1	04:09.7
NF1	n/a	Neurofibromin 1	16q14-q15	17q11.2	17:29.4
NPTN	Lgnuc606	Neuroplastin	27q13	15q22	16:13.8
PAX3	n/a	Paired box 3	5q33-q35	2q35	05:43.0
RAB38	n/a	RAB38, member RAS oncogene family	10q12-q14	11q14	12:27.8
RAG1	Lgnuc460	Recombination activating gene 1	10q25-q26	11p13	11:36.5
RALYL	n/a	RALY RNA binding protein-like	29q13	8q21.2	09:25.0
RB1CC1	n/a	RB1-inducible coiled-coil 1	29q15	8q11	08:53.5
SLC22A13	n/a	Solute carrier family 22 (organic anion transporter), member 13	17q13	3p21.3	03:38.3
SLC36A1	n/a	Solute carrier family 36 (proton/amino acid symporter), member 1	3q13-q16	5q33.1	07:30.8
SLC45A2	n/a	Solute carrier family 45, member 2	3q33-q34	5p13.2	05:33.9
SOX2	n/a	SRY (sex determining region Y)-box 2	1q21-q23	3q26.3-q27	06:01.4
STS-XY	n/a	Steroid sulfatase (microsomal), isozyme S	Xp16; Yq11	Xp22.32	X:0.7; Y:17.6
TGFBR3	n/a	Transforming growth factor, beta receptor III	9q25	1p33-p32	02:32.1
TRBV30	Lgnuc355	T cell receptor beta variable 30	7q24	7q34	09:22.5
TTR	Lgnuc409	Transthyretin	24q13-q14	18q12.1	18:29.1
TYRP1	n/a	Tyrosinase-related protein 1	4q21	9p23	09:12.6
Unknown transcript	Lgnuc134	Alpaca scaffold_48:270613:271380:1	2q33	4p15.3	4:00
Unknown transcript	Lgnuc681	Alpaca scaffold_374:105849:106822:1	30q12-q14	18q21	18:00

* ‘Lgnuc’ designate alpaca cDNA sequences (P. Perelman and J. Pontius, unpublished).

Table 2. List of all gene-specific PCR and overgo primers and the isolated BAC clones for each marker. BAC clones in bold were used for FISH mapping.

Gene Symbol	PCR primers 5'-3'	PCR Product Size (bp)	Overgo primers 5'-3'	CHORI-246 BAC clones
AGPAT2	F: GCAGGGACCATCAAGGTAGA R: AGGGTCCCTCTGTCCATTCT	232	F: TGAGGGCCTGCTTCTTCCACATAT R: TCCTGGGCTGTCTTGGATATGTGG	**101I14**, 135C03, 175I01
ARHGDIG	F: AGATGAGGCACTGGATGAGG R: AGCGCCTGCTTGTACTTGAC	117	F: ATCCAGCAGCTGGACCCAGACGAC R: ACTTGACCAGGCTCTCGTCGTCTG	**013L12**, 037H19, 110H09, 132K14, 147J04, 182A21
ASIP	F: ATGTCACCCGCCTCTTCCTA R: CCACAATAGAGACAGAAGGGAAA	156	F: AAGGAAGCCTGAGAAGCAACTCCT R: TCTAACAGGTTCTTGGAGGAGTTG	**018C13**, 133J24, 196A19, 227N08, 234F10
ATP6AP1	F: TCTTTGGCCTGAGAAGGAAA R: TAGGAAGGAAACCAGGAGCA	122	F: GAAAGGACCTCCATTCAGGTTGGG R: CAGTGATCACACCCAGCCCAACCT	**012K09**, 150E07, 150K15
BAG4	F: GGGACAAATACCGCCTCATA R: AAGCAGGATATCCGGGAAC	196	F: ACAAATACCGCCTCATACTCGGGG R: CAGGTGTGTAATAAGCCCCCGAGT	**229C18**
BRE	F: AGCACATATGGCTGCTTCCT R: GGTGTGCGCATTTATGTGTC	252	F: GGAAGCCATTGGAATGTCTTCATG R: GTGAGCGAAACGCTGCCATGAAGA	**074O15**, 062H18, 067N19
C6orf211	F: CCCAGAAAAGCTGTGTGTGA R: GGCAAAACGTTCAAAGGAAA	117	F: TGGGAACTCTCGGTTGCATAGAAA R: GAGTGCCTGTTAGATCTTTCTATG	**015P15**
CAT56	F: GCAGGGTCAGAGGTCTTGAG R: TAACAGCTGCAGGCACAACT	204	F: TCAGTTCATAAGTGCAAGCGCTGG R: TGGGTTTCCCGCCTCACCAGCGCT	**003A09**, 092P17
CDC42BPB	F: AAGTCATCGCTGGCATCTCT R: TAAGGAGCCTCGACTTCCAA	107	F: CAAACGCTGGTGTGTCTGCACTTG R: AAGGAGCCTCGACTTCCAAGTGCA	**076A14**
CSTF2T	F: GGAATGGAGACCTGCACAAT R: GGATTACCCACCCCTGAAAT	200	F: GGAAACGAGAGGCATTGATGCAAG R: CCTGATCTCCATTCCTCTTGCATC	**086G21**, 044C15, 088E24, 117M20, 167O13, 174H22
DSCC1	F: CCTTGGTGGACAGACACTCA R: TGGAGCAATATCTTCTTCTGTCC	134	F: GGTTTAGCCTTGGTGGACAGACAC R: TAATTTCCGGTCTTGAGTGTCTGT	**071D11**, 082O06, 103C18, 104P21, 144O11, 159H07, 170J24, 188H08, 222J12
DYRK1A	F: ACGCCAGAGCTGTTCTCAGT R: CTCCTCACTGTTCAGCACCA	148	F: GTCTGTGCTCTTCACTTTTGGACC R: AACACCCACCAGCACTGGTCCAAA	**118E13**, 078M24, 155F12

Table 2 continued

Gene Symbol	PCR primers 5'-3'	PCR Product Size (bp)	Overgo primers 5'-3'	CHORI-246 BAC clones
EDN3	F: AGGTCAGTTTGGGGAGCAG R: TCCAGATGATGTCCAGGTGA	146	F: CATGCCTTACCTACAAGGACAAGG R: CAATAGTAGACACACTCCTTGTCC	**125P19**, 262I23, 151M14, 194O05
FDFT1	F: GGGCAAGTACGTGAAGAAGC R: GCGGAATAGCGCAGAAGTTA	170	F: TGGGCGACTTTGCTAAGCCAGAGA R: ACGGCCACATCAACGTTCTCTGGC	**262L13**,166A12, 262J17, 282K10, 287K22, 482J17, 498J01, 528F19
FGF5	F: AAGAGGGGGAAAGCTAAACG R: TTCTCCGAGATGTGGAAAGG	184	F: ACACCTATGCCTCAGTGATACACA R: GATACACAGAACTGAGAACACGGG	**208O21**, 185E09
FGFR2	F: ACTGGACCAACACGGAAAAG R: TATAGCCTCCGATGCGATGT	159	F: GTCCAACACCAACTATGAGGTGGC R: TCCTTCCCGTTTTTCAGCCACCTC	**364G05**
GNB1L	F: CAGAACTCACCAGCTCACCA R: GTCCGCCAGTGAAACACAC	114	F: TCTCGATCACAGGTGCACAGAACT R: GGTTGGTGAGCTGGTGAGTTCTGT	**132C06**, 171L12
HEYL	F: CCTTCTTTCCCACCTCAACA R: TGGGGTAAGCAAGAGAGGAG	197	F: TTCCCACCTCAACAGCTATGCAGC R: TGAAGGCTCCATCTCGGCTGCATA	**020D23**, 023K10, 057P10, 067D21, 084E10, 084G08, 101A20
HS3ST3A1	F: TACCAACACAATCCCACACG R: TAGGACATGGACTCCCCATT	207	F: ACTGTGGACATCAGCATGATCCTC R: GACGGATGGCTTCTGAGAGGATCA	**080J08**, 106M05
HSD17B12	F: TCTTTGCCTAGGCTGTGGTT R: TTTTGAGGGTGCTAAATGCC	226	F: CCCAGGCTCTGAGTTTAGGTACCA R: TGTTTTGCAAATTTACTGGTACCT	**190I20**, 189J19, 217D24
KITLG	F: AGATGGTGGCACAGTTGTCA R: GTGTTCTTCCATGCACTCCA	140	F: GAGATGGTGGCACAGTTGTCAGTC R: GAAGATCAGTCAAGCTGACTGACA	**154O16**, 113O13, 117M20, 147G01, 155N05, 163N09
LARP4B	F: CAAAAACAAAAGCTCTCGGC R: TAGGAGTCAGTGCCATGCTG	232	F: TCATGTCTGGTGAGGCGCCAGGTT R: GAAGTGTAAGGGTGAGAACCTGGC	**325C10**, 213O02
Lgnuc134	F: GGTCCATTAGGAAGACAACTCA R: TGACACTTCATAAGAGGGGACAT	115	F: GCCAGTTATATCTGGACTTGGAGA R: GGGTGATACTATGGTGTCTCCAAG	**090N13**, 182M18
Lgnuc681	F: AAGCTCTGCCTTGTTCCAAA R: CACATTTCCATCCCCTGACT	189	F: ACTTGGATAAGGAAAGTAGTGCTC R: CATCCCCTGACTTACTGAGCACTA	**104K07**, 084P08
LMO3	F: GATTGACACGGGAACCAACT R: CAACTCTGAACTGGGGCAAT	111	F: CACCCCAGGTTCGCTGATCTATCA R: CTTAATGGGGTGATGTTGATAGAT	**058N01**

Table 2 continued

Gene Symbol	PCR primers 5'-3'	PCR Product Size (bp)	Overgo primers 5'-3'	CHORI-246 BAC clones
LPGAT1	F: TGGTCACATGTGGATCTGGT R: ACCCCACCCCCACTACTAAG	206	F: GGCTGGTCACATGTGGATCTGGTT R: CCCAAATGCATTCACAAACCAGAT	470A15, 497E10
MITF	F: CTCGAAAACCCCACCAAGTA R: ATAGCCATGGGGCTGTTG	185	F: CCACATACAGCAAGCCCAAAGGCA R: GTACTGCTTTACCTGCTGCCTTTG	033H02, 087L08, 119K06, 147I21, 211O19, 237C04, 277F19
NF1	F: CTGTTTTGGGGTTTTTGGAA R: TCCATTTGCTGATGGTGAAA	194	F: CTCTGATGTCACCCAAACACACAT R: GATGGTGAAACCCCTGATGTGTGT	429D06
NPTN	F: TTTACCTTCTGTTTTCAATGACCTT R: TGCCGGGTGAGAAAATCTAC	131	F: GGGTGCTCACACGCGGTACGTAAC R: AGTACATGCATCTACCGTTACGTA	**062D09**, 068E11
PAX3	F: AGTCCGATGAAGGCTCTGAC R: CAGCTCCTCCCTGGTGTAAA	158	F: GAAGGAGCCGAACCACCTTCACAG R: TCTTCAAGCTGTTCTGCTGTGAAG	**378C17**
RAB38	F: CAGCCACATTTGAAGCAGTG R: TGCAGAACTGGTCCATCTTG	159	F: AATGTGACCAGGGGAAGGATGTGC R: AGGCCATTGTTCACGAGCACATCC	176P13, 223K20, 238F03
RAG1	F: GGAATGAGCACAGACAAGCA R: GGAAGCCACGTTCTTCAGAG	255	F: CAGGACTGTGAAAGCCATCACGGG R: CTGGAAAATCTGCCTCCCCGTGAT	**084G11**, 108P04, 117M12, 206A18
RALYL	F: CTCCCCGCGCAGTAATTC R: GAGCCGGATGAAGAAGACC	127	F: AGGTCCACAGCCAGCGGGTCTTCT R: ATTTCGAGCCGGATGAAGAAGACC	**224G07**, 429H15, 429J17
RB1CC1	F: GCCAAGAAACTCTGCTCGTT R: TGGTAGGAATTTGAGCAATCC	110	F: GCCAAGAAACTCTGCTCGTTCTGT R: CGTGGACCAGACCTTCACAGAACG	**084G03**, 055C12, 153O01
SLC22A13	F: AGCTGCTGATCCTGGTGAGT R: AAGCTGTGGCTGAGGATGTC	250	F: CCTTCAACATGTTTGCCCAGGTCT R: TCATCCAGGACCATGAAGACCTGG	**035N16**
SLC36A1	F: GCCCCTTGAAAACAAAATGA R: ACAGTTGGGCAGGTTGAGAG	160	F: GGGGAGTCTGGGGTACCTGCAATT R: TTGGATATTTGCTCCAAATTGCAG	**080N08**, 270A14
SLC45A2	F: GGGTTACGTCTTGGGTGCTA R: GTGCGGGAATGTCTTTAAGC	164	F: TGCTCTCTTTGTGTTTTATCATCC R: GGGATACTGCACAGATGGATGATA	**220B13**, 068B04, 103P04, 124O10, 188G15, 244M04
SOX2	F: CACAACTCGGAGATCAGCAA R: CGGCAGCGTGTACTTATCCT	186	F: CTGCACATGAAGGAGCACCCGGAT R: GGGGCCGGTATTTATAATCCGGGT	**024K02**, 095J02, 119K06, 204G02, 192G17, 208C18, 272A05, 272G07

Table 2 continued

Gene Symbol	PCR primers 5'-3'	PCR Product Size (bp)	Overgo primers 5'-3'	CHORI-246 BAC clones
STS	F: ACGGATGCTTCTCCACACAC R: CCGTCCTGCTTTTCTCTGTC	301	F: AGCTGTCGCTGGGGAACATTATCT R: TGAAGCCACGGCTTCCAGATAATG	**119F05**, 074A23, 087H18
TGFBR3	F: ACACCATCCCTCCTGAGCTA R: GACCAGGAAACAGCTGCAAG	210	F: AACGGAGGCTTCCCCTTTCCTTTC R: TCCTGGCGATATCGGGGAAAGGAA	**046F06**, 060G12, 070H10, 126J10, 148M09, 168F06
TRBV30	F: AGCAGACTGTGGCTTCACCT R: AGCGCGAGGATAAAGAACAA	197	F: GTGCTCATGGCCATGGTTAAGAAA R: GGTCTCAGGAATCCTTTTTCTTAA	**010K16**, 089L16, 102A01, 142L12
TTR	F: GCCCCTACTCCTACTCCACC R: TTGTCTCTGCCCGAGTTTCT	174	F: GATCCAAAGGACGAGGGACAGGAT R: CTATCGGTTGCACGAAATCCTGTC	**248O05**, 254M24
TYRP1	F: CCACCAGGAGATCAGAGGAA R: CCACTTCACCAAAGCTCTCC	150	F: ATATTGGGGCCAGACGGCAACACG R: TGTTCTCAAATTGGGGCGTGTTGC	**129N17**, 176E21, 204G02, 277C17

Alpaca CHORI-246 BAC Library Screening and BAC DNA Isolation

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 25 or less 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 (Table 2) 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.

BAC DNA Labeling and FISH

The physical location of the genes was determined by fluorescence *in situ* hybridization (FISH) to alpaca metaphase and/or interphase chromosomes according to our protocols (Raudsepp & Chowdhary, 2008a). 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. Alpaca chromosomes were counterstained with 4'-6-diamidino-2-phenylindole (DAPI) and identified according to the nomenclature proposed by Balmus and colleagues (2007) with our modifications for LPA12, 24, 26, 27, 29, 33, 36, and Y (see Results).

Generation of Probes for LPA36, the Minute Chromosome, and the Sex Chromosomes

Flow sorted LPA36, LPAX and LPAY (Stanyon R, Perelman P, Stone G, unpublished data) were amplified and biotin- or DIG-labeled by degenerate oligonucleotide primed-PCR (DOP-PCR; Telenius *et. al.,* 1992; Rens *et. al.,*

2006). A probe for the abnormally small homologue of LPA36, the *minute* chromosome, was generated by chromosome microdissection, as previously described (Kubickova *et. al.,* 2002). Briefly, chromosome spreads from 3 animals carrying the *minute* chromosome were prepared on glass-membrane slides. Ten copies of the *minute* per animal were microdissected using the PALM MicroLaser system (P.A.L.M. GmbH, Bernried, Germany) and collected into a PCR tube containing 20 µL of 10 mmol Tris-HCl (pH 8.8). Chromosomal DNA was amplified and labeled with Spectrum Orange-dUTP (Vysis) by DOP-PCR (Telenius *et. al.,* 1992; Rens *et. al.,* 2006). Additionally, repeat-enriched blocking DNA was prepared by microdissection and DOP-PCR amplification of all alpaca centromeres. The labeled *minute* DNA was mixed with unlabeled centromeric DNA, denatured, preannealed to block repetitive sequences, and hybridized to normal and *minute*-carrying alpaca metaphase spreads as described earlier.

Comparative Genomic Hybridization

Genomic DNA from a normal male alpaca (control) and from 2 *minute* carriers (case) was isolated and directly labeled by nick translation (Abbott, Inc.) with SpectrumGreen-dUTP (Vysis) and SpectrumOrange-dUTP (Vysis), respectively. Labeled control and case DNA (each ~ 500ng) were mixed with 20µg of unlabeled alpaca repetitive DNA and 35µg of salmon sperm DNA (Sigma) and cohybridized to metaphase spreads of a normal male alpaca. The

comparative genomic hybridization (CGH) process and analysis of the results were carried out as described in detail by Hornak and colleagues (Hornak *et. al.,* 2009). For each CGH experiment, the red:green signal ratio was calculated for 10 metaphase spreads using the Isis-CGH software (MetaSystems, GmbH). A red:green ratio of >1.25:1 was indicative of chromosomal material gain, whereas a ratio of <0.75:1 indicated loss.

RESULTS

A Map of Molecular Cytogenetic Markers for the Alpaca Genome

The alpaca CHORI-246 genomic BAC library was screened with primers corresponding to 44 alpaca genes and expressed sequence tags. Altogether, 151 BAC clones were isolated and identified for the gene content (Table 2). Most of the genes were found in 2 or more clones, whereas each of the following 8 genes - *BAG4, C6orf211, CDC42BPB, FGFR2, LMO3, NF1, PAX3,* and *SLC22A13* - corresponded to only 1 BAC. One clone (that which gave the strongest and cleanest PCR amplification) for each of the 44 genes was selected for labeling and FISH mapping (Table 2). Each alpaca BAC clone produced a strong and clean FISH signal at 1 distinct location, and there were no chimeric clones or those that recognized multiple sites across the genome.

The 44 BACs were assigned to 31 alpaca autosomes and the sex chromosomes (Figure 7). The clone containing the steroid sulfatase (*STS*) gene

mapped to both the LPAXpter and Ypter and was considered pseudoautosomal (Figure 7, Figure 8e). Thus, the gene-specific BACs were assigned to 33 chromosomes, of which 11 chromosomes were demarcated by 2 distinctly located markers, either on the same arm (acrocentrics) or on 2 different arms (submetacentrics; LPA16 and LPAX). The relative order of all syntenic markers was determined by dual-color FISH (Figure 8). No markers were assigned to five chromosomes, namely, LPA14, 21, 22, 28, and 36 (Figure 7).

Precise cytogenetic locations of all BACs were determined by aligning the DAPI bands with the G-band nomenclature proposed by Balmus and colleagues (2007). However, we changed chromosome band numbering in compliance with the guidelines for human nomenclature (ISCN, 1995) by designating centromeres as p11/q11 and starting band numbering on both arms from the centromere. New ideograms were generated for LPA12, 24, 26, 27, 29, 33, 36 and Y (Figure 7), because LPA12, 29, 33 and 36 are submetacentric and not acrocentric as their counterparts in the dromedary camel karyotype (Balmus *et. al.,* 2007); LPAY is a small acrocentric compared to the submetacentric CDRY, and the banding pattern of LPA24, 26, and 27 differed from their CDR counterparts (Figure 7; Figure 9). Otherwise, the locations of all genes in the alpaca chromosomes were in agreement with the predictions of human-camel Zoo-FISH data (Balmus *et. al.,* 2007).

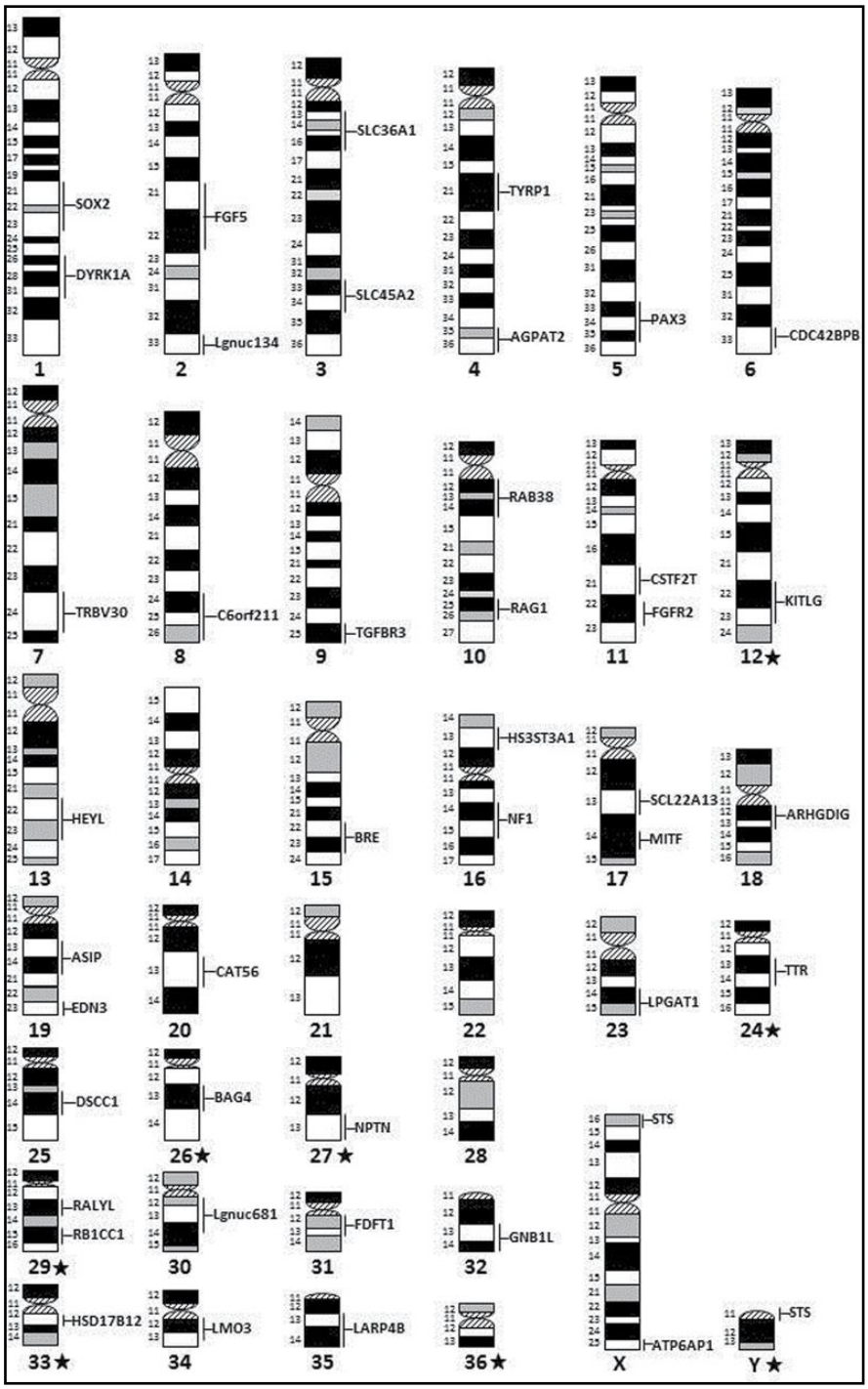

Figure 7. A cytogenetic gene map of the alpaca genome. Karyotype arrangement and ideograms are adapted from Balmus and colleagues (2007). The band nomenclature is corrected according to ISCN 1995. Chromosomes with ideograms adjusted for the alpaca are marked with a star.

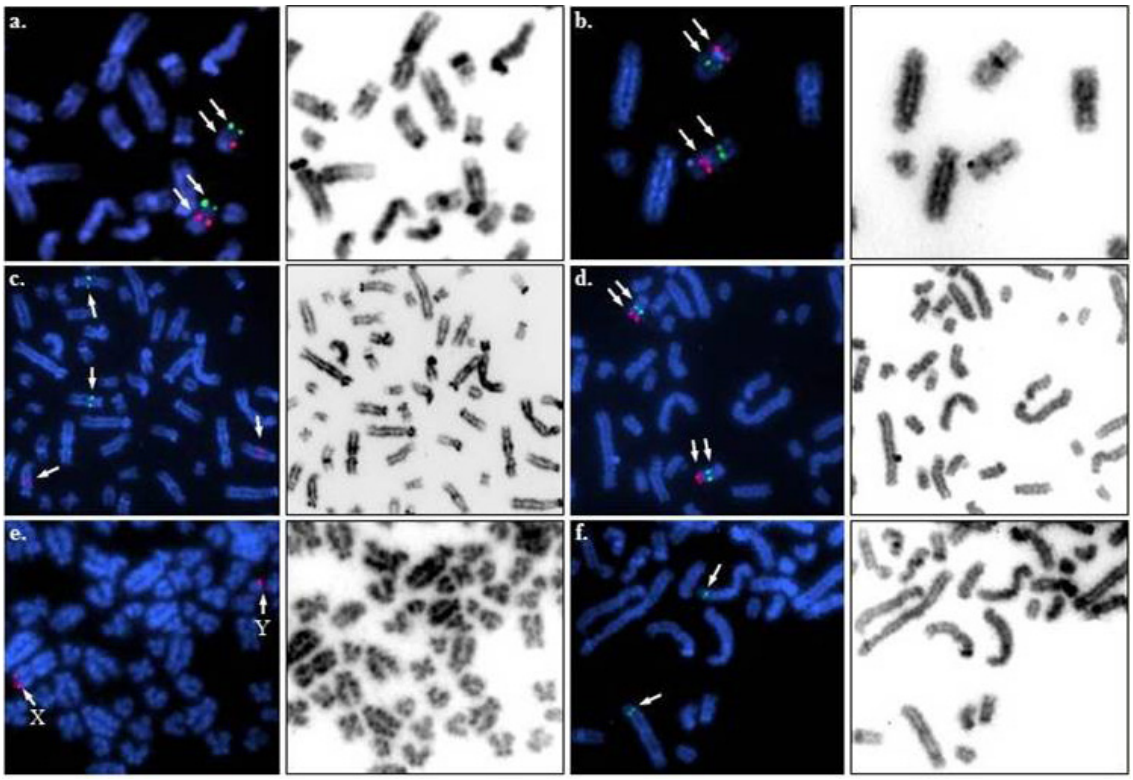

Figure 8. Partial alpaca metaphase spreads showing FISH results (left, arrows) and corresponding inverted-DAPI images (right) for selected markers mapped in this study. a. *EDN3* (green) and *ASIP* (red) on LPA19; b. *NF1* (green) and *HS3ST3A1* (red) on LPA16; c. *RAB38* (green) on LPA10 and *TYRP1* (red) on LPA4; d) *RALYL* (green) and *RB1CC1* (red) on LPA29; e. *STS* (red) on LPAX and LPAY; f. *FGFR2* (green) on LPA11.

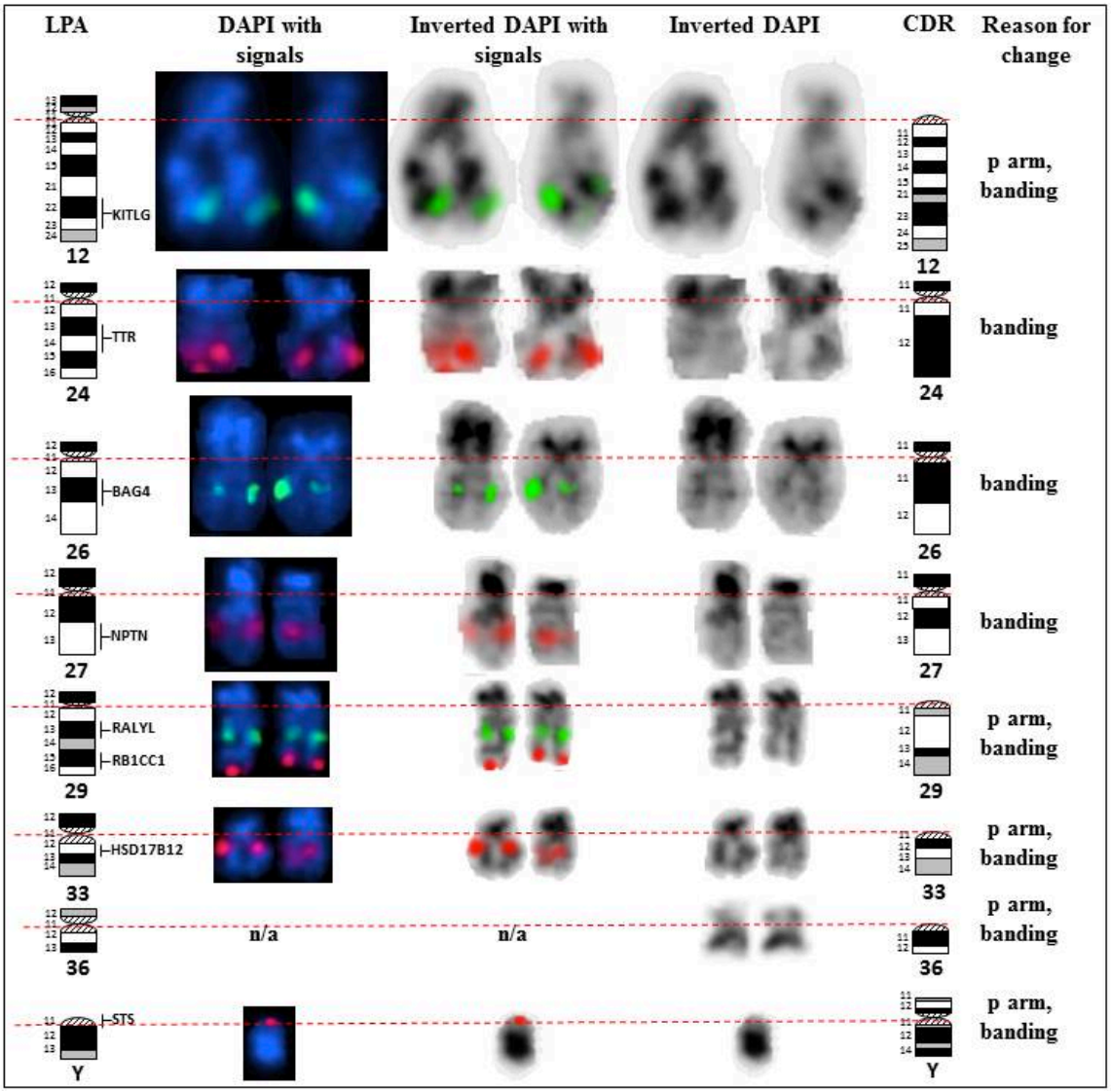

Figure 9. Improved ideograms for 8 alpaca chromosomes. Ideograms for LPA12, 24, 26, 27, 29, 33, 36 and Y (left) together with FISH signals, DAPI and inverted DAPI images for each are shown in relation to their CDR counterparts (right; adapted from Balmus *et. al*, 2007). Reasons for changes in LPA ideograms are indicated at the far right. Dotted red line demarcates the position of the centromere.

Cytogenetic Findings

In the past 7 years (2005-2011), the Molecular Cytogenetics and Genomics Laboratory at Texas A&M University (http://vetmed.tamu.edu/labs/cytogenics-genomics), in close collaboration with the Department of Animal Sciences at the Oregon State University, has received samples from 51 alpacas (both Suri and Huacaya) and 1 llama. The animals were referred for chromosome analysis due to various reproductive and/or developmental disorders, including abnormal sexual development, gonadal dysgenesis, subfertility, and sterility. Also, control samples were procured from a number of normal alpacas and llamas.

Among the phenotypically abnormal animals, chromosome abnormalities were detected in 12 cases (23%). Abnormal karyotypes included XX/XY chimerism, XY sex reversal, an autosomal translocation, and the presence of an abnormally small LPA36, also known as a *minute* chromosome (Drew *et. al.* 1999). Notably, the frequency of *minute* carriers was 17.7% of females with reproductive problems. A summary of the cytogenetic findings is presented in Table 3.

Table 3. Summary of cytogenetic finding in 51 alpacas and 1 llama subjected to chromosome analysis due to reproductive problems and/or abnormal sexual development.

Species	Karyotype	Chromosomal abnormality	Phenotype	Number of cases
Alpaca	74,XXm	*Minute* chromosome	Infertile female	8
Alpaca	74,XX/74,XY	Blood chimerism	Co-twin to a male	2
Alpaca	74,XY	Sex reversal	Female	1
Llama	73,XY(t20;?)	Autosomal translocation	Infertile male	1

Application of Molecular Tools in Camelid Clinical Cytogenetics

Autosomal translocation in a sterile male llama

A 10-year old male llama was presented for chromosome analysis due to infertility. Clinical examination showed that ~75% of his sperm had abnormal morphology (midpiece defects, nuclear and acrosomal vacuoles), whereas the testes and accessory glands appeared normal on ultrasound checkup.

Cytogenetic analysis determined that the llama had an abnormal karyotype 73,XY carrying an autosomal translocation. The derivative chromosome, as determined by G-banding, was submetacentric with size and morphology similar to the X chromosome (Figure 10a). The G-banding pattern suggested the probable involvement of LGL11 and LGL17 (Figure 10b), although cytogenetic identification of the origin of the translocation remained ambiguous.

Molecular cytogenetic analysis by FISH using LPAX and LPAY flow-sorted paints showed the presence of normal XY sex chromosomes and confirmed the autosomal origin of the derivative chromosome (Figure 10c). Dual-

color FISH with all 41 autosomal BAC clones refuted the involvement of LGL11 and LGL17 in the translocation. Instead, FISH revealed that the short arm of the derivative chromosome corresponds to LGL20 (Figure 10d), the chromosome carrying the MHC (Table 1). The origin of the long arm of the aberrant chromosome remains as yet undetermined.

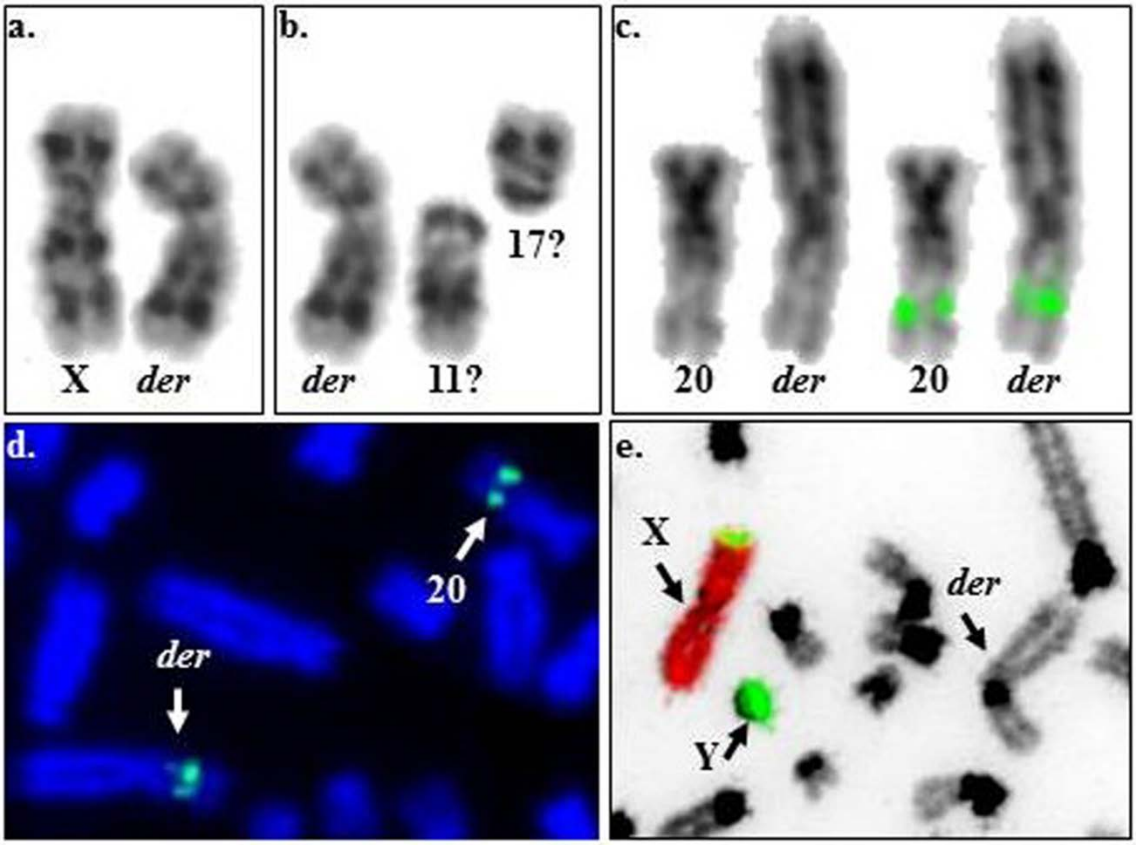

Figure 10. Autosomal translocation in a male llama. a) G-banded LGLX (left) and the derivative chromosome (*der*; right). b) G-banded *der* (left) and LGL11 and 17 (right) – thought to be involved in the formation of the *der*. c) side-by-side LGL20 and the *der* as inverted-DAPI image (left) and with *CAT56* signal (right). d) partial metaphase showing FISH signals by *CAT56* on LGL20 and the *der* (arrows). e) chromosome painting with LPAX (red) and Y (green) showing that *der* (arrow) is of autosomal origin.

The *minute* chromosome in infertile alpacas

Among the 11 infertile females, 8 animals had karyotypes with an extremely small LPA36 – the *minute* (Figure 11a). In all cases, the condition was heterozygous. Otherwise, chromosome number (74,XX) and gross morphology of other chromosomes in these animals were normal. Cytogenetic analysis determined that the *minute* is morphologically submetacentric, shows no distinct G-banding pattern, but stains positively by C-banding (Figure 11b), and is probably largely heterochromatic. However, it was not possible to identify the origin of the *minute* by conventional cytogenetic analysis.

Finally, FISH with two terminally located LPAX markers (*STS* and *ATP6AP1*) on metaphase spreads of *minute* carriers showed that the X chromosome in these animals is normal, thus challenging the hypothesis that the missing part of the *minute* has translocated to LPAX (Weber A, personal communication) (Figure 11c, 11d).

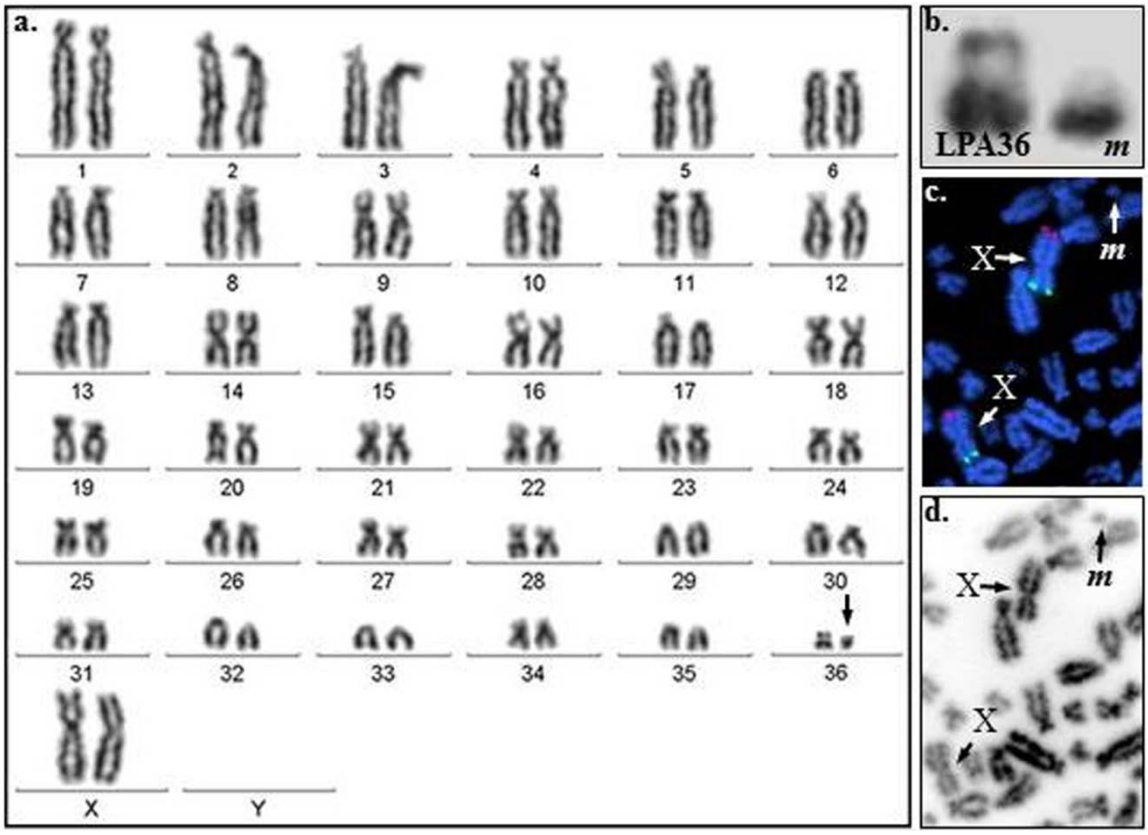

Figure 11. The *minute* chromosome. a) Karyotype of a female alpaca carrying the *minute* chromosome (arrow). b) G-banded LPA36 and the *minute* (*m*). c) FISH with *STS* (green) and *ATP6AP1* (red) on LPAX, and d. the same image as inverted-DAPI. The *minute* is shown as *m* (arrow).

Molecular hybridizations with flow sorted LPA36 and microdissected *minute* probes to metaphase spreads of a *minute* carrier showed FISH signals on LPA36 and the *minute* but also on all centromeres and intercalary heterochromatic regions (Figure 12a, 12b). In addition, the flow-sorted LPA36 contained DNA from another small autosome, LPA34 (Figure 12a, 12b). Although FISH results confirmed the largely heterochromatic nature of the

normal and *minute* LPA36, they did not bring us closer to understanding the origin of the abnormality.

Next, in order to test a working hypothesis that the *minute* results from a deletion rather than a translocation, CGH experiments were carried out on normal male metaphase spreads using genomic DNA from a normal male and a *minute*-carrying female as hybridization probes. No regions of genomic imbalance between the control and *minute*-carrying animal were detected, providing no experimental proof to the deletion theory (Figure 12c).

DISCUSSION

This study reports the generation of a genome-wide collection of 151 gene-containing BAC clones and the construction of a 44-marker cytogenetic map for the alpaca. According to our best knowledge, this is the first cytogenetic gene map for the alpaca or any other camelid species and the first application of the CHORI-246 alpaca genomic BAC library (http://bacpac.chori.org/library.php?id=448).

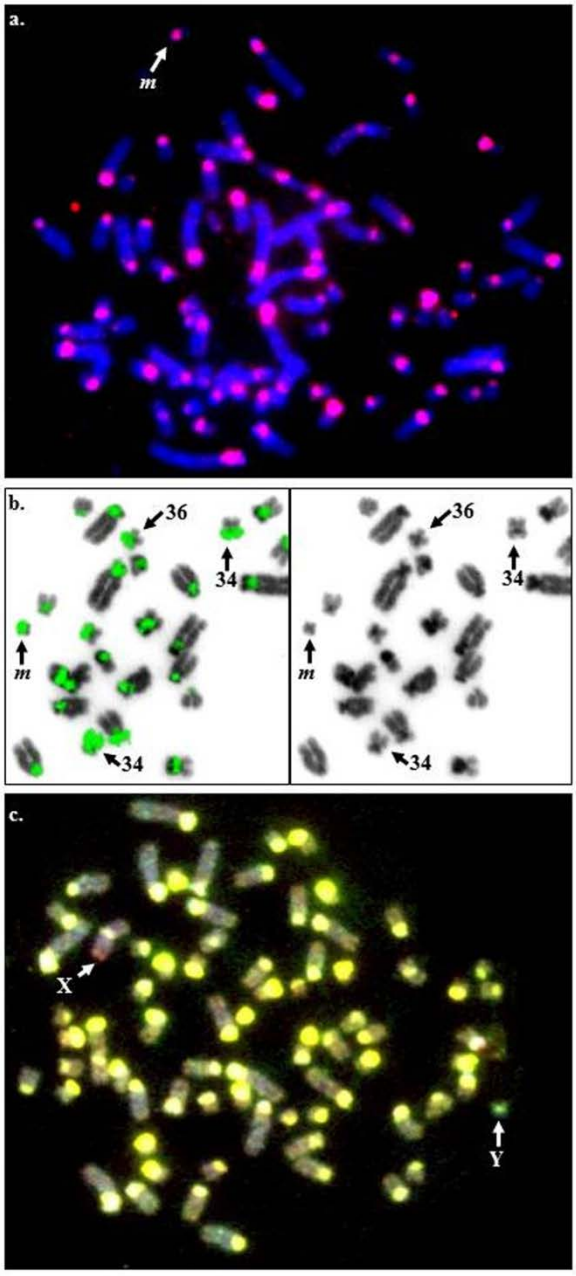

Figure 12. a) FISH with a microdissected *minute* probe on a metaphase spread of a *minute* carrier: signals are seen on all centromeres and the *minute* (*m*, arrow). b) FISH with a flow sorted LPA36+LPA34 probe on a *minute* carrier: the *minute*, LPA36 and LPA34 are indicated by arrows (left – FISH signals; right-inverted DAPI). c) CGH results with the genomic DNA of a normal male (green) and a female *minute* carrier (red). Arrows show the gain on the X and the loss on the Y chromosome.

Until now, the only molecular probes for camelids were whole chromosome paints from the flow karyotype of the dromedary camel, which have been used for camel-human, camel-cattle, and camel-pig Zoo-FISH studies (Balmus *et. al.,* 2007), for the study of chromosome evolution in Cetartiodactyla (Kulemzina *et. al.,* 2009) and ruminants (Kulemzina *et. al.,* 2011), as well as for the identification of the X and Y chromosomes in the alpaca karyotype (Di Berardino *et. al.,* 2006).

The BAC-based chromosome map, as presented in this study, confirms all and refines some of the known Zoo-FISH homologies. For example, assignment of 2 genes from HSA9 (*TYRP1*, HSA9p23; *AGPAT2*, HSA9q34.2) to LPA4 improved the demarcation of homologous regions between the human sequence map and the alpaca chromosome. Likewise, Zoo-FISH homologies were refined for 10 autosomes and the X chromosome by mapping 2 gene-specific markers on each (Figure 7, Table 1). In clinical cytogenetics, these markers will have a potential use for demarcating inversion and translocation breakpoints and determining the origin of complex rearrangements.

In some instances, particularly when 1 human chromosome shared evolutionary homology with 2 or more segments in the alpaca genome, the isolated BACs did not map to the expected alpaca chromosome. Instead, FISH signals were observed in another alpaca chromosome, which is homologous to the same human counterpart. This might be due to the relatively low resolution (~5Mb, Scherthan *et. al.,* 1994) and rather broad demarcation of evolutionary

breakpoints by Zoo-FISH. Therefore, no markers were assigned to LPA21, 22 and 28, which correspond to parts of HSA1, 5 and 2, respectively. In the case of LPA14, which corresponds one-to-one to HSA13 (Balmus *et. al.,* 2007), the BAC clone containing the pseudogene *ATP5EP2* mapped to a different alpaca chromosome (see Chapter IV).

Because the CHORI-246 BAC library was constructed from a female alpaca (http://bacpac.chori.org/library.php?id=448), we did not expect markers to be assigned to the Y chromosome. Nevertheless, a BAC clone for the *STS* gene produced FISH signals on both sex chromosomes, providing the first pseudoautosomal (PAR) marker for the alpaca genome. Interestingly, *STS* is an X-specific gene in humans (Skaletsky *et. al.,* 2003; Ross *et. al.,* 2005), and a non-PAR gene on horse sex chromosomes (Raudsepp & Chowdhary, 2008b), whereas in other non-rodent mammals studied so far, *STS* belongs to the PAR (Raudsepp & Chowdhary 2008b; Das *et. al.,* 2009; Raudsepp *et. al.,* 2011).

Thus, our results demarcate the location of the PAR in the alpaca sex chromosomes and provide the first gene-specific molecular marker for LPAY. Given that sex chromosome abnormalities are the most common viable cytogenetic defects associated with disorders of sexual development and reproduction in domestic animals (Villagomez & Pinton, 2008; Villagomez *et. al.,* 2009), including camelids (Fowler 1990; Wilker *et. al.,* 1994; Hinrichs *et. al.,* 1997; Drew *et. al.,* 1999; Hinrichs *et. al.,* 1999; Tibary 2008), the BACs

containing the *STS* gene will be of value for the identification of Y chromosome abnormalities in clinical studies.

Cytogenetic assignment of alpaca BAC clones in this study was carried out following the Giemsa (GTG)-banded chromosome nomenclature for the dromedary camel (Balmus *et. al.,* 2007) and not the one recently proposed for the alpaca (Di Berardino *et. al.,* 2006). Our primary argument was that the camel nomenclature is aligned with the human (Balmus *et. al.,* 2007) and other mammalian genomes (Kulemzina *et. al.,* 2009; Kulemzina *et. al.,* 2011), thus facilitating the development of gene-specific markers in the present and future studies. Also, Balmus and colleagues (2007) ordered chromosomes by size and not by morphological types as in the alpaca nomenclature (Di Berardino *et. al.,* 2006). The former seems to be the most logical approach in camelids, because heterochromatin and/or nucleolus organizer region (NOR) polymorphism in the short arms of some chromosomes (Bunch *et. al.,* 1985; Bianchi *et. al.,* 1986), combined with either ambiguous or too similar banding patterns in others, make morphological classification arbitrary. Furthermore, inverted-DAPI-banding patterns of alpaca chromosomes in this study corresponded well to the GTG-banded camel chromosomes and ideograms (Balmus *et. al.,* 2007), further justifying our approach. The few minor differences between the alpaca and dromedary camel homologues, namely, chromosomes 12, 24, 26, 27, 29, 33, 36 and Y, were adjusted in the resulting FISH map (Figure 7, Figure 9). However, despite the well-known evolutionary conservation of camelid karyotypes (Bianchi

et. al., 1986; Di Berardino *et. al.,* 2006; Balmus *et. al.,* 2007), it is anticipated that, with the expansion of the alpaca cytogenetic map, more differences between alpaca, dromedary camel and other camelid chromosomes will be revealed.

Successful identification of one of the chromosomes involved in an autosomal translocation in an infertile male llama (Figure 10d) demonstrated the immediate utility of the markers in camelid cytogenetics. Also, erroneous calling of the aberrant chromosomes by G-banding (Figure 10b) highlighted the limitations of conventional cytogenetic methods. This is in line with experiences from other domestic species, in which the development of molecular cytogenetic markers has considerably improved the quality and depth of clinical cytogenetic studies (Breen, 2008; Ducos *et. al.,* 2008; Lear & Bailey, 2008; Rubes *et. al.,* 2009; Raudsepp & Chowdhary, 2011). Efforts will be made to identify the other counterpart of the aberration; likely candidates could be LGL21 and 22. Interestingly, the translocation did not seriously affect meiosis because the animal produces sperm, though with morphological defects. The involvement of LGL20, the chromosome harboring the MHC (see Chapter IV) in the translocation is noteworthy, though studies are needed to elucidate the possible genetic consequences of this rearrangement.

As expected, no markers were assigned to LPA36 because, to date, there is no knowledge about mammalian homology to the smallest autosome present in the karyotypes of all 6 extant camelid species (Bianchi *et. al.,* 1986; Balmus

et. al., 2007). Zoo-FISH studies with flow sorted CDR36 in humans, pigs, cattle (Balmus *et. al.,* 2007), ruminants (Kulemzina *et. al.,* 2011) and other Cetartiodactyls (Kulemzina *et. al.,* 2009) concluded that the chromosome does not contain enough euchromatin to produce detectable FISH signals. Indeed, our cytogenetic studies and FISH results with normal and *minute* LPA36 paints support the idea that the chromosome is largely heterochromatic (Figure 12a, 12c).

The lack of LPA36-specific markers hinders the understanding of the origin of the *minute*. The *minute* might be either the result of a deletion or a translocation. Attempts to test the deletion theory by CGH were inconclusive because of the limited resolution of chromosome CGH. Similarly, the lack of specific markers for LPA36 did not allow testing the theory of a translocation. The only exception was the X chromosome, where FISH with markers from Xpter (*STS*) and Xqter (*ATP6AP1)* showed that both terminal segments were the same in *minute* carriers and controls and did not support LPA36/X translocation.

Because the *minute* is largely heterochromatic, we have considered the possibility that it is an accessory or a B chromosome. However, except for the heterochromatin, the *minute* in alpacas does not qualify as a typical B chromosome. In mammals, B chromosomes are found in some species, for example, canids; they are supernumerary to the standard karyotype, are completely heterochromatic or might contain amplified oncogenes, but are dispensable to the carrier (Vujosevic and Blagojevic 2004; Becker *et. al.,* 2011).

In contrast, the *minute* in alpacas is not completely heterochromatic (Figure 11), there is no variation in its numbers between individuals, and most importantly, it has been detected in infertile individuals. Furthermore, in all our cases, the *minute* was heterozygous, suggesting that homozygosity for the aberration might not be viable.

Despite these arguments, one cannot exclude the possibility that the *minute* is a normal size polymorphism of LPA36, which can be found at a certain frequency in the alpaca population, and the association of the *minute* with infertility is accidental. Testing this hypothesis needs large cohort karyotyping in alpacas with confirmed records of fertility. Yet, the *minute* is a unique feature of the alpaca genome, and further molecular studies, including direct sequencing of LPA36, are needed to determine the origin and molecular nature of this chromosome.

In summary, this collection of cytogenetically mapped markers forms a foundation for molecular and clinical cytogenetics in camelids. These and additional FISH-mapped markers will help the improvement and standardization of chromosome nomenclature for the alpaca and other camelids, as well as for anchoring and validating radiation hybrid maps and the genome sequence assembly (Breen 2008; Raudsepp *et. al.,* 2008; Lewin *et. al.,* 2009). This is of particular importance in alpacas, a species in which large sequence scaffolds have not yet been assigned to physical chromosomes (Ensembl: http://www.useast.ensembl.org/index.html). Finally, the 151 BAC clones

containing specific alpaca genes can be used as baits for target-enrichment capture and next generation sequencing (Mamanova *et. al.,* 2010; Horn 2012) to identify sequence variants and mutations associated with important health and disease phenotypes in these valued animals.

CHAPTER III

A CYTOGENETIC MAP FOR CAMELID CHROMOSOME 36: HOMOLOGY WITH HUMAN CHROMOSOME 7 AND THE ORIGIN OF THE *MINUTE* CHROMOSOME IN ALPACAS

INTRODUCTION

Cytogenetic maps constitute an important resource for studying the architecture and comparative organization of mammalian genomes, for chromosomal anchoring and integrating linkage, synteny, radiation hybrid (RH) and sequence maps (reviewed by Rubes *et. al.,* 2009; Graphodatsky *et. al.,* 2011; Raudsepp & Chowdhary, 2011; Das *et. al.,* 2012; Stanyon *et. al.,* 2012; Raudsepp & Chowdhary, 2013; Raudsepp, 2014), as well as for clinical cytogenetics (Avila *et. al.,* 2012; Fellows *et. al.,* 2012; Raudsepp & Chowdhary, 2013; Raudsepp, 2014).

In most domestic species, advanced whole genome integrated cytogenetic maps have been available for over a decade (reviewed by Chowdhary & Raudsepp, 2005; Rubes *et. al.,* 2009; Raudsepp & Chowdhary 2011; Raudsepp & Chowdhary, 2013). This is in sharp contrast with domesticated camelids, namely, the Bactrian camel (*Camelus bactrianus*, CBA), the dromedary camel (*Camelus dromedarius*, CDR), the alpaca (*Lama pacos*,

LPA), and the llama (*Lama glama*, LGL), in which gene mapping and molecular cytogenetic studies started only recently. The first and probably the most important breakthrough was the generation of a comparative chromosome painting (Zoo-FISH) map for the dromedary camel by determining its chromosomal homologies with humans, cattle and pigs (Balmus *et. al.*, 2007). This way, information from the advanced gene maps of these three species was indirectly transferred to the dromedary camel. Furthermore, as all extant camelids - domestic and wild - share outstanding similarities in their diploid chromosome number (2n=74), chromosome morphology and even banding patterns (Bianchi *et. al.*, 1986; Avila *et. al.,* 2012; reviewed by Raudsepp, 2014), the human-dromedary camel Zoo-FISH data largely apply also to alpacas and llamas. Indeed, Zoo-FISH data were recently validated and further refined by cytogenetic mapping of 44 gene-specific markers to alpaca chromosomes (Avila *et. al.*, 2012). The map provided molecular markers for 31 alpaca autosomes and the sex chromosomes, and represented the first gene map for the alpaca and for all camelid species (see Chapter II).

Despite this progress, no Zoo-FISH homology (Balmus *et. al.*, 2007) or gene map information (Avila *et. al.*, 2012) is available for the smallest camelid autosome, chromosome 36 (chr36). It is hypothesized that this is due to the small size and largely heterochromatic nature of this chromosome (Balmus *et. al.,* 2007; Kulemzina *et. al.*, 2009; Kulemzina *et. al.,* 2011; Avila *et. al.*, 2012). This lack of information hinders the completion of the alpaca cytogenetic map

and its integration with the genome sequence assembly (Avila *et. al.*, 2012; Perelman, 2011).

Development of molecular markers for chr36 is also needed for clinical cytogenetics, especially to study a condition in alpacas and llamas known as the *Minute* Chromosome Syndrome (MCS) (Drew *et. al.,* 1999; Avila *et. al.,* 2012; Fellows *et. al.,* 2012). The MCS is characterized by the presence of an abnormally small homologue of chr36 (thus called the *minute)* that is typically observed in subfertile or infertile female alpacas (Avila *et. al.,* 2012). Working hypotheses that the *minute* is the result of a translocation or a deletion involving a segment of chr36 have been tested, but not confirmed (Avila *et. al.,* 2012). An alternative hypothesis that the *minute* originates from a naturally occurring size polymorphism of chr36, as seen in other camelid chromosome pairs (Avila *et. al.*, 2012), will require large cohort karyotyping to be tested. Thus, the molecular origin of this chromosome remains unknown. Therefore, a cytogenetic map for chr36 is critical to properly characterize the MCS and devise tests for diagnostics.

In this study, we develop markers for alpaca chr36, construct a cytogenetic and comparative map for this chromosome, and apply it on the study of MCS. Markers for chr36 are developed by direct sequencing of normal chr36 and the *minute*, and by analyzing alpaca whole genome sequence scaffolds. The markers are mapped to alpaca, llama and dromedary camel chromosomes

by fluorescence *in situ* hybridization (FISH). The study of MCS also includes karyotyping of 100 normal alpacas.

MATERIALS AND METHODS

Animals, Chromosome Preparations and Karyotyping

The depository of fixed cell suspensions, DNA samples and chromosome slides of alpacas, llamas and dromedary camels of the Laboratory of Molecular Cytogenetics and Genomics at Texas A&M University was used for conventional and molecular cytogenetics analyses in this study. The depository comprises samples of 126 alpacas, 8 llamas and 2 dromedary camels, including 8 female *minute* carriers (Avila *et. al.*, 2012). All samples have been cytogenetically characterized, cataloged, and stored at -20ºC. Metaphase and interphase chromosome spreads were prepared from peripheral blood lymphocytes, stimulated with either concanavalin A (Sigma) or pokeweed (Sigma) mitogens according to standard protocols (Raudsepp & Chowdhary, 2008; Avila *et. al.*; 2012). Chromosomes were stained with Giemsa, counted and arranged into karyotypes using the Ikaros (MetaSystems GmbH) software. A minimum of 10 cells were analyzed per individual.

Development of Markers for Chromosome 36

Next generation sequencing of CDR36 and the alpaca *minute* chromosome

DNA from flow-sorted CDR36 (V. Trifonov, personal communication) and microdissected alpaca *minute* chromosome (Avila *et. al.*, 2012) was amplified by DOP-PCR (Telenius, 1992; Rens *et. al.*, 2006) and used to generate Illumina HiSeq paired-end sequencing libraries with the TruSeq DNA kit (Illumina). Briefly, overhangs were repaired using an exonuclease that removes 3' overhangs and a polymerase that fills 5' overhangs. The repaired fragments were purified using AMPure XP beads (Beckman Coulter). The fragments were adenylated at the 3' ends and adapters were ligated. The ligation products were electrophoresed on a 2% TAE agarose gel and stained with SYBR Gold Nucleic Acid Stain (Invitrogen). Fragments corresponding to 400-500 bp were excised and the DNA was purified using a QIAquick Gel Extraction kit (Qiagen). The DNA was enriched for adapter-ligated fragments by PCR (98°C for 30s; 10 cycles: 98°C for 10s, 60°C for 30s, 72°C for 30s, 72°C for 5min, hold 4°C), and purified with AMPure XP beads. The final products were analyzed for concentration and size using an agarose gel, a Bioanalyzer (Agilent), a NanoDrop ND-2000 spectrophotometer (Thermo Fisher), and a Qubit 2.0 flourometer (Life Technologies). Each library was normalized to 10 nM in Tris-HCl 10 mM pH 8.5, and sequenced on one lane of an Illumina Genome Analyzer

IIx with a paired-end 100 bp run at Texas A&M AgriLife Research Genomics and Bioinformatics Services Core Facilities.

The sequence reads were checked for quality and analyzed with the CLC Genomics Workbench, version 6 (CLC bio, www.clcbio.com). After masking repetitive sequences with RepeatMasker (http://www.repeatmasker.org), the reads were trimmed and assembled *de novo* with the CLC Genomics Workbench default settings (Minimum contig length = 100; Mismatch cost = 2; Insertion/Deletion cost = 3; Similarity fraction = 0.95). Unplaced alpaca genome sequence scaffolds were downloaded in FASTA format (ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes/vertebrates_mammals/Vicugna_pacos/Vicugna_pacos-2.0.1/Primary_Assembly/unplaced_scaffolds/FASTA/) and individually analyzed by BLAST against the assembled CDR36 and *minute* chromosome reads using the local BLAST tool in the CLC Genomics Workbench. The assembled sequences of CDR36 and the *minute* were also analyzed by BLAST for mammalian homology.

Analysis of alpaca genome sequence scaffolds

The second-generation alpaca RH map, with over 4,500 markers (Perelman, 2011; P. Perelman, personal communication), has been constructed by systematically mapping cDNA sequences based on human-camel comparative Zoo-FISH data (Balmus *et. al.* 2007), and DNA markers from alpaca genome sequence scaffolds (http://www.ncbi.nlm.nih.gov/assembly/

557168/). As a result, the majority of sequence scaffolds have been assigned to RH groups and the majority of RH groups have been assigned to chromosomes, though not to LPA36 (P. Perelman, personal communication). At the same time, a number of scaffolds remain unassigned due to the lack of RH markers and/or human homology. This implies that LPA36 sequences are likely among the unplaced alpaca genome sequence scaffolds (ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes/vertebrates_mammals/Vicugna_pacos/Vicugna_pacos-2.0.1/Primary_Assembly/unplaced_scaffolds/FASTA/).

Therefore, we selected 5 currently unplaced alpaca genome sequence scaffolds, with sizes ranging from 180 kb to 1.9 Mb, for the search of molecular markers for LPA36 (Table 4). The scaffolds were masked for repeats in RepeatMasker (http://www. repeatmasker.org/) and sequence homology with the human genome was determined by BLAST analysis (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The contents of protein coding genes, as well as comparative, cytogenetic and genome map locations of alpaca sequences in the human genome, were retrieved using the UCSC BLAT tool (http://genome.ucsc.edu/cgibin/hgBlat?command=start).

Table 4. List of markers mapped to alpaca (LPA) and dromedary camel (CDR) chromosomes, their cytogenetic location in both species and corresponding human (HSA) cytogenetic location. The markers are listed according to their location on HSA7.

Marker symbol	Gene name	Scaffold length (bp)	HSA cytogenetic location	LPA cytogenetic location	CDR cytogenetic location
ZNF853	Zinc finger protein 853	n/a	7p22.1	7p12	7p12
TMED4	Transmembrane emp24 protein transport domain containing 4	n/a	7p13	7q12	7q12
H2AFV	H2A histone family, member V	n/a	7p13	7q12	7q12
VWC2	Von Willebrand factor C domain containing 2	n/a	7p12.2	36q12-q13	36q12-q13
ZPBP	Zona pellucida binding protein 1	n/a	7p12.2	36q12-q13	36q12-q13
Scaffold 263	n/a	1,987,812	7p12.1	36q12-q13	36q12-q13
Scaffold 337	n/a	1,094,352	7p12.1	36q12-q13	36q12-q13
Scaffold 395	n/a	697,186	7p12.1	36q12-q13	36q12-q13
Scaffold 540	n/a	189,560	7p12.1	36q12-q13	36q12-q13
Scaffold 549	n/a	183,872	7p11.2	36q12-q13	36q12-q13
PSPH	Phosphoserine phosphatase	n/a	7p11.2	18q15	18q15
CCT6A	Chaperonin containing TCP1, subunit 6A (zeta 1)	n/a	7p11.2	18q16	18q16
SUMF2	Sulfatase modifying factor 2	n/a	7p11.2	18q16	18q16
GUSB	Glucuronidase, beta	n/a	7q11.21	18q16	18q16
CRCP	CGRP receptor component	n/a	7q11.21	18q16	18q16
TYW1	tRNA-yW synthesizing protein 1 homolog (*S. cerevisiae*)	n/a	7q11.21	18q16	18q16
AUTS2	Autism susceptibility candidate 2	n/a	7q11.22	18q16	18q16
WBSCR17	Williams-Beuren syndrome chromosome region 17	n/a	7q11.22	18q16	18q16
BCL7B	B-cell CLL/lymphoma 7B	n/a	7q11.23	18q16	18q16
MDH2	Malate dehydrogenase 2, NAD (mitochondrial)	n/a	7q11.23	18q16	18q16
MAGI2	Membrane associated guanylate kinase, WW and PDZ domain containing 2	n/a	7q21.11	7q14	7q14
HBP1	HMG-box transcription factor 1	n/a	7q22.3	7q22	7q22
TRBV30	T cell receptor beta variable 30	n/a	7q34	7q24-q25	7q24-q25

Development of markers orthologous to HSA7

Human Genome Map Viewer in NCBI (http://www.ncbi.nlm.nih.gov/ projects/mapview/map_search.cgi?taxid=9606&build=106.0) was used to select 18 protein-coding genes on human chromosome 7 (HSA7, Table 4) for comparative mapping in the alpaca and dromedary camel. Orthologous alpaca sequences for each gene were retrieved from the Ensembl Genome Browser (http://www. ensembl.org/index.html) for primer design.

Primer design

Overgo and PCR primers for the sequence scaffolds and for alpaca orthologs of HSA7 genes were designed as previously described (Avila *et. al.,* 2012). Briefly, the genomic sequence for each scaffold and gene was masked for repeats using RepeatMasker (http://repeatmasker.org), and used for the design of exonic PCR primers in the Primer3 software (Rozen & Skaletsky, 1998), as well as overgo primers in or around the PCR amplicons (Gustafson *et. al.*, 2003). Three pairs of overgo and 3 pairs of PCR primers were designed per scaffold, and one pair per each gene. Details of the selected scaffolds and genes, as well as the PCR and overgo primers, are presented in Table 4 and Table 5, respectively.

Table 5. List of all gene-specific PCR and overgo primers, and the isolated BAC clones. BAC clones in bold were used for FISH mapping.

Gene/ Scaffold symbol	PCR primers 5' - 3'	PCR product size (bp)	Overgo primers 5' - 3'	CHORI-246 BAC clones
Scaffold 263	F: CCAGATTGCCAGCTAGAGGT R: ATTGGGGCATTCAATACAGC F: CCAGCAAGATTCACGCATAA R: TCAAAAGAAGGGTGCTGACC F: GGCACCCTGTGTTGTCTCTT R: TGGGAGCCATTCATTTTCTC	154 165 151	F: TACATGTGGTTCGACAAGAGCTGG R: AAAGTAGAAGCGGGTGCCAGCTCT F: CCGTGATCGGCTGGTAAATGGAAA R: AAGATAAGAGCCCCGTTTTCCATT F: AAGTAGAGACTCAGAGTGATGAAG R: CAGGCACGCCCATTCTCTTCATCA	**037O03**, 097P07, 246A19, 262I13, 353G06, 376A22, **107I10**, 134P12, 139I04, 194N13, 298D08, 339O20
Scaffold 337	F:GATGCGCCATCTTAGTTCCT R: GAGACTGTCATTCATGCTGTT F: TGTTTGCCTTATCTACTTTGTAGCA R: GACCCCACCTGTCTGTGACT F: CGCAGGTGAAGCTCTGTTCT R: AAGAGGGTGTGTGAGGGTCA	173 158 162	F: CCTCTTGTCTTGAGCTTAATCCAA R: CTGTTGCCTCCCCTTCTTGGATTA F: GGTTGCATCCTACACGCCAAGGTG R: TTCTTTTCTCACCATGCACCTTGG F: CTGGTGGCTTCTGCACACCACGTA R: TTAAAAATGCACCCTGTACGTGGT	**096M08**, 097A22, 107O12, 109K04, 148D10, 153E12, 323N16, 341O07, **125E19**, 227G12, 332F18
Scaffold 395	F: AGATGGGAAGCACAGAGCAC R: CCTTAGGGCGGTTAGGAAAG F: CCACACAAATCGCTCCTACA R: GGACCCACTCACTGCTTTTG F: AAGGAGAGGGTGTCTGAGCA R: CCAGGGCATAAAGACAGGAA	206 169 207	F: AGGTGGACATGGCAATTCCGGGTC R: TCCGGGTCTCCTGGTCTTTCCTGA F: GGATCAAGCCTGGCTTTGCGACTC R: TGCGACTCTTAGACCCTCACGCAC F: CCCCTCTGCATAAAGCATCACCTT R: ATCACCTTCAGCCAAGGAGAGGGT	427M15, 439C16, 476N23, **515E22**, 216A12, 251B12, **281N10**, 216A12, **378G01**
Scaffold 540	F: AATGAGGTAAACACTGGTAATTGTG R: CGCCTTAATTAGGCAGCAGA F: AGGGAGGCAAACATCATCAA R: GAGGAAAACCCCCATTTCAT F: GAGAAGAAAGGTCATCTTTGAAGG R: GGGAGTGAAGTCCCCTCTTT	154 159 161	F: CGGTTGAGCTCCCCTTCCGAGTCT R: CATCAGTAAAATCCATAGACTCGG F: GCTTTAAGCATTTGCAGCACCTCT R: GCTGTGACCAGAAGCTAGAGGTGC F: ATGCGTGTCTTGGGGACAGCTGTC R: ATCTGACAACTTGAGTGACAGCTG	096M08, 097A22, 100A04, **107O12**, 109K04, 112N22, 148D10, 323N16, 341O07, 363A09, 141J02, 284F23, **059J14**, 092D03, 105L13, 153E12, 246A19, 247B04, **284F23**, 332P20, 333P12, 347B08, 379C24, 407F18

Table 5 continued

Gene/ Scaffold symbol	**PCR primers 5' - 3'**	**PCR product size (bp)**	**Overgo primers 5' - 3'**	**CHORI-246 BAC clones**
Scaffold 549	F: ACCCTGGGGAGCCATTAG R: CTCATTTCCGGAAGGACAGA F: GAACTCTGAAGGGCGAGGAG R: CTCAGAGGGCCATTCAAGAC F: TGATAGCAAGTTGTGGGTATCTG R: GTTTCGCGCTGGATTAGATG	168 166 155	F: CAGTTTCTTCCCAGGGGCCGGGTT R: TTACAAAAGTGTCAACAACCCGGC F: ACGGAGGTGCTTGCACATTTGGGG R: TTGAAGCCTTGTACTTCCCCAAAT F: TACAAGAACAGGAGGACCTGGACA R: TGCCCAGCTGTATTTCTGTCCAGG	**037P07**, 059K07, 083F15, 106C08, 121F04, 296J13, 313N10, 408P17, 148D10, **199B15**, 232E20, **239G17**, 274L19, 294C09
ZNF853	F: CCGATCCTTGTGCCTTACAT R: CTGGCCTCCTTTCTCAGATG	162	F: CCGATCCTTGTGCCTTACATTCTC R: AACTCAGGGCGTCAGGGAGAATGT	**074A23**
ZPBP	F: CCTGGTGTTCCAGAACTTCG R: GTACTTCAGCGGCAGCTTTC	106	F: CGTAACCTAAGCAGCTGATGGCCG R: GTACAGGGACAAGGAACGGCCATC	**003N14**, 012K21, 018E20, 100E03, 110F17, 134E17, 160O11, 178H11, 181B05
TMED4	F: CGTGTGCACCTAAACATCCA R: CTGTTCCACCTGATCAAGCA	120	F: ACCCTGAGATCGCTGCCAAGGATA R: TGCAGCTCTGTCAGTTTATCCTTG	**041D24**, 102E05, 136B02, 144H05, 154D01, 154D21, 156K05
VWC2	F: CCCAACTGCTTCGCAGAGA R: CTACATCTGCCTGCATTCGT	150	F: TCTATTTCAAGACATCCACAAGAA R: CTCCGATGGCCCCAAGTTCTTGTG	**003N14**, 012K21, 018E20, 100E03, 110F17,178H11, 181B05
H2AFV	F: CGCATCCACAGACATTTGAA R: TCAGCCGTGAGGTACTCCAG	101	F: TTTCCTGTGGGCCGCATCCACAGA R: TGCGAGTCTTCAAATGTCTGTGGA	041D24, **102E05**, 136B02, 144H05, 270M07
PSPH	F: CCCACTCAGAGCTGAGGAAA R: TTCTGACACAGCGTCCTCAA	131	F: CTGCTGATGCAGTGTGCTTCGATG R: ATGACTGTGCTATCGACATCGAAG	018I02, **057O17**, 119F15
SUMF2	F: GGAATGGGCAGACAACTAGC R: TGAAATCTTTGTTGGTGACAGG	156	F: TGGTGAAGGACCTGTCCGGGAAGT R: AAAAGGTTTTACTGCCACTTCCCG	018I02, 057O17, 119F15, **471B09**

Table 5 continued

Gene/ Scaffold symbol	PCR primers 5' - 3'	PCR product size (bp)	Overgo primers 5' - 3'	CHORI-246 BAC clones
CCT6A	F: TGTGTAGTTCCAGGTGCTGGT R: ATGAGCAATGCATCAGCAAA	122	F: TAAATACAAGCCCAGTGTAAAGGG R: TCCAAGTTGGGCCCTGCCCTTTAC	057O17, **083F20**, 119F15, 432O07
GUSB	F: GATGGGCCTGTGTCTGACTT R: CTGCATCCTCATGCTTGTTG	127	F: GCGGTCACAGAGAGCCAGTTTCTC R: AAGGTTTCCCATTGATGAGAAACT	127N21, **364B06**
CRCP	*F: GGAGAGTGAAGAACGGCTGA* *R: TACGCTGGGTCCTCTTCATC*	141	F: CAGGCCCAGAGGCAGAGCAGAAGA R: TCATTGTTGGTGTTTTTCTTCTGC	**127N21**
TYW1	*F: AAATTTGCGACAATTCTTGCTG* *R: CTTCTACTAGGTGATCATCTGGG*	100	F: TTACCTCCCTCGAGCTGCCTGTGG R: TTCAGATCAGTAATTGCCACAGGC	**018I02**, 050M03, 140J11
AUTS2	*F: GCACTTAAGCCTCAGGAACG* *R: CTGTCTGAGTCCCTTCAGCA*	204	F: AATGGCTTGTCCTTTCACTCCAAG R: GGCTGAGTCTGCTCTTCTTGGAGT	**403A12**, 422L15
WBSCR 17	*F: ACATCGAGCGGAAGAAGAAG* *R: AGTGGCAGATTCCACGCTAT*	130	F: CCATACAACAGCAACATCGGCTTC R: CGTTCCTCTTGGTGTAGAAGCCGA	**422L15**
BCL7B	*F: AGGTTGCTGAGGAAGAGGAA* *R: GGGTGGTGCTAGCTTTCTGA*	103	F: TTCTAGGTTGCTGAGGAAGAGGAA R: GTGGGGCACCTGAGTCTTCCTCTT	033O04, **098H24**
MDH2	*F: ATGACCCGGGATGACCTATT* *R: CTCACCGGGTTTGAAATGAT*	113	F: GAGGCCATGATCTGCATCATTTCA R: TGACACTCACCGGGTTTGAAATGA	**003P04**, 022H01
MAGI2	*F: ATGTGAGTGGGCCAGGAAC* *R: TGGTGATCTTGCTGGTCAGA*	155	F: GCATTCAGTAACACCTGTGGTTCT R: GGTGATCTTGCTGGTCAGAACCAC	**049H15**, 081H13, 100J24, 126A22
HBP1	*F: CCTGGATCACCACAGCTCTC* *R: TCTTTCCCTGGATACATCTGAG*	173	F: ACAGCTCAGGGACTGTGAGCGCCA R: CACTTGTTAGGAGAGGTGGCGCTC	**056N03**
TRBV30	*F: AGCAGACTGTGGCTTCACCT* *R: AGCGCGAGGATAAAGAACAA*	197	F: GTGCTCATGGCCATGGTTAAGAAA R: GGTCTCAGGAATCCTTTTTCTTAA	**010K16**, 089L16, 102A01, 142L12

Library screening and BAC DNA isolation

Overgo primers were radioactively labeled with [^{32}P] 2'-deoxyadenosine triphosphate (dATP) and [^{32}P] deoxycytidine triphosphate (dCTP) (Perkin Elmer), and hybridized to nylon high-density filters of the CHORI-246 alpaca bacterial artificial chromosome (BAC) library (https://bacpac.chori.org/library.php?id=448) as previously described (Avila *et. al.*, 2012). Positive BAC clones were identified and picked from the library. Their correspondence to individual genes or scaffolds was determined by PCR using gene- or scaffold-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 were evaluated by gel electrophoresis and NanoDrop spectrophotometry (Thermo Fisher Scientific).

Probe labeling, fluorescence *in situ* hybridization (FISH) and microscopy

Probe labeling and FISH mapping were carried out according to our protocols (Raudsepp & Chowdhary, 2008; Avila *et. al.*, 2012). Briefly, probe DNA (BAC DNA, rDNA) was labeled with biotin-16-deoxyuridine, 5'-triphosphate (dUTP) or digoxigenin (DIG)-11-dUTP, using Biotin- or DIG-Nick Translation Mix (Roche Diagnostics), respectively. The location of ribosomal RNA (rRNA) gene clusters was detected with a probe containing human 5.8S, 18S and 28S rDNA

(Maden*, et. al.*, 1987), a kind gift from Dr. Natalia Serdukova. Differently labeled probes were hybridized in pairs to metaphase chromosome or in groups of three to interphase chromosomes. Biotin and digoxigenin signals were detected with avidin-fluorescein isothiocyanate (FITC) (Vector Laboratories) and anti-DIG-rhodamine (Roche Applied Science), respectively. Images for a minimum of 10 metaphase spreads or 20 interphase nuclei were captured for each experiment, and analyzed with a Zeiss Axioplan 2 fluorescence microscope, equipped with the 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 *et. al.* (2007), with our modifications (Avila *et. al.*, 2012).

RESULTS

A Cytogenetic Map for Alpaca, Llama and Dromedary Camel Chr36

A total of 59 unique BAC clones were identified and isolated for unassigned sequence scaffolds, and 13 BACs (one clone per primer pair, Table 5) were used for FISH. Notably, all 5 hitherto unplaced alpaca sequence scaffolds mapped to alpaca, llama and dromedary camel chr36 - LPA/LGL/CDR36q12-q13 (Figure 13a, 13b), representing the first DNA sequences ever assigned to the smallest camelid autosome. Synteny of the

BACs was confirmed by dual-color metaphase FISH, and their relative order was determined by interphase FISH (Figure 13a, 13c).

Analysis of the 5 alpaca scaffolds by BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) showed similarity to sequences within a 4 Mb segment of the short arm of human chromosome 7 (HSA7p12.1-p11.2 at 51.5 Mb - 55.5 Mb). Importantly, this segment of HSA7 had been previously determined to be homologous to CDR18 by comparative chromosome painting (Balmus *et. al.*, 2007). Analysis of HSA7:51.5-55.5 Mb in Ensembl (http://www.ensembl.org/index.html) and UCSC Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) showed that the region contained RNA genes, pseudogenes and novel transcripts, as well as protein-coding genes. In order to test if the homologous synteny block (HSB) extended beyond the 4 Mb segment on HSA7, and to identify more putative genes on LPA/CDR36, we selected 5 genes downstream and 13 genes upstream of the segment, identified orthologous sequences in alpaca, and isolated and mapped corresponding BAC clones by FISH (Table 5, Table 6). Of the 18 selected genes, 2 - *ZPBP* and *VWC2* - mapped to LPA/CDR36q12-q13 (Figure 13a, 13b; Table 6). Dual-color FISH on metaphase and interphase chromosomes confirmed the synteny of *ZPBP* and *VWC2* with the 5 scaffolds, and mapped the 2 genes distal to those (Fig. 13a, 13d).

Thus, we successfully assigned the first molecular markers and protein-coding genes to camelid chr36 and revealed its homology to HSA7p12.1-p11.2.

Cytogenetic maps of chr36 in the alpaca, llama and dromedary camel comprise 7 markers and are essentially identical (Fig. 13d).

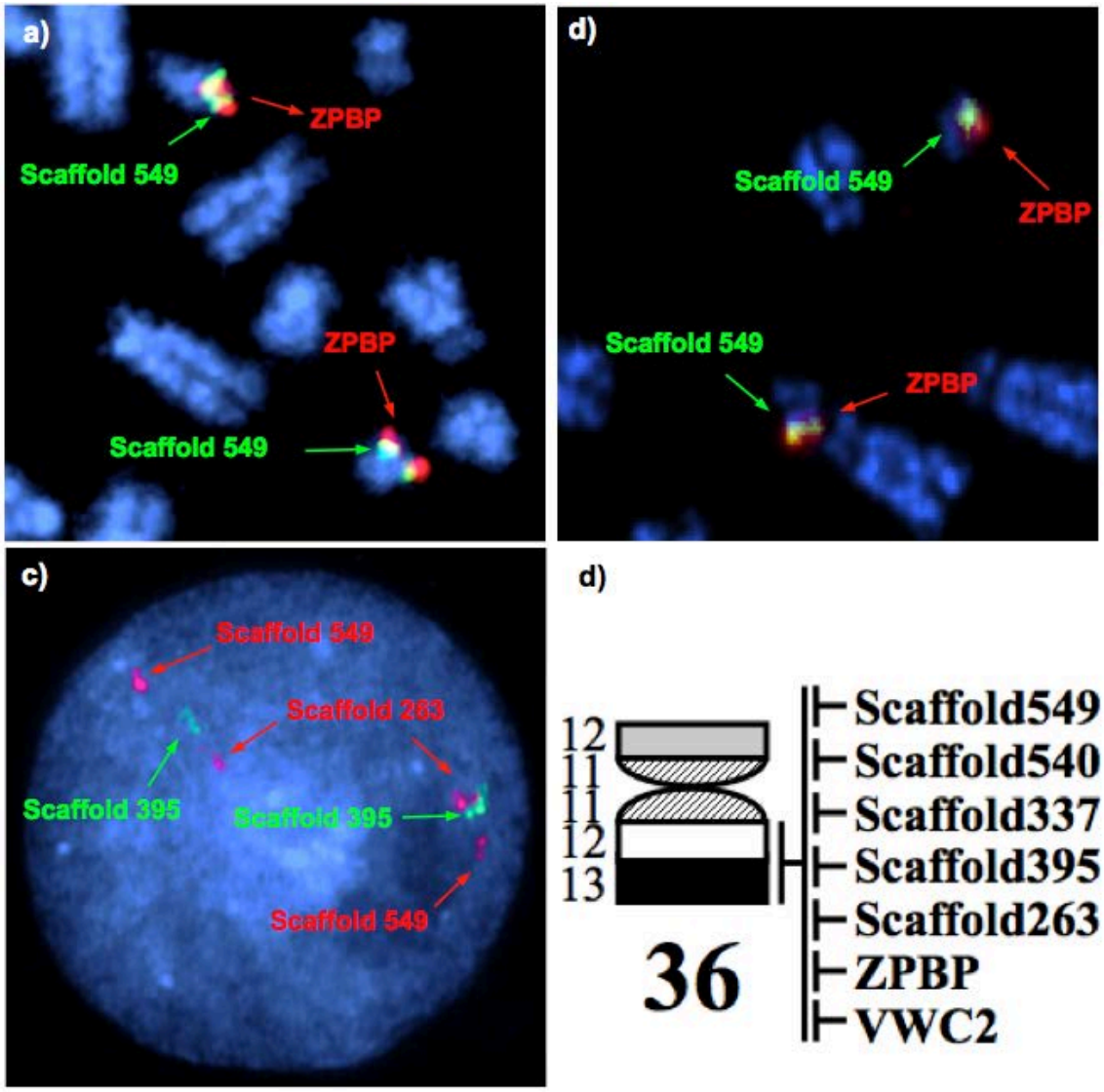

Figure 13. a) Dual-color metaphase FISH showing *ZPBP* (red signals) and Scaffold 549 (green signals) on LPA36. b) Dual-color metaphase FISH showing ZPBP (red signals) and Scaffold 549 (green signals) on CDR36. c) Interphase FISH showing the relative order of scaffolds 263, 549 (red signals) and 395 (green signal). d) A cytogenetic map for chr36. The ideogram represents LPA/LGL36, whereas CDR36 is acrocentric (Balmus *et. al.*, 2007).

Table 6. Chromosomal locations (chromosome:Mb position) of markers used in this study in different mammalian species and the chicken. The markers are listed according to their location on HSA7.

Gene/scaffold symbol	Human	Alpaca	Dromedary camel	Horse	Cow	Pig	Dog	Cat	Rat	Mouse	Chicken
ZNF853	7:6.6	7	7	13:1.9	25:39	n/a	6:11.9	E3:4.4	12:15.3	5:142.3	n/a
TMED4	7:44.6	7	7	4:15.3	4:77.5	18:55.6	n/a	A2:62.4	14:87.3	11:6.2	n/a
H2AFV	7:44.8	7	7	4:15.5	4:77.3	18:55.3	n/a	A2:62.5	14:80.3	11:6.4	n/a
VWC2	7:49.8	36	36	4:19.7	4:5.9	9:149.5	18:1.2	A2:65.6	14:90.9	11:11.1	2:80.4
ZPBP	7:49.9	36	36	4:19.8	4:5.7	9:149.6	18:1.3	A2:65.7	14:91.1	11:11.2	2:80.5
Scaffold 263	7:51.6	36	36	4:17.4	4:75.6	n/a	18:1.8	n/a	n/a	11:8.3	2
Scaffold 395	7:53.1	36	36	4:22.6	4:2.4	9:152.2	18:4.0	A2:68.5	n/a	n/a	2
Scaffold 337	7:53.5	36	36	4:23.1	4:1.9	9:152.5	18:4.4	A2:68.9	14:13.5	n/a	2
Scaffold 540	7:53.7	36	36	4:23.0	4:1.9	9:152.4	18:4.3	A2:68.9	n/a	11:15.0	2
Scaffold 549	7:55.5	36	36	4:25.2	n/a	9:154	18:6.3	A2:71.0	n/a	n/a	2
PSPH	7:56.0	18	18	13:18.0	25:27.9	3:17.1	6:0.49	E3:15.8	12:32.4	5:129.7	19:4.8
CCT6A	7:56.1	18	18	13:17.9	25:27.9	3:17.09	6:0.54	E3:15.8	12:32.4	5:129.8	19:4.8
SUMF2	7:56.1	18	18	13:17.9	25:27.9	3:17.08	6:0.55	E3:15.8	12:32.4	5:129.8	19:4.8
GUSB	7:65.4	18	18	13:17.7	25:28.1	3:16.9	6:0.73	E3:15.6	12:32.1	5:129.9	19:4.9
CRCP	7:65.5	18	18	13:17.6	25:29.7	3:16.7	6:0.81	E3:15.5	12:27.6	5:130	19:4.9
TYW1	7:66.5	18	18	13:16.9	25:28.6	3:16.2	6:1.2	E3:14.9	12:31.7	5:130.2	19:0.8
AUTS2	7:69.0	18	18	13:14.6	25:30.0	3:13.9	6:2.7	E3:13.5	12:25.1	5:131.4	19:1.3
WBSCR17	7:70.7	18	18	13:15.8	25:29.3	3:15.4	6:2.1	E3:13.8	12:30.3	5:130.8	19:1.0
BCL7B	7:72.9	18	18	13:11.2	25:34.1	3:10.5	6:6.6	E3:9.7	12:26.5	5:135.1	19
MDH2	7:75.6	18	18	13:10.4	25:34.7	3:9.9	6:7.3	E3:9.0	12:25.9	5:135.7	19:4.2
MAGI2	7:77.8	7	7	4:2.3	4:42.5	9:112.1	18:18	A2:83	4:10.9	5:19.2	1:11.9
HBP1	7:106.8	7	7	4:7.1	4:48.5	9:115.8	18:13.3	A2:78.1	6:59.9	12:31.9	1:14.3
TRBV30	7:142.5	7	7	4:95.9	4:106.9	18:7.7	16:6.7	A2:156.8	4:135.6	6:41.2	n/a

*n/a indicates that the map location could not be determined by BLAST analysis.

Next-Generation Sequencing of CDR36 and the Minute Chromosome Validates the Homology of Camelid Chromosome 36 with HSA7

In order to obtain unique sequences for chr36 and to investigate the molecular origin of the *minute* chromosome, CDR36 and the alpaca *minute* were sequenced by NGS. A total of 26,937,978 reads were produced for CDR36, assembled in 6,693 contigs with a minimum and maximum size of 337 and 5,148bp, respectively, and an average GC content of 42.1%. The N50 for this chromosome was 337 (Table 7). For the *minute* chromosome, 5,745,778 reads were produced, distributed in 1,251 contigs with a minimum and maximum size of 168 and 5,722 respectively. The GC content was slightly higher than that of CDR36, at 44.9%, and the N50 was 349 (Table 8). Analysis of the sequences by BLAST did not detect any annotated protein-coding genes, though both chromosomes aligned well with the 5 alpaca sequence scaffolds that mapped to chr36 (Figure 13). These findings corroborate and further validate the FISH results, confirming the presence of these previously unassigned genome sequence scaffolds on camelid chromosome 36.

Table 7. Summary of the CDR36 assembly statistics.

Total contig bases	1,855,867,775 (1.85Gb)
Total number of reads	26,937,978
Number of contigs	6,700
Minimum contig size	337
Maximum contig size	5,148
N50 contig length	337
Average GC content	42.10%

Table 8. Summary of the *minute* assembly statistics.

Total contig bases	396,127,198 (0.4Gb)
Total number of reads	5,745,778
Number of contigs	1,251
Minimum contig size	168
Maximum contig size	5,722
N50 contig length	359
Average GC content	44.90%

Comparative Map of HSA7 in Camelids

Cytogenetic mapping of *ZPBP* and *VWC2* to LPA/CDR36 not only assigned the first genes to camelid chr36, but also added a new segment to the previously known homology of HSA7 with CDR7 and CDR18 (Balmus et. al.,

2007). In order to precisely demarcate evolutionary breakpoints and refine the comparative organization of HSA7 in camelid genomes, another 16 genes (Table 6) were cytogenetically mapped in alpaca and dromedary camel. Of these, 6 mapped to LPA/CDR7 and were ordered as: ptel-*ZNF853*-cen-*TMED4*-*H2AFV*-*MAGI2*-*HBP1*-*TRBV30*-qtel (Figure 14a; see figure on page 106 and Discussion), corroborating and refining the previously established homology between camelid chromosome 7 and HSA7pter-p13 and HSA7q11.2-qter (Balmus *et. al.*, 2007). The other 10 genes (*PSPH*, *CCT6A, SUMF2, GUSB, CRCP, TYW1, AUTS2, WBSCR17, BCL7B,* and *MDH2*) mapped to LPA/CDR18q15-q16 (Figure 14b, Figure 15), confirming and refining the homology between LPA/CDR18q15-q16 and HSA7q11.2 (Figure 14b) but not with HSA7p12 (Balmus *et. al.,* 2007). The latter, as shown in this study, corresponds to LPA/CDR36 (Figure 13; see figure on page 103). Due to their close distance, LPA/CDR18 markers were ordered by interphase FISH (Figure 14c). Importantly, refined comparative mapping of HSA7 genes in camelid genomes precisely demarcated 3 evolutionary breakpoints. Two of those are located on HSA7p – one between *VWC2* and *H2AFV,* and another between Scaffold549 and *PSPH*; and one on HSA7q, between *MDH2* and *MAGI2* (see figures on pages 103 and 106).

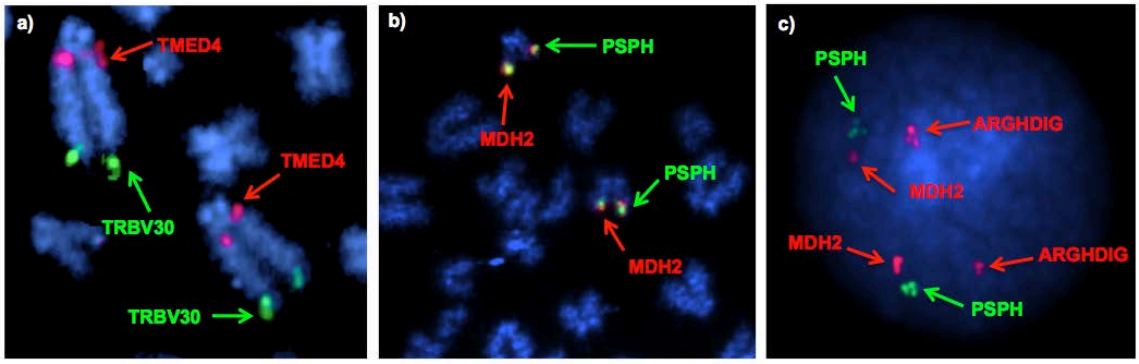

Figure 14. a) *TMED4* (red signals) and *TRBV30* (green signals) map to LPA/CDR7; b) *PSPH* (green) and *MDH2* (red) map to LPA/CDR18; c) Representative interphase FISH image showing the relative order of *ARGHDIG* (red), *PSPH* (green) and *MDH2* (red) in LPA/CDR18.

The Origin of the Minute Chromosome

The 7 molecular markers assigned to LPA36 were also used to investigate the *minute* chromosome. As a first step, the markers were FISH mapped to metaphase spreads of *minute* carriers, and confirmed that the *minute* is indeed an abnormally small LPA36 (Figure 15). As all 7 markers mapped to both the *minute* and the normal size LPA36, the results refuted, or at least did not confirm, the hypothesis that the *minute* is the result of a deletion or a translocation (Avila *et. al.,* 2012) (Figure 15).

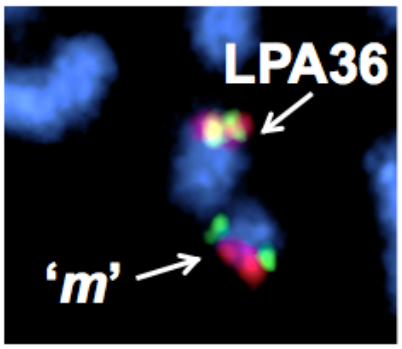

Figure 15. Representantive dual-color FISH image showing the co-localization of *ZPBP* (red) and Scaffold 549 (green) on the *minute* (*'m'*) chromosome.

Next, in order to test the hypothesis that the *minute* represents a naturally occurring chr36 size polymorphism in normal female alpacas (and that its association with infertility phenotypes is anecdotal), large cohort karyotyping was performed in 100 females with proven fertility records. Among those, the *minute* was identified in only one individual, ALP81 (Table 9), showing that its frequency in the normal alpaca population (1%) is negligible.

Finally, as CDR36 carries an NOR (Balmus et. al., 2007), we theorized that the *minute* might originate from an NOR size polymorphism in this pair of chromosomes. To test this hypothesis, we hybridized a probe containing human 5.8S, 18S and 28S rDNA to 6 infertile *minute* carriers and to 3 normal female alpacas with proven fertility records, including LPA81 with a *minute* chromosome (Table 9). Notably, NOR signals were not observed on LPA36 in any of the normal individuals, including the fertile ALP81 control carrying a *minute* (Table 9). This was in contrast to what was observed in dromedary camels, in which

CDR36 carries rDNA on both homologs (Balmus *et. al.,* 2007). In the case of infertile females with the *minute*, NOR signals were consistently observed in only the normal LPA36 and not in the *minute*, suggesting that the NOR was the cause of the size difference between the homologs, but failing to explain its association with infertility phenotypes. Intriguingly, these results show that it is the larger homologue of chr36, and not the *minute,* that is different between fertile and infertile alpacas.

Table 9. Results of NOR-FISH in infertile and control alpacas.

Alpaca ID	Karyotype*	NORs in LPA36 homologs	Phenotype
ALP9	74,XX*m*	NOR/neg	Infertile female
ALP14	74,XX*m*	NOR/neg	Suspected hermaphrodite
ALP25	74,XX*m*	NOR/neg	Ambiguous genitalia
ALP28	74,XX*m*	NOR/neg	n/a
ALP68	74,XX*m*	NOR/neg	Abnormal uterus and ovaries
ALP73	74,XX*m*	NOR/neg	Failure to breed
ALP78	74,XX	Neg/neg	Normal female, 1 cria
ALP81	74,XX*m*	Neg/neg	Normal female, 5 crias
ALP109	74,XX	Neg/neg	Normal female, 2 crias

*indicates the presence of a *minute* chromosome.

DISCUSSION

This study reports the development and mapping of the first DNA markers to the smallest autosome in camelids, chromosome 36, and the generation of a cytogenetic map comprising 7 ordered markers for LPA/CDR36 (Figure 13d).

These findings filled an important gap in the first cytogenetic map for the alpaca (Avila *et. al.,* 2012) and added molecular cytogenetics tools for comparative, evolutionary and clinical studies in camelids. Importantly, assignment of protein-coding genes to LPA/CDR36 'painted' the last 'white area' in the human-dromedary camel Zoo-FISH map (Balmus *et. al.,* 2007), and showed that the smallest camelid autosome is homologous to HSA7, revealing a new homologous synteny block (HSB) between HSA7p12.1-p11.2 and LPA/CDR36. This segment of HSA7 had been erroneously reported as being homologous to CDR18 (Balmus *et. al.,* 2007). This might be due to the relatively low resolution (~5Mb, Scherthan *et. al.,* 1994) and rather broad demarcation of evolutionary breakpoints by Zoo-FISH, particularly since the segment of homology between HSA7p and LPA/CDR36 was estimated to be around 6 Mb based on the human sequence map (Table 6). This demonstrates the importance of cytogenetic maps to verify and refine Zoo-FISH data (Avila *et. al.*, 2012; Raudsepp & Chowdhary, 2013; Raudsepp, 2014).

Furthermore, FISH mapping of 5 previously unassigned alpaca genome sequence scaffolds to chr36 demonstrated the use of molecular cytogenetics in improving genome sequence assemblies (Das *et. al.*, 2012; Raudsepp & Chowdhary, 2011), especially those generated by NGS platforms, as is the case of the alpaca genome (*VicPac2.0.1*, http://www.ncbi.nlm.nih.gov/genome/905). BLAST analysis of the 5 scaffolds against mammalian genomes showed the presence of other protein-coding genes orthologous to HSA7p12.2-p11.2. These

included *FIGNL1* (7:50.1Mb), *COBL* (7:51Mb), *IKZF1* (7:50.3Mb), *GRB10* (7:50.6Mb), *DDC* (7:50.5Mb), among others. The findings corroborate the FISH results and confirm the LPA36-HSA7 homology, but also indicate that chr36 may contain more protein-coding genes than those mapped in this study. The presence of these and additional genes will be tested by PCR with gene-specific primers on all BAC clones corresponding to the 5 scaffolds (Table 6) and genes.

Further, since all the 5 sequence scaffolds orthologous to the aforementioned genes mapped to LPA36q, it is likely that the rest of the chromosome is gene-poor or heterochromatic as previously indicated by C-banding (Avila *et. al.,* 2012). This is in agreement with the results of NGS of CDR36 and the *minute*, in which the relatively low contig N50 values might indicate the small size and predominantly heterochromatic nature of this chromosome (Balmus *et. al.,* 2007; Avila *et. al.,* 2012). An additional factor affecting N50 and the overrepresentation of heterochromatic sequences in the assembly was probably the method used for DNA amplification prior to sequencing. In order to obtain enough DNA for the library preparation (see Materials and Methods), both flow-sorted CDR36 and microdissected *minute* chromosomes were subjected to DOP-PCR amplification (Telenius *et. al.,* 1992; Rens *et. al.*, 2006), which inherently favors the amplification of repetitive over unique DNA sequences. The heterochromatic nature of chr36 might also explain the relatively low GC content calculated for CDR36 and the *minute* (42,1% and 44.9%, respectively; Table 7, Table 8). These values are comparable to the

average GC content found in gene-poor human chromosomes. For a comparison, the most gene-rich human chromosome, HSA19, has an average GC content of 48.5% (Bernardi *et. al.*, 1985; Constantini *et. al.*, 2006). Thus, cumulatively, the largely heterochromatic nature of chr36, the bias of DOP-PCR and the short reads of NGS likely affected the quality of the *de novo* assembly of these chromosomes. Despite this, the assembled sequences of flow sorted CDR36 and microdissected *minute* chromosomes showed significant similarity by BLAST to the 5 alpaca sequence scaffolds and consequently, the homology with HSA7.

With mapped markers and corresponding BAC clones available for chr36 (Table 6), a solution for improving the sequence assembly could be re-sequencing a BAC tiling path using long-read sequencing technology, such as the Pacific Biosciences (PacBio) single molecule real time (SMRT) sequencing. Efficiency of this approach for resolving problems in the assembly of 'difficult' sequences was recently shown for copy number variable and segmentally duplicated regions in the human genome (Huddleston *et. al.,* 2014).

An important outcome of this study was the discovery of the homology between LPA/CDR36 and a segment in HSA7p. This allowed for the comparison of camelid chr36 with the genomes of other 7 main domestic and model mammalian species, as well as the chicken (Figure 16; Table 6) using the available sequence map information from NCBI (http://www.ncbi.nlm.nih.gov/), Ensembl (http://www.ensembl.org/index.html), and UCSC (http://genome.ucsc.

edu/) Genome Browsers. It appears that the HSA7:49.8-55.5 Mb homologous segment is a separate chromosome only in camelids, whereas in all other species (including the chicken), these sequences are part of a larger chromosome, where they share conserved synteny with other segments of HSA7 (Figure 16). This means that genes and sequences located in LPA/CDR36 are exposed to a different environment in meiosis, chromatin remodeling and gene interactions, compared to the homologous segments in other species. Significance of these observations will be tested in the future, when the knowledge about the functions and mutations of LPA/CDR36 genes improves.

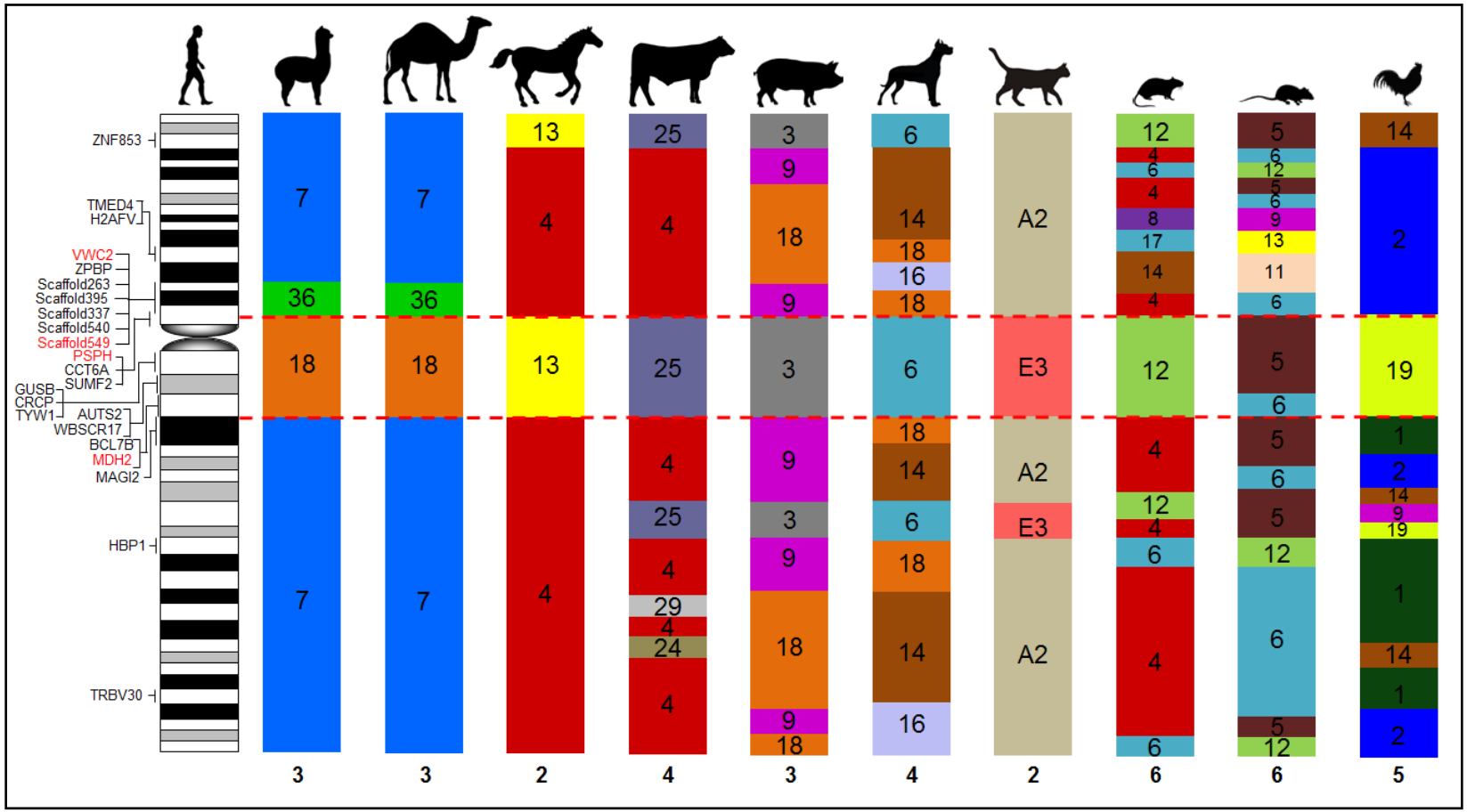

Figure 16. Homologous synteny blocks among HSA7 and 10 other vertebrate species (namely horse, cow, pig, dog, cat, rat, mouse, and chicken). Markers to the left of HSA7 were mapped in camelids; markers in red demarcate evolutionary breakpoints. The stippled red lines indicate breaks present in all analyzed vertebrate genomes. Numbers below each species correspond to the number of HSBs corresponding to HSA7.

Comparative mapping of 18 HSA7 genes in the alpaca and dromedary camel refined the evolutionarily conserved segmental homology between HSA7 and its camelid counterparts (Figure 17), as well as comparison with other mammalian and vertebrate species. Alignment of HSA7 with its counterparts in 10 species (Figure 16) showed that markers *VWC2* and Scaffold 549, located at the terminal ends of the HSB corresponding to LPA/CDR36, indeed demarcated ancestral evolutionary breakpoint regions (EBRs) (Murphy *et. al.,* 2005; Kemkemer *et. al.*, 2009; Larkin *et. al.*, 2009). The EBR defined by *VWC2* (HSA7:49.8Mb) is present in the alpaca, the dromedary camel, the pig, and the dog (Figure 16), thus being a "reuse" EBR. By definition, "reuse" breakpoints are those present in the same genomic intervals in at least three species from two different clades (Murphy *et. al.,* 2005), and correspond to approximately 7% of all EBRs found in mammals (Kemkemer *et. al.*, 2009; Larkin *et. al.*, 2009). The EBR demarcated by Scaffold 549 (HSA7:55.5Mb, Table 7), on the other hand, was found in all species analyzed, except the human. Thus, mapping Scaffold 549 to camelid chr36 in this study confirms that the region is a conserved EBR in vertebrates and that the configuration of HSA7 is specific to humans or primates only (*Kemkemer et. al.,* 2009). Also, our findings in camelids confirm another known vertebrate EBR located between *MDH2* and *MAGI2* (HSA7:75.9) (Kemkemer *et. al.*, 2009) in this lineage (Figure 16, Figure 17; Table 7). Comparison of the number of HSBs corresponding to HSA7 in different species showed that camelids - with 3 segments - closely resemble the pig (also with 3

segments), as opposed to 2 segments in horses and cats, 4 segments in cattle and dogs, 5 in chicken and 6 in rodents (Figure 16). However, the observed differences in intrachromosomal rearrangements within the HSBs in camels and pigs might be true, or due to the more advanced map information available for pigs, and thus needs further testing.

Because HSA7 is not an ancestral chromosome, it would be more appropriate to view the organization of camelid counterparts in relation to the putative ancestral eutherian karyotype (AEK) (Murphy *et. al.*, 2005, Svartman *et. al.* 2006, Ferguson-Smith & Trifonov 2007). The homology segment between HSA7 and camelid chromosome 18 (Figure 16, Figure 17) corresponds to the ancestral eutherian chromosome (AEC) 20 (Ferguson-Smith & Trifonov, 2007), comprised of HSA16p and a part of HSA7. Notably, this ancestral configuration has been retained in the majority of eutherian species studied so far, including camelids (Balmus *et. al.,* 2007). The few exceptions include humans and a few other primates, giant panda and tree shrew (Ferguson-Smith & Trifonov, 2007). The remaining of HSA7 corresponds to the entire AEC9 (Ferguson-Smith & Trifonov, 2007), which has undergone more rearrangements in different lineages (Figure 16). The correspondence of AEC9 to LPA/CDR7 and 36 was revealed in this study. However, the extent of intrachromosomal rearrangements in camelids can be determined only when detailed gene and sequence maps will be available for these chromosomes.

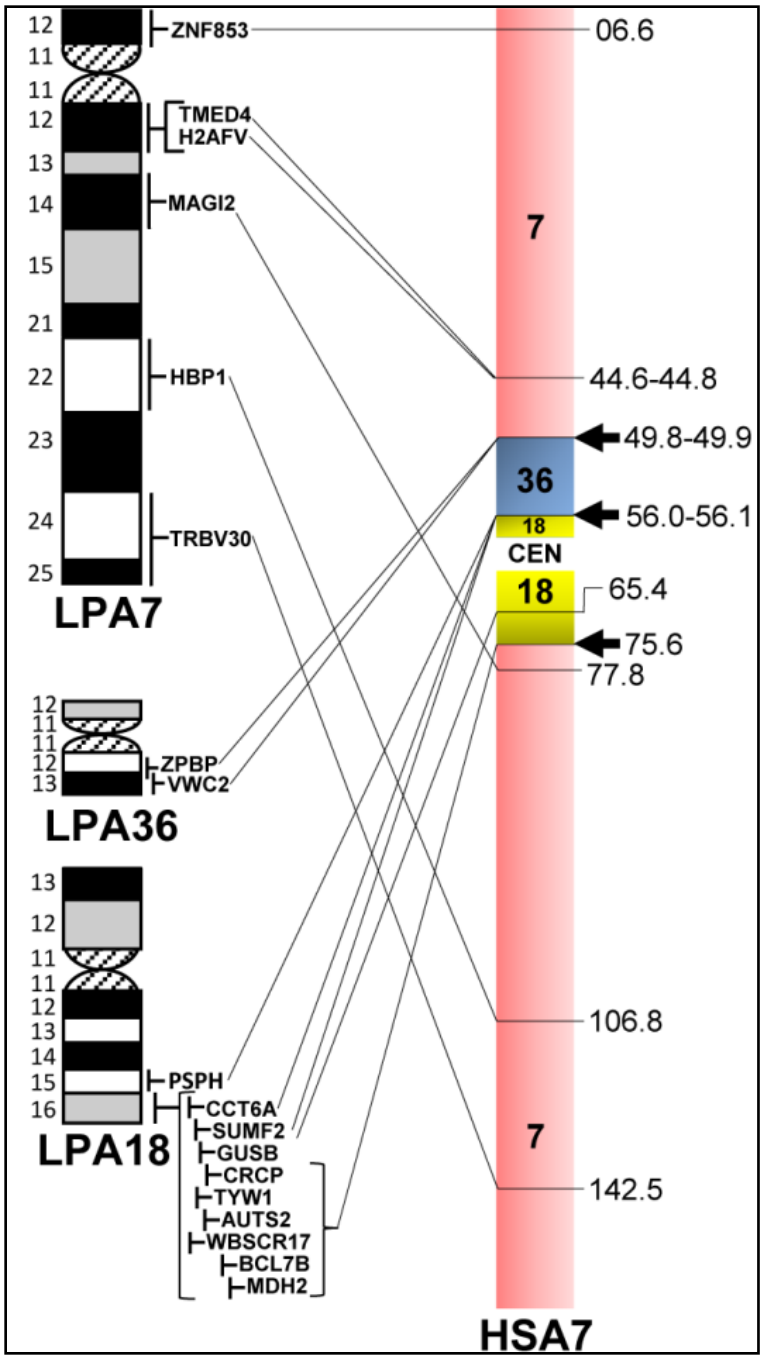

Figure 17. Schematic showing the FISH mapping results in alpaca chromosomes 7, 18 and 36 (left) and their approximate location in HSA7 in Mb (right). Arrows show the location of evolutionary breakpoints.

The development and mapping of markers to LPA36, as well as sequencing of a microdissected *minute* chromosome, helped to further explore the molecular origins of the *minute* chromosome in alpacas. We demonstrated that the *minute* is a derivative of LPA36 and that it is most likely not the result of a deletion or translocation, because all LPA36 markers also mapped to the *minute* (Figure 15). Also, by analyzing the sequences of CDR36 and the *minute* we concluded that the presence of additional protein-coding genes on the short arm of chr36 is unlikely. Next, we tested a possibility that the *minute* is a natural size polymorphism of LPA36 and should be present also in normal alpaca populations. However, cytogenetic analysis of 100 normal fertile individuals identified one female (1%) with a *minute*, in contrast to approximately 15% among infertile individuals (Avila *et. al.,* 2012), which indicates that the association of this chromosome with infertility phenotypes still holds true.

Finally, since chr36 is an NOR chromosome in the dromedary camel (Balmus *et. al.* 2007), we theorized that the size difference between chr36 homologs is due to copy number variation of rRNA genes. This was true for infertile alpacas where the NOR was consistently found in the larger homolog of chr36 but not in the *minute*. In contrast, NORs were not detected in chr36 of fertile females, suggesting that this is the normal status for alpacas. Significance of the chr36 NOR in infertile alpacas at this point remains enigmatic and requires further investigation. We hypothesize that the NOR polymorphism found in chr36 in affected individuals might be affecting the function of nearby reproduction-

related genes, or affecting the proper pairing between chr36 and the *minute* during female meiosis, thus leading to infertility/subfertility phenotypes. Thus, there might be a need to rename the *minute* as a *jumbo-chromosome*. In summary, the cytogenetic map for LPA/CDR36 generated in this study filled a small, but important gap in the whole genome and comparative map of the alpaca, and added to the existing tools for molecular and clinical cytogenetics in camelids.

CHAPTER IV

A 230-MARKER INTEGRATED CYTOGENETIC MAP FOR THE ALPACA GENOME

INTRODUCTION

Alpaca (*Lama pacos*, LPA) is one of the six extant species of camelids (family Camelidae, order Cetartidoactyla) (Stanley *et. al.*, 1994). These animals were domesticated in the Andes region of South America around 7,000 years ago (Wheeler, 1995; Kadwell *et. al.*, 2001), almost as early as cattle (Taberlet *et. al.,* 2011), goats (Schibler *et. al.*, 2009), and horses (Groeneveld *et. al.*, 2010). According to the FAO, the current world population of alpacas is estimated to be around 3.5 million (http://www.fao.org/docrep/004/X6500E/X6500E21.htm). These South American camelids are used both as production (for their meat and fiber) and companion animals. Despite their economic importance, analysis of the alpaca genome, including physical mapping and molecular cytogenetics, has progressed slowly (Avila *et. al.*, 2012; Raudsepp, 2014).

Cytogenetic studies in alpacas and other camelids date back to late 1960s, when the correct diploid number 2n=74 was established for all species (Hsu & Benirschke, 1967; Taylor *et. al.*, 1968; Koulischer *et. al.*, 1971; Hsu & Benirschke, 1974). It has been shown that despite distinct adaptations and

physiological and anatomical differences between the species, the karyotypes of all camelids have remained remarkably conserved (Bunch *et. al.*, 1985; Bianchi *et. al.*, 1986). The studies also showed that, due to the high diploid chromosome number, similar morphology and banding patterns between different chromosome pairs, and heterochromatin size polymorphism between homologous chromosomes, identification of individual chromosomes in alpacas and other camelids is challenging (Bunch *et. al.*, 1985; Bianchi *et. al.*, 1986; Avila *et. al.*, 2012; Raudsepp, 2014).

In order to provide molecular markers for chromosome identification, recently the first low resolution cytogenetic map was developed for the alpaca (Avila *et. al.,* 2012). The map comprises 44 gene-specific markers, distributed on 31 autosomes and the sex chromosomes; no markers were mapped to 5 autosomes. Despite the limited genome coverage, this collection of molecular markers has already been used for clinical cytogenetic analyses in alpacas and llamas (Avila *et. al.,* 2012) and has potential applications for all other camelid species (reviewed by Raudsepp, 2014).

Other recent developments in alpaca genomics include a 22X genome sequence assembly (*VicPac2.0.1* - http://www.ncbi.nlm.nih.gov/genome/ 905), a second-generation RH map composed of 4,590 markers (Perelman, 2011), and a comparative Zoo-FISH map revealing 48 evolutionary conserved synteny blocks between the dromedary camel and humans (Balmus *et. al.* 2007). Despite this, a number of sequence scaffolds and RH groups remain

chromosomally unassigned, and there is no common platform that integrates the information from various maps and aligns it with physical chromosomes. To achieve this, the current alpaca cytogenetic map (Avila *et. al.,* 2012) needs improvement by systematically mapping more markers selected from RH groups and sequence scaffolds. A cytogenetic map with evenly spaced markers on all chromosomes will also benefit clinical cytogenetics by providing better tools for chromosome identification and refined analysis of chromosomal aberrations.

The goal of this study is to generate an integrated cytogenetic map for the alpaca genome by assigning systematically selected genes and DNA markers from the RH, sequence and comparative maps to chromosomes by fluorescence *in situ* hybridization (FISH). The same markers are mapped to the dromedary camel to generate a comparative chromosome map between the two camelid species.

MATERIALS AND METHODS

Animals and Chromosome Preparations

Alpaca and dromedary camel chromosome preparations and fixed cell suspensions were obtained from the depository of the Molecular Cytogenetics and Genomics Laboratory at Texas A&M University. The collection comprises samples from 130 alpacas, 8 llamas and 2 dromedary camels, all cytogenetically characterized and cataloged before being stored at -20˚C.

Metaphase and interphase chromosome preparations were prepared from either peripheral blood lymphocytes or fibroblast cell cultures, according to standard protocols (Raudsepp & Chowdhary, 2008; Avila *et. al.*, 2012). In the case of camelid blood lymphocyte cultures, optimal results were obtained using Concanavalin A (Con A from *Canavalia ensiformis*, Sigma-Aldrich) as the mitogen. Cell suspensions were dropped on clean, wet glass slides and checked under phase contrast microscopy (x100) for quality and density. All cell suspensions used in this study were obtained from karyotypically normal individuals.

Marker Selection

In order to achieve the desired coverage across the genome, *i.e.,* at least two markers per chromosome, with markers assigned to all homologous segments between alpaca and human chromosomes (Balmus *et. al.*, 2007), as well as to properly integrate the RH and genome sequence maps with physical chromosomes, three different approaches were used to identify molecular markers for this study. In the first approach, protein-coding genes were selected from segments of the human genome that are homologous to individual alpaca chromosomes, or specific chromosome regions, based on comparative information (Balmus *et. al.,* 2007). Using this approach, human orthologs of 131 alpaca genes were identified in the National Center for Biotechnology Information (NCBI) Human Genome MapViewer (http://www.ncbi.nlm.nih.gov/

projects/genome/guide/human/). The alpaca genomic sequence for each gene was retrieved from the Ensembl Genome Browser (http://www.ensembl.org/index.html), and masked for repeats using the RepeatMasker software (http://www.repeatmasker.org/). Whenever possible, human genes were selected according to their putative association with reproduction, development, or other economically important traits or diseases in alpacas.

In the second approach, 91 markers were selected from the first- and second-generation alpaca RH maps (Perelman, 2011). These markers included alpaca complementary DNA (cDNA) sequences (generated by L. Wachter and kindly provided by J. Pontius and W.E. Johnson, unpublished data), SNPs from a custom Illumina alpaca 767 SNP chip and the non-homologous Bovine SNP50 v2 DNA Analysis BeadChip (Illumina), as well as oligos from a custom Illumina GoldenGate assay - all genotyped on the alpaca RH panel (Perelman, 2011; P. Perelman, unpublished data). These markers were used to anchor RH groups to physical chromosomes and to ascertain their order and orientation within each chromosome.

Lastly, molecular markers were generated from 8 previously unassigned genome sequence scaffolds (ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Eukaryotes/vertebrates_mammals/Vicugna_pacos/Vicugna_pacos-2.0.1/Primary_Assembly/unplaced_scaffolds/FASTA/), in order to integrate the cytogenetic map with the alpaca genome sequence. These large scaffolds,

ranging in size from 180kb to 1.9Mb, had not been assigned due to the lack RH markers and/or human homology (P. Perelman, personal communication).

Primer Design

Polymerase chain reaction (PCR) primers were designed for each marker using the Primer3 software (Rozen & Skaletsky, 1998), and overgo primers were designed in or around the PCR amplicons, according to our protocols (Avila *et. al.*, 2012). In the case of protein-coding genes, exonic sequences were selected for primer design, and 3 pairs of primers were designed for each sequence scaffold. Details of all selected markers, as well as the PCR and overgo primers sequences, are presented in Tables 10 and 11, respectively.

Table 10. List of markers used in this study (listed by their position on alpaca chromosomes), with their respective locations in alpaca (LPA) and human (HSA) chromosomes and in the human sequence map, and the BAC clone used for mapping. RH markers are indicated by their respective RH panel IDs, followed by their names in parenthesis. Markers indicated with an asterisk (*) were also mapped on dromedary camel chromosomes.

LPA Chr	Marker symbol	Gene symbol	Marker name	LPA Cytogenetic location	HSA Cytogenetic location	Human Mb	BAC mapped
1	GG_21	LOC102545896	Collagen alpha-4(VI) chain-like	1q12	3p25.1	15.2	274022
1	GG_796	n/a	*Homo sapiens* 3 BAC RP11-63K8	1q15-q16	3q26.1	165.4	150E01
1	n/a	*SOX2*	SRY (sex determining region Y)-box 2	1q18-q21	3q26.33	181.3	024K02
1	GG_1124	n/a	*Homo sapiens* 3 BAC RP11-307B12	1q21-q22	3q27.1	184.3	106N02
1	n/a	*ITGB5*	Integrin, beta 5	1q23	3q21.2	124.4	015K02
1	Lgnuc737	*DYRK1A*	Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A	1q23-q25	21q22.13	38.7	118E13
1	GG_410	n/a	*Homo sapiens* genomic DNA, chromosome 21q22.1, clone B382J1, SOD-AML region	1q33	21q22.11	31.9	071E14
2	n/a	*TLR2*	Toll-like receptor 2	2q13-q14	4q31.3	154.6	108K09
2	GG_1365	*NR3C2*	Nuclear receptor subfamily 3, group C, member 2	2q14	4q31.23	149.2	084B11
2	n/a	*FGF5*	Fibroblast growth factor 5	2q22	4q21.21	81.1	208O21
2	GG_326	n/a	*Homo sapiens* BAC clone RP11-176K15	2q23	4q21.23	85	114G20
2	n/a	*KIT*	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	2q23-q24	4q12	55.5	078N15
2	Lgnuc134	n/a	n/a	2q33	n/a	n/a	090N13
2	GG_250	n/a	*Homo sapiens* BAC clone RP11-673J15	2q33	4p15.31	18.5	050K11
3	GG_733	n/a	*Homo sapiens* chromosome 5 clone CTB-47B8	3q12	5q33.3	157.4	106P22
3	n/a	*SLC36A1*	Solute carrier family 36 (proton/amino acid symporter), member 1	3q12	5q33.1	150.8	080N08
3	n/a	*HMGXB3*	HMG box domain containing 3	3q12	5q32	149.3	207L08
3	GG_1378	n/a	*Homo sapiens* chromosome 5 clone CTC-352D11	3q22-q23	5q12.3	66.7	104I15
3	Lgnuc233	*C9*	Complement component 9	3q31-q32	5p13.1	39.2	063J01

Table 10 continued

LPA Chr	Marker symbol	Gene symbol	Marker name	LPA Cytogenetic location	HSA Cytogenetic location	Human Mb	BAC mapped
3	n/a	*SLC45A2*	Solute carrier family 45, member 2	3q32-q33	5p13.2	33.9	220B13
3	n/a	*TRIO*	Trio Rho guanine nucleotide exchange factor	3q34	5p15.2	14.1	043E23
4	n/a	*TYRP1*	Tyrosinase-related protein 1	4q21dist-q22	9p23	12.6	129N17
4	Lgnuc105	*PSAT1*	Phosphoserine aminotransferase 1	4q24-q31	9q21.2	80.9	118I18
4	Lgnuc400	*NCBP1*	Nuclear cap binding protein subunit 1, 80kDa	4q32-q33	9q22.33	100.3	059P19
4	GG_478	n/a	Human DNA sequence from clone RP11-58C3 on chromosome 9	4q34	9q33.1	118.9	071E21
4	GG_1338	n/a	Human DNA sequence from clone RP11-269P11 on chromosome 9	4q34-q35	9q33.3	128.2	135M08
4	GG_965	n/a	Human DNA sequence from clone RP11-336P12 on chromosome 9	4q35	9q33.3	128.7	026N11
4	Lgnuc411	*AGPAT2*	1-acylglycerol-3-phosphate O-acyltransferase 2	4q35-q36	9q34.3	139.5	101I14
4	Lgnuc416	*MRPL41*	Mitochondrial ribosomal protein L41	4q36	9q34.3	140.4	135B22
5	GG_559	n/a	*Homo sapiens* BAC clone RP11-252K7 from 2	5q23-q24	2q24.2	161.1	099P06
5	GG_634	n/a	*Homo sapiens* BAC clone RP11-458D8 from 2	5q25-q31	2q35	216.5	068C12
5	Lgnuc116	*TMEFF2*	Transmembrane protein with EGF-like and two follistatin-like domains 2	5q33-q34	2q32.3	192.8	104O06
5	n/a	*PAX3*	Paired box 3	5q35	2q36.1	223	378C17
5	Lgnuc131	*MYEOV2*	Myeloma overexpressed 2	5q36	2q37.3	241	109E06
6	CSNP_2782	n/a	*Homo sapiens* chromosome 15, clone RP11-625H11	6q12-q14	15q22.2	62.9	142G13
6	GG_398	*DAD1*	Defender against cell death 1	6q17	14q11.2	23	117N19
6	GG_705	n/a	Human chromosome 14 DNA sequence BAC R-785G15 of library RPCI-11	6q22	14q22.2	54.2	013L16
6	Lgnuc582	*C14ORF132*	Chromosome 14 open reading frame 132	6q24-q31	14q32.2	96.5	067N16

Table 10 continued

LPA Chr	Marker symbol	Gene symbol	Marker name	LPA Cytogenetic location	HSA Cytogenetic location	Human Mb	BAC mapped
6	GG_962	n/a	Human chromosome 14 DNA sequence BAC R-362L22 of library RPCI-11	6q31	14q32.2	100.7	136J13
6	Lgnuc584	*CDC42BPB*	CDC42 binding protein kinase beta (DMPK-like)	6q32	14q32.32	103.4	076A14
6	n/a	*IGHV@**	Immunoglobulin heavy locus, variable region	6q32	14q32.33	106	064M07
6	n/a	*IGHA@**	Immunoglobulin heavy locus, constant alpha	6q32	14q32.33	106	132H10
7	Lgnuc330*	*ZNF853*	Zinc finger protein 853	7p12	7p22.1	6.6	074A23
7	n/a	*TMED4**	Transmembrane emp24 protein transport domain containing 4	7q12	7p13	44.6	102E05
7	n/a	*H2AFV**	H2A histone family, member V	7q12	7p13	44.8	041D24
7	GG_1376*	*MAGI2*	Membrane associated guanylate kinase, WW and PDZ domain containing 2	7q14	7q21.11	78.9	049H15
7	n/a	*HBP1**	HMG-box transcription factor 1	7q22	7q22.3	106.8	056N03
7	Lgnuc355*	*TRBV30*	T cell receptor beta variable 30	7q24-q25	7q34	142.5	010K16
8	GG_585	*CD109*	CD109 molecule	8q12	6q13	74.4	020I02
8	GG_175	*SNAP91*	Synaptosomal-associated protein, 91kDa	8q13	6q14.2	84.2	152N09
8	GG_775	*SLC35F1*	Solute carrier family 35, member F1	8q21-q22	6q22.1	118.2	016L07
8	GG_932	n/a	Human DNA sequence from clone RP11-290P3 on chromosome 6	8q23-q24	6q23.3	137.7	162O12
8	Lgnuc618	*C6ORF211*	Chromosome 6 open reading frame 211	8q25	6q25.1	151.7	015P15
8	GG_1388	*PARK2*	Parkin RBR E3 ubiquitin protein ligase	8q26	6q26	162.9	044N05
9	Lgnuc716	*FTL*	Ferritin, light polypeptide	9p14	19q13.33	49.4	006O04
9	Lgnuc345	*CYP2A13*	Cytochrome P450, family 2, subfamily A, polypeptide 13	9p13	19q13.2	41.5	032D09
9	GG_1068	n/a	*Homo sapiens* chromosome 16 clone RP11-700H13	9q15	16q21	58.8	002N23
9	n/a	*CD101*	CD101 molecule	9q23	1p13.1	117.5	053I18
9	n/a	*TGFBR3*	Transforming growth factor, beta receptor III	9q25	1p22.1	92.1	046F06

Table 10 continued

LPA Chr	Marker symbol	Gene symbol	Marker name	LPA Cytogenetic location	HSA Cytogenetic location	Human Mb	BAC mapped
10	n/a	*RAB38*	RAB38, member RAS oncogene family	10q12dist-q14prox	11q14.2	87.8	176P13
10	n/a	*SF1*	Splicing factor 1	9q23	11q13.1	64.5	077C18
10	Lgnuc460	*RAG1*	Recombination activating gene 1	9q25-q26	11p12	36.5	084G11
10	n/a	*WT1*	Wilms tumor 1	9q27	11p13	32.3	032E05
11	n/a	*REN*	Renin	11p13	1q32.1	204.1	061D02
11	CSNP_29273	*TTC13*	Tetratricopeptide repeat domain 13	11p13-p12	1q42.2	231	116H15
11	Scaffold 421	*n/a*	n/a	11q12	n/a	n/a	074K20
11	Lgnuc28	*LDB3*	LIM domain binding 3	11q13	10q23.2	88.4	152A20
11	n/a	*CSTF2T*	Cleavage stimulation factor, 3' pre-RNA, subunit 2, 64kDa, tau variant	11q16	10q21.1	53.4	088E24
11	n/a	*ALDH18A1*	Aldehyde dehydrogenase 18 family, member A1	11q21	10q24.1	97.3	062G15
11	n/a	*TCF7L2*	Transcription factor 7-like 2 (T-cell specific, HMG-box)	11q22	10q25.2	114.7	072C10
11	n/a	*SLC18A2*	Solute carrier family 18 (vesicular monoamine transporter), member 2	11q22	10q25.3	119	375L16
11	n/a	*FGFR2*	Fibroblast growth factor receptor 2	11q23	10q26.13	123.2	364G05
12	GG_844	*BAZ2A*	Bromodomain adjacent to zinc finger domain, 2A	12q13	12q13.3	56.9	004J13
12	GG_51	*IFNG-AS1*	IFNG antisense RNA 1 (IFNG-AS1), transcript variant 2, long non-coding RNA	12q21	12q15	68.3	044K10
12	Lgnuc534	*KITLG*	KIT ligand	12q22	12q24	88.9	113O13
13	n/a	*LHX8*	LIM homeobox 8	13q13	1p31.1	75.6	480O10
13	GG_986	*NEGR1*	Neuronal growth regulator 1	13q14	1p31.3	72.1	038G12
13	Lgnuc30	*DHCR24*	24-dehydrocholesterol reductase	13q21-q22	1p32.3	55.3	094H18
13	Lgnuc27	*SCP2*	Sterol carrier protein 2	13q21-q22	1p32.3	53.4	094K01
13	n/a	*HEYL*	Hes-related family bHLH transcription factor with YRPW motif-like	13q23	1p34.3	40	020D23
13	n/a	*RSPO1*	R-spondin 1	13q23	1p34.3	38	023K10

Table 10 continued

LPA Chr	Marker symbol	Gene symbol	Marker name	LPA Cytogenetic location	HSA Cytogenetic location	Human Mb	BAC mapped
13	CSNP_20238	*KAZN*	Kazrin, periplakin interacting protein	13q24-q25prox	1p36.21	14.9	012O08
13	Lgnuc06	*DHRS3*	Dehydrogenase/reductase (SDR family) member 3	13q24dist-q25	1p36.22	12.6	054P17
14	n/a	*BRCA2*	Breast cancer 2, early onset	14p15	13q12.3	32.9	012E10
14	n/a	*FOXO1*	Forkhead box O1	14p14	13q14.11	41.2	128d05
14	n/a	*RB1*	Retinoblastoma 1	14p13	13q14.2	48.9	089N13
14	GG_558	n/a	Human DNA sequence from clone RP11-14B2	14q13prox	13q21.1-21.3	72.7	039N18
14	n/a	*EDNRB*	Endothelin receptor type B	14q13dist	13q22	78.4	138C23
15	n/a	*TACR1*	Tachykinin receptor 1	15q12	2p12	75.2	029E13
15	Scaffold 411	*n/a*	n/a	15q12	2p13.1	74.6	244L03
15	Lgnuc82	*BRE*	Brain and reproductive organ-expressed (TNFRSF1A modulator)	15q22	2p23.2	28.1	074O15
16	n/a	*ANKFN1**	Ankyrin-repeat and fibronectin type III domain containing 1	16p14dist	17q22	54.2	116C14
16	Lgnuc653*	*DDX52*	DEAD box polypeptide 52	16p14prox	17q12	35.9	018J07
16	n/a	*LHX1**	LIM homeobox 1	16p13	17q12	35.2	010N01
16	Lgnuc652*	*CCL16*	Chemokine (C-C motif) ligand 16	16p13	17q12	34.3	159H07
16	n/a	*AP2B1**	Adaptor-related protein complex 2, beta 1 subunit	16p13	17q12	33.9	156N10
16	n/a	*HS3ST3A1**	Heparan sulfate (glucosamine) 3-O-sulfotransferase 3A1	16p12	17p12	13.3	080J08
16	n/a	*IKZF3**	IKAROS family zinc finger 3 (Aiolos)	16q12	17q12	37.9	135E24
16	n/a	*VIM**	Vimentin	16q13	10p13	17.2	026J07
16	n/a	*NF1**	Neurofibromin 1	16q14prox	17q11.2	29.4	429D06
16	Lgnuc651*	*CDK5R1*	Cyclin-dependent kinase 5, regulatory subunit 1 (p35)	16q14dist	17q11.2	30.8	114G04
16	n/a	*KCNJ16**	Potassium inwardly-rectifying channel, subfamily J, member 16	16q16	17q24.3	68	408P06
16	n/a	*JMJD6**	Jumonji domain containing 6	16q17	17q25.1	74.7	296H06

Table 10 continued

LPA Chr	Marker symbol	Gene symbol	Marker name	LPA Cytogenetic location	HSA Cytogenetic location	Human Mb	BAC mapped
17	n/a	*SLC22A13*	Solute carrier family 22 (organic anion/urate transporter), member 13	17q12	3p22.2	38.3	035N16
17	GG_1414	n/a	*Homo sapiens* chromosome 3 clone RP11-413B21, complete sequence	17q13	3p21.2	51.6	021E03
17	n/a	*MITF*	Microphthalmia-associated transcription factor	17q14	3p13	69.9	033H02
17	GG_1079	n/a	*Homo sapiens* chromosome 3 clone RP11-140B10 map 3p, complete sequence	17q15	3p26.1	6.3	010N02
18	n/a	*RBFOX1**	RNA binding protein, fox-1 homolog (*C. elegans*)	18q12	16p13.3	7.3	118G19
18	Lgnuc612	*ARHGDIG**	Rho GDP dissociation inhibitor (GDI) gamma	18q13	16p13.3	0.33	013L12
18	n/a	*PSPH**	Phosphoserine phosphatase	18q15	7p11.2	56.1	018I02
18	n/a	*CCT6A**	Chaperonin containing TCP1, subunit 6A (zeta 1)	18q16	7p11.2	56.1	432O07
18	n/a	*SUMF2**	Sulfatase modifying factor 2	18q16	7p11.2	56.1	471B09
18	n/a	*GUSB**	Glucuronidase, beta	18q16	7q11.21	65.4	364B06
18	n/a	*CRCP**	CGRP receptor component	18q16	7q11.21	65.5	127N21
18	n/a	*TYW1**	tRNA-yW synthesizing protein 1 homolog (*S. cerevisiae*)	18q16	7q11.21	66.5	050M03
18	n/a	*AUTS2**	Autism susceptibility candidate 2	18q16	7q11.22	70.2	403A12
18	n/a	*WBSCR17**	Williams-Beuren syndrome chromosome region 17	18q16	7q11.22	70.7	422L15
18	n/a	*BCL7B**	B-cell CLL/lymphoma 7B	18q16	7q11.23	72.9	098H24
18	n/a	*MDH2**	Malate dehydrogenase 2, NAD (mitochondrial)	18q16	7q11.23	75.6	003P04
18	n/a	*ZCWPW1*	Zinc finger, CW type with PWWP domain 1	18q16	7q22.1	100	272O23
19	n/a	*ASIP*	Agouti signaling protein	19q12	20q11.2-q12	32.8	018C13
19	n/a	*BMP7*	Bone morphogenetic protein 7	19q22	20q13	55.7	093P06
19	n/a	*EDN3*	Endothelin 3	19q23	20q13.2-q13.3	57.8	125P19

Table 10 continued

LPA Chr	Marker symbol	Gene symbol	Marker name	LPA Cytogenetic location	HSA Cytogenetic location	Human Mb	BAC mapped
20	n/a	*CRISP3*	Cysteine-rich secretory protein 3	20q12	6p12.3	49.6	116A11
20	n/a	*HLA-F**	Major histocompatibility complex, class I, F	20q13	6p22.1	29.6	092P17
20	n/a	*HLA-G**	Major histocompatibility complex, class I, G	20q13	6p22.1	29.7	084M12
20	n/a	*PRR3*	Proline rich 3	20q13	6p21.33	30.5	003A09
20	n/a	*DDR1*	Discoidin domain receptor tyrosine kinase 1	20q13	6p21.3	30.8	003A09
20	n/a	*HLA-B**	Major histocompatibility complex, class I, B	20q13	6p22.1	31.2	092P17
20	n/a	*ATP6V1G2*	ATPase, H+ transporting, lysosomal 13kDa, V1 subunit G2	20q13	6p21.33	31.5	190I20
20	n/a	*CLIC1*	Chloride intracellular channel 1	20q13	6p21.33	31.6	104E06
20	n/a	*NOTCH4*	Notch 4	20q13	6p21.33	32.1	100F14
20	n/a	*C6orf10*	Chromosome 6 open reading frame 10	20q13	6p21.33	32.2	100F14
20	n/a	*COL11A2*	Collagen, type XI, alpha 2	20q13	6p21.33	33.1	060A13
20	n/a	*KIFC1*	Kinesin family member C1	20q13	6p21.33	33.3	100F14
21	n/a	*LAMC1*	Laminin, gamma 1	21q12	1q25.3	182.9	068G03
21	n/a	*MYOC*	Myocilin, trabecular meshwork inducible glucocorticoid response	21q13	1q23-q24	171.6	128F16
21	n/a	*USF1*	Upstream transcription factor 1	21q14	1q22-q23	161	056L05
21	n/a	*ADAR*	Adenosine deaminase, RNA-specific	21q15	1q21.3	154.5	090L23
22	n/a	*CANX*	Calnexin	22q12	5q35	179.6	006G05
22	n/a	*GDF15*	Growth differentiation factor 15	22q14	19p13.11	18.3	021E17
22	n/a	*SAFB*	Scaffold attachment factor B	22q15	19p13.3-p13.2	5.6	024E01
23	n/a	*PIGR*	Polymeric immunoglobulin receptor	23p12	1q32.1	207.1	238K08
23	n/a	*EXO1*	Exonuclease 1	23q13	1q43	242	119A07
23	n/a	*LBR*	Lamin B receptor	23q14prox	1q42.12	225.5	245P05
23	GG_1100	n/a	*Homo sapiens* chromosome 1 clone RP11-410C4	23q14dist	1q41	220.5	026L03
23	Lgnuc63	*LPGAT1*	Lysophosphatidylglycerol acyltransferase 1	23q15	1q32.3	211.9	470A15
23	n/a	*TROVE2*	TROVE domain family, member 2	23q15	1q31.2	193	068I08

Table 10 continued

LPA Chr	Marker symbol	Gene symbol	Marker name	LPA Cytogenetic location	HSA Cytogenetic location	Human Mb	BAC mapped
23	Lgnuc551	*ATP5EP2*	ATP synthase, H+ transporting, mitochondrial F1 complex, epsilon subunit pseudogene 2	23q15	13q12	28.5	212F18
24	Lgnuc409	*TTR*	Transthyretin	24q14	18q12.1	29.1	248O05
24	GG_1360	*GAREM*	GRB2 associated, regulator of MAPK1	24q15	18q12.1	30	036G20
24	GG_1521	*ANKRD12*	Ankyrin repeat domain 12	24q16	18p11.22	9.2	007F14
25	n/a	*ESRP1**	Epithelial splicing regulatory protein 1	25q12	8q22.1	95.6	509I21
25	n/a	*ATP6V1C1**	ATPase, H+ transporting, lysosomal 42kDa, V1 subunit C1	25q13	8q22.3	104	176E21
25	n/a	*ZFPM2**	Zinc finger protein, FOG family member 2	25q13	8q23.1	106.8	020K01
25	n/a	*RSPO2**	R-spondin 2	25q14	8q23.1	108.9	100M08
25	n/a	*DSCC1**	DNA replication and sister chromatid cohesion 1	25q15prox	8q24.2	120.8	071D11
25	n/a	*TRIB1**	Tribbles pseudokinase 1	25q15prox	8q24.3	126.4	428D22
25	n/a	*ASAP1**	ArfGAP with SH3 domain, ankyrin repeat and PH domain 1	25q15mid	8q24.21	131	162C17
25	n/a	*SLA**	Src-like-adaptor	25q15dist	8q24.22	134	027P17
25	n/a	*GRINA**	Glutamate receptor, ionotropic, N-methyl D-aspartate-associated protein 1	25q15ter	8q24.3	145	246A10
26	n/a	*BAG4**	BCL2-associated athanogene 4	26q13	8p11.23	38	229C18
26	GG_866*	n/a	*Homo sapiens* BAC clone RP11-376O6 from 4	26q14	4q34.3	178.2	280B09
27	GG_1438	n/a	*Homo sapiens* chromosome 15, clone RP11-236L14	27q12prox	15q26.2	96.5	033K18
27	GG_723	n/a	*Homo sapiens* chromosome 15, clone RP11-97O12	27q12dist	15q25.3	89	039N07
27	GG_1142	n/a	*Homo sapiens* chromosome 15, clone RP11-331H13	27q13prox	15q24.3	76.8	100J24
27	Lgnuc606	*NPTN*	Neuroplastin	27q13ter	15q24.1	73.8	062D09
28	GG_395	*IL18RAP*	Interleukin 18 receptor accessory protein	28q12	2q12	102.4	010O03

Table 10 continued

LPA Chr	Marker symbol	Gene symbol	Marker name	LPA Cytogenetic location	HSA Cytogenetic location	Human Mb	BAC mapped
28	n/a	*BUB1*	BUB1 mitotic checkpoint serine/threonine kinase	28q13	2q14	110.6	080G05
28	n/a	*MAL*	Mal, T-cell differentiation protein	28q14	2q11.1	95	014J18
28	n/a	*CAPG*	Capping protein (actin filament), gelsolin-like	28q15	2p11.2	85.3	001J16
28	n/a	*IGKV@**	Immunoglobulin kappa variable cluster	28q15	2p12	75	023D20
28	n/a	*IGKJ@**	Immunoglobulin kappa joining cluster	28q15	2p12	75	265O04
29	n/a	*RALYL**	RALY RNA binding protein-like	29q13	8q21.2	85	224G07
29	n/a	*ZFAND1**	Zinc finger, AN1-type domain 1	29q14	8q21.13	82.6	194O05
29	n/a	*RDH10**	Retinol dehydrogenase 10 (all-trans)	29q15	8q21.11	74.2	215C02
29	n/a	*RP1**	Retinitis pigmentosa 1	29q16	8q12.1	55.5	178K02
29	n/a	*RB1CC1**	RB1-inducible coiled-coil 1	29q16	8q11.23	53.5	084G03
30	Lgnuc683	*CYB5A*	Cytochrome b5 type A	30q12	18q22.3	71.9	215N04
30	Lgnuc682	n/a	*Homo sapiens* chromosome 18, clone RP11-1151B14	30q13	18q21.31	56.1	062D09
30	Lgnuc681	n/a	*Homo sapiens* chromosome 18, clone RP11-729L2	30q14	18q21.2	48.4	104K07
30	n/a	*IER3IP1*	Immediate early response 3 interacting protein 1	30q14	18q21.1	44.6	497E10
31	n/a	*RP1L1**	Retinitis pigmentosa 1-like 1	31q13prox	8p23.1	10.4	066P16
31	n/a	*C8ORF74**	Chromosome 8 open reading frame 74	31q13prox	8p23.1	10.5	028M08
31	n/a	*FDFT1**	Farnesyl-diphosphate farnesyltransferase 1	31q13prox	8p23.1	11.6	166A12
31	Lgnuc358*	*LGI3*	Leucine-rich repeat LGI family, member 3	31q13dist	8p21.3	22	118G23
31	n/a	*STC1**	Stanniocalcin 1	31q13-q14	8p21.2	23.7	522L05
31	Lgnuc360*	*ADRA1A*	Adrenoceptor alpha 1A	31q14prox	8p21.2	26.7	169G14
31	n/a	*UBE2K**	Ubiquitin-conjugating enzyme E2K	31q14ter	4p14	39.7	032H16
32	n/a	*IGLV@**	Immunoglobulin lamba variable cluster	32q12	22q11.2	23.2	171G07
32	n/a	*IGLJ@**	Immunoglobulin lambda joining cluster	32q12	22q11.2	23.2	057I04
32	n/a	*IGLC@**	Immunoglobulin lambda constant cluster	32q12	22q11.2	23.2	133M08

Table 10 continued

LPA Chr	Marker symbol	Gene symbol	Marker name	LPA Cytogenetic location	HSA Cytogenetic location	Human Mb	BAC mapped
32	Lgnuc743	*GNB1L*	Guanine nucleotide binding protein (G protein), beta polypeptide 1-like	32q13	22q11.2	19.7	132C06
32	Lgnuc545	*HIP1R*	Huntingtin interacting protein 1 related	32q14	12q24	123.3	203J11
33	GG_516	n/a	*Homo sapiens* chromosome 11 clone RP11-321E15	33q13-q14	11q24.1	123.5	022E07
33	GG_788	*SIK3*	SIK family kinase 3	33q13-q14	11q23.3	116.8	012H04
33	Lgnuc524	*HSD17B12*	Hydroxysteroid (17-beta) dehydrogenase 12	33q13-q14	11p11.2	43.7	190I20
34	n/a	*PKP2*	Plakophilin 2	34p12	12p11.21	32.9	012A14
34	n/a	*SOX5*	SRY (sex determining region Y)-box 5	34q12	12p12.1	23.6	047K13
34	Lgnuc510	*LMO3*	LIM domain only 3 (rhombotin-like 2)	34q13	12p12.3	16.7	058N01
34	n/a	*AICDA*	Activation-induced cytidine deaminase	34q13	12p13.31	8.7	065O12
35	n/a	*CREM*	cAMP responsive element modulator	35q13	10p11.21	35.4	126O20
35	n/a	*BAMBI*	BMP and activin membrane-bound inhibitor	35q14	10p12.3-p11.2	28.6	003B06
35	n/a	*OPTN*	Optineurin	35q14	10p13	13.1	003G02
35	n/a	*UCN3*	Urocortin 3	35q14	10p15.1	5.4	085N19
35	Lgnuc417	*LARP4B*	La ribonucleoprotein domain family, member 4B	35q14	10p15.3	0.8	325C10
36	Scaffold 549*	n/a	n/a	36q12-q13	7p11.2	55.5	037P07
36	Scaffold 540*	n/a	n/a	36q12-q13	7p12.1	53.7	107O12
36	Scaffold 337*	n/a	n/a	36q12-q13	7p12.1	53.5	125E19
36	Scaffold 395*	n/a	n/a	36q12-q13	7p12.1	53.1	251B12
36	Scaffold 263*	n/a	n/a	36q12-q13	7p12.1	51.6	037O03
36	n/a	*ZPBP**	Zona pellucida binding protein	36q12-q13	7p14.3	49.9	003N14

Table 10 continued

LPA Chr	Marker symbol	Gene symbol	Marker name	LPA Cytogenetic location	HSA Cytogenetic location	Human Mb	BAC mapped
36	n/a	*VWC2**	Von Willebrand factor C domain containing 2	36q12-q13	7p12.2	49.7	003N14
X/Y	n/a	*CSF2RA**	Colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)	Xp16	Xp22.33	1.38	021N01
X/Y	n/a	*ARSF**	Arylsulfatase F	Xp16	Xp22.33	2.9	055K02
X/Y	n/a	*STS**	Steroid sulfatase (microsomal), isozyme S	Xp16	Xp22.31	7	119F05
X/Y	n/a	*PNPLA4**	Patatin-like phospholipase domain containing 4	Xp16	Xp22.31	7.8	021D09
X/Y	n/a	*KAL1**	Kallmann syndrome 1 sequence	Xp16	Xp22.31	8.4	010N11
X/Y	n/a	*GPR143**	G protein-coupled receptor 143	Xp16	Xp22.2	9.6	075H21
X/Y	n/a	*SHROOM2**	Shroom family member 2	Xp16	Xp22.2	9.8	075H21
X/Y	n/a	*WWC3**	WWC family member 3	Xp16	Xp22.2	10	098E09
X/Y	n/a	*CLCN4**	Chloride channel, voltage-sensitive 4	Xp16	Xp22.2	10.1	098E09
X/Y	n/a	*MID1**	Midline 1 (Opitz/BBB syndrome)	Xp16	Xp22.2	10.4	156E13
X	Lgnuc760	*TSPAN7*	Tetraspanin 7	Xp13	Xp11.4	38.4	034N20
X	Lgnuc765	*HSD17B10*	Hydroxysteroid (17-beta) dehydrogenase 10	Xp13	Xp11.22	53.4	034A10
X	014E13	n/a	n/a	Xq14	n/a	n/a	014E13
X	Lgnuc166	*PGRMC1*	Progesterone receptor membrane component 1	Xq21	Xq24	118.3	054D17
X	Scaffold 368	n/a	n/a	Xq22	n/a	n/a	041H14
X	Lgnuc610	*ATP6AP1*	ATPase, H+ transporting, lysosomal accessory protein 1	Xq25	Xq28	153.5	012K09
X	n/a	*HUWE1*	HECT, UBA and WWE domain containing 1, E3 ubiquitin protein ligase	Xq25	Xq28	153.6	012K09

Table 11. List of PCR and overgo primers for the markers used in this study.

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
014E13	AATCCCCAAGATTTCAACCC TCCTCTTCCTCACCACCATC	154	CCCCAAGATTTCAACCCCCAGCAC CCTTCTCATCCAGGAAGTGCTGGG
ADAR	GGTGAGTTTCGAGCCATCAT GTGGTCCGCTCTGCTCTATC	178	TGAAGAACCCCATCAGCGGGCTGT AACTGAGCGTATTCTAACAGCCCG
ADRA1A	GTCAGTGGTCTGGACTGGGT CTCAAATCTCCATCCCCAGA	273	GACAACATAGAGGAAGAGGTGCCC CTCTGTCCGCAGTATAGGGCACCT
AGPAT2	GCAGGGACCATCAAGGTAGA AGGGTCCCTCTGTCCATTCT	232	TGAGGGCCTGCTTCTTCCACATAT TCCTGGGCTGTCTTGGATATGTGG
AICDA	ACTGCTCTTCCTCCGCTACA GCCGTGAAGATCCTCAGACT	159	TGGGAGGGGCTGCATGAAAATTCG GTCTGGACAGACGAACCGAATTTT
ALDH18A1	CAGCTCCTCTGCTGAAACG TCAGGGCGAGACTCAAAGAT	126	GAGCCTCTCCACGTCTAAACTGAA ACCGATGGCCAGGCTGTTCAGTTT
*ANKFN1**	GAACAGCCGTGTGAAAATGA ACTCAGCCGGTCAGTTCTTG	184	AACCTCTCCGAGAAGCTGAAGGGG CAAACGAAGAGTGGCTCCCCTTCA
ANKRD12	GCAAGTTTGTAGACCCTGGTT GCTGAAGTATGCAAGTGAGGAA	177	CACCTTGTTGTTACCTTTCTTTGA GCCAGACTGCTGGCAGTCAAAGAA
*AP2B1**	CACTGATGCAGGTGACAGC TCCTCCTCCCAAGAGATCAA	181	ACCAACCTGGAACAGCCTCAGGTT CACCTTGAGAGGGGATAACCTGAG
*ARHGDIG**	AGATGAGGCACTGGATGAGG AGCGCCTGCTTGTACTTGAC	117	ATCCAGCAGCTGGACCCAGACGAC ACTTGACCAGGCTCTCGTCGTCTG
*ARSF**	AAGTGGCATCAGGGCTTAGA CAACCAGAAGTAGCCCAGGA	294	AAGCAGACTCTGGCTCTGTGTGCA TGCCAAGATGACCAGCTGCACACA
*ASAP1**	ATCAAGAAAGAGCGCTCACC GGTGTTTCTGGGAGGCAGT	213	TTCACCAACCAGATCTTCGTCTCC GCGAGTCTGTGCTTGTGGAGACGA
ASIP	ATGTCACCCGCCTCTTCCTA CCACAATAGAGACAGAAGGGAAA	156	AAGGAAGCCTGAGAAGCAACTCCT TCTAACAGGTTCTTGGAGGAGTTG
ATP5EP2	TGGCAGCAACGTAAAAATTG TCATAGGTTGACGATGCCAA	173	CCCCAACTTCAGCTTGAAATGCTG CTTCCCCTCAGAAGGGCAGCATTT
ATP6AP1	TCTTTGGCCTGAGAAGGAAA TAGGAAGGAAACCAGGAGCA	122	GAAAGGACCTCCATTCAGGTTGGG CAGTGATCACACCCAGCCCAACCT
*ATP6V1C1**	TATCAGCTCCAGGGGAGAAA CAGGAATGTTGAACTTGGAGGT	105	GGGGAGAAAACATGTCAGCAAACG CATGCAACTTCTCCCACGTTTGCT

Table 11 continued

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
ATP6V1G2	CGGTATTGCTCCACTTCCAT GAAGAAATGGGTGGCTGTTG	199	CGGTATTGCTCCACTTCCATCTGT AGGCAAAGGAAGAAGCACAGATGG
*AUTS2**	GCACTTAAGCCTCAGGAACG CTGTCTGAGTCCCTTCAGCA	204	AATGGCTTGTCCTTTCACTCCAAG GGCTGAGTCTGCTCTTCTTGGAGT
*BAG4**	GGGACAAATACCGCCTCATA AAGCAGGATATCCGGGAAC	196	ACAAATACCGCCTCATACTCGGGG CAGGTGTGTAATAAGCCCCCGAGT
BAMBI	CACTGCGTGGCTACTGGTTA GAGTGGTTTCGTGCCTGTTT	161	CTCCAGACTCCTTGATCCTCAGAA GAGTGGGGAATTCGTGTTCTGAGG
BAZ2A	CATCCTCCCCACCTTAGACC GCGGAAACAAGAGACCAAGA	167	GCTTCCTTCTGCTTTAAGCTCTTG AGCGGAAACAAGAGACCAAGAGCT
*BCL7B**	AGGTTGCTGAGGAAGAGGAA GGGTGGTGCTAGCTTTCTGA	103	TTCTAGGTTGCTGAGGAAGAGGAA GTGGGGCACCTGAGTCTTCCTCTT
BMP7	CACTTCATCAACCCGGAAAC CGGACCACCATGTTTCTGTA	161	AAACGGTGCCTAAGCCCTGCTGTG TTGAGCTGAGTCGGAGCACAGCAG
BRCA2	AGCCGTGGTCGAACTTACAG ATTCTGGGGCTTCAAGAGGT	170	TGTGGGTCAGAAGGTCATTATTCA CACCAGCTCTGCTCCGTGAATAAT
BRE	AGCACATATGGCTGCTTCCT GGTGTGCGCATTTATGTGTC	252	GGAAGCCATTGGAATGTCTTCATG GTGAGCGAAACGCTGCCATGAAGA
BUB1	GTGCAACAGTGACCTGCATC GCCTGGTTTTGAATCCCTCT	159	AAAACCAGGCCGAACCCGGGGAGC TATTGCTGTTGCAGCAGCTCCCCG
C14ORF132	CTACTCAAGCCCCCAGGTTT TCGTGCTCCTTTGACTGTTG	115	GTAAAACTGTAGCCGTGGAATCGT CAACAGTCAAAGGAGCACGATTCC
C6orf10	TTCTCATCCCAAGAGGTTGC TGGGCCTAATATTTGCAAGG	197	ACTTCTCTGTGTACCCCTTATTGC GCAAGGCACTGTCCTAGCAATAAG
C6ORF211	CCCAGAAAAGCTGTGTGTGA GGCAAAACGTTCAAAGGAAA	117	TGGGAACTCTCGGTTGCATAGAAA GAGTGCCTGTTAGATCTTTCTATG
*C8ORF74**	CCAAAAGGTAGGGAGCACCT TTTCCTTGAGCAGCTCTTCTG	184	AAAAGGTAGGGAGCACCTACGGAG CTCCCAGTTCAACAGTCTCCGTAG
C9	CTGCCAAATGAAGAGCAACA AGCTGTCCTGGCATATTTGG	168	AGCTCTGAAGTCTCCTCTCCTGGC CATCATGGATCTAAGAGCCAGGAG
CANX	GTGTGTGTTGCTGGTCCTTG AAACCTTTGGAGATGGAGCA	158	GAGGTGGAAGACTCAAAACCGAAA CAGTGCTGGCATCTGGTTTCGGTT

Table 11 continued

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
CAPG	GGCAGATGAACCTGACCAAG CCACGTACCCTTCCAGATGT	125	CTGGACAACGGGCTCTGTGGCAAG CCTTCCAGATGTAGATCTTGCCAC
CCL16	GGCCTAGGTCACCGTCATTA GAAAATGCAGACCACCCTG	135	AGCGGAAGCCCAAGTTCAAGAATG GGGTGACACCTTTGTTCATTCTTG
*CCT6A**	TGTGTAGTTCCAGGTGCTGGT ATGAGCAATGCATCAGCAAA	122	TAAATACAAGCCCAGTGTAAAGGG TCCAAGTTGGGCCCTGCCCTTTAC
CD101	AGGGCACTTGGTTCTTCAAA TCTTGTCACCTCTGCCACTG	150	TAAGCCCCAAGGCTTTCTCTCTCA CCCACAGAGAAGATCTTGAGAGAG
CD109	ATCAGCCGTCTTCCTCAAAA GGTTCCATTCAGCTCCAAAA	159	GATCCAAGAGTTGCAAGAGGCAAT TAGGAGGAGAGCTGCTATTGCCTC
CDC42BPB	AAGTCATCGCTGGCATCTCT TAAGGAGCCTCGACTTCCAA	107	CAAACGCTGGTGTGTCTGCACTTG AAGGAGCCTCGACTTCCAAGTGCA
CDK5R1	GTGGGATTGTGAGGCAAGTT GTGGAGACGTGGAACTTGCT	133	GGTGCAAAGACGGTGTCCTGTTGT GTCGTGCATGGTCAAAACAACAGG
*CLCN4**	CCGCCTCGTTCTCTTCTATG ACAGAGCTGGGAGGACTCAA	301	TATCACACGCCCTGGTACATGGCC ACATGGCCGAGCTCTTCCCCTTCA
CLIC1	CACACTTCAAGCCCCTTTGT GGCTACCCAATGGACACACT	199	CATCACCCCACAGCAAGCCCTGCA GTACAGAGAATCCTGGTGCAGGGC
COL11A2	GTCTCAGCACCCCACTCACT ATATGGGGCAGTGGAATGAA	195	CACTCTCACACTCCTAGTCTTCAT CATATGGGGCAGTGGAATGAAGAC
*CRCP**	GGAGAGTGAAGAACGGCTGA TACGCTGGGTCCTCTTCATC	141	CAGGCCCAGAGGCAGAGCAGAAGA TCATTGTTGGTGTTTTTCTTCTGC
CREM	CACCGGTGACATGCCTACTT TTTTTCATCAGCCTCAGCTCT	150	CTGGAACCGTTGCGCCCAGGTTGT AAATGCCACACAACACACAACCTG
CRISP3	TACGAGGCAATGTGGTGAGA GCCCTACGCCATAGGTAAAA	149	CAAGTAGCCCGACTTCCTGGTCAA CAGCTTTGGATTGCATTTGACCAG
*CSF2RA**	CACAGCTGCCCAAAACTTCT GGATCGAGTCAAAGAACTGGA	367	TAACAGACGCGGGAACAAACGTGG TCTTGGAGGTGGCATCCCACGTTT
CSNP_2782	GCTGTGCAGGGTACACAGACT AACCCCATAAGCTGTTCACAT	100	CCAGCTGTGCAGGGTACACAGACT TGTTGCCCATTGTCTCAGTCTGTG
CSTF2T	GGAATGGAGACCTGCACAAT GGATTACCCACCCCTGAAAT	200	GGAAACGAGAGGCATTGATGCAAG CCTGATCTCCATTCCTCTTGCATC

Table 11 continued

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
CYB5A	AGCGCTTAAAACATGCCATC CCTCGCATCACACAGACAAC	132	TGTCTCTAACGTGTAGAATCGGCT CAATGACTGCCCAACAAGCCGATT
CYP2A13	GTCAGCCCAAGTTTGAGGAC GAAGAACGTGGGGTCTTTCA	200	TATACGGAGGCAGTGATCCATGAG CTCCAAATCGCTGGATCTCATGGA
DAD1	CGTGCCAACCATAATCAAAA ATCCTGCACCTCGTTGTCAT	161	CGGGAATGATTCAGCCAACGAAGT CACCTCGTTGTCATGAACTTCGTT
DDR1	GGGGAGGGGAACAGAGTTTA GGACATAATGGGTGGAGGTG	201	GGAGGGGAACAGAGTTTAGCCCAG TCCCAGGTACCAGACACTGGGCTA
DDX52	GCTGCAAGGTGCAAGAAATA CCCATTGACCAGTTGTTTGA	160	CTGGACTTCTGTGTACTGGTACTA ACAACTGGTCAATGGGTAGTACCA
DHCR24	TGAGTCCGGTAAGGCTCTGT TTTGGCACAGCATATCTTGG	199	TTGCCTGACAGCCAGAAGAGCTCC TTCCTTCTCTCCATGAGGAGCTCT
DHRS3	TCTATGCACATTGCTGAGGC AAAGCCATGGAACCTGAGTG	255	GGACAGGAGCAGTTCCGCTCACCC TTGACACACTTTTGCTGGGTGAGC
*DSCC1**	CCTTGGTGGACAGACACTCA TGGAGCAATATCTTCTTCTGTCC	134	GGTTTAGCCTTGGTGGACAGACAC TAATTTCCGGTCTTGAGTGTCTGT
DYRK1A	ACGCCAGAGCTGTTCTCAGT CTCCTCACTGTTCAGCACCA	148	GTCTGTGCTCTTCACTTTTGGACC AACACCCACCAGCACTGGTCCAAA
EDN3	AGGTCAGTTTGGGGAGCAG TCCAGATGATGTCCAGGTGA	146	CATGCCTTACCTACAAGGACAAGG CAATAGTAGACACACTCCTTGTCC
EDNRB	ATCGAGCTGTTGCTTCTTGG TGCATGAAGGCTGTTTTCTG	115	TTTTAATTTGGGTGGTCTCCGTGG TCAGGGACAGCCAGAACCACGGAG
*ESRP1**	AGCCCAATTTCTCTCCAAGG GTGCGTACTCCTCACAAGCA	203	GACGAGGTGGTATCCTTCTTCGGA TAATGGGGCAATGCTGTCCGAAGA
EXO1	TGCAGAAGAAGACAAGCCAAT CCCAACCATTTGTTACTTTGG	172	CTCAGAAGCCAGAGAGTGTTTCAC GATATTGATAGAACGGGTGAAACA
*FDFT1**	GGGCAAGTACGTGAAGAAGC GCGGAATAGCGCAGAAGTTA	170	TGGGCGACTTTGCTAAGCCAGAGA ACGGCCACATCAACGTTCTCTGGC
FGF5	AAGAGGGGGAAAGCTAAACG TTCTCCGAGATGTGGAAAGG	184	ACACCTATGCCTCAGTGATACACA GATACACAGAACTGAGAACACGGG
FGFR2	ACTGGACCAACACGGAAAAG TATAGCCTCCGATGCGATGT	159	GTCCAACACCAACTATGAGGTGGC TCCTTCCCGTTTTTCAGCCACCTC

Table 11 continued

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
FOXO1	CTTGGTGGATGCTCAATCCT GCTTGCAGGCCACTTAGAAA	191	GGACAATAACAGTAAATTTGCTAA AGCTCGGCCTCGGCTCTTAGCAAA
FTL	CTGCTATTGCCCTGGAGAAG CGGAGGTTAGTCAGGTGGTC	163	TGGATCTGCATGCCCTGGGTTCTG CAGAACCCAGGGCATGCAGATCCA
GAREM	TTTCATGAACTTTCCTGCCTTT GCCTTGGGTTACTCCTCCTT	154	GACTCCAGAAAAGGAGGAGTAACC AAGTCACAAGGCCTTGGGTTACTC
GDF15	GGGCAGCTACAGACACCACT TGGAGATATCCGGACTGGAG	163	AGTTGCGGAAACGCTACGAAAATT CGAAGCCGGGTCAGGAAATTTTCG
GG_1068	TCAAATGCAATGGACATGAAA GGAAAGGCTCCTGGCTTTTA	162	CATTGGATTCATTGGTTCAAGAGT CATTCCGTCAGCCAGAACTCTTGA
GG_1079	CTTATGGCCTGATCCTGGAA TTTGCCTCCTAAAGGACCTG	153	TGAGGCTTATGGCCTGATCCTGGA CATGTTCCAGGGGAATTCCAGGAT
GG_1100	TGCAAGATATGCCAAGGTCA ATTTGTTGTGCATAAACAGCTTG	150	CTCAGCACCATCTGGAGGCATTAG AGTGGATTCACTCATGCTAATGCC
GG_1124	GCACCTACTATGTGTTGGCATT ACAGCTCCATTTGCAGGAAG	176	CCACCAGAAGTGTCACATCCAACA GCTCCTTCATACACACTGTTGGAT
GG_1142	CCTTCCCCACCACATGAT GGAGTGGGCAGTAATTTTCC	180	TGATGGACACCACTGAGAGTCTGC CGTCTTTAGGTGGTGAGCAGACTC
GG_1338	CTACGTCAACGCCAGCATAC AAACAGACAGGGGTCTCGTG	140	AATGCAACTACAGTGTCTACACAG TGGTTTCCCCAGAGAGCTGTGTAG
GG_1376*	ATGTGAGTGGGCCAGGAAC TGGTGATCTTGCTGGTCAGA	155	GCATTCAGTAACACCTGTGGTTCT GGTGATCTTGCTGGTCAGAACCAC
GG_1378	GCATATGGCTGTGATGAATGA TGTGACGGTGTCACTTAGGG	127	AGCCTGAACTGTCGAAAGTTCCCT TGTGACGGTGTCACTTAGGGAACT
GG_1414	TCAAGGAAAACTGATCAGAAAGA TCCACATCTACCCCTGTTCC	173	TAGACAGTTGTGAGAGGTTCTGCT CATGGAAGGTGGAGCAAGCAGAAC
GG_1438	TGACAGTCAACCTCCAAGCA GCTTCTTCTCCACAGGCTGA	188	TGGTTGGGGGTGGTAGTGGATGGA GAATGGGCTATTGAATTCCATCCA
GG_250	TGTCAAAGGACAGACCAGCA GCTTGTTTACAGTAGCCCCAAA	103	CAGAGCTAAAACAAGGTTCTCTGA TAGTCGTCTTGGTGGGTCAGAGAA
GG_326	GCAAAACGCTGATAAGCTGAC TCCTGAGATTGCTCTGTTATTTT	125	GCAAAACGCTGATAAGCTGACTTT CTCAGTTGAAGCTGAGAAAGTCAG

Table 11 continued

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
GG_410	TGGTTGATTAGAAATGGATAATTTGA CCATTGGCTAATTTGCAGGA	197	CTCAATACCCCTGAGGCCATGACT CCATGACTTAAAACCAGGAGTCAA
GG_478	GAACACATGGGCAACTGAAT GATGAGGCTGCAAGGATCTC	190	TATTGGAAAGAACACTGGACTGGG CCCAGGTCTCTTGGTTCCCAGTCC
GG_516	AGCCAGCCTTTTCTCTCTCC TCTCTCCTTCCCACCCTTTC	174	TCTCTTCTGGGAAAGGGTGGGAAG GCATCTCTGATCTCTCCTTCCCAC
GG_558	GCTCAGAAACTCTATGGGAGCTA AAGTTGGTCTGGAATATGCTACAA	150	GCTCAGAAACTCTATGGGAGCTAC ACTTCACAGCAGAAGAGTAGCTCC
GG_559	ATCTTGGCTGGAGAGCTGAG CATGAGTAACCCCAGAGGAAG	124	GGCCATTCCCATCTTGGCTGGAGA AAATTCCACTCTCAGCTCTCCAGC
GG_634	TGAAAACATCACAGCAGCTTTT TCCTCGCATGTAATGGGAAT	114	CAGACAACACTCTGAGGTCAATGG GATCGTAACACTGACGCCATTGAC
GG_705	TTGCTCACCCAGGGAATATC AGGAATGCTTTATTTGGCAGA	163	AGGGAAAGTACGCTGAATTGGAAT GACCTGGCAAGCTTCGATTCCAAT
GG_723	TTACAACCCTGACCCGTCTC TGCTTTTCCAAAGGCTCAAA	150	CCTTCTTTAAGGTAGAGCCACAAC CTCTTTGAGGGCCAAGGTTGTGGC
GG_733	CCCAGAATGGGCAGTGATAA CAGGACCCCTTTATGAAGCA	129	CCCAGAATGGGCAGTGATAACATC GTCGACTGAGCTGTACGATGTTAT
GG_796	TGGATTTGCATTTTCTTCAGC ACCCAGAACTCAACCGATCA	102	GCTCTCCCCTGTACACTTACACTT AACAGGTCACCACATCAAGTGTAA
GG_866*	TAGCTGTGAAATGACAGCACCA GGTCTGTCAAGCTCAGGATG	100	TAGCTGTGAAATGACAGCACCAGA ATGCTAGCAAGGTCCATCTGGTGC
GG_932	TTTTAGGAGACAGTTGTGGGAGA AGGACTCATCTTTGGCCTCA	153	CCATTTAGTGCCAAGCTAACCAGA TCTTTGGCCTCAGAGCTCTGGTTA
GG_962	TGCAAAGTCAAGCATACAGCA CCTCTCACATCGTGACCTCTT	167	AAGAGGTCACGATGTGAGAGGTAA GTACTCAGTGAGTACTTTACCTCT
GG_965	CCTTTCCACTCTGTGCATGA ACCCCTCAGCACGTGTTTTA	151	ATATAAAACGTCAGCCTCGCGGCC GTTCTGTATGTGCTGTGGCCGCGA
GNB1L	CAGAACTCACCAGCTCACCA GTCCGCCAGTGAAACACAC	114	TCTCGATCACAGGTGCACAGAACT GGTTGGTGAGCTGGTGAGTTCTGT
*GPR143**	CAAGTCCAGCAGCAGCATC CAGGCCCACTTTGTAGAGGA	151	AAGCGGGACTCAGCTTATGGTTCA TGGAGCGCGTGGAGAATGAACCAT

Table 11 continued

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
*GRINA**	CAGCTGTCCGTGACTCTGTC GAGATTCCAGGGGTGCTTTC	171	ACCTACTATGTCTCCTACGCCGTC TGAGGGAGATGAAGAAGACGGCGT
*GUSB**	GATGGGCCTGTGTCTGACTT CTGCATCCTCATGCTTGTTG	127	GCGGTCACAGAGAGCCAGTTTCTC AAGGTTTCCCATTGATGAGAAACT
*H2AFV**	CGCATCCACAGACATTTGAA TCAGCCGTGAGGTACTCCAG	101	TTTCCTGTGGGCCGCATCCACAGA TGCGAGTCTTCAAATGTCTGTGGA
*HBP1**	CCTGGATCACCACAGCTCTC TCTTTCCCTGGATACATCTGAG	173	AAAAGCTGTCAAAAACCACAGCTC GGCGCTCACAGTCCCTGAGCTGTG
HEYL	CCTTCTTTCCCACCTCAACA TGGGGTAAGCAAGAGAGGAG	197	TTCCCACCTCAACAGCTATGCAGC TGAAGGCTCCATCTCGGCTGCATA
HIP1R	AGGCTCTGCGGAAGAAGTTT TAGCAGGGAAGAGGCACTGT	137	CATGTCAGAGGCTCTGCGGAAGAA CGGCAGTCTCAGAAACTTCTTCCG
*HLA-B**	TGGGCCAGTTATTCAACCTC CCATAGCACTCAGGGGAAAA	195	CAGGTCTCTTCTGTGCTCAGAGTC TGCGCTTAGAATCTGAGACTCTGA
*HLA-F**	GTGAGATCTCCGCAGGGTAG CCTGGTTAGCAGTCACACCA	201	ATGTGTGTCCTGGGAGGCTCTGTG TCCCTGCTTTGTCTACCACAGAGC
*HLA-G**	TTCATTACCAGTGCGGTTCA GACGACTACCTGAGGCTGGA	198	GTATCTCCTCAGCCACAGGACACA CTACTTGGAGAAGGAGTGTGTCCT
HMGXB3	ATCATCATCCCCAAGAGCAG GGCTGAGGATCTCTGAGGTG	244	AAAGGACCATCTCTTGTATCCAAC CTGTCTCCAGGGCAGTGTTGGATA
*HS3ST3A1**	TACCAACACAATCCCACACG TAGGACATGGACTCCCCATT	207	ACTGTGGACATCAGCATGATCCTC GACGGATGGCTTCTGAGAGGATCA
HSD17B10	TCATCTGGTACAGGCCATCA TTTTGTAACCTCTGCCCACC	122	GGCCATCATCGAGAACCCATTCAT GATGACCTCTCCATTGATGAATGG
HSD17B12	TCTTTGCCTAGGCTGTGGTT TTTTGAGGGTGCTAAATGCC	226	CCCAGGCTCTGAGTTTAGGTACCA TGTTTTGCAAATTTACTGGTACCT
HUWE1	ATCTTTGGCCTGAGAAGGAAG ATGGAAGGAAACCAGGAGCC	125	CAAAGGACCTCCATTCAGGTTGGC CGGTGATCACACCCAGCCCAACGT
IER3IP1	ACTAAGCAGCCACGCCTTG CAGCAGCGAGTACAGCGTAA	145	TTTTCCGGGAGCGCGTTTTGGGCC CTACGGCAGCCAGTGAGGCCCAAA
IFNG-AS1	GCAAAACGTGAGCTTGACAG CCCTTTTGGTGCTAGTCCAC	157	CCAGCAAAAGCTGTAGCCACTTGA CCAGGCTTGTTCTTCCTCAAGTGG

Table 11 continued

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
*IGHA@**	CCAAGGATGTGCTGGTTCG GGTCGATGGTCTTCTGTGTG	110	ACCCACCCATGTCAATGTGTCTGT CACCTCCGCCATGACAACAGACAC
*IGHV@**	AAGTGCAGCTGGTGGAGTCT GGAGTTCTTGGCGTTGTCTC	233	ATAGGCTATGCGGACTCTGTGAAG AGATGGTGAATCGGCCCTTCACAG
*IGKJ@**	GGTTTCTGCTGAGAGGCAAG TTTAGACCCAAATTTCAGAAAGG	121	CTGTGGATCACCTTTGGCAAAGGG TAATCTCCAGACGTGTCCCTTTGC
*IGKV@**	GGGCCAGTCAGAGTGTTAGC AGGCTGCTGATGGTGAGAGT	163	TTCACTCTCACCATCAGCAGCCTG CAAAATCTTCAGACTGCAGGCTGC
*IGLC@**	AGCAGCTACCTGAGCCTGAC GAGAAGGGCTGGATGACTTG	210	GGCCCCTACAGAATGTTCATAGGT GGGGTTAGAGTTGGGAACCTATGA
*IGLJ@**	GCAAGGGGTTTATGTTCGAG CCCAGTCAGAGCAGAGAGGA	152	GCAAGGGGTTTATGTTCGAGGCTG ACAGGACACAGTGATACAGCCTCG
*IGLV@**	ACAGTCACAGTGTTGGCAAC TAATCAGCCTCGTCCTCAGC	193	ATTATTACTGCTCAGCCTGGGACA TAAGCACTGAGGCTGCTGTCCCAG
*IKZF3**	CCATGGGAAATGCAGAAGAG TATGGCTCCGCTTATGAACC	195	GATAGTAGCAGGCCAACCAGTGGC CATCGCAGTTCATCTTGCCACTGG
IL18RAP	ACGTGACTGTGGGCAAAAC TATGTCAACGGACCCAGGAT	163	CTCCACGTGACTGTGGGCAAAACC TGAAAAAAGCTCTCAGGGTTTTGC
ITGB5	AGGTTCGTCAGGTGGAGGAT TGGTACCTCGGTGCTGTGTA	144	AGGAGATGAGGAAGCTCACCAGCA AACCCCAGTCGGAAGTTGCTGGTG
*JMJD6**	TTATAAGCCCGTGGTTCTGC AGATGTAAAGGGGGCTGTCA	200	AATGCCCAAGAAGGCTGGTCTGCG GAGTCCATTTCTCCTGCGCAGACC
*KAL1**	CCTCTGTTCCCCAAGAAGAA ACAGCCATTGGAACAGCACT	186	ACCAGCTGTGAGTTCCTCAAGTAC GCTTCACCGACAGGATGTACTTGA
KAZN	GCCATGTGGCAAGGAACT ACTTTGTGATTGTAGGATTGAGG	101	GCAAACAGCCATCAAGAGAACAGG TTGTAGGATTGAGGCCCCTGTTCT
*KCNJ16**	TTACGTCCTCTCCTGGTTGG CACAGAGCATTCTTCGGTGA	196	GGCATCACGCCTTGTGTCGACAAC CCGTGAAGGAGTGGACGTTGTCGA
KIFC1	TTTGTCCCTTATCCCCTTCC CTTTTTCTTGCTGCCTCCTG	202	TCACTCTCCAGGACTGGTCTACCT GAGACCCAAGGATGAGAGGTAGAC
KIT	ATCCCGAGAAGCTTTTCCTC CACGTTTCTGATGGTGATGC	168	CCAAGGCTGGCATCACCATCAGAA TACTCGCGCTTCACGTTTCTGATG

Table 11 continued

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
KITLG	AGATGGTGGCACAGTTGTCA GTGTTCTTCCATGCACTCCA	140	GAGATGGTGGCACAGTTGTCAGTC GAAGATCAGTCAAGCTGACTGACA
LAMC1	GTGTAACGGACACGCAAGTG CATTCGCTGGCACTCTCC	156	TTACGGGGTGGACTGTGAAAAGTG ATTGAAGAAGGGGAGACACTTTTC
LARP4B	CAAAAACAAAAGCTCTCGGC TAGGAGTCAGTGCCATGCTG	232	TCATGTCTGGTGAGGCGCCAGGTT GAAGTGTAAGGGTGAGAACCTGGC
LBR	TTCAGGCAAAGGAAAAGTGG CAGTACCAGCGGAGTCAGTTT	192	CCGCCTCTGCTTCGCACCAGGTTG CTTGCTTCCTTAATGTCAACCTGG
LDB3	TGCGACTGAGTGACTTGTGTC AGAATTCACACAGGCCAAAAA	197	GGATCTACAGTTCTTGTCTCCTCA AGTAAATGTCAACCAATGAGGAGA
LGI3	TGACCCTCTGGTCTTTCCTG AGGACGGTAGTGTGGGTGAG	191	TCAGTTACACGCACACACACCTGT GGAAATGGGTGTGTAGACAGGTGT
Lgnuc134	AGGTCCATTAGGAAGACAACTCA TGACACTTCATAAGAGGGGACAT	115	GCCAGTTATATCTGGACTTGGAGA GGGTGATACTATGGTGTCTCCAAG
Lgnuc681	AAGCTCTGCCTTGTTCCAAA CACATTTCCATCCCCTGACT	189	ACTTGGATAAGGAAAGTAGTGCTC CATCCCCTGACTTACTGAGCACTA
Lgnuc682	TCACCGAGGAAATTCTGGTC TGTGGAAAACTAAAACCGGG	126	ACCGAGGAAATTCTGGTCCTTTTG AGACTGGTTACGAGCACAAAAGGA
*LHX1**	GGAGGGCAATATTTGCTGAA GGAGCAGGGACATCTGAGAG	360	TGCAAGCTGTGGGCCACATCAGTG GGATTGACATGAGCTCCACTGATG
LHX8	GGTATGGAACTCGCTGCTCT TCCACCAAAGCAAACTCCTC	221	ATCCATTCTACTGACTGGGTCCGG CATTTCCCTTGGCCCTCCGGACCC
LMO3	GATTGACACGGGAACCAACT CAACTCTGAACTGGGGCAAT	111	CACCCCAGGTTCGCTGATCTATCA CTTAATGGGGTGATGTTGATAGAT
LOC102545896	CTGTGTGCACATCCAGAAGG CCCCACGTGGAACATCTTTA	124	CTAGGCTGTGTGCACATCCAGAAG GGAAGTAAATATCTGCCTTCTGGA
LPGAT1	TGGTCACATGTGGATCTGGT ACCCCACCCCCACTACTAAG	206	GGCTGGTCACATGTGGATCTGGTT CCCAAATGCATTCACAAACCAGAT
MAL	GACGCAGCTTACCACTGTGT ATGGCAGGTTGACTGTGGAC	150	CCTGTTTTACTTCGGTGCCTCCGT GGCAAAAGCTTCCAGGACGGAGGC
*MDH2**	ATGACCCGGGATGACCTATT CTCACCGGGTTTGAAATGAT	113	GAGGCCATGATCTGCATCATTTCA TGACACTCACCGGGTTTGAAATGA

Table 11 continued

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
*MID1**	CACAGCCTCTGCTTCAACTG CTGGTCACAGAACTGGCAGA	294	TGTCATCACTCTCAGCCAGCGAGG TTTGAGCCCGTCTAGACCTCGCTG
MITF	CTCGAAAACCCCACCAAGTA ATAGCCATGGGGCTGTTG	185	CCACATACAGCAAGCCCAAAGGCA GTACTGCTTTACCTGCTGCCTTTG
MRPL41	CGCCTGCTATTGAGAAGGAC ACAGATGCCCGAAACAAATC	259	CCTCGCTGCGTTTCTATTAAAGGC CTGTACGGTAGCGAAGGCCTTTAA
MYEOV2	GTGGTGGATCTCCTCAAAGG GGACAGAGGCTGGTCATGTT	150	GAGGACGATTTTAACATCTGGGTG CAGAGAGCAACAGGCCCACCCAGA
MYOC	AGACCACGTGGAGAATCGAC GTGCCCAGCTCATACCTGAT	205	GATGTGGAGAACTTGTGTGGGTAG AGCGTGACAGGCTCTCCTACCCAC
NCBP1	AATGTTGGGGGTTTTTAGGG CAGCCAGGATATGTGGGTCT	247	CCCTTTTTCCTGTGTCCACAGCAT GCTGGATTATTTGGTGATGCTGTG
NEGR1	GGTTTCCATGAACAGCTGGA AGAACAGGAACCGTGATTGC	155	AAGGCCTAGAGATAGAGTCGGTTT ATCCAGCTGTTCATGGAAACCGAC
*NF1**	CTGTTTTGGGGTTTTTGGAA TCCATTTGCTGATGGTGAAA	194	CTCTGATGTCACCCAAACACACAT GATGGTGAAACCCCTGATGTGTGT
NOTCH4	GGCCAAGTCTACGGACAAAG AACCACTGGGATCTGCTGAC	210	GGCCAAGTCTACGGACAAAGTCCG AGCCTGTCCGCAAGCTCGGACTTT
NPTN	TTTACCTTCTGTTTTCAATGACCTT TGCCGGGTGAGAAAATCTAC	131	GGGTGCTCACACGCGGTACGTAAC AGTACATGCATCTACCGTTACGTA
NR3C2	AGAATCTGGCAGGCCTCA CTCAATTTGCTTTGTTCTTTGC	100	GAGGATAACTACTCTATCTGGAAT TACTGTAAGTGGCCGGATTCCAGA
OPTN	AGCAGAGGCAGAACAGGAAG AGCCATTCCGATTTCAACAA	154	AGGCTGAAAAGGCAGATCTGCTGG AATTCAGACACGATGCCCAGCAGA
PARK2	GCAGTTTGTATTCCCATGTGC GCTTGCAGTACGCATTGGT	151	ACCCTGAGGAGCAGAGCATTTACC GCTTGCAGTACGCATTGGTAAATG
PAX3	AGTCCGATGAAGGCTCTGAC CAGCTCCTCCCTGGTGTAAA	158	GAAGGAGCCGAACCACCTTCACAG TCTTCAAGCTGTTCTGCTGTGAAG
PGRMC1	CAGCTGGCTTGCCTAAAAAC GCGATGAGAACTGCTTCCTC	200	GAATAGGAACAGGTGATTGCTCTC GCGTGAAGGCTCTTAGGAGAGCAA
PIGR	GGTGAAGAGTGTGCCCAAAT CCTCTTTCTTCAGGCCAGTG	167	CCTGACTTTGAGGGCAGGATCCTA TGTTGCCAGTAGTGAGTAGGATCC

Table 11 continued

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
PKP2	GCTGTGGAATTTGTCCTCCA TTAGACATCCCGTGACGTTG	173	AGACTACCCCAAAGCGAACGGCTT TATATCAAAGTCGAGCAAGCCGTT
*PNPLA4**	ACTTGTCCTTTGCAGCCTGT GTTAGCAGGACGGAAGCAAC	148	CTGTGGGTTTCTGGGCATTTACCA AGATGCTGCCCCCAAGTGGTAAAT
PRR3	GCAGGGTCAGAGGTCTTGAG TAACAGCTGCAGGCACAACT	204	TCAGTTCATAAGTGCAAGCGCTGG TGGGTTTCCCGCCTCACCAGCGCT
PSAT1	CAGTTCTCCCCTGAGCCATA GAGATCTGGCCTCTTTCACG	147	ACTAGACTTTTTCTAGTGCAAGTG GGCAGGGAGTGAGTTCCACTTGCA
*PSPH**	CCCACTCAGAGCTGAGGAAA TTCTGACACAGCGTCCTCAA	131	CTGCTGATGCAGTGTGCTTCGATG ATGACTGTGCTATCGACATCGAAG
RAB38	CAGCCACATTTGAAGCAGTG TGCAGAACTGGTCCATCTTG	159	AATGTGACCAGGGGAAGGATGTGC AGGCCATTGTTCACGAGCACATCC
RAG1	GGAATGAGCACAGACAAGCA GGAAGCCACGTTCTTCAGAG	255	CAGGACTGTGAAAGCCATCACGGG CTGGAAAATCTGCCTCCCCGTGAT
*RALYL**	CTCCCCGCGCAGTAATTC GAGCCGGATGAAGAAGACC	127	AGGTCCACAGCCAGCGGGTCTTCT ATTTCGAGCCGGATGAAGAAGACC
RB1	TTTTTGTTCCCAGGAAGGTT TGGCAGAGGCCTATACTGATG	163	GGCAGTTGACCTAGATGAGATGCC CTCAGTGAAAGTGAATGGCATCTC
*RB1CC1**	GCCAAGAAACTCTGCTCGTT TGGTAGGAATTTGAGCAATCC	110	GCCAAGAAACTCTGCTCGTTCTGT CGTGGACCAGACCTTCACAGAACG
*RBFOX1**	CTGGGTAATTGGCTCCCATC ATGTCCAGAGAGCAAAAGGAA	217	CACTGATGCAGGTGACAGC TCCTCCTCCCAAGAGATCAA
*RDH10**	CTGGGAATGGTGAGGAAGAA AGAAGGTGATGCCCAGACAC	175	TGAGAGAACCATGATGGTCAATTG CCAGAAGTGTGCGTGGCAATTGAC
REN	GAGGTGCATCCGTGATTTTT TTATCCAGTAGCCCCTGGTC	160	TTCACTTTTGCCCACTTCGATGGG AACCGAGGCCCAATACCCCATCGA
*RP1**	CAGTCTCCAGGCTGTGATCC TGACCTAGCATTTCCCTTGG	154	TGCTTCCTGCTAGATTACCAGGGA TGCACACGATGAGAGATCCCTGGT
*RP1L1**	GGACTCTCTGAAGGGTCTGCT TTGAAGTCTGCCCAGACAAA	128	ACTTTGTCTGGGCAGACTTCAAGG ACTCACCATACTTGTTCCTTGAAG
RSPO1	GAGCTCTGTTCGGAGGTCAA TGATGCACTTGTTCATGTCG	190	CTTAAGTGCTCGCCCAAGCTGTTC TCCTCTCCAGCAGGATGAACAGCT

Table 11 continued

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
*RSPO2**	ATCAAATCCCATTTGCAAGG GTGTCCGTAGTACCCGGATG	115	GTTGCTTGTCTTGCTCAAAAGACA CATCGGCTGCACCCGTTGTCTTTT
SAFB	CGCCAAGGTTGTGACAAAT TTTGTGCAGGTGGTTAATGC	106	AATGCATTAACCACCTGCACAAAA TTCCCATGGAGTTCTGTTTTGTGC
Scaffold 263*	CCAGATTGCCAGCTAGAGGT ATTGGGGCATTCAATACAGC	154	TACATGTGGTTCGACAAGAGCTGG AAAGTAGAAGCGGGTGCCAGCTCT
Scaffold 337*	GATGCGCCATCTTAGTTCCT GGAGACTGTCATTCATGCTGTT	173	CCTCTTGTCTTGAGCTTAATCCAA CTGTTGCCTCCCCTTCTTGGATTA
Scaffold 368	CCTGGAGGGCCATTACTTTT CTGGCTGGATATGTGGGACT	199	CACAGTGTCAAACCCAAGAATACC AGAATACCAGTCCCACATATCCAG
Scaffold 395*	AGATGGGAAGCACAGAGCAC CCTTAGGGCGGTTAGGAAAG	162	AGGTCTCGTAATGCACAGAGCAGC CCAAGGAGGCTAAAAAGCTGCTCT
Scaffold 540*	AATGAGGTAAACACTGGTAATTGTG CGCCTTAATTAGGCAGCAGA	159	GCTTTAAGCATTTGCAGCACCTCT GCTGTGACCAGAAGCTAGAGGTGC
Scaffold 549*	ACCCTGGGGAGCCATTAG CTCATTTCCGGAAGGACAGA	168	CAGTTTCTTCCCAGGGGCCGGGTT TTACAAAAGTGTCAACAACCCGGC
Scaffold411	GAGTCGGGGCAATACAGAGA ACCCTCAGATCCCACGTTCT	206	ACGTGTGCTCCCCTAGGCAAGCTT GGGGTTCAGCGCTATGAAGCTTGC
Scaffold421	GGTGATGGGAAAATCTGCAC TTTTCTCCCTCAGCTCTTGG	216	GGGGGTGATGGGAAAATCTGCA GTGCAGCTGCAGTGTGCAGATT
SCP2	GCTTCAACCAGGCAAAGCTA CACCCAGAGCTTAGGGAAAA	190	AGCTGTGAAGAACTACCTTTGCTC CCTGATTTTCAAAGGTGAGCAAAG
SF1	ACAACCACGTCTCCCAAAAG TTTCGTTCGCTCTCTGGATT	163	ACCACGAGTGCACCGTGTGCAAGA ACCTTGTTGAAACTGATCTTGCAC
*SHROOM2**	CTGAATCCAGCCCAAGGTT AGCAAGCTCAAGGCCTCAT	230	AATCCAGCCCAAGGTTTCCTGTTG CATAGAAGGCCAGGGACAACAGGA
SIK3	AAACTTCCACGGAATGATTTCT CATAAACATTTCAACAAAGCACTG	170	CAAGAAGAGTTTCCTGAATACCAA TGCCCGTAAGTTCTGCTTGGTATT
*SLA**	ATTCCCTCTCGGTGAGACAC GTGATTCACCAGGTCCTCCA	119	CCAAACAACTGGTATTACATTTCC GGAAGGTGAGCCTCGGGGAAATGT
SLC18A2	CAACTCCACCATGGTCACTG CTGCACGTTCTCATTCAGGA	201	CCGCCTCAGGCAGTTTCCAGAACA TCATAATAGGAGAAGATGTTCTGG

Table 11 continued

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
SLC22A13	AGCTGCTGATCCTGGTGAGT AAGCTGTGGCTGAGGATGTC	250	CCTTCAACATGTTTGCCCAGGTCT TCATCCAGGACCATGAAGACCTGG
SLC35F1	TGTATGCCATTTCCTATCTTCC GGTTTCAGATGCAACTCCTTG	151	TGTATGCCATTTCCTATCTTCCGC CAGCTCGCACTCAGAAGCGGAAGA
SLC36A1	GCCCCTTGAAAACAAAATGA ACAGTTGGGCAGGTTGAGAG	160	GGGGAGTCTGGGGTACCTGCAATT TTGGATATTTGCTCCAAATTGCAG
SLC45A2	GGGTTACGTCTTGGGTGCTA GTGCGGGAATGTCTTTAAGC	164	TGCTCTCTTTGTGTTTTATCATCC GGGATACTGCACAGATGGATGATA
SNAP91	GATGCTGTTCAAGGGTCTCC CACATGTGTTGGGGTGCAG	153	CTGGAACATCACTGTAACTTTGCA CCACAAACCTCCACCGTGCAAAGT
SOX2	CACAACTCGGAGATCAGCAA CGGCAGCGTGTACTTATCCT	186	CTGCACATGAAGGAGCACCCGGAT GGGGCCGGTATTTATAATCCGGGT
SOX5	TGCCCCACACAACTCATCTA GCTCTTCCATTTTCCTCTGC	158	AAGGCAGCTTAGCTGACGTCGTTG CTCTGCTTCAAGGTGTCAACGACG
*STC1**	AACATGGCCAGCCTCTTTC TTTCAGCTTCTGTGGCTCGT	213	AACAAAATTGGGCCCAACATGGCC GGATGTGAAAGAGGCTGGCCATGT
*STS**	ACGGATGCTTCTCCACACAC CCGTCCTGCTTTTCTCTGTC	301	AGCTGTCGCTGGGGAACATTATCT TGAAGCCACGGCTTCCAGATAATG
*SUMF2**	GGAATGGGCAGACAACTAGC TGAAATCTTTGTTGGTGACAGG	156	TGGTGAAGGACCTGTCCGGGAAGT AAAAGGTTTTACTGCCACTTCCCG
TACR1	GGTGACTATGAGGGGCTTGA AGAGGAGCCATTGGAGGTCT	180	AAATCCACTCGGTACCTCCAGACC TGTACACGCTGCCCTGGGTCTGGA
TCF7L2	TGCGTTCGCTACATACAAGG CTTCAGGGACAGCGAGAGAG	162	AAGGCAGCTGCCTCAGTCCACCCT AAGCTTCCATCTGAAGAGGGTGGA
TGFBR3	ACACCATCCCTCCTGAGCTA GACCAGGAAACAGCTGCAAG	210	AACGGAGGCTTCCCCTTTCCTTTC TCCTGGCGATATCGGGGAAAGGAA
TLR2	CAGGATTCCTACTGGGTGGA CACCACTCGCTCTTCACAAA	191	CAGGTTGTGTCTTCACAAGCGGGA TTTGCCAGGAATGAAGTCCCGCTT
*TMED4**	CGTGTGCACCTAAACATCCA CTGTTCCACCTGATCAAGCA	127	ACCCTGAGATCGCTGCCAAGGATA TGCAGCTCTGTCAGTTTATCCTTG
TMEFF2	AAGCATGGAAAGGCTTTGAC CCATAGAGCTGTGTCAGGCA	263	GTCCATCCTCCTTATCTTAAGCAT TGGTCAAAGCCTTTCCATGCTTAA

Table 11 continued

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
TRBV30	AGCAGACTGTGGCTTCACCT AGCGCGAGGATAAAGAACAA	197	GTGCTCATGGCCATGGTTAAGAAA GGTCTCAGGAATCCTTTTTCTTAA
*TRIB1**	AAGTGATCCTTGGGGAAACC CTCCTCCGTGGAGAAGACAA	209	ACACTACCAGGACAAAATCCGGCC TGGCAGCTGGATGTAAGGCCGGAT
TRIO	GCGTGGAGCTTCTTTGTGAC GCTGCTCAGCTGCTAAATCC	108	AGGACTTGCTGGAGTTTCTCCATG AGTTCCTGCTGCTTTTCATGGAGA
TROVE2	CCTGCACCGGTTCTTATGTT AGGACACTGGGAACAGATGG	214	AGTCAAGAAGGCAGAACCGCAAAG AGAGCAGAGGCTCCTGCTTTGCGG
TSPAN7	TAGTGTGGGAGGCTCCTGTT TTCTCCTCGCGGACAATTAC	200	CCCTTGATCTTGATTGATAAATGC ACATCTGGGGGAGGGGGCATTTAT
TTC13	TACTGAAATGTCTCTACTCTGCTCA AGCAGCAGAGCCTTCTGAAA	100	TTGAAGCAGCCACTGAGAGCTTTC AGCAGCAGAGCCTTCTGAAAGCTC
TTR	GCCCCTACTCCTACTCCACC TTGTCTCTGCCCGAGTTTCT	174	GATCCAAAGGACGAGGGACAGGAT CTATCGGTTGCACGAAATCCTGTC
TYRP1	CCACCAGGAGATCAGAGGAA CCACTTCACCAAAGCTCTCC	150	ATATTGGGGCCAGACGGCAACACG TGTTCTCAAATTGGGGCGTGTTGC
*TYW1**	AAATTTGCGACAATTCTTGCTG CTTCTACTAGGTGATCATCTGGGTCAT	100	TTACCTCCCTCGAGCTGCCTGTGG TTCAGATCAGTAATTGCCACAGGC
*UBE2K**	CCATGGGGCCAGTTTACTAC GAAAATAAAACCAAAGCAAATGG	154	CGATTAGCTCTCATAAGCCAGGCT GCAAATCTTGGGCTCCAGCCTGGC
UCN3	GTCTTCAGCTGCATCAGCAC GGCCCTTTTCTCCTTGTCTT	159	CATCCTCAGGAGAGGACGAGGAGG TTGTCTTGCTCCTCCTCCTCCTCG
USF1	CGAGTGATGATGCAGTTGACA CTGCCCCAGTAGTGCCTCT	155	CACAGAGGGGACAGCTGCTGAGAC GAAGTAAGTATAGTGCGTCTCAGC
*VIM**	ACCTCTCCGAGGCTGCTAAC GCATCCACTTCACAGGTCAG	106	GCGAAGCAGGAGTCCAATGAATAT ACTGCACCTGTCTCCGATATTCAT
*VWC2**	CCCAACTGCTTCGCAGAGA CTACATCTGCCTGCATTCGT	134	TCTATTTCAAGACATCCACAAGAA CTCCGATGGCCCCAAGTTCTTGTG
*WBSCR17**	ACATCGAGCGGAAGAAGAAG AGTGGCAGATTCCACGCTAT	130	CCATACAACAGCAACATCGGCTTC CGTTCCTCTTGGTGTAGAAGCCGA
WT1	GGAGTAGCCCCGACTCTTGT GGCTGTGCATCTGTAAATGG	151	CGCCCCTTCATGTGTGCTTACCCA ATCTCTTATTGCAGCCTGGGTAAG

Table 11 continued

Marker symbol	PCR primer sequences Forward (top); Reverse (bottom)	PCR Product Size (bp)	Overgo primer sequences Forward (top); Reverse (bottom)
*WWC3**	CAGAGCTTACCCAGATTGAGG GTATGAGCGGCTGTCGTCCT	200	ACAATGGAGACGAGAGCAGGAACG GTACTCCTTCAGCATCCGTTCCTG
ZCWPW1	TGACCCGGGATGACCTATTA CTCACCGGGTTTGAAATGAT	115	TGAGGCCATGATCTGCATCATTTC TGACACTCACCGGGTTTGAAATGA
*ZFAND1**	CCAAGACAGGAGAGACAGCA CCATCCGCATGCATCTTTA	100	CCAAGACAGGAGAGACAGCAAGTA GCACCTTTCCGCCGTTTACTTGCT
*ZFPM2**	CACCCTGGATGAATTCTGGT GGTCCATTAGGGTTGTGCTG	274	TGAATTCTGGTCCTTCAGAGAGTC ATGCCGTGGTAGGGCCGACTCTCT
ZNF853	CCGATCCTTGTGCCTTACAT CATCTGAGAAAGGAGGCCAG	146	AGAGCCTCATTCCCAGGGCAAGCA GGCCTCCTTTCTCAGATGCTTGCC
*ZPBP**	CCTGGTGTTCCAGAACTTCG GTACTTCAGCGGCAGCTTTC	113	CGTAACCTAAGCAGCTGATGGCCG GTACAGGGACAAGGAACGGCCATC

Alpaca CHORI-246 BAC Library Screening and BAC DNA Isolation

Bacterial artificial chromosome (BAC) clones corresponding to the selected markers were identified using overgo primers, radioactively labeled with [^{32}P] 2'-deoxyadenosine triphosphate (dATP) and [^{32}P] deoxycytidine triphosphate (dCTP) (Perkin Elmer), and hybridized to nylon high-density filters of the CHORI-246 alpaca BAC library (https://bacpac.chori.org/library.php?id=448) as previously described (Avila *et. al.*, 2012). Briefly, equal amounts of 24 or less overgo probes were pooled and added to the hybridization solution, containing 20X SSPE, 10% sodium dodecyl sulfate (SDS), 5% dry milk, 100X Denhardt's solution, and 50% formamide. The solution was denatured by boiling for 10min, chilled on ice for 5min, and hybridized to the 11 filters containing all 202,752 BAC clones for 16h at 42˚C. After incubation, the filters were washed 3 times in 2X SSPE at 55˚C, 3 times at 25˚C, and exposed to autoradiography films in intensifying cassettes for 2 to 3 days at -80˚C. The autoradiograms were developed and the positive BAC clones were identified and picked from the library. The BAC clones corresponding to each marker (Table 2) were identified by PCR using BAC cell lysates as templates and marker-specific PCR primers. Isolation of DNA from 100mL cultures of each BAC clone was carried out using the Plasmid Midi Kit (Qiagen), according to the manufacturer's protocol. The BAC DNA quality and quantity were evaluated by electrophoresis in 1% agarose gels and nanodrop spectrophotometry (NanoDrop 3300, Thermo Fisher Scientific).

BAC DNA Labeling, Fluorescence In Situ Hybridization (FISH) and Microscopy

Labeling of BAC DNA was carried out with biotin-16-deoxyuridine, 5'-triphosphate (dUTP) or digoxigenin (DIG)-11-dUTP, using Biotin- or DIG-Nick Translation Mix (Roche Diagnostics), respectively (Raudsepp & Chowdhary, 2008; Avila *et. al.*, 2012). The precise chromosomal location of each molecular marker was determined by hybridizing differently labeled probes in pairs to metaphase chromosomes or in groups of 3 to interphase chromosomes, according to our protocols (Raudsepp & Chowdhary, 2008; Avila *et. al.*, 2012). After overnight incubation, biotin and DIG signals were detected with avidin-fluorescein isothiocyanate (FITC) (Vector Laboratories) and anti-DIG-rhodamine (Roche Applied Science), respectively. Images for a minimum of 10 metaphase spreads or 20 interphase nuclei were captured for each experiment, and analyzed using a Zeiss Axioplan 2 fluorescence microscope, equipped with the Isis Version 5.2 (MetaSystems GmbH) software. Chromosomes were counterstained with 4'-6-diamidino-2-phenylindole (DAPI) and identified according to the nomenclature proposed by Balmus *et. al.* (2007), with our modifications (Avila *et. al.*, 2012).

RESULTS

A Comprehensive Whole Genome Cytogenetic and Integrated Map for the Alpaca

The alpaca CHORI-246 BAC library was screened with primers corresponding to a total of 230 markers: 192 protein-coding genes, 30 DNA or cDNA markers with unknown gene homology, and 8 previously unassigned sequence scaffolds (Figure 18, Table 10). A total of 91 markers were shared with the RH map: 61 genes and 30 aforementioned DNA or cDNA markers. Forty-four of the markers used in this study were present in the first cytogenetic map for the alpaca (Avila *et. al.*, 2012). Therefore, another 186 molecular markers were added to the previous alpaca cytogenetic map. Most of the markers were found, on average, in 5 BAC clones, but only one clone – the one with the cleanest and strongest PCR amplification – was selected for FISH mapping (Table 10). Only those clones that produced a clean FISH signal at a distinct chromosomal location were used in this study; chimeric BACs, or those that recognized multiple sites across the genome, were substituted (see Discussion).

The 230 BACs were assigned to all 36 alpaca autosomes and the sex chromosomes (Figure 18). The goal of this study was to map at least 2 markers per chromosome arm, whenever possible. This was accomplished for 5 chromosomes: LPA9, 11, 14, 16, and the X. For 3 autosomes, only 1 marker

was mapped to the short arm: LPA7, 23, and 34. In the remaining 29 autosomes, no markers could be assigned to the short arm, but least two markers were successfully mapped to the long arm. In the case of LPAY, all 10 genes assigned to this chromosome (*CSF2RA, ARSF, STS, PNPLA4, KAL1, GRP143, SHROOM2, VWC3, CLCN4, MID1)* were pseudoautosomal, thus also mapping to LPAXpter. No markers were mapped to the male-specific region of Y because the CHORI-246 BAC library originates from a female alpaca (see Chapter I).

Precise cytogenetic locations of all BAC clones were determined by aligning the DAPI bands with the chromosome nomenclature proposed by Balmus and colleagues (2007), with our modifications (Avila *et. al.*, 2012). The relative order of all syntenic markers was determined by dual-color FISH on metaphase chromosomes (Figure 19a) or, in the case of closely located markers, by interphase FISH (Figure 19b). In some regions, such as the PAR in LPAX and LPAY, the MHC in LPA20, immunoglobulin clusters in LPA6, 28 and 32 (Figure 18), in which the markers were very close and thus overlapping even in interphase, their relative order remained unresolved.

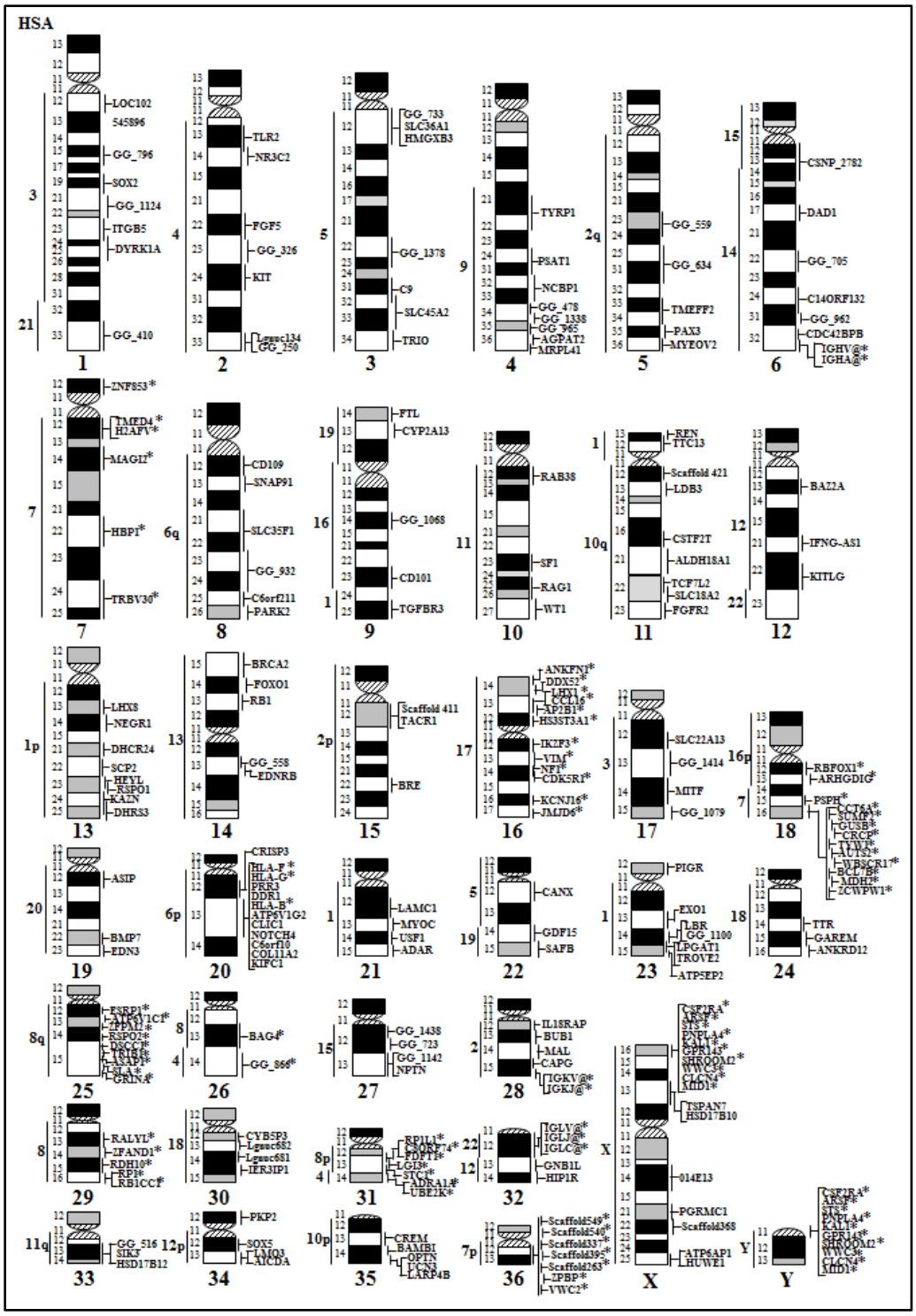

Figure 18. A 230-marker whole genome cytogenetic map for the alpaca. Bars to the left indicate chromosomal homology with human (HSA). Asterisks indicate markers mapped in the alpaca and the dromedary camel.

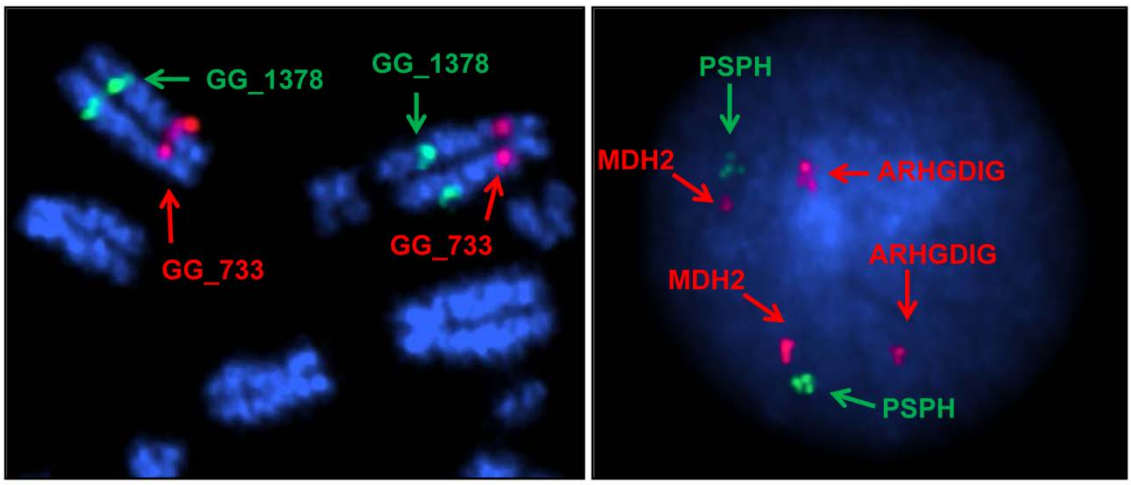

Figure 19. a) Representative metaphase FISH showing GG_733 (red signals) and GG_1378 (green signals) on LPA3. b) Representative interphase FISH showing the order of *ARHGDIG* (red), *PSPH* (green) and *MDH2* (red) on LPA18.

The chromosomal locations of the majority of markers were in agreement with, and in some instances refined, the human-camel Zoo-FISH data (Balmus *et. al.*, 2007), except for 3 regions in which new homologous synteny blocks (HSBs) were revealed. Two genes from HSA1qter, *REN* (HSA1q32.1) and *TTC13* (HSA1q42.2), mapped to LPA11q instead of LPA23 (Figure 18). Also, assignment of a gene from HSA4p13 (*UBE2K*) to LPA31q14 revealed a third segment of the alpaca genome homologous to this human chromosome, besides the previously reported LPA2 and LPA26 (Balmus *et. al., 2007*). (Figure 18; see figure on page 151). Most importantly, 7 markers were assigned to LPA36, thus filling the last gap in the human camel Zoo-FISH map (Balmus *et. al.,* 2007), and revealing the homology between LPA36 and HSA7p12.1-p11.2

(Figure 18; see Chapter III). This segment of HSA7 had been erroneously reported as being homologous to CDR18 by Zoo-FISH (Balmus *et. al.,* 2007).

Overall, all conserved synteny segments between human and camel from the Zoo-FISH data (Balmus *et. al.*, 2007) were confirmed by gene mapping, except for one - no markers were mapped to LPA12q23, which corresponds to HSA22q13.1-qter (Figure 18).

Besides the PAR (LPAXpter/LPAY), whose precise location in the sex chromosomes was determined in Chapter II, the chromosomal location of other regions of interest on the alpaca genome was determined in this study. Mapping of 3 Major Histocompatibility Complex (MHC) genes (*HLA-B*, *HLA-F* and *HLA-G*) assigned this region to LPA20q13. Likewise, FISH mapping of 7 immunoglobulin (Ig) genes revealed 3 Ig gene clusters on LPA6q32, LPA28q15ter, and LPA32q12 (Table 10, Figure 18).

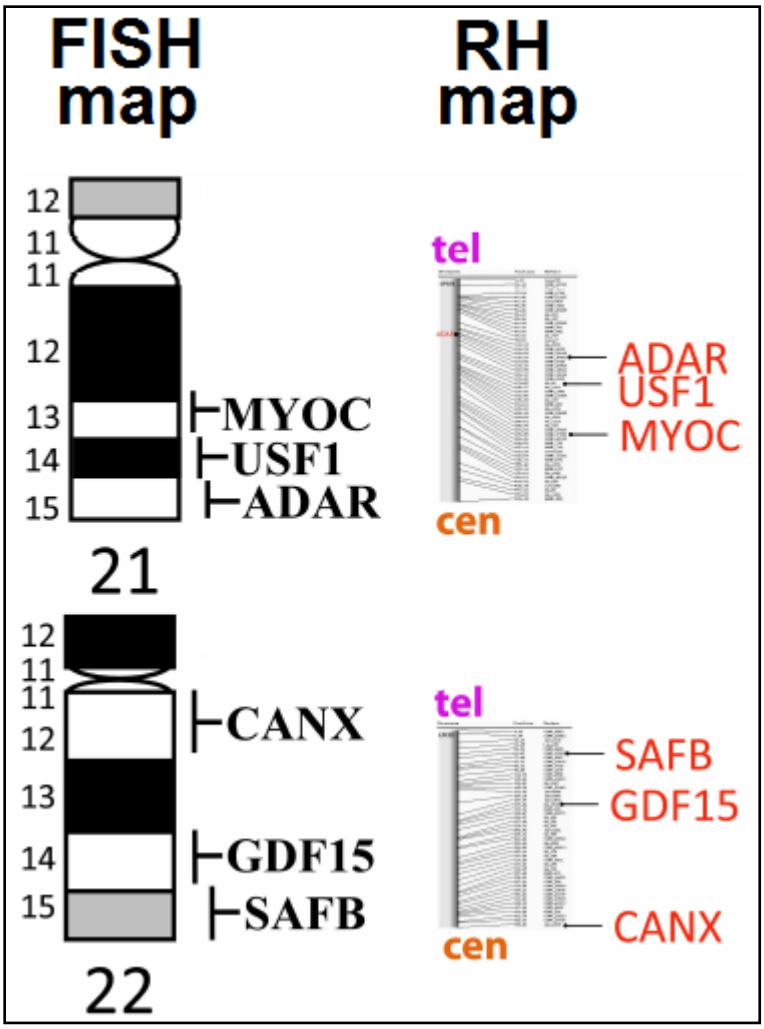

Figure 20. Representative examples of two RH groups (right) anchored and oriented on LPA21 (top) and LPA22 (bottom).

Integration of Cytogenetic, RH and Sequence Maps

A total of 91 markers from the alpaca RH map (P. Perelman, personal communication) were used to analyze the co-linearity and integrate the RH and cytogenetic maps. With this approach, RH groups (P. Perelman, unpublished

data) were successfully anchored to their respective chromosomal regions, and their telomere-centromere orientation was determined (Figure 20).

Additionally, the precise chromosomal locations of 8 previously unassigned alpaca genome sequence scaffolds were determined by FISH mapping (Table 10, Figure 18). Among those, 5 mapped to LPA36, namely scaffolds 263, 337, 395, 540 and 549, whereas scaffold 263 was assigned to LPAXq, scaffold 421 mapped to LPA11q, and scaffold 411 mapped to LPA15q (Figure 18).

Comparative Mapping between Alpaca and Dromedary Camel

The karyotypic conservation between camelids has been well established (Bianchi *et. al.*, 1986; Bunch *et. al.*, 1985; Di Berardino *et. al.*, 2006; Balmus *et. al.*, 2007; Avila *et. al.*, 2012; Raudsepp, 2014). The construction of a gene map for the alpaca genome in this study also demonstrated that Zoo-FISH-based conserved synteny blocks are equivalent between human and dromedary camel, and human and alpaca. To further investigate the degree of evolutionary conservation of gene synteny and linkage between these two camelid species, selected markers from the alpaca cytogenetic map were used for FISH mapping in the dromedary camel. A total of 81 genes, distributed across 13 alpaca chromosomes (Table 10) were also mapped in the dromedary camel. These markers included immunoglobulin genes on LPA/CDR6 (*IGHV@, IGHA@*), LPA/CDR28 (*IGKV@, IGKJ@*), AND LPA/CDR32 (*IGLV@, IGLJ@, IGLC@)*; MHC genes on LPA/CDR20 (*HLA-B, HLA-F, HLA-G)*; PAR genes on

LPA/CDRXpter / LPAY/CDRYp (see above); all 6 genes on LPA/CDR7 (*ZNF853, TMED4, H2AFV, MAGI2, HBP1* and TRBV30) (see Chapter III); all 12 genes mapped to LPA/CDR16 – *ANKFN1, DDX52, LHX1, CCL16, AP2B1, HS3ST3A1, IZKF3, VIM, NF1, CDK5R1, KCNJ16,* and *JMJD6)* (which is represented by one homologous block in humans (HSA17); all markers mapped to LPA/CDR25 (*ESRP1*, *ATP6V1C1, ZFPM2, RSPO2, DSCC1, TRIB1, ASAP1, SLA, GRINA)* LPA/CDR26 (*BAG4,* GG_866), LPA/CDR29 (*RALYL, ZFAND1, RDH10, RP1, RB1CC)*, and LPA/CDR31 (*RP1L1, C8ORF74, FDFT1, LGI3, STC1, ADRA1A,* and *UBE2K)* (which correspond to the ancestral eutherian combination of HSA4/HSA8); and all LPA/CDR36 markers (Table 10, Figure 18; see Chapter III). These results showed that, besides sharing a high degree of karyotypical conservation, alpaca and dromedary camel chromosomes show a remarkable degree of evolutionarily conserved linkage (Figure 21), despite these species having diverged around 15 MYA (Honey *et. al.,* 1998; Webb & Meachen, 2004). Also, exceptional conservation of gene order was observed between human chromosomes and their camelid counterparts, even though these species diverged around 100 MYA (Stanley *et. al.,* 1994).

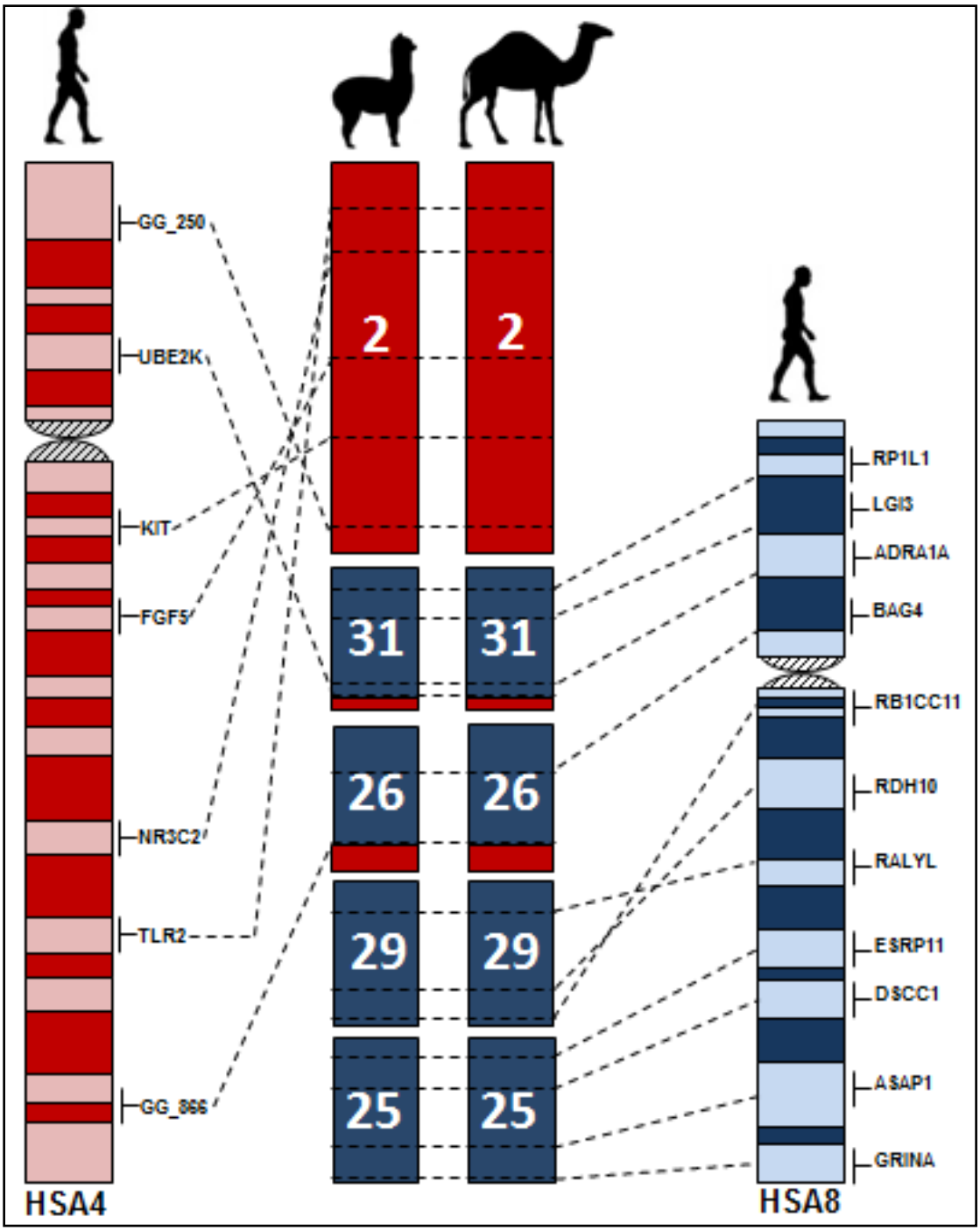

Figure 21. Chromosomal homologies between HSA4/8 and their camelid counterparts. Selected genes from the alpaca whole genome map show the remarkable conservation of gene synteny and linkage between alpaca and dromedary camel. HSA4 is represented by 3 segments in the camelid genome, and HSA8 corresponds to 4 segments.

DISCUSSION

This study reports the generation of an integrated whole genome cytogenetic map for the alpaca, composed of 230 markers distributed across all 36 alpaca autosomes and the sex chromosomes. The addition of 186 new markers essentially improved the previous 44-marker alpaca cytogenetic map (Avila *et. al.,* 2012), which represented the first genome map for a camelid species (see Chapter II). Notably, markers were assigned for the first time to LPA14, 21, 22, 28, and 36, and to large chromosomal regions that had only one or a few markers in the previous map (Figure 18; Avila *et. al.*, 2012). However, the main outcome of this study was the successful integration of the cytogenetic, RH and genome sequence maps, by anchoring RH linkage groups and sequence scaffolds to the alpaca karyotype - an important milestone in camelid genomics.

Mapping of sequence scaffolds by FISH is essential for the improvement of genome sequence assembly, by integrating the sequence map with the karyotype (Duke *et. al.*, 2007; Breen, 2008; Raudsepp *et. al.,* 2008; Lewin *et. al.*, 2009). For the first time, this was accomplished in camelids by anchoring 8 previously unassigned genome sequence scaffolds, with a total size of 6.1 Mb, to their respective chromosomal locations (Figure 18, Table 10). A similar strategy was successfully used to anchor and order 415 sequence scaffolds for the gray short-tailed opossum, *Monodelphis domestica* (Duke *et. al.*, 2007) and

to create the foundation for the integrated virtual map for the tammar wallaby, *Macropus eugenii* (Wang *et. al.*, 2011). This approach is especially useful for the alpaca, whose genome was sequenced and assembled before a physical map was made available, contrary to most domestic mammals (see Chapter I). Also, as the majority of markers in the second-generation alpaca RH map have been aligned with the *VicPac2.0.1* sequence scaffolds (Perelman, 2011; P. Perelman, personal communication), FISH mapping of 91 markers shared with the RH map in this study (Table 10, Figure 18) allowed simultaneous anchoring and orientation of RH groups and their corresponding sequence scaffolds, thus successfully integrating map information for the alpaca.

Further, as 192 of the mapped markers correspond to protein-coding genes, the cytogenetic map improved and refined the comparative information between camelid and human genomes. For example, the findings allowed for the precise demarcation of evolutionary breakpoint regions (EBRs) in the alpaca genome corresponding to HSA7 (see Chapter III), as well as to HSA4 and HSA8 (Table 10, Figure 21). In the putative ancestral eutherian karyotype, HSA4 is a part of ancestral eutherian chromosome (AEC) 2 and shares synteny with sequences corresponding to HSA8p, whereas HSA8q corresponds to AEC15 (Figure 20) (Svartman *et. al.*, 2006; Ferguson-Smith & Trifonov, 2007). Here we showed that in the alpaca genome, the ancestral synteny of HSA4/8 is present not once but twice – in LPA26 and LPA31 (Figure 21), whereas homology of LPA31 with HSA4 was detected for the first time. Altogether, mapping of 8

genes from HSA4 and 22 genes from HSA8 in this study confirmed the known HSBs (Balmus *et. al.*, 2007) and refined the location of the evolutionary breakpoints in the alpaca genome. Cytogenetic mapping of specific genes revised the Zoo-FISH (Balmus *et. al.*, 2007): HSA5pter-p14 (*TRIO*) was found to be homologous to LPA3q34 and not to LPA22 (Table 10, Figure 18); two genes from HSA1q32.1 and HSA1q42.2 (*TTC13* and *REN*, respectively) mapped to the short arm of LPA11, instead of LPA23 (Balmus *et. al.*, 2007). Also, as shown in Chapter II, HSA7p12.1-q11.2 corresponded to LPA36q12-q13 and not to HSA18. A few other inconsistencies between gene mapping and Zoo-FISH included *ATP5EP2* (HSA1p), which mapped to LPA23q15ter instead of LPA13; *VIM* (HSA10p), mapping to LPA16q13 and not to LPA35, and a BAC clone (014E13) which contained the *FTO* gene from HSA16 but mapped to LPAX instead of LPA9, as expected (Figure 18). These occurrences can be explained by novel HSBs – too small to be identified by Zoo-FISH. Though, it is plausible that in the case of *ATP5EP2*, the overgo primers picked a BAC clone containing another member of this large pseudogene family of ATP synthases. The BAC clone 014E13 probably contained some specific repeat sequences and gave a multicopy signal on LPAX, even though it was PCR-positive with primers for *FTO*. Nevertheless, the BAC was included in the map as an anonymous X-linked DNA marker denoted with its BAC ID. Due to the strong multicopy signal, it can be used for X-chromosome identification in clinical cytogenetic analyses.

Such minor discrepancies between Zoo-FISH and cytogenetic maps are expected due to the broad demarcation of evolutionary breakpoints stemming from the use of chromosomal paints (Scherthan *et. al.*, 1994). Similar disparities have been reported in several other species in which cytogenetic and Zoo-FISH maps are available (Chowdhary *et. al.*, 1998; Rubes *et. al.,* 2012; Graphodatsky *et. al.,* 2011). Apart from these few exceptions, gene mapping confirmed all the other Zoo-FISH homologies, especially in cases that one alpaca chromosome corresponded to 2 or more human HSBs, such as LPA9 (Figure 18, Table 10). The only exception was LPA12q23, corresponding to HSA22q13.1-qter (Balmus *et. al.*, 2007): despite the systematic development of markers for every human-camelid HSB, it was likely that this one was overlooked, and no markers were mapped to this region in the alpaca. Nevertheless, the overall good agreement between the cytogenetic and Zoo-FISH maps highlights the quality and robustness of the human-dromedary camel comparative map (Balmus *et. al.*, 2007). The fact that it was developed relatively late compared to other species (see Chapter I) was probably one of the contributing factors for its high quality, as it could build upon lessons from previous studies, combined with well-established techniques and the quality of chromosome paints.

It is worth mentioning that the only large region in the alpaca genome to which no genes were mapped was LPA4pter-q15 (Figure 18). This region was not painted by Zoo-FISH (Balmus *et. al.*, 2007) either, suggesting the prevalence of heterochromatic sequences in this segment of LPA4. The same was observed

for the majority of the short arms of the alpaca chromosomes. Despite having a goal to assign at least one marker per chromosomal arm, markers were mapped to the short arm of only 8 chromosomes: LPA7, 9, 11, 14, 16, 23, 34, and X. It is likely that this was because the short arms of many camelid chromosomes are heterochromatic, as previously shown by C-banding (Bianchi *et. al.*, 1986; Di Berardino *et. al.*, 2006; Avila *et. al.*, 2012; our unpublished data).

Besides refining and revising human-camelid Zoo-FISH homology, the map, with 192 linearly ordered genes, showed that linkage within HSBs is remarkably conserved. For example, LPA4 shares one-to-one conserved synteny with HSA9, and the 8 mapped genes are in the same linear order in the two species. The only outstanding exception was LPA16, which corresponds one-to-one to HSA17, but shows extensive intrachromosomal rearrangements compared to its human counterpart. Evolutionary rearrangements in the segment corresponding to HSA17 are quite common and present in many other species, such as horses (Chowdhary *et. al.*, 2002), dogs (Guyon *et. al.*, 2003), pigs (Liu et al. 2006) and cattle (Yang & Womack, 1998).

Whereas the gene order between the alpaca and human was overall well conserved, in some instances, there was substantial difference in the distance between the orthologs between the two species. For example, the markers GG_1338 and GG_965 in LPA4q34-q36 (Figure 18) were ordered by metaphase FISH, even though they are located only 500 kb apart in the human map (Table 10). Due to the resolution of this technique, the minimum distance between two

probes so that the signals can be distinguished is 1-5 Mb (Raudsepp & Chowdhary, 2008). Likewise, *AGPAT2* (HSA4:138.5 Mb) and *MRPL41* (HSA4:140.4 Mb) could be easily ordered in LPA4q36, despite their very close location (900 kb) in the human genome. The opposite occurred to two X-linked markers, *TSPAN7* (HSAX:38.4 Mb) and *HSD17B10* (HSAX:53.4 Mb): in the human X chromosome, the markers are located 15 Mb apart, but in LPAX their signals overlapped even in interphase. These differences might be due to variations in the position and/or amount of species-specific repetitive sequences located between such genes, or the presence of transposable elements.

This comprehensive integrated map for the alpaca, with at least two markers per chromosome, constitutes a much needed tool for clinical and molecular cytogenetics in all camelid species. Potential uses of these markers include the demarcation of inversion and translocation breakpoints, as well as the study of complex chromosomal rearrangements in these species. Markers from this map have already found application to investigate an autosomal translocation in a llama (Avila *et. al.*, 2012; see Chapter II) and the *Minute* Chromosome Syndrome (MCS) in alpacas (Avila *et. al.*, 2012; see Chapter II and III).

In summary, the whole genome cytogenetic map reported in this study represents an important advancement in camelid genomics and will expedite the availability of an improved alpaca genome sequence assembly which is aligned to individual chromosomes, as in other domestic mammalian species. The

collection of 230 cytogenetically mapped markers will be essential for the improvement and standardization of camelid chromosome nomenclature. Finally, for the development of this map, over 1,000 alpaca BAC clones were identified and isolated. This is a unique resource for targeted re-sequencing of regions of interest in the alpaca genome, for mutation discovery or to locally improve the genome sequence assembly.

CHAPTER V

SUMMARY AND CONCLUSION

In this study, information from the RH and genome sequence maps was integrated with cytogenetic data by assigning 230 molecular markers to the 36 alpaca autosomes and the sex chromosomes by FISH. In order to accomplish that, individual clones from the CHORI-246 alpaca BAC library were identified and isolated using primers corresponding to 192 protein-coding genes, 30 DNA or cDNA markers from the RH map and 8 previously unassigned genomic scaffolds. Altogether, 91 markers were shared with the RH map – 61 genes and 30 DNA or cDNA markers. The telomere-centromere orientation of RH groups within each chromosome was also determined. The assignment of 8 genome sequence scaffolds to their corresponding chromosomal locations represented the first step towards anchoring the alpaca genome sequence to the karyotype. In addition, the 91 RH markers mapped in this study will ultimately contribute to further integrate the sequence assembly with physical chromosomes in the alpaca, since each marker used to construct the RH map has been assigned to a scaffold from the *VicPac2.0.1* genome assembly. This information will be available once the alpaca RH map is published.

The alpaca cytogenetic map also serves as a tool for comparative studies in camelids. The molecular markers originated from the map were successfully

utilized to investigate the evolution of the camelid karyotype, as well as chromosomal homologies between camelids and several other vertebrate species. FISH mapping of selected markers from the whole genome map also allowed for the precise demarcation of evolutionary breakpoints in alpaca and dromedary camel chromosomes. Remarkable conservation of gene synteny and linkage was discovered between alpaca and dromedary camel, even though the two species diverged around 15 MYA, and despite their distinctive anatomical and physiological adaptations. It also revealed interesting evolutionary relationships between these species and between camelids and humans, as in the study case of camelid counterparts of human chromosomes 4, 7 and 8. Gene order was found to be overall conserved between humans and camelids, with few exceptions, despite around 100 million years of divergence between them, as well as between camelid species, confirming their karyotypical conservation.

The findings shown herein confirmed most and revised some of the Zoo-FISH data between humans and the dromedary camel. Refinements to the human-camel comparative map included the detection of 3 novel homologous synteny blocks (HSBs) in the camelid genome, and the precise demarcation of evolutionary breakpoint regions (EBRs), which can be accomplished with the use of molecular markers. Importantly, the only gap on the Zoo-FISH map was filled in this study, with the assignment of the first DNA sequences to camelid chromosome 36.

Molecular markers generated in this study were also used in clinical cytogenetics, to identify the chromosomes involved in an autosomal translocation in a male llama, and to investigate the *Minute* Chromosome Syndrome (MCS), characterized by the presence of an abnormally small chromosome 36 in infertile female alpacas. The importance of using molecular markers to identify numerical and structural chromosomal aberrations in camelids is accentuated by their particularly difficult karyotype to analyze. The relatively high number of chromosomes (2n=74), combined with overall similar chromosome morphology and banding patterns across the camelid karyotype and size polymorphism between homologs, hinder the efficiency of conventional cytogenetics to investigate chromosomal abnormalities in these species. Therefore, the set of cytogenetically mapped molecular markers for the alpaca can help with the identification of individual chromosomes or chromosomal segments in all camelid species.

Analysis of the *minute* chromosome in alpacas involved cohort karyotyping (n=100) of normal females, comparative genomic hybridization (CGH), FISH mapping of markers from chromosome 36 and rDNA probes in normal and affected individuals, and next-generation sequencing of flow-sorted chromosome 36 from dromedary camel and microdissected *minute.* Even though the results confirmed the homology between chromosome 36 and the *minute*, as well as its association with infertility phenotypes, the origin of this chromosome still needs to be further investigated. The hypothesis proposed in

this study indicated that size polymorphism identified on chromosome 36 originates from the presence of NOR sequences on one of the homologues, but not on the other, in affected individuals. However, NOR signals were not found on chromosome 36 in normal individuals. Further testing needs to be done in order to confirm this hypothesis.

The results of this study constitute an important step towards effectively bringing the alpaca into the genomics era. Findings from this study contributed for ideogram improvement in the alpaca, and laid the framework for the much needed standardization of chromosome nomenclature for the alpaca and other camelids. This gene map also plays an important role in facilitating subsequent studies aimed at the discovery of genes of interest for camelid diseases and economically important phenotypic traits. Also, the BAC clones identified in this study can be used for targeted re-sequencing of regions of interest in the alpaca genome, for mutation discovery or to locally improve the alpaca genome sequence assembly.

REFERENCES

Aba, M.A. 2014. **Anatomy and physiology of reproduction in the female llama and alpaca.** In: *Llama and Alpaca Care: Medicine, Surgery Reproduction, Nutrition and Herd Health (Ed. Christopher Cebra),* Saunders, Elsevier Inc., St. Louis, Missouri, USA, Chapter 14, pp. 140-150.

Achour, I.; Cavelier, P.; Tichit, M.; Bouchier, C.; Lafaye, P.; Rougeon, F. 2008. **Tetrameric and homodimeric camelid IgGs originate from the same IgH locus.** *J Immunol.* **181**(3): 2001-2009.

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:** 964-985.

Airmet, K.W.; Hinckley, J.D.; Tree, L.T.; Moss, M.; Blumell, S.; Ulicny. K.; Gustafson, A.K.; Weed, M.; Theodosis, R.; Lenhardt, M.; Genho, J.; Stevens, M.R.; Kooyman, D.L. 2012. **Construction of a llama bacterial artificial chromosome library with approximately 9-fold genome equivalent coverage.** *J Biomed Biotechnol.* Vol. 2012 Article ID 371414.

Allen, W.R.; Short, R.V. 1997. **Intraspecific and extraspecific pregnancies in equids: anything goes.** *J. Hered.* **5:**384-92.

Al-Swailem, A.M.; Shehata, M.M.; Abu-Duhier, F.M.; Al-Yamani, E.J.; Al-Busadah, K.A.; Al-Arawi, M.S. Al-Khider, A.Y.; *et.al*. 2010. **Sequencing, analysis and annotation of expressed sequence tags for *Camelus dromedarius*.** *PLoS One.* **5**(5):e10720.

Arneson, N.; Moreno, J.; Iakovlev, V.; Ghazani, A.; Warren, K.; McCready, D.; Jurisca, I.; Done, S.J. 2012. **Comparison of whole genome amplification methods for analysis of DNA extracted from microdissected early breast lesions in formalin-fixed paraffin-embedded tissue.** *ISRN Oncol.* Vol. 2012 Article ID 710692.

Arrighi, F.E. & Hsu, T.C. 1971. **Localization of heterochromatin in human chromosomes.** *Cytogenetics.* **10:** 81-86.

Avila, F.; Das, P.J.; Kutzler, M.; Owens, E.; Perelman, P.; Rubes, J.; Hornak, M.; Johnson, W.E.; Raudsepp, T. 2012. **Development and application of**

camelid molecular cytogenetic tools. *J Hered*. Epub ahead of print. 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:** 550-557.

Balmus, G.; Trifonov, V.A.; Biltueva, L.S.; O'Brien, P.C.; Alkalaeva, E.S.; Fu, B.; Skidmore, J.A.; Allen, T.; Graphodatsky, A.S.; Yang, F. *et. al.* 2007. **Cross-species chromosome painting among camel, cattle, pig and human: further insights into the putative Cetartiodactyla ancestral karyotype.** *Chromosome Res.* **15:** 499-515.

Bactrian Camels Genome Sequencing and Analysis Consortium, Jirimutu, *et. al.* 2012. **Genome sequences of wild and domestic bactrian camels.** *Nat Commun.***3:**1202.

Barreta, J.; Gutiérrez-Gil, B.; Iñiguez, V.; Saavedra, V.; Chiri, R.; Latorre, E.; Arranz, J.J. 2012. **Analysis of mitochondrial DNA in Bovian llama, alpaca and vicuna populations: a contribution to the phylogeny of the South American camelids.** *Anim Genet.* **44:** 158-168.

Becker, S.E.; Thomas, R.; Trifonov, V.A.; Wayne, R.K.; Graphodatsky, A.S.; Breen, M. 2011. **Anchoring the dog to its relatives reveals new evolutionary breakpoints across 11 species of the Canidae and provides new clues for the role of B chromosomes.** *Chromosome Res.* **19:** 685-708.

Bernardi, G.; Olofsson, B.; Filipski, J.; Zerial, M.; Salinas, J.; Cuny, G.; Meunier-Rotival, M.; Rodier, F. 1985. **The mosaic genome of warm-blooded vertebrates.** *Science*. **228**(4702): 953-958.

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

Breen, M. 2008. **Canine cytogenetics--from band to basepair.** *Cytogenet Genome Res.* **120:** 50-60.

Breen, M.; Bullerdiek, J.; Langford, C.F. 1999. **The DAPI banded karyotype of the domestic dog (*Canis familiaris*) generated using chromosome-specific paint probes.** *Chromosome Res.* **7:** 401-406.

Breen, M.; Hitte, C.; Lorentzen, T.D.; Thomas, R.; Cadieu, E.; Sabacan, L.; Scott, A.; Evanno, G.; Parker, H.G.; Kirkness, E.F.; *et al.*, 2004. **An**

integrated 4249 marker FISH/RH map of the canine genome. *BMC Genomics.* **5:** 65.

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. & Plamieri, N. 2013. **Estimating the population mutation rate from a *de novo* assembled bactrian camel genome and cross-species comparison with dromedary ESTs.** *J Hered.* Epub ahead of print. doi: 10.1093/jhered/est005.

Capanna, E. & Civitelli, M.V. 1965. **The chromosomes of three species of neotropical Camelidae**. *Mamm Chrom Newsl.* **17:** 75-79.

Chaves, R.; Frönicke, L.; Guedes-Pinto, H.; Wienberg, J. 2004. **Multidirectional chromosome painting between the Hirola antelope (*Damaliscus hunteri*, Alcelaphini, Bovidae), sheep and human.** *Chromosome Res.* **12:** 495-503.

Chowdhary, B.P. & Raudsepp, T. 2000. **HSA4 and GGA4: remarkable conservation despite 300-Myr divergence.** *Genomics.* **61**(1): 102-105.

Chowdhary, B.P. & Raudsepp, T. 2001. **Chromosome painting in farm, pet and wild animal species.** *Methods Cell Sci.* **23**(1-3): 37-55.

Chowdhary, B.P. & Raudsepp, T. 2005. **Mapping genomes at the chromosomal level.** In: *Mammalian Genomics (Eds: A. Ruvinsky and J. Marshall Graves)*, CAB International, Wallingford, Oxfordshire OX10 8DE, UK, pp. 23-65.

Chowdhary, B.P.; Raudsepp, T.; Frönicke, 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.

Chowdhary, B.P.; Raudsepp, T.; Kata, S.R.; Goh, G.; Millon, L.V.; Allan, V.; Piumi, F.; Guérin, G.; Swinburne, J.; Binns, M.; Lear, T.L.; Mickelson, J.; Murray, J.; Antczak, D.F.; Womack, J.E.; Skow, L.C. 2003. **The first-generation whole-genome radiation hybrid map in the horse identifies conserved segments in human and mouse genomes.** *Genome Res.***13**(4): 742-751.

Conrath, K.E.; Wernery, U.; Mulydermans, S.; Nguyen, V.K. 2003. **Emergence and evolution of functional heavy-chain antibodies in Camelidae.** *Dev Comp Immunol.* **27:** 87-103.

Constantini, M.; Clay, O.; Auletta, F.; Bernardi, G. 2006. **An isochore map of human chromosomes.** *Genome Res*. **16:** 536-541.

Criscitiello, M.F. 2014. **What the shark immune system can and cannot provide for the expanding design landscape of immunotherapy.** *Expert Opin Drug Discov.* **9**(7):725-739.

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

Das, P.J.; Chowdhary, B.P.; Raudsepp, T. 2009. **Characterization of the bovine pseudoautosomal region and comparison with sheep, goat, and other mammalian pseudoautosomal regions.** *Cytogenet Genome Res.* **126:** 139-147.

Das, P.; Lyle, S.K.; Beehan, D.; Chowdhary, B.P.; Raudsepp, T. 2012. **Cytogenetic and molecular characterization of Y isochromosome in a 63XO/64Xi(Yq) mosaic karyotype of an intersex horse.** *Sex Dev.* **6**(1-3): 117-127.

Das, P.; Mishra, D.K.; Ghosh, S.; Avila, F.; Johnson, G.A.; Chowdhary, B.P.; Raudsepp, T. 2013. **Comparative organization and gene expression profile of the porcine pseudoautosomal region.** *Cytogenet Genome Res.* **141:** 26-36.

Davis, B.W.; Raudsepp, T.; Pearks Wilkerson, A.J.; Agarwala, R.; Schaffer, A.A.; Houck, M.; Chowdhary, B.P.; Murphy, W.J. 2009. **A high-resolution cat radiation hybrid and integrated FISH mapping resource for phylogenomic studies across Felidae.** *Genomics.* **93:** 299-304.

Deakin, J.E.; Delbridge, M.L.; Koina, E.; Harley, N.; Alsop, A.E.; Wang, C.; Patel, V.S.; Graves, J.A.M. 2013. **Reconstruction of the ancestral marsupial karyotype from comparative gene maps.** *BMC Evol Biol.* **13:** 258.

De Genst, E.; Saerens, D; Mulydermans, S.; Conrath, K. 2006. **Antibody repertoire development in camelids.** *Dev Comp Immunol.* **30:** 187-198.

Di Berardino, D.; Nicodemo, D.; Coppola, G.; King, A.W.; Ramunno, L.; Cosenza, G.F.; Iannuzzi, L.; Di Meo, G.P.; Balmus, G.; Rubes, J. 2006.

Cytogenetic characterization of alpaca (*Lama pacos*, fam. Camelidae) prometaphase chromosomes. *Cytogenet Genome Res.* **115:** 138-144.

Di Meo, G.P.; Goldammer, T.; Perucatti, A.; Genualdo, V.; Iannuzzi, A.; Incarnato, D.; Rebl, A.; Di Berardino, D.; Iannuzzi, L. 2011. **Extended cytogenetic maps of sheep chromosome 1 and their cattle and river buffalo homoeologues: comparison with the OAR1 RH map and human chromosomes 2, 3, 21 and 1q.** *Cytogenet Genome Res.* **133:** 16-24.

Di Meo, G.P.; Perucatti, A.; Floriot, S.; Hayes, H.; Schibler, L.; Incarnato, D.; Di Berardino, D.; Williams, J.; Cribiu, E.; Eggen, A.; *et al.*, 2008. **An extended river buffalo (*Bubalus bubalis*, 2n = 50) cytogenetic map: assignment of 68 autosomal loci by FISH-mapping and R-banding and comparison with human chromosomes.** *Chromosome Res.* **16:** 827-837.

Di Meo, G.P.; Perucatti, A.; Floriot, S.; Hayes, H.; Schibler, L.; Rullo, R.; Incarnato, D.; Ferretti, L.; Cockett, N.; Cribiu, E.; *et al.*, 2007. **An advanced sheep (*Ovis aries*, 2n = 54) cytogenetic map and assignment of 88 new autosomal loci by fluorescence in situ hybridization and R-banding.** *Anim Genet.* **38:** 233-240.

Di Rocco, F.; Zambelli, A.; Maté, L.; Vidal-Rioja, L. 2010. **The complete mitochondrial DNA sequence of the guanaco (Lama guanicoe): comparative analysis with the vicuña (Vicugna vicugna) genome.** *Genetica.* **138:** 813-818.

Dong, Y.; Xie, M.; Jiang, Y.; Xiao, N. Du, X.; Zhang, W.; Tosser-Klopp, G.; Wang, J.; Yang, S.; Liang, J.; *et. al.,* 2013 **Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*).** *Nat Biotechnol.* **31:** 135-141.

Drew, M.L.; Meyers-Wallen, V.N.; Acland, G.M.; Guyer, C.L.; Steinheimer, D.N. 1999. **Presumptive Sry-negative XX sex reversal in a llama with multiple congenital anomalies.** *J Am Vet Med Assoc.* **215:** 1134-1139.

Ducos, A.; Revay, T,; Kovacs, A.; Hidas, A.; Pinton, A.; Bonnet-Garnier, A.; Molteni, L.; Slota, E.; Switonski, M.; Arruga, M.V.; *et al.*, 2008. **Cytogenetic screening of livestock populations in Europe: an overview.** *Cytogenet Genome Res.* **120:** 26-41.

Duke, S.E.; Samollow, P.B.; Mauceli, E.; Lindblad-Toh, K.; Breen, M. 2007. **Integrated cytogenetic BAC map of the genome of the gray, short-tailed opossum, *Monodelphis domestica*.** *Chromosome Res.* **15**(3): 361-370.

El-Wishy, A.B. 1987 **Reproduction in the female dromedary (*Camelus dromedarius*): a review.** *Anim Reprod Sci.* **15:** 273-297.

El-Wishy, A.B. 1988 **A study of genital organs of female dromedary (*Camelus dromedarius*).** *J Reprod Fert.* **82:** 587-593.

Even-Desrumeaux, K.; Fourquet, P.; Secq, V.; Baty, D.; Chames, P. 2012. **Single-domain antibodies: a versatile and rich source of binders for breast cancer diagnostic approaches.** *Mol BioSyst.* **8:** 2385-2394.

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

Fellows, E.; Kutzler, M.; Avila, F.; Das, P.J.; Raudsepp, T. 2012. **Ovarian dysgenesis in an alpaca with a *minute* chromosome 36.** *J Hered.* Epub ahead of print. doi: 10.1093/jhered/ess069.

Ferguson-Smith, M.A. & Trifonov, V. 2007. **Mammalian karyotype evolution.** *Nat Rev Genet.* **8:** 950-962.

Fowler, M.E. 2008. **Camelids are not ruminants.** In: *Zoo and Wild Animal Medicine (Eds: Fowler, M.E. & Miller, R.E.)* Current Therapy, 6th ed. Saint Louis: Elsevier, pp. 3375-3385.

Fowler, M. E. 2010. **Congenital and hereditary conditions of camelids.** In: *Medicine and Surgery of Camelids.* Third Edition, Blackwell Publishing, Inc., Ames, Iowa, USA, Chapter 48, pp. 391-403.

Fowler, M. 1990. **Twinning in llamas.** *Int Camelid J.* **4:** 35-38.

Franklin, W.L. 1982. **Biology, ecology, and relationship to man of the South American camelids.** In: *Mammalian Biology in South America (Eds: Mares, M.A.; Genoways, H.H.)* Pymatuning Laboratory of Ecology, University of Pittsburgh. pp. 457-487.

Frönicke, L. & Wienberg, J. 2001. **Comparative chromosome painting defines the high rate of karyotype changes between pigs and bovids.** *Mamm Genome.* **12:** 442-449.

Galibert, F.; Quignon, P.; Hitte, C.; Andre, C. 2011. **Toward understanding dog evolutionary and domestication history.** *C R Biol.* **334:** 190-196.

Gallardo, M.H.; González, C.A.; Cebrián, I. 2006. **Molecular cytogenetics and allotetraploidy in the red vizcacha rat, *Tympanoctomys barrerae* (Rodentia, Octodontidae).** *Genomics.* **88**(2): 214-221.

Gartler, S.M. 2006. **The chromosome number in humans: a brief history.** *Nat Rev Genet.* **7:** 655-660.

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

Goldammer, T.; Brunner, R.M.; Rebl, A.; Wu, C.H.; Nomura, K.; Hadfield, T.; Maddox, J.F.; Cockett, N.E. 2009. **Cytogenetic anchoring of radiation hybrid and virtual maps of sheep chromosome X and comparison of X chromosomes in sheep, cattle, and human.** *Chromosome Res.* **17:** 497-506.

Graphodatsky, A.S.; Trifonov, V.A.; Stanyon, R. 2011. **The genome diversity and karyotype evolution of mammals.** *Mol Cytogenet.* **4:** 22.

Gray, A.P. 1972. **Mammalian hybrids: a check list with bibliography.** Technical Communication Number 10 (Revised). Commonwealth Bureau of Animal Breeding and Genetics, Farnham Royal Slough SL2, 3BN, Edingburgh, England.

Groeneveld, L.F.; Lenstra, J.A.; Eding, H.; Toro, M.A.; Scherf, B.; Pilling, D.; Negrini, R.; Finlay, E.K.; Jianlin, H.; Groeneveld, E.; *et al*., 2010. **Genetic diversity in farm animals--a review.** *Anim Genet.* **41:** 6-31.

Gustafson, A.L.; Tallmadge, R.L.; Ramlachan, N.; Miller, D.; Bird, H.; Antczak, D.F.; Raudsepp, T.; Chowdhary, B.P.; Skow, L.C. 2003. **An ordered BAC contig map of the equine major histocompatibility complex.** *Cytogenet Genome Res.* **102:** 189-195.

Guyon, R.; Kirkness, E.F.; Lorentzen, T.D.; Hitte, C.; Comstock, K.E.; Quignon, P.; Derrien, T.; André, C.; Fraser, C.M.; Galibert, F.; Ostrander, E.A. 2003. **Building comparative maps using 1.5x sequence coverage: human chromosome 1p and the canine genome.** *Cold Spring Harb Symp Quant Biol.* **68:**171-7.

Hamers-Casterman, C.; Atarhouch, T.; Mulydermans, S.; Robinson, G.; Hamers, C.; Songa, E.B.; Bendahman, N.; Hamers, R. 1993. **Naturally occuring antibodies devoid of light chains.** *Nature.* **363**(6428): 446-448.

Hassanin, A.; Delsuc, F.; Ropiquet, A.; Hammer, C.; Jansen van Vuuren, B.; Matthee, C.; Ruiz-Garcia, M.; Catzeflis, F.; Areskoug, V.; Nguyen, T.T.; 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:** 32-50.

Hinrichs, K.; Buoen, L.C.; Ruth, G.R. 1999. **XX/XY chimerism and freemartinism in a female llama co-twin to a male.** *J Am Vet Med Assoc.* **215:** 1140-1141.

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:** 1503-1504.

Honey, J.G.; *et. al.* 1998. **Family Camelidae. In:** *The Evolution of Artiodactyls (Ed. Donald R. Prothero and Scott E Foss).* **14:** 177-178.

Horn, S. 2012. **Target enrichment via DNA hybridization capture.** *Methods Mol Biol.* **840:** 177-188.

Hornak, M.; Hulinska, P.; Musilova, P.; Kubickova, S.; Rubes, J. 2009. **Investigation of chromosome aneuploidies in early porcine embryos using comparative genomic hybridization.** *Cytogenet Genome Res.* **126:** 210-216.

Hsu, T.C. & Benirschke, K. 1967-1977. **An Atlas of Mammalian Chromosomes.** New York: Springer-Verlag. Volumes 1-10.

Hsu, T.C. & Benirschke, K. 1967. **An atlas of mammalian chromosomes.** New York: Springer-Verlag. **1:** folio 40.

Hsu, T.C. & Benirschke, K. 1974. **An atlas of mammalian chromosomes.** Berlin (Germany):Springer-Verlag. **1:** folio 389.

Huddleston, J.; Ranade, S.; Maliq, M.; Antonacci, F.; Chaisson, M.; Hon, L.; Sudmant, P.H.; Graves, T.A.; Alkan, C.; *et. al.* 2014. **Reconstructing complex regions of genomes using long-read sequencing technology.** *Genome Res.* **24**(4): 688-696.

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

Iannuzzi, L. & Di Berardino, D. 2008. **Tools of the trade: diagnostics and research in domestic animal cytogenetics.** *J Appl Genet.* **49**(4): 357-366.

Iannuzzi, A.; Perucatti, A.; Di Meo, G.P.; Schibler, L.; Incarnato, D.; Cribiu, E.P. 2003. **Chromosomal localization of sixty autosomal loci in sheep (*Ovis Aries*, 2n=54) by fluorescence *in situ* hybridization and R-banding.** *Cytogenet Genome Res.* **103**(1-2): 135-138.

Iannuzzi, A.; Perucatti, A.; Genualdo, V.; Pauciullo, A.; Incarnato, D.; Musilova, P.; Rubes, J.; Iannuzzi, C. 2014. **The utility of chromosome microdissection in clinical cytogenetics: a new reciprocal translocation in sheep.** *Cytogenet Genome Res.* **142:** 174-178.

ISCNDB: **International System for Chromosome Nomenclature of Domestic Bovids,** Cribiu, E.P. *et. al.* 2000. *Cytogenet Cell Genet.* **92**(3-4): 283-299.

ISCNH: **International System for Chromosome Nomenclature of the Domestic Horse.** Bowling, A.T. *et. al.* 1997. *Chromosome Res.* **5:** 433-443.

ISCN: **International System for Human Cytogenetic Nomenclature**. 1995. Basel (Switzerland), Karger.

Ji, R.; Cui, P.; Ding, F.; Geng, J.; Gao, H.; Zhang, H.; Yu, J.; Hu, S.; Meng, H. 2009. **Monophyletic origin of domestic bactrian camel (*Camelus bactrianus*) and its evolutionary relationship with the extant wild came (*Camelus bactrianus ferus*).** *Anim Genet.***40:** 377-382.

Johnson, W.E. 2011. **Alpaca comparative genomics: recent developments and future expectations.** *Conference on Camelid Genetics & Reproduction, September 15 – 18, Houston, Texas, USA.*

Jürgens, K.D.; Pietschman, M.; Yamaguchi, K.; Keinschmidt, T. 1988. **Oxygen binding properties, capillary densities, and heart weights in high altitude camelids.** *J Comp Physiol B.* **158:** 469-477.

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.** *Proc Biol Sci.* **268:** 2575-2584.

Kemkemer, C.; Kohn, M.; Cooper, D.N.; Froenicke, L.; Högel, J.; Hameister, H.; Kehrer-Sawatzki, H. 2009. **Gene synteny comparisons between different vertebrates provide new insights into breakage and fusion events during mammalian karyotype evolution.** *BMC Evol Biol.* **9:** 84.

Kleinschmidt, T.; Marz, J.; Jürgens, K.D.; Braunitzer, G. 1986. **The primary structure of two Tylopoda hemoglobins with oxygen affinity: Vicugna (*Lama vicugna*) and Alpaca (*Lama pacos*).** *Biological Chemistry Hoppe-Seyler.* **367:** 153-160.

Krallinger, H.F. 1931. **Cytologische Studien an einigen Haussaugetieren**. In: Mangold, E., Zorn, W. (Eds.). Archiv fur Tierernahrung und Tierzucht. Belrin, Verlage von Julius Springer, pp. 127-187.

Koulischer, L.; Tijskens, J.; Mortelmans, J. 1971. **Mammalian cytogenetics. IV. The chromosomes of two male Camelidae: *Camelus bactrianus* and *Lama vicugna*.** *Acta Zool Pathol Antverp.* **52:** 89-92.

Kubickova, S.; Cernohorska, H.; Musilova, P.; Rubes, J. 2002. **The use of laser microdissection for the preparation of chromosome-specific painting probes in farm animals.** *Chromosome Res* **10:** 571-577.

Kulemzina, A.I.; Trifonov, V.A.; Perelman, P.L.; Rubtsova, N.V.; Volobuev, V.; Ferguson-Smith, M.A.; Stanyon, R.: Yang, F.; Graphodatsky, A.S. 2009. **Cross-species chromosome painting in Cetartidactyla: reconstructin the karyotype evolution in key phylogenetic lineages.** *Chromosome Res.* **17:** 419-436.

Kulemzina, A.I.; Yang, F.; Trifonov, V.A.; Ryder, O.A.; Ferguson-Smith, M.A.; Graphodatsky, A.S. 2011. **Chromosome painting in Tragulidae facilitates the reconstruction of Ruminantia ancestral karyotype.** *Chromosome Res* **19:** 531-539.

Larkin, D.M.; Pape, G.; Donthu, R.; Auvil, L.; Welge, M.; Lewin, H.A. 2009. **Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories.** *Genome Res*. **19**(5): 770-777.

Lear, T.L. & Bailey, E. 2008. **Equine clinical cytogenetics: the past and future.** *Cytogenet Genome Res.* **120:** 42-49.

Lewin, H.A.; Larkin, D.M.; Pontius, J.; O'Brien, S.J. 2009. **Every genome sequence needs a good map.** *Genome Res.* **19:** 1925-1928.

Liu, G.E,; Hou, Y.; Zhu, B.; Cardone, M.F.; Jiang, L.; Cellamare, A.; Mitra, A.; Alexander, L.J.; *et. al.* 2010. **Analysis of copy number variations among diverse cattle breeds.** *Genome Res.* **20**(5): 693-703.

Maden, B.E.; Dent, C.L.; Farrell, T.E.; Garde, J.; McCallum, F.S.; Wakeman, J.A. 1987. **Clones of human ribosomal DNA containing the complete 18 S-rRNA and 28 S-rRNA genes. Characterization, a detailed map of the human ribosomal transcription unit and diversity among clones.** *Biochem J.* **246**(2): 519-527.

Mamanova, L.; Coffey, A.J.; Scott, C.E.; Kozarewa, I.; Turner, E.H.; Kumar, A.; Howard, E.; Shendure, J.; Turner, D.J. 2010. **Target-enrichment strategies for next-generation sequencing.** *Nat Methods* **7:** 111-118.

McCoy, L.E.; Quigley, A.F.; Strokappe, N.M.; Bulmer-Thomas, B.; *et al.* 2012. **Potent and broad neutralization of HIV-1 by a llama antibody elicited by immunization.** *J Exp Med*. **209**(6):1091-103.

Murphy, W.J.; Larkin, D.M.; Everts-van der Wind, A.; Bourque, G.; Tesler, G.; Auvil, L.; Beever, J.E.; Chowdhary, B.P.; Galibert, F.; Gatzke, L.; Hitte, C.; Meyers, S.N.; Milan, D.; Ostrander, E.A.; Pape, G.; Parker, H.G.; Raudsepp, T.; Rogatcheva, M.B.; Schook, L.B.; Skow, L.C.; Welge, M.; Womack, J.E.; O'Brien, S.J.; Pevzner, P.A.; Lewin, H.A. 2005. **Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps.** *Science.* **309:** 613-617.

Musilova, P.; Kubickova, S.; Vahala, J.; Rubes, J. 2013. **Subchromosomal karyotype evolution in Equidae.** *Chromosome Res.* **2:**175-87.

O'Brien, S.J.; *et. al.* 2006. **Atlas of mammalian chromosomes**. Wiley-Liss, John Wiley & Sons, Inc. Hoboken, NJ.

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.

Painter, T.S. 1923. **Further observations on the sex chromosomes of mammals.** *Science.* **58**(1500): 247-248.

Pearson, L.K.; Rodriguez, J.S.; Tibary, A. 2014. **Breeding management.** In: *Llama and Alpaca Care: Medicine, Surgery Reproduction, Nutrition and Herd Health (Ed. Christopher Cebra),* Saunders, Elsevier Inc., St. Louis, Missouri, USA, Chapter 16, pp. 162-168.

Perelman, P. 2011. **Second generation of alpaca radiation hybrid map: use of high-throughput genotyping.** *Conference on Camelid Genetics & Reproduction, September 15 – 18, Houston, Texas, USA.*

Raudsepp, T. 2012. **Camelid genomics and its applications in clinical cases.** *Proceedings from the International Camelid Health Conference for Veterinarians.* March 21-25, Columbus, Ohio, USA.

Raudsepp, T. 2014. **Cytogenetics and infertility.** In: *Llama and Alpaca Care: Medicine, Surgery Reproduction, Nutrition and Herd Health (Ed. Christopher Cebra),* Saunders, Elsevier Inc., St. Louis, Missouri, USA, Chapter 21, pp. 243-249.

Raudsepp, T. & Chowdhary, B.P. 1999. **Construction of chromosome-specific paints for meta- and submetacentric autosomes and the sex chromosomes in the horse and their use to detect homologous chromosomal segments in the donkey.** *Chromosome Res.* **7**(2): 103-114.

Raudsepp, T. & Chowdhary, B.P. 2008a. **FISH for mapping single copy genes.** *Methods Mol Biol.* **422:** 31-49.

Raudsepp, T. & Chowdhary, B.P. 2008b. **The horse pseudoautosomal region (PAR): characterization and comparison with the human, chimp and mouse PARs.** *Cytogenet Genome Res.* **121:** 102-109.

Raudsepp, T. & Chowdhary, B.P. 2011. **Cytogenetics and physical chromosome maps.** In: *The Genetics of the Pig (Eds. A. Ruvinsky and M. Rothschild)*, CAB International, Chapter 7, pp. 134-178.

Raudsepp, T. & Chowdhary, B.P. 2013. **Physical and comparative map.** In: *Equine Genomics (Ed. Bhanu P. Chowdhary),* John Wiley & Sons, Inc., Wiley-Blackwell, Oxford, UK, Chapter 3, pp. 49-72.

Raudsepp**,** T.; Mariat, D.; Guérin, G.; Chowdhary, B.P. 2001. **Comparative FISH mapping of 32 loci reveals new homologous regions between donkey and horse karyotypes.** *Cytogenet Cell Genet.* **94**(3-4):180-5.

Raudsepp, T.; Das, P.J.; Avila, F.; Chowdhary, B.P. 2011. **The pseudoautosomal region and sex chromosome aneuploidies in domestic species.** *Sex Dev.* **6**(1-3):72-83.

Raudsepp, T.; Durkin, K.; Lear, T.L.; Das, P.J.; Avila, F.; Kachroo, P.; Chowdhary, B.P. 2010. **Molecular heterogeneity of XY sex reversal in horses.** *Anim Genet.* **2:** 41-52.

Raudsepp, T.; Gustafson-Seabury, A.; Durkin, K.; Wagner, M.L.; Goh, G.; Seabury, C.M.; Brinkmeyer-Langford, C.; Lee, E.J.; Agarwala, R.; Stallknecht-Rice, E.; *et al.*, 2008. **A 4,103 marker integrated physical and comparative map of the horse genome.** *Cytogenet Genome Res.* **122:** 28-36.

Raudsepp, T.; Lee, E.J.; Kata, S.R.; Brinkmeyer, C.; Mickelson, J.R.; Skow, L.C.; Womack, J.E.; Chowdhary, B.P. 2004. **Exceptional conservation of horse-human gene order on X chromosome revealed by high-resolution radiation hybrid mapping.** *Proc Natl Acad Sci U S A.* **101**(8): 2386-2391.

Rens, W.; Fu, B.; O'Brien, P.C.; Ferguson-Smith, M. 2006. **Cross-species chromosome painting.** *Nat Protoc.* **1:** 783-790.

Rothschild, M.F. & Ruvinsky, A. **The genetics of the pig.** Oxfordshire, UK, CABI, pp. 134-178.

Ross, M.T.; Grafham, D.V.; Coffey, A.J.; Scherer, S.; McLay, K.; Muzny, D.; Platzer, M.; Howell, G.R.; Burrows, C.; Bird, C.P.; *et al.*, 2005. **The DNA sequence of the human X chromosome.** *Nature.* **434:** 325-337.

Rozen, S. & Skaletsky, H. 1998. **Primer3.** Code available at http://www.genome.wi.mit.edu/genome_software/other/primer3.html.

Rubes, J.; Musilova, P.; Kopecna, O.; Kubickova, S.; Cernohorska, H. 2012. **Comparative molecular cytogenetics in Cetartiodactyla.** *Cytogenet Genome Res.* **137:** 194-207.

Rubes, J.; Pinton, A.; Bonnet-Garnier, A.; Fillon, V.; Musilova, P.; Michalova, K.; Kubickova, S.; Ducos, A.; Yerle, M. 2009. **Fluorescence in situ hybridization applied to domestic animal cytogenetics.** *Cytogenet Genome Res.* **126:** 34-48.

Saitone, T.L. & Sexton, R.J. 2012. **The alpaca bubble revisted**. In: *Review of Agricultural Economics Update.* University of California, Giannini Foundation of Agricultural Economics. **15**(5): 9-11.

Samman, M.A.; Al-Saleh, A.A.; Sheth, K. 1992. **The karyotype of the Arabian camel, *Camelus dromedarius*.** *Jornal of King Saud University, Science.* **5**(1).

Scherthan, H.; Cremer, T.; Arnason, U.; Weier, H.U.; Lima-de-Faria, A.; Fronicke, L. 1994. **Comparative chromosome painting discloses homologous segments in distantly related mammals.** *Nat Genet.* **6:** 342-347.

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:** 77-85.

Schreck, R.R. & Distèche, C.M. 2001. **Chromosome banding techniques.** *Curr Protoc Hum Genet.* **4:** 4.2.

Seabright, M. 1971. **A rapid banding technique for human chromosomes.** *Lancet.* **2:** 971-972.

Skaletsky, H.; Kuroda-Kawaguchi, T.; Minx, P.J.; Cordum, H.S.; Hillier, L.; Brown, L.G.; Repping, S.; Pyntikova, T.; Ali, J.; Bieri, T.; *et al.*, 2003. **The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes.** *Nature.* **423:** 825-837.

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:** 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*.** *Proc R Soc Lond.* **266:** 649-656.

Sørensen, K.M.; Jespersgaard,C.; Vuust, J.; Hougaard, D.; Nørgaard-Pedersen, B.; Andersen, P.S. 2007. **Whole genome amplification on DNA from filter paper blood spot samples: an evaluation of selected systems.** *Genet Test.* **11**(1): 65-71.

Speicher, M.R. & Carter, N.P. 2005. **The new cytogenetics: blurring the boundaries with molecular biology.** *Nat Rev Genet.* **6**(10): 782-792.

Stanley, H.F.; Kadwell, M.; Wheeler, J.C. 1994. **Molecular evolution of the family Camelidae: a mitochondrial DNA study.** *Proc Biol Sci.* **256:** 1-6.

Stanyon, R.; Rocchi, M.; Bigoni, F.; Archidiacono, N. 2012. **Evolutionary molecular cytogenetics of caterrhine primates: past, present and future.** *Cytogenet Genome Res.* **137:** 273-284.

Strokappe, N.; Szynol, A.; Aasa-Chapman, M.; Gorlani, A.; *et al*. 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): 1-11.

Svartman, M.; Stone, G.; Stanyon R. 2006. **The ancestral eutherian karyotype present in Xenarthra.** *PLoS Genet.* **2**(7): 109.

Taberlet, P.; Coissac, E.; Pansu, J.; Pompanon, F. 2011. **Conservation genetics of cattle, sheep, and goats.** *C R Biol.* **334:** 247-254.

Taylor, K.M.; Hungerford, D.A.; Snyder, R.L.; Ulmer, F.A. Jr. 1968. **Uniformity of karyotypes in the Camelidae.** *Cytogenetics.* **7:** 8-15.

Telenius, H.; Carter, N.P.; Bebb, C.E.; Nordenskjold, M.; Ponder, B.A.; Tunnacliffe, A. 1992. **Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer.** *Genomics.* **13:** 718-725.

Tibary, A. 2008. **Reproductive disorders in alpacas and llamas.** Proceedings of the 1st International Workshop on Camelid Genetics, February 22-24, Scottsdale, Arizona, USA.

Tijo, J.H. & Levan, A. 1956. **The chromosome number of man.** *Hereditas.* **42:** 1-6.

Trask, B.J. 2002. **Human cytogenetics: 46 chromosomes, 46 years and counting.** *Nat Rev Genet.* **3:** 769-778.

Trifonov, V.A.; Stanyon, R.; Nesterenko, A.I.; Fu, B.; Perelmen, P.L.; O'Brien, P.C.; Stone, G.; Rubtsova, N.V.; Houck, M.L.; Robinson, T.J.; Ferguson-Smith, M.A.; Dobigny, G.; Graphodatcky, A.S.; Yang, F. 2008. **Multidirectional cross-species painting illuminates the history of karyotypic evolution in Perissodactyla.** *Chromosome Res.* **16:** 89-107.

Tornquist, S.J. & Cebra, C. 2014. **Clinical pathology.** In: *Llama and Alpaca Care: Medicine, Surgery Reproduction, Nutrition and Herd Health (Ed. Christopher Cebra),* Saunders, Elsevier Inc., St. Louis, Missouri, USA, Chapter 31, pp. 338-350.

Vidal-Rioja, L.; Larramendy, M.L.; Semorile, L. 1989. **Ag-NOR staining and in situ hybridization of rDNA in the chromosomes of the South American camelids.** *Genetica.* **79:** 215-222.

Villagómez, D.A.; Parma, P.; Radi, O.; Di Meo, G.; Pinton, A.; Iannuzzi, L.; King, W.A. 2009. **Classical and molecular cytogenetics of disorders of sex development in domestic animals.** *Cytogenet Genome Res*. **126:** 110-131.

Villagómez, D.A. & Pinton, A. 2008. **Chromosomal abnormalities, meiotic behavior and fertility in domestic animals.** *Cytogenet Genome Res*. **120:** 69-80.

Vujosević, M. & Blagojevic, J. 2004. **B chromosomes in populations of mammals.** *Cytogenet Genome Res.* **106:** 247-256.

Wade, C.M.; Giulotto, E.; Sigurdsson, S.; Zoli, M.; Gnerre, S.; Imsland, F.; Lear, T.L.; Adelson, D.L. Bailey, E.; Bellone, R.R; *et. al.* **Genome sequence, comparative analysis, and population genetics of the domestic horse.** *Science.* **326**(5954): 865-867.

Wang, C.; Deakin, J.E.; Rens, W.; Zenger, K.R.; Belov, K.; Graves, J.A.M.; Nicholas, F.W. 2011. **A first-generation integrated tammar wallaby map and its use in creating a tammar wallaby first-generation virtual genome map.** *BMC Genomics.* **12:** 422.

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.

Wheeler, J.C. 1984. **On the origin and early development of camelid pastoralism in the Andes.** In: *Animals and archaeology: Vol. 3, early herders and their flocks (Eds: Clutton-Brock, J & Grigson, C).* Oxford: BAR International Series **202:** 395-410.

Wheeler, J.C. 1995. **Evolution and present situation of the South American Camelidae.** *Biol J Linn Lond.* **54:** 271-295.

Wheeler, J.C. 2006 **Historia natural de la Vicuña.** Vilá B (ed) Investigación. *Conservatión y Manejo de Vicuñas.* Proyecto MACS. pp. 25-36.

Wilker, C.E.; Meyers-Wallen, V.N.; Schlafer, D.H.; Dykes, N.L.; Kovacs, A.; Ball, B.A.; 1994. **XX sex reversal in a llama.** *J Am Vet Med Assoc.* **204:** 112-115.

Wurster, D.H. & Benirschke, K. 1970. **Indian momtjac, *Muntiacus muntiak*: a deer with a low diploid chromosome number.** *Science.* **168**(3937): 1364-1366.

Yang, Y.P. & Womack, J.E. **Parallel radiation hybrid mapping: a powerful tool for high-resolution genomic comparison.** *Genome Res.* **7:**731-6.

Yang, F.; Fu, B.; O'Brien, P.C.; Nie, W.; Ryder, O.A.; Ferguson-Smith, M.A. 2004. **Refined genome-wide comparative map of the domestic horse, donkey and human based on cross-species chromosome painting: insight into the occasional fertility of mules.** *Chromosome Res.***12:** 65-67.

Zhang, Q.L.; Dong, C.S.; He, J.P.; He, X.Y.; Fan, R.W.; Geng, J.J.; Ren, Y.H. 2005. **Study on the chromosomal karyotype and G-banding of alpacas (*Lama pacos*).** *Yi Chuan.* **27:** 221-226.

Zhou, X.; Xu, S.; Yang, Y.; Zhou, K.; Yang, G. 2011. **Phlyogenomic analyses and improved resolution of Cetartiodactyla.** *Mol Phylogenet Evol.* **61**(2): 255-264.