GENOMIC ANALYSIS OF THE HORSE Y CHROMOSOME

A Dissertation

by

AVNI BHAWAN SANTANI

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

DOCTOR OF PHILOSOPHY

December 2004

Major Subject: Genetics

GENOMIC ANALYSIS OF THE HORSE Y CHROMOSOME

A Dissertation

by

AVNI BHAWAN SANTANI

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Bhanu Chowdhary (Chair of Committee)	Loren Skow (Member)
James Womack (Member)	Dickson Varner (Member)
Geoffrey Kapler (Chair of Genetics Faculty)	Evelyn Castiglioni (Head of Department)

December 2004

Major Subject: Genetics

ABSTRACT

Genomic Analysis of the Horse Y Chromosome. (December 2004)

Avni Bhawan Santani, B.S., University of Mumbai;

M.S., University of Aberdeen

Chair of Advisory Committee: Dr. Bhanu P. Chowdhary

Stallion fertility is of significant economic importance in the multibillion dollar equine industry. Presently, the underlying genetic causes of infertility in stallions are unknown. Analysis of the human genome has shown that in more than 25% of cases, male infertility is associated with deletions/rearrangements in the Y chromosome. Presently there is no gene map for the Y chromosome in the horse. Therefore, the primary aim of this study is to build a detailed physical map of the chromosome with a long-term aim to identify and analyze Y-specific factors affecting fertility in stallions.

To materialize this, we constructed the first radiation hybrid and FISH map of the euchromatic region of the horse Y chromosome. This basic map was used to obtain Y-specific BAC clones that provided new STS markers from the end sequences. Chromosome walking provided 73 BACs comprising 7 contigs that were built across the euchromatic region using 124 markers for content mapping. The results were validated by restriction fingerprinting and Fiber FISH. The map is presently the most informative among the domestic species and second to only human and mouse Y chromosome maps.

The construction of this map will pave the way for isolation and functional characterization of genes critical for normal male fertility and reproduction and will in the future lead to the development of a diagnostic test to facilitate early identification of deletions/rearrangements on the Y chromosome of potentially affected foals/stallions.

The second part of the study comprised the first extended investigation to assess genetic variation in the horse Y chromosome. Approximately 4.5Mb of the euchromatic region was screened for polymorphic microsatellite markers. Of the 27 markers that were characterized and screened for polymorphism in 14 breeds of the domestic horse and eight extant equids, only one was polymorphic in the domestic horse, suggesting a low level of genetic variation on the chromosome. However, 21 of the markers showed noteworthy variation (on average four alleles/marker) among the eight equids. These markers will be vital in future studies aimed at elucidating the genetic relationships between the various equids through phylogenetic analysis.

DEDICATION

This work is dedicated to my mother, Mrs. Rekha Bhawan Santani, a strong and gentle soul who taught me that so much could be done with little, and whose laughter and sense of fairness made the world a better place. You are a part of every page, every thought.

ACKNOWLEDGMENTS

Many people have been a part of my graduate education, as friends, teachers, and colleagues. Dr. Bhanu Chowdhary, first and foremost, has been all of these. The best advisor that I could have asked for, he is actively involved with his students, and has their best interests in mind. Thank you for your enthusiastic supervision, guidance and support throughout the duration of this project and for withstanding my constant interruptions in your office with a smile. Thank you for also keeping an endless supply of freshly baked bread and inviting me over to your home for discussions over cups of *chai* and home-made food. Also, for stimulating discussions and help I acknowledge the members of my graduate committee, Dr. James Womack, Dr. Loren Skow and Dr. Dickson Varner.

This work would have not been possible without the immense help and guidance of Dr. Terje Raudsepp, the unofficial co-advisor for all my projects.

For the mapping project, I gratefully acknowledge the contributions of co-authors Dr. Loren Skow, Dr. Barbara Wallner, Dr. Hong-Bin Zhang and Dr. Cheng-wei Ren. I would also like to recognize the assistance and support of Elaine Owens, Dee Honeycutt and Ashley Gustafson. I would also like to thank Dr. Varner for productive discussions on stallion fertility and for providing blood samples from subfertile stallions.

For the Y microsatellite project, I gratefully acknowledge the help of Dr. Loren Skow, Dr. Clare Gill and Colette Abbey for technical assistance with construction of subclone libraries. I would especially like to thank Colette who has spent countless days

providing assistance and support for this project. I thank Dr. James Derr for advice with genotyping analysis. Thanks to Dr. Gus Cothran, Dr. Oliver Ryder and Leona Chemnick for providing DNA samples of equids. Thanks to the DNA Technologies Core Lab of the Department of Veterinary Pathobiology for technical assistance.

For the past three years I have shared office space with several graduate students and scientists who have contributed to this work and to whom I am grateful. Dr. Eun-Joon Lee, Candice Brinkmeyer, Chris Childers, Glenda Goh, Keith Durkin, and Caroline Kotzot were an incredible group with whom I had several productive discussions, friendly debate and laughter over barbecue and salad. A special thanks to our resident computer wizard Chris Childers for withstanding my extreme ignorance about bio-informatics with endless patience, understanding, and humor. In addition to the people at Texas A&M University, I have been lucky enough to have the support of many good friends. I am grateful to all my friends in the United States, UK, and India for providing moral support and encouragement, especially, Rohan, for giving me the strength to pursue my goals and Sid and Anjali for their faith in me.

A special thanks to Dr. James Womack and Dr. Loren Skow for the delightfully insightful nicknames that they bestowed upon me at regular intervals throughout my graduate career.

Finally, I would like to thank those closest to me, whose presence helped make the completion of my graduate work possible. Thanks to Pranaya Kondekar, my best friend, for pancakes at two in the morning, proof reading my dissertation a million times, cooking countless meals and being a pillar of support through all the rough times. A

special thanks to Neeraj, for enduring hair-raising driving lessons and my temper tantrums with supreme calm and patience; also for unconditional love and support. Most of all, I would like to thank my family, those present, and those who have passed away. Papa, whom I have adored and idolized since birth. Ba and Bapuji, my fantastic grandparents who have taught me so much. Most of all I would like to thank my mother, and brother Dev, who risked financial security in order to provide education for me. For patiently enduring the longest separation in 22 years and for their absolute confidence in me….the knowledge that they will always be there to pick up the pieces is what allowed me to venture beyond the safety harness of home.

Look back at man's struggle for freedom,

Trace his present day strength to its source,

And you'll find that his pathway to glory

Is strewn with the bones of the horse.

-- Anonymous

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

STALLION FERTILITY

A 1996 report by the American Horse Council showed that the number of horses in the United States increased by 47% (to 6.9 million horses) during the previous two decades. It was evident that annually, the horse industry had a total impact of $112.1 billion on the country's gross domestic product, with the breeding industry contributing $6.42 billion (http://www.horsecouncil.org/statistics.htm). This industry ranked fifth and was above the national apparel industry. In many states like California, Texas, Kentucky, Indiana, Pennsylvania, Minnesota, etc., the economic impact of the equine industry is considerable. For example, in Pennsylvania, the total economic impact during the year 2002 was estimated to be $1 billion dollars (http://www.wiwfarm.com/pdf/PAEquineSurveyResults-Academic.pdf). A 1998 report by Texas A&M University states that the total economic impact of the Texas horse industry amounts to more than $11 billion dollars annually (http://animalscience.tamu.edu/ansc/publications/horsepubs/hrg014-industry.pdf). These statistics demonstrate that nationwide, the equine industry is a viable enterprise that strongly influences the economy at various levels.

Due to a constantly increasing demand for top quality breeding stock worldwide, a number of stallions enter the ranks of breeding sires annually. These stallions are oftentimes not subjected to a detailed breeding soundness examination prior to standing

---

This dissertation follows the style of *Genome Research*.

at stud (Colenbrander et al. 2003). These examinations traditionally entail semen analysis, examination of external genitalia and a general physical evaluation.

It is only later when desired foaling rate is not achieved that the stallion's fertility records are thoroughly examined and it becomes apparent that there is some fertility problem. In addition, it is commonplace for stallions to be insured for first-year congenital infertility insurance coverage without a detailed breeding soundness examination. In essence, lack of adequate examinations to critically evaluate fertility of stallions destined for long-term breeding is risky because it could lead to huge losses connected to hundreds of thousands of dollars in stud fees and income associated with the foal-crop.

Stallion fertility is usually predicted primarily by direct means like through the analysis of various semen parameters that may include sperm number, sperm motility and sperm morphology (Graham 1996; Madill 2002; Colenbrander et al. 2003; Love et al. 2003). In-depth assays such as the sperm chromatin structure assay and the acrosomal responsiveness assay, are being used with increasing frequency to more critically evaluate sperm functionality (Love and Kenney 1998; Varner et al. 2000). Indirect measures like per cycle pregnancy rate are also used to predict stallion fertility (Colenbrander et al. 2003; Love et al. 2003). But these factors are influenced by management of both mares and stallions (Rousset et al. 1987; Love et al. 2003). For example, inadequate management practices, such as improper teasing of mares, inadequate management of reproductive tract disease, inaccurate time of ovulation, can

influence the overall fertility of a stallion simply by reduced mare fertility or suboptimal management factors (Dowsett and Pattie 1982; van Buiten et al. 1998; Morris and Allen 2002; Colenbrander et al. 2003). Since, such factors are not indicative of the *actual* fertility of the stallion, they may easily distort evaluations. Despite availability of a range of direct and indirect approaches to characterize and evaluate fertility, no single test effectively and comprehensively determines fertility in stallions. The reason for this could be - for example with regards to semen quality – the fact that the variety of assays used for drawing conclusions is restricted to only a small sample of spermatozoa, which may not be completely representative to detect some of the compartmental defects. Due to this complexity and lack of reliability in judgment, some breeders tend to overlook laboratory based fertility tests. Moreover, time and expense associated with the entire procedure are yet other factors that may deter breeders from choosing systematic fertility evaluations of their stallions (Hurtgen 1992; Graham 1996).

As mentioned above, defining and measuring fertility in stallions is indeed difficult. Nevertheless, attempts are ongoing to catalogue and analyze all likely factors that could influence fertility. These include variations which result in phenotypes ranging from normal fertile individuals to those that may be completely sterile. Additionally, internal factors such as size and position of testicles and external factors including, influence of administered hormones, injury, inflammation, age, season, environmental temperature, drugs (steroids, antibiotics, etc) also impact the overall assessment of a prospective breeding stallion (Varner et al. 2000; Madill 2002; Colenbrander et al. 2003; Love et al. 2003). Improved stallion and mare management

has contributed to an improvement in breeding capacities of stallions, in some cases for even those with relatively low/reduced fertility. However the downside to these advancements has been the proliferation of fertility problems, especially if it is passed over to the next generation through a 'defective' gene pool. A classical example for this is that horsemen tend to select primarily for traits such as athletic ability and conformation. Due to this, they oftentimes inadvertently disregard reproductive potential of a stallion. Hence certain fertility defects (e.g. undescended or underdeveloped/small testicles, etc.) are carried along the sire lines, thus increasing the probability of a pool of inherently subfertile horses. This point of view generates support from the fact that wild horses have higher fertility than horses kept in captivity (McDonnell 2000) possibly because nature self limits those with inherent reproductive problems.

Fertility is a highly complex trait that is governed by a range of factors. Traditionally, in horses, these factors are largely considered to be environmental and those related to management regimes. However, one important component that is key to the regulation and expression of fertility potential- the genetic factor – has largely been ignored. This is primarily because the resources required for study of this component in horses, have been quite limited. Studies in humans and mice have clearly demonstrated a strong influence of genetic factors in regulating fertility (Mahadevaiah et al. 1998; Kuroki et al. 1999; Kuroda-Kawaguchi et al. 2001; Krausz et al. 2003; Kujawski et al. 2004). Similarly, research carried out in pigs and cattle show that an improved understanding of factors governing fertility can be obtained by identifying the genes

regulating fertility and analyzing their individual or joint functions (Rothschild et al. 1997; Liefers et al. 2002; Ashwell et al. 2004).

The basic ingredient for analyzing the *genetic component* of fertility is establishment of a gene map. A map is essential for identifying and localizing genes on chromosomes. Using these maps as a platform, investigations can be conducted to catalogue and analyze genes regulating important traits like fertility that are governed in a quantitative or cumulative manner. For example, high resolution gene maps of all autosomes and the sex chromosomes in humans and mice have led to the identification of a number of genes associated with fertility and pathways to sex determination (Mahadevaiah et al. 1998; Mazeyrat et al. 1998; Kuroda-Kawaguchi et al. 2001; Skaletsky et al. 2003). However, as far as known, no organized research has been conducted in horses to identify the possible genetic mechanisms governing stallion infertility or subfertility. One important reason for this is that the horse gene map has been inadequately developed. Consequently very little is known about genes that may be involved in regulating fertility and determination of phenotypic sex in the horse.

The genes governing fertility are distributed on autosomes as well as the sex chromosomes. Among the sex chromosomes, in females, these genes are located only on the X chromosomes while in males these genes are present both on the X and the Y chromosomes. Though rapid progress has been made during recent years to develop gene maps for autosomes and X chromosome in the horse (Chowdhary et al. 2003; Raudsepp et al. 2004a), no gene map for the Y chromosome has yet been generated. A similar situation exists in other domestic species also. This lack of information on the Y

chromosome is partly attributed to the long standing misconception that the chromosome is extremely gene poor, with male sex determination as its only primary function. However extensive research in humans and mice clearly shows that in addition to genes on the autosomes and X chromosome, genes located on the Y chromosome play an important role in regulating male fertility (Krausz and McElreavey 1999; Foresta et al. 2001; Fox and Reijo Pera 2001; Kuroda-Kawaguchi et al. 2001). Realizing the significance of this, we embarked on developing a detailed map of the euchromatic region of the horse Y chromosome. The project thus marks the first systematic effort to generate a complete map of the horse Y chromosome that will lead to discovery of genes regulating stallion fertility and those involved in the pathway towards sex determination.

The primary aim of this project is to assemble a physical map of the horse Y chromosome (ECAY) by using a combination of mapping approaches such as radiation hybrid mapping, STS content mapping, FISH mapping etc. At the conclusion of this project, a detailed map of the horse Y chromosome will be built that will lay the foundations for initiating studies aimed at identifying male fertility genes that reside on this chromosome. The information generated through this work will also facilitate studies aimed at examining interactions between Y specific genes and other fertility genes on the autosomes and X chromosome. In the long run, the information thus generated will be key for the development of advanced diagnostic and therapeutic options that will help to identify and manage stallions with fertility problems.

The following paragraphs provide an overview of research carried out to date to analyze the Y chromosome in various mammalian species. The summary also provides

how this analysis led in some species (primarily humans and mice) to the identification of genes that play an important role in regulating male fertility. This information is critical in understanding that in addition to genes on the autosomes and the sex chromosomes, male specific genes on the Y chromosome of other species also need to be studied in greater details to facilitate identification of Y-specific causes of reduced fertility in males.

OVERVIEW OF Y CHROMOSOME RESEARCH IN DIFFERENT MAMMALS

Analysis of the human Y chromosome

Among mammals, the human Y chromosome has been extensively studied. One of the first studies showing correlation between the human Y chromosome and infertility was reported almost three decades ago. Deletions involving the entire distal heterochromatic region of the long arm of the Y chromosome was consistently observed in several infertile men (Tiepolo and Zuffardi 1976). Based on these findings, it was hypothesized that certain genetic factors, termed as the Azoospermia factors (AZF), controlling spermatogenesis reside on the long arm of the chromosome (Tiepolo and Zuffardi 1976). The discovery of genes like ZFY, that was initially thought to be the testis determining factor (Page et al. 1987), and more importantly SRY (sex determining region on the Y; (Sinclair et al. 1990) were critical in initiating studies that led to generation of a gene map for this chromosome.

One of the early significant results in humans include generation and use of 76 Yq loci for deletions in 370 azoospermic/severly oligospermic men. The results led to identification of three non overlapping regions of Yq that were deleted in men suffering

from partial or complete infertility. These regions were termed as AZFa, AZFb, and AZFc (Vogt et al. 1996). Interestingly, each of the regions contains candidate genes related to male infertility. Since their discovery, the three regions have been extensively studied in terms of association with male infertility. For example, mutations in the genes from the AZFa region (USP9Y and DBY) were believed to result in spermatogenic failure (Sargent et al. 1999; Sun et al. 1999; Foresta et al. 2000). Mutations in the RBMY gene family residing in the AZFb region were considered as a candidate for spermatogenic failure. The gene demonstrated testis specific expression and was recurrently deleted in azoospermic men. It is homologous to the mouse *Rbm* gene, mutations in which causes spermatogenic arrest (Elliott et al. 1997; Mahadevaiah et al. 1998). Lastly, deletions of the AZFc regions are found to be a major cause of spermatogenic failure with *de novo* deletions arising in around 1 in 4000 males (Raicu et al. 2003).

The main candidate gene in the AZFc region is the DAZ (Deleted in Azoospermia) cluster, a set of genes transcribed in the testis and expressed exclusively in germ cells (Reijo et al. 1995; Saxena et al. 1996; Menke et al. 1997). The role of the human DAZ gene was proven by the observation that a human DAZ transgene was capable of partially rescuing the sterile phenotype of a mouse knockout for the homologous gene *Dazl* (Slee et al. 1999). This likelihood was further strengthened by the high homology of human DAZ with a Drosophila male infertility gene, *boule*, mutation of which causes spermatogenic arrest (Burgoyne 1996; Eberhart et al. 1996).

One of the important breakthroughs in understanding the role of Y specific factors in regulating male fertility in humans came from a deletion map wherein 200 sequence tagged sites (STSs) uniformly distributed over the Y chromosome were used to identify deletions and rearrangements (Vergnaud et al. 1986). This map divided the Y chromosome into seven deletion intervals. Subsequently the map was further refined and expanded to include 43 sub intervals (Vollrath et al. 1992). In addition, clones from a human XYYYY male YAC library were screened with Y specific STSs to identify YACs containing corresponding sequences. This ultimately led to the development of the first physical map of the human Y chromosome with 196 overlapping YAC clones (Foote et al. 1992). The map later provided an important framework for obtaining the complete sequence of the human Y chromosome (Foote et al. 1992; Skaletsky et al. 2003).

Sequence analysis of the male specific region of the human Y chromosome (MSY) revealed presence of 156 transcription units, of which 78 were protein coding (Skaletsky et al. 2003). Of great interest was the identification of gene families present in several copies scattered throughout the euchromatic region of the chromosome. Out of the nine multicopy gene families, eight had members located in palindromes (a palindrome is a sequence of nucleotides that reads the same in either direction e.g. GAATTC). It was noteworthy that the palindromes encompassed AZF regions that have been found to show massive deletions (6.2 – 7.7Mb) leading to spermatogenic failure (Repping et al. 2002).

The rapid expansion of knowledge about the structure and organization of the human Y chromosome combined with the availability of a high resolution STS map has made detection of molecular lesions on this chromosome relatively easy. Analysis by PCR is now routinely performed on this collection of STSs by laboratories worldwide to identify microdeletions associated with a range of male-specific subfertile and infertile phenotypes (Vergnaud et al. 1986; Dada et al. 2003). The high frequency of *de novo* Y chromosome deletions in humans indicates the susceptibility of this chromosome to spontaneous loss of genetic material.

Involvement of the Y chromosome in other conditions such as cancer, aging, and sex reversal has been reported. For example, frequent loss of the complete Y chromosome has been observed in squamous cell carcinomas (Kujawski et al. 2004) and esophageal cancer (Hunter et al. 1993). Gene mapping studies have identified a gonadoblastoma locus on the Y chromosome that causes testicular cancer and tumorigenesis in gonads of XY sex reversed patients (Page 1987; Lau et al. 2000; Lau et al. 2003). A recent study suggested the role of RBMY, a Y chromosome specific locus, in causing male liver cancer (Tsuei et al. 2004).

Other instances where the involvement of the Y chromosome is noted to cause a variety of health related conditions are as follows: complete loss of Y chromosome and deleterious mutations on it have been found to be associated with reduced life span of males compared to females (Stone and Sandberg 1995; Cortopassi 2002). Further, mutations in the SRY gene has been associated with sex reversal (sex reversal is the difference between chromosomal sex and phenotypic sex, e.g. XX males, XY females)

and gonadal dysgenesis (a syndrome characterized by "streak gonads" in a phenotypic female with a 46,XY karyotype) (Shahid et al. 2004). Lastly, mosaicism of Y; variation in the number of chromosomes in the body's cells where certain percentage of cells contain a different karyotype such as 46XX, 46XY has been reported in several patients suffering from Turner's syndrome (45XO karyotype) (Hook and Warburton 1983; Nagafuchi et al. 1992) and it is believed that the Y chromosome or the genes therein have some role to play.

Analysis of the mouse Y chromosome

Following humans, the only other mammalian species where the Y chromosome has been analyzed is the mouse. Efforts are underway to obtain the complete sequence of the mouse Y chromosome. Interesting parallels have been drawn while comparing the structure and organization of Y chromosome between humans and mouse. While the euchromatic region on the mouse Y chromosome is condensed on the short arm the human euchromatic region is more expanded (Bishop and Mitchell 1997). Moreover, there are differences in the gene content of the pseudoautosomal region (Ellison et al. 1996; Gianfrancesco et al. 2001; Perry et al. 2001). Lastly, some genes such as DAZ that are Y specific in one species are autosomal in the other (Disteche et al. 1992; Ellison et al. 1996; Saxena et al. 1996; Perry et al. 2001) .

Mouse spermatogenesis is broadly comparable to human spermatogenesis and therefore mouse is presently the primary model for expression studies of Y related genes causing azoospermia. Deletion in the *Sxr*^{b} region from Yp, which is homologous to AZFa in humans (Mazeyrat et al. 1998) prevents survival of spermatogonia (Sutcliffe

and Burgoyne 1989). Further, deletion of the *Rbm* gene in mouse, which bears a strong homology to the human RMBY gene (Elliott et al. 1996), has been associated with male sterility (Laval et al. 1995; Mahadevaiah et al. 1998). Next, the mouse protein *PL10* bears a strong homology to DBY and its expression is testis specific (Foresta et al. 2000). Comparative mapping studies between human and mouse have demonstrated the presence of *Dby* and *Uty* genes in the Sxr^b interval (Mazeyrat et al. 1998). The Ssty (spermatid-specific transcripts) gene in mouse has testis specific expression, has more than 100 copies on the Y chromosome, and plays an important role in spermatid differentiation (Conway et al. 1994). Consistent with its expression/role, Yq deletions in mouse are associated with lack of protein and consequently sperm abnormalities (Styrna et al. 1991a; Styrna et al. 1991b; Burgoyne et al. 1992). Thus, in close association with the rapid development of the Y chromosome map in the mouse, the function(s) of the male specific genes are being deciphered, and their role in regulating male fertility is being understood.

Analysis of the Y chromosome in domestic species

As mentioned earlier, the study of the structure and organization of the mammalian Y chromosome was limited primarily to human and mouse. The unusually high heterochromatic content of the Y chromosome in most species led numerous researchers to believe that the Y chromosome is very gene poor, hence not worth investigating. It therefore comes as no surprise that even until late 1990s, no planned research was initiated to analyze the Y chromosome in any of the domesticated species. Consequently, we know almost nothing about the organization and function of the genes

of the Y chromosome in domestic animals, which is in stark contrast to the development in humans and mice during recent years.

In order to understand the complexity and unusual structure of the Y chromosome in mammals, a closer look at the evolution of the sex chromosomes in mammals will be useful. This will also help to understand why, comparative mapping tools were unsuccessful in determining the gene identity and content of the Y chromosome in different species. Although X and Y chromosomes are postulated to have evolved from a homologous pair of autosomes (Ohno 1967), the Y has adopted an independent pathway in evolution. Compared to the X chromosome which is highly conserved across species (Charlesworth 1991; Murphy et al. 1999; Raudsepp et al. 2004a), Y chromosomes have undergone extensive degeneration, heterochomatinization, duplication, transposition, inversion and loss of gene content (Graves 1995; Graves 1998; Graves 2001; Waters et al. 2001; Skaletsky et al. 2003). Hence the likelihood of Y-sequence homology between species is negligible and comparative mapping is most likely not a viable alternative.

Early evidence for this effect was provided by ZOO-FISH studies that indicated lack of homology for this chromosome across most mammals. In contrast to other chromosomes, the Y chromosome specific heterologous painting probes failed to work in other species. Among the few exceptions to this phenomenon were the evolutionarily closely related species; within primates and equids (Wienberg et al. 1992; Koehler et al. 1995; Solinas-Toldo et al. 1995; Richard et al. 1996; Chowdhary et al. 1998; Muller and Wienberg 2001). This lack of homology has therefore been the major factor in

preventing transfer of information on the Y chromosome across species, for example from humans/mice to the domesticated animals.

During recent years, efforts have been made to analyze the Y chromosome in cat, pig, dog and cattle. In most species these efforts have been restricted to the mapping of only a few genes. As far as known, no large scale efforts have as yet been initiated to generate a detailed map of the Y chromosome in any of the species. In the following paragraphs, a summary of Y chromosome research in the domesticated species is presented.

Cat

Presently only a basic radiation hybrid (RH) map is available for the Y chromosome in the cat (Murphy et al. 1999). The map comprises of eight genes and provides a preliminary comparison with the existing human and mouse Y chromosome maps. Conserved physical order was found for three of the loci across all three species. Three of the feline loci (USP9Y, DBY, and UTY) primarily correspond to AZFa in human and *Sxr*^{b} in mouse (deletions in both are known to block spermatogenesis). In addition to mapping, genetic variation in ZFX/ZFY genes was examined in 34 felid species. Comparative sequencing showed that Y chromosome in felids is less conserved than the X chromosome with greater mutation rate in males (Pecon Slattery and O'Brien 1998). Analyses of X-Y homologous genes in 26 felid species revealed two ancestral episodes of gene conversion from X to Y (Pecon Slattery et al. 2000). It was postulated that ectopic gene conversion may be a mechanism to prevent degeneration of genes that reside in the non-recombining region of the Y chromosome. Additionally introns of three

Y-specific genes were examined in 20 felid species to study pattern of SINE distribution (Slattery et al. 2000). The absence of SINEs in the ZFX homolog of all felids is in agreement with the theory that repetitive DNA is predominant in regions that do not undergo recombination. It was postulated that non-homologous recombination between the SINE segments in the Y chromosome may result in the extensive intrachromosomal rearrangements that is observed in the eutherian Y (Graves 1995; Slattery et al. 2000).

Cattle

The first RH map of the bovine Y chromosome (Liu et al. 2002) included 49 microsatellites, three genes and ten STSs, making it one of the most comprehensive maps of the Y chromosome in domestic species. Additionally, few studies comparing sequence of homologous genes have been carried out across cattle, humans and mice. A homology of 60% was found 300 bp upstream of exon 1 of amelogenin genes (AMELY) in cattle and mouse (Gibson et al. 1998). Cloning and comparative analysis of ZFX and ZFY genes was recently reported for bovine, porcine and equine homologs (Poloumienko 2004). A high degree of conservation was found between the bovine ZFX and ZFY genes and homologous genes from the other species.

Among bovids, Y-specific microsatellite markers have been used to study population structure and diversity in cattle and river buffalo (Edwards et al. 2000; Hanotte et al. 2000; Van Hooft et al. 2002; Liu et al. 2003). Further, bovine Y- specific sequences have been used for sexing embryos (Alves et al. 2003), and for sexing evolutionarily closely related species like sheep and goat (Di Berardino et al. 2004). Y chromosome specific repetitive DNA has been cloned and sequenced using PCR in river

buffalo for sexing preimplantation embryos (Appa Rao and Totey 1999). Recently, Y markers were used for sex determination in water buffaloes using the PCR approach (Rao et al. 1993; Manna et al. 2003). Finally, fluorescent in situ hybridization (FISH) with bovine Y chromosomes probes led to sexing of spermatozoa of other species of the family Bovidae including river buffalo, goat and sheep (Di Berardino et al. 2004).

Dog

One of the first Y chromosome studies in dog described cloning and mapping of four pseudoautosomal genes on the X and Y chromosomes (Toder et al. 1997). A limited number of male specific markers including the SRY gene were later assigned to the canine Y chromosome using FISH and RH mapping approaches (Olivier et al. 1999; Guyon et al. 2003). Two nucleotide sequences of length 658bp and 989bp were identified on the canine Y chromosome using RAPD (Random Amplified Polymorphic DNA) PCR in 34 male dogs (Olivier and Lust 1998). In 1999, the 658bp sequence along with the SRY gene was localized on the canine Y chromosome using FISH (Olivier et al. 1999). Subsequently, this sequence was amplified from 285 dogs from 18 different breeds. The analysis yielded a marker that could be successfully used for detecting males in different breeds of domestic dogs. In the same study two polymorphic CA repeat microsatellites were identified among 12 different dog breeds (Olivier et al. 1999). On similar lines, a canine Y-specific marker of 800bp length was isolated using RAPD (Wang et al. 1999). Later, Y-specific sequences were also used to evaluate percentage of chimerism from male to female bone marrow transplantations (Fiegler et al. 2002). Lastly, Y chromosome markers were instrumental in identifying the presence of wolf-

dog hybrids in natural wolf populations (Vila et al. 2003) and in cases of sex reversal (Selden et al. 1978; Meyers-Wallen et al. 1995; Meyers-Wallen et al. 1999).

Pig

The first porcine Y chromosome map (Quilter et al. 2002) comprised of seven X/Y homologous genes and one Y-specific gene. The gene order was determined by RH mapping and later confirmed with dual color FISH. Consistent with the findings in humans, mice and cats, a conserved order of USP9Y-DBY-UTY loci was also observed in the pig. Gene order of the pseudoautosomal loci between the X and the Y chromosomes was found to be completely conserved suggesting that the porcine sex chromosomes were analogous to the ancestral sex chromosome (Quilter et al. 2002). In earlier studies, a few Y-specific markers were developed by various groups with an aim to obtain male specific markers for sexing embryos (Pomp et al. 1995; Castellanos et al. 1996). The entire Y chromosome was also used as a probe to identify host and donor cells in transplants (Braun et al. 1997). Later, sexing of porcine embryos and spermatozoa was carried out by FISH using Y-specific probes (Kawarasaki et al. 1998; Kawarasaki et al. 2000). Such probes were also used to detect aneuploidy in porcine sperm (Rubes et al. 1999). Thus, despite limited sporadic efforts to obtain male specific markers and to develop a preliminary map, no serious investigations have yet been carried out to get a broader picture of the structure and organization of the Y chromosome in the pig. Hence, like in other domestic species, little is known about the significance of the genes present on this chromosome.

OTHER USES OF Y MARKERS

The Y chromosome is different from the X chromosome and autosomes because of its overall non-recombining nature with the X chromosome (excluding the small pseudoautosomal regions or PARs). Consequently, the Y chromosome acts as a single non-recombining unit that is male specific and in effect haploid, thus ensuring the preservation of mutational events as a single haplotype via male lines.

In humans, a number of SNPs (Single Nucleotide Polymorphisms), indels (insertions/deletions), microsatellites, and minisatellites have been identified on the Y chromosome (Jobling and Tyler-Smith 2003). In the absence of recombination, the Y chromosomal haplotypes are strictly inherited in paternal lineages and change only by mutation. The mutation rate distinguishes between the slowly mutating markers (e.g. SNPs and indels) and rapidly evolving, hypervariable, microsatellites and minisatellites.

On the basis of the mutation rate, combinations of biallelic markers are used to define stable lineages of Y chromosomes that are informative for defining related groups of chromosomes called haplogroups. Diversities within each haplogroup can then be examined using rapidly evolving microsatellite and minisatellite markers.

There are other factors that make Y chromosomal markers an important tool for conducting population studies. One is the low population size of the Y chromosome. Assuming that the number of males and females are equal in a population, the number of Y chromosomes is one quarter of the number of autosomes. The low population size makes the Y chromosome more susceptible to genetic drift. Drift increases the diversity between populations with different Y chromosomes; which in turn is extremely effective

for investigating past events in a population. In addition, the Y chromosome shows greater geographic structuring than mitochondrial DNA and autosomal chromosomes due to differences in male and female behaviors. This structuring may be due to greater rate of female migration as compared to male migration rate (patrilocality) (Seielstad 2000). Genetically, this results in strong geographical clustering of Y chromosomal variants which are useful for haplotype and association studies. In contrast, due to migrations of females, the mitochondrial DNA gets homogenized over several generations resulting in reduced geographical clustering.

The lack of recombination, small population size and the practice of patrilocality has made the Y chromosome an immensely powerful tool to study the origins of humans and histories of different populations as summarized below (Hammer 1995; Jobling and Tyler-Smith 1995; Hurles et al. 1998; Seielstad et al. 1998; Bhattacharyya et al. 1999; Hammer et al. 2001; Stumpf and Goldstein 2001; Jobling and Tyler-Smith 2003; Manni et al. 2004).

The following paragraphs highlight the application of Y-specific markers in various key studies related to population structure and forensics.

Analysis of populations using Y-specific markers

Haplotype analysis using Y chromosomal markers in/across populations has shown that some haplotypes are shared by Asians and in Northern European populations including Finns, Buryats, and Yakuts (Zerjal et al. 1997). Y-chromosomal haplotypes constructed for seven Jewish populations indicated that Jews descended from common Middle Eastern ancestral populations (Hammer et al. 2000). Very recently, sequence

variation in the Y chromosome was used to suggest a common ancestry of Samaritans and Jews (Shen et al. 2004). Analysis of two isolates one from the central valley of Costa Rica and the other from Antioquia in northwest Columbia - were characterized using mitochondrial and Y chromosome markers (Carvajal-Carmona et al. 2000). Analysis showed that 94% of the male founders were European while 90% of the mitochondrial DNA gene pool was of Amerind origin. The contrasting results indicate that the founding population consisted of an admixture of immigrant men and native Amerind women.

Finland is home to some very rare genetic recessive disorders. This is attributed to a founder effect that took place ~2000 years ago (Nevanlinna 1972). Autosomal, mitochondrial and Y chromosomal markers were used to asses the level of genetic variation in Finland and test the bottleneck theory related to founding of the extant population (Kittles et al. 1999). Low variability in the Y chromosome along with evidence for distinct Y lineages proved that founding of this population was due to at least 2 bottleneck events and that multiple founding events were involved.

Studies using Y chromosome markers have facilitated the examination of the origin of founders in the New World (Bianchi et al. 1998; Schurr and Sherry 2004). These markers have also helped to analyze social hierarchical structure within the Hindu caste system (Basu et al. 2003), and provided strong evidence for an influx of Y chromosomes from the Indian subcontinent to Australia during Holocene (Redd et al. 2002).

Admixture studies using Y markers

Sex-biased admixture has been detected in Tibeto-Burman populations using Y chromosomal markers (Wen et al. 2004b). Strong evidence for male-biased admixture in Greenlandic Innuit from Europeans was recently shown (Bosch et al. 2003). The origin of Roma (Gypsys) was investigated using Y chromosomal and mitochondrial DNA and the results showed Asian-specific maternal and paternal lineages (Gresham et al. 2001). Differential admixture from around Europe was also demonstrated due to similar patterns of Y haplotypes of Roma and clinal distributions in Europe. Despite controversies in interpretations, several studies have attempted to elucidate the contributions of Neolithic and Paleolithic populations to the existing populations in Europe using Y-specific markers (Semino et al. 2000; Chikhi et al. 2002; Arredi et al. 2004; Semino et al. 2004).

MRCA (Most Recent Common Ancestor)

A lot of molecular genetic evidence has accumulated to support the "*out of Africa model"* for modern human origins. Nucleotide sequence diversity in the Y chromosome in chimpanzees and humans has dated the ancestral haplotype to 188,000 years (Hammer 1995). Contrary to this, sequence variation in three Y chromosomal genes in another study showed the time to the most common ancestor of humans to be more recent, i.e., around 59,000 years (Thomson et al. 2000). MRCA has also been calculated within population isolates using Y chromosomal markers. For example in the Samaritan isolate, 12 Y-STRs were examined in 124 DNA samples from different pedigrees, and the MRCA was estimated to be ~ 2000 years (Bonne-Tamir et al. 2003).

The results corroborated the genealogical records maintained by the community for the past several generations.

Paternity analysis

Y chromosomal markers have been frequently used for paternity testing when a male offspring is in question (Gusmao et al. 1999; Thangaraj et al. 2004). Especially significant is the fact that even though the father may be deceased, information on his Y DNA can be ascertained by testing his male relatives. One of the most well-known examples for this kind of study is the use of Y chromosomal markers that proved that US president Thomas Jefferson fathered a son by a slave (Foster et al. 1998). In another interesting study, a population sharing a common last name (Sykes) was studied using four Y-specific microsatellite markers. The average estimation of non paternity was found to be 1.3% per generation over 700 years (Sykes and Irven 2000).

Forensic analyses

A number of highly polymorphic Y-specific microsatellites are used for personal identification and forensic analyses (Iwasa et al. 2003; Johnson et al. 2003; Schoske et al. 2004). For this, STRs or Short Tandem Repeats are commonly used due to their high levels of diversity. Along with autosomal STRs, they are used to build highly discriminative haplotypes that are extremely useful in forensic investigations. In criminal investigations especially rape cases where the assailant's cells may have been mixed with that of the victim, it is much easier to type Y chromosomal markers than autosomal markers to differentiate between the two samples. In instances involving multiple assailants or cases where body fluids are mixed (blood-blood, blood-saliva), Y

chromosomal markers are found to be extremely useful in distinguishing subjects (Jobling et al. 1997).

Sex-differentiated behaviors

Mitochondrial polymorphisms and Y-specific microsatellites have high mutation rates and hence their patterns of variation are likely to reflect relatively recent divergence in populations. On the other hand biallelic polymorphisms (on the autosomes or the X chromosome) represent slowly mutating DNA sites and hence are useful to give better evidence for long term subdivision among populations (Hurles and Jobling 2001). Several recent studies have combined mitochondrial and Y chromosomal markers for examining the sex-specific behaviors in various populations (Wen et al. 2004a; Wilder et al. 2004). For example, the use of the two types of markers helped to understand that the expansion of the Han culture was attributed to a larger participation of males in this expansion than the females (Wen et al. 2004a). A recent study showed that contrary to popular opinion, Y chromosomal and mitochondrial genetic structure in humans is not influenced by the high rate of female migration (Wilder et al. 2004).

Likely applications of Y-specific markers in studying animal populations

Y chromosomal markers have been helpful in constructing informative haplotypes that are significantly useful for elucidating the origins and natural populations of humans. Hence, study of Y chromosome haplotypes would be expected to be immensely useful to study natural populations of domestic species. However such studies have not yet been very successful in domestic species due to lack of significant genomic resources such as a well developed Y chromosome map and availability of

polymorphic markers. Nevertheless a few studies have recently been initiated in cattle, wolves, dog and horse. Several polymorphic microsatellite markers have been identified in cattle and dog (Olivier et al. 1999; Edwards et al. 2000; Liu et al. 2003). These markers have been assessed for variability in different breeds though no phylogenetic analysis has been carried out yet. Identification of wolf-dog hybrids has been successful using polymorphic canine Y-markers (Vila et al. 2003). Finally, in horse three studies have so far unsuccessfully tried to identify SNPs and polymorphic microsatellites on the Y chromosome. The results appear to state that low levels of genetic variation exist on the domestic Y chromosome, seemingly due to a strong sex bias in mating (Wallner et al. 2003; Lindgren et al. 2004; Wallner et al. 2004). As more genomic resources are developed and polymorphic markers are isolated, it will be possible to study domestication and origin of livestock species.

SUMMARY OF RESEARCH ON THE HORSE Y CHROMOSOME

The horse Y chromosome remains the most understudied chromosome in the equine genome. This is evident from the current status of the horse gene map, where radiation hybrid maps of medium resolution are available for all autosomes and the X chromosome (Chowdhary et al. 2003). However, there is no map for the Y chromosome. This lack of information leaves a major gap in our understanding of its structure and organization and prevents studies aimed at analyzing the functional role of Y-specific genes that regulate male fertility and probably other biologically important functions. Below is a brief overview of studies hitherto carried out on the horse Y chromosome.

The very first elementary map consisting of three type I genes was built for the horse Y chromosome using synteny mapping (Shiue et al. 2000). Using a somatic cell hybrid (SCH) panel, synteny between STS-Y, SRY and ZFY in the horse Y chromosome was demonstrated. Later, localization of SRY and ZFY to this chromosome was confirmed using fluorescent *in situ* hybridization (FISH) (Hirota et al. 2001).

Thus far, partial genomic and/or cDNA sequences of five Y-specific genes have been reported. These genes include: AMELY, TSPY, SMCY, SRY and ZFY. RT-PCR (Reverse Transcription Polymerase Chain Reaction) was used to obtain 5' end sequence of the horse SMCY cDNA (Agulnik et al. 1997). Analysis of the synonymous substitutions and sequence comparisons showed that Y-linked genes were evolving much faster than the X-linked homologs. This led to the conclusion that mammalian evolution is male driven. Later a sex-specific *Hae*III RFLP was demonstrated in the horse ZFX gene (Senese et al. 1999). The ZFY gene lacks the *Hae*III site thus yielding sex specific banding patterns that can be used to identify males, females, and individuals with sex reversal. Further, a combination of RT PCR and RACE (Rapid amplification of cDNA Ends) was used to clone the equine SRY gene transcript (Hasegawa et al. 1999). Comparison of the coding region of the horse AMELY showed that the sequence differed by 24 bases with respect to the AMELX (the counterpart on the X chromosome). During recent years, sequence data from SRY and AMELY was used to develop a PCR test aimed at identifying the presence/absence of the two genes in XY mares (Hasegaw et al. 2000). Lastly, partial sequencing of the TSPY gene was carried out by Manz et al and PCR primers designed for sex determination in embryos. The

expression of TSPY was shown to be restricted to testicular tissue using RT-PCR (Manz, Vogel et al. 1998). Partial cDNA sequences were also obtained for the ZFY and STS-Y genes. Overall, the sequences specific to the horse Y chromosome have been utilized for determination of sex and chromosomal abnormalities (Pailhoux et al. 1994; Pailhoux et al. 1995; Abe et al. 1999; Vaughan et al. 2001; Bugno et al. 2003).

The microdissection of equine Y was first carried out by us to develop a whole chromosome painting probe (Raudsepp and Chowdhary 1999). DNA from the microdissected Y was also used for constructing a library that did not yield any microsatellites in initial analysis (Raudsepp and Chowdhary; unpublished results).

In contrast to the extensive studies of genetic variation in human populations using Y chromosome markers, there are only two studies focusing on genetic variation in the horse Y chromosome (Wallner et al. 2003; Lindgren et al. 2004). Using representational differential analysis, around ten Y-specific STSs and six microsatellite markers were developed and characterized for variation in 14 different breeds of horses (Wallner et al. 2003). Additionally, SNP analysis conducted over a 14.3kb of non coding region of the horse Y chromosome in 15 different domestic breeds (Lindgren et al. 2004) showed no variation. These results were consistent with earlier studies mentioned above.

From the above summary, it is obvious that the overall published information on the structure, organization and gene content of the horse Y chromosome is sparse. Consequently, very little is known about the function/role of the genes in regulating male fertility.

RATIONALE OF THE STUDY

Stallion infertility/subfertility is an extremely significant component of the equine industry. It is associated with millions of dollars of lost revenue annually. The underlying genetic causes for this are still largely unknown. It is evident from the above review that in addition to autosomes and the X chromosome, a proportion of genes governing stallion fertility are present also on the Y chromosome. Unfortunately, none of these genes have yet been identified or analyzed for their role and function. The amount of information presently available on the structure, organization and gene content of the equine Y chromosome is obviously the major limiting factor. To date, no organized studies have been undertaken to analyze the potential involvement of Y chromosomal genes in regulating stallion fertility. Hence, the current most important need is the development of a detailed map that could serve as a platform for launching the next level of studies aimed at identifying genes intricately involved in male reproduction/fertility.

To fulfill this goal, we herein report in the beginning, the first organized attempt to generate a detailed physical map of the horse Y chromosome. This map will form the basis to facilitate comprehensive investigations focused at identifying molecular causes associated with reduced fertility in stallions. As exemplified in humans, the focus of research in horses has to be on deletions, mutations and rearrangements in the Y chromosome and their effects on stallion infertility/subfertility. In humans, the detection of such ‘defects’ or ‘mutations’ was considerably facilitated by the generation of a dense marker map of the Y chromosome.

Taking a cue from these developments, Y-specific markers will be developed also in the horse using various tools and resources. These markers will be used to develop a basic physically ordered radiation hybrid map and FISH map of ECAY. The integrated FISH and RH map will provide a backbone to build contigs of BAC clones across the euchromatic region. Y-specific BACs will be isolated by screening three equine genomic BAC libraries. Subsequently, contigs will be built spanning the euchromatic region of the Y chromosome using STS content mapping, restriction fingerprinting and fiber-FISH analyses. At the conclusion of this project, a reasonable number of BAC clones will be ordered to obtain a moderate coverage of the euchromatic region of ECAY. The map will serve as a structural framework for obtaining a complete sequence of the euchromatic region of the Y chromosome in the future. This will lead to identification of genes involved with male fertility and sex determination. Additionally, the presence of a finely ordered and uniformly distributed STS map will aid in the formation of a molecular diagnostic test in the future to identify foals likely to have fertility problems when at stud.

In contrast to the extensive number of investigations in studying the origins and populations of humans, Y chromosomal haplotype information is scarce in livestock species. As indicated earlier, this is once again due to lack of mapping and sequence information on the Y chromosome. The second study marks the first systematic effort aimed at generating and characterizing microsatellite markers in order to study the genetic diversity among domestic breeds as well as among different equids. The generation of BAC contigs spanning almost the entire euchromatic region of the horse Y

chromosome in the first study will provide a valuable resource to isolate microsatellites and study genetic variability in a region that constitutes more than 25% of the male-specific region of the Y. Isolation of these markers can provide valuable insights into the signatures punctuating domestication of the horse and also evolution of different equids.

.

A DETAILED PHYSICAL MAP OF THE HORSE Y CHROMOSOME*

INTRODUCTION

Stallion fertility is of tremendous economic importance to horse owners and breeders. Breeding is a significant component of the multi-billion dollar equine industry with millions of dollars invested in stud fees alone. Stallion infertility is therefore of major concern to the equine industry. Male fertility in horses is a complex trait that is regulated by environmental, management, and genetic factors. So far intensive efforts have been invested by veterinarians to regulate environmental and management factors. However the genetic component of fertility which is shown to be of considerable significance in humans and mice (Vogt et al. 1996; Mahadevaiah et al. 1998; Mazeyrat et al. 1998; Foresta et al. 2001) has been completely neglected in the horse. In addition to autosomes and the X-chromosome, genes related to fertility and sex determination are present also on the Y chromosome. Hence a detailed study of the horse Y chromosome will be of immense significance in providing a complete genetic resource to comprehensively identify genes related to male fertility, and in the future their complex interactions. This will lead to improved diagnostic and therapeutic options for sub fertile stallions in the equine industry.

---

*This section is reproduced with permission from Proceedings of the National Academy of Sciences, USA, A detailed physical map of the horse Y chromosome, Terje Raudsepp, Avni Santani, Barbara Wallner, Srinivas R. Kata, Chengwei Ren, Hong-Bin Zhang, James E. Womack, Loren C. Skow and Bhanu P. Chowdhary, volume 101, 9321-9326, © 2004 by the National Academy of Sciences, U.S.A.

In humans, more than 25% of the infertile men show microdeletions in the Y chromosome (Vogt 1997). Research during the last 20 years has demonstrated the presence of more than 27 gene families out of which about 11 are expressed specifically in the testes suggesting that these genes are necessary for sexual development and spermatogenesis (Foote et al. 1992; Saxena et al. 1996; Foresta et al. 2000; Foresta et al. 2001; Ferlin et al. 2003; Skaletsky et al. 2003). Extensive mapping efforts in the frequently deleted regions of the human Y have yielded a dense marker map of this chromosome and ultimately the complete sequence of the Y chromosome (Tilford et al. 2001; Skaletsky et al. 2003). This information along with the use of molecular analytical techniques has facilitated identification of those regions of the Y chromosome that are prone to deletions and thus manifesting infertility (Vergnaud et al. 1986; Dada et al. 2003).

On the other hand, the horse Y chromosome remains the most understudied chromosome in the equine genome - a situation common to other livestock species. This is evident from the status of the horse gene map. Despite mapping of more than 2000 markers on the horse genome, less than 10 markers are mapped to the Y chromosome (Chowdhary et al. 2002; Chowdhary et al. 2003). On the whole, the published information on the horse Y chromosome is scarce. As far as known, no organized steps have as yet been undertaken to generate a map for this chromosome. The primary objective of this study therefore is to build a detailed physical map of the horse Y

chromosome. These objectives will be achieved through the following four specific objectives

1. Develop a basic RH (radiation hybrid) map and FISH (fluorescent *in situ* hybridization) map of ECAY using available and newly generated markers.
2. Isolate Y-specific BAC clones by screening three horse genomic BAC libraries with Y chromosome markers. Additionally end sequence BACs to generate new STS markers.
3. Build contigs of clones across the euchromatic region of ECAY using STS content mapping. Verify clone overlaps, orientation and physical order using restriction fingerprinting analyses and Fiber FISH.
4. Develop a detailed STS map across the euchromatic region of the Y chromosome.

The resulting ECAY map will be the most comprehensive Y-map among domestic species and second to only humans and mice. This study will be critical for identifying genes that are involved in male fertility and their interactions with genes on autosomes and X chromosome. Development of contigs and low resolution STS map will provide an essential diagnostic tool to identify regions commonly deleted in infertile stallions. Identification of these regions will aid in developing a PCR based test to evaluate the fertility of a stallion especially at a younger age. Such information could help in determining which stallion to purchase and the number of mares that should be booked for a particular stallion. Such a test would also help in eliminating stallions with genetic

defects or low fertility very early in the breeding program and alert owners regarding potential problems before the stallion embarks on his breeding career.

MATERIALS AND METHODS

Primer design, PCR optimization, and sequence verification

Primer pairs were developed from 15 Y-specific sequence tagged sites (STSs) and five partial cDNA sequences including the genes ZFY, AMELY, SMCY, TSPY and SRY (Wallner et al. 2003; Wallner et al. 2004) (Table 1). In addition, using human and porcine Y-specific gene sequences three sets of heterologous primers were designed. Finally, 57 end sequences from equine Y-chromosome specific BAC clones isolated in this study were used for primer design. All primers were designed using PRIMER3 software (http://www-genome.wi.mit.edu/cgibin/primer/primer3_www.cgi) & optimized on male horse, female horse, and hamster genomic DNA (for RH mapping, see below) to develop male specific markers. Primers were used for STS content mapping and bi-directional chromosome walking by PCR (as shown in Results). Detailed information on primers sequences and PCR conditions is presented in Table 1.

Table 1. Detailed information on markers, primers, PCR conditions and BAC clones

Marker symbol reference, Acc. No.	**Primers 5' – 3'**	**PCR product (bp)**	**Ta**	**BAC clone symbol**
Adlican (DKFZp564I1922) (3), CL525862	F: TTTGCCAATGGGACCCTG R: CGGACTGCATGAAGGAGTT	274	64	***TAMU 74.3H5***
AMELY (4), AB032194	F: CCAACCCAACACCACCAGCCAAACCTCCCT R: AGCATAGGGGGCAAGGGCTGCAAGGGGAAT	M: 160 F: 184	65	***ECY041, ECY042, ECY043***
CLY001 BV140811	F: CCACATTTTACACATGCCACA R: CCCAGAAAGACACCGTAACAA	117	62	ECY001, ECY002
CLY002 BV140812	F: TGCAACAGAAAGCACAGAGG R: TGCCAGGACCATCTCAGG	174	60	ECY002, ECY003
CLY003 BV140813	F: GCACCAATACGCTAGAGTCCA R: GCTTGGCCATGTTAAGTGCT	176	60	ECY002, ECY003, ECY004
CLY004 BV140854	F: TTACAACGGGAAGACCAACC R: ACCCAAGTGAAATCAACACC	228	60	ECY003, ECY004, ECY005, ECY006
CLY005 BV140814	F: GATTAAGAGCCCAGGAGAGG R: TTGGCTGGAGACTTGGTAGG	250	62	ECY003, ECY004, ECY005, ECY006, ECY007
CLY006 BV140815	F: GTTGGATCCTGCTGTGGACT R: ACAGCGTCCTTGATGCTTCT	212	62	ECY003, ECY004, ECY005, ECY006, ECY007
CLY007 BV140855	F: AGAAGGAGCCTGAGACAAGG R: CTCAACACCCAGTCTTTTTGG	154	60	ECY005, ECY006, ECY007
CLY008 BV140816	F: TGGCCTTTTGTGTGGTGTAA R: TCCACAGACCCATGCAAATA	238	60	ECY006, ECY007, ECY008
CLY009 BV140772	F: GACAGGAGGCACGTAAAGGA R: ATCTCCCCATCCCAAACTTC	171	67	ECY007, ECY008, ECY009
CLY010 BV140817	F: GTCAGCACGACAGCTCACAG R: ACAGCAATAGTCCACCAGCA	247	65	***ECY008, ECY009, ECY010***
CLY011 BV140818	F: TTGGGGAGACTTACCCACCT R: AGAGGCGTTAGGGTTGGTTT	100	60	ECY008, ECY009, ECY010
CLY012 BV140819	F: GCGAGTTCTGAGGACCAGAG R: GACCTGCCAACCAGTGATCT	173	60	ECY010, ECY011, ECY012, ECY013, ECY072
CLY013 BV140820	F: TGAAATCACAAAGCTCCAAATG R: GGTTACGAATGGCTCCTTGA	172	60	ECY010, ECY011, ECY012, ECY013, ECY072
CLY014 BV140821	F: AGTCCACACCACCACAGTGA R: TGATGGGAGTGGGAGTTTTC	248	60	ECY011, ECY013, ECY072

Table 1 continued

Marker symbol reference, Acc. No.	**Primers 5' – 3'**	**PCR product (bp)**	**Ta**	**BAC clone symbol**
CLY015 BV140874	F: CAGCATGGCTTCACTATTTCC R: TTTAGACAGGCAGAATCAGTTCC	164	66	ECY014, ECY015
CLY016 BV140774	F: TGTGACGGAGGCCAAAATTA R: ACTCCACATCAGGGTTGGTT	110	58	ECY014, ECY015, ECY016
CLY017 BV140856	F: TGTGGCAGAGTGAGCTTTCC R: CATTCCAGTAGAGGGCTTCC	348	60	ECY014, ECY015, ECY016, ECY017
CLY018 BV140775	F: CTGATGCAGGTTTCACTGG R: GAGTCAGAGAGCTGGAACTGG	109	67	***ECY015,*** ECY016, ECY017, ECY018
CLY019 BV140857	F: CCAAAACTAATGAGAGCCGTGA R: CACAGACTCAGGAGCAGGAA	173	60	ECY016, ECY017, ECY018, ECY019
CLY020 BV140776	F: CAGCTACTTGTGGTCTGGTCA R: GGACTGACTGTAAGAGCCCACT	103	65	ECY016, ECY019
CLY021 BV140777	F: GCAAGATTGGAAACATGAAGC R: TCACATTTCTTCTTTGCTCTATGC	177	65	ECY019
CLY022 BV140822	F: GGAGGCCACAGAGTGTTTTT R: GAAAGGTTGTCTCCATCTTTCCT	185	60	ECY020
CLY023 BV140778	F: CTCAGTGTCAGCAGGTTCCA R: TGCAGCTCTCTATCAGAACAGG	185	58	ECY020, ECY021, ECY022
CLY024 BV140779	F: GGGCTACAGGAGGACATGAG R: GCACAAGTTTGGCCGATT	177	67	ECY020, ECY021, ECY022, ECY023
CLY025 BV140858	F: TTAGAATGGGCTTGGCTCCT R: ACAGCTTTGAGGGATGGTTC	179	60	ECY020, ECY021, ECY022, ECY023, ECY024
CLY026 BV140823	F: TCGGTGACATCAGCAAAATG R: TACCCCGAATCCAGATCCTC	192	55	ECY020, ECY021, ECY022, ECY023, ECY024, ECY025
CLY027 BV140859	F: TTGAGCCCTTGTTTCTGTCC R: CAGGGAAGTAGGGAGTGACC	165	60	ECY020, ECY021, ECY022, ECY023, ECY024, ECY025
CLY028 BV140824	F: CCCATGACCTGTCCATACTG R: AACCAAGCCACATTTCATCG	182	60	ECY020, ECY021, ECY022, ECY023, ECY025
CLY029 BV140825	F: CTCCCCTCCTCCCACATTAT R: GGCAGCAGATCAACTACCTG	153	60	ECY021, ECY022, ECY023, ECY025
CLY030 BV140780	F: TCAGGTGAAGATTAAAGGAAGC R: GGGAATCCAACCAATAAGGAA	250	58	ECY021, ECY022, ECY023
CLY031 BV140781	F: CCATTTCAACTCCTGATCCA R: AAGGGTAGGTCAGAGAGGAAGG	101	65	ECY026
CLY032 BV140826	F: GGTCCAGAATGCCTGAGTAA R: AGAGACCTTTTGTGGGTGGA	396	65	ECY026, ECY027, ECY028

Table 1 continued

Marker symbol reference, Acc. No.	Primers 5' – 3'	PCR product (bp)	Ta	BAC clone symbol
CLY033 BV140827	F: AGCGAGGTCTGCACTTTCC R: GCGGGGAGTACATCAGTTCC	156	66	ECY026, ECY027, ECY028, ECY029
CLY034 BV140828	F: CAAGGGAAATGGAGTCAAGG R:AAAGAAAGTGTGTGTGTGTCAG G	201	65	ECY026, ECY027, ECY028, ECY029, ECY030
CLY035 BV140829	F: ATATGCCAGACATGGCACTG R: GCCCTGATCTAACTACTGCCAATC	145	60	ECY026, ECY027, ECY028, ECY029, ECY030
CLY036 BV140830	F: CTGCCTATCTCCATTCTTCATAC R: AGTGTTTTGGGGCAAGTGTT	153	65	ECY028, ECY029, ECY030
CLY037 BV140831	F:CCCCTCTTGCTCAGTTTTTG R: ACAGCAAACCAAACCCTACG			ECY029, ECY030, ECY031, ECY032
CLY038 BV140832	F: CTACGTGAAGGAATGTGTCTGG R: GATGTTTCTAAAGTCCAGCAAGG	150	60	ECY029, ECY030, ECY031, ECY032
CLY039 BV140782	F: AACCCATTTTCTCACAGTCTTG R: CAACACAAGTTGGAATGAGATG	106	60	ECY030, ECY031, ECY032, ECY033
CLY040 BV140860	F: ACTTTGCACTTGCCCCATAA R: TCCCAGCTAAAAAGGTACTCC	120	60	ECY030, ECY031, ECY032, ECY033, ECY034, ECY035
CLY041 BV140833	F: TGGGTTTCTTGGAATCCTGA R: CCAGTGAAGGGGGACATCA	249	60	ECY030, ECY031, ECY032, ECY033, ECY034, ECY035, ECY036
CLY042 BV140783	F: CAGACCAGAAGCTGAAGAAGAG R: GGGCTGCATACAAGGAAAGT	331	65	ECY032, ECY033, ECY035, ECY036, ECY037
CLY043 BV140784	F: GCGGTCTCAGTTTCTCTTCC R: CTGCTTCTTTCGCCTCTCC	242	65	ECY032, ECY033, ECY035, ECY036, ECY037
CLY044 BV140861	F: AGCTGATCGAACCCATAACC R: CTCTCCCAATGCCCTTCC	173	60	ECY033, ECY035, ECY036, ECY037
CLY045 BV140785	F: CAGAGGAGCGTCTTCCAGTT R: GGGTTTTTCCCCCAGTTTT	159	60	ECY033, ECY035, ECY037
CLY046 BV140786	F: TCAAGACTCTATCACAGCACTAAA CAG R: AAGGGTCCACCTCAGTCACA	100	58	ECY033, ECY037
CLY047 BV140834	F: GGGCCAGAATATGCAAGGA R: GATGTGTTTGTGTGCCTCTGTT	182	60	ECY038
CLY048 BV140835	F: AACAGAACCACTGCACTAAACC R: CAGATTCCCTTGGCTGACC	358	60	ECY038, ECY039, ECY040
CLY049 BV140836	F: TGTCAGCTTTGCCATTGTCT R: TCTGCCTGAAATGAAAGGAA	226	62	ECY038, ECY039, ECY040

Table 1 continued

Marker symbol reference, Acc. No.	**Primers 5' – 3'**	**PCR product (bp)**	**Ta**	**BAC clone symbol**
CLY050 BV140837	F: GCCTCAAGTAGAACCACATCC R: GCATCCAGAACAGCAAACC	300	60	ECY038, ECY040
CLY051 BV140787	F: ATAAATGCCGGAATCCATGC R: CCCCATGGGAATGGTAAAGT	181	52	ECY041
CLY052 BV140788	F: AGTGGTTAAGGCACAATCCA R: CCCCTACATCAGTTCAACTTTTT	166	60	ECY041, ECY042
CLY053 BV140838	F: ACAGTGCGTTCTGTGGTGAT R: AGTGAGCTGAGAATGCTTTGG	174	60	ECY041, ECY042, ECY043
CLY054 BV140789	F: TGCCCTTATCTACGTTTTGG R: TCTGCAAAGCTGGATCTCTT	123	58	ECY041, ECY042, ECY043, ECY044
CLY055 BV140839	F: GAACCAGCACTGCTCATCAA R: CCTCCAGAATGTCTCCTCCA	210	60	ECY041, ECY042, ECY043, ECY044
CLY056 BV140790	F: CCCGTCCTTGTACTTTGGAG R: ATTTCACATGGCCCCTAATC	128	60	ECY041, ECY042, ECY044
CLY057 BV140791	F: TTGCTTGTGGTATTTCCTTGC R: CCTTTCTTCTTTCCTTACATGC	116	58	ECY042, ECY044
CLY058 BV140840	F: GTGGAACTGCGCACTGCTTA R: CAGGAACATTAGGCCTCAGC	150	65	ECY045, ECY046, ECY047, ECY048, ECY050, ECY052, ECY057, ECY058, ECY059, ECY060, ECY063
CLY059 BV140841	F: AGGGAGGTCTATGGAGAAGG R: ATTTTTAGCTTGCCCTTTGG	296	65	ECY045, ECY046, ECY047, ECY048, ECY050, ECY052, ECY057, ECY058, ECY059, ***ECY060,*** ECY063
CLY060 BV140842	F: TGGAGAATTCACTGGCTGTC R: AAATGAAATAGCTGCAATGAAGT	102	60	ECY045, ECY057, ECY060, ECY059
CLY061 BV140862	F: CCCACCTCAGGATATTGCAT R: GCAGTGGTGTACAAAGACAGCA	214	65	ECY045, ECY047,ECY048, ECY050, ECY051, ECY054, ECY057, ECY058, ECY059, ECY060, ECY061, ECY063, ECY064, ECY065, ECY066

Table 1 continued

Marker symbol reference, Acc. No.	**Primers 5' – 3'**	**PCR product (bp)**	**Ta**	**BAC clone symbol**
CLY062 BV140863	F: ACACAGCTGTTGGTGTGCAG R: CCCTTTGCCTTCTGATGTTC	201	65	ECY045, ECY047, ECY048, ECY050, ECY051, ECY054, ECY057, ECY058, ECY059, ECY060, ECY061, ECY063, ECY064, ECY065, ECY066
CLY063 BV140864	F: CAGTCCTGTCAGTGCTCCAA R: CATGGCACAATGCAACTAGG	157	60	ECY045, ECY047,ECY048, ECY057, ECY058, ECY059, ECY060, ECY063
CLY064 BV140865	F: CTGCTGAGCAAGGGGTTAAG R: CGAGTGTTTCAGCAAACAGG	195	65	ECY050
CLY065 BV140866	F: GTGCCTGGGGATTCTCAGAT R: GTATAGTTTGGAAGGTGCTAAAC ACAG	150	60	ECY050
CLY066 BV140843	F: CCGATTCCAAACCATGAGAT R: AGGAAGTCAGCACCTTGCAT	138	65	ECY052
CLY067 BV140792	F: ACCCAGCCAAAAGCAGATAC R: CTGAGGCCAGGTGTGGAG	171	65	ECY045, ECY046, ECY048, ECY051, ECY054, ECY055, ECY056, ECY057, ECY061, ECY065
CLY068 BV140793	F: CAGTCTTCTTGTGGCGGACT R: TGTCGGTTAAGCTACTGGACCT	100	67	ECY046, ECY047, ECY048, ECY049, ECY054, ECY055, ECY056, ECY057, ECY061, ECY065,
CLY069 BV140794	F: GCCAGGATCGTGAGATATGG R: AACGTGCACGAGATAAGATGG	151	67	ECY046, ECY047, ECY048, ECY049, ECY054, ECY055, ECY056, ECY057, ECY061, ECY065
CLY070 BV140795	F: TCAACTTAGCGACTTCCTAGCC R: TTCAAAAATTCGATGTTGTCC	351	63	ECY048, ECY051, ECY054, ECY055, ECY056, ECY057, ECY061, ECY065

Table 1 continued

Marker symbol reference, Acc. No.	**Primers 5' – 3'**	**PCR product (bp)**	**Ta**	**BAC clone symbol**
CLY071 BV140796	F: GAACTCGCCTGTGGTTTCA R: GCCTGAATAAGATGCTGTCAAG	142	65	ECY045, ECY046, ECY048, ECY049, ECY054, ECY056, ECY057, ECY061, ECY062, ECY063, ECY064, ECY065, ECY066
CLY072 BV140797	F: CCATAGAGTGAGAGCTGATTGG R: TTGTTTTCCTTTTGGACTGG	238	62	ECY045, ECY046, ECY047, ECY052, CY057, ECY058, ECY059, ECY060, CY063
CLY073 BV140798	F: AGTTGAAGCAAGTGTGTGTGG R: CGGACTACGCCAAGAAAAGG	188	65(25 x)	ECY054, ECY061, ECY062, ECY063, ECY064, ECY065, ECY066
CLY074 BV140799	F: AATATCCACAACCCCTCTTCC R: CGGAGCATAGTAGCCAGACC	155	67	ECY054, ECY056, ECY061, ECY065, ECY066, ***ECY064***
CLY075 BV140800	F: GGGGAAAACTCCAAAGCAG R: CAGGGGAACAAAGCCAGAG	133	65	ECY051, ECY054, ECY062, ECY065, ECY066
CLY076 BV140875	F: ACACACATACCTGGCTGTCC R: GGTGAGCAGATCCCACTTCC	112	67	ECY045, ECY046, ECY047, ECY048, ECY051, ECY054, ECY056, ECY057, ECY058, ECY059, ECY060, ECY061, ECY062, ECY063, ECY064, ECY065, ECY066
CLY077 BV140801	F: AGCCTGGGGTTGATATGG R: CTGTTCGAGATTCAGGTTGG	216	65	***ECY066***
CLY078 BV140802	F: GACTCGGCCTGAAGCTAATG R: TCTGGGTTCTGGATCTGACTG	173	65	ECY066
CLY079 BV140803	F: GTGTGATCTGCTGTGCTTGG R: GCAGTTGCTGTGTGACTGTAGG	163	65	ECY051
CLY080 BV140804	F: CCCGCCAAGTCTATTTCC R: CAGTTAGTGGGAGGTGAGACG	156	58	ECY068
CLY081 BV140805	F: TCAGAGATGGGCATTTTCAC R: CCTTATTCAAACAGCGTCCA	212	65	ECY069
CLY082 BV140806	F: AGTGGGTTTCAGGCAAAATG R: GGCACTTGTCTTCTCGGTGT	209	65	ECY070
CLY083 BV140807	F: GCTGCCTGAAACCTGACTTC R: TACTCCAAGTGCCAAGCTCA	176	65	ECY070

Table 1 continued

Marker symbol reference, Acc. No.	Primers 5' – 3'	PCR product (bp)	Ta	BAC clone symbol
CLY084 BV140844	F: TGGCTGGGTCCAATGTTTAT R: CAGCCAAAAACTGGTAGAATCA	138	65	ECY071
CLY085 BV140845	F: TCTGGGTTCTGGATCTGACTG R: GACTCGGCCTGAAGCTAATG	173	65	ECY071
CLY086 BV140867	F: CTTCCCAGTCCCTTCTCCAT R: CGAGAGTTGAGTGCCAGTGA	241	65	ECY067
CLY087 BV140868	F: TATGGAACCCCTTCTGCAAG R: CCTGCCGACATGACAAAATA	209	65	ECY067
CLY088 BV140869	F: GCGAGGGGTGAGTTATTGAA R: TTAGCTGGCCCTTCCCTTAT	193	60	ECY067
CLY089 BV140870	F: GAGGAGAAGTGCCTTAAATTCC R: CCCAGAAAATCGTTCATTCC	177	60	ECY049
CLY090 BV140846	F: TCATCTCGTATCTCCTCATATCC R: TCCCTATCCTTGTTGAAAATCC	495	60	ECY053
CLY091 BV140847	F: ACAGTCTCCTGCTGGTTTCC R: AATTTCAGCCTCCTTTCCA	156	60	ECY053
CLY092 BV140871	F: CGTACTGCTTTTAGCTTGAGACC R: AAGAGGTGAGTCCGTTGACC	236	60	ECY055
CLY093 BV140872	F: GTTGCTCTGGGTTTTCTTCC R: ACTCACCATCACACGACACC	247	60	ECY055
CLY094 BV140873	F: CCTCTGCTGGTCACCTTCTG R: TTCCACATAGAAGCCCCCTA	242	60	ECY056
CLY095 BV140848	F: TCACAACCTGCTTGTTCTTCC R: TGAATGGGAATGCATACATGA	199	55	ECY020
CLY096 BV140849	F: GGAATGGGCAGACTTAACCA R: GCAAACCAAACCCTACGAAA	185	60	ECY031
CLY097 BV140850	F: AGTGTCTTGCCAAGGAATGC R: AACCAAATGTCCCTGCAAAC	214	60	ECY031
CLY098 BV140851	F: AGAAGCATCAAGGACGCTGT R: GCTAAGGGCCAATTTTCCTC	163	60	ECY004
CLY099 BV140852	F: GGGTGCAGAACCAGCACT R: TCGTCCCTCAGCTAACATCTG	162	60	ECY043
CLY100 BV140853	F: GGCCCCATCGTCACTTATTT R: CAGAAAAGGCCAGTCCACAC	166	60	ECY018
CLY101 BV140808	F: ATCTGCTTCGGCCTTCTCTT R: GGTGTACCCTGCTTCTCGT	101	65	ECY072, ECY012
CLY102 BV140809	F: GTGGGGGCAATCTCAAGG R: CCAACTGGCAATCACAAGG	282	65	ECY072, ECY012
CLY103 BV140810	F: TTTGGTTTCCCACTCTCTGTG R: CTTCCTTTGCAGGCTAGTGA	190	58	ECY019

Table 1 continued

Marker symbol reference, Acc. No.	**Primers 5' – 3'**	**PCR product (bp)**	**Ta**	**BAC clone symbol**
DDX3Y CL525856	F: TGCATATAGCAGTTTTGGGTCT R: CATAGTCACTCCGTCCACGA	M: 900 F: 1000	58	***ECY073***
SH2-A-1 (24), BV005744	F: CGGTGTCAGGTTTTGGACTT R: AAGGATTCTGCTGCCCTCAT	746	64	***ECY046, ECY048, ECY049, ECY050, ECY053, ECY067***
SH3-B-14 (24), BV005745	F: GTGACCTCCCAGGAGCTGT R: TCTGCCTATGCTCTGGTGAA	486	64	ECY055, ECY056
SH3-B-6 (24), BV005718	F: AGAGTGCGATTTGTGATGG R: AGAGTCAGAAGAAAGCGTTGAT	492	64	ECY055, ***ECY056***
SH3-B-7 (24), BV005719	F: TGAAAGGGCTAATGAACGAT R: CCTCAGGTTTCGGGATT	436	58	***ECY046, ECY052***
SMCY U52364	F: TGGAGGTGCCCRAARTGT R: AACTCTGCAACLYTRTACTCCT	400	58	***ECY007, ECY008, ECY009***
SRY (6), AB004572	F: TCATGGTGTGGTCTCGTGAT R: CCGGGTATTTCTCTTGATGC	200	58	***ECY010, ECY011, ECY012, ECY013, ECY072***
TSPY AF026959	F: CAGAGGAGAAGCCAGAGCAG R: ATGGCGCTTCTCCGATCTA	196	58	***ECY046,*** ECY063
USP9Y CL525857	F: TCACACTTTTACGGCACCTTC R: CCTTTTGAGCCAATCAATTTC	101	58	ECY030, ECY031, ECY032, ***ECY033,*** ECY034, ECY035, ECY036
UTY (18)	F: CAGCTGTTTTCGGTGATGAG R: GCCTCCTTCTCTTCGGTTG	101	54	-
Y2B17 (23, 24), G72335	F: TTCAGTCCTGCTTTCTCCTCA R: CAGGATGTGCCATGTGATTG	528	58	***ECY020, ECY021, ECY022, ECY023, ECY024, ECY025***
Y3B1 (23, 24), G72336	F: TGGGTTAATGGGATTTGGTG R: CAAGCACAGCTCTGTATCAA	508	58	***ECY030, ECY031, ECY032, ECY033, ECY034, ECY035, ECY036***
Y3B12 (23, 24), G72338	F: GGGAGGCACTGGAAAGTACA R: GGTGGAGGAATCAGCTGGAG	400	58	***ECY015, ECY016, ECY017, ECY018***
Y3B19 (23, 24), BV005720	F: AAGCCTTTCATGGAAATTGG R: TTACGCAGACATCCTGGACA	255	58	ECY015, ECY016, ECY017, ECY018
Y3B8 (23, 24), G72337	F: CCCAAGTTCCTTGCCATC R: AAATTGAAGAGGCCCCAAAG	472	58	ECY004, ECY005, ***ECY006, ECY007***

Table 1 continued

Marker symbol reference, Acc. No.	**Primers 5' – 3'**	**PCR product (bp)**	**Ta**	**BAC clone symbol**
YA16 (24), BV005729	F: TGACTGGAAATTGAAGATG R: TTGTAGCAACAAAGTAACAC			-
YE1 (24), BV005727	F: CTTCACTCCCGACCAAGAGA R: GTGTGTCGTGCCGTGTTTAC	199	58	***ECY020, ECY021, ECY022, ECY023, ECY024, ECY025***
YH12 (24), BV005747	F: CGAACAGGTGACGAAGCATC R: GCAGACATGCACACCAACC			-
YJ10 (24), BV005728	F: AGTTCCCCTGCACACCT R: TGCCTCCCACAGCCATAC	215	64	ECY046
YM2 (23,24), BV005725	F: TGGTTCAGATGGTGTATTTTGTT R: TTTGCAGCCAGTACCTACCTT	119	58	***ECY003,*** ECY004, ECY005, ***ECY006***
YP9 (24), BV005726	F: AAGCACTGCCTTTTGGAATC R: AACCCTGGACTTTCTTTTGAA	216	60	ECY010, ECY011, ECY012, ECY013, ECY072
ZFY (6)	F: TGCACATTTCCTTTAATCT R: GCACATTAAAGAGAAACCTT	342	58	***ECY038, ECY039, ECY040,***

BAC clones in bold/italics represent clones that have been used for FISH analysis. – represents markers for which no BAC clones could be identified despite repeated screening attempts.

RH mapping

Primers from eight genes and 15 STS markers were typed in duplicate on a 5,000- rad horse x hamster RH panel (Chowdhary et al. 2002). PCR was carried out in 10 µl reactions containing 50 mM KCl, 10 mM Tris HCl (pH 8.3), 1.5 mM $MgCl_2$, all four dNTPs (each at 0.2 mM), 0.3pmols/µl of each primer, 50 ng of BAC DNA as template, and 0.025 unit of *Taq* polymerase (Sigma). After initial denaturation at 95°C for one min, samples were amplified by 30 cycles as follows: 30s at 94°C (DNA denaturation), 30s at 55 - 65°C (primer annealing), and 30s at 72°C (primer extension) followed by a 5-min final primer extension at 72°C. PCR products were separated on a 2.0% agarose gel containing ethidium bromide (0.5 µg/ml) and visualized under UV light. Scoring of positive clones and RH data analysis was carried out as described in detail earlier (Chowdhary et al. 2002; Chowdhary et al. 2003).

BAC library screening and end sequencing

Three equine Bacterial Artificial Chromosome (BAC) libraries: Texas A&M University equine BAC library (http://hbz7.tamu.edu/homelinks/bac_est/bac.htm), Children's Hospital Oakland Research Institute 241 (CHORI 241) (http://bacpac.chori.org/equine241.htm) and Institut National dela Recherche Agronomique-complemented BAC library (INRA) (Milenkovic et al. 2002) (http://dga.jouy.inra.fr/grafra/INRA_libraries_database_simplified.htm) were screened by male specific primers to isolate Y-chromosome specific BAC clones using PCR. Initial screening was carried out for 23 markers that included 15 STSs and eight genes (as shown above). Subsequent screenings were carried out using primer pairs obtained

from end sequences of BACs obtained in this study (Table 1). The screening proceeded from superpools to plate pools and then to individual plates. Positive clones thus identified were streaked on LB/chloramphenicol plates and verified again for the presence of individual markers by PCR. A single positive clone was selected for culture. Subsequently DNA extraction was carried out using a Qiagen midi prep kit (Qiagen, Chatsworth, CA) according to manufacturer's instructions. BAC end sequencing was carried out using universal end sequencing primers (T7 forward, M13 reverse, SP6 reverse). Dye terminator sequencing reactions (total volume 10µl) were set up as follows: 1µg BAC DNA, 2µL standard sequencing primer (10 µM solution), 2µL BigDye (Big Dye Terminator v2.0 Cycle Sequencing Kit Applied Biosystems), 2µl half BigDye (*half*BD^{TM}, Genetix), 0.5µl Master Amp (Amp PCR Enhancer, Epicentre, Madison, WI). Cycle sequencing was performed in a Gene Amp PCR system 9700 thermocycler using the following cycling conditions: 96°C for 2min followed by eight cycles of 96°C for 30 sec, 58°C to 50°C touchdown for 10 sec, 65°C for 4 min, followed by 60 cycles of 96°C for 30 sec, 50°C for 30 sec, 60°C for 25 sec, and one cycle of 96°C for one min, 50 °C for one min, 65°C for 15 min. Reaction products were purified using spin columns (Spin-50, Bio Max, Odenton, MD) and loaded on an ABI 3100 automated capillary sequencers (PE Applied Biosystems) for analysis.

BAC fingerprinting and contig assembly

BAC DNA was isolated and fingerprinted as shown in (Ren et al. 2003). Briefly, BAC clones were inoculated in 96-deep well plates containing one ml LB medium and 15 µg/ml chloramphenicol. The cultures were grown overnight at 37°C with constant

shaking at 250 rpm. BAC DNA was isolated and purified using a modified alkaline lysis method (Chang et al. 2001). The DNA was double-digested with *Hin*dIII and *Hae*III, end labeled with [^{32}P]dATP using reverse transcriptase at 37° for two hrs, and then subjected to 3.5% (w/v) polyacrylamide DNA sequencing gel electrophoresis at 90 W for ~100 min. The gel was dried and autoradiographed. The fingerprints on the autoradiographs were scanned into image files using a UMAX Mirage D-16L scanner. The image of the fingerprints was size adjusted to 1.1 MB, transferred to a computer workstation (SUN Microsystems, Utra10), and edited using the Image 3.10b software (Sulston et al. 1989). Because of the lower resolution of the higher-molecular-weight bands at the top of the gels, only the bands ranging from 58 to 773 bp were used for contig assembly. Vector bands were removed manually from the data files. All digitized band data were standardized against the λ-DNA/*Sau*3AI marker and converted from base pairs into migration rates. Contigs were automatically assembled by FPC PROGRAM Version 6 (Soderlund et al. 2000). The tolerance was fixed at 2, and the cutoff value was set to 1e^{-06} by a series of tests.

FISH

Peripheral blood from a stallion of known fertility was collected in Na-heparin vacutainers. The blood was used for (i) short-term pokeweed (Sigma) stimulated lymphocyte cultures followed by metaphase/interphase chromosome slide preparation according to standard procedures and (ii) preparation of agarose imbedded DNA plugs and slides with mechanically stretched DNA fibers (Heiskanen et al. 1994). Lymphocytes from whole blood were isolated using Lymphoprep TM (Axis-Schield

PoC AS) and imbedded in 1% low melting point agarose (NuSieve® GTG®) plugs using CHEF molds (BioRad). Cells were lysed by incubating the plugs at 50°C in 1mg/ml Proteinase K (Sigma), 50 mM EDTA and 1% N-laurosylsarcosine (Sigma) for 48h. The plugs were washed four times in 1X TE (1mM EDTA, 10mM Tris-HCl, ph 7.5) and incubated at 50°C in 40 ug/ml phenylmethylsulfonyl fluoride (Sigma), 100% ethanol solution for 50 min. Finally, the plugs were washed four more times in 1X TE and stored at 4°C in 1X TE until further use. DNA fiber slides were prepared as following: a small piece of agarose imbedded DNA was placed at the end of poly-L-lysine (Sigma) coated glass slide, melted in microwave oven for one min and mechanically stretched with an aid of another slide. The slides were air-dried and used for hybridization the same day.

BAC DNA was labeled with biotin-16-UTP and digoxiegenin-11-dUTP by nick translation using Biotin- or DIG-Nick Translation Mix (Roche Molecular Biochemicals). Hybridization and signal detection was carried out on metaphase/interphase chromosomes and DNA fiber slides as described earlier (Raudsepp and Chowdhary 1999). Briefly, 2µl of probe DNA (50ng/ul) was denatured and hybridized to denatured chromosome or DNA fiber slides. After overnight hybridization, the slides were washed three times in 50% formamide 2XSSC at 40°C and four times in 4XSSC, 0.01% Tween20 at RT. Biotin labeled probes were detected with two layers of avidin-FITC (Vector) and one layer of biotinylated anti-avidin (Vector). Detection of dig-labeled probes was carried out using mouse anti-digoxigenin-rhodamine (Roche) antibodies. The slides were mounted in DAPI-antifade solution coverslip and stored at -20°C until microscope analysis.

FISH results were examined and analyzed with a Zeiss Axioplan2 fluorescent microscope and CytoVision/Genus application software version 2.7 (Applied Imaging). At least 30 metaphase spreads and ~50 interphase or DNA fiber hybridizations were analyzed and captured for each experiment.

RESULTS

BAC library screening, end sequencing and bidirectional chromosome walking

Three BAC libraries (Texas A&M University, CHORI-241, and Institut National de la Recherche Agronomique) were screened with a set of 23 markers that included eight genes and 15 STS. Three markers (*UTY, YA16*, and *YH12)* failed to isolate a BAC and a total of 37 BAC clones representing 20 of the markers were isolated. Direct end sequencing of each clone provided nucleotide sequence information used to derive novel sequence tagged sites. Each sequence was analyzed against Repeat Masker and BLAST before designing primers. This analysis provided nucleotide sequence information averaging 928 bp (range, 300–1,400 bp), which was used to develop 60 previously uncharacterized unique STSs. The remaining sequences were predominantly long interspersed repeats and therefore could not be used for further analysis.

Screening the BAC libraries with recently developed STS markers provided additional depth to the foundation contigs and also added few BACs to some of the termini. On the basis of the marker content, the BACs were manually grouped into nine contigs. Expansion of contigs facilitated by bi directional chromosome walking with these markers derived from the end sequence of BAC clones resulted in closure of two

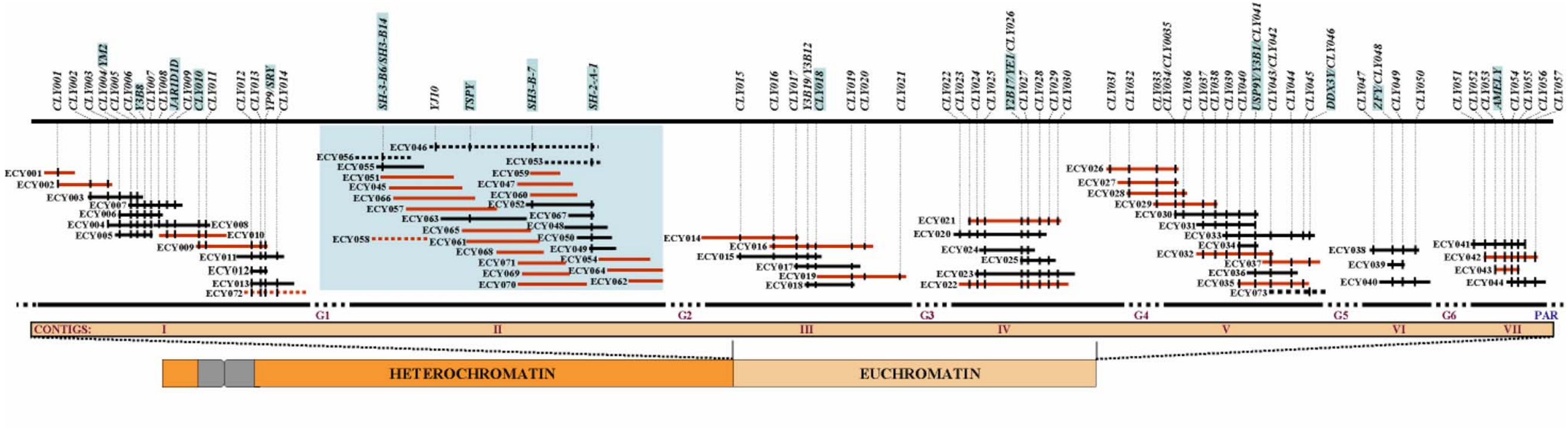

Figure 1 A map of the euchromatic region of ECAY. (*Top*) A contig map of 73 BACs over the euchromatic region of ECAY. Top row represents 77 STS loci and genes used to generate the seven contigs (I–VII). Loci shaded blue are also present in the RH map. Contig II (in blue rectangle) shows BACs ordered only by fingerprinting. G1–G6 represent gaps in the contigs. BACs shaded black represent clones isolated using the first set of 23 markers. BACs shaded red represent additional clones isolated by chromosome walking.

gaps between *SRY-JARID1D* and *YM2-JARID1D* (see Figure 1). Two subsequent rounds of chromosome walking resulted in isolation of 36 new BACs (highlighted red).

These BACs were end sequenced to develop new STS markers. End sequencing of these BACs yielded 43 additional STS. Thus, the entire experiment of end sequencing and chromosome walking provided a total of 103 STS. However, 32 of the STS amplified a similar-size DNA band in the control female in addition to the male, whereas the remaining 71 were male specific. The average size of the STS markers was 186 bp (range 100–495 bp).

STS content mapping, restriction fingerprinting and contig assembly

Analysis by PCR of the 73 BACs with 23 initial markers and 103 newly developed STSs provided seven groups of overlapping BAC contigs. Six of the contigs (I and III–VII; Figure 1) were comprised of 3–13 BACs. Contig II, referred to as the putative "multicopy" region (discussed later), had 27 BACs. Within each of the former six contigs, the likely order, orientation, and overlap of BACs were determined on the basis of the presence or absence of specific markers. A summary of results showing individual contigs (I and III—VII, oriented centromere to telomere) and all markers tested for content mapping of BACs are presented in Figure 1. BACs in the "multicopy group" (contig II; light-blue shaded region, Figure 1) were ordered by using a finger-printing approach. The physical order of the contigs was derived by using a combination of RH and interphase-/fiber-FISH approaches and is described in pertinent sections.

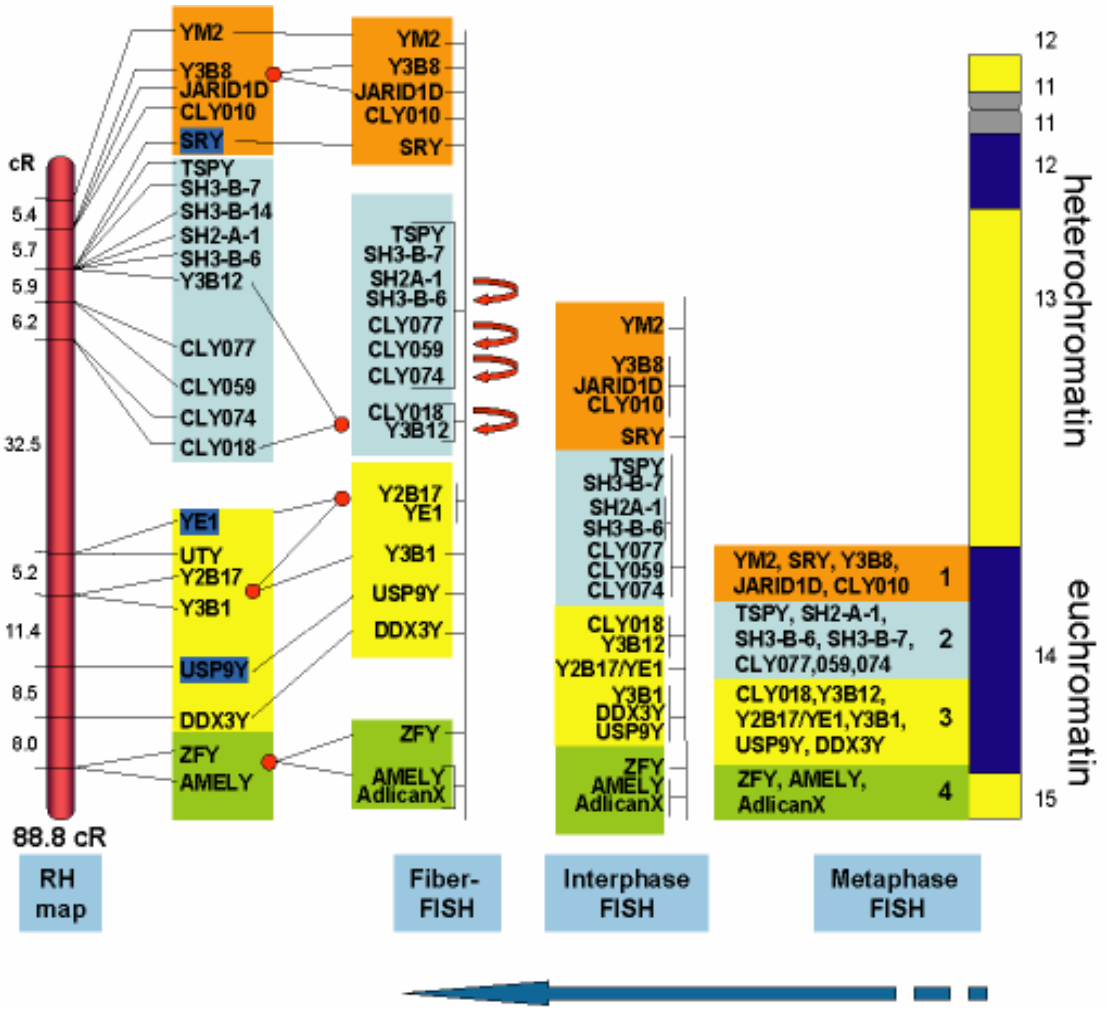

Figure 2 RH and cytogenetic FISH map of ECAY. 23 markers (8 genes and 15 STSs) were mapped on ECAY using RH and FISH mapping. The RH map is correlated with maps obtained by various FISH approaches through the four color codes for groups of loci (1–4 beside Y chromosome ideogram). Fiber and interphase FISH resolved order for several markers. Semicircular arrows in the fiber-FISH map indicate partial overlaps between copies of some of the markers. Blue arrow represents increasing level of resolution.

Restriction digestion and fingerprint analysis of 67 BACs (6 BACs did not grow) gave a total of 2,466 unique bands at a tolerance of 2 and a cutoff at $1e^{-06}$ and allowed the BACs to be assembled into seven contigs. One clone remained as a singleton. Overall results show that the fingerprint map (not presented) is in close agreement with the contigs obtained by STS content mapping.

RH mapping

Twenty three markers (8 genes and 15 STSs) were typed on the 5,000-rad horse x hamster RH panel with an average retention frequency of 11%. The analysis provided a map that spanned 88 centirays and covered almost the entire euchromatic region of the ECAY (Figure 2). Two-point linkage analysis of typing data resulted in a single linkage group at a logarithm of odds score of 7 (2PT-RHMAP) (Liu et al. 1998). Anchoring the RH map to chromosome by FISH enabled us to orient the map and indicated that *YM2, JARID1D, SRY*, and *TSPY* are proximally located within the euchromatic region, whereas *USP9Y*, *DDX3Y, ZFY*, and *AMELY* are distal. At seven of the map locations, two or more markers clustered at the same position (Figure 2), but the relative order of these loci could not be determined. The small size of the mapped segment and the suggested ~1Mb resolution limit of the 5000 rad panel are several explanations as to why the relative order of closely clustered loci could not be resolved. However, other approaches described below were used to resolve physical order for these loci.

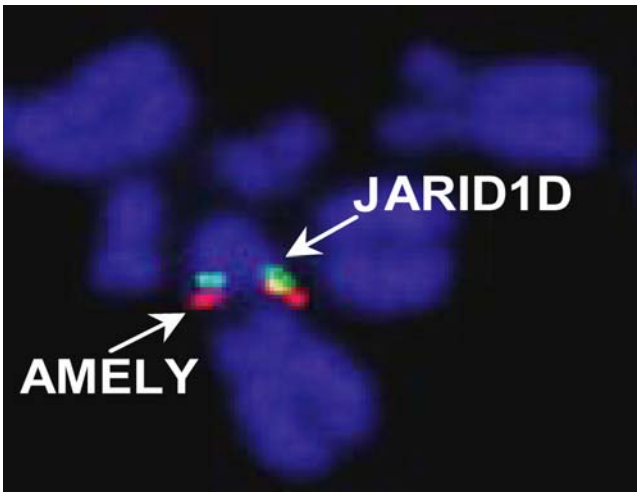

Figure 3 Cohybridization showing the order of JARID1D and AMELY. Double color FISH on metaphase chromosomes shows a gap between the two markers and also elaborated their proximal and distal orientation on the chromosome.

FISH mapping by metaphase FISH

Of the 114 BAC clones isolated in this study, 67 hybridized specifically to the Y chromosome, whereas four hybridized to both X and Y. Six of the clones hybridized both to Y and an autosome. Of the remaining clones, 22 showed signals on only autosomes, two only on the X chromosome, and 13 did not give a specific signal. These hybridizations helped to verify the origin of individual clones. BAC screening from the three equine genomic libraries resulted in one or more clones for each marker. After the initial control hybridizations with all isolated BACs, a total of 22 clones were selected and used for cytogenetic localization of the RH mapped markers (see Table 1). Markers *Y2B17/YE1* and *SH3-B-14/SH3-B6* were present in the same BACs. Single and double-color metaphase FISH with multiple pairwise marker combinations helped to divide the 22 loci into four major physically ordered groups in the euchromatic region of ECAY (groups 1–4; see Figure 2). For example, cohybridization of *JARID1D* (group 1) and *AMELY* (group 4) clearly showed a noteworthy gap between them and also elaborated their proximal and distal orientation on the chromosome (Figure 3). Due to the diminutive size of ECAY euchromatin and the resolution limit of ~2–3 Mb for metaphase FISH (Trask et al. 1993), it was not possible to resolve the order of markers within the groups.

Interphase FISH

Double-color hybridizations on interphase nuclei with labeled probes used in different combinations confirmed the deduced physical organization of the four groups and

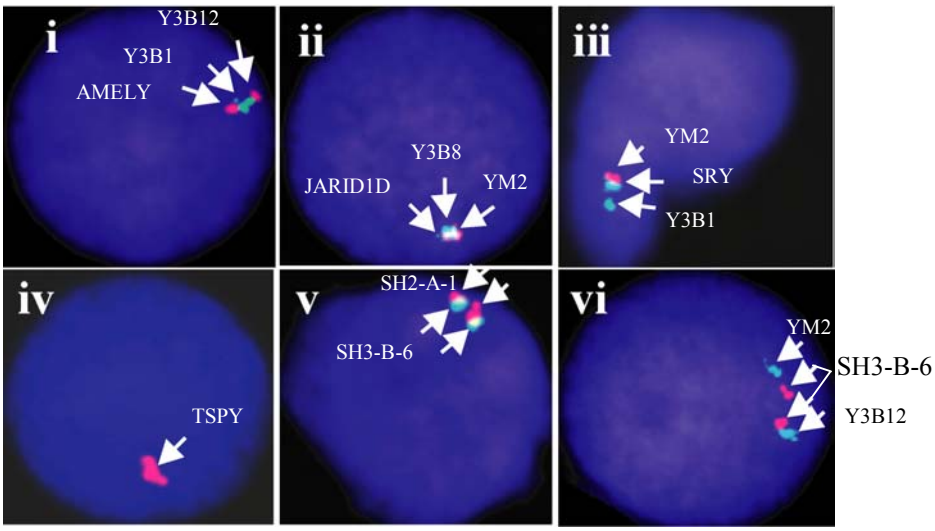

Figure 4 Interphase FISH to resolve order of markers. Double-color hybridizations on interphase nuclei with labeled probes used in different combinations confirmed the deduced physical organization of the four groups.

allowed us to resolve the physical order of the majority of the loci within each group, in particular the distal groups 3 and 4 (Figures 2 & 4). For example, *Y3B12* was found to be the most proximal marker in group 3. In group 4, *AMELY* was the most terminal marker (Figure 4i), and *ZFY* was distinctly proximal to *AMELY*. However, BACs for *Y3B1, USP9Y*, and *DDX3Y* in group 3 and for *AMELY* and *AdlicanXY* in group 4 gave strongly overlapping signals on interphase chromosomes, hence their order remained ambiguous. Similarly, signals for markers in the most proximal group (*YM2, Y3B8, JARID1D*, and *SRY*) mostly overlapped (Figure 4*ii*). We could nevertheless show that the end markers of this group were *YM2* and *SRY* by cohybridization of BACs containing these two loci with markers from other groups. The results revealed that *YM2* is the most proximal and *SRY* is the most distal within this group (Figure 4iii). It is remarkable that all markers in group 2 (Figures 2 & 4) showed two or more hybridization signals in interphase nuclei, indicating the presence of more than one copy of these sequences on ECAY. The BAC containing *TSPY* showed the strongest hybridization signal (Figure 4iv). Other markers gave ≈ 2–10 signals in interphase cells. Therefore, cohybridization of markers from this region could not resolve their relative order (Figure 4v). Nonetheless, it appears that signals from these BACs are regionally clustered and the majority of the multiple copies are located between group 1 (e.g., *SRY*) and group 3 (e.g., *Y3B12*; Figure 4vi). However, the presence of these copies in other regions of the ECAY euchromatin cannot be excluded.

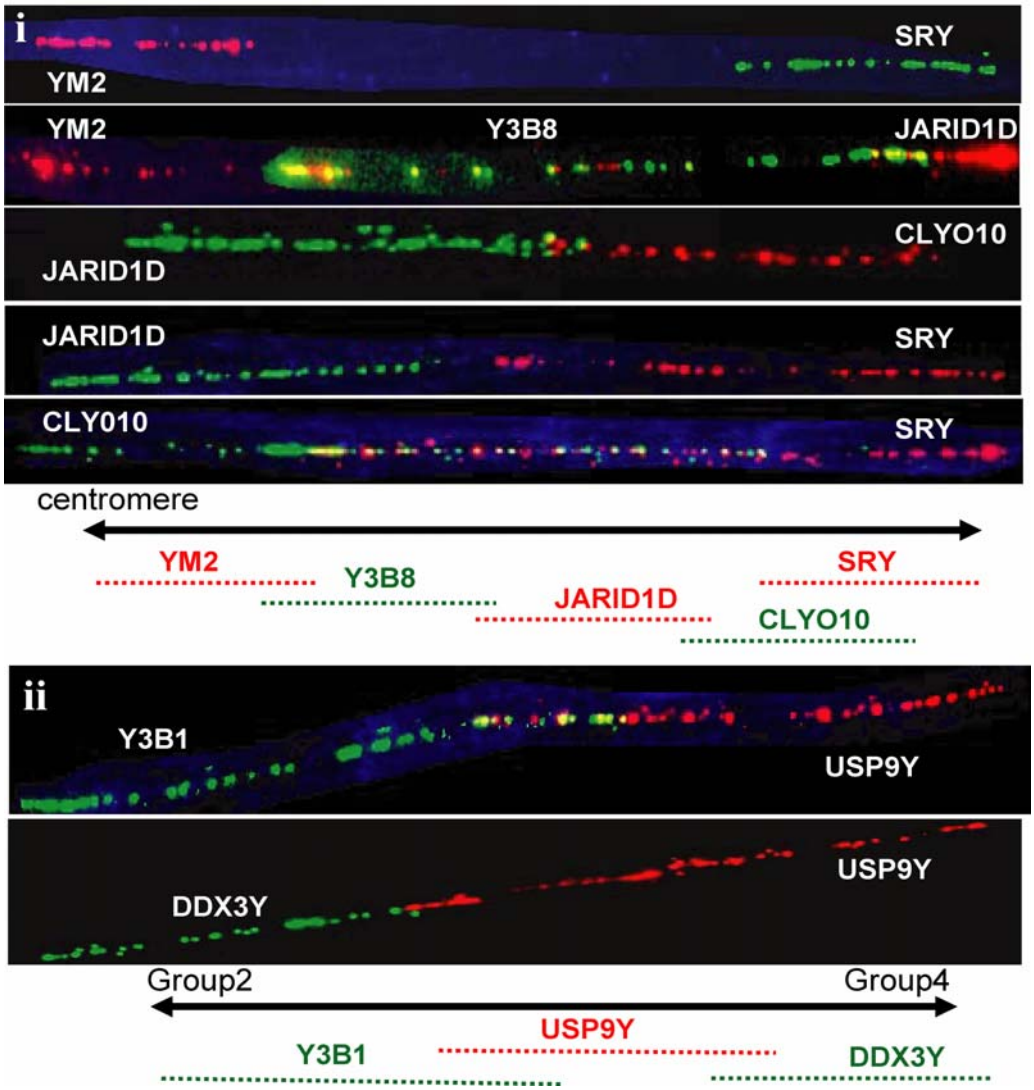

Figure 5 Fiber FISH to resolve order of loci. Two examples whereby fiber FISH resolved order of loci, as depicted in schematic drawings below each set of experiments. i) precise determination of the order of five markers as: prox-*YM2-Y3B8-JARID1D-CLY010-SRY*-dist ii) the relative order of markers in group 3 was resolved as: prox-*Y3B1-USP9Y-DDX3Y*-dist.

Fiber FISH

Hybridizations with combinations of two or three probes on mechanically stretched DNA fibers from the most proximal cluster (group 1) enabled precise determination of the order of five markers as: prox-*YM2-Y3B8-JARID1D-CLY010-SRY*-dist (Figure 5i). Similarly, the relative order of markers in group 3 was resolved as: prox-*Y3B1-USP9Y-DDX3Y*-dist (Figure 5ii). However, BAC clones for markers in the terminal group 4 (*AdlicanX* and *AMELY*) showed neither proximity nor overlaps on DNA fibers. Hence their relative order could not be resolved.

As expected, ordering markers in the multicopy cluster was difficult. Nonetheless, we could confirm and refine interphase results concerning close proximity of *SH2-A-1* and *SH3-B6*, showing that at least one copy of *SH2-A-1* partially overlapped with at least one copy of *SH3-B6* on fibers (Figure 6). Fiber FISH with *TSPY* confirmed our interphase observations that the gene has several copies on ECAY. Further, cohybridizations indicated that some copies of marker pairs *CLY077-CLY059* and *CLY059-CLY07*, respectively, either overlap or are in close proximity to each other. However, due to the presence of numerous non-overlapping copies, the physical order of these STSs could not be resolved (Figure 1).

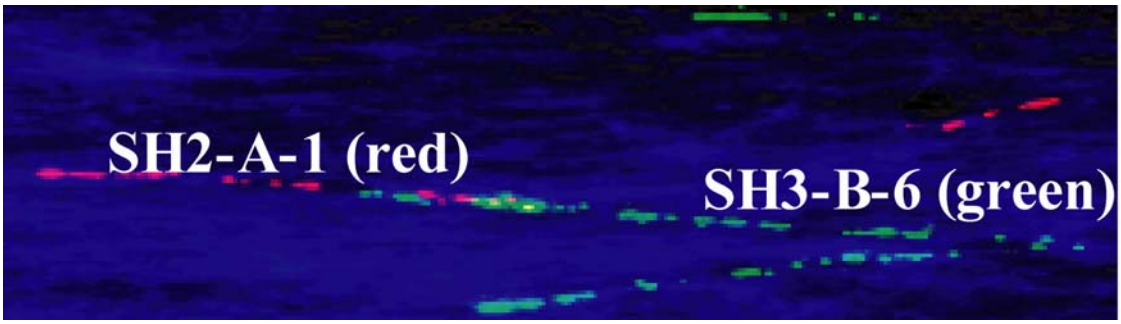

Figure 6 Hybridization with two BACs containing multicopy sequences. Results show that at least one copy of *SH2-A-1* partially overlapped with at least one copy of *SH3-B*$_6$ on fibers

DISCUSSION

The study provides a detailed physical map of the ECAY. Until now, cytogenetic, linkage and RH maps were available for all the autosomes and the X chromosome and ECAY has been completely neglected (as have most Y chromosomes in domestic species). The lack of recombination in the majority of the region of the chromosome prevents construction of a linkage map. Additionally, the small size of the chromosome and the difficulty in developing primers using human and mouse sequence data caused severe setbacks in RH and FISH mapping. Hence developing a map of this chromosome required integration of a number of mapping approaches to validate and confirm our results.

RH, FISH and contig analyses

The 5000rad horse x hamster RH panel allowed the resolution of 23 markers on the Y chromosome which formed the foundation for future analysis. Limitations on the number of markers that could be placed on the RH map were mainly due to the small size of the chromosome, the resolution limit of the panel, and presence of multicopy markers. Such difficulties have also been reported during the mapping of Y chromosomes in other species especially humans (Tilford et al. 2001). In humans a combination of fingerprinting (Tilford et al. 2001), optical mapping (Giacalone et al. 2000), fiber-FISH mapping (Rottger et al. 2002) and STS content mapping (Foote et al. 1992; Ferlin et al. 2003) was used to overcome these problems. Similarly in analyzing the horse Y chromosome integrating RH map along with various FISH approaches allowed us to verify marker, order, orientation, number of copies and distribution.

The markers, together with the additional set developed in this study by us (total 113; 104 STS and nine genes) allowed us to generate an elaborate contig map containing seven groups of 73 overlapping BAC contigs spread over the euchromatic region of the ECAY. Overall, the study represents a major development, because gene maps are presently available for only equine autosomes and the X chromosome (Chowdhary et al. 2003).

Information on the organization of the Y chromosome in mammals other than human and mouse is sparse. At present, a total of 62 loci (primarily microsatellites) are known to have been mapped in cattle (Liu et al. 2002; Liu et al. 2003); eight genes in the cat (Murphy et al. 1999); and ten loci in the dog, of which *SRY* is the only gene (Guyon et al. 2003); and ten genes in the pig (Quilter et al. 2002). Even in rats, (http://ratmap.gen.gu.se/SearchList.html), only nine genes have been provisionally assigned to the Y chromosome. Considering this, the ECAY map presented in this study signifies a substantial advancement over corresponding maps in other species, making it third in mapped loci after human and mouse Y maps.

Organization of ECAY

Until now ECAY consisted of only a rudimentary map based on indirect synteny mapping (Shiue et al. 1999) and FISH mapping of two genes (Hirota et al. 2001). ECAY is one of the smallest chromosomes in the karyotype. It is sub metacentric and largely heterochromatic (Power 1988). ECAY most likely spans ≈45–50 Mb. This is estimated from the relative length of 1.62% (Hansen. 1984) against a 3,000-Mb size of the equine genome. The euchromatic part spans the terminal region of the Yq and is expected to be

≈15 Mb. Developing seven contigs with 73 Y-specific BACs provides a coverage of the ECAY euchromatic region beyond that reported in any other domesticated species. The BACs spanning the minimum tiling path for each of the contigs indicate a cumulative coverage of ≈4 Mb that corresponds to almost 20–25% of the euchromatic region.

Synaptonemal complex analysis suggests the equine PAR on Y to be located on the terminal part of the long arm of Y chromosome (Power et al. 1992). Comparative studies in several mammalian species including primates, cattle, pig, and horse, demonstrate that the 3' end of the *AMEL* locus spans the PAR boundary (Iwase et al. 2003). Hence mapping the *AMLEY* gene to Yqter strongly supports the theory that the likely location of ECAY-PAR is towards the distal end of the euchromatic region. However, more genes need to be mapped to further elucidate this region.

Map alignment and contig development

Markers common to the contigs and the RH/FISH map (highlighted in Figure 2) were critical in aligning the two maps and in deducing the physical order of the seven contigs within the euchromatic region. The STS content map has 7 gaps outside of the heterochromatic region. Two gaps are near the terminal region of the Y chromosome close to the AMELY and ZFY group. We have been unable to identify unique probes to close these gaps despite repeated screenings of TAMU and CHORI libraries with end sequence markers. The apparent absence of BACs covering these two gaps may reflect random fluctuation in the distribution of clones, or the sequences that cannot be cloned in to the BAC vector. Markers that lay at the ends of contigs and STSs that did not fall in

to contigs are being used in further rounds of library screening to close the remaining gaps.

Nevertheless STS content mapping has provided an ordered map of 77 markers spanning from near the proximal to near the terminal end of the euchromatic region. The order of the markers, however, is preliminary, because other than in contigs I and V, the proposed centromere-telomere orientation of the loci could be reversed. However, within individual contigs, the order was cross-verified by STS content mapping and/or fiber FISH. It is noteworthy that the contigs are fairly uniformly distributed across the euchromatic region. Hence, despite the six gaps between the contigs, the map provides a valuable platform for further chromosome walking and contig expansion.

Multiple approaches to develop a Y map

A combination of mapping approaches was simultaneously used to develop the ECAY map. The unusual organization of this chromosome necessitated cross-verification and confirmation of results, because observations from a single approach were inconclusive. For example, the RH map required verification and refinement with a range of FISH mapping techniques, because at seven of the locations, marker order was unclear (see RH map, Figure 2). This is not unexpected, considering that within a small genomic region (15–20 Mb for euchromatic region of Y) the resolution limit of mapping in a 5,000-rad RH panel might easily be saturated with ~20 markers due to insufficient number of breaks in the haploid Y component. In such instances, interphase and fiber FISH proved instrumental in resolving the order of 17 of the 23 mapped loci, thus refining and validating RH and metaphase-FISH data. The approaches were also critical

in confirming the *cen-tel* orientation of the seven BAC contigs. Further, STS content mapping permitted verification of the results. For example, the order of loci *YM2-Y3B8-JARID1D-CLY10-SRY* deduced from the FISH map (Figure 2) was supported independently by STS content mapping (contig I, Figure 1). Similarly, the STS content mapping order of *Y3B1-USP9Y-DDX3Y* (contig V, Figure 1) corresponded to that obtained by fiber FISH (Figure 5). Last, fingerprint analysis of the BACs provided vital supplementary confirmation regarding the grouping and ordering of BACs/markers.

Putative Fertility related Region(s) on ECAY

FISH analyses demonstrated that *TSPY* and several STSs were multi copy sequences clustered in the middle of the euchromatic region. In humans several such multicopy gene families exist and these have been called as ampliconic regions. These regions exhibit a large number of testis specific genes with more than nine protein coding families and have been associated with spontaneous deletions in the Y chromosome (Skaletsky et al. 2003). Additionally such regions have been identified in mouse (Toure et al. 2004) and cattle (Liu et al. 2002) and primates (Charlesworth 2003). Hence it was essential to identify and map this ampliconic region in the horse Y chromosome in order to assist in the search for male-fertility related genes.

Mapping the ampliconic region (shaded block in Figure 1) was extremely difficult even after integration of several approaches (viz. STS content mapping, interphase FISH, or fiber FISH). A number of markers from the ampliconic region cross amplified BACs in other regions when PCR was used. Eight of the STS from this region showed PCR amplification on BAC templates from contigs I, V, and VI. Further, 11

markers from contigs I, IV, and V showed PCR amplification on BACs from contig II. Though this conclusively proved that contig II bore all of the hallmarks of a "multicopy or ampliconic region," the definite order of BACs and markers could not be derived. Therefore, restriction fingerprinting was the only way to deduce a putative order of clones in this region. Despite this, the physical placement of the contig in relation to other contigs is definitive based on (*i*) the location of five of the six markers from this contig in the RH map and (*ii*) interphase and fiber-FISH orientation of BACs from this region, in relation to BACs from contigs I and III.

At present, it is difficult to ascertain the number, distribution, and size of ampliconic region(s) in the horse. Nonetheless, the contig II (Figure 1) assembled by fingerprinting indicates the region to be ~1.2 Mb. Expansion of this contig and identification of other ampliconic regions will help to provide an accurate cumulative estimate. Additionally the proximal part of the region contained the sex determining gene and the distal part included *USP9Y* and *DBY* genes. A number of sex reversal, and intersexuality cases have been associated with translocations of the *SRY* gene on the X chromosome in humans (Zenteno-Ruiz et al. 2001; Li et al. 2004a) and horses.

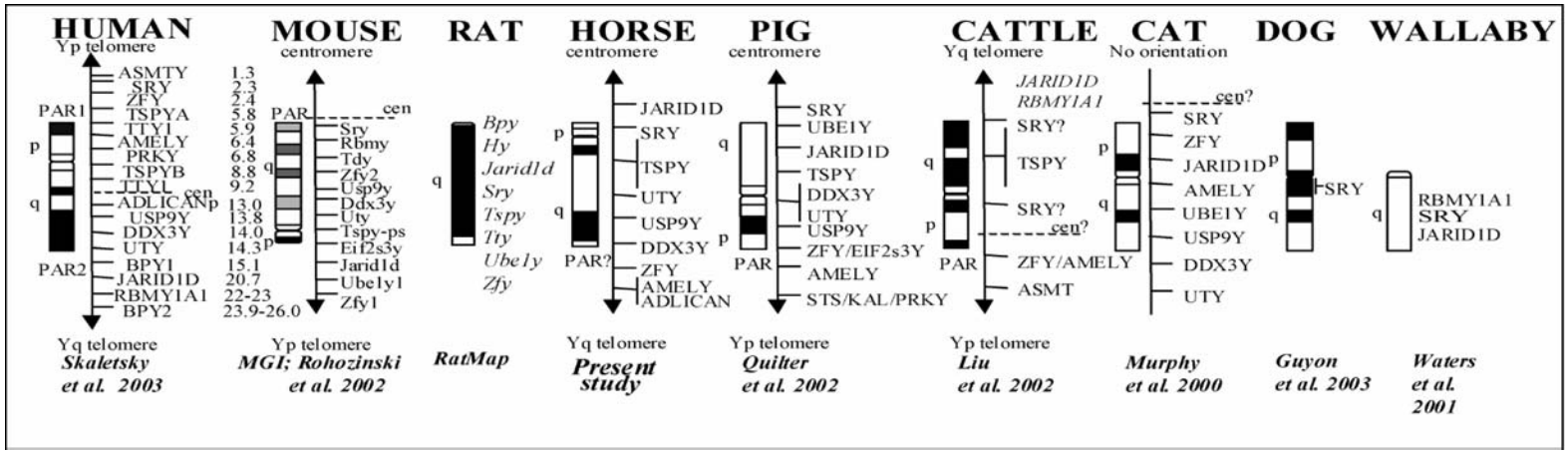

Figure 7 An overview of the comparative status of Y chromosome maps in several mammalian species. ECAY most closely resembles the porcine counterpart. n both species, *ZFY* and *AMELY* are telomeric/distal (like in cattle), whereas *SRY* and *JARID1D* are centromeric/proximal (as observed in cat and mouse but not in cattle and humans). Additionally, *TSPY, UTY, DDX3Y*, and *USP9Y* are clustered together between the two terminal groups in both species, and their order is also essentially conserved.

Comparative map

Earlier cross-species comparisons of Y chromosome by Zoo FISH suggested lack of homology across evolutionarily distantly related species (e.g., humans vs. mouse, cattle pig, dog, or horse; (Chowdhary et al. 1998) but the presence of homology across closely related species, e.g., within primates (Koehler et al. 1995), bovids (unpublished results), and equids (Raudsepp and Chowdhary 1999). Comparison of the relative order of the eight genes on ECAY to that in human, mouse, cattle, cat and pig demonstrates that the Y chromosome of all species differs in morphology and gene order. Morphologically, the Y chromosome is almost meta-/submetacentric in human, cattle, cat, dog, and pig. However, it is almost acrocentric in mouse, rat, and horse. Not more than 10 Y specific genes have been mapped until now in nonhuman/mouse species. A comparative overview of the physical order of these genes in different species (Figure 7) indicates that ECAY most closely resembles the porcine counterpart. In both species, *ZFY* and *AMELY* are telomeric/distal (like in cattle), whereas *SRY* and *JARID1D* are centromeric/proximal (as observed in cat and mouse but not in cattle and humans). Additionally, *TSPY, UTY, DDX3Y*, and *USP9Y* are clustered together between the two terminal groups in both species, and their order is also essentially conserved. Incidentally, these four loci are also clustered together in human and mouse. Other than this, no clear conservation in gene order was observed for genes mapped in different species. It is noteworthy that *DAZ*, which maps to the Y chromosome in humans, is autosomal in the horse and is located on ECA16q22.3 (unpublished results).

Overall, our data provides the first detailed physical map of the horse Y chromosome. This map is presently the most informative among domesticated species and is second to only human and mouse Y maps. This map identifies those regions that may be associated with fertility (ampliconic regions, USP9Y, DBY), intersexuality (SRY) and other biologically important functions (PAR). The presence of the contigs lays an important framework to eventually build a minimum tiling path of BACs across the euchromatic region. This will facilitate the complete sequencing of the Y chromosome. Additionally the map provides locations of several genes which might prove useful in identifying regions that are preferentially deleted in subfertile animals.

GENETIC DIVERSITY IN THE HORSE Y CHROMOSOME

INTRODUCTION

Genetic diversity and domestication of livestock species

Anthropologists, and archaebiologists have traditionally dealt with discerning the evolutionary history, distribution, origin and classification of human beings, plants and animals. By examining remains of archaelogical sites, an anthropologist studies fossil records to analyze the past and present of various current and extinct forms of life, in an attempt to understand their biological, social, and sometimes even cultural variations across a long period of time. But archaelogical records do not necessarily distinguish between wild and domesticated forms of species, making elucidation of events surrounding the history and origin of various breeds of livestock and companion species unclear. Fortunately in the past two decades, a range of molecular markers from the nuclear and mitochondrial genomes have emerged as powerful tools used by geneticists to decipher the domestication of livestock species and to identify their course of evolution from the wild ancestors. Possibilities to use these set of markers has offered a unique perspective not only on the origins of livestock and pet species but also on the origins and distribution of different breeds within each species. This is evident from some recent studies that summarize the use of nuclear and mitochondrial markers toassess the genetic diversity in a number of species including cattle, goat, dog, cat and horse (Bruford et al. 2003; Li et al. 2004b; Moioli et al. 2004; Parker et al. 2004). While polymorphisms on the mitochondrial loci are of specific interest in tracing maternal

origins, the nuclear markers, especially those located on the Y chromosome, provide a unique platform to assess the origins on the paternal side. A summary of the use of various autosomal and X chromosome specific molecular markers in studying evolution and analyzing genetic diversity in different domesticated species is provided below. Horse is discussed separately. Versatility of the markers originating from the Y chromosome (where available) will be discussed thereafter.

Autosomal microsatellite markers and their use in studying domestic animal populations

Microsatellites or STRs (Short Tandem Repeats) are codominantly inherited markers comprising tandemly repeated units of 1-6bp length. Microsatellite mutation rates range from 10^{-6} to 10^{-2} per generation which is significantly higher than base substitution rates (Ellegren 2004). They are highly variable and their polymorphisms are mainly derived from differences in length rather than changes in the DNA sequence. The main mutational mechanism that causes changes in length is proposed to be replication slippage. During DNA replication, the nascent and template strand misalign, which alters the repeat number of the microsatellite sequence (Schlotterer and Tautz 1992). Due to their highly polymorphic nature, the primary use of STRs over the years has been in developing linkage maps using pedigree material, and more lately in genome scan panels for conducting analysis leading to detection of association between markers/haplotypes and diseases/traits of economic significance in domestic species (Liefers et al. 2002; Ashwell et al. 2004). Additionally, microsatellite markers have since long been found to be extremely useful to study genetic variation between and among species (Moioli et al. 2004). They have been abundantly used to detect admixture in

populations (Kumar et al. 2003), and also to identify specific individuals, breeds, species, or lineages (as reviewed before, see previous section) (Ciampolini et al. 1995; Williams et al. 1997). However a few studies have explored the likely involvement of STRs in transcription (Gebhardt et al. 1999; Gebhardt et al. 2000; Iglesias et al. 2004), translation (Chen et al. 2003) and even possible influence on recombination (Gendrel et al. 2000). A brief overview on the use of autosomal microsatellite markers in studying populations of domestic species is presented below.

Cattle/bovids

Twenty highly variable autosomal bovine microsatellites were used to detect the rate and extent of gene flow and the degree of admixture in different African populations of cattle, and to determine the phylogeny of 20 different cattle populations from Africa, Europe and Asia (MacHugh et al. 1997). Marked differences between distribution of alleles in the taurine and zebu cattle indicated that independent domestication events were involved. Further, using allelic frequencies of three bovine microsatellite loci, genetic distance between water buffalo and different bovids was calculated to position buffalo on the phylogenetic tree in relation to cattle, sheep, horse, pig, goat, and sheep. Consequently, the estimated time of divergence between water buffalo and cattle was calculated to be 21million years (Mattapallil and Ali 1999). Surprisingly enough, even though the buffalo shows closer phenotypic identity to cattle, phylogenetically it was shown to share a closer ancestry with sheep and goat.

Genetic variation using autosomal microsatellites was employed to study the genetic structure of different breeds of cattle including European, Asian and North

American populations (Ciampolini et al. 1995; Williams et al. 1997; MacHugh et al. 1998; Vallejo et al. 2003; Mateus et al. 2004a; Mateus et al. 2004b; Moioli et al. 2004). For example, 30 microsatellite markers were used to study genetic diversity between ten Portuguese cattle breeds (Mateus et al. 2004b). More than 350 alleles from these markers were used to build a phylogenetic tree which proved that the genetic relationships between the breeds were consistent with the historical origins of the populations.

Pig

Porcine microsatellite markers have been used to identify parental origins of various commercial breeds, to study the diversity among different European, American and Asian breeds and to estimate the level of introgression between populations (Paszek et al. 1998a; Paszek et al. 1998b; Lemus-Flores et al. 2001; Babicz et al. 2003; Vernesi et al. 2003; Li et al. 2004b). For example, significant associations between allele frequencies of nine autosomal microsatellites markers in diverse pig breeds such as Yorkshire, Hampshire and Meishan breeds was demonstrated. In the same study, genetic distances between the breeds were calculated; divergence time between Meishan and Hampshire (~2000 years) breeds was found to be the oldest (Paszek et al. 1998b). In yet another study, extensive genetic diversity (average heterozygosity 0.73) was detected in 177 unrelated hairless Mexican pigs using allele frequencies of 10 microsatellite loci. The results showed a large genetic distance between these pigs and four commercial breeds (111 pigs from Landrace, Large White, Hampshire, and Duroc). The study proved that Mexican hairless pig was genetically divergent from the commercial breeds (Lemus-Flores et al. 2001). Next, in the Italian wild boar, moderate level of introgression

from Hungarian populations was detected using variance data on nine autosomal microsatellite markers. These findings of introgression in protected populations were found to be significant for implementing better conservation and management strategies for the Italian wild boars (Vernesi et al. 2003).

Dog

Phylogenetic classification (Parker et al. 2004), identification of breeds of origin (Koskinen 2003), and genetic relationships between different dog breeds has been demonstrated in some of the recent studies using microsatellite markers (Kim et al. 2001; Irion et al. 2003; Parker et al. 2004). Allele frequencies of ten microsatellite loci in 250 dogs from five breeds were used to calculate the genetic distance between breeds. Based on the results, it was possible to genetically distinguish different breeds (Koskinen 2003). In another major study, variations in 96 microsatellite loci were assessed in 414 dogs representing 85 breeds (Parker et al. 2004) to evaluate breed specificity. The results led to the detection of breed-specific alleles that served to genetically distinguish diverse dog breeds. In addition, phylogenetic analysis served to separate the ancient dog breeds from the more recent ones. Interestingly, the phylogenetic tree showed that Asian spitzes represented the oldest dog breed.

The above studies demonstrate increasing use of autosomal microsatellite markers for phylogenetic analysis, detection of genetic diversity within and between populations, and for genetic profiling/characterization of individuals.

Mitochondrial DNA markers and their use in studying domestic animal populations

In most mammals, mitochondrial (mt) DNA is less than 20kb in length and is highly polymorphic. The popularity of mtDNA markers is in particular due to a number of factors. These include: lack of recombination, uniparental inheritance (maternal inheritance only), a rapid and constant rate of evolution, and adequate variability across geographically distinct species (Bruford et al. 2003). In humans, and livestock species, the cytochrome b gene and the control region has been extensively analyzed for polymorphic markers since the region appears to be the most variable (Bruford et al. 2003). For example, in humans, more than 500 haplotypes within the mtDNA control region have been identified (Handt et al. 1998). Discovery of similar variation in animals has served as a powerful tool to identify phylogenetic relationships and genetic diversity in different domestic species. A brief summary about the applications of mitochondrial markers in studying the domestication and origin of livestock species is presented below.

Cattle

In cattle, ~ 900bp of the D loop (Displacement loop; the most variable region of mtDNA), was sequenced in 26 animals representing six European, four African, and three Indian breeds (Loftus et al. 1994). The analysis identified 63 polymorphisms. Subsequent phylogenetic analysis demonstrated two distinct mitochondrial lineages: one for the Indian zebu cattle and the other for the European, African taurine, and African zebu cattle (Loftus et al. 1994). The analysis also revealed that the two lineages diverged 200,000 to possibly 1 million years ago. Sequence divergence estimates in the control region of mtDNA led to the conclusion that independent domestication events occurred

in each continental population (Loftus et al. 1994; Bradley et al. 1996). More recently, another study examined 201bp of control region of mtDNA for variation in 392 cattle representing 34 breeds from Europe, Africa, and Asia, along with four extinct British wild cattle (Troy et al. 2001). In the modern cattle from all three continents, 77 polymorphic markers led to construction of 152 haplotypes. Sequence comparisons proved that the wild cattle are distinct from the European modern cattle, and more closely related to cattle near the Middle East.

Pig

A number of studies have examined the control region of mtDNA in domestic pigs, wild boars, and extinct pigs. The entire mtDNA sequence (~15kb) was compared between one European domestic pig, one Asian domestic pig, and two European wild boars (Kijas and Andersson 2001). Considerable sequence divergence was present in the three clades (0.8-1.2%). Phylogenetic analysis showed the presence of two domestication events, one from a European wild boar species and the other from an Asian species. Similar results were demonstrated in other mtDNA studies as well (Giuffra et al. 2000; Kijas and Andersson 2001; Watanobe et al. 2002).

Dog

A 261bp stretch of the control region of mtDNA was sequenced in wolves and dogs (140 dogs representing 67 breeds) from 27 geographical regions worldwide (Vila et al. 1997). A total of 27 & 26 haplotypes were found in wolves and dogs, respectively. Phylogenetic analysis demonstrated that wolves were ancestral to dogs and the origin of dogs was estimated to have occurred 135,000 years ago. A more recent study analyzed

582bp of mitochondrial DNA for variation in 654 dogs from diverse breeds. The presence of four phylogenetic clades indicated that multiple domestication events occurred in the domestic dog population. In addition, the analysis suggested that the geographic origin of domestic dog occurred in East Asia. This was attributed to larger genetic variation (20 haplotypes within 51 East Indian dogs) found in the dogs of this region, compared to other regions (16 haplotypes in 51 Southwest Asian dogs) (Savolainen et al. 2002).

Goat

In goats, 481bp of control region sequence was analyzed in 406 individuals representing 88 breeds. The 331 haplotypes observed for the region led to identification of three divergent lineages (Luikart et al. 2001). The results indicated that due to high divergence in the multiple maternal lineages multiple domestication events had occurred.

In conclusion, studies involving analyses of variation in mitochondrial loci in various domestic species like cattle, goats and pigs have helped identify the occurrence of at least two independent domestication events in East Asia and Europe. In comparison, the domestication of dogs occurred as a number of separate events at different locations around the world.

Analysis of genetic diversity within the domestic horse

Microsatellite markers have been extensively used to study genetic diversity, population sub divisions and extent of gene flow among different equids. For example, comparable levels of diversity was demonstrated between *Equus caballus* and *Equus przewalskii* using 13 autosomal microsatellite markers (Breen et al. 1994). Next, 13

microsatellite loci used in 211 Thoroughbreds horses revealed low levels of genetic diversity. The high level of allele sharing, which is an indicator of low genetic diversity was 0.469; illustrating the closed nature of the Thoroughbred breeding population (Cunningham et al. 2001). Along with pedigree information the data also provided an estimate of the genetic diversity within the founder group (which was lower than the diversity seen in Egyptian and Turkish breeds) and the individual contribution of each founder to the current genetic pool of Thoroughbreds.

Microsatellite variation in several Asian, European, and African breeds of domestic horse has been analyzed to study the genetic diversity between the different breeds (Kelly et al. 2002; Bjornstad et al. 2003; Tozaki et al. 2003; Achmann et al. 2004). For instance, 561 Lippizan horses from several European subpopulations were examined for variation in 18 microsatellite loci (Achmann et al. 2004). As a result, it was possible to genetically distinguish between the different subpopulations. The level of genetic diveristy found (0.675) was comparable to the data seen in other horse breeds reported elsewhere (Cunningham et al. 2001).

In another study, the genetic relationships between 60 Mongolian, 135 Japanese, 21 Korean, 25 Thoroughbreds, and 18 Anglo Arab horses was studied using 20 microsatellite markers (Tozaki et al. 2003). The average heterozygosity was estimated to be 0.62 in each population. Phylogenetic analysis revealed three distinct clusters separating the European and the Asian lineages. The results also showed that Japanese horse originated from Mongolian horses. In addition, genetic relationships between

individual Japanese breeds corresponded to the geographical distribution of each population.

Finally, mitochondrial DNA analyses have revealed a complex pattern of domestication in horse (Vila et al. 2001; Jansen et al. 2002). In the first study, sequence analysis of 616bp of mitochondrial DNA from ancient horse DNA and current domestic breeds revealed at least 6 divergent maternal clades. (Vila et al. 2001). In the second study, 247bp of control region was analyzed and 652 mitochondrial haplotypes were obtained in 318 unrelated horses representing 25 breeds (Jansen et al. 2002). This led to the formation of 17 maternal lineages. To explain the high levels of diversity it was postulated that at least 77 mares were involved in domestication. The underlying conclusion from both studies is that the modern horse had been domesticated a number of times at different locations.

Though the above studies show the use of autosomal and mitochondrial markers to study genetic diversity and domestication in different breeds of domestic horse, a complete overview of the genetic relationship between different horse breeds still remains incomplete because of the lack of the patrilinear-specific component. Isolation and characterization of Y specific markers would provide an informative paternal analogue to study the origin of domestic horse, genetic composition of breeds and the evolution of equid species. The following is a brief summary on the use of Y chromosomal markers in studying genetic variation in natural populations of humans and livestock species.

The study of genetic variation using Y chromosomal markers

The proposed lack of recombination in the male specific region of the Y (MSY) implies that the Y chromosome is inherited as a single block of haplotype. The significance of Y polymorphic markers has been recognized to be ideally suited for population studies and to analyze genetic diversities within and between populations in humans (Jobling and Tyler-Smith 1995; Mukherjee et al. 2001). For Y-chromosomal population studies four types of markers have been generally used. These include indels, SNPs, microsatellites and minisatellites (Hurles and Jobling 2001). Indels are insertions or deletions of the DNA at a particular location on the chromosome. SNPs or single nucleotide polymorphisms are variations that occur due to a single nucleotide change in the genome sequence. Microsatellites or simple sequence repeats (SSRs) are tracts of DNA composed of 1-6 base pairs units that are tandemly repeated whereas minisatellites are repetitive units that are 10-60 bp long. Due to lack of recombination all the above markers are linked along the entire length of the Y chromosome. A set of linked markers called as a haplotype records the evolutionary history of a particular Y chromosome on which it is located and is an important tool for investigating population or evolutionary processes.

In humans, Y-specific polymorphic microsatellite markers have been used to develop a number of haplotypes that reflect ancient as well as recent evolutionary events in the history of human origin/evolution (see section titled ‘Introduction’) (Jobling and Tyler-Smith 1995; Whitfield et al. 1995; Mukherjee et al. 2001; Stumpf and Goldstein 2001; Zerjal et al. 2002; Saha et al. 2003; Cordaux et al. 2004a; Cordaux et al. 2004b;

Manni et al. 2004; Shen et al. 2004). Microsatellite based Y haplotype distribution has also been used for detecting association of male infertility in specific populations in Japan (Kuroki et al. 1999). In addition, the distribution of Y chromosomal diversity in Europe, America and parts of Asia has been used in various genealogical studies (Karafet et al. 2002; Lell et al. 2002; Arredi et al. 2004; Schurr and Sherry 2004; Wen et al. 2004a) to analyze social hierarchical structure and to trace paternal lineages along ethnic boundaries in these populations (Stumpf and Goldstein 2001; Basu et al. 2003).

Analysis of genetic variation in domestic species using Y chromosome markers

The eutherian Y has undergone extensive rearrangements during the course of evolution (Graves 2001; Waters et al. 2001; Skaletsky et al. 2003). Also, due to the generally non-recombining nature, the Y chromosome is evolutionarily labile and thus sequence homology between different species is extremely limited. Consequently, transfer of gene map information for the Y chromosome across species is not easy. This is evident from the fact that Y chromosomal studies related to variation between evolutionarily distantly related species are scarce and very little is known about the cross-species diversity of the Y chromosome in domestic animals.

SNP screening for non coding regions of X and Y chromosomes in five mammalian species (lynx, wolf, cattle, reindeer, and field vole) revealed low levels of diversity in all five Y chromosomes as compared to the X chromosomes (Hellborg and Ellegren 2004). Next, four Y chromosomal microsatellites were used in a range of bovids to assess genetic variability in 56 animals representing African, Indian, Asian bovids includuing river buffalo, bison, and yak (Edwards et al. 2000). A total of 21

alleles were detected in various populations of cattle and other species. Two markers were found to give zebu and taurine specific alleles that could potentially be used for differentiating between the two populations of cattle. In another study, ~40 Y chromosomal microsatellites were evaluated for polymorphisms in 17 unrelated bulls and fourteen polymorphic microsatellites were isolated (Edwards et al. 2000; Liu et al. 2003). As yet, no phylogenetic studies have been initiated in cattle using Y chromosomal polymorphisms.

Two Y-specific microsatellites were used to study polymorphism in 13 unrelated canid males. This led to the identification of 3-5 alleles for each of the markers (Olivier et al. 1999). A single polymorphic Y-specific marker, along with 18 autosomal polymorphic microsatellites and mtDNA loci were used to identify wolf-dog hybrids (Vila et al. 2003). Nine Y-alleles were observed in wolves and eight alleles were seen in the dog. Analysis of the results permitted clear differentiation between wolves and dogs, indicating that crosses between the two species (classified primarily due to their habitat and behavior) were relatively rare events.

In another study, four polymorphic Y chromosome microsatellites were used for constructing 17 different Y haplotypes in a sample of 100 wolves to determine the origins and evolving genetic structure of Scandinavian wolves. Only two haplotypes were found common within the Scandinavian wolves indicating presence of two founders and a low level of genetic migration between wolves from Scandinavia and the neigbouring countries (Sundqvist et al. 2001). No studies have yet been published using Y chromosomal markers to study the evolution of different breeds of dogs.

In sheep, nine Y chromosomal sequences were analyzed for SNPs in 14 rams from seven different breeds (Meadows et al. 2004). A single nucleotide variation (A/G) was identified in four animals, whereas the remaining eight markers were monomorphic.

Lastly, in the domestic horse two published studies have been carried out to identify polymorphisms in the euchromatic region of the ECAY. Both studies failed to detect variation in a wide variety of breeds within the domestic horse (Wallner et al. 2003; Lindgren et al. 2004). In another recent study, six Y-specific microsatellite markers formed a single haplotype in the domestic horse (Wallner et al. 2004). When the same markers were used to determine diversity between different species of equids, polymorphisms were observed on three of the loci, leading to the identification of eight SNPs (Wallner et al. 2003; Lindgren et al. 2004; Wallner et al. 2004). The data was used to construct haplotypes and conduct phylogenetic analysis between *Equus przewalskii* and *Equus caballus*. The findings showed that the two 'species' were in fact sister taxa. However, with regards to the other equids, their genetic relationships could not be elucidated with the tested markers.

To summarize, even though studies using Y-chromosomal markers have been initiated in domestic species to examine genetic diversity, these are not comparable to the extensive number of investigations that have taken place in humans. The lack of data in animals is mainly due to the absence of a detailed Y chromosome map (hence inavailability of markers) in most of the species. Construction of a detailed map of the horse Y chromosome described in the previous chapter (Raudsepp et al. 2004b) has provided us with a valuable resource to initiate development of microsatellite markers in

the horse to assess the level of genetic variation in different breeds of domestic horse as well as the range of diversity in equids.

Rationale of the study

In horses, analysis of mitochondrial DNA has demonstrated a high diversity of matrilines among modern horses (Ishida et al. 1995; Vila et al. 2001; Jansen et al. 2002). The level of genetic variability across different breeds suggests that a large number of wild mares were involved as founders (of the current population of modern horse) during multiple geographically distinct domestication events. In addition to mtDNA, the genetic relationships of various horse populations have also been investigated using autosomal microsatellites (Tozaki et al. 2003; Achmann et al. 2004). Though a marked divergence has been shown between breeds in most of the studies, no clear cut phylogenetic relationships has yet been elucidated between different breeds.

It should be noted that results obtained using mitochondrial and autosomal markers reflect the evolutionary processes that affect both sexes (autosomal markers) or only the females (mitochondrial). However, markers differentiating male and female specific lineages are still missing. Ideally Y-specific markers would supplement the paternal analogue and would be significantly useful to comprehensively elucidate the patterns of domestication in the horse and evolution in other equids.

Recent studies have shown that Y chromosomal markers provide information on male specific lineages and are an important tool for studying natural populations in humans (Hurles and Jobling 2001). Hence to study the genetic contribution of stallions

during horse domestication, Y chromosomal markers demonstrating variance across breeds are essential.

As far as known, we are unaware of any detailed investigations initiated to isolate polymorphic markers in the horse Y chromosome. In fact, previous attempts to isolate Y-specific polymorphic markers (SNPs and microsatellites) within the domestic horse breeds have been singularly unsuccessful (Wallner et al. 2003; Lindgren et al. 2004; Wallner et al. 2004). It is possible that the domestic horse has an inherently low variability within the Y chromosome. However, it is also possible that the less than 1% of the Y chromosome analyzed by the two research groups (Lindgren et al. 2004; Wallner et al. 2004) may not be representative of the entire chromosome. Another reason could be the limited number of motifs tested for the isolation of microsatellite markers. This is evident from the study by the Austrian group where only (CA) motifs were used to screen for markers (Wallner et al. 2003). From the six microsatellites analyzed, the repeat length of four markers was less than 24bp. It has been postulated that short microsatellite repeats are stable and less prone to polymorphisms than those with more repeats; this could probably also be one of the causes for lack of variation seen on ECAY (Rose and Falush 1998). A more detailed analysis using microsatellites with a number of different motifs may help to circumvent this problem. Lastly, lack of variation in the horse Y may also originate from the sparse or limited distribution of patrilines in the course of domestication of the horse. Irrespectively, the lack of sufficient number of Y-specific markers hinders obtaining a true estimate of patrilinear variability (that

arising from the Y chromosome) within and across populations. It also is an obstacle in obtaining similar estimates for the 8 other extant equids.

The proposed study is therefore the first methodical attempt aimed at isolating microsatellite markers originating from di, tri, tetra, and pentanucleotide motifs across the euchromatic region of the horse Y chromosome. The generation of seven BAC contigs spanning a considerable region of the Y chromosome recently developed by us (Raudsepp et al. 2004b) provides a valuable resource to isolate microsatellites and study genetic variability in a region that constitutes more than 25% of the euchromatic part of the Y chromosome versus the less than 1% studied earlier. Additionally, the use of these microsatellite markers across various breeds, as well as a range of equids, will help to identify the degree of variation in the Y chromosomal haplotypes. This will particularly help to shed light on the extent of patrilinear diversity during domestication of the horse as well as during breed diversification, and in general also on the evolution of equids.

The goal for the proposed part of the project is to screen for microsatellite markers in BACs from the seven contigs presently available on the Y chromosome. These contigs are uniformly distributed across the euchromatic region of the horse Y chromosome. Genotyping markers obtained from these regions in geographically and phenotypically distinct horse breeds and various extant equids will provide an overview of between-breeds and between-species male specific genetic variation in equids. These goals will be accomplished through the following specific objectives:

1. Generate 5X coverage subclone libraries for individual BACs from the minimum tiling path on parts of the euchromatic region of the horse Y chromosome.

2. Screen the subclone libraries with di, tri, and tetranucleotide repeat motifs, and identify and sequence subclones containing the repeats.
3. Design primers flanking the microsatellite repeats and optimize PCR conditions on male and female horse genomic DNA.
4. Genotype the markers in diverse breeds of domestic horse and also in a range of equids.

MATERIALS AND METHODS

Pulse field gel electrophoresis

DNA for individual BACs selected from the minimum tiling path of the seven contigs on the euchromatic region of the horse Y chromosome was isolated using Montage^{TM} BAC_{96} Miniprep Kit (Millipore Billerica, MA). Briefly, inserts from individual BAC clones were released by *Not*I (New England Biolabs) digested, and subjected to PFGE analysis on a CHEF-DR II system (Biorad, Hercules, CA) on 1% agarose gels. The following conditions were used: 6V/cm, pulse time 5-15s, 0.5X TBE for 16 hours at a reorientation angle of 120^{o}. Low range PFG marker (New England Biolabs, Beverly, MA) was used as a DNA size standard.

Pooled BAC DNA subcloning

On the basis of insert sizes, 200ng of individual BAC DNA was pooled together into four groups. Subsequently, a limited *Sau*3AI digest was carried out. Briefly, pooled BAC DNA was digested for 15 minutes with *Sau*3AI (Promega, WI**)** (1unit/µg of DNA) in a total volume of 75µl. The reaction was inactivated at 65°C. Reactions were purified using QIAEXII kit (Qiagen, CA) and separated on a 0.65% agarose gel containing

ethidium bromide (0.5 μg/ml) and visualized under UV light. pBluescript Vector DNA was digested with *Bam*HI (Promega, WI**)** and dephosphorylated using Shrimp Alkaline Phosphatase (USB, OH). Ligations were carried out using T4 DNA Ligase (NEB, MA) and subsequently transformed into electrocompetent cells. The cells were then transferred to 1ml of SOC medium and incubated at 37°C for 45 minutes. 50ul and 100ul dilutions of each transformation mixture were plated on to LB agar plates containing 50 ug/ml ampicillin (Sigma, MO) and X-gal and IPTG (MoleculA, Sterling, VA) to get an estimate of the titer. Subsequently each mixture was plated on seven LB agar plates with ampicillin 50 ug/ml to give ~500 colonies per plate.

Screening and isolation of microsatellites

After an overnight incubation of bacterial plates at 37°C, filters were prepared using the colony lift approach. Filters were denatured and neutralized using Southern OH and Southern Tris buffers. Fixation of single stranded DNA onto filters was carried out at 80^{o}C for 60 min. All filters were washed and treated with 1X pre-hybridization solution at 65°C prior to hybridization.

Filters were subjected to a single round of screening with a mixture of different microsatellites viz. $(CA)_{12}$, $(TATC)_6$, $(GAAA)_4$ as probes. The probes were labeled in a 10ul reaction containing 1uM forward primer, 1 uM reverse primer, 150 Ci/mmol each of ^{32}P dATP and dCTP (Amersham Biosciences, Piscataway, NJ), 2U Klenow fragment DNA polymerase (Roche, Indianapolis, IN) and 1X DNA polymerase buffer (Promega, Madison, WI) at 37^{o}C for 30 min. Unincorporated nucleotides were removed using Sephadex G-10 gravity flow columns. The labeled microsatellite probes were pooled and

hybridized onto filters at 42°C for 16h. Each filter was washed two times for 15 min in 2X SSPE. The filters were exposed to film over intensifying screens for 24-48h at –80°C and the autoradiograms developed. Each positive clone was subsequently picked and rescreened to identify and confirm positives.

Construction of individual BAC subclone libraries

24 BACs that form the minimum tiling path in each of the seven contigs developed by us (Raudsepp et al. 2004b) were selected for construction of subclone libraries. Briefly, DNA was extracted from individual BACs using the alkaline/lysis procedure, and organic extraction. The purified BAC DNA was digested with *Pst*1 enzyme and ligated into the *Pst*1 site of dephosphorylated *pBluescript* SK+. The vector and insert construct were then used to transform competent *E. coli* cells using electroporation. Transformed cells were plated on LB agar containing selective antibiotic (50 μg/ml ampicillin), IPTG and X-gal (MoleculA, Sterling, VA). Blue/White colony selection procedure was used to pick 1000 clones per BAC into 384 well plates using the Q-bot system (Genetix, UK).

Screening and isolation of microsatellites from individual subclone libraries

Each subclone library was duplicated and spotted on nylon membrane filters (Amersham Hybond N+). DNA on filters was denatured and neutralized using Southern OH, and Southern Tris solutions respectively. Fixation of single stranded DNA onto filters was carried out at 80°C for 60 mins. Screening for microsatellites was carried out as described above using the following seven motifs $(CA)_{12}$, $(ATT)_8$, $(ATA)_8$, $(GATA)_6$, $(TAGA)_6$, $(CAGA)_6$, $(GAAA)_4$ as probes.

Each positive subclone was identified and picked onto a 96-well plate for subsequent DNA isolation using Millipore Montage plasmid 96 kits on a Qiagen bioRobot 3000. Subsequently all positive clones were pooled in to six groups and subcloned using *Sau*3AI as described above to get insert sizes of 1-3 kb. ~6000 clones were picked in to 384 well plates (10 subclones per positive) and filters were prepared by dot blotting. Filters were re-screened to identify positive clones as described above.

Sequencing and analysis

Plasmid end sequencing was carried out using universal end sequencing primers (T7 forward, M13 reverse, SP6 reverse). Dye terminator sequencing reactions (total volume 10ul) were set up as follows: 1µg plasmid DNA, 2µL standard sequencing primer (10 µM solution), 2µL BigDye (Big Dye Terminator v2.0 Cycle Sequencing Kit Applied Biosystems), 2µL half BigDye (*half*BD^{TM}, Genetix), 0.5µL Master Amp (Amp PCR Enhancer, Epicentre, Madison, WI). Cycle sequencing was performed in a Gene Amp PCR system 9700 thermocycler using the following cycling conditions: 96°C for 2min followed by 8 cycles of 96°C for 30 sec, 58°C to 50°C touchdown for 10 sec, 65°C for 4 min, followed by 60 cycles of 96°C for 30 sec, 50°C for 30 sec, 60°C for 25 sec, and one cycle of 96°C for 1 min, 50 °C for 1 min, 65°C for 15 min. Reaction products were purified using spin columns (Spin-50, Bio Max) and loaded on an ABI 3100 automated capillary sequencer (PE Applied Biosystems).

Selected sequences were base called and analyzed to remove the vector sequences using the PHRED (Ewing and Green 1998) and PHRAP program (http://www.phrap.com). Sequences were compared against each other to remove

redundant clones. Unique sequences were classified by their similarities into different sequence families using Sputnik. Sequences were compared for homologies to known sequences using the BLASTN program (http://www.ncbi.nih.gov/BLAST/). Selected sequences flanking the microsatellite repeat were then chosen for primer design using the program Primer 3 (http://www-genome.wi.mit.edu/). PCR was carried out in 10 µl reactions containing 50 mM KCl, 10 mM Tris HCl (pH 8.3), 1.5 mM $MgCl_2$, all four dNTPs (each at 0.2 mM), 0.3pmols/µl of each primer, 50 ng of plasmid DNA as template, and 0.025 unit of *Taq* polymerase (Sigma). After initial denaturation at 95°C for one min, 30 cycles of amplification was carried out as follows: 30s at 94°C (DNA denaturation), 30s at 55 - 65°C (primer annealing), and 30s at 72°C (primer extension) followed by a 5-min final primer extension at 72°C. PCR products were separated on a 2.0 % agarose gel containing ethidium bromide (0.5 µg/ml) and visualized under UV light.

Genotyping

Genotyping was carried out as described in (Oetting et al. 1995). The 5' end of the forward primer was tailed with a 19-bp extension, identical to the sequence of an M-13 primer 5'TTTCCCAGTCACGACGTTG3'. Genotyping was carried out with the addition of an M-13 primer-dye conjugate as the sole primer conjugated to the fluorescent dye. PCR was performed in a 10 ul volume containing 50 ng of DNA, 2 µM of forward tailed primer, 5 vM of reverse primer, 5 µM of M-13 primer-dye (FAM) conjugate, 450 µM of dNTPs, 10 uM of buffer, 0.5 units of Taq (Qiagen Hot Star). After initial denaturation at 95°C for 20 min, 30 cycles of amplification were carried out as

follows: 30s at 94°C (DNA denaturation), 30s at 56°C (primer annealing), and 30s at 72°C (primer extension) followed by a 5-min final primer extension at 72°C. Alleles were sized relative to the internal size standard Rhodamine Mapmarker (Bioventures, TN) using Genescan version 3.7 (Applied Biosystems).

Animal populations

Microsatellite variability was evaluated on DNA from 33 male horses belonging to 14 different domestic horse breeds (Standardbred AngloArab, Wielkopolska/ Hanoverian, Thoroughbred, QuarterHorse, Selle Francais, Draft (cross), German riding horse, Fresian, Caspian, Haflinger, Suffolk, Akhal Teke, Shetland Pony). In addition, DNA samples from the following eight equids were analyzed: 1 Przewalski's horse (*Equus przewalskii),* 1 Donkey *(Equus asinus),* 1 Onager *(Equus hemionus onager),* 1 Kiang *(Equus kiang),* 1Grevy's Zebra *(Equus grevyi),* 1 Hartmann's Zebra *(Equus zebra hartmannae),* 1 Damara Zebra *(Equus burchelli antiquorumi),* 1 Grant's Zebra *(Equus burchelli boehmi).*

RESULTS

Isolation of microsatellites from sub clone libraries of pooled BAC DNA

Isolation of ECAY-specific microsatellite markers was carried out using two different approaches. In the first approach 40 BAC clones were pooled into four groups on the basis of insert sizes to ensure equimolar concentrations of BACs. Subsequently, subclone libraries were constructed for each group of BACs. Approximately $1.5x10^4$ subclones were blotted on filters using the colony lift approach. Hybridization of three microsatellites probes $(CA)_{12}$, $(TATC)_6$, $(GAAA)_4$ on these filters yielded 125 positives.

A second round of screening was carried out to in order to verify the positive clones. A total of 79 subclones were picked for further analysis. After DNA isolation and sequencing (see section on sequence analysis), seven novel microsatellite markers were identified.

Isolation of microsatellites from subclone libraries of individual BACs

Individual subclone libraries were constructed for 24 BACs that formed the minimum tiling path across the analyzed euchromatic region of the horse Y chromosome (refer previous section) (Raudsepp et al. 2004b). Subcloning efficiency varied for each BAC. Majority of the BACs (16) were successfully subcloned to give on average 1152 sub-clones per clone. However, for the remaining 8 BACs, the number of subclones ranged from 384-768. Nevertheless with an average insert size of 5-10 kb, more than 9X coverage was obtained for each of the BAC clones.

In the first round of screening, a total of 2.3 x 10^4 subclones were hybridized with seven microsatellite probes $(CA)_{12}$, $(ATT)_8$, $(ATA)_8$, $(GATA)_6$, $(TAGA)_6$, $(CAGA)_6$, $(GAAA)_4$. 558 positive subclones were isolated and re-screened to verify the positives. Since the insert sizes of the clones ranged from 3-10kb, isolation of microsatellites by end sequencing was difficult. Hence further subcloning was carried out. In the final round of screening of ~6000 subclones (1-3 kb in size), 144 subclones were identified for DNA isolation and sequencing from both ends.

Sequence analysis and primer design

The raw traces from positive subclones obtained from both approaches were base-called and trimmed to remove low quality sequences using Phred (Ewing and

Green 1998). Vector trimmed sequences were subsequently used to create a database to identify redundant sequences amongst the group. Out of the 288 sequences analyzed, 142 unique sequences were selected for further analysis. These sequences were screened for repeats using RepeatMasker and the masked sequences were compared with the NCBI (National Center for Biotechnology Information) by the use of BLASTN to identify previously isolated markers. 75 unique sequences were then chosen for microsatellite identification using Sputnik. A total of 43 unique microsatellite markers were isolated. These markers contain 38 dinucleotide motifs, four trinucleotide motifs, 18 tetranucleotide motifs and two pentanucleotide motifs (Table 2). 20 markers had a repeat length > 24bp, and 23 markers had a repeat length ranging from 10-22bp.

Further, 23 markers contained perfect repeats while the rest contained imperfect repeats (Table 2). Primer sequences were designed for 32 of these markers; 11 sequences that did not contain flanking sequence for primer design or those that contained LINE1 elements, were not included for genotyping.

Table 2. Summary of characterization of microsatellite markers isolated from the horse Y chromosome

Criteria for characterization	**Number**
Total number of microsatellite loci	43
Perfect repeats	23
Imperfect repeats	20
Length ≥ 24bp	20
Length < 24bp (10-22bp)	23
Dinucleotide repeats	38
Trinucleotide repeats	4
Tetranucleotide repeats	18
Pentanucleotide repeats	2
Primers and optimization	32
Genotyped markers	27
Male specific markers	21

Marker optimization and mapping

Of the thirty two loci, twenty seven markers were successfully optimized on male and female horse DNA and 21 markers proved to be male specific. All optimized markers were mapped onto BAC contigs constructed previously by us (refer previous section) (Raudsepp et al. 2004b). 28 loci were assigned successfully while the location of four loci remained ambiguous (Figure 8). The distribution of loci was essentially random. Of the nine loci that mapped to the ampliconic region (Figure 8), eight contained dinucleotides; only one locus contained tetranucleotide repeats. Six loci mapped to contig I. Three of these loci contained tetranucleotide repeats.

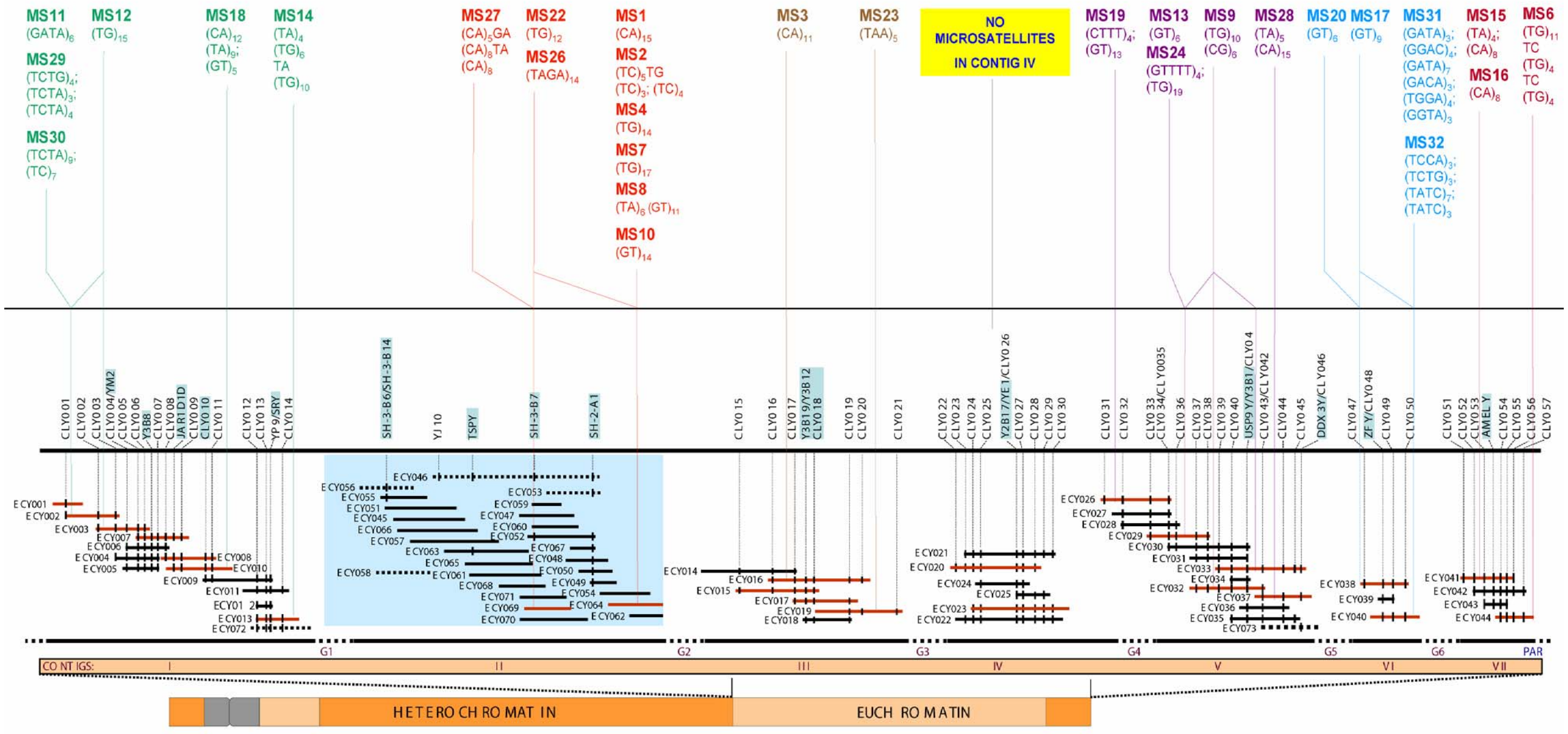

Figure 8 Localization and ordering of microsatellite loci over the euchromatic region of ECAY. (Top) Represents microsatellite loci ordered on the BAC contigs using PCR. (Below) A Contig and STS map of 73 BACs over the euchromatic region of ECAY. BACs utilized for subclone library construction are shaded red.

Table 3. The breeds and number of horses in each assessed for genetic variability

Breed	**Number of Horses**
Standardbred	2
AngloArab	1
Wielkopolska/Hanoverian	1
Thoroughbred	4
QuarterHorse	4
Selle Francais	1
Draft (cross)	1
German riding horse	2
Fresian	3
Caspian	3
Haflinger	3
Suffolk	2
Akhal Teke	3
Shetland Pony	3
14 breeds	**33 animals**

Five loci one of which contains a tetranucleotide, and one a pentanucleotide repeat, mapped to contig V. Three loci each mapped to contigs VI (one tetranucleotide) and VII respectively. Only two loci mapped to contig III, of which one was a trinucleotide repeat. Finally, no microsatellites were isolated from contig IV.

Genotyping across different horse breeds

Twenty-seven Y chromosomal markers were genotyped using the 5' M13-tailed primer protocol (Oetting et al. 1995). Microsatellite variability was evaluated with DNA from different breeds of domestic horse that represented substantial geographic and phenotypic variation (Table 3). Thirteen markers were genotyped on 33 male animals representing 14 breeds, while the rest were genotyped in a sample of 20 domestic horses representing 8 breeds. Although males from 14 different breeds were analyzed, only one marker, MS30 was found to be polymorphic across different breeds of domestic horse (Table 4). This marker was mapped by PCR to a BAC on contig I (refer previous section) (Figure 8). Even though this marker was not male specific, three Y–specific alleles (257, 261, 269) could be detected with certainty (Table 4).

Table 4. The MS30 [$(TCTA)_9$; $(TC)_7$] genotyping results observed in horses of 15 different domestic horse breeds

Number	Species	X	Y
1	Fresian	261	261
2	Fresian	258/262*	258/262*
3	Fresian	261	261
4	Thoroughbred	261	261
5	Thoroughbred	261	261
6	Thoroughbred	257/261*	257/261*
7	Caspian	261	261
9	Caspian	257	257
10	Haflinger	261	261
11	Haflinger	261	261
12	Haflinger	253/261*	253/261*
14	Suffolk	257	257
15	Suffolk	258/269*	258/269*
16	Akhal Teke	257/261*	257/261*
17	Akhal Teke	257/261*	257/261*
18	Akhal Teke	261/265*	261/265*
19	Shetland Pony	257/269*	257/269*
20	Shetland Pony	269	269
21	Shetland Pony	261	261
A10	Standard Bred	258/262*	258/262*
B20	Anglo-Arab	257/261*	257/261*
C28	Wielkopolska/Hanoverian	257	257
E10	Thoroughbred	257/261*	257/261*
F42	Quarter Horse	261	261
G35	Selle Francais	257	257
H1	Standardbred	261	261
K21	Draft (cross)	257/261*	257/261*
L5	German riding horse	262	262

Table 4 continued

Number	Species	X	Y
N13	German riding horse	261	261
RH male	Quarter Horse	257/261*	257/261*
E male	Quarter Horse	261	261
B male	Quarter Horse	258/269*	258/269*
C male	Quarter Horse	261	261
L female	Quarter Horse	261/265*	261/265*
F female	Quarter Horse	261	261
G female	Quarter Horse	261	261
N female	Quarter Horse	261/269*	261/269*

*represents alleles whose assignment to X or the Y chromosomes is ambiguous. Three alleles 257, 261, and 269 could be assigned to the Y chromosome with certainty.

Eight markers also produced a pattern with multiple peaks across all breeds of domestic horse; six markers were male specific (Table 5).

Variation in Y-specific markers across different equids

DNA from *Equus przewalskii, Equus asinus, Equus burchelli antiquorumi, Equus burchelli boehmi, Equus grevyi, Equus hemionus onager, Equus kiang* and *Equus zebra hartmannae* (one animal each) was tested for variability. A single variable marker (MS27) in *Equus przewalskii* could be distinguished from the domestic horse. For the remaining seven breeds, out of 27 markers, 24 markers could be analyzed under less stringent PCR conditions. MS4, MS9, and MS10 failed to amplify all but *Equus caballus, and Equus przewalskii* even under less stringent conditions. Genotyping revealed 21 novel microsatellites markers that demonstrated considerable variation across all species.

Table 5. Observed allele sizes for Y chromosome specific microsatellites in ten different species of equids

Species	***Equus caballus***	***Equus przewalskii***	***Equus asinus***	***Equus hemionus onager***	***Equus kiang***	***Equus grevyi***	***Equus zebra hartmannae***	***Equus burchelli antiquorumi***	***Equus burchelli boehmi***	***Marker Designation***
MS14	302	302	285	285	285	294	285	283	285	Single copy
MS15	372	372	370	374	374	372	372	372	372	Single copy
MS16	201	201	203	199	199	199	199	201	201	Single copy
MS17	217	216	213	218	218	214	218	218	220	Single copy
MS19	211	211	231	216	218	213	216	228	222	Single copy
MS20	170	170	172	172	172	170	172	172	172	Single copy
MS21	206	206	202	202	202	205	202	202	202	Single copy
MS27*	188	186	180	166	164	180	176	178	178	Single copy
MS31*	420	420	440	452	453	396	400	400	404	Single copy
MS1	195, 199	199, 201	197, 203	193, 197	202	199, 203	202, 204	202	204,206,208,210	Multicopy
MS2*	318, 320	320	324	320	318	320	316, 320, 324	320	320	Multicopy
MS3*	231, 235, 237, 241	231, 235, 237, 241	230	230	230	230	230	230	230	Multicopy
MS6*	196	196	_	192	192	200,202	196,198	181	181	Multicopy
MS8*	288, 293	291, 293	_	_	288,292,294	289	289	287,291	289	Multicopy
*MS11**	216, 200, 220,229,	212, 220, 229	_	221, 229	229	225	221, 229	230	230	Multicopy
MS18*	303, 305, 325	303	299, 313	305, 313	305, 315	319	319	325	327, 333	Multicopy
MS22	388	388	_	380,382	382	_	_	384	_	Multicopy
MS28*	188	189,191	190	192	194	196	187	_	_	Multicopy
*MS29 **	399, 403	399	399	399	399	399	399	399	384	Multicopy
MS30	253, 257, 261, 265, 269	257	241	247, 249	249	267, 271	251, 255	251, 254	251	Multicopy
MS32*	316	316	337	349	349	292	296, 300	296	300	Multicopy
MS23*	244	244	244	244	244	244	244	244	244	Monomorphic
MS13	261	261	261	261	261	261	261	261		Monomorphic
MS25*	254	254	254	254	254	254	254	254	254	Monomorphic

Table 5 continued

Species	***Equus caballus***	***Equus przewalskii***	***Equus asinus***	***Equus hemionus onager***	***Equus kiang***	***Equus grevyi***	***Equus zebra hartmannae***	***Equus quagga antiquorumi***	***Equus burchelli boehmi***	***Marke Designationr***
MS4*	260, 262	260, 265	_	_	_	_	_	_	_	Negative
MS9	174	174	_	_	_	_	_	_	_	Negative
MS10	187	187	_	_	_	_	_	_	_	Negative

Allele designations correspond to allele sizes measured by capillary electrophoresis (ABI 3100; Applied Biosystems, CA). Microsatellite markers in italics show a multiple peak pattern. Markers in bold are male specific. * identifies markers that were genotyped on 20 horses representing 8 breeds. The remaining markers were typed on 33 animals representing 14 breeds. – represents unsuccessful amplification of markers.

across species (see Table 5). Nine single copy markers demonstrated ~ 37 alleles between various species. The number of alleles ranged from 2-7 in these markers. The most informative microsatellites were MS19 and MS27 that demonstrated 7 alleles each across the nine species of equids. Three additional markers (MS13, MS23, and MS25) demonstrated no variation across different equids.

DISCUSSION

Comparison of the two approaches to develop subclone libraries from BACs

Initially, screening of pooled BAC DNA subclone libraries yielded seven novel microsatellite markers (MS15, MS16, MS17, MS18, MS19, MS20 & MS21). All microsatellites isolated using this approach contained only dinucleotide repeats. The obvious disadvantage of this procedure was the colony lift approach due to which every new screening or isolation of new markers required preparation of new filters. Nevertheless this approach was highly efficient and cost effective in terms of rapidly isolating markers. As compared to this, the approach wherein subclone libraries from

individual BACs were used for screening, the advantage was that firstly, more number of microsatellites could be obtained, and secondly, the motifs were not restricted to only dinucleotide repeats; microsatellites with tri, tetra and pentanucleotide repeats could also be obtained. Moreover, every marker in the latter approach could be readily mapped back to the parent BAC. However, the drawback with this approach was that it was labor intensive, time consuming and expensive. In addition, since the insert sizes of the subclones from these libraries ranged from 3-10kb, direct end sequencing did not always reveal microsatellites, and further subcloning was necessary.

Applications of individual BAC subclone libraries

The subclone libraries will be valuable for generating new STS markers. It is evident that a high resolution STS map will be essential for screening a panel of subfertile stallions to identify deletions or rearrangements that maybe present in the affected animals. In addition generation of new markers along with the 77 STSs already isolated (refer previous section) (Raudsepp et al. 2004b) can be used to conduct a detailed analysis for SNPS in the horse Y chromosome. The subclone libraries will also provide a significant resource for fulfilling our long term goal of generating a complete sequence of the Y chromosome.

Characterization and distribution of microsatellites across the euchromatic region of the horse Y chromosome

Screening ~ 4.5Mb of the ECAY chromosomal region led to the isolation of 43 microsatellite loci. Five of these loci could not be analyzed because it was not possible to design suitable primers for them. This left us with 37 novel markers. Optimization of

the primer pairs from these markers on control male and female DNA showed that 21 of them amplified a male specific band, 11 gave a X/Y pattern with same band sizes in both sexes, while 5 showed no amplification. Thus, 32 new microsatellites were obtained by us, increasing the number of presently available loci in the horse Y chromosome from 6 to 38. This work thus provides a basic resource for future expansion in this area.

The newly identified microsatellites originate from different segments of the euchromatic region of the Y chromosome. The distribution in terms of the contig map we presently have is as follows: ~18% originate from BAC contig I, 32% from the putative ampliconic region (contig II), 7% from contig III, 17% from contig V, and 10% each from contigs VI and VII (Figure 8) (for localization to individual BACs see Figure 8) (Raudsepp et al. 2004b). The absence of microsatellite markers from contig IV, three BACs (ECY1, 7, 8) from contig I and one BAC (ECY15) from contig III is intriguing. This could possibly be due to a high number of repetitive sequences that would result in decreased cloning efficiency. This is validated by the fact that only 384 subclones were obtained for BACs ECY1, 8 and 15 compared to ~1000 subclones that were obtained for the rest of the BACs.

Secondly, it is possible that compared to the other regions analyzed, the number of microsatellites are lower in these regions. For ECY7 for which more than a 1000 subclones were obtained, an inherent lack of microsatellites seems like a valid explanation. Finally, given that two previously isolated microsatellites (Wallner et al. 2004) have been mapped to contig IV by us (Figure 1) (Raudsepp et al. 2004b), it is likely that no more microsatellite repeats are present in that region..

Presence of multicopy microsatellite markers

Electropherograms of eight markers (MS1, MS2, MS3, MS4, MS8, MS11, MS18, MS29) showed multiple amplification peaks (Table 5 and Figure 9). Of these, 6 markers, viz., MS1, MS2, MS3, MS4, MS8, MS18 amplified male specific bands while the remaining two, MS11, and MS30 gave similar amplifications in males & females. One of the possible explanations for multiple amplification products could be that the markers originate from sequences still shared between X and Y chromosomes – degenerate sequences – as indicated earlier in humans (Wolf et al. 1992). Alternatively, these markers may be present in multiple copies throughout the Y chromosome. If present in the pseudoautosomal region, they may result in amplification of both male and female DNA, as seen for two of the markers.

In humans, duplications and triplications of Y chromosomal microsatellites and SNPs has been widely reported (Santos et al. 1996; Butler and Schoske 2004; Sanchez et al. 2004). These markers have been shown to be highly polymorphic, and have been used for constructing haplogroups (Sanchez et al. 2004). Multicopy Y chromosomal markers have also been isolated in cattle and horse (Liu et al. 2003; Wallner et al. 2004). As indicated in humans, precise asignment of such markers is difficult because they are scattered across the chromosome (Jobling and Tyler-Smith 2003; Sanchez et al. 2004). Hence, sequencing of each variant will probably help to design locus-specific primers, which in turn will ensure identification of unique loci.

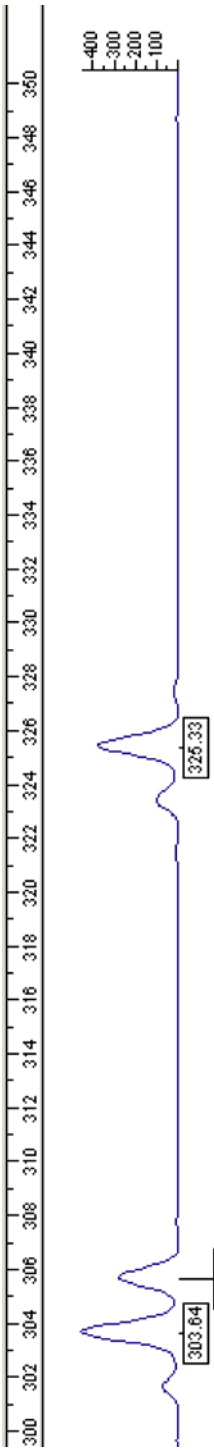

Figure 9 Electropherogram showing multiallelic peak pattern of MS 18.

Polymorphism – or lack of - on the horse Y chromosome

Twenty six of newly characterized microsatellite markers were monomorphic in different horse breeds. The only exception was marker MS30 (Table 4). This marker originates from BAC ECY002 that maps to contig I of the euchromatic region on the Y chromosome and has been previously localized by FISH to both X and Y chromosomes. Overall, the findings reflect similar observations reported in other recent studies (Wallner et al. 2003; Lindgren et al. 2004; Wallner et al. 2004) except that we report the first polymorphic microsatellite marker isolated from the chromosome. It was impossible to characterize the heterozygotes (X/Y) because of difficulties of distinguishing the X and Y specific alleles (Table 4). Sequencing of amplification products from homozygous individual for three different alleles of MS30 showed variability in the length of a tetranucleotide repeat $(TCTA)_9$ with the motif being repeated 8, 9 and 11 times. Of the thirty three animals tested, thirteen were homozygotes for $(TCTA)_9$ allele, four for $(TCTA)_8$ & one for $(TCTA)_{11}$. No obvious geographical distribution could be seen other than the fact that $(TCTA)_9$ seemed to be most common allele across all breeds, and $(TCTA)_{11}$ the least common allele, was found only in Suffolk and Shetland pony.

The lack of polymorphisms on the Y chromosome shows a clear contrast when compared to the amount of genetic variability observed on the rest of the nuclear genome, as indicated through variability in autosomal microsatellites and the mitochondrial component (Vila et al. 2001; Jansen et al. 2002; Kelly et al. 2002; Bjornstad et al. 2003). The results are also significantly different from the genetic variation that has been observed on the Y chromosome in humans, where more than 160

polymorphic Y microsatellite markers have been identified (Kayser et al. 2004). Moreover, in domestic species such as sheep, cattle and dog, polymorphic Y-specific markers has been reported (Olivier et al. 1999; Edwards et al. 2000; Hurles and Jobling 2001; Jobling and Tyler-Smith 2003; Meadows et al. 2004).

A number of factors influence the extent of genetic variation on a chromosome. These include recombination, selection, and effective population size. The presence of a predominantly single Y chromosome haplotype in domestic horse reflects a significantly lower contribution of the stallions to the gene pool than the mares. This hypothesis is further reinforced by the fact that traditionally a select number of stallions are used to cover more than 15-20 mares. For example, just 10 of the 80 founding stallions of the current thoroughbred population have contributed >50% of genes to the entire population (Bailey 1998), with the three main founding stallions being Byerley Turk, the Darley Arabian and the Godolphin Arabian (Cunningham et al. 2001). Moreover, after 175 years of increasing paternal lineages, 95% of males in the current thoroughbred population lead to a single stallion (Darley Arabian) (Cunningham et al. 2001). In essence, the number of males contributing to the subsequent generations is considerably low, thereby drastically reducing the number of variant Y chromosomes in the population. This is obviously expected to lead to low variability on the Y, as has been observed by us after analyzing a sizable region of euchromatic region in the horse. Interestingly, the low variability evidenced through the analysis of the Y chromosome is in sharp contrast to the high levels of diversity reported earlier in the domestic horse population, based on analysis of mitochondrial loci (Vila et al. 2001). This is obviously

expected because the female breeding population contributing to the next generation in horses far exceeds the male, thus reflecting mitochondrial diversity.

It has been proposed that the use of male horses for meat consumption and females for reproduction (in some European countries) would have also resulted in a strong sex bias leading to reduced number of patrilines (Lindgren et al. 2004). Since the Y chromosome is hemizygous and restricted to only males, the effective population size of the Y chromosome is 1/3 compared to X chromosome and 1/4 compared to autosomes (under random mating conditions). Moreover, as indicated above, a strong sex bias towards females (more females, fewer males) in mating further reduces the effective population size of the horse Y chromosome, thus limiting the number of Y chromosomes circulating in the population.

Since the euchromatic region of the Y chromosome does not recombine with the X chromosome, it essentially acts as a single haplotype unit on which evolutionary forces act in a uniform manner. As such, selection would be an important force that affects the diversity of the Y chromosome. For example in positive selection, if a new mutation confers an advantage on a previously neutral marker (Rice 1987); due to lack of recombination, this selection may affect all the markers on that chromosome and result in a selective sweep throughout the population. Background selection (Charlesworth et al. 1995) would cause those chromosomes to be removed from the population that contained deleterious mutations (Charlesworth and Charlesworth 1998). Hence less genetic variability is expected on the Y chromosome as compared to

autosomes and X chromosome. In horse, along with the sex bias in mating, this phenomenon will result in even lower genetic variation on the Y chromosome.

Irrespective of the exceptionally low variability observed on the Y chromosome in the horse, the degree of diversity observed on the mammalian Y chromosome, in general, is significantly low compared to the rest of the genome, as is evident through observations hitherto made in humans, apes (Hammer 1995), lynx, wolf, reindeer, cattle, field vole (Hellborg and Ellegren 2004), Drosophila (McAllister and Charlesworth 1999) and plants (Filatov et al. 2000). The results of this study add to the emerging view that the genetic variability of the sex limited chromosome is inherently low indicating a possible common mechanism affecting all species (Montell et al. 2001). Positive and background selection has been identified as an important factor responsible for reduced diversity in the Y and W chromosome (Hellborg and Ellegren 2004). However in the future, more studies need to be carried out to elucidate the mechanisms affecting genetic variability on the sex limited chromosome.

Variability in Y-specific markers across equids

Out of the 27 microsatellite markers isolated in this study, 24 markers were successfully analyzed on most of the equids, demonstrating sequence conservation (Table 5). For three of the markers, PCR amplification could not be obtained even under less stringent conditions. Lack of amplification for these markers can be attributed to insufficient sequence conservation across the species. Of the 24 markers showing amplification in all species analyzed, three were monomorphic (MS13, MS23 and MS25) and four were dimorphic (MS15, MS20, MS21 and MS29). The remaining

markers showed a range of alleles varying from 3 to 7. The highest degree of polymorphism was demonstrated by two of the markers, viz., MS19 and MS27. Excluding the monomorphic markers, we observed an average number of four alleles per marker. This study increased the number of variable Y microsatellite markers across equids from three (Wallner et al. 2004) to 24. These polymorphic markers will be extremely useful for constructing informative haplotypes for each of the equid species and for conducting a detailed phylogenetic analysis –something that has not been carried out earlier due to lack of polymorphic Y-specific markers.

The markers analyzed in this study can be classified into two categories: those having a distinct allele from each marker across the nine species and those having multiple alleles for a marker in some of the species. While the former is understandable as being the allele present for a particular marker in a species (occurring once on the Y chromosome), the amplification of different size bands for the same marker in a single individual/species is puzzling. The only explanation to these strange amplifications can be associated to the unusual structure/organization of the Y chromosome. As reiterated throughout the text elsewhere, two phenomena are noticed on the Y chromosome: first, a range of sequences can be repeated intermittently across the chromosome; second, they can also be repeated tandemly across the chromosome. It could very well be that in instances where more than one peaks were observed for a marker in a species, the primers for the marker amplified more than one site, and that each site had a different size fragment reflecting polymorphism within the chromosome. An alternative explanation for the presence of more than one allele in a single individual is the

likelihood of the presence of the Y-sequence also on the X chromosome. This possibility cannot be ruled out completely because: of the 24 markers used for analysis, 19 were amplifying a male specific fragment while the remaining five showed an amplification of the same size also in the female controls in the horse.

While fossil records do help to analyze the divergence of different species within the equid family, controversy still remains on the degree of relationships/closeness between them. Phylogenies derived by morphometric analysis of cranial and dental characters of the different equids are not in agreement with each other (George and Ryder 1986). Earlier studies to phylogenetically define the various species of equids using mitochondrial DNA variations have been unsuccessful in clarifying the relationships between all equids (George and Ryder 1986; Ishida et al. 1995). Variation in 21 Y-specific markers provides a valuable resource for phylogenetic analyses of equids. We compared the allele sizes of these markers across the different equids to determine common haplotypes among the species that would give an indication about the genetic relationship of the nine equids.

Of the 24 markers used for analysis on different species, only nine showed a single allele/peak in each of the species studied. The other markers showed multiple alleles/peaks predominantly in domestic and Przewalski horse (e.g., MS3 marker), while in some cases also in other equids (e.g., MS1, MS2 MS8, MS11, MS18 & MS30). Thus the markers with a single allele in all species will be, for convenience in discussion, referred to as ‘single copy’ markers.

Analysis of Y chromosomal markers in Equus hartmannae, Equus grevyi, Equus burchelli antiquorumi, Equus burchelli boehmi

Examination of the levels of variation among the four zebra species for the 9 single copy Y chromosomal markers (MS14, MS15, MS16, MS17, MS19, MS20, MS21, MS27 & MS31) showed that only one marker (MS15) shared the same allele size in all the four zebra species. Further, *Equus grevyi* emerged out as the species with least number of shared/common alleles across the four species. Among the 'single copy' markers, it shared only one additional allele with *Equus zebra hartmannae* (MS16) and none with *Equus burchelli antiquorum & Equus burchelli boehmi*. A similar trend (of increased allele sharing of *Equus grevyi* with *Equus zebra hartmannae* as compared to other zebra species) was visible when comparison was extended to the rest of the markers that show more than one allele/peak in some of the species. This leads us to speculate that as far as Y chromosomal diversity is concerned, *Equus grevyi* is probably the most diverged among the four zebra species.

With regards to the remaining three species, a clear consensus does not emerge with the limited data because *Equus burchelli antiquorumi & Equus burchelli boehmi* share four common alleles for MS16, MS20, MS21, MS27 markers; *Equus burchelli antiquorumi* and *Equus zebra hartmannae* share four common alleles for MS14, MS17, MS20, MS21 markers and *Equus zebra hartmannae* and *Equus burchelli boehmi* share three common alleles for MS14, MS20, MS21 markers. Though data on more animals is needed within each of the species to draw valid conclusions, the preliminary observations of this study are similar to those reported earlier using mitochondrial DNA

analysis, where three zebra species *Equus zebra hartmannae, Equus grevyi, and Equus burchelli antiquorumi* formed a monophyletic group (George and Ryder 1986).

Analysis of Y chromosomal markers in Equus hemionus onager, Equus kiang *and* Equus asinus

Haplotypes of the nine 'single copy' markers in *Equus hemionus onager, Equus kiang* and *Equus asinus* showed that six markers (MS14, MS15, MS16, MS17, MS20, MS21) shared a similar allele in *Equus hemionus onager and Equus kiang*. A similar trend was observed for the 'multi-copy' markers, with 7 of the 11 markers showing common alleles. This suggests that, in general, the two species may be more closely related with each other for the Y-haplotypes, than with the other equids. Next, comparison of Y-specific alleles for the 21 markers in *Equus asinus* with the corresponding alleles in rest of the equids shows that *Equus asinus* shared 7 common alleles with *Equus hemionus onager* (MS1, MS3, MS18, MS14, MS20, MS21 & MS29) and 5 with and *Equus kiang* (MS3, MS14, MS20, MS21, MS29). Among zebras, it shares most alleles with *Equus zebra hartmannae* (6; MS2, MS3, MS14, MS20, MS21 & MS29) than with *Equus burchelli antiquorumi* (4; MS3, MS20, MS21 & MS29), *Equus grevyi* (4; MS1, MS3, MS27& MS29) & *Equus burchelli boehmi* (4; MS3, MS14, MS20 & MS21). Interestingly, when comparison was extended to domestic and Przewalski's horses, donkey showed a common allele only for one (MS29) of the 21 markers tested. Whether this implies that donkey Y is more closely related to other equids than to the two horse species, needs verification with more number of individuals. However our preliminary observations are in agreement with previous mitochondrial DNA analyses

that reports a closer relationship between donkey and zebra than to horse (George and Ryder 1986; Xu et al. 1996).

Analysis of Y chromosomal markers in Equus caballus *and* Equus przewalskii

The Przewalski's horse became extinct from the wild in the 1960s. Presently a captive population is being maintained that has mainly descended from 13 individuals. The phylogenetic relationship between *Equus caballus* and *Equus przewalskii* has always been controversial. Karyotypically *Equus przewalskii* (2n = 66) can be distinguished from the domestic horse (2n = 64) due to a Robertosonian translocation of two pairs of acrocentric chromosomes (EPR23 & 24) of *Equus przewalskii* that resulted in a pair of submetacentric chromosomes in the domestic horse, i.e., ECA 5 (Benirschke et al. 1965; Myka et al. 2003). The presence of the two extra chromosomes in *Equus przewalskii* has led some to believe that it was ancestral to the domestic horse (Benirschke et al. 1965). Genetic studies using blood markers also distinguished between the two (Bowling and Ryder 1987). However, this is contradicted by analysis of mitochondrial D-loop, that suggests that of the four extant female lineages in Przewalski's, the variation between and within Przewalski and domestic horse is not different (Oakenfull and Ryder 1998).

In the present study, comparison of *Equus caballus* and *Equus przewalskii* alleles demonstrated the presence of only one 'single-copy' marker (MS27) that showed distinct alleles in the two species. For all other markers, the two species share at least one common allele and are therefore predominantly similar. Next, for five of the markers (MS1, MS3, MS4, MS8 & MS11), both species show multiple alleles (ranging from 2 –

4 alleles). Lastly, three of the markers show more alleles in the horse, but only one in *Equus przewalskii* (MS2, MS18 & MS29); a reverse situation is seen only for marker (MS28) where Przewalskii has two alleles and the domestic horse has only one. In brief, *Equus caballus* and *Equus przewalskii* share considerable similarity with each other than with the rest of the equids. Nonetheless, some of the markers can be readily used to differentiate between the two, in particular the 'single-copy' marker MS27. More results based on larger sample size will, however, be needed to firstly assess the frequency of the segregating alleles in the two species and secondly to establish the alleles that clearly distinguish them from each other.

Two previous studies have identified eight SNPs on the Y chromosomes of *Equus caballus* and *Equus przewalskii (Wallner et al. 2003; Lindgren et al. 2004)*. The results suggested that the Przewalski's haplotype was distinct from that of the domestic horse Y, but the two species were still considered to belong to sister taxa. Phylogenetic analysis using mitochondrial DNA variation clustered two different maternal lineages of Przewalski horse with the domestic horse (George and Ryder 1986; Ishida et al. 1995; Vila et al. 2001), indicating that *E przewalskii* is not a evolutionary distinct from the dometic horse. However, which of the two species is indeed ancestral, still needs to be resolved.

The basic information obtained over here for the 24 markers is indeed limited. However, the primary objective in this study was first to obtain Y-specific markers in the horse and then to test the likelihood of their use in other equids. This was essential to verify whether or not the domestic horse markers are also polymorphic in other eight

equids, and if they could be used across them for Y-related phylogenetic analysis. To this extent our experiments have been successful because, except for 3 markers, the remaining 21 markers demonstrated polymorphism across different equids. Hence, these markers serve as a starting point for an organized extended analysis where more unrelated representatives from each of the species will be genotyped in order to accurately develop phylogenetic relationships between them.

Overview

In conclusion, the present study provides the first extended evaluation for genetic variability on the Y chromosome across nine species of equids. Approximately 4.5 Mb of euchromatic region was screened for polymorphic microsatellite markers. One polymorphic marker with 3 alleles has been identified for the first time on the horse Y chromosome. Absence of polymorphisms in the remaining 26 markers supports earlier notion that the level of genetic diversity in the Y chromosome is inherently low (Shen et al. 2000; Montell et al. 2001; Hellborg and Ellegren 2004). A strong sex bias in mating is perhaps the major factor attributed to the low variability seen in the domestic horse Y chromosome.

Next, across the nine extant equids, as yet only three polymorphic microsatellite markers were reported (Wallner et al. 2004). The identification of 21 variable markers in this study offers a valuable resource for initiating phylogenetic studies in equids, and to elucidate relationships between them. The identification of fewer allelic differences between *Equus caballus* and *Equus przewalskii* as compared to other equids clearly shows that the two species are genetically closely related and may represent sister taxa.

Finally, the construction of subclone libraries will serve a useful resource for generating new STS markers, conducting SNP analyses, and for obtaining complete sequence of the entire euchromatic region of the horse Y chromosome in the future.

CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK

Presently, medium resolution genetic maps are available for all equine autosomes. For some of the autosomes e.g., ECA17 and 22, and the X chromosome, 1marker/Mbp resolution maps are available. These maps are already facilitating analysis of coat color traits like Appalousa, Grey etc., and disease traits like severe combined immunodeficiency (Bernoco and Bailey 1998) and Epitheliogenesis imperfecta (Lieto and Cothran 2003) to identify genes and the variations that contribute to their expression or manifestation. Despite these developments, very little is currently being done to analyze complex traits – those governed by several genes. One of these traits is fertility. Genes governing fertility in males and females are present on autosomes and the X chromosome in the females and, additionally on the Y chromosome in the males. Unfortunately almost nothing is known about the structure and genetic organization of the horse Y chromosome. Consequently, the knowledge about the Y-specific genes and their function in regulating male fertility is sparse. This thesis was therefore conceived to lay the foundation of that knowledge base by creating the most basic but essential tool: a detailed gene map of the horse Y chromosome. By no means developing a map must imply finding immediate answers to all male fertility related questions. It should only be looked upon as one of the first crucial steps without which questions of practical relevance would never be answered. This thesis is thus a foundation to a complete sequence map of the horse Y chromosome, which in turn will lead to studies aimed at

isolating genes that contribute to overall manifestation of male fertility and understanding their role and function in the future.

The first study represents the first systematic effort to generate a detailed map of the horse Y chromosome. Due to the unusual organization and gene content of the horse Y chromosome (as in most mammalian Y chromosomes), mapping this chromosome posed several challenges; a host of repetitive sequences and multicopy genes to name just a few. To resolve these problems, unlike autosomes, a number of mapping techniques were used to develop the ECAY map. The first integrated radiation hybrid map and FISH map laid the groundwork of the study, on the basis of which seven BAC contigs were constructed across ~4.5 Mb of the euchromatic region of the Y chromosome. In order to validate and cross-verify our results, STS content mapping, restriction fingerprinting, and Fiber FISH were conducted. The analyses led to the first detailed physical map of the horse Y chromosome. This map will serve as a foundation to develop a complete high resolution map that will be critical for obtaining a complete sequence of the chromosome and for initiating comprehensive functional studies in the future to identify Y-specific genes associated with spermatogenic failure and stallion subfertility.

The second study is a spin-off from the results of the first study, and is aimed at i) examining the level of genetic variation specific to the Y chromosome across different breeds of domestic horse by generating and analyzing microsatellite markers and ii) exploring the possibility of the use of these markers in the extant 9 equids. Subclone libraries were constructed for BACs from the minimum tiling path of the the 7 contigs

generated in the first study. These BACs cover ~4.5Mb of the euchromatic region of the horse Y chromosome. Libraries were screened with di, tri, tetra, and pentanucleotide repeat motifs and in total 32 new microsatellite loci were identified and mapped to the Y chromosome. Of these, 27 could be used to assess the extent of variation in these markers across 14 diverse breeds of domestic horse and eight extant equids. Only one microsatellite marker was identified to be polymorphic in different breeds. This amazingly low level of genetic variation in the horse Y chromosome is attributed to the small effective population size of males that have contributed to past and current generations. Contrary to these findings, the use of the same microsatellites in eight other equids, led to the detection of polymorphism on all but one locus. Though the findings are preliminary and were aimed to test the hypothesis that there must be variation on the Y chromosome across the equids, they will form the corner-stone for a detailed analysis with a larger and representative population size that will be crucial for adding a new dimension (the Y chromosome specific) to our current understanding of the evolution among equids. Finally, the subclone libraries generated in this part of the work will significantly contribute towards generation of new markers essential for developing a high resolution map of ECAY.

REFERENCES

Abe, S., Miyake, Y.I., Kageyama, S.I., Watanabe, G., Taya, K., and Kawakura, K. 1999. Deletion of the Sry region on the Y chromosome detected in a case of equine gonadal hypoplasia (XY female) with abnormal hormonal profiles. *Equine Vet J* **31:** 336-338.

Achmann, R., Curik, I., Dovc, P., Kavar, T., Bodo, I., Habe, F., Marti, E., Solkner, J., and Brem, G. 2004. Microsatellite diversity, population subdivision and gene flow in the Lipizzan horse. *Anim Genet* **35:** 285-292.

Agulnik, A.I., Bishop, C.E., Lerner, J.L., Agulnik, S.I., and Solovyev, V.V. 1997. Analysis of mutation rates in the SMCY/SMCX genes shows that mammalian evolution is male driven. *Mamm Genome* **8:** 134-138.

Alves, B.C., Hossepian de Lima, V.F., Teixeira, C.M., and Moreira-Filho, C.A. 2003. Use of primers derived from a new sequence of the bovine Y chromosome for sexing *Bos taurus* and *Bos indicus* embryos. *Theriogenology* **59:** 1415-1419.

Appa Rao, K.B. and Totey, S.M. 1999. Cloning and sequencing of buffalo male-specific repetitive DNA: sexing of *in vitro* developed buffalo embryos using multiplex and nested polymerase chain reaction. *Theriogenology* **51:** 785-797.

Arredi, B., Poloni, E.S., Paracchini, S., Zerjal, T., Fathallah, D.M., Makrelouf, M., Pascali, V.L., Novelletto, A., and Tyler-Smith, C. 2004. A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. *Am J Hum Genet* **75:** 338-345.

Ashwell, M.S., Heyen, D.W., Sonstegard, T.S., Van Tassell, C.P., Da, Y., VanRaden, P.M., Ron, M., Weller, J.I., and Lewin, H.A. 2004. Detection of quantitative trait loci affecting milk production, health, and reproductive traits in Holstein cattle. *J Dairy Sci* **87:** 468-475.

Babicz, M., Kuryl, J., and Walkiewicz, A. 2003. Evaluation of the genetic profile of the Pulawska breed. *J Appl Genet* **44:** 497-508.

Bailey, E. 1998. Odds on the FAST gene. *Genome Res* **8:** 569-571.

Basu, A., Mukherjee, N., Roy, S., Sengupta, S., Banerjee, S., Chakraborty, M., Dey, B., Roy, M., Roy, B., Bhattacharyya, N.P., Roychoudhury, S., and Majumder, P.P. 2003. Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Res* **13:** 2277-2290.

Benirschke, K., Malouf, N., Low, R.J., and Heck, H. 1965. Chromosome complement: differences between *Equus caballus* and *Equus przewalskii*, Poliakoff. *Science* **148:** 382-383.

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

Bhattacharyya, N.P., Basu, P., Das, M., Pramanik, S., Banerjee, R., Roy, B., Roychoudhury, S., and Majumder, P.P. 1999. Negligible male gene flow across ethnic boundaries in India, revealed by analysis of Y-chromosomal DNA polymorphisms. *Genome Res* **9:** 711-719.

Bianchi, N.O., Catanesi, C.I., Bailliet, G., Martinez-Marignac, V.L., Bravi, C.M., Vidal-Rioja, L.B., Herrera, R.J., and Lopez-Camelo, J.S. 1998. Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. *Am J Hum Genet* **63:** 1862-1871.

Bishop, C. and Mitchell, M. 1997. Mouse Y chromosome. *Mamm Genome* **7:** S327-329.

Bjornstad, G., Nilsen, N.O., and Roed, K.H. 2003. Genetic relationship between Mongolian and Norwegian horses? *Anim Genet* **34:** 55-58.

Bonne-Tamir, B., Korostishevsky, M., Redd, A.J., Pel-Or, Y., Kaplan, M.E., and Hammer, M.F. 2003. Maternal and paternal lineages of the Samaritan isolate: mutation rates and time to most recent common male ancestor. *Ann Hum Genet* **67:** 153-164.

Bosch, E., Calafell, F., Rosser, Z.H., Norby, S., Lynnerup, N., Hurles, M.E., and Jobling, M.A. 2003. High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. *Hum Genet* **112:** 353-363.

Bowling, A.T. and Ryder, O.A. 1987. Genetic studies of blood markers in Przewalski's horses. *J Hered* **78:** 75-80.

Bradley, D.G., MacHugh, D.E., Cunningham, P., and Loftus, R.T. 1996. Mitochondrial diversity and the origins of African and European cattle. *Proc Natl Acad Sci U S A* **93:** 5131-5135.

Braun, J., Hazekamp, M.G., Koolbergen, D.R., Sugihara, H., Goffin, Y.A., Huysmans, H.A., and Cornelisse, C.J. 1997. Identification of host and donor cells in porcine homograft heart valve explants by fluorescence in situ hybridization. *J Pathol* **183:** 99-104.

Breen, M., Downs, P., Irvin, Z., and Bell, K. 1994. Intrageneric amplification of horse microsatellite markers with emphasis on the Przewalski's horse (*E. przewalskii*). *Anim Genet* **25:** 401-405.

Bruford, M.W., Bradley, D.G., and Luikart, G. 2003. DNA markers reveal the complexity of livestock domestication. *Nat Rev Genet* **4:** 900-910.

Bugno, M., Klukowska, J., Slota, E., Tischner, M., and Switonski, M. 2003. A sporadic case of the sex-reversed mare (64,XY; SRY-negative): molecular and cytogenetic studies of the Y chromosome. *Theriogenology* **59:** 1597-1603.

Burgoyne, P.S. 1996. Fruit(less) flies provide a clue. *Nature* **381:** 740-741.

Burgoyne, P.S., Mahadevaiah, S.K., Sutcliffe, M.J., and Palmer, S.J. 1992. Fertility in mice requires X-Y pairing and a Y-chromosomal "spermiogenesis" gene mapping to the long arm. *Cell* **71:** 391-398.

Butler, J.M. and Schoske, R. 2004. Duplication of DYS19 flanking regions in other parts of the Y chromosome. *Int J Legal Med* **118:** 178-183.

Carvajal-Carmona, L.G., Soto, I.D., Pineda, N., Ortiz-Barrientos, D., Duque, C., Ospina-Duque, J., McCarthy, M., Montoya, P., Alvarez, V.M., Bedoya, G., and Ruiz-Linares, A. 2000. Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am J Hum Genet* **67:** 1287-1295.

Castellanos, C., Barragan, C., and Rodriguez, M.C. 1996. Detection of four porcine Y-specific markers by RAPD. *Anim Genet* **27:** 433-434.

Chang, Y.L., Tao, Q., Scheuring, C., Ding, K., Meksem, K., and Zhang, H.B. 2001. An integrated map of *Arabidopsis thaliana* for functional analysis of its genome sequence. *Genetics* **159:** 1231-1242.

Charlesworth, B. 1991. The evolution of sex chromosomes. *Science* **251:** 1030-1033.

Charlesworth, B. 2003. The organization and evolution of the human Y chromosome. *Genome Biol* **4:** 226.

Charlesworth, D. and Charlesworth, B. 1998. Sequence variation: looking for effects of genetic linkage. *Curr Biol* **8:** R658-661.

Charlesworth, D., Charlesworth, B., and Morgan, M.T. 1995. The pattern of neutral molecular variation under the background selection model. *Genetics* **141:** 1619-1632.

Chen, L.S., Tassone, F., Sahota, P., and Hagerman, P.J. 2003. The (CGG)n repeat element within the 5' untranslated region of the FMR1 message provides both positive and negative cis effects on *in vivo* translation of a downstream reporter. *Hum Mol Genet* **12:** 3067-3074.

Chikhi, L., Nichols, R.A., Barbujani, G., and Beaumont, M.A. 2002. Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci U S A* **99:** 11008-11013.

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

Chowdhary, B.P., Raudsepp, T., Honeycutt, D., Owens, E.K., Piumi, F., Guerin, G., Matise, T.C., Kata, S.R., Womack, J.E., and Skow, L.C. 2002. Construction of a 5000(rad) whole-genome radiation hybrid panel in the horse and generation of a comprehensive and comparative map for ECA11. *Mamm Genome* **13:** 89-94.

Chowdhary, B.P., Raudsepp, T., Kata, S.R., Goh, G., Millon, L.V., Allan, V., Piumi, F., Guerin, G., Swinburne, J., Binns, M., Lear, T.L., Mickelson, J., Murray, J., Antczak, D.F., Womack, J.E., and 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:** 742-751.

Ciampolini, R., Moazami-Goudarzi, K., Vaiman, D., Dillmann, C., Mazzanti, E., Foulley, J.L., Leveziel, H., and Cianci, D. 1995. Individual multilocus genotypes using microsatellite polymorphisms to permit the analysis of the genetic variability within and between Italian beef cattle breeds. *J Anim Sci* **73:** 3259-3268.

Colenbrander, B., Gadella, B.M., and Stout, T.A. 2003. The predictive value of semen analysis in the evaluation of stallion fertility. *Reprod Domest Anim* **38:** 305-311.

Conway, S.J., Mahadevaiah, S.K., Darling, S.M., Capel, B., Rattigan, A.M., and Burgoyne, P.S. 1994. Y353/B: a candidate multiple-copy spermiogenesis gene on the mouse Y chromosome. *Mamm Genome* **5:** 203-210.

Cordaux, R., Aunger, R., Bentley, G., Nasidze, I., Sirajuddin, S.M., and Stoneking, M. 2004a. Independent origins of Indian caste and tribal paternal lineages. *Curr Biol* **14:** 231-235.

Cordaux, R., Weiss, G., Saha, N., and Stoneking, M. 2004b. The northeast Indian passageway: a barrier or corridor for human migrations? *Mol Biol Evol* **21:** 1525-1533.

Cortopassi, G.A. 2002. Fixation of deleterious alleles, evolution and human aging. *Mech Ageing Dev* **123:** 851-855.

Cunningham, E.P., Dooley, J.J., Splan, R.K., and Bradley, D.G. 2001. Microsatellite diversity, pedigree relatedness and the contributions of founder lineages to thoroughbred horses. *Anim Genet* **32:** 360-364.

Dada, R., Gupta, N.P., and Kucheria, K. 2003. Molecular screening for Yq microdeletion in men with idiopathic oligozoospermia and azoospermia. *J Biosci* **28:** 163-168.

Di Berardino, D., Vozdova, M., Kubickova, S., Cernohorska, H., Coppola, G., Enne, G., and Rubes, J. 2004. Sexing river buffalo (*Bubalus bubalis* L.), sheep (*Ovis aries* L.), goat (*Capra hircus* L.), and cattle spermatozoa by double color FISH using bovine (*Bos taurus* L.) X- and Y-painting probes. *Mol Reprod Dev* **67:** 108-115.

Disteche, C.M., Brannan, C.I., Larsen, A., Adler, D.A., Schorderet, D.F., Gearing, D., Copeland, N.G., Jenkins, N.A., and Park, L.S. 1992. The human pseudoautosomal GM-CSF receptor alpha subunit gene is autosomal in mouse. *Nat Genet* **1:** 333-336.

Dowsett, K.F. and Pattie, W.A. 1982. Characteristics and fertility of stallion semen. *J Reprod Fertil Suppl* **32:** 1-8.

Eberhart, C.G., Maines, J.Z., and Wasserman, S.A. 1996. Meiotic cell cycle requirement for a fly homologue of human Deleted in Azoospermia. *Nature* **381:** 783-785.

Edwards, C.J., Gaillard, C., Bradley, D.G., and MacHugh, D.E. 2000. Y-specific microsatellite polymorphisms in a range of bovid species. *Anim Genet* **31:** 127-130.

Ellegren, H. 2004. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* **5:** 435-445.

Elliott, D.J., Ma, K., Kerr, S.M., Thakrar, R., Speed, R., Chandley, A.C., and Cooke, H. 1996. An RBM homologue maps to the mouse Y chromosome and is expressed in germ cells. *Hum Mol Genet* **5:** 869-874.

Elliott, D.J., Millar, M.R., Oghene, K., Ross, A., Kiesewetter, F., Pryor, J., McIntyre, M., Hargreave, T.B., Saunders, P.T., Vogt, P.H., Chandley, A.C., and Cooke, H. 1997. Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. *Proc Natl Acad Sci U S A* **94:** 3848-3853.

Ellison, J.W., Li, X., Francke, U., and Shapiro, L.J. 1996. Rapid evolution of human pseudoautosomal genes and their mouse homologs. *Mamm Genome* **7:** 25-30.

Ewing, B. and Green, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8:** 186-194.

Ferlin, A., Moro, E., Rossi, A., Dallapiccola, B., and Foresta, C. 2003. The human Y chromosome's azoospermia factor b (AZFb) region: sequence, structure, and deletion analysis in infertile men. *J Med Genet* **40:** 18-24.

Fiegler, H., Knabel, M., Franz, M., Kolb, H.J., and Just, U. 2002. Determination of donor-type chimerism using a semi-quantitative PCR-based method in a canine model for bone marrow transplantation. *Vet Immunol Immunopathol* **84:** 61-70.

Filatov, D.A., Moneger, F., Negrutiu, I., and Charlesworth, D. 2000. Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* **404:** 388-390.

Foote, S., Vollrath, D., Hilton, A., and Page, D.C. 1992. The human Y chromosome: overlapping DNA clones spanning the euchromatic region. *Science* **258:** 60-66.

Foresta, C., Ferlin, A., and Moro, E. 2000. Deletion and expression analysis of AZFa genes on the human Y chromosome revealed a major role for DBY in male infertility. *Hum Mol Genet* **9:** 1161-1169.

Foresta, C., Moro, E., and Ferlin, A. 2001. Y chromosome microdeletions and alterations of spermatogenesis. *Endocr Rev* **22:** 226-239.

Foster, E.A., Jobling, M.A., Taylor, P.G., Donnelly, P., de Knijff, P., Mieremet, R., Zerjal, T., and Tyler-Smith, C. 1998. Jefferson fathered slave's last child. *Nature* **396:** 27-28.

Fox, M.S. and Reijo Pera, R.A. 2001. Male infertility, genetic analysis of the DAZ genes on the human Y chromosome and genetic analysis of DNA repair. *Mol Cell Endocrinol* **184:** 41-49.

Gebhardt, F., Burger, H., and Brandt, B. 2000. Modulation of EGFR gene transcription by secondary structures, a polymorphic repetitive sequence and mutations--a link between genetics and epigenetics. *Histol Histopathol* **15:** 929-936.

Gebhardt, F., Zanker, K.S., and Brandt, B. 1999. Modulation of epidermal growth factor receptor gene transcription by a polymorphic dinucleotide repeat in intron 1. *J Biol Chem* **274:** 13176-13180.

Gendrel, C.G., Boulet, A., and Dutreix, M. 2000. (CA/GT)(n) microsatellites affect homologous recombination during yeast meiosis. *Genes Dev* **14:** 1261-1268.

George, M., Jr. and Ryder, O.A. 1986. Mitochondrial DNA evolution in the genus Equus. *Mol Biol Evol* **3:** 535-546.

Giacalone, J., Delobette, S., Gibaja, V., Ni, L., Skiadas, Y., Qi, R., Edington, J., Lai, Z., Gebauer, D., Zhao, H., Anantharaman, T., Mishra, B., Brown, L.G., Saxena, R., Page, D.C., and Schwartz, D.C. 2000. Optical mapping of BAC clones from the human Y chromosome DAZ locus. *Genome Res* **10:** 1421-1429.

Gianfrancesco, F., Sanges, R., Esposito, T., Tempesta, S., Rao, E., Rappold, G., Archidiacono, N., Graves, J.A., Forabosco, A., and D'Urso, M. 2001. Differential divergence of three human pseudoautosomal genes and their mouse homologs: implications for sex chromosome evolution. *Genome Res* **11:** 2095-2100.

Gibson, C.W., Collier, P.M., Yuan, Z.A., and Chen, E. 1998. DNA sequences of amelogenin genes provide clues to regulation of expression. *Eur J Oral Sci* **106** Suppl 1**:** 292-298.

Giuffra, E., Kijas, J.M., Amarger, V., Carlborg, O., Jeon, J.T., and Andersson, L. 2000. The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics* **154:** 1785-1791.

Graham, J.K. 1996. Analysis of stallion semen and its relation to fertility. *Vet Clin North Am Equine Pract* **12:** 119-130.

Graves, J.A. 1995. The origin and function of the mammalian Y chromosome and Y-borne genes--an evolving understanding. *Bioessays* **17:** 311-320.

Graves, J.A. 1998. Evolution of the mammalian Y chromosome and sex-determining genes. *J Exp Zool* **281:** 472-481.

Graves, J.A. 2001. From brain determination to testis determination: evolution of the mammalian sex-determining gene. *Reprod Fertil Dev* **13:** 665-672.

Gresham, D., Morar, B., Underhill, P.A., Passarino, G., Lin, A.A., Wise, C., Angelicheva, D., Calafell, F., Oefner, P.J., Shen, P., Tournev, I., de Pablo, R., Kucinskas, V., Perez-Lezaun, A., Marushiakova, E., Popov, V., and Kalaydjieva, L. 2001. Origins and divergence of the Roma (gypsies). *Am J Hum Genet* **69:** 1314-1331.

Gusmao, L., Brion, M., Gonzalez-Neira, A., Lareu, M., and Carracedo, A. 1999. Y chromosome specific polymorphisms in forensic analysis. *Leg Med (Tokyo)* **1:** 55-60.

Guyon, R., Lorentzen, T.D., Hitte, C., Kim, L., Cadieu, E., Parker, H.G., Quignon, P., Lowe, J.K., Renier, C., Gelfenbeyn, B., Vignaux, F., DeFrance, H.B., Gloux, S., Mahairas, G.G., Andre, C., Galibert, F., and Ostrander, E.A. 2003. A 1-Mb resolution radiation hybrid map of the canine genome. *Proc Natl Acad Sci U S A* **100:** 5296-5301.

Hammer, M.F. 1995. A recent common ancestry for human Y chromosomes. *Nature* **378:** 376-378.

Hammer, M.F., Karafet, T.M., Redd, A.J., Jarjanazi, H., Santachiara-Benerecetti, S., Soodyall, H., and Zegura, S.L. 2001. Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol* **18:** 1189-1203.

Hammer, M.F., Redd, A.J., Wood, E.T., Bonner, M.R., Jarjanazi, H., Karafet, T., Santachiara-Benerecetti, S., Oppenheim, A., Jobling, M.A., Jenkins, T., Ostrer, H., and Bonne-Tamir, B. 2000. Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Proc Natl Acad Sci U S A* **97:** 6769-6774.

Handt, O., Meyer, S., and von Haeseler, A. 1998. Compilation of human mtDNA control region sequences. *Nucleic Acids Res* **26:** 126-129.

Hanotte, O., Tawah, C.L., Bradley, D.G., Okomo, M., Verjee, Y., Ochieng, J., and Rege, J.E. 2000. Geographic distribution and frequency of a taurine *Bos taurus* and an

indicine *Bos indicus* Y specific allele amongst sub-saharan African cattle breeds. *Mol Ecol* **9:** 387-396.

Hasegawa, T., Sato, F., Ishida, N., Fukushima, Y., and Mukoyama, H. 2000. Sex determination by simultaneous amplification of equine SRY and amelogenin genes. *J Vet Med Sci* **62:** 1109-1110.

Hasegawa, T., Ishida, M., Harigaya, T., Sato, F., Ishida, N., and Mukoyama, H. 1999. Linear SRY transcript in equine testis. *J Vet Med Sci* **61:** 97-100.

Heiskanen, M., Karhu, R., Hellsten, E., Peltonen, L., Kallioniemi, O.P., and Palotie, A. 1994. High resolution mapping using fluorescence *in situ* hybridization to extended DNA fibers prepared from agarose-embedded cells. *Biotechniques* **17:** 928-929, 932-923.

Hellborg, L. and Ellegren, H. 2004. Low levels of nucleotide diversity in mammalian Y chromosomes. *Mol Biol Evol* **21:** 158-163.

Hirota, K., Piumi, F., Sato, F., Ishida, N., Guerin, G., Miura, N., and Hasegawa, T. 2001. FISH assignment of two equine BAC clones containing SRY and ZFY. *Anim Genet* **32:** 326-327.

Hook, E.B. and Warburton, D. 1983. The distribution of chromosomal genotypes associated with Turner's syndrome: livebirth prevalence rates and evidence for diminished fetal mortality and severity in genotypes associated with structural X abnormalities or mosaicism. *Hum Genet* **64:** 24-27.

Hunter, S., Gramlich, T., Abbott, K., and Varma, V. 1993. Y chromosome loss in esophageal carcinoma: an *in situ* hybridization study. *Genes Chromosomes Cancer* **8:** 172-177.

Hurles, M.E., Irven, C., Nicholson, J., Taylor, P.G., Santos, F.R., Loughlin, J., Jobling, M.A., and Sykes, B.C. 1998. European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. *Am J Hum Genet* **63:** 1793-1806.

Hurles, M.E. and Jobling, M.A. 2001. Haploid chromosomes in molecular ecology: lessons from the human Y. *Mol Ecol* **10:** 1599-1613.

Hurtgen, J.P. 1992. Evaluation of the stallion for breeding soundness. *Vet Clin North Am Equine Pract* **8:** 149-165.

Iglesias, A.R., Kindlund, E., Tammi, M., and Wadelius, C. 2004. Some microsatellites may act as novel polymorphic cis-regulatory elements through transcription factor binding. *Gene* **341:** 149-165.

Irion, D.N., Schaffer, A.L., Famula, T.R., Eggleston, M.L., Hughes, S.S., and Pedersen, N.C. 2003. Analysis of genetic variation in 28 dog breed populations with 100 microsatellite markers. *J Hered* **94:** 81-87.

Ishida, N., Oyunsuren, T., Mashima, S., Mukoyama, H., and Saitou, N. 1995. Mitochondrial DNA sequences of various species of the genus *Equus* with special reference to the phylogenetic relationship between Przewalskii's wild horse and domestic horse. *J Mol Evol* **41:** 180-188.

Iwasa, M., Koyama, H., Tsuchimochi, T., Maeno, Y., Isobe, I., Seko-Nakamura, Y., Monma-Ohtaki, J., Matsumoto, T., and Nagao, M. 2003. Y-chromosomal short tandem repeats haplotyping from vaginal swabs using a chelating resin-based DNA extraction method and a dual-round polymerase chain reaction. *Am J Forensic Med Pathol* **24:** 303-305.

Iwase, M., Satta, Y., Hirai, Y., Hirai, H., Imai, H., and Takahata, N. 2003. The amelogenin loci span an ancient pseudoautosomal boundary in diverse mammalian species. *Proc Natl Acad Sci U S A* **100:** 5258-5263.

Jansen, T., Forster, P., Levine, M.A., Oelke, H., Hurles, M., Renfrew, C., Weber, J., and Olek, K. 2002. Mitochondrial DNA and the origins of the domestic horse. *Proc Natl Acad Sci U S A* **99:** 10905-10910.

Jobling, M.A., Pandya, A., and Tyler-Smith, C. 1997. The Y chromosome in forensic analysis and paternity testing. *Int J Legal Med* **110:** 118-124.

Jobling, M.A. and Tyler-Smith, C. 1995. Fathers and sons: the Y chromosome and human evolution. *Trends Genet* **11:** 449-456.

Jobling, M.A. and Tyler-Smith, C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* **4:** 598-612.

Johnson, C.L., Warren, J.H., Giles, R.C., and Staub, R.W. 2003. Validation and uses of a Y-chromosome STR 10-plex for forensic and paternity laboratories. *J Forensic Sci* **48:** 1260-1268.

Karafet, T.M., Osipova, L.P., Gubina, M.A., Posukh, O.L., Zegura, S.L., and Hammer, M.F. 2002. High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Hum Biol* **74:** 761-789.

Kawarasaki, T., Matsumoto, K., Chikyu, M., Itagaki, Y., and Horiuchi, A. 2000. Sexing of porcine embryo by in situ hybridization using chromosome Y- and 1-specific DNA probes. *Theriogenology* **53:** 1501-1509.

Kawarasaki, T., Welch, G.R., Long, C.R., Yoshida, M., and Johnson, L.A. 1998. Verification of flow cytometorically-sorted X- and Y-bearing porcine spermatozoa and reanalysis of spermatozoa for DNA content using the fluorescence *in situ* hybridization (FISH) technique. *Theriogenology* **50:** 625-635.

Kayser, M., Kittler, R., Erler, A., Hedman, M., Lee, A.C., Mohyuddin, A., Mehdi, S.Q., Rosser, Z., Stoneking, M., Jobling, M.A., Sajantila, A., and Tyler-Smith, C. 2004. A comprehensive survey of human Y-chromosomal microsatellites. *Am J Hum Genet* **74:** 1183-1197.

Kelly, L., Postiglioni, A., De Andres, D.F., Vega-Pla, J.L., Gagliardi, R., Biagetti, R., and Franco, J. 2002. Genetic characterisation of the Uruguayan Creole horse and analysis of relationships among horse breeds. *Res Vet Sci* **72:** 69-73.

Kijas, J.M. and Andersson, L. 2001. A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. *J Mol Evol* **52:** 302-308.

Kim, K.S., Tanabe, Y., Park, C.K., and Ha, J.H. 2001. Genetic variability in East Asian dogs using microsatellite loci analysis. *J Hered* **92:** 398-403.

Kittles, R.A., Bergen, A.W., Urbanek, M., Virkkunen, M., Linnoila, M., Goldman, D., and Long, J.C. 1999. Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: evidence for a male-specific bottleneck. *Am J Phys Anthropol* **108:** 381-399.

Koehler, U., Bigoni, F., Wienberg, J., and Stanyon, R. 1995. Genomic reorganization in the concolor gibbon (*Hylobates concolor*) revealed by chromosome painting. *Genomics* **30:** 287-292.

Koskinen, M.T. 2003. Individual assignment using microsatellite DNA reveals unambiguous breed identification in the domestic dog. *Anim Genet* **34:** 297-301.

Krausz, C., Forti, G., and McElreavey, K. 2003. The Y chromosome and male fertility and infertility. *Int J Androl* **26:** 70-75.

Krausz, C. and McElreavey, K. 1999. Y chromosome and male infertility. *Front Biosci* **4:** E1-8.

Kujawski, M., Jarmuz, M., Rydzanicz, M., Szukala, K., Wierzbicka, M., Grenman, R., Golusinski, W., and Szyfter, K. 2004. Frequent chromosome Y loss in primary, second primary and metastatic squamous cell carcinomas of the head and neck region. *Cancer Lett* **208:** 95-101.

Kumar, P., Freeman, A.R., Loftus, R.T., Gaillard, C., Fuller, D.Q., and Bradley, D.G. 2003. Admixture analysis of South Asian cattle. *Heredity* **91:** 43-50.

Kuroda-Kawaguchi, T., Skaletsky, H., Brown, L.G., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Silber, S., Oates, R., Rozen, S., and Page, D.C. 2001. The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat Genet* **29:** 279-286.

Kuroki, Y., Iwamoto, T., Lee, J., Yoshiike, M., Nozawa, S., Nishida, T., Ewis, A.A., Nakamura, H., Toda, T., Tokunaga, K., Kotliarova, S.E., Kondoh, N., Koh, E., Namiki, M., Shinka, T., and Nakahori, Y. 1999. Spermatogenic ability is

different among males in different Y chromosome lineage. *J Hum Genet* **44:** 289-292.

Lau, Y., Chou, P., Iezzoni, J., Alonzo, J., and Komuves, L. 2000. Expression of a candidate gene for the gonadoblastoma locus in gonadoblastoma and testicular seminoma. *Cytogenet Cell Genet* **91:** 160-164.

Lau, Y.F., Lau, H.W., and Komuves, L.G. 2003. Expression pattern of a gonadoblastoma candidate gene suggests a role of the Y chromosome in prostate cancer. *Cytogenet Genome Res* **101:** 250-260.

Laval, S.H., Glenister, P.H., Rasberry, C., Thornton, C.E., Mahadevaiah, S.K., Cooke, H.J., Burgoyne, P.S., and Cattanach, B.M. 1995. Y chromosome short arm-Sxr recombination in XSxr/Y males causes deletion of Rbm and XY female sex reversal. *Proc Natl Acad Sci U S A* **92:** 10403-10407.

Lell, J.T., Sukernik, R.I., Starikovskaya, Y.B., Su, B., Jin, L., Schurr, T.G., Underhill, P.A., and Wallace, D.C. 2002. The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet* **70:** 192-206.

Lemus-Flores, C., Ulloa-Arvizu, R., Ramos-Kuri, M., Estrada, F.J., and Alonso, R.A. 2001. Genetic analysis of Mexican hairless pig populations. *J Anim Sci* **79:** 3021-3026.

Li, J.H., Huang, T.H., Jiang, X.W., and Xie, Q.D. 2004a. 46, XX male sex reversal syndrome. *Asian J Androl* **6:** 165-167.

Li, S.J., Yang, S.H., Zhao, S.H., Fan, B., Yu, M., Wang, H.S., Li, M.H., Liu, B., Xiong, T.A., and Li, K. 2004b. Genetic diversity analyses of 10 indigenous Chinese pig populations based on 20 microsatellites. *J Anim Sci* **82:** 368-374.

Liefers, S.C., te Pas, M.F., Veerkamp, R.F., and van der Lende, T. 2002. Associations between leptin gene polymorphisms and production, live weight, energy balance, feed intake, and fertility in Holstein heifers. *J Dairy Sci* **85:** 1633-1638.

Lieto, L.D. and Cothran, E.G. 2003. The epitheliogenesis imperfecta locus maps to equine chromosome 8 in American Saddlebred horses. *Cytogenet Genome Res* **102:** 207-210.

Lindgren, G., Backstrom, N., Swinburne, J., Hellborg, L., Einarsson, A., Sandberg, K., Cothran, G., Vila, C., Binns, M., and Ellegren, H. 2004. Limited number of patrilines in horse domestication. *Nat Genet* **36:** 335-336.

Liu, W.S., Beattie, C.W., and Ponce de Leon, F.A. 2003. Bovine Y chromosome microsatellite polymorphisms. *Cytogenet Genome Res* **102:** 53-58.

Liu, W.S., Mariani, P., Beattie, C.W., Alexander, L.J., and Ponce De Leon, F.A. 2002. A radiation hybrid map for the bovine Y chromosome. *Mamm Genome* **13:** 320-326.

Liu, W.S., Soldatov, N.M., Gustavsson, I., and Chowdhary, B.P. 1998. Fiber-FISH analysis of the 3'-terminal region of the human L-type Ca2+ channel alpha 1C subunit gene. *Hereditas* **129:** 169-175.

Loftus, R.T., MacHugh, D.E., Bradley, D.G., Sharp, P.M., and Cunningham, P. 1994. Evidence for two independent domestications of cattle. *Proc Natl Acad Sci U S A* **91:** 2757-2761.

Love, C.C. and Kenney, R.M. 1998. The relationship of increased susceptibility of sperm DNA to denaturation and fertility in the stallion. *Theriogenology* **50:** 955-972.

Love, C.C., Thompson, J.A., Brinsko, S.P., Rigby, S.L., Blanchard, T.L., Lowry, V.K., and Varner, D.D. 2003. Relationship between stallion sperm motility and viability as detected by two fluorescence staining techniques using flow cytometry. *Theriogenology* **60:** 1127-1138.

Luikart, G., Gielly, L., Excoffier, L., Vigne, J.D., Bouvet, J., and Taberlet, P. 2001. Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proc Natl Acad Sci U S A* **98:** 5927-5932.

MacHugh, D.E., Loftus, R.T., Cunningham, P., and Bradley, D.G. 1998. Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. *Anim Genet* **29:** 333-340.

MacHugh, D.E., Shriver, M.D., Loftus, R.T., Cunningham, P., and Bradley, D.G. 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics* **146:** 1071-1086.

Madill, S. 2002. Reproductive considerations: mare and stallion. *Vet Clin North Am Equine Pract* **18:** 591-619.

Mahadevaiah, S.K., Odorisio, T., Elliott, D.J., Rattigan, A., Szot, M., Laval, S.H., Washburn, L.L., McCarrey, J.R., Cattanach, B.M., Lovell-Badge, R., and Burgoyne, P.S. 1998. Mouse homologues of the human AZF candidate gene RBM are expressed in spermatogonia and spermatids, and map to a Y chromosome deletion interval associated with a high incidence of sperm abnormalities. *Hum Mol Genet* **7:** 715-727.

Manna, L., Neglia, G., Marino, M., Gasparrini, B., Di Palo, R., and Zicarelli, L. 2003. Sex determination of buffalo embryos (*Bubalus bubalis*) by polymerase chain reaction. *Zygote* **11:** 17-22.

Manni, F., Toupance, B., Sabbagh, A., and Heyer, E. 2004. New method for surname studies of ancient patrilineal population structures, and possible application to improvement of Y-chromosome sampling. *Am J Phys Anthropol* (available: http://www3.interscience.wiley.com/cgi-bin/jhome/28130).

Mateus, J.C., Eding, H., Penedo, M.C., and Rangel-Figueiredo, M.T. 2004a. Contributions of Portuguese cattle breeds to genetic diversity using marker-estimated kinships. *Anim Genet* **35:** 305-313.

Mateus, J.C., Penedo, M.C., Alves, V.C., Ramos, M., and Rangel-Figueiredo, T. 2004b. Genetic diversity and differentiation in Portuguese cattle breeds using microsatellites. *Anim Genet* **35:** 106-113.

Mattapallil, M.J. and Ali, S. 1999. Analysis of conserved microsatellite sequences suggests closer relationship between water buffalo *Bubalus bubalis* and sheep *Ovis aries*. *DNA Cell Biol* **18:** 513-519.

Mazeyrat, S., Saut, N., Sargent, C.A., Grimmond, S., Longepied, G., Ehrmann, I.E., Ellis, P.S., Greenfield, A., Affara, N.A., and Mitchell, M.J. 1998. The mouse Y chromosome interval necessary for spermatogonial proliferation is gene dense with syntenic homology to the human AZFa region. *Hum Mol Genet* **7:** 1713-1724.

McAllister, B.F. and Charlesworth, B. 1999. Reduced sequence variability on the Neo-Y chromosome of *Drosophila americana*. *Genetics* **153:** 221-233.

McDonnell, S.M. 2000. Reproductive behavior of stallions and mares: comparison of free-running and domestic in-hand breeding. *Anim Reprod Sci* **60-61:** 211-219.

Meadows, J.R., Hawken, R.J., and Kijas, J.W. 2004. Nucleotide diversity on the ovine Y chromosome. *Anim Genet* **35:** 379-385.

Menke, D.B., Mutter, G.L., and Page, D.C. 1997. Expression of DAZ, an azoospermia factor candidate, in human spermatogonia. *Am J Hum Genet* **60:** 237-241.

Meyers-Wallen, V.N., Bowman, L., Acland, G.M., Palmer, V.L., Schlafer, D., and Fajt, V. 1995. Sry-negative XX sex reversal in the German shorthaired pointer dog. *J Hered* **86:** 369-374.

Meyers-Wallen, V.N., Schlafer, D., Barr, I., Lovell-Badge, R., and Keyzner, A. 1999. Sry-negative XX sex reversal in purebred dogs. *Mol Reprod Dev* **53:** 266-273.

Milenkovic, D., Oustry-Vaiman, A., Lear, T.L., Billault, A., Mariat, D., Piumi, F., Schibler, L., Cribiu, E., and Guerin, G. 2002. Cytogenetic localization of 136 genes in the horse: comparative mapping with the human genome. *Mamm Genome* **13:** 524-534.

Moioli, B., Napolitano, F., and Catillo, G. 2004. Genetic diversity between Piedmontese, Maremmana, and Podolica cattle breeds. *J Hered* **95:** 250-256.

Montell, H., Fridolfsson, A.K., and Ellegren, H. 2001. Contrasting levels of nucleotide diversity on the avian Z and W sex chromosomes. *Mol Biol Evol* **18:** 2010-2016.

Morris, L.H. and Allen, W.R. 2002. Reproductive efficiency of intensively managed Thoroughbred mares in Newmarket. *Equine Vet J* **34:** 51-60.

Mukherjee, N., Nebel, A., Oppenheim, A., and Majumder, P.P. 2001. High-resolution analysis of Y-chromosomal polymorphisms reveals signatures of population movements from Central Asia and West Asia into India. *J Genet* **80:** 125-135.

Muller, S. and Wienberg, J. 2001. "Bar-coding" primate chromosomes: molecular cytogenetic screening for the ancestral hominoid karyotype. *Hum Genet* **109:** 85-94.

Murphy, W.J., Sun, S., Chen, Z.Q., Pecon-Slattery, J., and O'Brien, S.J. 1999. Extensive conservation of sex chromosome organization between cat and human revealed by parallel radiation hybrid mapping. *Genome Res* **9:** 1223-1230.

Myka, J.L., Lear, T.L., Houck, M.L., Ryder, O.A., and Bailey, E. 2003. FISH analysis comparing genome organization in the domestic horse (*Equus caballus*) to that of the Mongolian wild horse (*E. przewalskii*). *Cytogenet Genome Res* **102:** 222-225.

Nagafuchi, S., Tamura, T., Nakahori, Y., Takano, K., Nishi, Y., Iwatani, N., Kitao, M., Hori, Y., Konda, S., Hasegawa, T., and et al. 1992. The majority of the marker chromosomes in Japanese patients with stigmata of Turner syndrome are derived from Y chromosomes. *Hum Genet* **89:** 590-592.

Nevanlinna, H.R. 1972. The Finnish population structure. A genetic and genealogical study. *Hereditas* **71:** 195-236.

Oakenfull, E.A. and Ryder, O.A. 1998. Mitochondrial control region and 12S rRNA variation in Przewalski's horse (*Equus przewalski*i). *Anim Genet* **29:** 456-459.

Oetting, W.S., Lee, H.K., Flanders, D.J., Wiesner, G.L., Sellers, T.A., and King, R.A. 1995. Linkage analysis with multiplexed short tandem repeat polymorphisms using infrared fluorescence and M13 tailed primers. *Genomics* **30:** 450-458.

Olivier, M., Breen, M., Binns, M.M., and Lust, G. 1999. Localization and characterization of nucleotide sequences from the canine Y chromosome. *Chromosome Res* **7:** 223-233.

Olivier, M. and Lust, G. 1998. Two DNA sequences specific for the canine Y chromosome. *Anim Genet* **29:** 146-149.

Page, D.C. 1987. Hypothesis: a Y-chromosomal gene causes gonadoblastoma in dysgenetic gonads. *Development* **101 Suppl:** 151-155.

Page, D.C., Mosher, R., Simpson, E.M., Fisher, E.M., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A., and Brown, L.G. 1987. The sex-determining region of the human Y chromosome encodes a finger protein. *Cell* **51:** 1091-1104.

Pailhoux, E., Cribiu, E.P., Chaffaux, S., Darre, R., Fellous, M., and Cotinot, C. 1994. Molecular analysis of 60,XX pseudohermaphrodite polled goats for the presence of SRY and ZFY genes. *J Reprod Fertil* **100:** 491-496.

Pailhoux, E., Cribiu, E.P., Parma, P., and Cotinot, C. 1995. Molecular analysis of an XY mare with gonadal dysgenesis. *Hereditas* **122:** 109-112.

Parker, H.G., Kim, L.V., Sutter, N.B., Carlson, S., Lorentzen, T.D., Malek, T.B., Johnson, G.S., DeFrance, H.B., Ostrander, E.A., and Kruglyak, L. 2004. Genetic structure of the purebred domestic dog. *Science* **304:** 1160-1164.

Paszek, A.A., Flickinger, G.H., Fontanesi, L., Beattie, C.W., Rohrer, G.A., Alexander, L., and Schook, L.B. 1998a. Evaluating evolutionary divergence with microsatellites. *J Mol Evol* **46:** 121-126.

Paszek, A.A., Flickinger, G.H., Fontanesi, L., Rohrer, G.A., Alexander, L., Beattie, C.W., and Schook, L.B. 1998b. Livestock variation of linked microsatellite markers in diverse swine breeds. *Anim Biotechnol* **9:** 55-66.

Pecon Slattery, J. and O'Brien, S.J. 1998. Patterns of Y and X chromosome DNA sequence divergence during the Felidae radiation. *Genetics* **148:** 1245-1255.

Pecon Slattery, J., Sanner-Wachter, L., and O'Brien, S.J. 2000. Novel gene conversion between X-Y homologues located in the nonrecombining region of the Y chromosome in Felidae (Mammalia). *Proc Natl Acad Sci U S A* **97:** 5307-5312.

Perry, J., Palmer, S., Gabriel, A., and Ashworth, A. 2001. A short pseudoautosomal region in laboratory mice. *Genome Res* **11:** 1826-1832.

Poloumienko, A. 2004. Cloning and comparative analysis of the bovine, porcine, and equine sex chromosome genes ZFX and ZFY. *Genome* **47:** 74-83.

Pomp, D., Good, B.A., Geisert, R.D., Corbin, C.J., and Conley, A.J. 1995. Sex identification in mammals with polymerase chain reaction and its use to examine sex effects on diameter of day-10 or -11 pig embryos. *J Anim Sci* **73:** 1408-1415.

Power, M.M. 1988. Y chromosome length variation and its significance in the horse. *J Hered* **79:** 311-313.

Power, M.M., Gustavsson, I., Switonski, M., and Ploen, L. 1992. Synaptonemal complex analysis of an autosomal trisomy in the horse. *Cytogenet Cell Genet* **61:** 202-207.

Quilter, C.R., Blott, S.C., Mileham, A.J., Affara, N.A., Sargent, C.A., and Griffin, D.K. 2002. A mapping and evolutionary study of porcine sex chromosome genes. *Mamm Genome* **13:** 588-594.

Raicu, F., Popa, L., Apostol, P., Cimponeriu, D., Dan, L., Ilinca, E., Dracea, L.L., Marinescu, B., and Gavrila, L. 2003. Screening for microdeletions in human Y chromosome--AZF candidate genes and male infertility. *J Cell Mol Med* **7:** 43-48.

Rao, K.B., Pawshe, C.H., and Totey, S.M. 1993. Sex determination of *in vitro* developed buffalo (*Bubalus bubalis*) embryos by DNA amplification. *Mol Reprod Dev* **36:** 291-296.

Raudsepp, T. and 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:** 103-114.

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

Raudsepp, T., Santani, A., Wallner, B., Kata, S.R., Ren, C., Zhang, H.B., Womack, J.E., Skow, L.C., and Chowdhary, B.P. 2004b. A detailed physical map of the horse Y chromosome. *Proc Natl Acad Sci U S A* **101:** 9321-9326.

Redd, A.J., Roberts-Thomson, J., Karafet, T., Bamshad, M., Jorde, L.B., Naidu, J.M., Walsh, B., and Hammer, M.F. 2002. Gene flow from the Indian subcontinent to Australia: evidence from the Y chromosome. *Curr Biol* **12:** 673-677.

Reijo, R., Lee, T.Y., Salo, P., Alagappan, R., Brown, L.G., Rosenberg, M., Rozen, S., Jaffe, T., Straus, D., Hovatta, O., et al. 1995. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nat Genet* **10:** 383-393.

Ren, C., Lee, M.K., Yan, B., Ding, K., Cox, B., Romanov, M.N., Price, J.A., Dodgson, J.B., and Zhang, H.B. 2003. A BAC-based physical map of the chicken genome. *Genome Res* **13:** 2754-2758.

Repping, S., Skaletsky, H., Lange, J., Silber, S., Van Der Veen, F., Oates, R.D., Page, D.C., and Rozen, S. 2002. Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. *Am J Hum Genet* **71:** 906-922.

Rice, W.R. 1987. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics* **116:** 161-167.

Richard, F., Lombard, M., and Dutrillaux, B. 1996. ZOO-FISH suggests a complete homology between human and capuchin monkey (Platyrrhini) euchromatin. *Genomics* **36:** 417-423.

Rose, O. and Falush, D. 1998. A threshold size for microsatellite expansion. *Mol Biol Evol* **15:** 613-615.

Rothschild, M.F., Messer, L.A., and Vincent, A. 1997. Molecular approaches to improved pig fertility. *J Reprod Fertil Suppl* **52:** 227-236.

Rottger, S., Yen, P.H., and Schempp, W. 2002. A fiber-FISH contig spanning the non-recombining region of the human Y chromosome. *Chromosome Res* **10:** 621-635.

Rousset, H., Chanteloube, P., Magistrini, M., and Palmer, E. 1987. Assessment of fertility and semen evaluations of stallions. *J Reprod Fertil Suppl* **35:** 25-31.

Rubes, J., Vozdova, M., and Kubickova, S. 1999. Aneuploidy in pig sperm: multicolor fluorescence *in situ* hybridization using probes for chromosomes 1, 10, and Y. *Cytogenet Cell Genet* **85:** 200-204.

Saha, A., Udhayasuriyan, P.T., Bhat, K.V., and Bamezai, R. 2003. Analysis of Indian population based on Y-STRs reveals existence of male gene flow across different language groups. *DNA Cell Biol* **22:** 707-719.

Sanchez, J.J., Brion, M., Parson, W., Blanco-Verea, A.J., Borsting, C., Lareu, M., Niederstatter, H., Oberacher, H., Morling, N., and Carracedo, A. 2004. Duplications of the Y-chromosome specific loci P25 and 92R7 and forensic implications. *Forensic Sci Int* **140:** 241-250.

Santos, F.R., Gerelsaikhan, T., Munkhtuja, B., Oyunsuren, T., Epplen, J.T., and Pena, S.D. 1996. Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism DYS19. *Hum Genet* **97:** 309-313.

Sargent, C.A., Boucher, C.A., Kirsch, S., Brown, G., Weiss, B., Trundley, A., Burgoyne, P., Saut, N., Durand, C., Levy, N., Terriou, P., Hargreave, T., Cooke, H., Mitchell, M., Rappold, G.A., and Affara, N.A. 1999. The critical region of overlap defining the AZFa male infertility interval of proximal Yq contains three transcribed sequences. *J Med Genet* **36:** 670-677.

Savolainen, P., Zhang, Y.P., Luo, J., Lundeberg, J., and Leitner, T. 2002. Genetic evidence for an East Asian origin of domestic dogs. *Science* **298:** 1610-1613.

Saxena, R., Brown, L.G., Hawkins, T., Alagappan, R.K., Skaletsky, H., Reeve, M.P., Reijo, R., Rozen, S., Dinulos, M.B., Disteche, C.M., and Page, D.C. 1996. The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nat Genet* **14:** 292-299.

Schlotterer, C. and Tautz, D. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* **20:** 211-215.

Schoske, R., Vallone, P.M., Kline, M.C., Redman, J.W., and Butler, J.M. 2004. High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays. *Forensic Sci Int* **139:** 107-121.

Schurr, T.G. and Sherry, S.T. 2004. Mitochondrial DNA and Y chromosome diversity and the peopling of the Americas: evolutionary and demographic evidence. *Am J Hum Biol* **16:** 420-439.

Seielstad, M. 2000. Asymmetries in the maternal and paternal genetic histories of Colombian populations. *Am J Hum Genet* **67:** 1062-1066.

Seielstad, M.T., Minch, E., and Cavalli-Sforza, L.L. 1998. Genetic evidence for a higher female migration rate in humans. *Nat Genet* **20:** 278-280.

Selden, J.R., Wachtel, S.S., Koo, G.C., Haskins, M.E., and Patterson, D.F. 1978. Genetic basis of XX male syndrome and XX true hermaphroditism: evidence in the dog. *Science* **201:** 644-646.

Semino, O., Magri, C., Benuzzi, G., Lin, A.A., Al-Zahery, N., Battaglia, V., Maccioni, L., Triantaphyllidis, C., Shen, P., Oefner, P.J., Zhivotovsky, L.A., King, R., Torroni, A., Cavalli-Sforza, L.L., Underhill, P.A., and Santachiara-Benerecetti, A.S. 2004. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* **74:** 1023-1034.

Semino, O., Passarino, G., Quintana-Murci, L., Liu, A., Beres, J., Czeizel, A., and Santachiara-Benerecetti, A.S. 2000. MtDNA and Y chromosome polymorphisms in Hungary: inferences from the palaeolithic, neolithic and Uralic influences on the modern Hungarian gene pool. *Eur J Hum Genet* **8:** 339-346.

Senese, C., Penedo, M.C., Shiue, Y.L., Bowling, A.T., and Millon, L.V. 1999. A HaeIII PCR-RFLP in the ZFY/ZFX genes of horses. *Anim Genet* **30:** 390-391.

Shahid, M., Dhillion, V.S., Jain, N., Hedau, S., Diwakar, S., Sachdeva, P., Batra, S., Das, B.C., and Husain, S.A. 2004. Two new novel point mutations localized upstream and downstream of the HMG box region of the SRY gene in three Indian 46,XY females with sex reversal and gonadal tumour formation. *Mol Hum Reprod* **10:** 521-526.

Shen, P., Lavi, T., Kivisild, T., Chou, V., Sengun, D., Gefel, D., Shpirer, I., Woolf, E., Hillel, J., Feldman, M.W., and Oefner, P.J. 2004. Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. *Hum Mutat* **24:** 248-260.

Shen, P., Wang, F., Underhill, P.A., Franco, C., Yang, W.H., Roxas, A., Sung, R., Lin, A.A., Hyman, R.W., Vollrath, D., Davis, R.W., Cavalli-Sforza, L.L., and Oefner, P.J. 2000. Population genetic implications from sequence variation in four Y chromosome genes. *Proc Natl Acad Sci U S A* **97:** 7354-7359.

Shiue, Y.L., Bickel, L.A., Caetano, A.R., Millon, L.V., Clark, R.S., Eggleston, M.L., Michelmore, R., Bailey, E., Guerin, G., Godard, S., Mickelson, J.R., Valberg, S.J., Murray, J.D., and Bowling, A.T. 1999. A synteny map of the horse genome comprised of 240 microsatellite and RAPD markers. *Anim Genet* **30:** 1-9.

Shiue, Y.L., Millon, L.V., Skow, L.C., Honeycutt, D., Murray, J.D., and Bowling, A.T. 2000. Synteny and regional marker order assignment of 26 type I and microsatellite markers to the horse X- and Y-chromosomes. *Chromosome Res* **8:** 45-55.

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell-Badge, R., and Goodfellow, P.N. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346:** 240-244.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S.F., Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T.,

Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfing, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S.P., Waterston, R.H., Wilson, R.K., Rozen, S., and Page, D.C. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423:** 825-837.

Slattery, J.P., Murphy, W.J., and O'Brien, S.J. 2000. Patterns of diversity among SINE elements isolated from three Y-chromosome genes in carnivores. *Mol Biol Evol* **17:** 825-829.

Slee, R., Grimes, B., Speed, R.M., Taggart, M., Maguire, S.M., Ross, A., McGill, N.I., Saunders, P.T., and Cooke, H.J. 1999. A human DAZ transgene confers partial rescue of the mouse Dazl null phenotype. *Proc Natl Acad Sci U S A* **96:** 8040-8045.

Soderlund, C., Humphray, S., Dunham, A., and French, L. 2000. Contigs built with fingerprints, markers, and FPC V4.7. *Genome Res* **10:** 1772-1787.

Solinas-Toldo, S., Lengauer, C., and Fries, R. 1995. Comparative genome map of human and cattle. *Genomics* **27:** 489-496.

Stone, J.F. and Sandberg, A.A. 1995. Sex chromosome aneuploidy and aging. *Mutat Res* **338:** 107-113.

Stumpf, M.P. and Goldstein, D.B. 2001. Genealogical and evolutionary inference with the human Y chromosome. *Science* **291:** 1738-1742.

Styrna, J., Imai, H.T., and Moriwaki, K. 1991a. An increased level of sperm abnormalities in mice with a partial deletion of the Y chromosome. *Genet Res* **57:** 195-199.

Styrna, J., Klag, J., and Moriwaki, K. 1991b. Influence of partial deletion of the Y chromosome on mouse sperm phenotype. *J Reprod Fertil* **92:** 187-195.

Sulston, J., Mallett, F., Durbin, R., and Horsnell, T. 1989. Image analysis of restriction enzyme fingerprint autoradiograms. *Comput Appl Biosci* **5:** 101-106.

Sun, C., Skaletsky, H., Birren, B., Devon, K., Tang, Z., Silber, S., Oates, R., and Page, D.C. 1999. An azoospermic man with a de novo point mutation in the Y-chromosomal gene USP9Y. *Nat Genet* **23:** 429-432.

Sundqvist, A.K., Ellegren, H., Olivier, M., and Vila, C. 2001. Y chromosome haplotyping in Scandinavian wolves (*Canis lupus*) based on microsatellite markers. *Mol Ecol* **10:** 1959-1966.

Sutcliffe, M.J. and Burgoyne, P.S. 1989. Analysis of the testes of H-Y negative XOSxrb mice suggests that the spermatogenesis gene (Spy) acts during the differentiation of the A spermatogonia. *Development* **107:** 373-380.

Sykes, B. and Irven, C. 2000. Surnames and the Y chromosome. *Am J Hum Genet* **66:** 1417-1419.

Thangaraj, K., Reddy, A.G., and Singh, L. 2004. Mutation in the STR locus D21S1 1 of father causing allele mismatch in the child. *J Forensic Sci* **49:** 99-103.

Thomson, R., Pritchard, J.K., Shen, P., Oefner, P.J., and Feldman, M.W. 2000. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc Natl Acad Sci U S A* **97:** 7360-7365.

Tiepolo, L. and Zuffardi, O. 1976. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Hum Genet* **34:** 119-124.

Tilford, C.A., Kuroda-Kawaguchi, T., Skaletsky, H., Rozen, S., Brown, L.G., Rosenberg, M., McPherson, J.D., Wylie, K., Sekhon, M., Kucaba, T.A., Waterston, R.H., and Page, D.C. 2001. A physical map of the human Y chromosome. *Nature* **409:** 943-945.

Toder, R., Glaser, B., Schiebel, K., Wilcox, S.A., Rappold, G., Graves, J.A., and Schempp, W. 1997. Genes located in and near the human pseudoautosomal region are located in the X-Y pairing region in dog and sheep. *Chromosome Res* **5:** 301-306.

Toure, A., Grigoriev, V., Mahadevaiah, S.K., Rattigan, A., Ojarikre, O.A., and Burgoyne, P.S. 2004. A protein encoded by a member of the multicopy Ssty gene family located on the long arm of the mouse Y chromosome is expressed during sperm development. *Genomics* **83:** 140-147.

Tozaki, T., Takezaki, N., Hasegawa, T., Ishida, N., Kurosawa, M., Tomita, M., Saitou, N., and Mukoyama, H. 2003. Microsatellite variation in Japanese and Asian horses and their phylogenetic relationship using a European horse outgroup. *J Hered* **94:** 374-380.

Trask, B.J., Allen, S., Massa, H., Fertitta, A., Sachs, R., van den Engh, G., and Wu, M. 1993. Studies of metaphase and interphase chromosomes using fluorescence in situ hybridization. *Cold Spring Harb Symp Quant Biol* **58:** 767-775.

Troy, C.S., MacHugh, D.E., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., Chamberlain, A.T., Sykes, B.C., and Bradley, D.G. 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410:** 1088-1091.

Tsuei, D.J., Hsu, H.C., Lee, P.H., Jeng, Y.M., Pu, Y.S., Chen, C.N., Lee, Y.C., Chou, W.C., Chang, C.J., Ni, Y.H., and Chang, M.H. 2004. RBMY, a male germ cell-specific RNA-binding protein, activated in human liver cancers and transforms rodent fibroblasts. *Oncogene* **23:** 5815-5822.

Vallejo, R.L., Li, Y.L., Rogers, G.W., and Ashwell, M.S. 2003. Genetic diversity and background linkage disequilibrium in the North American Holstein cattle population. *J Dairy Sci* **86:** 4137-4147.

van Buiten, A., Remmen, J.L., and Colenbrander, B. 1998. Fertility of Shetland pony stallions used in different breeding systems: a retrospective study. *Vet Q* **20:** 100-103.

Van Hooft, W.F., Groen, A.F., and Prins, H.H. 2002. Phylogeography of the African buffalo based on mitochondrial and Y-chromosomal loci: Pleistocene origin and population expansion of the Cape buffalo subspecies. *Mol Ecol* **11:** 267-279.

Varner, D.D., Blanchard, T.L., Brinsko, S.P., Love, C.C., Taylor, T.S., and Johnson, L. 2000. Techniques for evaluating selected reproductive disorders of stallions. *Anim Reprod Sci* **60-61:** 493-509.

Vaughan, L., Schofield, W., and Ennis, S. 2001. SRY-negative XX sex reversal in a pony: a case report. *Theriogenology* **55:** 1051-1057.

Vergnaud, G., Page, D.C., Simmler, M.C., Brown, L., Rouyer, F., Noel, B., Botstein, D., de la Chapelle, A., and Weissenbach, J. 1986. A deletion map of the human Y chromosome based on DNA hybridization. *Am J Hum Genet* **38:** 109-124.

Vernesi, C., Crestanello, B., Pecchioli, E., Tartari, D., Caramelli, D., Hauffe, H., and Bertorelle, G. 2003. The genetic impact of demographic decline and reintroduction in the wild boar (*Sus scrofa*): a microsatellite analysis. *Mol Ecol* **12:** 585-595.

Vila, C., Leonard, J.A., Gotherstrom, A., Marklund, S., Sandberg, K., Liden, K., Wayne, R.K., and Ellegren, H. 2001. Widespread origins of domestic horse lineages. *Science* **291:** 474-477.

Vila, C., Savolainen, P., Maldonado, J.E., Amorim, I.R., Rice, J.E., Honeycutt, R.L., Crandall, K.A., Lundeberg, J., and Wayne, R.K. 1997. Multiple and ancient origins of the domestic dog. *Science* **276:** 1687-1689.

Vila, C., Walker, C., Sundqvist, A.K., Flagstad, O., Andersone, Z., Casulli, A., Kojola, I., Valdmann, H., Halverson, J., and Ellegren, H. 2003. Combined use of maternal, paternal and bi-parental genetic markers for the identification of wolf-dog hybrids. *Heredity* **90:** 17-24.

Vogt, P.H. 1997. Molecular basis of male (in)fertility. *Int J Androl* **20** Suppl **3:** 2-10.

Vogt, P.H., Edelmann, A., Kirsch, S., Henegariu, O., Hirschmann, P., Kiesewetter, F., Kohn, F.M., Schill, W.B., Farah, S., Ramos, C., Hartmann, M., Hartschuh, W., Meschede, D., Behre, H.M., Castel, A., Nieschlag, E., Weidner, W., Grone, H.J., Jung, A., Engel, W., and Haidl, G. 1996. Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum Mol Genet* **5:** 933-943.

Vollrath, D., Foote, S., Hilton, A., Brown, L.G., Beer-Romero, P., Bogan, J.S., and Page, D.C. 1992. The human Y chromosome: a 43-interval map based on naturally occurring deletions. *Science* **258:** 52-59.

Wallner, B., Brem, G., Muller, M., and Achmann, R. 2003. Fixed nucleotide differences on the Y chromosome indicate clear divergence between *Equus przewalskii* and *Equus caballus*. *Anim Genet* **34:** 453-456.

Wallner, B., Piumi, F., Brem, G., Muller, M., and Achmann, R. 2004. Isolation of Y chromosome-specific microsatellites in the horse and cross-species amplification in the genus *Equus*. *J Hered* **95:** 158-164.

Wang, X., Miller, A.B., Lepine, A.J., Scott, J.D., and Murphy, K.E. 1999. Analysis of randomly amplified polymorphic DNA (RAPD) for identifying genetic markers associated with canine hip dysplasia. *J Hered* **90:** 99-103.

Watanobe, T., Ishiguro, N., Nakano, M., Takamiya, H., Matsui, A., and Hongo, H. 2002. Prehistoric introduction of domestic pigs onto the Okinawa Islands: ancient mitochondrial DNA evidence. *J Mol Evol* **55:** 222-231.

Waters, P.D., Duffy, B., Frost, C.J., Delbridge, M.L., and Graves, J.A. 2001. The human Y chromosome derives largely from a single autosomal region added to the sex chromosomes 80-130 million years ago. *Cytogenet Cell Genet* **92:** 74-79.

Wen, B., Li, H., Lu, D., Song, X., Zhang, F., He, Y., Li, F., Gao, Y., Mao, X., Zhang, L., Qian, J., Tan, J., Jin, J., Huang, W., Deka, R., Su, B., Chakraborty, R., and Jin, L. 2004a. Genetic evidence supports demic diffusion of Han culture. *Nature* **431:** 302-305.

Wen, B., Xie, X., Gao, S., Li, H., Shi, H., Song, X., Qian, T., Xiao, C., Jin, J., Su, B., Lu, D., Chakraborty, R., and Jin, L. 2004b. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *Am J Hum Genet* **74:** 856-865.

Whitfield, L.S., Sulston, J.E., and Goodfellow, P.N. 1995. Sequence variation of the human Y chromosome. *Nature* **378:** 379-380.

Wienberg, J., Stanyon, R., Jauch, A., and Cremer, T. 1992. Homologies in human and *Macaca fuscata* chromosomes revealed by *in situ* suppression hybridization with human chromosome specific DNA libraries. *Chromosoma* **101:** 265-270.

Wilder, J.A., Kingan, S.B., Mobasher, Z., Pilkington, M.M., and Hammer, M.F. 2004. Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. *Nat Genet* **36:** 1122-1125.

Williams, J.L., Usha, A.P., Urquhart, B.G., and Kilroy, M. 1997. Verification of the identity of bovine semen using DNA microsatellite markers. *Vet Rec* **140:** 446-449.

Wolf, U., Schempp, W., and Scherer, G. 1992. Molecular biology of the human Y chromosome. *Rev Physiol Biochem Pharmacol* **121:** 147-213.

Xu, X., Gullberg, A., and Arnason, U. 1996. The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA comparisons among four closely related mammalian species-pairs. *J Mol Evol* **43:** 438-446.

Zenteno-Ruiz, J.C., Kofman-Alfaro, S., and Mendez, J.P. 2001. 46,XX sex reversal. *Arch Med Res* **32:** 559-566.

Zerjal, T., Dashnyam, B., Pandya, A., Kayser, M., Roewer, L., Santos, F.R., Schiefenhovel, W., Fretwell, N., Jobling, M.A., Harihara, S., Shimizu, K., Semjidmaa, D., Sajantila, A., Salo, P., Crawford, M.H., Ginter, E.K., Evgrafov, O.V., and Tyler-Smith, C. 1997. Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* **60:** 1174-1183.

Zerjal, T., Wells, R.S., Yuldasheva, N., Ruzibakiev, R., and Tyler-Smith, C. 2002. A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. *Am J Hum Genet* **71:** 466-482.

WEB SITE REFERENCES

http://www.horsecouncil.org/statistics.htm/; American Horse Council Home Page.
http://www.wiwfarm.com/pdf/ PAEquineSurveyResults-Academic.pdf/; Pennsylvania Equine Survery Results.

http://animalscience.tamu.edu/ansc/publications/horsepubs/ hrg014-industry.pdf/; Texas Equine Survery Results.

http://www-genome.wi.mit.edu/cgibin/primer/primer3_www.cgi/; Primer3 home page.
http://hbz7.tamu.edu/homelinks/bac_est/bac.htm/; Texas A&M University equine BAC library home page.

http://bacpac.chori.org/equine241.htm/; Children's Hospital Oakland Research Institute 241 Library home page.

http://dga.jouy.inra.fr/grafra/INRA_libraries_database_simplified.htm/; Institut National dela Recherche Agronomique-complemented BAC library (INRA) home page.

http://ratmap.gen.gu.se/SearchList.html/; Rat Map Database.

http://www.phrap.com/; PHRAP home page.

http://www.ncbi.nih.gov/BLAST/; NCBI BLAST home page.

VITA

Avni Bhawan Santani
Abhishek 203,
Andheri West,
Mumbai 400058
India

EDUCATION/TRAINING

University of Mumbai, India, B.S., Microbiology & Biochemistry, May 1998, **GPA 4.0**
University of Mumbai, India, Diploma., Medical Laboratory Technology, May 2000, **GPA 4.0**
University of Aberdeen, UK, M.S., Medical Molecular Genetics, August 2001, **GPA 3.7**
Texas A&M University, US, Ph.D., Genetics, September 2001-December 2004, **GPA 3.8**

PROFESSIONAL EXPERIENCE

1999-1999	Intern at the Jaslok Hospital, Departments of Blood Banking, Biochemistry and Histopathology, Mumbai, India; *Conducted clinical pathological analyses in the Intensive Care Unit.*
2001- 2001	Research Assistant at Aberdeen University for Dr. Al Cumming. *Research Topic: Isolation of polymorphic markers in the APO AI-CIII-AIV gene cluster and haplotype association study in relation to hypercholesterolemia and high lipid profiles*.
2002-2003	Teaching assistant in *Biomedical Genetics* Gene 320, Fall 2002, for Dr. Bhanu Chowdhary, VAPH, CVM, Texas A&M University.
2001-2004	Research Assistant at Texas A&M University, in Dr. Chowdhary's lab. Research Project: Developing a map of the horse Y chromosome to detect molecular causes of stallion infertility.

MANUSCRIPTS

Santani, A., Raudsepp, T., and Chowdhary, B.P. 2002. Interstitial telomeric sites and NORs in Hartmann's zebra (*Equus zebra hartmannae*) chromosomes. *Chromosome Research* **10:** 527-534.

Raudsepp, T., Santani, A., Wallner, B., Kata, S.R., Ren, C., Zhang, H.B., Womack, J.E., Skow, L.C., and Chowdhary, B.P. 2004. A detailed physical map of the horse Y chromosome. *Proc Natl Acad Sci U S A* **101:** 9321-9326.