DISSECTING GENOMIC FACTORS OF STALLION FERTILITY

A Dissertation

by

CAITLIN PATRICIA CASTANEDA

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Terje Raudsepp
Committee Members,	Brian Davis
	Dickson Varner
	E. Gus Cothran
Head of Department,	Todd O'Hara

August 2022

Major Subject: Biomedical Sciences

Copyright 2022 Caitlin P. Castaneda

ABSTRACT

Stallion fertility is a complex trait, affected by a considerable genetic component. Knowledge about the genetic factors contributing to reduced fertility in stallions is currently limited. The goal of this Dissertation was to investigate select autosomal and Y-linked factors of stallion fertility to enhance knowledge about the genetic component of stallion fertility.

For the select autosomal factors, we validated the association of the *FKBP6* gene with stallion subfertility due to impaired acrosome reaction (IAR). We developed two TaqMan genotyping assays for the IAR-associated SNPs in the *FKBP6* gene and confirmed significant association ($P < 0.0001$) between low per-cycle pregnancy rate and the IAR-susceptibility *FKBP6* genotype A/A-A/A in Thoroughbreds. Through whole genome sequencing, we identified a 171 Kbp haplotype block specific to subfertile Thoroughbreds with the *FKBP6* A/A-A/A genotype. Testis transcriptome analysis of *FKBP6* in select Thoroughbreds revealed that the gene is expressed from both alleles and not monoallelically as thought previously.

Regarding the male-specific Y chromosome (MSY), we developed droplet-digital PCR assays for copy number (CN) analysis of 7 multi-copy genes and *SRY* to determine CN variation (CNV) of these genes in a global equine population and compare CNV with MSY haplotypes (HTs). We show that *TSPY* is the most variable gene among individuals and breeds. The *SRY* gene is a single-copy gene in most horses but may have additional copies in indigenous breeds. Comparison of MSY gene CNV with MSY HTs, revealed no correlation between the two forms of MSY variation. Additionally, we conducted CN analysis of these genes in horses with disorders of sex development or subfertility. We observed significantly lower CNs in *TSPY* and *ETSTY2* within cryptorchid males, though no significant CNVs were observed in subfertile males. CN analysis of these genes indicated that the ampliconic assembly in eMSYv3 needs

improvement. Thus, we re-sequenced and re-assembled this region utilizing Nanopore technology. We generated an improved 1.53 Mbp assembly of the ampliconic MSY and closed one of the three gaps in single-copy Y. Finally, we conducted a detailed molecular cytogenetic analysis of a reciprocal translocation between horse chromosomes Y and 13 in a Friesian stallion with complete azoospermia.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Dr. Terje Raudsepp (advisor), Dr. E. Gus Cothran, and Dr. Brian Davis of the Department of Biomedical Sciences and Dr. Dickson Varner of the Department of Large Animal Clinical Sciences.

The fertility data analyzed for section 2.2 was provided by Professor Dr. Dickson Varner. The staring material for section 3.2.1 was provided by Dr. Barbara Wallner of the Institute for Animal Breeding and Genetics, University of Veterinary Medicine, Vienna. The haplotype analysis presented in section 3.3.2 was completed in part by Lara Radovic of the Institute for Animal Breeding and Genetics, University of Veterinary Medicine, Vienna.

Samples utilized for the studies in sections 2.2, 3.3, and 3.4 were obtained from the Molecular Cytogenetics and Animal Genetics Laboratory at Texas A&M University and Institute of Animal Breeding and Genetics at the Veterinary University of Vienna repositories.

All other work conducting for the dissertation was completed by the student independently.

Funding Sources

This work was made possible in part by the Morris Animal Foundation under Grant Number D19Eq-051 and the United States Department of Agriculture under the USDA-AFRI Grant Number 2018-06521.

NOMENCLATURE

BAC	Bacterial Artificial Chromosome
BES	BAC end sequences
CN(s)	Copy number(s)
CNV(s)	Copy number variation(s)
ddPCR	Droplet-digital PCR
ECA	*Equus caballus*, ECA
EVD	Equine Variant Database
FISH	Fluorescence *in situ* hybridization
gDNA	Genomic DNA
GWAS	Genome wide association study
HG	Haplogroup
HT(s)	Haplotype(s)
HWE	Hardy Weinberg Equillibrium
IAR	Impaired Acrosome Reaction
Kbp	Kilo base pair
MSCI	Meiotic sex chromosome inactivation
MSY	Male specific Y chromosome
PAR	Pseudoautosomal Region
PCPR	Per-cycle pregnancy rate
PSPR	Per-season pregnancy rate
scY	Single copy Y chromosome region
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
mcY	Multi-copy/ampliconic Y chromosome region
WGS	Whole genome sequencing

TABLE OF CONTENTS

Page

ABSTRACT ii

CONTRIBUTORS AND FUNDING SOURCES iv

NOMENCLATURE v

LIST OF FIGURES viii

LIST OF TABLES x

1. INTRODUCTION 1

1.1. Stallion fertility and cytogenetics 1
1.2. Contribution of autosomal factors to stallion fertility. 11
1.3. The male specific Y chromosome and stallion fertility 15

2. IMPAIRED ACROSOME REACTION IN STALLIONS AND THE *FKBP6* GENE 21

2.1. Objectives 21
2.2. Thoroughbred stallion fertility and the association with *FKBP6* genotypes 22
2.2.1 Experimental methods and design 22
2.2.2 Results 24
2.2.3 Discussion 30
2.3. Accessing the relationship of *FKBP6* and the IAR using whole genome sequencing 33
2.3.1 Experimental methods and design 33
2.3.2 Results 36
2.3.3. Discussion 41
2.4. Accessing the relationship of *FKBP6* and the IAR through testis RNA-sequencing 44
2.4.1. Experimental methods and design. 44
2.4.2. Results 46
2.4.3 Discussion 48
2.5. Conclusions and future studies for the *FKBP6* gene and the IAR 49

3. GENOMICS OF MALE FERTILITY WITHIN THE CONTEXT OF THE Y CHROMOSOME 53

3.1 Objectives 53
3.2. Copy number variation in horse Y chromosome genes 54
3.2.1 Experimental methods and design 54

3.2.2 Results ... 58
3.2.3. Discussion ... 74
3.3. Improvement of the current Y reference sequence eMSYv3 ... 84
3.3.1. Improvement of the multi-copy region of the horse Y chromosome ... 85
3.3.2. Improvement of the single copy region of the horse Y chromosome ... 89
3.3.3. Improving the current Y reference sequence eMSYv3 discussion ... 96
3.4. Y chromosomal rearrangements influence male fertility ... 101
3.4.1. Experimental methods and design ... 101
3.4.2. Results ... 105
3.4.3. Discussion ... 111

4. CONCLUDING REMARKS ... 118

REFERENCES ... 121

APPENDIX A: SUPPLEMENTARY FIGURES ... 153

APPENDIX B: SUPPLEMENTARY TABLES ... 158

LIST OF FIGURES

Page

Figure 1. Comparative results of FKBP6 exon 5 SNP 11,353,372G>A genotyping by two methods in three different stallions Horse #1, Horse #2, Horse #3 25

Figure 2. The distribution of per-cycle pregnancy rates (%; y-axis) of 150 stallions in relation to their genotypes for A: SNP G>A, B: SNP A>C, and C: genotype combinations (x-axis) 28

Figure 3. Genotype Plot of a 182 Kbp region depicting a 171 Kbp identity by decent region specific to case Thoroughbreds with the A/A-A/A FKBP6 genotype. 37

Figure 4. A 110 Kbp region surrounding the FKBP6 tagging SNPs in 21 individuals with the double homozygous A/A-A/A FKBP6 genotype.. 40

Figure 5. Box and whisker plot illustrating the range of copy number variation of horse MSY multi-copy genes and SRY in a multi-breed cohort of 209 normal male horses. 61

Figure 6. Haplotype (HT) distribution and frequency plot based on MSY tree modified from Felkel et al. 2019. 67

Figure 7. Schematic representation of the SRY-region in horse MSY proposing a mechanism for SRY/RBMY CNV. 80

Figure 8. Current published sequence map of eMSYv3 showing different sequence classes, annotated genes, their copy numbers, evolutionary origin, and direction of transcription (orientation of arrowheads). 84

Figure 9. Published horse MSY contig and BAC map from Supplementary Figure S1 in Janecka et al. 2018.. 86

Figure 10. Detailed BAC map of Gap 1 in eMSYv3. 93

Figure 11. Examples of metaphase, interphase, and DNA Fiber-FISH used to assemble BAC clones across Gap 1. 96

Figure 12. Determining the extent of genetic exchange between ECAY and ECA13. 106

Figure 13. Partial metaphase spreads showing dual-color FISH results determining the size and orientation of translocated segments between ECAY and ECA13. 108

Figure 14. Schematic representation of the reciprocal translocation between ECAY and ECA13 109

Figure 15. Schematic presentation of a hypothetical quadrivalent that may form in meiosis prophase of some primary spermatocytes of the Friesian stallion. 113

LIST OF TABLES

Page

Table 1. Summary table of chromosomal aberrations in genetically male horses........................ 8

Table 2. Allele, genotype and combined genotype frequencies of the SNPs in FKBP6 exon 5 for the cohort of 518 Thoroughbreds and the chi-square P-value for HWE statistics.. 26

Table 3. Sample identification, genotype, fertility status and breed of individuals used for whole genome sequencing. 34

Table 4. Detailed information on the 38 putative IAR-associated variants identified within the 171 Kbp IBD block. 38

Table 5. Sample identification, genotype, and breed type of individuals used for RNA sequencing.. 45

Table 6. Individual testis RNA-sequencing FKBP6 SNP genotype and cooresponding allelic depth parameters per SNP.. 47

Table 7. Droplet digital PCR assays for all Y chromosome genes analyzed in this study 55

Table 8. Comparison of gene copy numbers between eMSYv3 (Janecka et al. 2018) and ddPCR analysis of the reference male Bravo and a multi-breed cohort 60

Table 9. Average copy number (CN) and corresponding standard deviation (SD) of 7 MSY ampliconic genes and SRY across horse breeds and related equids. 62

Table 10. Copy number variation of SRY and RBMY in 21 males across 4 indigenous breeds. 64

Table 11. MSY gene average CN variation between cryptorchid and normal American Quarter Horses. 70

Table 12. MSY gene CN and HT comparison between related males. 73

Table 13. Newly identified Y chromosome BACs and newly sequenced BACs used in this study. 94

Table 14. Information about all BACs used for FISH in this study. 103

Table 15. CN analysis of seven MSY multi-copy genes and single copy SRY gene in the Friesian stallion (H787) and 16 control males. 110

1. INTRODUCTION

1.1 Stallion fertility and cytogenetics

Horses are an economically and culturally important domestic species as they have served humans in agriculture, warfare, and transportation throughout history. Today, horses continue to be used in sports, transportation, and agriculture, and are valued as companions. Currently, the equine industry contributes approximately $50 billion in direct impact to the US economy with a employment of nearly 1 million jobs (American Horse Council: https://www.horsecouncil.org/resources/economics/).

Stallion fertility is a core component of the equine industry with stallion subfertility as a recognized problem. Because one stallion typically sires numerous foals, reduced fertility could result in major financial losses due reduced foal crop (Colenbrander *et al.* 2003). Furthermore, stallions are selected as breeding sires based on their pedigree, athletic performance and conformation, rather than fertility (Colenbrander *et al.* 2003). This is probably a major reason why, 36%-43% of prospective breeding stallions fail in breeding soundness tests (Blanchard & Varner 1997; Woods *et al.* 2000), and stallions that have passed the breeding soundness evaluation, sometimes turn out to be subfertile. As a result, reduced fertility among breeding stallions has emerged as a significant issue for the equine industry (Leeb *et al.* 2005; Leeb 2007).

Male fertility is a complex trait influenced by multiple elements (behavioral, environmental, and physiological), including a considerable genetic component (Matzuk & Lamb 2008). For this reason, it is important to improve our knowledge about the factors contributing to reduced fertility so that improved diagnostic and preventive approaches can be developed. Today, the only routinely conducted genetic test for stallion reproductive soundness

is karyotyping for the detection of chromosome rearrangements (Raudsepp & Chowdhary 2016). However, the majority of subfertile or infertile males subjected for cytogenetic diagnostic testing through the Texas A&M Molecular Cytogenetics laboratory, have a normal 64,XY karyotype (Bugno-Poniewierska & Raudsepp 2021). In the rare event that an infertile/subfertile stallion has a chromosomal abnormality, cases are both clinically and cytogenetically unique. Male reproductive disorders caused by chromosome mutations can involve both the sex chromosomes and autosomes.

Cases associated with stallion subfertility or infertility vary from sex chromosome and autosome aneuploidies (i.e., abnormal chromosome numbers instead of the standard 64,XY), to structural rearrangements within a chromosome or involving two different chromosomes. The latter can be autosomes or a combination of an autosome and the Y chromosome (Bugno-Poniewierska & Raudsepp 2021). A handful of cases of infertile stallions have non-mosaic or mosaic sex chromosome aneuploidies (Table 1) with karyotypes such as (but not limited to): 65,XXY; 63XO/64,X,i(Yq); 64,XY/65,XXY; 63,XO/65,XYY; 66,XXXY. Males who harbor abnormal sex chromosomes due to aneuploidies tend to have phenotypes consistent with intersex syndrome or pseudo-hermaphroditism involving ambiguous external genitalia (Table 1).

To date, three cases of horses with either mosaic or non-mosaic 65,XXY karyotypes - the equivalent of human Klienfelter's syndrome – have been described (Kubien *et al.* 1993; Makinen *et al.* 2000). Phenotypes of the 3 cases vary from aberrant external genitalia, to normal male genitalia, but all had abnormal sperm production. One case involved a draft horse who had a normal penis and sheath, but the scrotum was absent, and the individual had two small gonads and a large epididymis (Kubien *et al.* 1993). The second case, a French Trotter, had normally descended testis, but a small penis and unusual azoospermic ejaculate (Makinen *et al.* 2000). The

third case had a mosaic karyotype 64,XY/65,XXY and normal external male genitalia, but was azoospermic. The incidence of XXY sex chromosome aneuploidy in horses is rare (0.025%) based on a large population study targeting sex specific microsatellite markers (Kakoi *et al.* 2005). Compared to humans, the frequency of the XXY sex chromosome aneuploidy is higher than that in equine, with Klinefelter's syndrome affecting anywhere from 0.1% to 0.2% of live born males resulting in hypogonadism and fertility problems (Los & Ford 2019). XXY sex chromosome aneuploidies is rare in horses and other domestic animals, but in humans, XXY is the most common sex chromosome abnormality (Los & Ford 2019). When compared to other mammals, XXY aneuploidy frequency remains low, with only a handful of cases reported in domestic bovids (reviewed by (Iannuzzi *et al.* 2021)), 6 cases in domestic dogs (reviewed by (Szczerbal & Switonski 2021)). Conversely, XXY aneuploidy is more common in domestic cat, with incidences occurring at a frequency of 16.5%, resulting in a male calico cat (our unpublished data and (Pedersen *et al.* 2014; Szczerbal *et al.* 2018; Szczerbal & Switonski 2020)).

Most XXY cases in domestic bovids have been mosaic with underdeveloped reproductive organs (Rieck *et al.* 1969; Dain & Bridge 1978; Rieck *et al.* 1982; Takebayashi & Jorg 1986). The only non-mosaic case of XXY, however, had a Robertsonian translocation between chromosome 1 and 29 (Schmutz *et al.* 1994). In goats, The only reported case was mosaic XX/XXY and fertile (Bhatia & Shankar 1992). (Bhatia & Shankar 1992). In dogs, the XXY abnormalities have been associated with hyperplastic testes and overall feminization, sterility, cryptorchidism, and testicular tumors (see (Szczerbal & Switonski 2021)). It is noteworthy that wile XXY aneuploidy is rare in domestic animal, including horses, it is the most common sex chromosome abnormality in humans affecting 0.15% of live male births (Bonomi *et al.* 2017; Kanakis & Nischlag 2018; Mahyari *et al.* 2021).

On the other hand, horses carrying and extra Y chromosome (65,XYY) –the equivalent of Jacob's syndrome in humans – are rare, with only two cases reported to date, and associated with various forms of disorders of sexual development (Hohn *et al.* 1980; Paget *et al.* 2001). Two cases with mosaic 63,XO/65,XYY have been reported: the first case, was identified as a pseudohermaphrodite (Hohn *et al.* 1980); the second, more recent case, was found in an Arabian initially identified as female until the horse started to exhibit stallion-like behavior as a yearling (Paget *et al.* 2001). Further clinical and cytogenetic analysis categorized the individual as a male-pseudohermaphrodite (Paget *et al.* 2001). In humans, the XYY syndrome is he next common sex chromosome abnormality after Klinefelter syndrome (Gekas *et al.* 2001; Shi & Martin 2001; Rives *et al.* 2005), and affects 1 of 1000 male births (0.1%) (Morel *et al.* 1999). The phenotype may involve "over-masulinization" phenotypes such as increased height and stature (Sood & Clemente Fuentes 2022), though often there are no obvious phenotypic changes. Men with 47,XYY syndrome can have variable sperm counts, ranging from normal to azoospermia (Wu *et al.* 2016). The condition, however, is rarely found in other mammals, with only 4 cases of XYY aneuploidy reported in bovids (see (Iannuzzi *et al.* 2021)), and no such cases identified in in domestics cats and dogs. All cases so far identified in bovids are mosaic for XYY cells and have testicular hypoplasia (Miyake *et al.* 1981; Jaszczak *et al.* 2003; Iannuzzi *et al.* 2021), although one ram was identified with mosaic XYY and did not have noticeable phenotypic affects (Moraes *et al.* 1980).

Other sex chromosome anueploidies are in male horses are scarce; only one non-mosaic individual with 66,XXXY has been described to date, and had an intersex phenotype with underdeveloped genitalia of both sexes (Glugovsci *et al.* 1970). Finally, there have only been two cases where a horse was mosaic for an isochrome Y chromosome with the karyotype

63,XO/64X,i(Yq) and two Y chromosomes joined together at the centromere (Herzog *et al.* 1989; Das *et al.* 2012). Both individual horses were described as intersex, or pseudo-hermaphrodites with multiple genital deformities, including testis-like gonads that were not descended, along with abnormal clitoris and vulva (Herzog *et al.* 1989; Das *et al.* 2012).

Several cases of liveborn male horses with autosomal aneuplodies have been described (Table 1). All cases with autosomal aneuploidies involve the smallest of the equine chromosomes (ECA) – ECA23, 26, 27, 28, 30, 31 (Power 1987; Klunder & McFeeley 1989; Zhang *et al.* 1992; Buoen *et al.* 1997; Lear *et al.* 1999). This is because the genetic imbalance for larger chromosomes is likely not viable. Live born male horses with autosomal aneuploides typically have multiple congenital defects, including cryptorchidism, and are sterile (in cases where they are kept alive till sexual maturity). For example, investigation of meiosis in a stallion with trisomy ECA28 and azoospermia confirmed that the aneuploidy was the cause of azoospermia (Power *et al.* 1992).

Male subfertility due to structural chromosomal rearrangements (i.e., translocations, duplications, deletions, and inversions) are equally limited. Translocations involve nonhomologous chromosomes which exchange genetic content, or fuse together, giving rise to reciprocal or non-reciprocal translocations (Morin *et al.* 2017). In individuals where these translocations are balanced, carriers are phenotypically normal but have reduced fertility owing to the fact that not all gametes produced by the translocation carrier will be genetically balanced and do not result in viable offspring. Individuals who carry the genetically balanced translocation can pass the translocation on to the next generation if the gamete with a balanced translocation is fertilized. Conversely, fertilization of unbalanced gametes results in early embryonic loss or

death of the fetus, lowering fertility rates of the individual (Raudsepp & Chowdhary 2016; Gosh *et al.* 2020).

Balanced translocations are some of the few chromosomal aberrations which can be transferred to the offspring and will result in similar subfertility in the next generation (Raudsepp & Chowdhary 2016). There are currently two cases of "translocation families" involving elite stallions who carry genetically balanced autosomal translocations and have passed the chromosomal aberration to a number of offspring (Table 1) (Durkin *et al.* 2011; Gosh *et al.* 2020). One such familial case of a recurrently transmitted balanced translocation was initially identified in a phenotypically normal Warmblood stallion who had a history of producing foals with congenital abnormalities (Gosh *et al.* 2020). It was later identified that this stallion carried a translocation between ECA4 and ECA30, that was then transmitted to the next generation in 5 out of 9 offspring studied, where all but one offspring was phenotypically normal (Gosh *et al.* 2020). This particular case was submitted for chromosomal analysis due to poor fertility records and because of familial congenital defects: analysis of the breeding records showed that live foals were only achieved 61% of the time, and that almost 50% of the foals born had congenital abnormalities and were later euthanized due to poor thriving (Gosh *et al.* 2020). The second family of translocations was discovered in an elite Thoroughbred stallion who passed the chromosomal abnormality to 8 offspring (Durkin *et al.* 2011). This stallion was subject to cytogenetic analysis because of subfertility due to recurrent early embryonic loss (Durkin *et al.* 2011). This particular Thoroughbred carried a balanced translocation between ECA5 and ECA16 with a derived marker chromosome (Durkin *et al.* 2011). As a consequence of the translocation, two-thirds of the gametes produced by the stallion are unbalanced and cause infertility or early embryonic loss, whereas the viable gametes produced will continue to propagate the t(5;16)

chromosomal abnormality and phenotypic subfertility. Both cases involved desirable stallions based on their athletic performance and are good examples of why systematic chromosome analysis is a necessary part of breeding soundness evaluation to reduce economic losses due to the propagation of such translocations over generations.

In addition to the two breeding stallions carrying and propagating balanced autosomal translocations, only two other cases of autosomal translocations have been described in male horses. First, a male Thoroughbred with a translocation between ECA1 and ECA30 that was subjected to chromosomal analysis due to reduced fertility (Long 1996). Second, a de novo translocation was found in an Arabian male that was produced by somatic cell nuclear transfer (SCNT) and subjected for chromosome analysis due to an abnormal reproductive clinical report (Gosh *et al.* 2020). The cloned Arabian was found to carry a translocation between ECA12 and ECA25, and was phenotypically normal aside from small testis (Gosh *et al.* 2020). Notably, a cloned brother of this horse that was produced by SCNT from the same donor, was chromosomally normal.

While balanced structural rearrangements of autosomes typically produce phenotypically normal individuals, this is not the case with structural rearrangements of the Y chromosome which affect spermatogenesis causing azoospermia or result in various disorders of sexual development (Durkin *et al.* 2011; Raudsepp & Chowdhary 2016; Ruiz *et al.* 2019; Castaneda *et al.* 2021b). One such case has been described in a Shetland Pony with a 64,XY male karyotype but abnormal external genitalia with no penis. The horse had an unusually small Y chromosome due to a large deletion in the heterochromatic portion (Bugno-Poniewierska & Raudsepp 2021) which is comprised of massively amplified sequences of an equine testis-specific transcript in Y 7 (*ETSTY7*) of hitherto unknown functions (Janecka et al. 2018).

When it comes to Y-autosome rearrangements, the associated phenotypes are largely dependent on which region of the Y chromosome is involved, whether the associated autosome is an acrocentric or a non-acrocentric, and whether the translocation is reciprocal or non-reciprocal (Castaneda *et al.* 2021b). For example, in humans, non-reciprocal translocations of the distal heterochromatic portion of Yq to an acrocentric autosome does not affect the phenotype or fertility and has been transmitted as chromosomal variants both by men and women (Beruitex *et al.* 1979; Morel *et al.* 2002). In contrast, 80% of men with balanced reciprocal Y-autosome translocations involving euchromatic portion of the Y, have non-obstructive oligozoospermia or azoospermia (Morel *et al.* 2002; Sun *et al.* 2005; Wang *et al.* 2017). The genetic content of the Y chromosome and the type of the autosome involved in Y-autosome translocations directly affect the meiotic behavior of the aberrant chromosome and the functional status of both the Y-linked and autosomal genes (Sun *et al.* 2005; Barasc *et al.* 2012; Mary *et al.* 2018).

To date, only one case of stallion azoospermia has been associated with Y-autosome rearrangements involving a Friesian stallion with Y-ECA13 translocation (Ruiz *et al.* 2019; Castaneda *et al.* 2021b). This stallion had normal external genitalia, normal erection, and normal ejaculation. However, no pregnancies were achieved due to the complete lack of sperm in the ejaculate (Ruiz *et al.* 2019). Based on Y-autosome translocation studies in human and pig, Castaneda *et al* (2021b) speculated that the Y-autosome translocation creates aberrant sex body formation in male meiosis prophase, affecting male specific meiotic sex chromosome inactivation (MSCI), and resulting in sterility. This is the first case of a phenotypically normal, but sterile stallion with a large rearrangement involving the Y chromosome.

Table 1. Summary table of chromosomal aberrations in genetically male horses.

Individual Karyotype	Chromosome abnormality	Phenotype	Fertility Status	Offspring karyotypes	Reference
65,XY+27 (3 cases)	Autosomal aneuploidies	Congenital defects	n/a	n/a	(Zhang *et al.* 1992; Buoen *et al.* 1997; Brito *et al.* 2008)
65,XY+30 (2 cases)	Autosomal aneuploidies	Congenital defects	n/a	n/a	(Brito *et al.* 2008; Bugno-Poniewierska & Raudsepp 2021)
64,XY,i(26q)	Autosomal aneuploidies	Congenital defects	n/a	n/a	(Brito *et al.* 2008)
65,XY+23	Autosomal aneuploidies	Congenital defects	n/a	n/a	(Klunder & McFeeley 1989)
65,XY+28	Autosomal aneuploidies	Congenital defects	n/a	n/a	(Power 1987)
65,XY+31	Autosomal aneuploidies	Congenital defects	n/a	n/a	(Lear *et al.* 1999)
64,XY,t(1;30)	Autosomal translocation	n/a	Subfertile	n/a	(Long 1996)
64,XY,t(5;16)+mar	Autosomal translocation; familial case	Normal	Subfertile	8 with translocations; 2 normal	(Durkin *et al.* 2011)
64, XY;t(4;30)	Autosomal translocation; familial case	Normal	Subfertile	5 with translocations; 4 normal	(Gosh *et al.* 2020)
64,XY,t(12;25)	Autosomal translocation	Small testes	n/a	n/a	(Gosh *et al.* 2020)

65,XXY (2 cases)	Sex chromosome and ploidy mosaicism	Abnormal genitalia	Infertile	n/a	(Kubien *et al.* 1993; Makinen *et al.* 2000)
63XO/64,X,i(Yq) (2 cases)	Sex chromosome and ploidy mosaicism	Abnormal genitalia	Infertile	n/a	(Herzog *et al.* 1989; Das *et al.* 2012)
64,XY/65,XXY	Sex chromosome and ploidy mosaicism	Abnormal genitalia	Infertile	n/a	(Makinen *et al.* 2000)
63,XO/65,XYY (2 cases)	Sex chromosome and ploidy mosaicism	Male pseudo-hermaphrodite	Infertile	n/a	(Hohn *et al.* 1980; Paget *et al.* 2001)
64,XYdel(Y)	Y chromosome structural rearrangement	No penis	Infertile	n/a	(Bugno-Poniewierska & Raudsepp 2021)
64,XY,t(Y;13)	Y;autosome translocation	Azoospermic	Infertile	n/a	(Ruiz *et al.* 2019; Castaneda *et al.* 2021b)

In addition to chromosomal aneuploidies and structural rearrangements, many cases have been described where a horse has a normal *SRY*-positive 64,XY male karyotype but is phenotypically abnormal showing a spectrum of disorders of sex development and sterility (reviewed by (Bugno-Poniewierska & Raudsepp 2021)). Most cases are described as intersex/ambiguous sex, or pseudohermaphrodites, some are female-like horses with stallion-like behavior, and most have abnormal genital tract with ovotestes, testicular feminization or

rudimentary abdominal testes (see review (Bugno-Poniewierska & Raudsepp 2021). Despite the range of female-to-male like appearances, these individuals are genetically male with an intact Y chromosome and the male sex determining gene, *SRY* (Raudsepp *et al.* 2010). As such, chromosomal abnormalities identified through clinical cytogenetic analysis reveal only a fraction of genetic defects that may cause reduced stallion fertility or infertility. Thus, the genetic causes of the majority of subfertile phenotypes remain undefined. This is largely because little is known about the molecular underpinnings of male fertility in mammals.

1.2. Contribution of autosomal factors to stallion fertility.

It is estimated that about 10-20% of all genes in the mammalian genome are involved in male fertility (Carrell 2007). Due to the large number of genes involved, the molecular research of stallion fertility requires genome wide approaches. Thus, great expectations are placed on the use of horse whole genome sequence information (Kalbfleisch *et al.* 2018) and the ongoing functional annotation of the horse genome (Andersson *et al.* 2015; Giuffra *et al.* 2019; Peng *et al.* 2021a; Peng *et al.* 2021b). Availability of these essential tools has promoted organized studies of the horse genome (Janecka *et al.* 2018; Kalbfleisch *et al.* 2018), which have resulted in identification of a number of genetic variants that influence pigmentation, disease and performance traits (reviewed by (Raudsepp *et al.* 2019)). Despite the decade of genomic searches for candidate male fertility genes, only a few loci or genomic regions have been associated with male fertility parameters and phenotypes (reviewed by (Raudsepp 2020; Laseca *et al.* 2021)).

The most common gene associated with an infertility in horses is the male sex determining gene *SRY* as its deletion is associated with 64,XY *SRY*-negative disorders of sexual development (Raudsepp *et al.* 2010). Mutations and deletions in the X-linked androgen receptor

gene (*AR*) have previously been associated with androgen insensitivity syndrome (AIS) and cause genetically male horses to be sterile and appear female (Revay *et al.* 2012; Bolzon *et al.* 2016; Welsford *et al.* 2017). However, the two most recent point mutations and deletions found within the *AR* gene are distinct from those previously identified (Villagomez *et al.* 2020), suggesting that *AR* is another heterogeneous gene possibly responsible for a number of 64,XY *SRY*-positive disorders of sex development. On a similar note to the *AR* gene and its involvement with 64,XY-*SRY* positive disorders of sexual development, a 200kbp deletion in the *AKR1C* gene cluster in ECA29 has been considered a risk factor for 64,XY-*SRY* positive disorders of sexual development and cryptorchidism in male horses (Gosh *et al.* 2014a; Gosh *et al.* 2014b; Gosh *et al.* 2016).

In the context of stallion fertility, several autosomal candidate or high impact genes have been identified in multiple European horse breeds (Schrimpf *et al.* 2016). In addition to the high impact genes, 4 autosomal genes have been associated with the paternal component of pregnancy rate per estrus: *CRISP3* (Hamann *et al.* 2014), *PRLR* (Giesecke *et al.* 2010a), *SPATA1* (Giesecke *et al.* 2009), and *INHBA* (Giesecke *et al.* 2010b). Additional autosomal candidate fertility genes are *CRISP1* and *STK31*, with the former associated with accurate sperm-egg fusion mechanisms (Giese *et al.* 2002), and later to be involved with the reorganization on sperm chromatin during spermatogenesis (Sabeur *et al.* 2008). However, among these, the most intriguing is the autosomal FKBP prolyl isomerase family member 6 (*FKBP6*) because of its conflicting associations in Thoroughbreds and Hanoverians. In Thoroughbreds, *FKBP6* is considered as a susceptibility gene for stallion subfertility owing to an impaired acrosome reaction (IAR) (Raudsepp *et al.* 2012).

A key factor in gamete interaction and successful fertilization is the acrosome reaction (AR) which facilitates penetration of the spermatozoa through the zona pellucida (protective coat) of the oocyte. The genetics underlying the AR event is complex and involves many genes including those encoding neurotransmitter receptors (Sato *et al.* 2000; Meizel & Son 2005), calcium channels (Fukami *et al.* 2001; Jin *et al.* 2005; Gibbs *et al.* 2011) and those facilitating membrane fusion (Tanigawa *et al.* 2008) and exocytosis (Kitamura *et al.* 2005). More recently, studies of the AR have focused on the sperm proteome to understand the complex mechanism (Lin *et al.* 2007; Tanigawa *et al.* 2008; Satouh *et al.* 2012; Zhang *et al.* 2019b; Fujihara *et al.* 2020). The few known genetic mutations associated with the impaired acrosome reactions (IAR) have been identified in mice involving a missense mutation change within the alpha subunit of Glycine Receptor protein, which results in a truncation protein (Sato *et al.* 2000; Meizel & Son 2005). More recently, mutations within pachytene PIWI-interacting RNAs (piRNAs) in the pi6 gene have been associated with male fertility, where the mutant sperm penetrated the zona pellucida only 5% of the time (Wu *et al.* 2020). The molecular causes of IAR are complex and mouse models suggest that the success of the AR is a result of several genes (Mayernia *et al.* 2003).

Stallion subfertility associated with acrosomal dysfunction was first reported in the late 1990s by analyzing the acrosomal reaction of both fertile and subfertile stallions whose seasonal pregnancy rates were < 30% (Meyers *et al.* 1995; Meyers *et al.* 1996). In 2001, stallions with very low per-cycle pregnancy rates (< 15%), despite having sperm quality parameters comparable for fertile stallions (Kenney *et al.* 1983) and normal testicular size, produced sperm which failed to undergo the acrosomal reaction test (Varner *et al.* 2001). As sperm characteristics are like that of fertile stallions (i.e., sperm motility, morphology, total sperm number), the IAR

condition remains unnoticed during routine breeding soundness evaluation which do not test for acrosomal function (Varner *et al.* 2000). Stallions with possible IAR are not identified until later due to their poor breeding records or classified as idiopathic subfertile (Varner *et al.* 2000; Brinsko *et al.* 2007). The stallion subfertility phenotype of low per-cycle pregnancy rates associated with the IAR is variable (Castaneda *et al.* 2021a) because 97% of the sperm fails to acrosome react, where the remaining 3% can perform normal acrosomal reaction (Brinsko *et al.* 2007).

The Thoroughbred is so far the only horse breed where IAR has been confirmed as a cause of stallion subfertility (Varner *et al.* 2001). Furthermore, the affected stallions described in the initial study were closely related, indicating that the condition might have a genetic basis. This notion led to a genome-wide association study (GWAS) using a single nucleotide polymorphism (SNP) microarray platform (Equine 50K Beadchip (McCue *et al.* 2012)). The GWAS involved 44 Thoroughbred stallions, 7 with a confirmed IAR phenotype and 37 controls, and resulted in the discovery of an IAR susceptibility gene (Raudsepp *et al.* 2012). The GWAS revealed significant association (*P*-value = 4.93E-11) between the IAR phenotype and a double homozygous A/A-A/A genotype for two SNPs in *FKBP6* gene exon 5: a synonymous SNP (EquCab3.0 assembly; chr13:11,353,372G>A) and a non-synonymous (missense mutation) SNP (EquCab3.0 assembly; chr13:11,353,436C>A) (p.167H>N). The two SNPs individually were also significantly associated with the IAR, particularly the synonymous SNP chr13:11,353,372G>A (*P*-value=3.20E-07), though not as strongly as the double homozygous genotype A/A-A/A (Raudsepp *et al.* 2012). Additionally, a large cohort analysis of 265 male horses of different breeds showed that the IAR-susceptibility genotype is present in 100% of stallions with confirmed IAR, but only in 2% of the general male population (Raudsepp *et al.* 2012). It is,

however, necessary to underline that fertility data for the general male population was largely unavailable.

Conversely, a similar study in 216 Hanoverian stallions did not detect any significant interaction effect between the two *FKBP6* exon 5 SNPs and did not find a significant association between stallion fertility and the double homozygous A/A-A/A genotype (Schrimpf *et al.* 2015). In contrast to the study in Thoroughbreds (Raudsepp *et al.* 2012), this group reported higher conception rates of Hanoverian stallions homozygous (A/A) for the SNP resulting in the missense mutation. Thus, follow-up studies in a controlled cohort of Thoroughbred breeding stallions with access to breeding data are needed to validate the association of the double homozygous A/A-A/A *FKBP6* genotype with the IAR phenotype identified by Raudsepp and colleagues (2012).

Based on the data presented by Raudsepp *et al*. (2012), it was proposed that IAR displays incomplete penetrance (i.e., some individuals will not express the phenotype despite carrying the susceptibility genotype). However, incomplete penetrance can lead to phenotyping variation. This is true for affected stallions with the IAR phenotype: while all affected stallions present a subfertility phenotype, the degree of subfertility is variable (Castaneda *et al.* 2021a). Therefore, determining the prevalence of the IAR susceptibility genotype in larger Thoroughbred population is needed to properly estimate the risk of IAR in Thoroughbred breeding populations.

1.3. The male specific Y chromosome and stallion fertility

The male specific Y chromosome is thought to have a unique contribution to male development and fertility, though molecular details of this contribution largely remain enigmatic in most species, including the stallion. The Y chromosome is one of the most structurally,

functionally, and evolutionarily distinct regions in the mammalian genome. During the evolution of the eutherian sex chromosomes from the same autosomal ancestor, the Y acquired a dominant testis-determining locus, which led to gradual cessation of X-Y recombination (Lahn & Page 1999; Bellot *et al.* 2014; Waters & Ruiz-Herrera 2020b). Reduced X-Y recombination was accompanied by a cascade of other evolutionary events in the Y such as an increase of structural rearrangements (inversions), gradual loss of ancestral genes, and reduction in size (Graves 2006; Bellot *et al.* 2014; Hughes *et al.* 2015; Hughes *et al.* 2020). On the other hand, the lack of recombination and male specific transmission, favored the acquisition and expansion of male-benefit genes in the Y, several of which became multi-copy or ampliconic with high (>99%) sequence identity between the copies (Skaletsky *et al.* 2003; Hughes *et al.* 2020). The presence of high identity euchromatic repeats has made Y chromosome sequence assembly challenging, thus, only four species have finished Y assemblies: human (Skaletsky *et al.* 2003), chimp (Hughes *et al.* 2010), rhesus macaque (Hughes *et al.* 2012; Hughes *et al.* 2015), and mouse (Soh *et al.* 2014). Though, draft Y assemblies are available for several species, including domestic animals such as cat and dog (Li *et al.* 2013; Janecka *et al.* 2018), pig (Skinner *et al.* 2016), bull (Bellot *et al.* 2014; Hughes *et al.* 2020), horse (Janecka *et al.* 2018), goat (Xiao *et al.* 2021), and sheep (Li *et al.* 2020). The increasing number of high-quality Y assemblies expands the scope of comparative studies across eutherian species ((see for example (Li *et al.* 2013; Bellot *et al.* 2014; Cortez *et al.* 2014; Janecka *et al.* 2018; Martinez-Pacheco *et al.* 2020), but also allows the study of intraspecific Y sequence variation. Of particular interest are copy number variations (CNVs) of multi-copy and ampliconic genes which, due to high sequence identity between the copies, are prone for non-allelic homologous recombination resulting in deletions and duplications. (Vogt *et al.* 1996; Repping *et al.* 2006; Lange *et al.* 2009).

In humans, male specific Y (MSY) CNVs have been studied for decades as some have direct link to male fertility. These CNVs are represented by three critical regions in human MSY, known as Azoospermia Factors AZFa, AZFb, and AZFc (Reijo *et al.* 1996; Saxena *et al.* 1996; O'Brien *et al.* 2010) which harbor multiple single- and multi-copy genes that are essential for sperm development. Various microdeletions in these regions affect the efficiency of spermatogenesis causing azoospermia, oligozoospermia, or oligoasthenozoospermia in men (O'Brien *et al.* 2010). The severity of a subfertility phenotype in each case depends on the size of the deletion, in addition to the exact region where it is located. Even though the organization and gene content of the Y chromosome is different across eutherians (Martinez-Pacheco *et al.* 2020), similar association between MSY CNVs and male fertility has been observed in other species. One example is found in mouse, were deletions in the ampliconic long arm of the mouse Y chromosome leads to decreased sperm quality and infertility (Ellis *et al.* 2005; Toure *et al.* 2005; Grzmil *et al.* 2007). Studies of MSY CNVs, particularly of amplicon variation, have also been initiated in primates (Ghenu *et al.* 2016; Oetjens *et al.* 2016; Tomaszkiewicz *et al.* 2016; Vegesna *et al.* 2020), murine subspecies (Morgan & Pardo-Manuel de Villena 2017), groups of bovids (Mukherjee *et al.* 2013; Mukherjee *et al.* 2015; Oluwole *et al.* 2017; Pei *et al.* 2019; Zhang *et al.* 2019a), dogs (Krzeminska *et al.* 2021), felids (Janecka *et al.* 2018), and the donkey (Han *et al.* 2017b). Though, aside from men and mice, there is limited knowledge about the association between MSY CNVs and male fertility in other species. In cattle, it is suggested that a lower number of *TSPY* copies could affect semen quality (Mukherjee *et al.* 2015), and in dogs, fewer *SRY* copies is thought to be associated with an increased risk of disorders of sex development (DSDs) (Krzeminska *et al.* 2021). Despite the economic importance of stallion fertility, CNV studies have not yet been conducted for the 15 multi-copy genes in horse MSY

(Janecka *et al.* 2018). Though, their novel acquisition and amplification in MSY, and testis-specific transcription suggest a role in male reproduction (Paria *et al.* 2011). Along with functions in spermatogenesis and male fertility, it is proposed that the Y chromosome carries important single copy genes which ensure male viability during the developmental process (Bellot *et al.* 2014).

Another important form of MSY intraspecific variation are single nucleotide variants (SNVs) which are excellent markers for determining Y haplotypes (HT) and tracing the history of patrilines (Jobling & Tyler-Smith 2003). MSY haplotype data have been widely used to infer the paternal ancestry of populations in human (Poznik *et al.* 2016; Jobling & Tyler-Smith 2017; Grugni *et al.* 2019), and several species such as cattle (Edwards *et al.* 2011), dog (Ding *et al.* 2012), goats (Vidal *et al.* 2017), and pigs (Guirao-Rico *et al.* 2018). The unprecedented low nucleotide variation of the horse MSY originally hindered the development of horse Y haplotyping (Wallner *et al.* 2003; Wallner *et al.* 2004; Wutke *et al.* 2018), however because of newly available sequencing tools, HT data for horses has significantly expanded over the last decade (Wallner *et al.* 2017; Felkel *et al.* 2018; Castaneda *et al.* 2019; Felkel *et al.* 2019; Liu *et al.* 2020). Fine scale mapping of Y chromosome haplotypes by genotyping for SNVs separated domestic horses into two groups denoted as “Crown” or “outgroup/Non-crown” (Wallner *et al.* 2017; Felkel *et al.* 2019). The majority of domestic horses fall within the “Crown” group and have conserved SNVs rAX and rAY (Wallner *et al.* 2017). However, select indigenous European and Asian breeds fall outside of the Crown group and carry an ancestral rAX allele and derived rAY allele (Felkel *et al.* 2018; Castaneda *et al.* 2019; Liu *et al.* 2020). The Przewalski’s Horse is used as the outgroup for all domestic horse Y haplotype genotyping because the species carries both a derived/ancestral rAX and rAY allele. Horse Y haplotype analysis of more genetically

diverse breed groups, such as those with Asian descent, have proven beneficial to the development of a refined horse Y haplotype map (Felkel *et al.* 2018; Liu *et al.* 2020). Few studies in human and primates have integrated MSY CNV and HT data to characterize the diversity the Y chromosome (Ye *et al.* 2018) and determine whether there is any correlation between the two forms of variation (Vegesna *et al.* 2019), however such studies have not been initiated in the horse.

Compared to the studies of autosomal factors for stallion fertility, identifying candidate loci in the Y chromosome has remained more elusive. This is, in part, due to the fact that genomic tools necessary to study the horse Y chromosome in a genomic context have only recently become available (Janecka *et al.* 2018). However, the current eMSYv3 assembly contains gaps and the assembly is only tentative in bioinformatically laborious areas such as amplified gene families, segmental duplications and palindromes that complicate the sequence assembly. This portion contains highly repetitive sequences of 99.9% or greater sequence similarity, multiple gene copies, and palindromic (mirror) sequences which create problems during sequencing and remain a bioinformatic challenge for assembly. Most commonly, the presence of many highly identical repeats will cause the *de novo* assembly to collapse multiple repeats into a single short contig (Bellot *et al.* 2018).These problematic areas can be improved only by using cutting-edge ultra-long molecule and long-read sequencing technology-based approaches.

Previously, a bacterial artificial chromosome (BAC) clone contig map formed the basis for horse MSY sequencing, and BACs located within multi-copy region of the horse Y chromosome (mcY) were sequenced on a multi-level platform utilizing both short read Illumina 2x250 paired-end MiSeq and 1 long-read PacBio SMRT cell (Janecka *et al.* 2018). These

combined approaches were, however, not sufficient to produce high quality assembly for the horse mcY. Today, the lowered sequencing costs have enabled long-read sequencing of high molecular weight DNA molecules of mcY BACs utilizing Oxford Nanopore technology (Hu *et al.* 2021). It is anticipated that long-read re-sequencing of known horse MSY multi-copy BACs and their alignment with eMSYv3, will generate an improved hybrid assembly for the horse Y chromosome, thus providing a better tool for the study of MSY role in stallion biology.

While the horse Y chromosome sequence is nearly complete, it is also important to close the 3 gaps located in the single copy portion of MSY to ensure a complete Y chromosome gene catalogue. A recent study targeting Y polymorphism by WGS of males, identified single copy Y sequences which were missing from eMSYv3 (Felkel *et al.* 2019). These sequences are of utmost importance as they may fall inside one of the 3 gaps in the eMSYv3 and will be useful for creating a higher quality reference.

The overall goal of the following studies is to improve our currently limited knowledge about select autosomal and Y-linked factors underlying stallion fertility. The autosomal component of stallion fertility of this dissertation will focus on the *FKBP6* gene as a susceptibility locus for IAR. For this, one aim is to determine the frequency and statistical significance of the IAR susceptibility genotype in a large cohort of Thoroughbred stallions with well-documented fertility data. The second is to identify other associated or causative factors of IAR through whole genome sequencing. As to the Y-linked factors, the first steps are to: 1) identify natural ranges of Y chromosome sequence variability in normal horse populations and 2) improve the horse Y chromosome assembly using state-of-the-art technologies. Only thereafter can Y chromosomal variability (CNVs, SNPs, chromosomal aberrations) be studied in relation to subfertile/infertile phenotypes and XY disorders of sex development (XY DSDs).

2. IMPAIRED ACROSOME REACTION IN STALLIONS AND THE *FKBP6* GENE[*]

2.1. Objectives

Follow-up studies using a large cohort of stallions with well-recorded reproductive data are needed to validate and determine the significance of the *FKBP6* SNPs identified by Raudsepp et al. (2012) in relation to other genetic factors in Thoroughbreds (Castaneda *et al.* 2021a). Even though the current findings clearly point at *FKBP6* as a susceptibility gene for IAR in stallions, we still lack a clear mechanistic and functional explanation. It is possible that the IAR-associated genotype A/A-A/A is tagging an underlying functional variant, and that *FKBP6* relates to IAR via regulatory or modifying pathways. To identify other genes and mutations contributing to IAR and ascertain causative relationship between *FKBP6* and IAR, we performed whole genome sequencing (WGS) of stallions with confirmed IAR. There are two main objectives which will be discussed in this section.

Objective 1: Validate the association between Thoroughbred stallion fertility and the IAR-association *FKBP6* genotypes and access the relationship between inbreeding levels (FIS) and fertility data (Castaneda et al 2021).

Objective 2: Access the relationship of *FKBP6* and the IAR using whole genome sequencing methods.

---

[*] Sections of this chapter are reprinted from the published work found in Animal Genetics under the Creative Commons Attribution License: Castaneda C., Juras R., Kjollerstrom J., Hernandez Aviles C., Teague S.R., Love C.C., Cothran E.G., Varner D.D. & Raudsepp T. (2021a) Thoroughbred stallion fertility is significantly associated with FKBP6 genotype but not with inbreeding or the contribution of a leading sire. *Animal Genetics* **52**, 813-23.

2.2. Thoroughbred stallion fertility and the association with *FKBP6* genotypes

The following sections, section 2.2.1, 2.2.2, and 2.2.3 will focus on the materials and methods, results, and a brief discussion about achieving the Objective 1 goal mentioned in 2.1. This work is done, in part, for the published work "*Thoroughbred stallion fertility is significantly associated with FKBP6 genotype but not with inbreeding or the contribution of a leading sire*" (Castaneda *et al.* 2021a).

2.2.1 Experimental methods and design*

Samples and DNA isolation

Here, we utilized a study cohort of 518 male and female Thoroughbreds from seven countries, primarily originating in the United States, to obtain the frequency of *FKBP6* genotypes. Of these, 350 individuals were available from the archives of the Molecular Cytogenetics and Animal Genetics Laboratories, Texas A&M University. Many of these horses lacked information about their fertility, except a few individuals previously identified as subfertile due to IAR (Raudsepp et al. 2012) or were subjected for genotyping due to idiopathic subfertility (Supplementary Table S1). Additional samples were obtained from a select cohort of 168 Thoroughbred breeding stallions. Fertility records such as per-cycle pregnancy rate (PCPR), per-season pregnancy rate (PSPR), and the total number of mares bred for that season, i.e., Mare

* Reprinted from the published work found in Animal Genetics under the Creative Commons Attribution License: Castaneda C., Juras R., Kjollerstrom J., Hernandez Aviles C., Teague S.R., Love C.C., Cothran E.G., Varner D.D. & Raudsepp T. (2021a) Thoroughbred stallion fertility is significantly associated with FKBP6 genotype but not with inbreeding or the contribution of a leading sire. *Animal Genetics* **52**, 813-23.

book, were obtained for 150 (Supplementary Table S2) out of the 168 stallions. For these, five-generation pedigree data was obtained from Pedigree Online Thoroughbred Database (https://www.pedigreeonline.com/ped/renew.php). Fertility data was not available for 18 out of 168 stallions and these horses were used for the analysis of total allele and genotype frequencies in the 518 Thoroughbred cohort. Genomic DNA was extracted from hair follicles or peripheral blood with Gentra Puregene Tissue or Blood kit (Qiagen, Germantown, MD, USA), respectively, following the manufacturer's protocols.

Development of TaqMan assays and genotyping

For large scale genotyping, we designed TaqMan™ assays for the SNPs chr13:11,353,372G>A and chr13:11,353,436A>C in *FKBP6* exon 5 based on EquCab3 (Kalbfleisch *et al.* 2018) using Custom TaqMan® SNP Genotyping Assays. The assays for the two *FKBP6* SNPs were designed separately using the guidelines provided by Applied BioSystems (ThermoFisher). TaqMan™ allelic discrimination reactions were conducted using the BioRad CFX96 Real-Time PCR instrumentation and corresponding analysis software. Genotyping assays were validated by direct Sanger sequencing of *FKBP6* exon 5 PCR amplicons of select individuals as described earlier (Raudsepp *et al.* 2012). After assay optimization and validation, the two genotyping assays were used to generate a combined *FKBP6* genotype for the 518 Thoroughbreds. Each animal was genotyped twice to confirm genotype results. For simplicity, SNP chr13:11,353,372G>A will be referred as SNP1 or SNP G>A and SNP chr13:11,353,436A>C will be referred as SNP2 or SNP A>C here out.

Pedigree and statistical analysis

The 150 Thoroughbred stallions (Supplementary Table S2) from the United States that had detailed breeding records, were analyzed for pedigree-based inbreeding coefficients using studbook data over 5 generations and PEDIGRAPH software (Garbe & Da 2003). For the total cohort of 518 Thoroughbreds, total allele and genotype frequencies were calculated for the two *FKBP6* exon 5 SNPs separately and together. Statistical analysis was carried out using JMP v. 15 (JMP®, Version 15. SAS Institute Inc., Cary, NC). Nominal logistic regression model was used to examine the relationship between fertility data (low fertility vs. high fertility), pedigree data (inbred vs. not inbred), and *FKBP6* exon 5 SNP genotypes, both separate and together. Individuals were considered to have low fertility if their per-cycle pregnancy rate (PCPR) was less than or equal to 46% (Love 2011), and were considered inbred if their pedigree based inbreeding coefficient was greater than 4% which is a strict threshold compared to previous pedigree-based Thoroughbred studies (Fawcett *et al.* 2019). Contingency analysis between pedigree based inbreeding coefficients, fertility data (PCPR) and *FKBP6* exon 5 SNP genotypes were performed to identify relationship between the inbreeding rate of a stallion and the genotype. The results of these combined statistical tests were used to determine whether the phenotype (per-cycle pregnancy rates and fertility) is independent from genotypes and inbreeding levels.

2.2.2 Results*

---

* Reprinted from the published work found in Animal Genetics under the Creative Commons Attribution License: Castaneda C., Juras R., Kjollerstrom J., Hernandez Aviles C., Teague S.R., Love C.C., Cothran E.G., Varner D.D. & Raudsepp T. (2021a) Thoroughbred stallion fertility is significantly associated with FKBP6 genotype but not with inbreeding or the contribution of a leading sire. *Animal Genetics* **52**, 813-23.

Allele and genotype frequencies of the *FKBP6* SNPs

The design of TaqMan™ allelic discrimination assays for the two *FKBP6* exon 5 SNPs were first validated by genotyping stallions whose *FKBP6* genotypes were previously determined by direct Sanger sequencing (Figure 1 in (Castaneda *et al.* 2021a)).

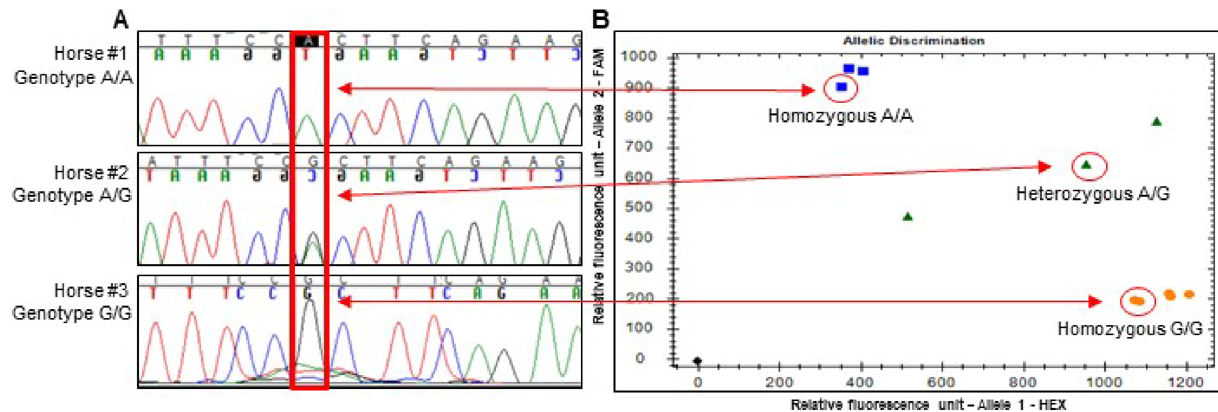


Figure 1. Comparative results of FKBP6 exon 5 SNP 11,353,372G>A genotyping by two methods in three different stallions Horse #1, Horse #2, Horse #3. **(A)** Direct sequencing by Sanger method. The three vertical panels show genotypes of the SNP (red rectangle) in three different stallions; (B) Genotyping results of the same SNP in the same stallions by TaqMan assay; horizontal two-headed arrows indicate correspondence between genotypes determined by Sanger sequencing and TaqMan assay (Castaneda et al. 2021a).

Allelic frequencies for the synonymous SNP G>A were 0.68 for the major allele (G) and 0.32 for the minor allele (A). Most of the individuals had a G/G genotype (47.3%), and the frequency of homozygous A/A genotype was 11.6%. The allelic frequencies for the non-synonymous SNP A>C were 0.61 for the major allele (A) and 0.39 for the minor allele (C). Most of the stallions had a A/C genotype (50.2%) and the homozygous A/A genotype frequency was 36.1%. In this cohort, the most frequent combined genotype for the two SNPs was G/G-A/C which was found in 120 individuals (23.2%) and the least frequent combined SNP genotypes

were A/A-C/C (2.7%). The IAR susceptibility genotype double homozygous A/A-A/A was found in 21 individuals (4.1%) of the 518 Thoroughbred cohort, 18 were males and 3 were females (Table 2; Table 1 in (Castaneda *et al.* 2021a)).

Table 2. Allele, genotype and combined genotype frequencies of the SNPs in FKBP6 exon 5 for the cohort of 518 Thoroughbreds and the chi-square P-value for HWE statistics. n/a – not applicable. Modified Table 1 from (Castaneda et al. 2021a).

SNP G>A synonymous change					
Allele/ Genotype	**Observed allele/genotype frequency**	**No. of observed individuals**	**No. of expected individuals**	**Chi-square *P*-value**	**HWE**
G	0.679	n/a	n/a	n/a	n/a
A	0.321	n/a	n/a	n/a	n/a
GG	0.473	245	238.5	0.24	in HWE
GA	0.411	213	226		
AA	0.116	60	53.5		
SNP A>C non-synonymous change					
Allele/ Genotype	**Observed allele/genotype frequency**	**No. of observed individuals**	**No. of expected individuals**	**Chi-square *P*-value**	**HWE**
A	0.612	n/a	n/a	n/a	n/a
C	0.388	n/a	n/a	n/a	n/a
AA	0.361	187	194	0.43	in HWE
AC	0.502	260	246		
CC	0.137	71	78		
Two-Locus *FKBP6* Genotypes					
Genotype	**Observed allele/genotype frequency**	**No. of observed individuals**	**No. of expected individuals**	**Chi-square *P*-value**	**HWE**
A/A-A/A	0.041	21	20.0	0.03	not in HWE
A/A-A/C	0.048	25	25.4		
A/A-C/C	0.027	14	8.1		
A/G-A/A	0.124	64	84.6		
A/G-A/C	0.222	115	107.3		
A/G-C/C	0.066	34	34.0		
G/G-A/A	0.197	102	89.3		

G/G-A/C	0.232	120	113.3
G/G-C/C	0.044	23	35.9

Chi-square testing of genotype frequencies using SNP1 and SNP2 individually, did not deviate from HWE in this cohort; however, the two-locus genotype significantly deviated from HWE (P = 0.03) when using a full chi-square model.

Analysis of fertility data and *FKBP6* genotypes in 150 breeding stallions

Utilizing PCPR as a measurement of stallion fertility, 13 out of 150 stallions were considered to have low fertility rates because their PCPR were less than the threshold of 46% (Love 2011). We used both a contingency analysis and Student's t-test to study the relationship between PCPR and genotypes of the two *FKBP6* exon 5 SNPs separately and as a two-locus genotype (Figure 2 in (Castaneda *et al.* 2021a)).

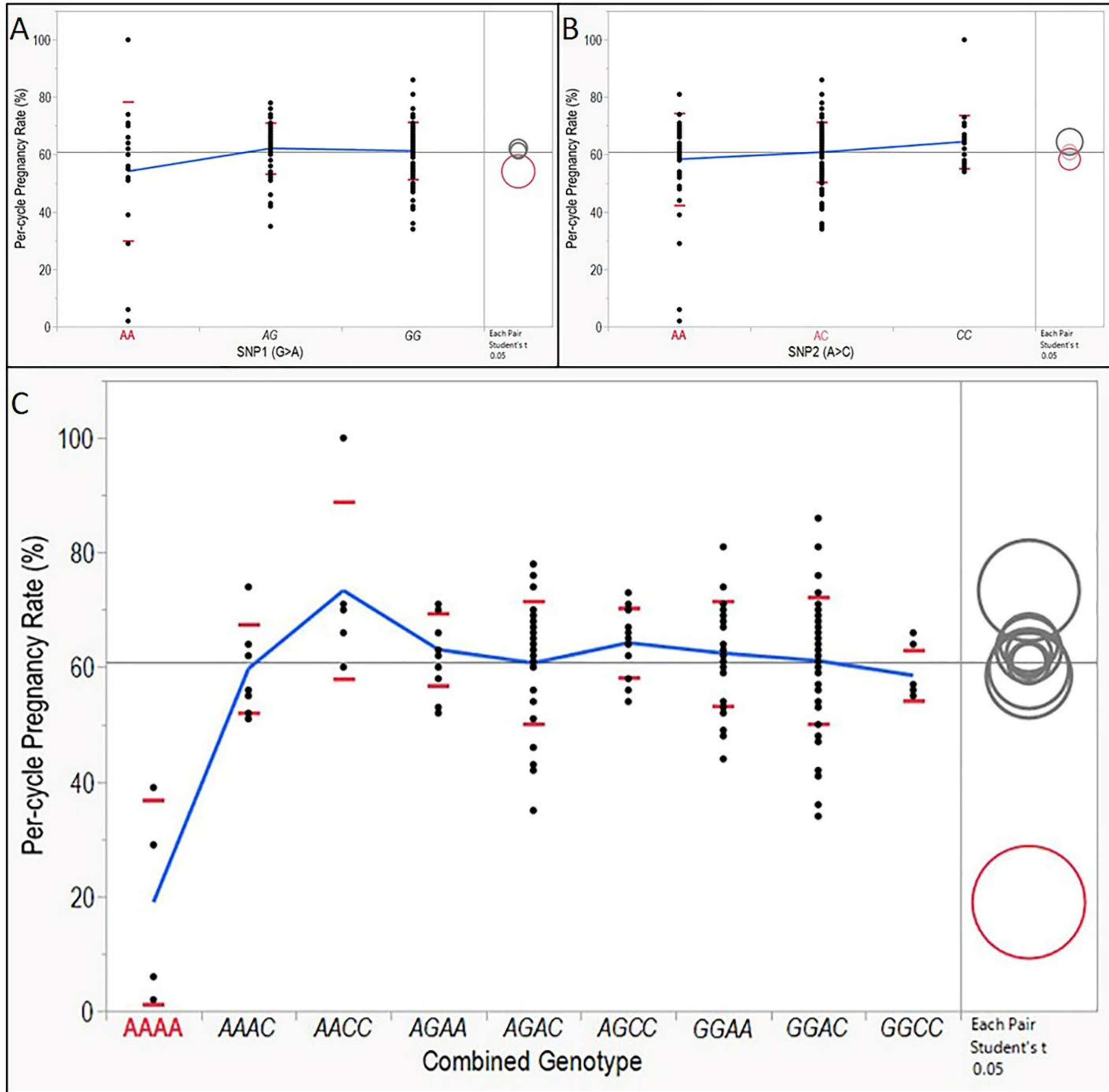


Figure 2. The distribution of per-cycle pregnancy rates (%; y-axis) of 150 stallions in relation to their genotypes for A: SNP G>A, B: SNP A>C, and C: genotype combinations (x-axis). Each dot corresponds to an individual stallion; the horizontal grey line denotes the total mean value for PCPR; the blue line denotes the means of each genotype group and connects the means of each group; the red lines represent the upper and lower standard error bars for each genotype. Student's t test is represented by circles, each circle represents a genotype and its *P*-value in

comparison to each genotype. The A/A genotype for SNP1 and SNP2 as well as the double homozygous A/A-A/A genotypes are in red bold font. The corresponding A/A and double homozygous A/A-A/A circles are also highlighted in red. In B, the genotypes A/A and A/C are in red indicating that they are not significantly different from each other as demonstrated by the *P*-value in Table 2 (Castaneda *et al.* 2021a).

PCPRs were significantly different between A/A and A/G, as well as between A/A and G/G genotypes for SNP G>A. For SNP A>C, a significant difference was only documented between the A/A and C/C genotypes. In the two-locus *FKBP6* genotype analysis, the statistically most significant relationship ($P < 0.0001$) was found between PCPR and the combined double homozygous genotype A/A-A/A (Table 2 in (Castaneda *et al.* 2021a)). On an average, stallions with the double homozygous A/A-A/A genotype had significantly lower average PCPR and per-season pregnancy rates (PSPR) compared to stallions with other genotypes. On average, their PCPR was 19% (ranging from 2 to 36%) and PSPR 31% (ranging from 2 to 60%), compared to an average of 62% PCPR (ranging from 53 to 71%) and an average of 88% PSPR (ranging from 81 to 95%) in stallions with other genotypes (Supplementary Table S2). Finally, we used Fisher's exact statistics to study the 13 stallions with PCPR lower than 46%. Here, we found that there was a significant association between the low fertility and the individual's combined genotype ($P = 0.0007$). On an individual SNP level, the low fertility status was not significantly associated with SNP G>A ($P = 0.0971$) or SNP A>C ($P = 0.1995$) (Supplementary Table S3).

<u>Analysis of per-cycle pregnancy rates, inbreeding, and *FKBP6* genotypes</u>

Finally, we conducted a nominal logistical regression analysis to determine if *FKBP6* exon 5 genotype or pedigree-based inbreeding level had a significant relationship with whether

the stallion had low fertility (PCPR < 46%). As already mentioned, 13 individuals were considered to have low fertility. Inbreeding rate was based on pedigree analysis and a horse was considered inbred if the inbreeding coefficient was greater than 4% (n=2; Supplementary Table S2). Based on the effect likelihood ratio test within the Nominal Logistical Fit analysis in JMP Pro 15, the only statistically significant association was between low fertility and the *FKBP6* genotype (P = 0.002). Lastly, we conducted contingency analysis for inbreeding vs. two-locus *FKBP6* genotype and inbreeding vs. low fertility, using a range of inbreeding levels from 1% to 5%, but did not detect any statistically significant associations.

2.2.3 Discussion[*]

This section discusses a follow up study to an earlier genome-wide association study (GWAS) which originally identified *FKBP6* as a susceptibility locus for stallion subfertility due to IAR (Raudsepp *et al.* 2012). Here, we controlled for a fertility phenotype by acquiring sound fertility data such as per-cycle and per-season pregnancy rates (PCPR and PSPR), and the number of mares bred in 150 Thoroughbred breeding stallions. While all three parameters are important, genotype-phenotype associations were based on PCPR as it is a more accurate representation of stallion fertility than PSPR and the number of mares bred (Brinsko *et al.* 2007; Love 2011).

---

[*] Reprinted from the published work found in Animal Genetics under the Creative Commons Attribution License: Castaneda C., Juras R., Kjollerstrom J., Hernandez Aviles C., Teague S.R., Love C.C., Cothran E.G., Varner D.D. & Raudsepp T. (2021a) Thoroughbred stallion fertility is significantly associated with FKBP6 genotype but not with inbreeding or the contribution of a leading sire. *Animal Genetics* **52**, 813-23.

Despite using different criteria for determining the phenotype and low fertility rates in this study (PCPR < 46%) and the earlier analysis (IAR; (Raudsepp *et al.* 2012)), the results were consistent. Both studies showed that the highest statistically significant association is between *FKBP6* exon 5 combined SNP genotype A/A-A/A and low fertility in stallions ($P < 0.0001$, both studies), regardless of how the low fertility was defined. However, it must be noted that in the cohort of 150 stallions with available fertility data, only 4 out of 13 stallions with low fertility carried the A/A-A/A genotype (Supplementary Table S2), indicating that the genotype is associated with just a fraction of low fertility phenotypes in stallions. Conversely, among the 518 Thoroughbred horses involved in this study, 18 males carried the A/A-A/A genotype and 15 (83.3%) of these had low fertility and the remaining 3 males had no fertility information available (Supplementary Table S1). This observation strongly supporting the observed genotype-phenotype association in our previous (Raudsepp *et al.* 2012) and the present study (Castaneda *et al.* 2021a).

A similar study in 216 Hanoverian stallions did not detect any significant interaction effect between the two *FKBP6* exon 5 SNPs and did not find a significant association between stallion fertility and the double homozygous A/A-A/A genotype (Schrimpf *et al.* 2015). Instead, that study showed significant association of the non-synonymous SNP A>C, with improved conception rates in Hanoverian stallions. One can argue that the two studies were not directly comparable because the standard breeding practices in the two breeds are different and there was a difference in fertility evaluation parameters. However, a more plausible explanation for the contrasting differences between the two studies is that the two SNPs in *FKBP6* exon 5 are not causative for stallion fertility or subfertility. It is more likely that the two SNPs are tagging a larger haplotype, a different one in different breeds, carrying causative structural or regulatory

variant(s) affecting the phenotype. Further details about this observation and supporting material will be discussed in the next section.

Besides confirming significant association of *FKBP6* exon 5 combined SNP genotype A/A-A/A with low fertility in Thoroughbred stallions, we investigated whether this genotype and low fertility was influenced by additional factors such as inbreeding. Inbreeding in Thoroughbreds has been proposed to decrease foaling rates (Todd *et al.* 2020). Despite this, fertility status and *FKBP6* genotype in our study cohort were independent from pedigree-based inbreeding rate (Supplementary Table S2). For example, stallions FK161_JM and FK072_DV127 with the highest inbreeding coefficients 5.08% and 4.88%, respectively, had normal fertility rates and no IAR susceptibility genotype (Supplementary Table S2). In contrast, the four stallions with the A/A-A/A genotype and low fertility rates had low inbreeding values (0 to 0.391%) (Supplementary Table S2). Combined analysis of low fertility rates, genotype, and inbreeding contribution, as well as comparison tests of genotype and low fertility rates against different inbreeding levels, further strengthened our conclusion that among all considered factors, significant association is only between the *FKBP6* genotype and low fertility rates in Thoroughbred stallions.

An important outcome of this study was successful development of a reliable and cost-effective molecular diagnostic test using TaqMan allelic discrimination assays for genotyping. This study is currently the most extensive genetic analysis of *FKBP6* and provides compelling evidence for significant association between low fertility rates in Thoroughbred stallions and the combined A/A-A/A genotype of SNPs chr13:11,353,372G>A and chr13:11,353,436A>C in *FKBP6* exon 5, thus confirming and refining earlier findings by GWAS (Raudsepp *et al.* 2012). Further studies are needed for two reasons, firstly, because the associated sequence variants are

not causative, and secondly, because the *FKBP6* genotype is associated with stallion subfertility only in Thoroughbreds but not in other horse breeds.

Since this study, groundwork for a detailed whole genome sequence-based analysis of Thoroughbreds, Hanoverians, and individuals of other breeds with the combined A/A-A/A genotype has begun and will be discussed in the next section (2.3). To understand the associated genotype to phenotype relationship from a functional perspective, elect Thoroughbreds were subject for testis RNA-sequencing and will be discussed in section 2.4.

2.3. Accessing the relationship of *FKBP6* and the IAR using whole genome sequencing

The following sections will focus on the materials and methods, results, and discussion about achieving the Objective 2 mentioned in 2.1.

2.3.1 Experimental methods and design

Samples

For whole genome sequencing (WGS), gDNA of 10 male horses (n=9 Thoroughbreds; n=1 Friesian) with the A/A-A/A *FKBP6* two-locus genotype were obtained from the Molecular Cytogenetics and Animal Genetics repository. Of the 9 Thoroughbreds sequenced, one sample (H054) was used in the initial discovery GWAS study (Raudsepp *et al.* 2012), and one sample (H815) had a slightly different phenotype than the other Thoroughbreds sequenced for this study. In addition to harboring the A/A-A/A *FKBP6* genotype and low per-cycle pregnancy rates, H815 sperm analysis suggested a certain degree of testicular dysfunction (unpublished data; (Castaneda *et al.* 2021a)). The 9 case Thoroughbreds used in this study were identified either by

diagnosis of subfertility due to an impaired acrosome reaction, by having poor per-cycle pregnancy rates, or by large cohort *FKBP6* genotyping (Raudsepp *et al.* 2012; Castaneda *et al.* 2021a). The common factor between all cases is that they carry the susceptibility A/A-A/A *FKBP6* genotype. Additionally, this study includes a Friesian (H510) because it is the only other breed which carried the A/A-A/A *FKBP6* genotype found within the Molecular Cytogenetics and Animal Genetics repository. Sample H510 is used as a control to determine if the haplotype block associated with the A/A-A/A *FKBP6* genotype and IAR phenotype is specific to the Thoroughbred breed. Library preparations and sequencing were performed at Texas A&M Molecular Genomics Workspace and each individual was sequenced to 30x coverage using the HiSeq Illumina 2x150 bp paired-end platform. Table 3 outlines the individuals used for WGS analysis.

Table 3. Sample identification, genotype, fertility status and breed of individuals used for whole genome sequencing.

Sample	Genotype	Fertility Status	Breed
H302	A/A-A/A	Low fertility	Thoroughbred
H650	A/A-A/A	Subfertile	Thoroughbred
H652	A/A-A/A	Subfertile	Thoroughbred
H698	A/A-A/A	Subfertile	Thoroughbred
H815	A/A-A/A	Low fertility	Thoroughbred
H860	A/A-A/A	Low fertility	Thoroughbred
H875	A/A-A/A	Low fertility	Thoroughbred
H940	A/A-A/A	n/a	Thoroughbred
H510	A/A-A/A	Control	Friesian
H054	A/A-A/A	Subfertile due to IAR	Thoroughbred

Equine WGS alignment and variant database for mutation discovery

In order to generate a large equine variant database (EVD) for a case-control analysis, Illumina paired-end WGS data from our unpublished work and those publicly available through NCBI's Sequence Read Archive (SRA) database were assembled against EquCab3 (Kalbfleisch *et al.* 2018) using SpeedSeq0.1.2. Only WGS data with greater than 10X coverage was obtained from the SRA database to use in the EVD. In total, 428 individuals from over 30 domestic horse breeds comprise the EVD. Once assembled and aligned, single nucleotide variants and small indels were called with GATK HaplotypeCaller v 4.1.4.1 (Poplin *et al.* 2017) for each chromosome. Joint calling across all 428 individuals was performed using the genotypegVCF function of GATK GenomicsDB for each chromosome (Poplin *et al.* 2017). For each variant, allele frequencies were calculated using VCFTools v 0.1.16 (Danecek *et al.* 2011) and predictive genetic variant effects were obtained using SnpEff v 5.1 (Cingolani *et al.* 2012). The bioinformatics were performed on the Texas A&M High Performance Computing Cluster, with subsequent computation on private, dedicated servers in Dr. Brian Davis' lab.

Identity by decent analysis, candidate SNP identification, and conservation score analysis

To determine the extent of the non-recombining haplotype within cases, 8447 biallelic SNPs and indels were visually inspected for recombination patterns between the case Thoroughbreds, the Friesian with the A/A-A/A *FKBP6* genotype, and control EVD individuals. Plots of haplotype blocks were visualized using the GenotypePlot v 0.2.1 (Whiting 2022) function in R. Variants associated or specific to the case Thoroughbreds were identified based on two criteria: first, if all case Thoroughbreds contained a homozygous alternate allele or indel and second, if a SNP had an allele frequency less than 20% when calculated in the total cohort of 428 horses (419 EVD horses + 9 case Thoroughbreds). PhyloP conservation scores from were

obtained from UCSC Genome Browser for variants which met both criteria to determine how conserved a base pair site is across mammalian species (Rhead *et al.* 2010). PhyloP scored greater than 1 denote a conserved site, and those less than one denotes a variable site across species.

2.3.2 Results

A 171 Kbp non-recombining haplotype block in IAR Thoroughbreds

Identity by decent (IBD) analysis was conducted between the 9 case Thoroughbreds sequenced for this study and the remaining 419 individuals in the EVD. We determined a 171 Kbp region (EquCab3: ECA13: 11,279,424 – 11,450,501) encompassing a total of 8447 variants which is non-recombining in the case Thoroughbred stallions (Figure 3). One of the Thoroughbreds in the large cohort of 419 individuals contained the *FKBP6* IAR-associated genotype (Sample ID: TB_EAV003; SRA ID: SRR2103372). This sample was used in a study which focused on identifying genetic factors responsible for the equine arteritis virus, and not for a fertility study (see BioProject ID SRR2103372).

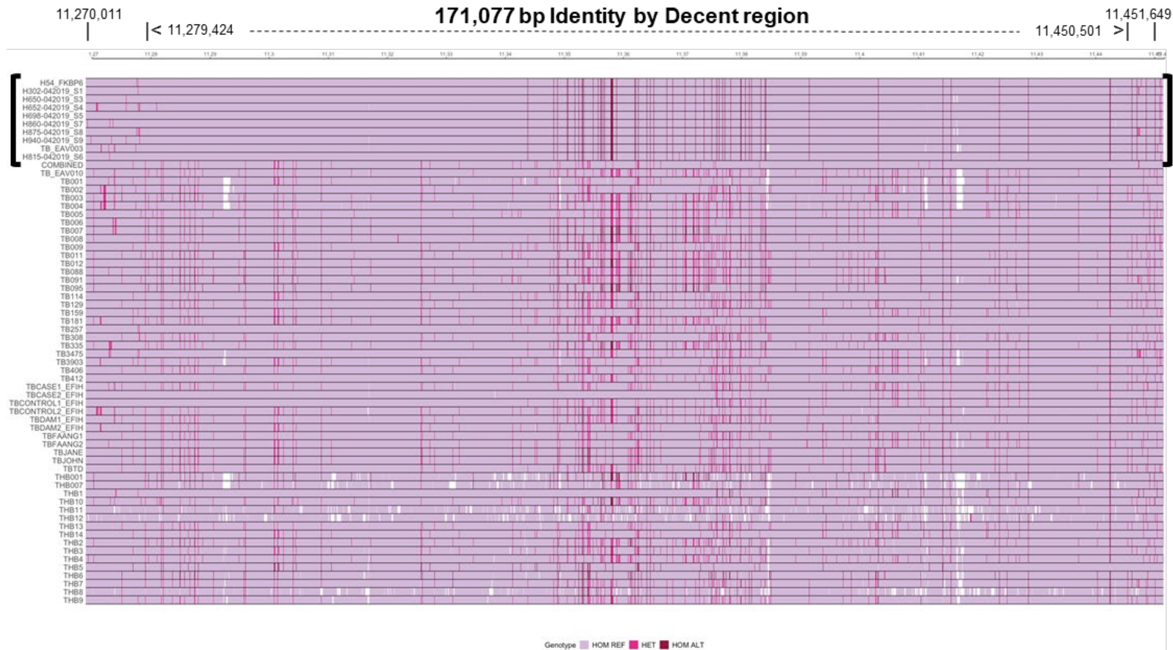


Figure 3. Genotype Plot of a 182 Kbp region depicting a 171 Kbp identity by decent region specific to case Thoroughbreds with the A/A-A/A FKBP6 genotype. Each lane shows sequence variants in individual Thoroughbreds: 9 case Thoroughbreds (top rows; brackets) and 55 Thoroughbreds from EVD. Pink coloring denotes a homozygous reference genotype, hot pink represents a heterozygous genotype, and purple represents a homozygous alternate allele genotype.

<u>Putative causal SNPs identified within the IBD area of IAR Thoroughbred cases</u>

A total of 38 putative SNPs were found in 5 genes: *POM121C*, *FKBP6*, *TRIM50*, *BAZ1B*, and *HIP1*. Of the 38 candidate SNPs, only one was exonic: SNP chr13:11,345,375C>T (Variant #3165). Variant #3165 produces a missense mutation at the 19th residue of *TRIM50* (Valine to Methionine), has a total allele frequency of 11.6% in the EVD, and conservation score of -0.48.

Another notable variant is chr13:11,302,833C>T (Variant #1004) in the intronic portion of *POM121C*. Variant #1004 was found to have the lowest total allele frequency in the EVD

with 6% and had a positive conservation score of 0.21. Interestingly, one of the case stallions sequenced (Sample ID H815) is heterozygous for this variant while the remaining case Thoroughbreds are the only homozygous individuals within the EVD. This difference may be because H815 has a different subfertility phenotype than the remaining case Thoroughbred males.

Additionally, three variants had a conservation score greater than 1, signifying that the SNPs are constrained based on the UCSC Genomic Evolutionary Rate Profiling method (threshold range: -4.5 to 7.5). Two of these variants were in the intronic region of the *FKBP6* gene: SNP chr13:11,373,240G>A (Variant #4464) and SNP chr13:11,374,173C>A (Variant #4548). Variant #4464 has a total allele frequency of 11.4% and conservation score of 1.272, and Variant #4548 had allele frequency of 12% and a conservation score of 2.425. The variant with the highest conservation score of 4.135 was within the large intron of *HIP1* (chr13:11,385,148G>A), with allele frequency of 10.6%. The remaining SNPs are highlighted in Table 4.

Table 4. Detailed information on the 38 putative IAR-associated variants identified within the 171 Kbp IBD block. Detailed information on the 38 putative IAR-associated variants identified within the 171 Kbp IBD block. Position in ECA13 is based on EquCab3 horse genome reference.

Variant #	ECA13 Position	Variant Type	ALT Allele Frequency	Fixed Case Genotype	Gene	Location	PhyloP Score
195	11285221	SNP	0.135	T/T	*POM121C*	Intron	n/a
1004	11302833	SNP	0.067	T/T	*POM121C*	Intron	0.21
3165	11345375	SNP	0.116	T/T	*TRIM50*	Exon	-0.48
3624	11352564	SNP	0.132	C/C	*FKBP6*	Intron	0.651
3737	11356059	SNP	0.158	T/T	*FKBP6*	Intron	n/a
3741	11356104	SNP	0.172	C/C	*FKBP6*	Intron	-0.54
3773	11357232	SNP	0.153	A/A	*FKBP6*	Intron	0.072
4199	11368611	SNP	0.159	G/G	*FKBP6*	Intron	n/a
4434	11373067	SNP	0.112	A/A	*FKBP6*	Intron	-2.53
4464	11373240	SNP	0.114	A/A	*FKBP6*	Intron	1.272
4465	11373251	SNP	0.118	T/T	*FKBP6*	Intron	-1.764

4548	11374173	SNP	0.120	A/A	*FKBP6*	Intron	2.425
4676	11376135	SNP	0.108	T/T	*FKBP6*	Intron	n/a
4755	11379096	SNP	0.116	T/T	n/a	n/a	n/a
4799	11380257	SNP	0.106	C/C	n/a	n/a	-0.246
4844	11381489	SNP	0.108	T/T	n/a	n/a	-0.281
4853	11381937	SNP	0.098	T/T	n/a	n/a	-4.803
4923	11384733	SNP	0.144	A/A	*HIP1*	Intron	0.03
4964	11385148	SNP	0.106	A/A	*HIP1*	Intron	4.135
4996	11385428	SNP	0.141	T/T	*HIP1*	Intron	0.544
5082	11387853	SNP	0.105	G/G	n/a	n/a	n/a
5353	11393497	Indel	0.179	G/G	*HIP1*	Intron	n/a
5500	11395467	SNP	0.193	G/G	*HIP1*	Intron	n/a
5528	11395738	SNP	0.172	T/T	*HIP1*	Intron	n/a
5647	11397744	SNP	0.182	A/A	*HIP1*	Intron	n/a
5701	11399250	SNP	0.179	C/C	*HIP1*	Intron	n/a
5735	11400582	SNP	0.189	G/G	*HIP1*	Intron	n/a
5841	11403674	SNP	0.178	A/A	*HIP1*	Intron	0.33
5889	11404804	SNP	0.169	G/G	*HIP1*	Intron	0.441
5909	11405431	SNP	0.164	A/A	*HIP1*	Intron	-0.077
6014	11408706	SNP	0.146	G/G	*HIP1*	Intron	n/a
6067	11410808	SNP	0.100	T/T	n/a	n/a	n/a
6384	11420095	Indel	0.178	T/T	*HIP1*	Intron	n/a
6545	11424126	SNP	0.151	C/C	*HIP1*	Intron	n/a
6546	11424134	Indel	0.151	G/G	*HIP1*	Intron	n/a
8064	11444802	SNP	0.137	G/G	*BAZ1B*	Intron	n/a
8244	11447372	SNP	0.137	A/A	*BAZ1B*	Intron	0.64
8402	11450501	SNP	0.143	G/G	*BAZ1B*	Intron	n/a

ALT = Alternate allele

Comparing individuals with the *FKBP6* A/A-A/A genotype

While identifying the 171 Kbp IBD region in the 9 case Thoroughbreds, it became clear that Sample H510, the Friesian, which was sequenced for this project, did not share the 171 Kbp IBD region despite harboring the double homozygous A/A-A/A *FKBP6* genotype. Therefore, we compared the case Thoroughbreds to other individuals of different breeds in the EVD who shared the double homozygous A/A-A/A genotype (n=21; 20 horses (case Thoroughbreds included), and 1 Donkey). As indicated by the 171 Kbp IBD region, it was expected that the 9 case Thoroughbreds would have identical genetic sequences around the target *FKBP6* SNPs. However, the remaining individuals with the double homozygous A/A-A/A genotype were genetically different, as indicated by the dendrogram of a 110 Kbp region in Figure 4.

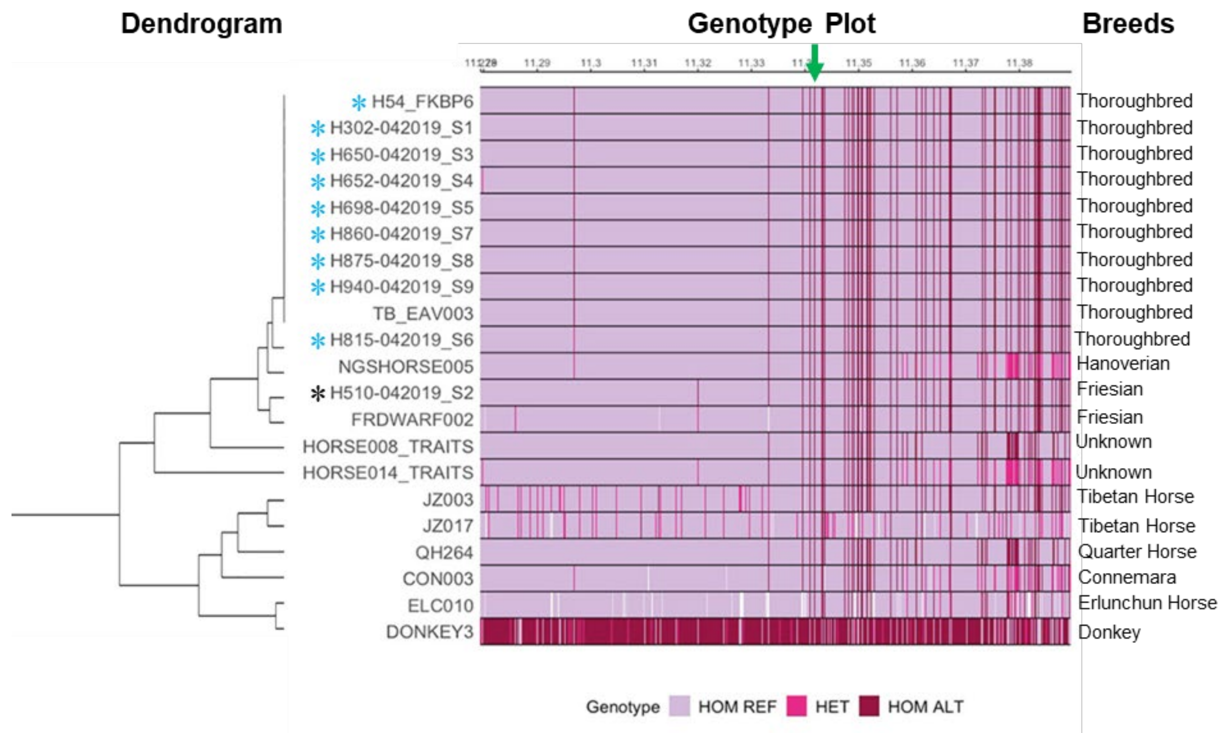


Figure 4. A 110 Kbp region surrounding the FKBP6 tagging SNPs in 21 individuals with the double homozygous A/A-A/A FKBP6 genotype. A 110 Kbp region surrounding the FKBP6 tagging SNPs in 21 individuals with the double homozygous A/A-A/A FKBP6 genotype. **Left**: a dendrogram representation of the 110 Kbp region (chr13:11,279,424 – 11389895) depicting the sequence similarity of the 9 case Thoroughbred cases compared to the remaining individuals. **Middle**: Genotype Plot of the same region; genotype coloring patterns are the same as in Figure 3. **Right**: Breed details of the individuals with the A/A-A/A *FKBP6* genotype. Individuals sequenced for this study are denoted by asterisks. Blue asterisks represent the 9 case Thoroughbreds, and the black asterisk represents the Friesian sequenced. The green arrow indicates the location of the tagging *FKBP6* SNPs within the genotype plot.

2.3.3. Discussion

Genome wide association study (GWAS) is an effective method for identifying the chromosomal location of disease genes in horses (Raudsepp *et al.* 2019). However, GWAS is often inadequate for identifying disease alleles with a frequency of less than 5% (minor allele frequency, MAF <0.05) because rare variation is not well represented on SNP genotyping arrays (Hoglund *et al.* 2019). These challenges and limitations fully apply to the GWAS that implicated *FKBP6* as a susceptibility gene for IAR (Raudsepp *et al.* 2012). Despite the strong genotype-phenotype association, the study failed to reveal the causative functional mutation in *FKBP6* or whether mutations in other genes are involved. Complementary to GWAS, whole-genome sequencing (WGS) enables highly efficient allele discovery and elucidation of the nature of all genetic variation and greatly improves the power and precision of identifying causal/associated disease variants, small insertions and deletions, and structural variations (Hoglund *et al.* 2019; Mahmoud *et al.* 2019).

To use WGS for effective discovery of DNA sequence variants that underlie specific phenotypes in horses, a large-scale variant catalog of common and rare genetic variants including single SNPs, insertions, deletions, inversions, and structural variants is needed (Mu *et al.* 2011). Such a catalogue was recently used to discover rare and common variants in Thoroughbred racehorse (Tozaki *et al.* 2021), and a large genetic variation catalog for horses was recently published utilizing 543 individuals (Durward-Akhurst *et al.* 2021). For this study, we generated a similar Equine Variant Database (EVD) catalog of 428 individuals utilizing unpublished and public WGSs of individual horses, the latter being available through the NCBI Sequence Read Archive (SRA: https://www.ncbi.nlm.nih.gov/sra). To increase the sample size and statistical power of this study, our case cohort consisted of IAR-confirmed individuals from both the

original 2012 (Raudsepp *et al.* 2012) study, and those identified through low per-cycle pregnancy rates and TaqMan genotyping (Castaneda *et al.* 2021a).

Our goal was to use WGS to identify breed-specific haplotype blocks around the *FKBP6* locus and other genomic signatures unique to Thoroughbred stallions with confirmed IAR, idiopathic subfertility, or low per-cycle pregnancy rate. As shown by the previous GWAS, haplotypes in this region are different even between Thoroughbred cases and controls: all IAR cases in the 2012 study have a large 3.31 Mb haplotype block which is broken down in controls, and within that haplotype block there is a smaller single haplotype which is highly associated with the IAR phenotype (Raudsepp et al. 2012). Through WGS, we refined the haplotype highly associated with the IAR phenotype through identity by decent (IBD) analysis of the 9 case Thoroughbreds with the IAR-susceptibility genotype compared to the the remaining horses in the EVD as controls.

Within the 171 Kbp IBD region identified, a total of 38 candidate variants were identified by targeting biallelic SNPs or indel variants with an allele frequency less than 20% and were homozygous across cases. Unfortunately, most candidate variants were located within noncoding regions. The two most promising variants (Variant #1004 and Variant #4946) were not within the *FKBP6* gene, but within introns of neighboring genes *POM121C* and *HIP1*. *POM121C* (POM121 transmembrane nucleoporin C) is a nuclear porin which is not well annotated in the horse with no known functional. However, a recent review highlights the emerging roles and effects that nucleoporins have in oogenesis and spermatogenesis using *Drosophila* and mouse models (reviewed by (Preston *et al.* 2019). The same can be said for *HIP1* (Huntingtin Interacting Protein 1), for which there is no functional annotation in the horse, but mouse *HIP1* knock out models suggest that the gene may have functions in spermatid development and post

meiotic spermatid viability (Rao *et al.* 2001; Khatchadourian *et al.* 2006). It is striking that of the 38 variants detected, only one is in an exon of *TRIM50* (Tripartite motif containing 50), however the resulting missense mutation does not cause an apparent change in the protein structure (Supplementary Figure S1).

It is possible that some of the candidate variants detected within the IBD region, despite being in gene introns or intergenic regions, are functionally important and affect regulatory elements. Since the human Encyclopedia of DNA Elements (ENCODE) project was initiated in 2003 (Consortium 2004), the knowledge and importance of noncoding variants, their association with transcription factors/regulatory elements, and involvement in complex traits and diseases has increased exponentially (Consortium *et al.* 2020; Yan *et al.* 2021). Such public tools are now being constructed for agricultural animals (Andersson *et al.* 2015; Giuffra *et al.* 2019), including the horse. Previously, equine FAANG models were focused on two female Thoroughbreds (Kingsley *et al.* 2019), however work is ongoing to generate functional analysis on male specific tissues (Donnelly *et al.* 2021). Therefore, testis tissue expression and regulatory elements are currently lacking. On the other hand, it is equally possible that the causal variant for the IAR phenotype is outside of the IBD region but affects candidate IAR-associated genes identified in this study and the initial GWAS study in 2012.

In addition to identifying candidate variants in a large haplotype block specific to the case Thoroughbreds, using the EVD as the control cohort allowed for a direct comparison of individuals with the IAR-associated A/A-A/A genotype across breed groups, specifically in the Hanoverian. This is of significant interest because of contrasting associations: in Thoroughbreds it is considered as a susceptibility gene for stallion subfertility owing to an impaired acrosome reaction (IAR) (Raudsepp *et al.* 2012) while in Hanoverians, it is associated with improved

conception rates (Schrimpf *et al.* 2015). Because there is no similar association between the IAR-associated A/A-A/A genotype and stallion subfertility in Hanoverians, it is suggested that the two SNPs are tagging a breed-specific haplotype with genetic variants unique to Thoroughbreds (Castaneda *et al.* 2021a). The IBD sequence differences between sample H510 (Friesian) and the remaining sequenced case Thoroughbreds was the first indication that the case Thoroughbreds harbor a different haplotype than other males with the A/A-A/A *FKBP6* genotype. Sequence differences in the 171 Kbp IBD region of the 9 case Thoroughbreds and other individuals with the A/A-A/A *FKBP6* genotype in the EVD, which includes one Hanoverian male (sample ID: NGSHOORSE005) and supports the breed-specific haplotype hypothesis (see Figure 4).

2.4. Accessing the relationship of *FKBP6* and the IAR through testis RNA-sequencing

The following sections will focus on the next phase of *FKBP6* research: using RNA-sequencing analysis to study the transcriptome of *FKBP6* and other genes of interest with the 171 Kbp IBD region in Thoroughbred stallions with the IAR-associated A./A-A/A genotype.

2.4.1. Experimental methods and design.

<u>Samples and RNA Isolation</u>

For testis RNA-sequencing, we chose two Thoroughbred males with A/A-A/A genotype in *FKBP6* and with confirmed IAR phenotype, along with one Thoroughbred male per each of other possible combined *FKBP6* genotypes: A/G-A/A; G/G-A/A; G/G-A/C; A/A-A/C; A/A-C/C; A/G-C/C; A/G-A/C; G/G-C/C, thus a total of 10 individuals. Samples were obtained from the

repository at the Texas A&M Molecular Cytogenetics Laboratory. Table 5 represents the individuals utilized for testis RNA-sequencing and their *FKBP6* genotypes.

Table 5. Sample identification, genotype, and breed type of individuals used for RNA sequencing.

Sample	Genotype	Breed
H054	A/A-A/A	Thoroughbred
H860	A/A-A/A	Thoroughbred
H449	A/A-A/C	Thoroughbred
H383	A/A-C/C	Thoroughbred
H542	A/G-A/A	Thoroughbred
H469	A/G-A/C	Thoroughbred
H472	A/G-C/C	Thoroughbred
H435	G/G-A/A	Thoroughbred
H438	G/G-A/C	Thoroughbred
H630	G/G-C/C	Thoroughbred

Total RNA was extracted from testis tissue using TRIzol Reagent (Invitrogen Life Technologies) according to the manufacturer's protocol. RNA quality and quantity was evaluated with TapeStation (Agilent Technologies) and Qubit fluorimeter with Qubit RNS HS Assay Kit (Life Technologies).

RNA-sequencing and Analysis

For each sample, mRNA was separated from the total RNA by PolyA selection and individually barcoded. Illumina cDNA libraries were produced and sequenced using 2x300 base-pair read lengths on the Illumina MiSeq. RNA-sequencing (RNASeq) library preparations and sequencing were done at Texas A&M Molecular Genomics Workspace. Testis transcriptome was assembled and aligned with EquCab3 using HiSat2v2.2.1 (Kim *et al.* 2019) and the genotypes for each sample were visually confirmed using IGV (Integrated Genomics Viewer) (Robinson *et al.* 2011). To increase the control non-A/A-A/A *FKBP6* genotype sample size, individuals

sequenced for this study were compared to other horse testis transcriptomes available through SRA (n=13; non-A/A-A/A testis samples), creating a total sample size of n=23. Specific focus was on the region around the *FKBP6* gene, comparing the transcriptomic data between A/A-A/A cases and Thoroughbred controls with other *FKBP6* genotypes. Special interest was on the tagging *FKBP6* SNPs to either confirm or refute the monoallelic cDNA amplification patterns identified in the original study (Raudsepp *et al.* 2012). Tagging *FKBP6* SNPs were visually inspected for monoallelic bias in IGV using the allelic depth (AD) per nucleotide measurement (Robinson *et al.* 2011).

2.4.2. Results

Small sample size presents analysis difficulties

It was expected that obtaining a statistically powerful sample size for case Thoroughbred stallions with the IAR phenotype would be problematic. Samples from stallions with the IAR phenotype or IAR-associated *FKBP6* genotype are difficult to collect for two reasons. One, because testis samples from breeding stallions are unobtainable unless they are gelded and two, a testis biopsy is can cause health risks for the animal. The small case sample size for this study (n=2) creates a large margin of error when comparing RNAseq data of case Thoroughbreds to control Thoroughbreds. Unexpectedly, two testis RNA samples were contaminated during the sequencing process, including one of the samples with the A/A-A/A *FKBP6* genotype, lowering our case sample size to n=1. The small sample size limited the power of the transcriptome analysis between the single case Thoroughbred and remaining control Thoroughbreds. The two samples which were contaminated (H054 and H435; Table 5) will be re-sequenced for future

analysis, however obtaining testis samples for RNAseq from additional Thoroughbreds with a confirmed IAR phenotype remains difficult.

Evidence supports biallelic expression of *FKBP6* SNPs

The original GWAS study in 2012 suggested monoallelic expression pattern of the *FKBP6* gene (Raudsepp *et al.* 2012). Using transcriptome data of individuals who were heterozygous for one or both *FKBP6* SNPs in this study refuted the monoallelic expression previously identified through Sanger Sequencing (Raudsepp *et al.* 2012). Visual inspection of individual RNAseq data in IGV and comparison of AD measurements for each *FKBP6* allele supported biallelic expression of the two *FKBP6* SNPs, with no allelic bias (Table 6). Though, it must be noted that the testis control samples in this study were not the same individuals as those used for allelic expression in the original study (see figure 5 and Figure S7 in (Raudsepp *et al.* 2012).

Table 6. Individual testis RNA-sequencing FKBP6 SNP genotype and cooresponding allelic depth parameters per SNP. Samples in red font were contaminated and not analyzed further.

Sample	Exp. Genotype	Obs. Genotype	SNP1 AD (G, A)	SNP2 AD (A, C)
H860	AA-AA	AA-AA	0, 982	0, 1044
H449	AA-AC	AA-AC	0, 7	3, 2
H383	AA-CC	AA-CC	0, 1430	0, 1417
H542	AG-AA	AG-AA	722, 729	1541, 0
H469	AG-AC	AG-AC	44, 41	41, 62
H472	AG-CC	AG-CC	82, 59	0, 123
H438	GG-AC	GG-AC	76, 0	43, 32
H630	GG-CC	GG-CC	1354, 0	1398, 0
H054	AA-AA	n/a	n/a	n/a

H435	GG-AA	n/a	n/a	n/a

Exp. = Expected; Obs. = Observed; AD = Alleleic depth.

2.4.3 Discussion

RNA-sequencing is beneficial for investigating the transcriptome and connect genomic sequence with functional annotation. Genomic studies of Thoroughbreds with the IAR-susceptibility genotype in *FKBP6* (see 2.3) identified select genes within a 171 Kbp IBD region. The testis transcriptome study by RNA-sequencing focused on the transcription of these genes in cases and controls. Previous studies identified that *FKBP6* transcription is specific to testis and sperm (Raudsepp *et al.* 2012), therefore RNA was isolated from two confirmed IAR samples and 8 controls. Analyzing the testis transcriptome between case Thoroughbreds with the IAR-associated A/A-A/A *FKBP6* genotype against control Thoroughbreds without the IAR-associated genotype will allow us to identify differences between the two transcriptomes which may be associated with the IAR phenotype. Additionally, by using testis from individuals who are heterozygous for the "A" allele at one or both of the *FKBP6* SNPs will confirm or refute previous findings of monoallelic expression within the *FKBP6* exon 5 SNPs (Raudsepp *et al.* 2012). Unfortunately, functional annotation of the horse genome is not complete, and the relevance of transcriptome variants identified within candidate genes will remain speculative.

One notable observation is that there does not appear to be any allelic bias in the *FKBP6* tagging SNPs. In the original GWAS study, cDNA for heterozygous control animals (genotype G/A-A/C) showed only the G allele at SNP1 G>A and the C allele at SNP2 A>C which suggested monoallelic expression patterns of *FKBP6* (Raudsepp *et al.* 2012). However, this study utilized 11 samples which were heterozygous at one or both *FKBP6* SNPs and both alleles were clearly represented in the RNA-sequencing data. It is likely that the monoallelic expression

observed in the original study is an artifact because attempts to recreate the experiment with the same testis cDNA samples did not show monoallelic expression (our unpublished data). Furthermore, in humans, *FKBP6* (alias *FKBP36*) has shown to be a gene affected by imprinting and allele-specific expression in early embryonic tissues, but biallelic expression patterns in adult testis tissues (Strogantsev *et al.* 2015).

One major hurdle for this phase of the project was the low number of testis samples available with the IAR-associated *FKBP6* A/A-A/A genotype. This is due partly because only samples available from the Molecular Cytogenetics repository were utilized and partly due to the rarity of the A/A-A/A genotype, which was identified in 21 out of 428 (4.9%) individual, multi-breed cohort (see section 2.3). Another limitation is that the IAR has been identified as a cause for stallion subfertility only in the Thoroughbred (Brinsko *et al.* 2007) and Friesian (Hernandez-Aviles et al. 2022) breed, this creates a breed restriction., but the IAR-susceptibility genotype in *FKBP6* is specific to Thoroughbreds and found only in 4% of the Thoroughbred population (Castaneda *et al.* 2021a). However, there are even less individuals identified through routine diagnostic tests (our unpublished data) as most stallions are gelded and not used for breeding purposes. Finally, an additional factor is that if an A/A-A/A stallion is identified at the Molecular Cytogenetics laboratory, it is up to the owner's discretion on if testis samples can be obtained. Due to the limitations of sample acquisition, it is imperative to re-sequence sample H054 to improve the testis transcriptome analysis of individuals with the IAR-susceptibility genotype.

2.5. Conclusions and future studies for the *FKBP6* gene and the IAR

Stallion subfertility associated with acrosomal dysfunction was first reported in the late 1990s (Meyers *et al.* 1995; Meyers *et al.* 1996). Stallions with acrosomal dysfunction have

normal seminal parameters (i.e., normal sperm morphology, motility, and quantity) but consistently low pregnancy rates when compared to fertile stallions (Kenney *et al.* 1983; Varner *et al.* 2001). As routine breeding soundness exams do not include acrosome reaction tests, stallions with possible IAR remain undetected and are classified as idiopathic subfertile (Varner *et al.* 2000; Brinsko *et al.* 2007). Due to the ambiguity of the IAR phenotype, and the absence of routine acrosomal reaction analysis in the breeding soundness exams, an important outcome of this study was the development of a reliable and feasible diagnostic Taqman test (Castaneda *et al.* 2021a). The newly established diagnostic test efficiently determines if the stallion has the IAR-susceptibility genotype in *FKBP6* (see Figure 1) and allows for large cohort genotyping to identify additional case Thoroughbred stallions for future studies. In addition to creating an effective diagnostic test for the IAR-susceptibility genotype, Castaneda *et al* (2021a) broadened the IAR phenotype of affected Thoroughbreds by associated the IAR-susceptibility genotype with low per-cycle pregnancy rates.

Based on the data presented by Raudsepp et al in the original GWAS study, it was proposed that the IAR displays incomplete penetrance (i.e., some individuals will not express the phenotype despite carrying the allele) (Raudsepp *et al.* 2012). Further support for this theory is the large range of fertility parameters in stallions identified with the A/A-A/A genotype in the follow up study performed in 2021 (Castaneda *et al.* 2021a). In a later study, stallions with IAR-susceptibility genotype A/A-A/A had a PCPR range from 2 to 36% and PCSR from 2 to 60%, suggesting that in some affected stallions a large amount of sperm is capable of performing a normal acrosome reaction (Castaneda *et al.* 2021a). Hernandez-Aviles et al (2022) clinically confirmed this phenomena, indicated that IAR affected stallions have two phenotypes for the sperm: one where very little sperm can capacitate and have “very low to low” PCPR scores and

the second where a moderate amount of sperm successfully undergo the AR and have “low to moderate” PCPR scores.

While the genetic cause of the IAR phenotype remains elusive, we are taking important steps to narrow down the associated region around *FKBP6* by detailed variant analysis between cases and controls using WGS and the EVD established for this project. It has been speculated that the IAR-associated *FKBP6* SNPs are “tagging” a specific IAR haplotype (Castaneda *et al.* 2021a). With the EVD we were able to identify a 171 Kbp haplotype specific to subfertile Thoroughbreds with the IAR-susceptibility genotype and identified additional putative IAR-associated SNPs (see Table 4) in 5 genes surrounding *FKBP6*. We will continue to work on refining and identifying variants in and around the 171 Kbp region by generating long read PacBio HiFi sequencing data for IAR affected Thoroughbreds. This will allow the detection of large structural variants that were not revealed by short read Illumina sequencing. One the other hand, transcriptome analysis of *FKBP6* and other candidate genes in the region in IAR cases remains preliminary until the procurement of more testis samples from individuals with that IAR-susceptibility genotype. However, gene expression analysis through quantitative PCR of the genes found within the 171 Kbp region can provide insight by identifying genes with an increased or decreased expression profile in IAR stallions. Unfortunately, this type of analysis still requires additional case samples to reach a statistically significant power but will be beneficial if used in conjunction with testis RNA-sequencing data.

The studies involving the IAR-susceptibility *FKBP6* genotype outlined in this chapter result in more questions than answers, as the genetic cause of the IAR phenotype remains undefined. However, these studies (see sections 2.3 and 2.4) generated important conclusions and steppingstones for future studies involving the IAR. First, Castaneda *et al* (2021) confirmed

the previous association by Raudsepp *et al* (2012) by showing that there is an association between low per-cycle pregnancy rates and the IAR-susceptibility *FKBP6* genotype in Thoroughbreds. Second, the later study generated a necessary tool for wide scale genotyping which will aid in identifying additional individuals with the *FKBP6* A/A-A/A genotype. Third, by integrating WGS of case Thoroughbreds with the IAR-susceptibility genotype a 171 Kbp haplotype block was identified to be specific to cases. The haplotype block encompasses several candidate variants and genes possibly associated with the IAR phenotype. Finally, analysis of the testis transcriptome in Thoroughbreds that were heterozygous for the "A" allele in one or both *FKBP6* SNPs refuted the previously speculation that *FKBP6* showed monoallelic expression patterns in horses (Raudsepp *et al.* 2012).

3. GENOMICS OF MALE FERTILITY WITHIN THE CONTEXT OF THE Y CHROMOSOME*

3.1 Objectives

The goals of the research on the Y chromosome and the role it plays in stallion fertility will focus on the improving the current equine Y assembly (eMSYv3) and identifying natural MSY sequence variations between horse populations and related equids. Additionally, we aim to analyze MSY CNVs in a subset of horses with disorders of sexual development (DSDs) and/or subfertility. Finally, we cytogenetically characterizing a rare Y-autosome translocation in a stallion with azoospermia to demonstrate the effect the Y chromosome on fertility.

Objective 1: Determine horse Y chromosome sequence variation across horse breeds, related equids and select cohorts of subfertile/infertile individuals utilizing droplet digital PCR (ddPCR) for absolute quantification of gene copy numbers and compare gene copy number variation with MSY haplotypes (HTs).

Objective 2: Improve the current Y reference sequence eMSYv3 (Janecka *et al.* 2018) utilizing sequenced contigs not found in eMSYv3 (Felkel *et al.* 2019) to identify new Y specific bacterial artificial chromosomes (BACs), and by resequencing BACs of the Y multi-copy region (mcY) with Nanopore technology.

Objective 3: Cytogenetically characterize the first Y-autosome translocation in an azoospermic Friesian stallion.

---

* This chapter contains sections reprinted from MDPI-GENES under Creative Commons CC BY 4.0 license: Castaneda C., Ruiz A.J., Tibary A. & Raudsepp T. (2021b) Molecular Cytogenetic and Y Copy Number Analysis of a Reciprocal ECAY-ECA13 Translocation in a Stallion with Complete Meiotic Arrest. *Genes* **12**.

3.2. Copy number variation in horse Y chromosome genes

The following section will focus on the development of Y copy number (CN) ddPCR assays for horse multi-copy genes to determine the range of MSY CNVs in a global equine population and some wild equids. Once the baseline for equine MSY "natural variation" is established, we compare MSY CNV patterns with MSY haplotypes for correlation. Finally, we evaluate MSY gene CNs in a group of abnormal males to identify CNVs associated with the phenotypes.

3.2.1 Experimental methods and design

Animals and samples

Genomic DNA (gDNA) samples of 289 male horses and equids were available from the repositories of Molecular Cytogenetics and Animal Genetics Laboratories at Texas A&M University and Institute of Animal Breeding and Genetics at the Veterinary University of Vienna. The samples included a normal control cohort of 216 male equids: 209 domestic horses (*Equus caballus*) of 22 breeds or breed mixes, 5 Przewalski's horses (*Equus caballus przewalskii*) and 2 kulans (*Equus hemionus kulan*) (Supplementary Table S4) Additionally, we used gDNA from 73 abnormal male horses: 24 cryptorchid Quarter Horses (6 bilateral cryptorchids; 18 unilateral cryptorchids), 29 horses of ambiguous sex with confirmed *SRY*-positive or *SRY*-negative 64,XY disorders of sex development (DSDs), 12 males with heterogeneous subfertility phenotypes, and 8 males within two families produced by somatic cell nuclear transfer (SCNT) (Supplementary Table S5).

Droplet-digital PCR (ddPCR) analysis

Droplet digital PCR assays were designed and optimized and ddPCR reactions were conducted as previously described (Castaneda *et al.* 2021b). Briefly: ddPCR assays were designed for 7 horse MSY multi-copy genes, single-copy *SRY*, and single-copy autosomal control genes *MYOZ1* and/or *MSTN* (Table 7).

Table 7. Droplet digital PCR assays for all Y chromosome genes analyzed in this study. Information includes the forward and reverse primer sequences, probe sequences, product size in DNA base pairs, the optimized restriction enzyme used with the gene, the location of the assays in the gene, and the expected number of copies based on the MSY reference eMSYv3 by Janecka *et al* 2018.

Gene name	**Forward Primer**	**Reverse Primer**	**Probe**	**Product size (bp)**	**Restriction enzyme**	**Location**	**Expected # of copies**
ETSTY1	GACGGACGACCTTGTGTT	ACGCTCACAGATGACAGTAG	TGTCCCGGCCACCTCAGGGC	166	NSPI	Exon 1	3
ETSTY2	TTGTTGTTAGGCTACCTGGC	AAGGGCAAACCATAACCTCC	TGGGCAAGCTTCTCCATGGTTGCTGCA	106	ECORI	Exon 1	7
ETSTY5	GAGGCAGGTACTTCGTTACC	TCACTCACAAAGTCAACGCT	TGCCGTGAGCTTGAGGGCGAA	216	NSPI	Exon 1	8
HSFY	AGGCTTTCTCCACTGGTTTC	GAGGCTGTCCCGAACTTTTA	CCCCTGCTCTAAAGTGCTTCCTGTCG	169	ECORI	Exon 1	3
MYOZ1	GACTTTCCAGATGCCCAAGT	ACCAGAACCTCTCCAACAGGCCTTCT	GCTCCTCTGTTTCTCCATCC	182	ECORI/NSPI	chr 1	2
MSTN	ACTGCGCCTGGAAACAGCTCC	TGTTTCCGTCGTCGCGTGGT	CCCAAAGCTCCTCCACTCCGGG	163	ECORI/NSPI	chr 18	2

RBMY	GAAGCTCCA CAACTTGAG GT	CTCTGACCT ATGATGGAA GCA	TGTCTGCCAC CATGCTCACG ACCA	214	ECORI	Exon 1	2
SRY	TTCTGTGAT CTATGCTGG CG	TTACCCTCC GGACTTTCT CA	AACAGGGACT CTGCCGCCAC CA	225	ECORI	Exon 1	1
TSPY	CATAGTGGA GGAAGAGG ATGAAA	GGCAATGGT TTAACCCTG AAA	CTCTTTCTGGG AGACCTGCCC TTT	75	NSPI	Intron 4	13
UBA1Y	TTTCTGTTG TCTGGACGG AG	CTCCACGGA TGTAGTCAG AG	AGCAGAGGCC TCCTGTGTCTG AGCT	218	ECORI	Exon 7	8

Primers were designed with Primer3 software (Untergasser *et al.* 2012) using reference sequences for the horse MSY (Janecka *et al.* 2018) and EquCab3 (Kalbfleisch *et al.* 2018) so that the size of PCR products was in the range of 75-200 bp. Fluorescently labeled (FAM for MSY genes, VIC for autosomal *MYOZ1* and *MSTN*) hydrolysis probes (TaqMan) were designed with PrimerQuest™ tool (Integrated DNA Technologies). The template gDNA was cleaved with EcoRI (Invitrogen) or NspI (New England Biolabs) restriction enzymes into < 5 kb fragments to fit into individual droplets. The restriction enzyme chosen for the experiment was dependent on the MSY gene sequence. The ddPCR reactions were carried out on C1000Touch (Bio-Rad) platform in 25 µL volume containing (final concentration) 1 X ddPCR Supermix for Probes no-UTP, 10 µM forward and reverse primers for an MSY gene and the control gene, 250 nM TaqMan probe for an MSY gene and the control gene, one of the two restriction enzymes (diluted 1:1 in water), and 1-10 ng of undigested gDNA as a template. Droplets were generated using the QX200™ (Bio-Rad) automated droplet generator and manufacturer's protocol. Cycling parameters were carried out using the recommended protocol for performing genomic enzymatic digestion during the PCR experiment. The PCR plate was transferred to QX200™ (Bio-Rad) droplet reader and the data were analyzed using the associated Quantasoft software. The results

were presented as number of copies per µL of the final 1 X ddPCR reaction. The male control sample, (Thoroughbred stallion *Bravo* - the DNA donor for the horse eMSYv3 assembly) (Janecka *et al.* 2018), the female control, (Thoroughbred mare *Twilight* - the DNA donor for the horse reference genome) (Kalbfleisch *et al.* 2018), and a water control were present in all experiments. Any sample with questionable CN results (i.e., high standard error, low droplet generation, or noticeably low or high CN) were subject to retesting.

<u>Copy number variation (CNV) statistical analysis</u>

Statistical analysis of gDNA copy number variations (CNVs) between various cohorts was carried out using JMP v 15 (JMP®, Version 15. SAS Institute Inc., Cary, NC). Oneway ANOVA was used to generate F-statistic P-values to determine if there is statistically significant CN variation within the 216-male cohort when divided by breed or Y-HT. Similar methods were used to compare the MSY gene CNs of 24 cryptorchid Quarter Horses with those of the 28 Quarter Horses in the normal cohort (Supplementary Table S2). *P*-values were not generated for the analysis of the 64,XY DSD cohort or the subfertile cohort. Instead, these individuals were compared to their corresponding breed group in the normal, control cohort to identify outstanding CNVs potentially associated with the subfertility or DSD phenotype.

<u>MSY genotyping</u>

We inferred MSY haplotypes (HTs) of 216 male equids. For genotyping, we selected 30 HT determining variants from the previously described horse Y phylogeny described (Felkel *et al.* 2019). Information about the variant markers, 29 single nucleotide variants (SNVs) and one short indel, are given in Supplementary Table S6. The variant markers chosen created a

condensed horse Y-HT tree, which served as a backbone for the HT analysis performed in this study (Figure 5). Genomic DNA (gDNA) was diluted with TE to a concentration of 5 ng/µl. For genotyping, competitive allele-specific PCR SNV genotyping assays (KASP™, lgcgroup.com) were used. KASP™ genotyping was performed on a CFX96 Touch® Real-Time PCR machine (BioRad) using the standard KASP™ genotyping protocol (lgcgroup.com). Each run included samples with their allelic state known as positive controls, while DNA from females and non-template controls were used as negative controls. Raw data were analysed with Bio-Rad CFX Manager 3.1® software (BioRad).

Genotyping was conducted sequentially, following the hierarchical backbone tree. First, we determined whether samples belong to the Crown haplogroup (HG) by genotyping the Crown determining variant rAX. If a sample carried the derived allele [C], which indicates that it belongs to the Crown, clustering of the sample into HGs T, A, and H was performed by testing variants rA, rW, and fYR. Based on the outcome, we genotyped the sample for the variants informative for the substructure of the HGs they cluster into. For samples carrying the ancestral allele for rAX [T], we genotyped 14 variants that determine the HGs outside the Crown. For HT reconstruction, the information of the 30 markers were concatenated and allelic states of markers not tested, were imputed according to the HTs previously defined (Felkel *et al.* 2019) (Supplementary Table S7). We constructed a haplotype frequency plot with draw.io platform (diagrams.net, 14.6.13). The phylogenetic relationships in the plot were based on MSY tree from (Felkel *et al.* 2019), and the circle radiuses were scaled to the respective number of samples with RStudio 4.0.3. (RStudio Team, 2020).

3.2.2 Results

Horse MSY gene copy number assays

We aimed to design ddPCR assays for all 15 multi-copy genes, which are annotated in the current horse MSY assembly eMSYv3 (Janecka *et al.* 2018). However, following assay requirements (Digital Droplet PCR Application Guide, BioRad) and MSY sequence properties, we were able to design assays for only 9 multi-copy genes. Of these, the assays for two autosomal transposed genes - *HTRA3Y* and *SH3TC1Y*, were not male specific and were not used for CN analysis. All in all, we succeeded to design and optimize male-specific ddPCR assays for only 7 MSY multi-copy genes. These included four amplified gametologs - *TSPY*, *RBMY*, *HSFY,* and *UBA1Y*, and three novel Y-born testis-specific transcripts - *ETSTY1*, *ETSTY2*, and *ETSTY5* (Janecka et al. 2018). In addition, ddPCR assay was successfully designed for the single-copy *SRY*. Detailed information about the ddPCR assays used in this study is presented in Table 6 (see section 3.2.1).

Comparison of gene copy numbers between ddPCR results and the MSY reference assembly

As the first step, we determined CNs of 7 multi-copy genes and *SRY* in a multi-breed cohort of 209 normal male horses and compared the results with the CNs in horse MSY reference assembly eMSYv3 (Janecka *et al.* 2018). It is important to note that the male control sample for every ddPCR experiment, a Thoroughbred stallion *Bravo*, was also the DNA donor for the MSY reference, thus allowing direct comparison between the MSY reference CN and those generated through ddPCR. As the male control was the same throughout the study, *Bravo's* gene CNs were averaged across 30 ddPCR experiments. For most genes, *Bravo's* CNs determined by ddPCR were notably different from those in the MSY reference (Table 8). Five genes (*TSPY*, *ETSTY2*, *ETSTY5*, *HSFY*, and *UBA1Y*) had almost half less copies by ddPCR

compared to eMSYv3, while *ETSTY1* had 5 copies by ddPCR compared to 3 copies in eMSYv3. Only two genes, *SRY* (CN=1) and *RBMY* (CN=2) showed consistent CN between ddPCR and MSY reference.

Table 8. Comparison of gene copy numbers between eMSYv3 (Janecka et al. 2018) and ddPCR analysis of the reference male Bravo and a multi-breed cohort.

	ETSTY1	*ETSTY2*	*ETSTY5*	*HSFY*	*RBMY*	*SRY*	*TSPY*	*UBA1Y*
CN eMSYv3 (Janecka *et al.* 2018)	3	7	8	3	2	1	13	8
CN reference *Bravo;* ddPCR	5	4	4	1	2	1	8	3
209 cohort average CN	5	5	4	1	2	1	10	4
209 cohort minimum CN	1.96	2.8	1.88	0.22	0.6	0.46	5.55	1.12
209 cohort maximum CN	10.8	14	29	3	2.93	2.7	38	12.3
209 cohort standard deviation	1.09	1.47	2.25	0.27	0.40	0.32	4.03	1.01

Similar disparities were observed when gene CNs of eMSYv3 were compared with averaged gene CNs determined by ddPCR in a large multi-breed cohort of 209 male horses including *Bravo* (Table 8). Like when comparing the CNs of *Bravo's* sequence assembly with *Bravo's* ddPCR, the only genes with concordant CNs between eMSYv3 and multi-breed cohort were *SRY* (CN=1) and *RBMY* (CN=2). Again, CN of *ETSTY1* was higher in the large cohort and CNs of the remaining five genes (*TSPY*, *ETSTY2*, *ETSTY5*, *HSFY*, and *UBA1Y*) were almost half less than in MSY reference (Table 8). At the same time, ddPCR-determined average gene CNs of *Bravo* and the 209-male cohort were identical for *ETSTY1*, *ETSTY5*, *SRY*, *RBMY*, and *HSFY* and very similar for *TSPY* (8 vs. 10), *ETSTY2* (4 vs. 5), and *UBA1Y* (3 vs. 4). Due to an outlier (Yakutian; TR028; Supplementary Table S4) with exceptionally high CNs of multiple genes, the

largest and smallest CNs per gene varied in a broad range, with *TSPY* having the largest range (from 5.5 to 28 copies) (Figure 5, Table 8; Supplementary Table S1). However, the outlier did not affect the overall average CN in the population and was, therefore, not excluded from analysis. Regardless whether the outlier TR028 was included or not, *TSPY* remained the most variable multi-copy gene tested in this study (Figure 5).

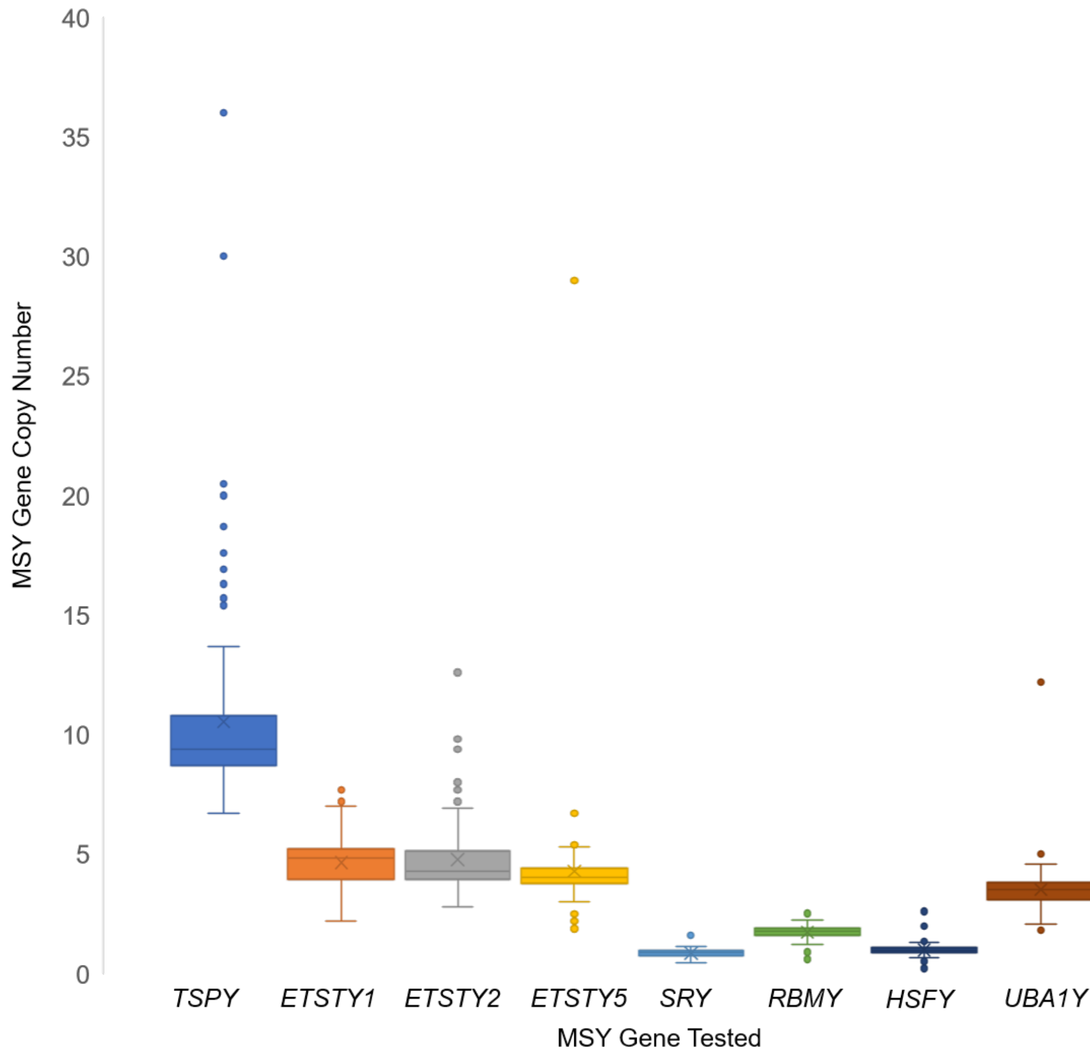


Figure 5. Box and whisker plot illustrating the range of copy number variation of horse MSY multi-copy genes and SRY in a multi-breed cohort of 209 normal male horses.

MSY gene CNVs across horse breeds and related equids

The 209 normal male horse cohort comprised of 22 breeds or breed mixes allowing for the comparison of MSY gene CNs across breeds (Table 9). Samples per breed ranged from a single individual (Friesian, American Paint, and Quarter Horse-Morgan mix) to 47 individuals for Thoroughbreds. The second most represented breeds were the Estonian Native horse and American Quarter Horse with 29 and 28 individuals respectively. For each breed group, we calculated the average CN along with standard deviation and generated F-statistic *P*-values to determine if CN variation between breeds was significant (Table 9). We found statistically significant ($P < 0.001$) CN differences between breeds for multi-copy genes *ETSTY1*, *ETSTY2*, *RBMY*, and *TSPY* and the single-copy *SRY*. However, the studied horse breeds did not significantly differ for *ETSTY5*, *HSFY* and *UBA1Y* CNs (Table 9).

Similarly, to gene CN variation between individuals in the large male cohort (Table 9), *TSPY* was also the most variable gene between breeds having a minimum of 5.55 copies in the Friesian and a maximum of 20.5 copies in the Tennessee Walking horse. The least variable MSY gene across breeds was *HSFY* (Table 9).

Table 9. Average copy number (CN) and corresponding standard deviation (SD) of 7 MSY ampliconic genes and SRY across horse breeds and related equids. CNs of the Przewalski's horse and kulan are presented separately as an outgroup.

Horse breed	**N**	***ETSTY1*** **CN (SD)**	***ETSTY2*** **CN (SD)**	***ETSTY5*** **CN (SD)**	***HSFY*** **CN (SD)**	***RBMY*** **CN (SD)**	***SRY*** **CN (SD)**	***TSPY*** **CN (SD)**	***UBA1Y*** **CN (SD)**
American Paint	1	3.60 (na)	4.48 (n/a)	4.26 (n/a)	1.04 (n/a)	1.77 (n/a)	0.90 (n/a)	9.80 (n/a)	3.48 (n/a)
Arabian	12	4.49 (0.77)	4.33 (1.09)	4.10 (0.37)	1.03 (0.17)	1.85 (0.20)	0.90 (0.09)	10.03 (3.43)	2.94 (0.87)

Caspian	6	4.73 (0.48)	4.53 (0.69)	4.08 (0.21)	0.97 (0.07)	1.88 (0.21)	0.84 (0.12)	9.67 (1.17)	3.57 (0.55)
Dales Pony	4	5.30 (1.58)	6.35 (0.79)	4.80 (0.51)	0.79 (0.22)	1.58 (0.49)	0.89 (0.16)	8.28 (0.59)	3.48 (0.40)
Estonian Native	28	4.49 (1.26)	5.93 (2.48)	5.76 (5.90)	1.15 (0.48)	2.15 (0.43)	1.21 (0.46)	12.98 (7.04)	4.09 (2.39)
Friesian	1	1.96 (na)	3.78 (n/a)	2.62 (n/a)	0.85 (n/a)	1.91 (n/a)	0.78 (n/a)	5.55 (n/a)	3.13 (n/a)
Haflinger	2	4.12 (0.13)	3.73 (0.06)	4.07 (0.02)	0.99 (0.02)	1.99 (0.16)	0.96 (0.07)	9.25 (0.35)	4.11 (0.23)
Icelandic	6	4.83 (0.26)	3.94 (0.47)	4.05 (0.23)	1.01 (0.08)	1.77 (0.25)	0.90 (0.15)	9.48 (0.86)	3.71 (0.37)
Lipizzan	10	5.10 (0.36)	3.85 (0.28)	3.90 (0.09)	0.95 (0.09)	1.91 (0.17)	0.96 (0.03)	8.73 (0.31)	3.99 (0.22)
Miniature	2	3.71 (0.08)	3.91 (0.01)	3.94 (0.34)	0.82 (0.01)	1.63 (0.17)	0.86 (0.08)	8.95 (0.49)	3.45 (0.05)
Mongolian	10	4.46 (0.54)	4.16 (0.30)	3.79 (0.23)	0.96 (0.11)	1.67 (0.55)	1.26 (0.51)	8.47 (0.84)	3.74 (0.34)
Noriker	4	4.91 (0.65)	3.92 (0.12)	3.89 (0.05)	0.98 (0.07)	1.92 (0.11)	0.98 (0.06)	8.68 (0.49)	4.00 (0.08)
Quarter Horse	28	4.64 (0.75)	4.59 (0.85)	4.21 (0.49)	1.00 (0.15)	1.79 (0.23)	0.83 (0.11)	10.66 (2.62)	3.43 (0.52)
Quarter Horse-Morgan mix	1	3.59 (na)	4.50 (n/a)	4.12 (n/a)	0.98 (n/a)	1.79 (n/a)	0.93 (n/a)	10.00 (n/a)	3.51 (n/a)
Shetland pony	5	5.10 (0.20)	4.25 (0.69)	4.21 (0.44)	0.94 (0.05)	1.73 (0.43)	0.80 (0.27)	8.37 (0.50)	3.80 (n/a)
Standardbred	8	5.08 (1.04)	5.11 (1.00)	4.62 (0.94)	1.11 (0.17)	1.41 (0.27)	0.61 (0.08)	9.00 (1.16)	2.98 (0.24)
Suffolk Punch	8	3.89 (0.17)	4.50 (0.31)	4.12 (0.24)	1.02 (0.11)	1.87 (0.13)	0.82 (0.09)	9.32 (1.19)	3.44 (0.32)
Heck horse (Heck 1952)	7	5.86 (1.26)	6.17 (0.81)	4.99 (0.40)	1.10 (0.08)	1.34 (0.46)	0.74 (0.11)	15.36 (2.64)	3.56 (0.45)
Tennessee Walking	5	5.60 (1.11)	6.02 (1.04)	4.61 (0.36)	1.07 (0.15)	1.62 (0.18)	0.79 (0.22)	16.92 (2.59)	3.56 (0.64)
Thoroughbred	47	4.49 (0.91)	4.35 (0.75)	3.74 (0.62)	0.91 (0.27)	1.79 (0.21)	0.87 (0.14)	9.35 (1.19)	3.50 (0.63)
Yakutian	4	5.97 (3.37)	6.41 (5.07)	5.58 (3.42)	1.22 (0.65)	1.78 (0.31)	1.40 (0.87)	15.72 (14.88)	3.59 (0.52)
Zemaitukai	10	5.39 (0.41)	5.02 (0.76)	4.70 (0.32)	0.92 (0.08)	2.32 (0.54)	1.27 (0.42)	11.62 (1.09)	3.62 (0.28)
***P*-value**	**209**	**0.0003**	**< 0.0001**	**0.64**	**0.19**	**< 0.0001**	**< 0.0001**	**< 0.0001**	**0.59**

Difference between Max and Min CN		4.01	2.68	3.14	0.43	0.98	0.79	11.37	1.17
Equid species	**N**	***ETSTY1***	***ETSTY2***	***ETSTY5***	***HSFY***	***RBMY***	***SRY***	***TSPY***	***UBA1Y***
Przewalski's horse	5	3.65 (1.85)	3.18 (0.40)	4.11 (0.31)	1.02 (0.15)	1.10 (0.23)	0.93 (0.14)	8.40 (0.77)	3.04 (0.58)
Kulan	2	2.29 (0.25)	4.10 (0.85)	3.50 (0.85)	1.69 (0.03)	0.82 (0.09)	0.75 (0.13)	n/a	n/a
Horse average from Table 1	209	5	5	4	1	2	1	10	4

N – number of individuals; Numbers in green font denote the highest and numbers in red font denote the lowest CN value per gene.

Interestingly, *SRY* was a single-copy gene in most breeds and individuals used in this study (Table 9; Supplementary Table S4), though we identified 21 individuals from 4 indigenous breeds (Estonian Native horse, Mongolian, Yakutian and Zemaitukai) with 2 or 3 copies of *SRY* (Table 10; Supplementary Table S1). Most of these 21 individuals had also an increased number of *RBMY* copies (CN=3), resulting in significant ($P < 0.0001$) *SRY* and *RBMY* CN differences between breed groups (Table 9). However, 3 of the 21 males had a decreased *RBMY* CN=1 (Table 10).

Table 10. Copy number variation of SRY and RBMY in 21 males across 4 indigenous breeds.

Breed	**Horse ID**	***SRY* CN**	***RBMY* CN**
Estonian Native	BP364	1.96	2.81
Estonian Native	BP378*	2	3
Estonian Native	BP379	1.74	2.685
Estonian Native	BP383	1.88	2.92
Estonian Native	BP384	1.49	2.66

Estonian Native	BP385	1.9	2.5
Estonian Native	BP386*	1.58	2.35
Estonian Native	BP387*	1.77	2.6
Estonian Native	BP388*	1.55	2.85
Estonian Native	BP395*	1.6	2.52
Estonian Native	BP399*	1.76	2.53
Estonian Native	BP400	1.98	2.61
Mongolian	BP298	1.99	2.785
Mongolian	TR020	1.87	1.06
Mongolian	TR021	1.95	1.01
Yakutian	TR028	2.7	1.37
Zemaitukai	121576*	1.7	2.8
Zemaitukai	121579	1.48	2.84
Zemaitukai	121581	1.66	2.77
Zemaitukai	121587	1.58	2.79
Zemaitukai	121589	1.89	2.93

* Confirmed fertile breeding stallions

In addition to the breeds of the domestic horse (*Equus caballus*), we used the optimized ddPCR assays for MSY gene CN analysis in two other equid species – the Przewalski's horse (*Equus przewaskii*) - a closely related caballine to the domestic horse, and kulan (*Equus hemionus kulan*) – an equid from the ass/onager group (Table 9). Assays for all 8 genes worked in the Przewalski's horse and showed CN similar to or lower than domestic horse averages (Table 9). In the kulan, ddPCR results were obtained for 6 genes, while the assays for equine *TSPY* and *UBA1Y* did not work in kulan (Table 9), likely due to MSY sequence divergence. Both wild equids, like most domestic horses, had a single copy of the *SRY* gene.

MSY haplotype analysis

We genotyped 30 MSY polymorphic markers (Supplementary Table S6) in 209 normal male horses, 5 Przewalski's horses and 2 kulans (outgroup) and assigned individuals to haplogroups (HG), haplotypes (HT) and sub-haplotypes according to Felkel *et al.* (2019). The 209 domestic horses and 5 Przewalski's horses separated into 20 HTs and 36 sub-haplotypes (Figure 6; Supplementary Table S7) and the kulans, as expected, formed an outgroup. We assigned 190 horses (including the Przewalski's Horse) to 14 previously defined HTs and 24 males were placed into internal nodes of the backbone topology DW1 (Domestic West 1), DW3, DW4, Tb, and Tb-1 (Figure 6). This inner clustering of samples is due to unresolved HTs with unknown private SNVs.

Most domestic horses (80%; 168/209) clustered into the Crown HG (Figure 6, Supplementary Table S7). Within the Crown HG, we distinguished 10 HTs (including two not fully ascertained HTs). The most represented HT was Tb-d (n=87) which enclosed 45 of 46 analyzed Thoroughbreds, but also 23 Quarter Horses, 8 Standardbreds, 5 Tennessee Walking horses, 3 Caspian ponies, 1 Paint, 1 Quarter-Morgan mix and 1 Estonian Native horse. The next abundant HTs were Ad and Tb-o with 24 and 21 horses, respectively. Crown HT Tu comprised of 8 Estonian Native horses and Hs of 4 Lipizzans. A single Caspian horse had Am and a single Zemaitukai horse had Ta HTs. Details about individuals, breeds and corresponding HGs, HTs and sub-haplotypes are presented in Supplementary Table S7. Out of 209 genotyped horses, 50 (24%) carried the ancestral allele rAX variant (Felkel *et al.* 2019), were placed outside the large Crown HG and categorized as "Non-Crown". Nordic breeds belonged to I and N HTs, and Asian horses grouped into O, M and Y HTs. Six Mongolian horses could not be attributed to any

ascertained HTs and were clustered basally into DW1 (n=1) and DW3 (n=5). Finally, 11 Estonian Native and 5 Zemaitukai horses clustered into DW4 having a derived allele at rAY and the ancestral allele at rAX (Figure 6).

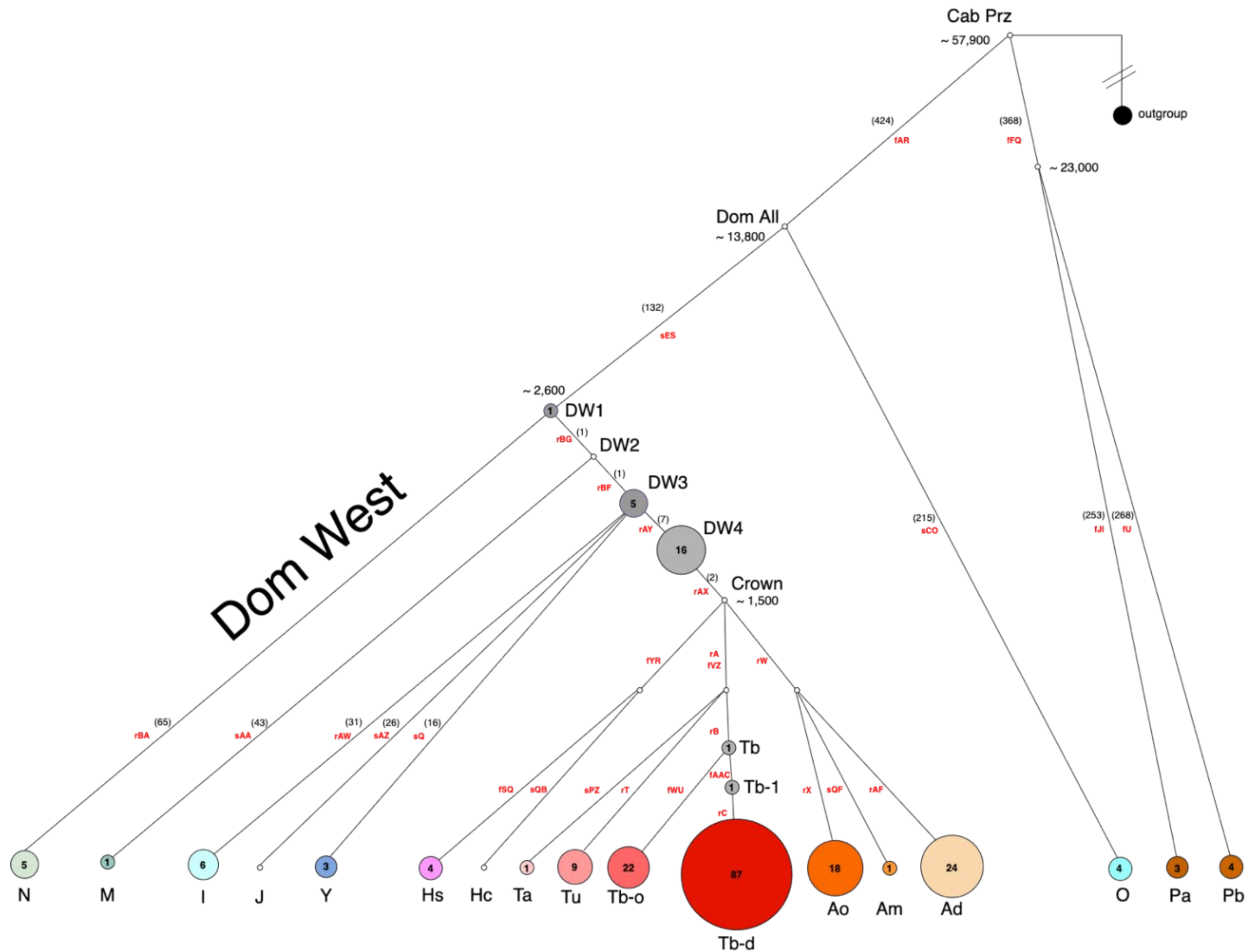


Figure 6. Haplotype (HT) distribution and frequency plot based on MSY tree modified from Felkel et al. 2019. HTs are given as circles with HT symbols below and circle radius corresponds to the number of clustered individuals. Absolute numbers of individuals are given inside the circles. Different colors and shades correspond to HTs represented in the dataset, while non colored points express HTs that were not detected in the sample set. Number of mutations on Non-Crown branches are denoted in brackets; in Crown HG it ranges 5-26 (Felkel *et al.* 2019).

Markers used for genotyping are in red font; Domestic West is abbreviated as DW. Estimated splitting times for branching points (years before present) are from Felkel *et al.* (2019).

As expected, the 5 Przewalski´s horses separated from domestic horses and fell into the previously identified P HT (Felkel *et al.* 2019) (Figure 6). However, while the Pb sub-HT comprised of 4 Przewalski's horses, the Pa group comprised of a single Przewalski's horse and two Heck horses, a horse linneage created by integrating domestic horses with the Przewalski's horses to "breed back" the extinct Tarpan horse (Heck 1952) (Supplementary Table S7).

<u>MSY gene CNV across MSY haplotypes</u>

We generated for each of the 21 MSY HTs (including the kulan outgroup) an average CN of MSY genes (Supplementary Table S8) and used F-statistics to determine if there was a significant CN variation between HTs. Similar to gene CN variation across breeds (Table 9), significant CN differences between HTs were observed for *ETSTY1* ($P < 0.0001$), *ETSTY2* ($P < 0.0001$), *HSFY* ($P = 0.0183$), *RBMY* ($P < 0.0001$), *TSPY* ($P = 0.0014$), and *SRY* ($P < 0.0001$) (Supplementary Table S8). Copy numbers of *ETSTY5* and *UBA1Y* were not significantly different between HTs. Like in breed comparison, *TSPY* had the broadest range of variation with a minimum CN=8.29 in the Y HT and maximum CN=16.79 in the Tu HT. *ETSTY1* CNs ranged from 2.29 (kulan outgroup) to 7.32 (Pa), *ETSTY2* from 3.03 (Pb) to 8.41 (Tu), *ETSTY5* from 3.5 (kulan outgroup) to 7.6 (Tu), *SRY* from 0.65 (Tb) to 2.17 (O), *RBMY* from 0.82 (kulan outgroup) to 2.58 (DW4), *HSFY* from 0.87 (DW1) to 1.69 (kulan outgroup), and *UBA1Y* from 2.67 (Tb-1) to 4.73 (Tu) (Supplementary Table S8). Interestingly, an average rounded CN of *SRY* in O HT was 2 and CN of *RBMY* in DW4 was 3. In this study, there are clear CN differences between Y HTs, however, we did not find correlation between SNV-based HTs and CNV pattern (i.e., the

two forms of sequence variation were independent). The simplest example of such discordance is in the Estonian Native breed where individuals with 2 copies of SRY represent 4 different HTs. Another example is the Lipizzan breed (n=10) TSPY CN which ranges from 8.2 to 9.1, and represent 3 HTs (Ao, Hs, and Tb-o). Therefore, individual CNs do not correspond with individual HTs.

MSY gene copy number variation between normal and abnormal male horses

We investigated MSY gene CNs in 3 groups of abnormal male horses. The first group consisted of 24 American Quarter Horses with bilateral (n=6) or unilateral (n=18) cryptorchidism (CO). MSY CNs were compared within this group (unilateral vs. bilateral CO) as well as with the 28 normal American Quarter Horses (QH) from the large male cohort (Table 11). There was no significant CN variation between the bilateral and unilateral cryptorchids, nor between bilateral cryptorchids and the normal cohort (Table 11). However, there was a significant CN difference in *TSPY* (P = 0.0452), *SRY*, (P = 0.0027) and *RBMY* (P = 0.0467) between unilateral cryptorchid and normal males. Copy numbers of the same three genes and *ETSTY2* (P = 0.0392) were significantly different between all cryptorchid and normal American Quarter Horses (Table 11). The *TSPY* gene showed the most notable CN change having 2 copies less in cryptorchids (rounded average CN=9) than normal males (rounded average CN=11). The average rounded CN of *ETSTY2* was 4 in cryptorchids compared to 5 copies in normal males. It must be noted that while statistical analysis showed significant CN differences also for *SRY* and *RBMY*, their average rounded CNs in cryptorchid and normal males were the same – 1 copy for *SRY* and 2 copies for *RBMY* (Table 11). Individual CNs for the 52 Quarter Horses used for this analysis are presented in Supplementary Table S9.

Table 11. MSY gene average CN variation between cryptorchid and normal American Quarter Horses.

Horse groups	N	ETSTY1	ETSTY2*	ETSTY5	HSFY	RBMY*	SRY**	TSPY*	UBA1Y
Bi-CO	6	4.16	4.22	4.01	1.03	1.92	0.9	9.11	3.63
Uni-CO	18	4.69	4.17	4.21	1.06	1.95	0.96	9.22	3.7
CO all	24	4.55	4.18	4.16	1.05	1.94	0.94	9.19	3.68
Normal	28	4.67	4.58	4.21	0.99	1.78	0.83	10.65	3.42
F-statistic *P*-value									
Bi-CO vs. Uni-CO		0.4708	0.7998	0.5978	0.6795	0.85497	0.4017	0.883	0.7249
Bi-CO vs. normal		0.1338	0.3067	0.3576	0.6445	0.1896	0.1484	0.167	0.3688
Uni-CO vs. normal		0.9654	0.0642	0.9794	0.2206	**0.0467**	**0.0027**	**0.045**	0.0746
CO all vs. normal		0.7213	**0.0392**	0.7595	0.2208	**0.0303**	**0.0025**	**0.019**	0.0583

Bi-CO – bilateral cryptorchid; Uni-CO – unilateral cryptorchid; * $P < 0.05$; ** $P < 0.005$

The second abnormal group of horses comprised of 29 individuals from 7 breeds or breed mixes with 64,XY karyotype and various forms of disorders of sex development (DSDs) (Supplementary Table S10). Of these, 4 individuals had cytogenetically detectable Y chromosome deletions (64,XYdel) and female-like or intersex phenotypes,12 individuals were XY females with *SRY*-negative male-to-female sex reversal condition (Raudsepp *et al.* 2010; Bugno-Poniewierska & Raudsepp 2021), 6 were female-like horses with *SRY*-positive sex reversal, and 7 individuals were phenotypically intersex with normal *SRY*-positive 64,XY male karyotype.

None of the 8 ddPCR assays amplified in the 4 individuals with Y deletions indicating complete loss of these sequences. All 12 XY *SRY*-negative sex reversal females had only 1 copy of *RBMY* instead of the expected 2 copies, suggesting that one copy of *RBMY* was lost together with the loss of *SRY*. The remaining 13 horses with *SRY*-positive XY DSDs did not show noticeably higher or lower CNs for the genes tested when compared to their corresponding breed group average in the large male cohort (Table 9). Individual CNs for the DSD group are presented in Supplementary Table S10.

The third group of abnormal horses contained 14 male horses of 6 breeds with variable subfertility/infertility phenotypes (Supplementary Table S5). Comparison of individual MSY CNs in this group to their breed average in the large normal male cohort (Table 9) did not reveal any statistically significant differences. Though, we noticed that CNs of two Arabians in this group slightly deviated from breed average: one (H963) with idiopathic subfertility and autosomal translocation had higher *TSPY* (CN=14.2 vs. 10.03), *ETSTY2* (CN=4.87 vs. 4.33), and *ETSTY5* (CN=5.86 vs. 4.10) CNs, while another (H284) with idiopathic subfertility had a lower *UBA1Y* CN (CN=1.42 vs. 2.94) CN. Individual CNs for the subfertile/infertile male group are presented in Supplementary Table S11

MSY gene copy numbers and haplotypes of closely related males

Within the 209 normal male cohort, we identified 4 sets of directly related male individuals with available MSY CN and HT information. These included two sire-son pairs, one grandsire-son pair, and one grandsire-sire-son trio. In addition, MSY CN data was generated for two cloned Arabians (group 5) and 6 cloned American Quarter Horses (group 6), all produced by somatic cell nuclear transfer (SCNT) (Table 12). This allowed us to investigate the dynamics of

MSY gene CN and HT between generations and MSY CN across genetically identical individuals. As expected, MSY HGs, HTs and sub-haplotypes were conserved in all paternal lineages (Table 12). In contrast, MSY gene CNs were not conserved, and we observed duplications and/or deletions in every generation including the cloned horses.

In group 1 (Table 12), we determined that *TSPY* underwent major deletions between the 3 generations, losing 6 copies from Grand sire (CN=16) to Son (CN=10). Likewise, *ETSTY2* lost 2 copies over 3 generations - from CN=6 in Grand sire, CN=5 in Sire to CN=4 in Son. Different dynamics was observed for *ETSTY1*, which had CN increase from Grand sire (5 copies) to Sire (7 copies) but reduced to 4 copies in Son. Copy numbers of *SRY* (CN=2) and *RBMY* (CN=3), however, remained the same over the 3 generations. Because of limited amount of DNA, we were not able to obtain CN data for *ETSTY5*, *HSFY*, and *UBA1Y* for the Son (BP364), though CNs of these three genes did not differ between the Grand sire and Sire.

In group 2 (Table 12), we compared MSY gene CN between a Grand sire and Son and observed CN increase from Grand sire to Son in *TSPY* (+1 copy), *ETSTY1* (+2 copies), *SRY* (+1 copy), and *RBMY* (+1 copy) and a decrease in *ETSTY2* (-1 copy) and *ETSTY5* (-1 copy). No CN changes were observed for *HSFY* and *UBA1Y*. Here, the most intriguing over generations CN change was for *SRY* and *RBMY* because as presented in MSY reference assembly (Janecka et al. 2018) and confirmed by ddPCR in this study (Table 8), most male horses have 1 copy of *SRY* and 2 copies of *RBMY*.

Groups 3 and 4 were both Sire-Son pairs, and we observed more MSY CN variation in group 3 over a single generation (Table 12). *TSPY, ETSTY1*, and *ETSTY2* each lost one copy, and *ETSTY5* and *UBA1Y* gained one copy from the sire to son. MSY CNs were more stable in group 4, where the only difference between generations was an extra copy of *TSPY* in Son.

Group 5 comprised of genetically identical Arabians derived from the same somatic cell donor (DNA not available) by SCNT. The two clones differed by 1 copy for *ETSTY5* and 2 copies for *TSPY*. In group 6, we compared CN of 5 cloned American Quarter Horses and their somatic cell donor and observed a range of CNV for *TSPY* (12-14 copies) and some CN differences for *ETSTY1* and *ETSTY2* (Table 12).

Table 12. MSY gene CN and HT comparison between related males.

Group	**Relation**	**Breed**	**HT**	***ETSTY1***	***ETSTY2***	***ETSTY5***	***HSFY***	***RBMY***	***SRY***	***TSPY***	***UBA1Y***	**ID**
1	Grand-sire	ENH	DW4	5.3	6	3.7	0.94	2.6	1.77	15.8	3.5	BP387
1	Sire	ENH	DW4	6.8	4.9	4	0.79	3	2	14.7	3.18	BP378
1	Son	ENH	DW4	3.8	3.7	n/a	n/a	2.81	1.96	10.1	n/a	BP364
2	Grand-sire	ENH	Ad	4	5.4	4.6	0.88	1.92	0.79	11.7	2.9	BP282
2	Son	ENH	Ad	5.6	4.5	4	0.99	2.85	1.55	12.9	3.07	BP388
3	Sire	ENH	DW4	5	5	3.3	1.23	2.53	1.76	10.4	3.06	BP399
3	Son	ENH	DW4	4.4	4.2	3.7	0.94	2.61	1.98	9.2	3.7	BP400
4	Sire	HH	Pa	7.3	7.2	5.4	1.17	0.6	0.68	17.8	3.6	15758
4	Son	HH	Pa	7.7	6.6	5.3	1.05	0.79	0.73	18.7	3.4	21150
5	Cloned brother	AR	n/a	4.37	4.62	4.73	1.07	1.74	0.99	11.9	3.11	H962
5	Cloned brother	AR	n/a	4.27	4.87	5.86	0.99	1.7	0.86	14.2	3.1	H963
6	Original Donor	QH	n/a	4.9	5.6	4.52	0.93	1.89	0.77	14.3	3.2	H396
6	Cloned brother	QH	n/a	4.69	4.6	4.41	1.06	1.69	0.84	12.8	3.35	H391
6	Cloned brother	QH	n/a	4.55	5.2	4.09	0.98	1.68	0.79	12	3.01	H392
6	Cloned brother	QH	n/a	4.54	5.3	4.32	1.02	1.54	0.77	14.4	2.98	H393
6	Cloned brother	QH	n/a	4.29	4.9	4.45	1.15	1.79	0.8	13.5	3.1	H394

6	Cloned brother	QH	n/a	4.44	5.1	4.44	0.96	1.92	0.78	11.6	2.85	H395

ENH – Estonian native horse; HH – Heck horse; AR – Arabian horse; QH – Quarter Horse

3.2.3. Discussion

Here we present the first comprehensive droplet digital PCR-based copy number analysis of 7 horse Y chromosome multi-copy genes and *SRY*. We established a baseline CN for these genes, allowing critical evaluation of the current horse MSY sequence assembly eMSYv3 (Janecka *et al.* 2018) and to study MSY gene CN variation in large horse populations, and males with disorders of sex development and reproduction. For the first time, the dynamics of horse MSY variation was compared at gene CN and single nucleotide variation (SNV) levels.

Copy number analysis of genes in the structurally complex Y chromosome relies heavily on the availability of high-quality reference assembly. The first annotated reference sequence of the horse MSY, eMSYv3 (Janecka *et al.* 2018), presents a high-quality assembly of single-copy regions but remains tentative for the ampliconic MSY – the region where multi-copy genes reside. The tentative nature of the ampliconic MSY assembly complicated the design of ddPCR assays for CN analysis in this and previous studies (Castaneda *et al.* 2019; Castaneda *et al.* 2021b) where one of the main limitations was inability to find a single shared male-specific sequence across all copies of a gene. Therefore, we were able to develop CN assays for only 7 (*ETSTY1*, *ETSTY2*, *ETSTY5*, *HSFY*, *RBMY*, *TSPY*, and *UBA1Y*) out of the 15-known horse MSY multi-copy genes (Janecka et al. 2018), as well as for the single-copy *SRY*. The study of the remaining 8 multi-copy genes, will require substantial improvement of the assembly of the ampliconic region of horse MSY.

In a way, improvement of the current eMSYv3 already started in this study by comparing CNs of 8 MSY genes in the same individual horse - a Thoroughbred stallion *Bravo* who was the DNA donor for eMSYv3 (Janecka *et al.* 2018) and the reference male for all ddPCR experiments (Table 8). Since multiple ddPCR experiments gave consistent CNs for MSY genes and autosomal control genes in the reference horse, we considered ddPCR results reliable. It is therefore noteworthy that only two genes, *SRY* and *RBMY*, both located in a transitional region between single-copy and multi-copy MSY (Janecka *et al.* 2018), had the same CN in eMSYv3 and by ddPCR (Table 8) confirming correct assembly of this MSY region. In contrast, *ETSTY2*, *ETSTY5*, *HSFY*, *TSPY*, and *UBA1Y* had almost twice as many copies in eMSYv3 than detected by ddPCR, suggesting over-assembly of the corresponding regions. Conversely, slightly lower CN for *ETSTY1* in eMSYv3 (CN=3) compared to ddPCR (CN=5) indicated that MSY reference is likely missing some copies of this equine testis-specific transcript. While our findings strongly support the accuracy of ddPCR results over eMSYv3, we cannot exclude a possibility that the designed ddPCR assays did not target all copies of some genes due to incomplete or diverged sequences.

The development of CN assays and determining baseline CN for select MSY genes in the reference male *Bravo*, allowed us to expand CN analysis to large multi-breed horse populations and related equids. To date, this is the most extensive MSY gene CN study in equids encompassing 282 domestic horses (209 normal and 73 with disorders) from 22 breeds, the Przewalski's horse and kulan (Supplementary Tables S4, S5). The only study of similar scope has been conducted in 263 donkeys of 13 breeds (Han *et al.* 2017a) where CNs of 5 MSY genes (*CUL4BY*, *ETSTY1*, *ETSTY4*, *ETSTY5*, and *SRY*) were evaluated by qPCR which is a relative quantitation method. Due to different methodological approaches (qPCR vs. ddPCR), the results

of the donkey study are too different for any meaningful comparison with our data. For example, the donkey study documented *SRY* CN range from 1 to 152. This is in stark contrast with this study where we show that *SRY*, together with *RBMY*, were the only genes with consistent average CN across all study cohorts and the eMSYv3 reference (Table 8, 9). An additional support for the accuracy of ddPCR was the fact that average CNs of the remaining 6 genes were the same (*ETSTY1, ETSTY5, HSFY*) or similar (*ETSTY2, TSPY, UBA1Y*) between the reference male *Bravo* and the multi-breed cohort of 209 normal horses (Table 8). Also, previous studies have indicated high degree of cytogenetic and sequence conservation between the horse and donkey MSYs (Paria *et al.* 2011). Therefore, it is unlikely that the CN differences between this study and that by Han *et al.* (2017) were caused by extensive divergence of equine and asine Y chromosomes. This is further supported by our results in the kulan, another equid from the asine group, where *SRY* CN was consistently one (Table 9).

Much more gene CN variation was observed when the 209-horse cohort was broken down into breeds showing statistically significant CN differences for 5 of the 8 genes studied (Table 9). However, significant inter-breed CN differences of *SRY* and *RBMY* were exclusively caused by a few individuals from indigenous breeds (Table 10). Otherwise, as mentioned above, CNs of these two genes were stable across most breeds (Table 9) and individuals (Supplementary Table S4). Likewise, significant inter-breed CN variation of *ETSTY2* was caused by one Yakutian (TR028) and 2 Estonian Native horses (BP379 and BP380), having 14 or 13 copies, respectively (Supplementary Table S4), compared to the horse cohort average of 5 (Table 8). In fact, the same Yakutian horse (TR028) showed extremely high CNs for all genes studied, except *RBMY*, suggesting that the horse may have a cytogenetic abnormality with an extra Y chromosome. Though, we could not verify this because cytogenetic information was available

only for the 73 abnormal males (Supplementary Table S5) but not for most individuals in the 209-horse cohort. Regardless of the inclusion of sample TR028, the truly most CN variable gene across breeds and individuals was *TSPY*, the gene which also had the highest CN (average 10, lowest 6, highest 38) among all MSY genes (Table 8, 9; Supplementary Table S4). Higher variability between individuals within larger ampliconic gene families (specifically *TSPY*) have also been reported in humans (Skov *et al.* 2017; Lucottee *et al.* 2018; Ye *et al.* 2018; Vegesna *et al.* 2019) and great apes (Oetjens *et al.* 2016; Tomaszkiewicz *et al.* 2016; Vegesna *et al.* 2020) and is because multi-copy genes with higher CN have an increased probability of being involved in intra-chromosomal rearrangements compared to genes with lower CNs (Ghenu *et al.* 2016).

In this context, *TSPY* is also a good example for other species because it is a multi-copy gene in nearly all mammalian Y chromosomes (Bellot *et al.* 2014; Cortez *et al.* 2014), but shows different degree of CNV in different species depending on the baseline CN. For example, the estimated CN of cattle *TSPY* is 50 to 200 and the gene shows significant CNV between individuals, breeds, and subspecies (taurus and indicus) (Hamilton *et al.* 2009; Mukherjee *et al.* 2013; Hughes *et al.* 2020). Also, *TSPY* is highly amplified in the domestic cat (~ 100 copies) and shows considerable CNV between felids (Janecka *et al.* 2018). In contrast, no CNV between individuals or breeds has been observed for pig *TSPY* which has just 3 copies (Quach *et al.* 2015). On the other hand, our results across all study cohorts strongly suggest that there is only one copy of *HSFY* in horse MSY and not 3 copies as presented in eMSYv3 (Janecka *et al.* 2018). Single-copy *HSFY* in horses is more similar to the 2 copies in humans (Tomaszkiewicz *et al.* 2016; Vegesna *et al.* 2020) and 6 copies in gorilla (Tomaszkiewicz *et al.* 2016), but in stark contrast to cattle and pig, where *HSFY* is massively amplified (Skinner *et al.* 2015; Hughes *et al.* 2020).

One of the most intriguing findings of this study was documenting 21 horses within the 209 normal male cohort with 2 or 3 copies of *SRY* (Table 10). Eighteen of these horses also had an extra copy of *RBMY* (CN=3), though 3 horses had a single *RBMY* instead of the normal 2. Not coincidentally, the cohort of 12 abnormal horses with *SRY*-negative XY DSD, had lost together with *SRY*, a copy of *RBMY* (Supplementary Table S10). Interrelationship of *SRY* and *RBMY* CNs is the consequence of the specific features of horse MSY structure where the single-copy *SRY* is embedded between almost 100% identical direct repeats, including two copies of *RBMY* (Janecka *et al.* 2018) (Figure 7). Because MSY is not recombining, it maintains genetic integrity by other mechanisms, of which one is homologous repair between sister chromatids. However, in structurally complex regions containing palindromes, inverted and direct repeats, exchange may happen between geographically distant repeats (non-allelic homologous repair), resulting in intra-chromosomal structural rearrangements (Lange *et al.* 2009).

For example, non-allelic homologous repair between chromatids in the horse *SRY*-region, may remove a segment with *SRY* and one copy of *RBMY* from one chromatid and add it to the other chromatid (Figure 7). In meiosis, this will result in two different sperm: one with 2 copies of *SRY* and 3 copies of *RBMY*, another with a single *RBMY* and no *SRY* (Figure 7), the latter will lead to *SRY*-negative XY DSD (also known as male-to-female sex reversal) (Raudsepp *et al.* 2010). This scenario was initially proposed as a likely mechanism to explain the relatively high incidence of *SRY*-negative XY DSD in horses compared to other domestic species (Raudsepp *et al.* 2010). Structural complexity and likely instability of the *SRY*-region was further confirmed by the horse MSY reference assembly (Janecka *et al.* 2018) and is consistent with ddPCR results in *SRY*-negative XY DSD horses in this study (Supplementary Table S10). However, until the

development of ddPCR assays, there have been no accurate tools to identify male horses with increased *SRY* and *RBMY* CN.

It is certainly noteworthy that the 21 horses with more than one copy of *SRY* were in the normal male cohort and 7 were confirmed breeding stallions (Figure 10), suggesting that elevated *SRY/RBMY* CNs have no negative phenotypic effect on fertility. It is though, puzzling that all males in this group were from small indigenous breeds (Table 10), while the *SRY*-negative XY DSD condition (Figure 7), has been found in many common breeds (Raudsepp *et al.* 2010; Bugno-Poniewierska & Raudsepp 2021). At present, we do not have any plausible explanation why we did not detect any horses among common commercial breeds with elevated *SRY/RBMY* CN. We can only speculate that this may be associated with subtle phenotypic changes affecting human selection decisions in commercial breeds but have no importance in less-controlled indigenous horses

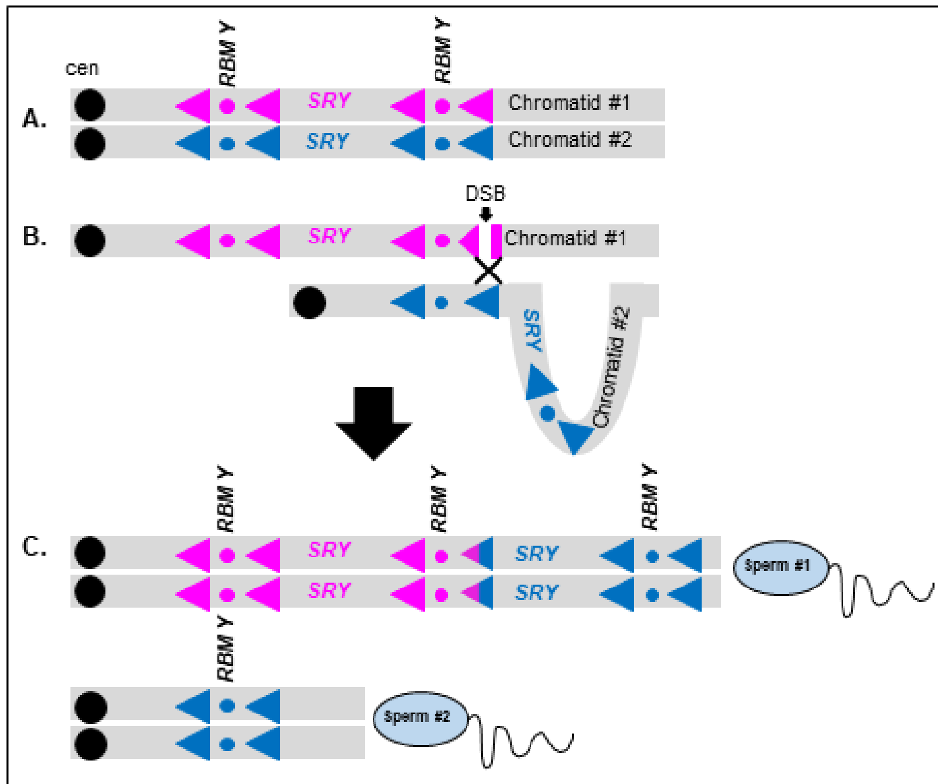


Figure 7. Schematic representation of the SRY-region in horse MSY proposing a mechanism for SRY/RBMY CNV. **A**. Normal Y chromosome with two identical sister chromatids; highly similar directional repeats flanking *SRY* are denoted with arrowheads in pink and blue color to distinguish the same repeats in sister chromatids; the 2 copies of *RBMY* are indicated with pink and blue dots and the single-copy SRY in pink and blue font in the two sister chromatids, respectively. **B**. Proposed mechanism for the repair of double stranded break (DBS) in one chromatid (arrow) by non-allelic homologous exchange (black cross) with an identical, but geographically distant repeat in the other chromatid as shown by sister chromatid misalignment and looping of Chromatid #2. **C**. Outcomes of the non-allelic homologous exchange shown in B. after sister chromatid separation in Meiosis II, followed by DNA replication and sperm formation: sperm #1 carries Y chromosome with 2 copies of *SRY* and 3 copies of *RBMY*, while sperm #2 has a single *RBMY* and no *SRY*. The latter leads to SRY-negative XY sex reversal. The

idea of non-allelic homologous exchange between Y sister chromatids is adopted from Lange et al. (2009) and Teitz et al. (2018).

To a very limited extent, we investigated MSY CNVs in related equids – 5 Przewalski's horses and 2 kulans and observed lower overall CNV compared to the domestic horse. A notable difference from the domestic horse was that all 7 wild equids had a single copy of *SRY* and *RBMY* (Table 9; Supplementary Table S4), suggesting that the structure of this region in these species may be different from horse MSY. Definite answers, however, need additional studies with more individuals and equid species. Otherwise, the successful use of all 8 ddPCR assays in the Przewalski's horse and 6 assays (except *TSPY* and *UBA1Y*) in kulan, suggests high degree of sequence conservation between these Y chromosomes.

Previously, sequence variation in the horse Y chromosome has been studied at single nucleotide level, which compared to other domestic species and wild equids, is outstandingly low (Wallner *et al.* 2003; Lindgren *et al.* 2004; Wallner *et al.* 2017). Nevertheless, the identified SNVs have allowed to determine MSY HGs and HTs, trace the origin of patrilines and gain information about the relationships between horse breeds (Wallner *et al.* 2017; Felkel *et al.* 2019). Here, we generated information for another form of MSY variation – CNV of multi-copy genes and showed that there is no correlation between Y CNs and HTs. For example, Estonian Native horses had similar CN patterns (see Table 9) but separated into both Crown and non-Crown HGs based on SNVs. Likewise, the two Estonian Native horses (BP379 and BP380) with over two times higher than average CN for *ETSTY2*, belonged to the most common Crown HG (Supplementary Table S4). Conversely, individuals from non-Crown HG, did not necessarily stand out regarding their CN patterns, except the above discussed outlier -Yakutian horse TR028.

Our observations are consistent with those in human and primates showing that SNV-based haplogroups do not cluster with CNV-based haplogroups (Ye *et al.* 2018; Vegesna *et al.* 2020). Also, similarly to primates, the studied 209-horse population showed much more diversity in MSY CNs compared to nucleotide diversity which defined only 2 HGs (Crown and non-Crown) and 20 HTs (Figure 6; Supplementary Table S8). The same discordance between CNVs and SNVs was evident in successive male generations and between genetically identical horses generated by SCNT (Table 12) - CN showed variation, while HTs remained the same. It is rather that this and previous studies attempted to compare *'apples and oranges'* because the sequence properties, molecular mechanisms, and evolutionary dynamics underlying CNVs and SNVs are different. Single nucleotide variants that determine HGs reside in MSY non-genic single-copy sequences and are mainly influenced by mutations which rate is as low as 2.91×10^{-8} bp/generation (Wallner *et al.* 2017). This is clearly different from CNVs of functional genes in structurally complex ampliconic sequences which are prone for structural rearrangements by inter- and intra-chromatid exchanges and gene conversion (Lange *et al.* 2009) (Figure 7). Also, our findings of trans-generational CN changes, as well as of CNV between cloned horses suggest that these structural rearrangements can be of both meiotic and mitotic origin.

In humans where Y chromosome research is currently the most advanced, analysis of high-throughput sequencing data from over 1200 males has allowed to accurately detect CN of MSY ampliconic genes in each individual, but also to determine the ancestral reference CN for each gene (Teitz *et al.* 2018). It appears that even though there is CNV between individuals, the reference (ancestral) CN of each ampliconic gene is rigorously maintained, indicative of mutation-selection balance. The presence of selective constraints on amplicon CN in human Y chromosome, suggests that MSY CNVs have phenotypic effects, most likely on spermatogenesis

(Teitz *et al.* 2018). It is too early to comment about whether and how this may apply to horse MSY CNVs, but the idea is important regarding stallion fertility and worth pursuing in future research. For example, even though in this study we determined the baseline CN for 8 MSY genes, a much larger and more diverse equine population is needed to find out whether the determined baseline CN is also the ancestral condition. The lack of such information combined with the overall limited structural and functional knowledge about the horse MSY ampliconic region, also sets limits to interpret CN analysis results in cryptorchid and infertile/subfertile males.

The observed lower CN of *TSPY* and *ETSTY2* in cryptorchid American Quarter Horses (Table 11) left only questions. On the other hand, the same two genes were most variable also in the normal cohort (Table 9), thus it is possible that the small sample size and the known heterogeneity of the cryptorchid phenotype (Amann & Veeramachaneni 2007) may have affected the statistics. However, if the association is true, we have no knowledge about the functions of the equine-specific transcript *ETSTY2* or the horse *TSPY* gene. Copy number of the latter has been associated with subfertility phenotypes in men (reviewed by (Rogers 2021)) and lower semen quality in bulls (Mukherjee *et al.* 2015) but not with cryptorchidism. The fact that we did not detect any significant CNV among subfertile/infertile stallions is likely the consequence of too many diverse phenotypes and very small sample size per each. Furthermore, most MSY multi-copy and ampliconic genes have not yet been functionally annotated in horses or any other domestic species, which greatly limits the understanding of their role in stallion biology.

This study created a unique opportunity to compare the CNs of horse MSY multi-copy genes in the reference assembly eMSYv3 (Janecka *et al.* 2018) with those determined by ddPCR. The observed disparities between the two indicate that the current assembly of the horse MSY

ampliconic region requires improvement. Improved reference sequence is also necessary for the design of ddPCR assays for the remaining 8 multi-copy MSY genes and, thus, expand gene CN analysis in the horse Y chromosome.

3.3. Improvement of the current Y reference sequence eMSYv3

The following sections (3.3.1 and 3.3.2) will discuss our attempts to improve the current horse Y chromosome reference eMSYv3 (Figure 8) using a combination of traditional and state-of-the-art approaches.

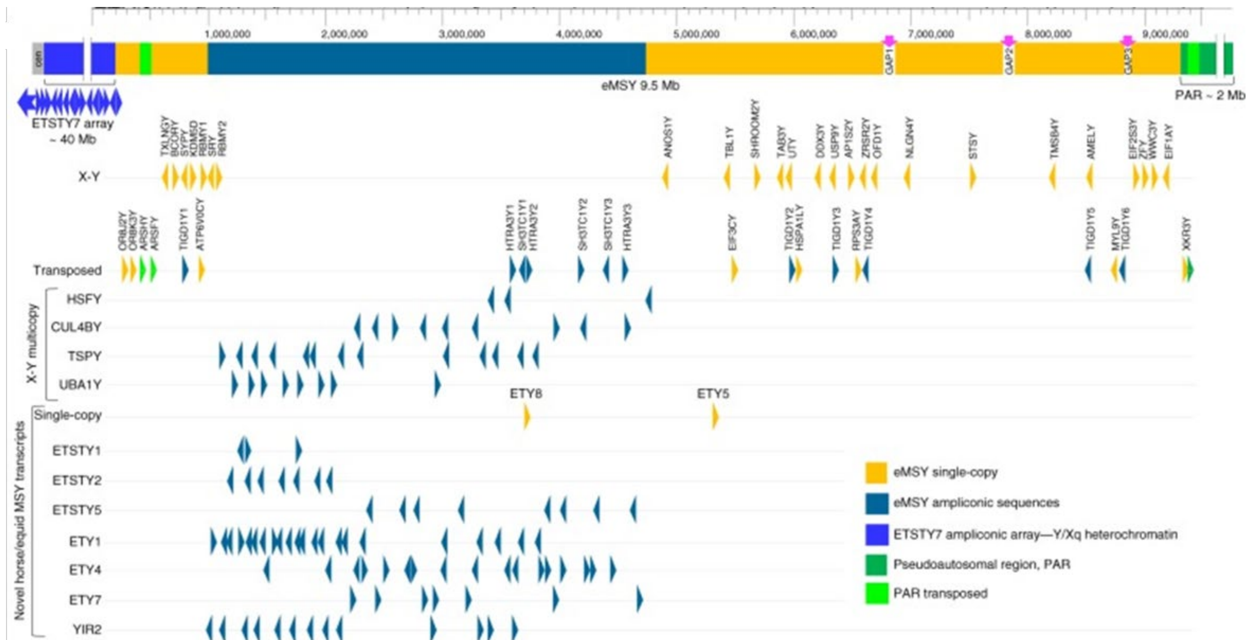


Figure 8. Current published sequence map of eMSYv3 showing different sequence classes, annotated genes, their copy numbers, evolutionary origin, and direction of transcription (orientation of arrowheads). Pink arrows identify gaps located in the single copy region (see Figure 1 in (Janecka et al. 2018)).

3.3.1. Improvement of the multi-copy region of the horse Y chromosome

The following section focuses on the improvements of the multi-copy region of the Y chromosome (mcY) by re-sequencing and re-assembly of known mcY BACs using long-read technology.

3.3.1.1. Experimental methods and design

Isolating BAC DNA for sequencing

In total, there are 49 individual Bacterial Artificial Chromosome (BAC) clones which belong to the mcY region of the MSY assembly. Previously, 43 out of 49 were sequenced for the current eMSYv3 assembly (Janecka *et al.* 2018). For this study, all 49 mcY BACs in addition to 3 flanking BACs in contig Ic (Figure 9; Supplementary Table 12) were grown in 2YT (Life Technologies) media with chloramphenicol (Sigma). High molecular weight BAC DNA was isolated using the Roche High Pure Plasmid Isolation kit (Roche) following the manufacturer's protocol. BAC DNA quality and quantity was measured using TapeStation (Agilent Technologies) and Qubit fluorimeter with Qubit BR DNA Assay Kit (Life Technologies).

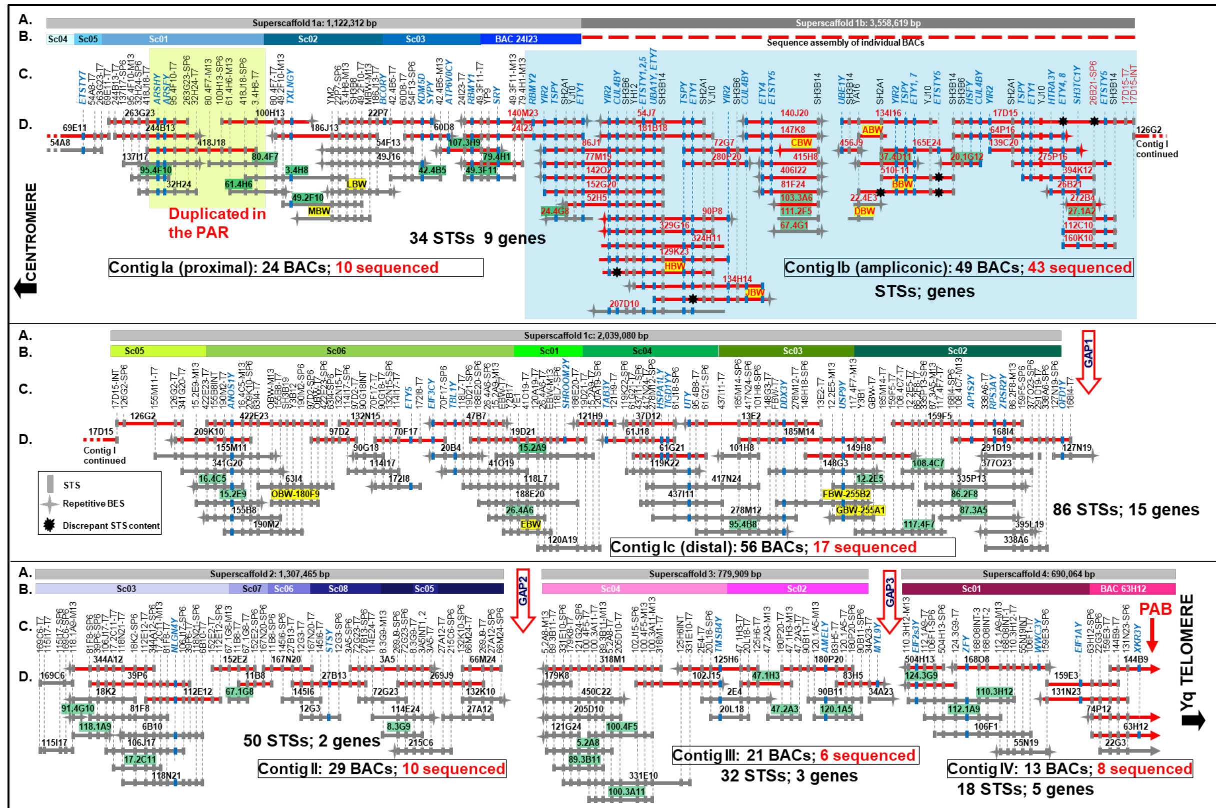


Figure 9. Published horse MSY contig and BAC map from Supplementary Figure S1 in Janecka et al. 2018. Red horizontal lines represent BAC clones sequenced to generate eMSYv3 and make up the 94-BAC tiling path for the horse Y chromosome. The multi-copy region of the Y chromosome is highlighted in blue (Contig Ib). Grey horizontal lines are BAC clones not sequenced for that study.

Sequencing BAC DNA

High molecular weight mcY BAC DNA was sequenced using long-read Oxford Nanopore technology. High molecular weight DNA ranges from 10 Kbp to over 100 Kbp in size and is frequently used for "third-generation sequencing" techniques which produce large sequencing reads (>10 Kbp long) which can span complex genomic sequences. For this study, clones were sequenced using the Flongle (Oxford Nanopore) so that sequencing reads could span

indirect and direct repeats native to the mcY. Individual clones were individually barcoded using the Oxford Nanopore Rapid barcoding kit before creating pools of maximum 4 BACs per pool. BACs were pooled together if they did not appear to share sequences based on the tiling path and eMSYv3 assembly. The Flongle generated ~ 1.5 Gb of data per flow cell with each BAC sequenced to a minimum depth of 100X coverage.

Bioinformatic processes

Raw sequences of individuals BACs were filtered for both vector and *Escherichia coli* contaminants using the BBDuk program in BBTools (BBMap; sourceforge.net/projects/bbmap/). After filtering, individual BACs were *de novo* assembled using Canu (http://canu.readthedocs.io/en/stable/). Sequences were validated using known mcY BAC end sequences (BES) and the BAC sequences previously used for eMSYv3. BAC sequences were aligned using MAUVE (Darling *et al.* 2004), the lamassemble tool in MAFFTv7 (Fritch *et al.* 2021), and D-GENIES (Cabanettes & Klopp 2018) to identify BACs which carry the same genetic sequence. The new mcY reference assembly was annotated for known multi-copy genes using the ddPCR target gene sequences outlined in section 3.2.1 (see Table 7).

3.3.1.2. Results

De novo assembly of individual BACs

A total of 52 BACs (49 mcY and 3 flanking BACs from contig Ic) were sequenced using the Flongle and assembled individually *de novo.* The average size of the assembled BACs was 170,143 base pairs (bp). Six BACs were highly fragmented and generated many contigs, instead of the ideal range of 1 to 3 contigs per BAC. These 6 BACs were subject to resequencing and

were not used in the assembly of the new mcY reference. Two BAC DNA samples (BAC ID: 24.4G8 and 67.4G1) failed the quality control prior sequencing and were, therefore, excluded from this study. The final assembly was based on long-read sequences of 41 BACs from mcY and 3 single-copy BACs from Contig Ic.

Generating a new mcY assembly

The new mcY assembly was methodically generated using different multiple alignment tools, with the goal of preventing indirect and direct repeats from collapsing. First, individual mcY BACs were aligned pairwise and in groups using MAUVE to determine which BACs contained the same sequences. After the initial BAC alignment (pairwise and groups), lamassemble was used to generate a larger contiguous consensus sequence of the select BACs. D-GENIES was then used to determine the accuracy of the consensus sequence by cross referencing the consensus sequence with individuals BACs that made up the target sequence. This process generated five contigs spanning across the mcY region and into the flanking contig Ic. These 5 contigs were then aligned and integrated together to form the final 1.53 Mbp size mcY assembly (Supplementary Figure S2).

Comparison of the new mcY assembly with published MSY reference

The multi-copy gene ddPCR assays described in section 3.2.1 were used to validate and annotate the new mcY assembly by gene CN comparison. In section 3.2.2., we showed that mcY gene CNs in eMSYv3 were higher than in the same reference male (Thoroughbred stallion *Bravo*) by ddPCR. However, similar CN comparison with the new mcY assembly generated in

this study, showed that the new assembly contained less copies of the target genes compared to ddPCR results in the same reference male (see Table 7).

Except for the *RBMY* and *SRY* genes, which CNs were consistent between the studies, and the *HSFY* gene, which CN was the same by ddPCR and in the new assembly, all other target gene CNs were essentially reduced in the mcY assembly generated in this study. It is likely that despite our efforts to prevent this, many repetitive sequences still collapsed during the *de novo* assembly, resulting in an under-assembled mcY sequence (Supplementary Figure S2).

3.3.2. Improvement of the single copy region of the horse Y chromosome

The following section focuses on the improvements of the single copy region of the Y chromosome (scY). Here we focus on using sequences missing from eMSYv3 to close the 3 gaps (Figure 8) found in the scY and identify other areas which need assembly correction.

3.3.2.1. Experimental methods and design

<u>Filtering of unidentified scY contigs</u>

A total of 179 male-specific contigs which are not present in the eMSYv3 assembly were obtained from a collaboration with Dr. Barbara Wallner at the Institute of Animal Breeding and Genetics, the Veterinary University of Vienna, Austria. These sequences are available from NCBI PRJNA428358 (Felkel *et al.* 2019). First, the contigs were repeat masked using RepeatMasker (http://www.repeatmasker.org/) to identify nonrepetitive sequences for primer design. The 179 contigs were then aligned against eMSYv3 using BLASTn and MegaBLAST (https://blast.ncbi.nlm.nih.gov/) tools in NCBI to validate that there was no sequence similarity

between the contigs and Y reference (Janecka *et al.* 2018). Contigs were considered "high interest" if they were not completely repetitive and had no sequence similar to eMSYv3. After identifying high interest contigs, a minimum of 2 primers were designed per contig using Primer3Plus software (Untergasser *et al.* 2012) to amplify ~300-500 base pair products.

<u>BAC library and BES screening for contig placement</u>

First, the primers designed from the contigs of high interest, were used to screen by PCR the 94-BAC tiling path of the eMSYv3 to verify BLASTn results (see Figure 9). Second, the primers that did not amplify from any BACs in the MSY tiling path, were used to screen the CHORI-241 male horse BAC library (Thoroughbred "*Bravo*") (http://bacpacresources.org/). The screening was hierarchical, proceeding from superpools to plate pools and then from individual plates to specific clones. As some newly identified BAC clones fell within the gapped regions of the scY (pink arrows in Figure 8), subsequent screening was done by using primer pairs obtained from the end sequences of these BACs ((Leeb *et al.* 2006) NCBI nucleotide: https://www.ncbi.nlm.nih.gov/nuccore)) (Supplementary Table 13).

<u>BAC cultures, DNA isolation and florescent *in situ* hybridization (FISH)</u>

Once identified, the BAC clones that were not part of the 94-BAC tiling path were grown in culture for BAC DNA isolation and Y chromosome origin of all BACs was validated by fluorescence *in situ* hybridization (FISH).

The BACs were grown in 2YT (Life Technologies) media with chloramphenicol (Sigma). High molecular weight BAC DNA was isolated using the Roche High Pure Plasmid Isolation kit (Roche) following the manufacturer's protocol. Isolated DNA quality and quantity were

evaluated on TapeStation (Agilent Technologies) and Qubit fluorimeter with Qubit HS DNA Assay Kit (Life Technologies). To ensure that the BAC clones identified through PCR screening originate from the Y chromosome, DNA from individual BAC clones were labeled with biotin-16-dUTP and/or digoxigenin-11-dUTP nick translation kits (Roche Diagnostics) to create FISH probes. Combinations of two differently labeled probes were hybridized to metaphase and interphase chromosomes or mechanically stretched DNA fibers following our standard protocols (Raudsepp & Chowdhary 2008a). Biotin-labeled probes were detected with Alexa Fluor® 488 streptavidin conjugate (Molecular Probes, Life Technologies, Carlsbad, CA, USA) and digoxigenin-labeled probes with DyLight®594 anti-digoxigenin conjugate (Vector Laboratories, Burlingame, CA, USA). Chromosomes/DNA fibers were counterstained with 4′,6-diamidino-2-phenylindole (DAPI). Images of at least 10 cells or 30 DNA fibers were captured and analyzed for each experiment using a Zeiss Axio Imager M2p fluorescent microscope and Isis v 5.2 (MetaSystems GmbH) software.

Sequencing of newly identified Y BACs and bioinformatic processes

Newly identified MSY BAC DNA was sequenced using Nanopore technology. Clones were combined into pools and sequenced to a depth of >100X coverage and *de novo* assembled using Canu software (http://canu.readthedocs.io/en/stable/). Raw sequences of individuals BACs were filtered for both vector and *Escherichia coli* contaminants using the BBDuk program in BBTools (BBMap; sourceforge.net/projects/bbmap/). Sequences were validated using known BES ((Leeb *et al.* 2006); NCBI nucleotide: https://www.ncbi.nlm.nih.gov/nuccore) and select eMSYv3 scY contigs from Felkel *et al* (2019) that are located within the BAC clone. Sequences

which fell into the gaps of the scY, were manually integrated into the eMSYv3 assembly (our unpublished data).

3.3.2.2. Results

Contig placement within the current eMSYv3 tiling path

Intensive filtering of the initial 179 contigs identified 144 contigs that were not found in eMSYv3. Contig size ranged from 770 bp to 41,683 bp. After screening the 94-BAC tiling path by PCR, 51 contigs were placed within the published eMSYv3 assembly despite not having sequence similarities by BLAST analysis. Briefly, 1 contig was placed in the proximal portion of contig Ia (I proximal), 1 contig hit both the proximal (Ia) and distal (Ic) portion of contig I, 20 contigs were placed in the distal portion of contig I (Ic), 2 contigs have sequences which hit both the I multi-copy and I distal portion (Ib and Ic), 8 contigs were placed in contig II, and 19 contigs were placed in contig III of the Y chromosome (Supplementary Figure S3 and Supplementary Table S14). However, 35 of 179 contigs were completely repetitive and therefore uninformative for sequence placement and 5 out of the high interest 144 contigs could not be screened due to unaccounted repetitive regions or amplification errors.

BAC library screening and chromosome walking

The remaining 88 out of the 144 contigs of interest did not amplify any BACs within the eMSYv3 tiling path and were subject to CHORI-241 horse BAC library screening. Screening these contigs resulted in identifying 15 novel Y chromosome BACs spanning from contig Ic terminal BAC 127N19 (i.e., BACs located at the end of contigs) to contig II terminal BACs, 169C6 and 115I17, effectively spanning Gap 1 of eMSYv3 (Figure 8, 10, Supplementary Table

S14). Additionally, sequence tagged site analysis of BES by PCR (also referred to as BAC-walking), allowed for a BAC tiling path to be generated for Gap 1 (Figure 10). BES primers that were used to build the tiling path along with corresponding sequence tagging site GenBank Accession numbers can be found in Supplementary Table S13.

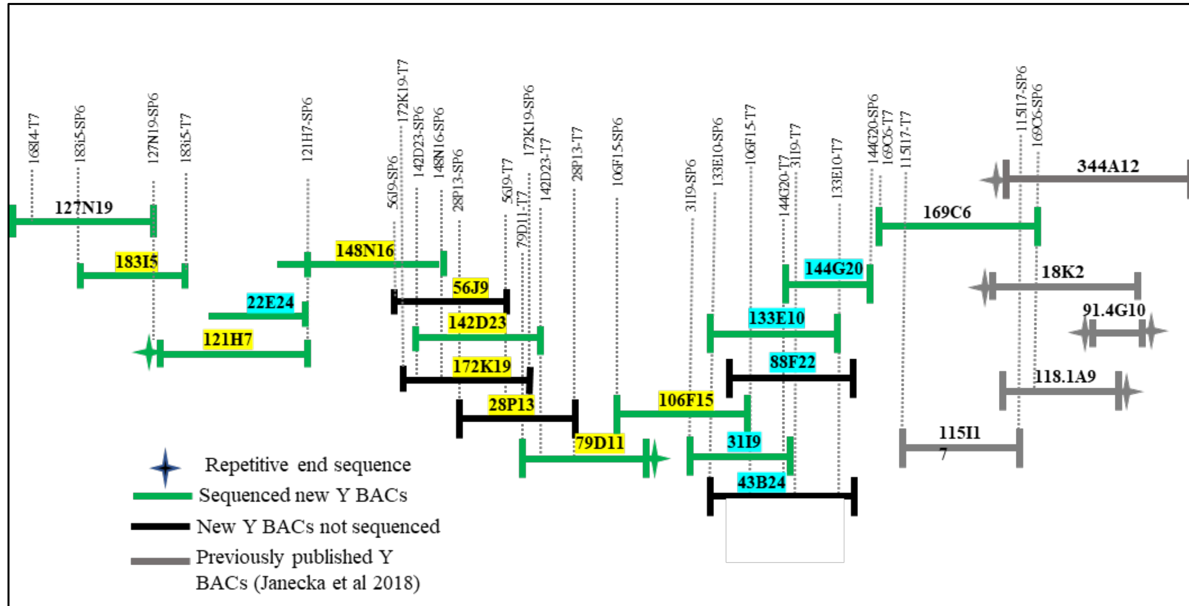


Figure 10. Detailed BAC map of Gap 1 in eMSYv3. Sequence tagged sites from BES that were used to build the BAC tiling path are presented at the top and connected to the corresponding BACs (horizontal lines) by vertical dotted lines; BACs with yellow IDs were Y specific; BACs with blue IDs amplified both the Y and an autosome; BACs denoted with green lines were sequenced to close Gap 1; BACs denoted with black lines were new but not sequenced, and BACs with grey lines were previously sequenced for the eMSYv3 assembly.

Of these 15 BACs, 10 were subject to sequencing in addition with contig Ic terminal BAC 127N19 and contig II terminal BAC 169C6 with the intention of extending the BAC tiling path across Gap 1 (Table 13).

Further, multiple male-specific contigs from Felkel *et al.* (2019) fell within the terminal BACs of contig III which boarder Gap 2 (179K8 and 121G24) and Gap 3 (34A23). Therefore, we sequenced the terminal BACs flanking Gaps 2 and 3 to use the outermost sequences for bidirectional expansion into the gaps. However, BAC library screening with primers designed from the sequences of gap 2 and 3 flanking BACs 179K8, 121G24, and 34A23 did not result in identifying new Y chromosome BACs. In total, for closing the 3 gaps in eMSYv3, we sequenced 15 Y chromosome BACs (10 new and 5 from eMSYv3) on the Flongle but were able to close only Gap 1.

Table 13. Newly identified Y chromosome BACs and newly sequenced BACs used in this study.

BAC ID	FISH Location	Sequenced	Reference
022 E24	ECAY + ECA16	This study	This study
028P13	ECAY	n/a	This study
031I9	ECAY + ECA18	This study	This study
034A23	ECAY	This study	Janecka *et al*. 2018
043B24	ECAY + ECA18	n/a	This study
056J9	ECAY	n/a	This study
079D11	ECAY	This study	This study
088F22	ECAY + ECA18	n/a	This study
106F15	ECAY	This study	This study
121G24	ECAY	This study	Janecka *et al*. 2018
121H7	ECAY	This study	This study
127N19	ECAY	This study	Janecka *et al*. 2018
133 E10	ECAY + ECA18	This study	This study
142D23	ECAY	This study	This study
144G20	ECAY + ECA18	This study	This study
148N16	ECAY	This study	This study
169C6	ECAY	This study	Janecka *et al*. 2018
172K19	ECAY	n/a	This study
179K8	ECAY	This study	Janecka *et al*. 2018
183I5	ECAY + ECA18	This study	This study

FISH mapping and BAC sequence validation

Each new BAC identified through the CHORI BAC library screening and subsequent chromosome walking was validated by FISH. New BACs were hybridized to metaphase chromosomes together with a control, previously published, Y chromosome BAC clones (Figure 11 and Supplementary Figure S4). 7 of the 15 BACs hybridized to both the Y chromosome and an autosome. The latter was identified as ECA16 and ECA18 using ECA16- and ECA18-specific BAC clones from horse genome EquCab3 BAC track (NCBI genome: https://www.ncbi.nlm.nih.gov/genome/?term=domestic+horse). The remaining clones hybridized specifically to the Y chromosome. Interphase and DNA fiber-FISH were used to order the Y BACs across Gap 1 (Figure 11B and 11C). Dual color BAC combinations on interphase nuclei deduced the physical location of the new Y clones, and dual color probe combinations on mechanically stretched DNA fibers validated BAC partial overlaps in Gap 1.

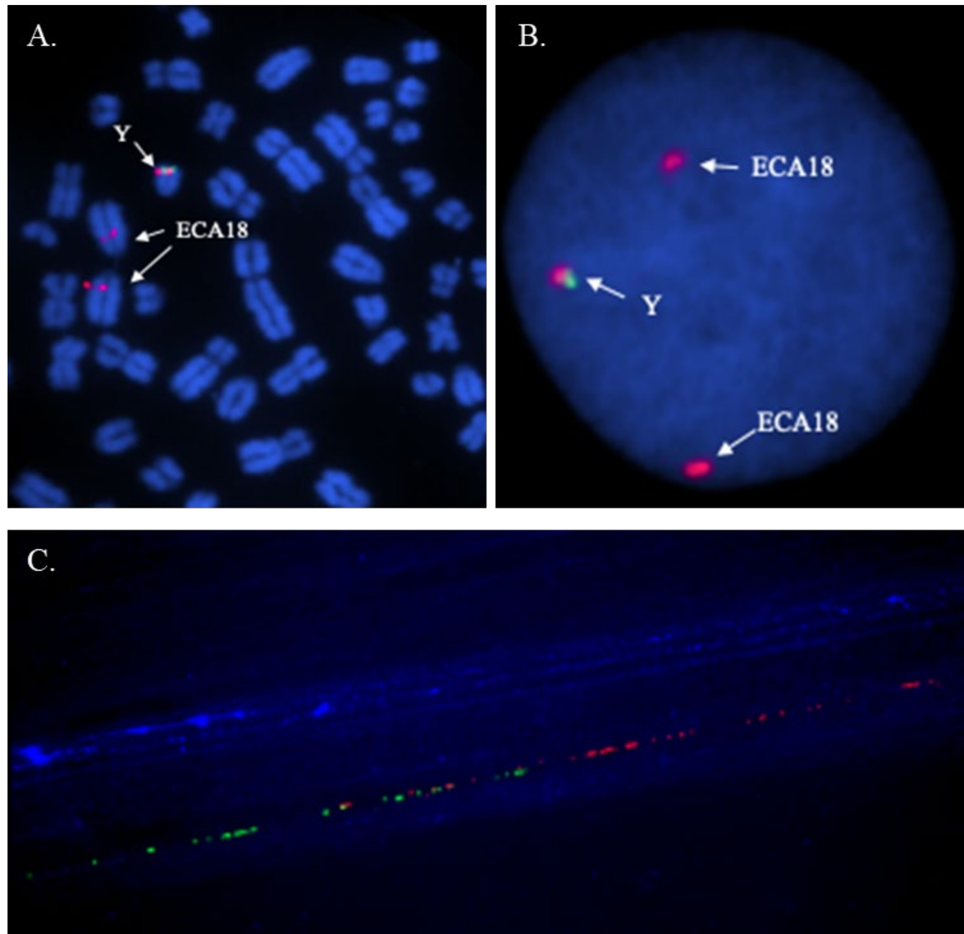


Figure 11. Examples of metaphase, interphase, and DNA Fiber-FISH used to assemble BAC clones across Gap 1. **A)** Partial metaphase spread of dual colored FISH probes; green signal represents the Y chromosome control BAC 54J7 and the red signal is the new BAC clone 183I5 that hybridizes to both the Y chromosome (Y) and chromosome 18 (ECA18). **B)** Interphase nuclei showing the proximity of terminal Contig I BAC clone 127N19 (green signal) and Gap 1 clone 183I5 (red signal). **C)** DNA Fiber-FISH showing that the terminal Contig I BAC clone 127N19 (green signal) has overlapping sequences with Gap 1 clone 183I5 (red signal).

3.3.3. Improving the current Y reference sequence eMSYv3 discussion

This study aimed to improve the Y chromosome reference assembly using a combination of high molecular weight DNA and ultra-long read sequencing technology, in addition to classical BAC-clone chromosome walking. While the single-copy portion of eMSYv3 reference is of high quality and a unique addition to horse whole genome reference, as well as an important tool for the horse genomics community, there are areas which need improvement. For example, the difference between the multi-copy Y (mcY) gene CN in the MSY reference and those quantified by the ddPCR assays in section 3.2.2 lead us to believe that the ampliconic region of the MSY is over assembled. Additionally, there are currently 3 gaps in the single copy region of the Y chromosome (scY) which need closing. The latter are likely part of the large number of contigs which are missing from eMSYv3 but identified in a recent publication regarding Y haplotypes (Felkel *et al.* 2019).

Ampliconic regions of the mammalian Y chromosome are incredibly repetitive and complex – they contain directional and inverted repeats, tandem repeats and palindromes with high sequence similarity and are, therefore a challenge for assembly even with the most advanced genomics technologies (Janecka *et al.* 2018; Teitz *et al.* 2018; Torresen *et al.* 2019). The mcY region of the horse MSY is assembled in eMSYv3 only tentatively. This assembly used a combination of Roche 454 medium-size reads, Illumina MiSeq short reads, and low coverage of PacBio long reads (Janecka *et al.* 2018). However, when utilized to sequence highly repetitive areas, these platforms have high error rates (Hu *et al.* 2021). Here, in order to increase the chances for proper assembly of the most complex MSY region, we sequenced all known mcY BAC clones and select scY BACs on the Flongle using Oxford Nanopore Technology platform. The platform generates ultra-long reads which not only span repetitive regions but may encompass the entire BAC clone in a single read (Hu *et al.* 2021). To maximize the reads

generated for each clone, we pooled no more than 4 clones per flow cell, still managing to maintain high coverage rates. Apart from the sequencing platform used in this study, *de novo* assembly and alignments of individual clones were very similar to the approaches used to generate eMSYv3 (Janecka et al. 2018).

While it was appealing to create a single *de novo* assembly using all the raw reads generated for the mcY clones, this may have easily collapsed the repetitive regions of the mcY, resulting in drastic under-assembly of the area as observed in our results. To prevent sequence collapse as much as possible, each clone was individually *de novo* assembled and the clones were methodically aligned to one another to create larger contigs using multiple alignment tools. Previously, individual clones were aligned and placed in a tentative tiling path and multi-copy genes were annotated by rigorous bioinformatic processes across the path (Janecka *et al.* 2018). Here, we generated the first large contiguous sequence of 1.53 Mb for the mcY and updated BAC tiling path (Supplementary Figure S2), creating the next steppingstone for improving the assembly of this complex region of horse MSY.

As the ddPCR experiments described in section 3.2 indicated a lower number of gene copies in the amplicon region compared to eMSYv3, we utilized these target sequences to annotate the newly assembled mcY region. In doing so, it became apparent that the 1.53 Mb mcY assembly generated in this study is under assembled, as only one or two copies of each mcY gene were present. Despite the efforts, multiple repetitive sequences collapsed during the *de novo* assembly process, resulting in a shorter and possibly less accurate assembly of mcY region than anticipated. Nevertheless, the generated 1.53 Mb assembly for the ampliconic region is an improvement over eMSYv3 (Janecka et al. 2018) and serves as a platform for further improvements. The latter will need the use of cutting-edge ultra-long sequencing platforms,

Bionano optical mapping technology and the rapidly increasing collection of publicly available whole genome sequences of male horses.

On the other hand, efforts to improve the scY region were more successful. Here, we obtained male specific contigs of sequences which were not incorporated into the MSY reference assembly (Felkel *et al.* 2019). By screening the contigs across the 94-BAC tiling path (Janecka *et al.* 2018), we identified areas within the published scY which needed improvement (Supplementary Figure S3). For example, sequences of BACs located in the junction of mcY and scY, between contigs Ib and Ic (BACs 209K10, 155M11, and 126G2, Figure 9), are likely tentatively assembled and possibly have duplicated sequences in the distal part of contig Ic (our unpublished data). This is one reason why these clones were sequenced alongside with known mcY clones. *De novo* assemblies of these 3 BACs were confirmed to be correct because they aligned well with some of the 179 newly found Y contigs by Felkel et al (2019). Likewise, multiple newly found Y contigs (Felkel et al. 2019) were placed in near-terminal or at the terminal ends of MSY contigs II and III (Supplementary Figure Figure 3) – an indication that there are more regions in the horse Y chromosome assembly that need improvement.

The single-copy sequence assembly in eMSYv3 is solely based on sequencing and assembly of a tiling path of overlapping BAC clones in these regions (Janecka *et al.* 2018). The reason there are gaps in the scY is because it was not possible to design primers from repetitive BES, thus stopping the search for new BAC clones for continuous chromosome walking (Janecka *et al.* 2018). Fortunately, a handful of new Y sequence contigs by Felkel *et al.* (2019) fell within Gap 1, the largest of the 3 gaps, spanning ~1 Mbp (Janecka *et al.* 2018). Combining chromosome walking by PCR with BAC-FISH at different resolution levels (metaphase, interphase, and DNA fiber) allowed for the close of Gap 1 and create a new BAC tiling path

which spans from the most distal BAC in contig Ic (127N19) to the most proximal clones in contig II (169C6 and 115I17). A total of 15 new Y BACs were identified, with 7 clones also sharing sequences with ECA16 (BAC 22E24) or ECA18 (31I9, 43B24, 88F22, 133E10, 144G20, 183I5). This is not completely unexpected because the presence of autosomal transposed and retro-transposed sequences in horse MSY has been described before (Janecka *et al.* 2018). The region in ECA18 (*EquCab3:* chr18: 13,498,631-13,803,111), which shares sequence homology with multiple MSY BACs, is very gene poor suggesting that no autosomal genes have been transposed to MSY from ECA18.

With regards to the BAC clone 022E24, which has sequence homology on both the Y chromosome and ECA16, it could not be determined which region of ECA16 has shared Y sequences. BLAST and D-GENIES alignment analysis of the *de novo* 022E24 clone sequence against eMSYv3 and mcY BACs suggests that the clone belongs within mcY and not scY, where the clone is identified. The sequence of 022E24 *de novo* assembly suggested a sample mix-up during either HMW DNA isolation or Flongle library preparation because select Y contigs from Felkel *et al.* (2019) used to identify 022E24 were not present in the *de novo* assembly. Re-sequencing of this BAC clone and additional FISH experiments are needed to both confirm the sequence of clone 022E24 and determine the approximate location of ECA16 homology identified through FISH. Unfortunately, no new BAC clones specific to Gaps 2 and 3 were identified and it is likely that multiple terminal clones of contigs II and III need to be re-sequenced to make progress in closing these two gaps.

In summary, improving the assembly of one of the most complex regions in the mammalian genome – the Y chromosome, is not an easy task. Nevertheless, the efforts of this work partially improved the assembly of mcY and closed one of the 3 gaps in scY. These are

small but important steppingstones for future progress. We are confident that with continued collaborative efforts and integration of state-of-the-art genomics tools, a new version of the horse Y chromosome assembly will become available in the upcoming years.

3.4. Y chromosomal rearrangements influence male fertility

The following sections, section 3.4.1, 3.4.2, and 3.4.3 will focus on the materials and methods, results, and a discussion about achieving the Objective 3 goal mentioned in 3.1. This work is done for the published work "*Molecular cytogenetic and Y copy number analysis of a reciprocal ECAY-ECA13 translocation in a stallion with complete meiotic arrest".* Section 3.4.3 includes additional discussion points about meiotic behavior of aberrant chromosomes, which were not included in the published paper.

3.4.1. Experimental methods and design*

Animal and Sample

Sodium heparin- and EDTA-stabilized blood samples were obtained from a previously described Friesian stallion (case ID H787) with azoospermia and a cytogenetically confirmed Y;13 reciprocal translocation (Ruiz *et al.* 2019).

Chromosome Preparations for Molecular Cytogenetic Analysis

* Reprinted from MDPI-GENES under Creative Commons CC BY 4.0 license: Castaneda C., Ruiz A.J., Tibary A. & Raudsepp T. (2021b) Molecular Cytogenetic and Y Copy Number Analysis of a Reciprocal ECAY-ECA13 Translocation in a Stallion with Complete Meiotic Arrest. *Genes* **12**.

Chromosome preparations were obtained from short-term blood lymphocyte cultures following standard procedures described elsewhere (Raudsepp & Chowdhary 2008a). Briefly, 1 mL of sodium heparin stabilized peripheral blood was mixed with 9 ml culture medium containing RPMI-1640 with Glutamax (Gibco), 30% fetal bovine serum (R&D Systems Inc.), 1 X antibiotic-antimycotic (Invitrogen, USA), and 1.4 µg/mL pokeweed mitogen (Sigma Aldrich). The cultures were grown for 72 h, harvested with demecolcine solution (final conc. 0.1µg/mL; Sigma Aldrich), treated with optimal hypotonic solution (Rainbow Scientific), and fixed in 3:1 methanol:acetic acid. Chromosome preparations were made on clean wet slides, air dried and stored at -20oC until needed.

Selection of Probes for FISH

CHORI 241 (CH241) Bacterial Artificial Chromosome (BAC) library (https://bacpacresources.org/, accessed on 18 September 2021) clones spanning horse (Equus caballus, ECA) chromosome 13 (ECA13) were identified from the CH241 genomic clone track of the horse reference genome EquCab3 in NCBI Genome (https://www.ncbi.nlm.nih.gov/genome/, accessed on 18 September 2021) or from the integrated physical map of the horse genome (Raudsepp *et al.* 2008). Information for horse Y chromosome BAC clones was retrieved from the BAC tiling path of the ECAY sequence map (Janecka *et al.* 2018) and BACs corresponding to the pseudoautosomal region (PAR) from the horse PAR BAC tiling path (Raudsepp & Chowdhary 2008b). The summary of information about the BAC clones used for FISH in this study is presented in Table 14. In addition to BACs, we used a biotin-labeled microdissected ECA13-specific painting probe (Raudsepp & Chowdhary 1999).

Table 14. Information about all BACs used for FISH in this study. Table 1 in (Castaneda et al. 2021b)

Marker ID	CH241 BAC	Cytogenetic location	Location in EquCab3 or ECAY BAC contig map	Reference marker	Reference
Y-1	069E11	Yq proximal 2/3	Y and Xq heterochromatin	*ETSTY7* ampliconic array	(Janecka *et al.* 2018)
Y-2	022P7	Yqdistal	Y Contig Ia, single copy	*KDM5D*	(Janecka *et al.* 2018)
Y-3	140M23	Yqdistal	Y Contig Ib, multi-copy (MC)	*SRY*	(Janecka *et al.* 2018)
Y-4	017D15	Yqdistal	Y Contig Ib, multi-copy (MC)	*TSPY*	(Janecka *et al.* 2018)
Y-5	090G18	Yqdistal	Y Contig Ic, single copy	n/a	(Janecka *et al.* 2018)
Y-6	112E12	Yqdistal	Y Contig II, single copy	*NLGN4Y*	(Janecka *et al.* 2018)
Y-7	011B8	Yqdistal	Y Contig II, single copy	n/a	(Janecka *et al.* 2018)
Y-8	125H6	Yqdistal	Y Contig III, single copy	*TMSB4Y*	(Janecka *et al.* 2018)
Y-9	102J15	Yqdistal	Y Contig III, single copy	*TMSB4Y*	(Janecka *et al.* 2018)
Y-10	106F1	Yqdistal	Y Contig IV single copy	*ZFY*	(Janecka *et al.* 2018)
PAR	194E12	Xpter/Yqter	chrX:3,945-246,703	PLCXD1	(Raudsepp & Chowdhary 2008b)
13-1	078E13	13p15	chr13:5,913,678-6,105,097	*GPER1*	This study
13-2	060D24	13p13	chr13:11,481,169-11,662,804	*ELN*	This study
13-3	158P20	13q12	chr13:18,059,033-18,254,381	LEX041	(Raudsepp *et al.* 2008)

Genomic and BAC DNA Isolation

Genomic DNA was isolated from EDTA-stabilized peripheral blood using the Gentra Puregene Blood Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. The DNA was checked for quality and quantity with the Nanodrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). ECA13, ECAY, and PAR BAC clones were picked from the CHORI 241 BAC library (https://bacpacresources.org/, accessed on 18 September 2021). The BACs were grown overnight in 100 mL 2YT (Life Technologies, Carlsbad, CA, USA) supplemented with 30 mg/mL chloramphenicol (Sigma Aldrich) and BAC DNA was isolated using the Plasmid Midiprep kit (Qiagen).

FISH analysis

Dual color FISH analysis was performed using the protocol outlined in previous sections using differently labeled combinations of two or three probes. The probes were labeled by nick translation either with biotin or digoxigenin using the BIO- or DIG-Nick Translation Mix (Roche, Basel, Switzerland), respectively. Biotin-labeled probes were detected with Alexa Fluor® 488 streptavidin conjugate (Molecular Probes, Life Technologies, Carlsbad, CA, USA) and digoxigenin-labeled probes with DyLight®594 anti-digoxigenin conjugate (Vector Laboratories, Burlingame, CA, USA). Chromosomes were counterstained with 4′,6-diamidino-2-phenylindole (DAPI). Images of at least 20 metaphases were captured and analyzed for each experiment using a Zeiss Axio Imager M2p fluorescent microscope and Isis v 5.2 (MetaSystems GmbH) software.

ddPCR CN analysis

This study performed ddPCR analysis of select mcY genes outlined in section 3.2. The results were compared with previously available MSY CN data for 16 normal control male horses (Castaneda *et al.* 2018; Castaneda *et al.* 2021b). A statistical analysis of CNVs between the Friesian stallion (H787, Table 15) and a control cohort of 16 males was conducted using methods described in section 3.2.

3.4.2. Results*

Molecular Cytogenetic Analysis of ECAY and ECA13 Reciprocal Translocation

The first set of FISH experiments determined the overall extent of the genetic exchange between ECAY and ECA13. We selected markers from the horse Y chromosome sequence map (Janecka *et al.* 2018) to represent the linear order of all the main regions in the horse Y chromosome, the proximal *ETSTY7* ampliconic array (Y heterochromatin), different MSY contigs which included the multi-copy region in contig Ib and the PAR (Figure 12A, Table 14). In a series of FISH experiments, we co-hybridized individual Y markers or pooled two or more markers from the same contig with a microdissected ECA13 painting probe. The results showed that the cells of the Friesian stallion carry a normal ECA13 and two derivative chromosomes designated as Y;13p and 13q;Y (Figure 12B). The derivative chromosome Y;13p was an acrocentric, which proximal part corresponded to Y chromosome material - *ETSTY7* ampliconic array and contigs Ia and MC-Ib, whereas the distal portion corresponded to ECA13 material

---

* Reprinted from MDPI-GENES under Creative Commons CC BY 4.0 license: Castaneda C., Ruiz A.J., Tibary A. & Raudsepp T. (2021b) Molecular Cytogenetic and Y Copy Number Analysis of a Reciprocal ECAY-ECA13 Translocation in a Stallion with Complete Meiotic Arrest. *Genes* **12**.

(Figure 12B). The second derivative chromosome 13q;Y was a small sub-metacentric with the long arm corresponding to ECA13 material and the short arm corresponding to the distal region of MSY and the PAR (Figure 12B). The only MSY region present on both derivative chromosomes was the multi-copy contig Ib, suggesting that the translocation breakpoint in ECAY is in contig Ib. Since these experiments used the whole chromosome painting probe, it was not possible to determine the translocation breakpoint in ECA13.

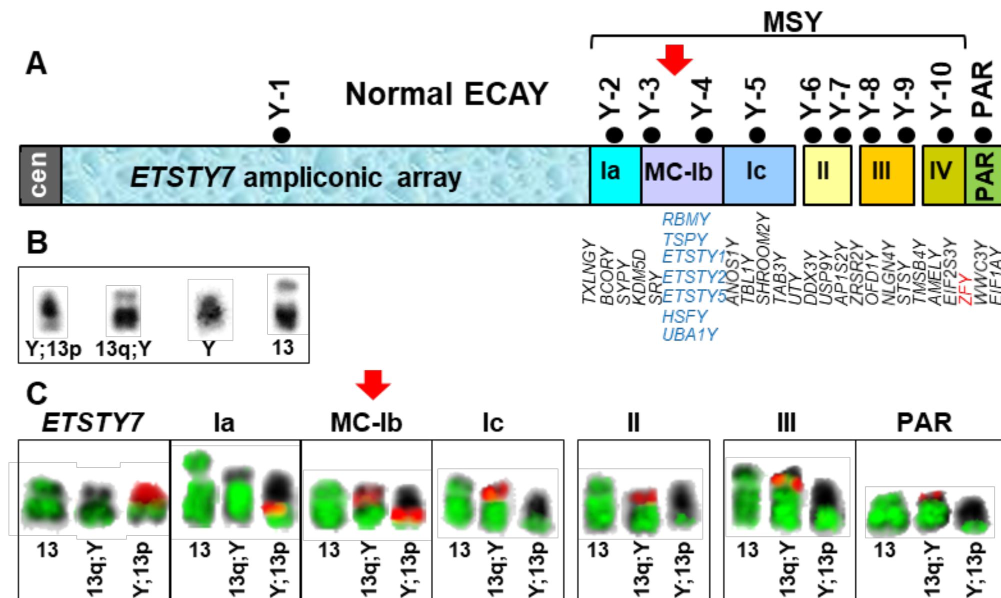


Figure 12. Determining the extent of genetic exchange between ECAY and ECA13. **(A)** ECAY sequence map (Janecka *et al.* 2018) showing Y heterochromatin (ETSTY7), contigs I-IV and the PAR; contig I is divided into three regions: single-copy Ia, multi-copy (MC) Ib, and single-copy Ic; contigs II-IV are single-copy; black dots with marker IDs above each region/contig correspond to the FISH markers used in this study (see Table 14); single copy gametologs are presented in map order below Y contigs in black font and sideways orientation; location of *ZFY* is highlighted in red font; contig MC-Ib multi-copy genes are in blue font and horizontally stacked with no known map order; **(B)** Inverted DAPI images of the derivative chromosomes Y;13p and 13q;Y, normal ECAY (Y) from a control horse, and normal ECA13 (13) **(C)** FISH

results with ECAY markers (red) and ECA13 painting probe (green) representing different regions of the chromosome. Note that only ECAY multi-copy contig MC-Ib marker hybridizes to both aberrant chromosomes, thus marking the translocation breakpoint in ECAY (red arrows). Figure 1 in (Castaneda *et al.* 2021b).

Next, we co-hybridized pairwise or in 3-probe combinations ECA13p and ECA13q markers (Table 14) with the ECAY markers (Table 14, Figure 12A). The FISH results confirmed and refined the initial findings with the ECA13 painting probe. We showed that ECA13p was located with the *ETSTY7* array and MSY contigs Ia and MC-Ib in derivative chromosome Y;13p (Figure 13A–D; Supplementary Figure S5), while ECA13q was with MSY contigs MC-Ib–IV and the PAR in derivative chromosome 13q;Y (Figure 13E,G,I,J). Based on this, we assigned the translocation breakpoint in ECA13 to the centromere. In line with the initial FISH results (Figure 12B), the translocation breakpoint in ECAY stayed in the multi-copy region since both MC-Ib markers, Y-3, and Y-4 (Figure 12A, Table 14), provided hybridization signals on both derivative chromosomes (Figure 13C,D). However, because of the multi-copy nature of these sequences, it was not possible to further narrow down the breakpoint in contig MC-Ib using FISH. This also means that we were not able to determine the location of the single copy equine *SRY* gene because it is embedded in the multi-copy sequences in marker Y-3 (BAC 140M23, Table 14) (Janecka *et al.* 2018).

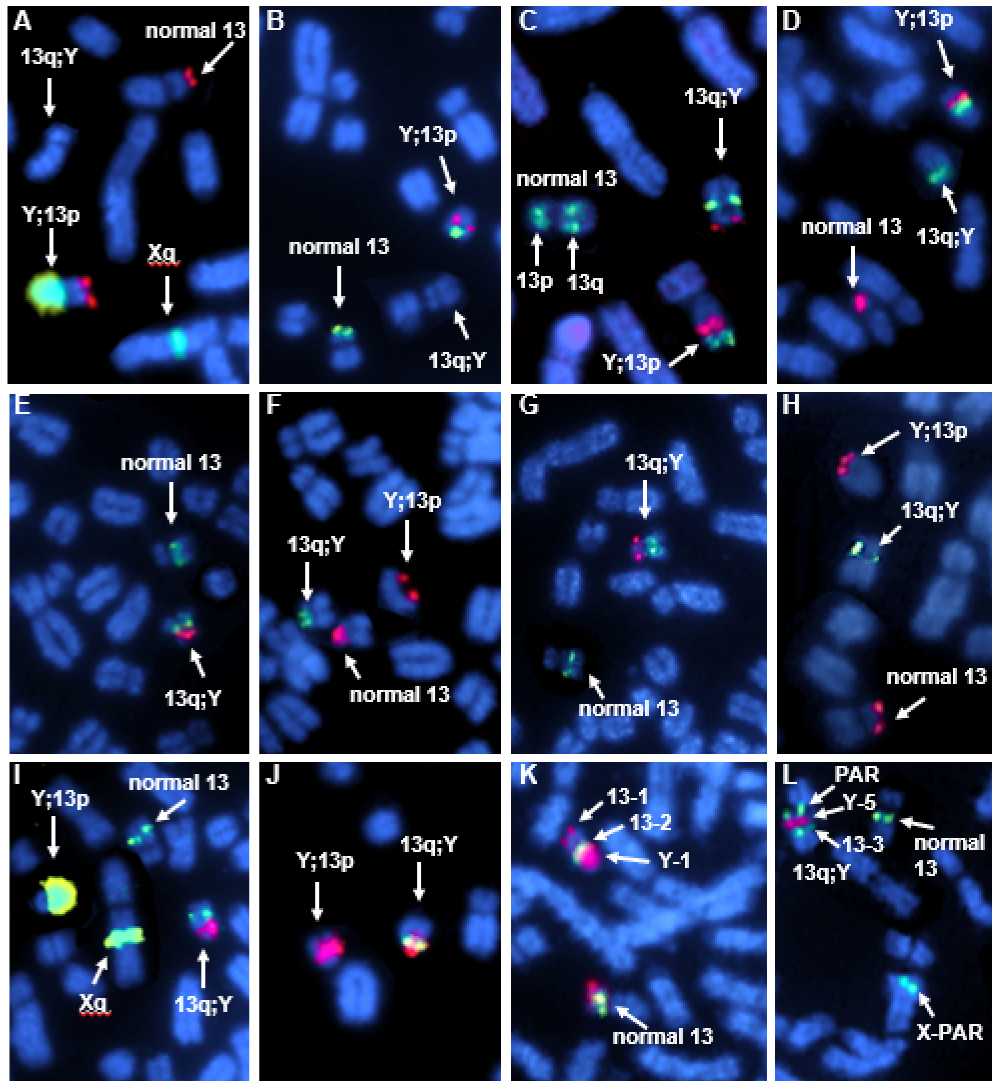


Figure 13. Partial metaphase spreads showing dual-color FISH results determining the size and orientation of translocated segments between ECAY and ECA13. **(A)** 13-1 red/ Y-1 green; full metaphase spreads of H787 and a control male horse with the results of this FISH experiment are presented in Supplementary Figure S5; **(B)** 13-1 green/ Y-2 red; **(C)** 13-2 green/ 13-3 green/ Y-3 red; **(D)** 13-1 red/Y-4 green; **(E)** 13-3 green/ Y-6 red; **(F)** 13-1 red/ Y-7 green; **(G)** 13-3 green/ Y-8 red**; (H)** 13-1 red/ Y-9 green**; (I)** Y-1 green/ Y-10 red/ 13-3 green**; (J)** Y-3 red/ Y-5 green**;** **(K)** Y-1 red/ 13-2 green/ 13-1 red**; (L)** 13-3 green/ Y-5red/ PAR green. See Figure 2 in (Castaneda et al. 2021b).

Finally, we used select markers to determine the orientation of reciprocally translocated segments of MSY and ECA13p (Figure 13K,L). We showed that in Y;13p, the ECA13p segment is attached to MSY contig MC-Ib by the proximal region with ECA13p15 remaining terminal in this derivative chromosome (Figures 13L and 13A). Likewise, the terminal end of the short arm of the derivative chromosome 13q;Y corresponded to PAR with MSY contigs IV-III-II-Ic-MC-Ib, located proximally (Figures 13K and 14B).

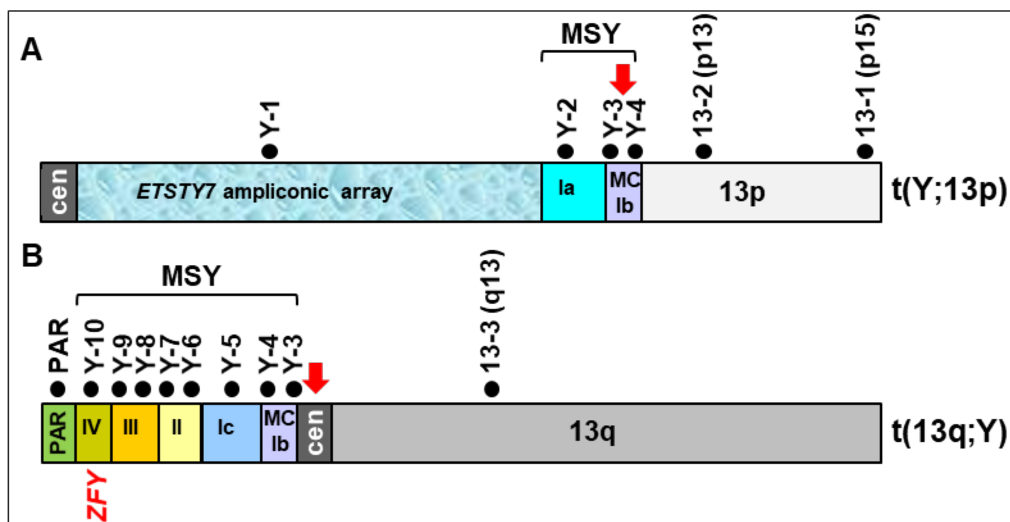


Figure 14. Schematic representation of the reciprocal translocation between ECAY and ECA13. **(A)** Aberrant chromosome Y;13p comprised of the proximal *ETSTY7* ampliconic array, MSY contigs Ia, part of MC-Ib and ECA13p; **(B)** Aberrant chromosome 13q;Y comprised of ECA13q and the distal portion of MSY, including part of multi-copy contig MC-Ib and single-copy contigs Ic, II, III, IV and PAR; black dots with marker IDs above each chromosome denote ECAY and ECA13 markers that were used for refined FISH analysis (see Table 14 and Figures 12, 13); red arrows indicate translocation breakpoints in ECAY contig MC-Ib and at ECA13 centromere; the location of *ZFY*, a candidate meiotic executioner gene, is indicated. See Figure 3 in (Castaneda *et al.* 2021b).

A summary of the ECAY-ECA13 reciprocal translocation is presented in Figure 14. The results show that the heterochromatic (*ETSTY7* ampliconic array) portion of ECAY together with a small single copy and partial multi-copy region have joined with the short arm of ECA13 (Figure 4A), whereas most of the single copy MSY together with partial multi-copy region and the PAR have relocated to join the long arm of ECA13 (Figure 14B). Translocation breakpoints were assigned to MSY multi-copy region and ECA13 centromere.

<u>Copy number analysis of horse MSY multi-copy genes</u>

After revealing that MSY multi-copy contig Ib sequences were present in both derivative chromosomes and that the translocation breakpoint in the Y chromosome was in the multi-copy region, we further studied this region for gene copy number (CN) variation to see whether the translocation had affected the CN of known MSY multi-copy genes. We determined absolute copy numbers of 7 MSY multi-copy and testis-specific genes (*TSPY*, *RBMY*, *ETSTY1*, *ETSTY2*, *ETSTY5*, *HSFY,* and *UBA1Y*) and *SRY* in the Friesian stallion and chromosomally normal control males. The latter also included the DNA donor for horse MSY reference assembly, a Thoroughbred *Bravo*. The results showed that CNs of 4 multi-copy genes (*RBMY, ETSTY2, HSFY* and *UBA1Y*) and the *SRY* were not statistically different between the Friesian stallion and controls (Table 15). However, CNs of a protein coding gene *TSPY* and two testis-specific transcripts, *ETSTY1* and *ETSTY5*, were significantly (P<0.05) lower in the Friesian stallion, with the most significant (P=0.004) CN reduction for *ETSTY5* (Table 15).

Table 15. CN analysis of seven MSY multi-copy genes and single copy SRY gene in the Friesian stallion (H787) and 16 control males.

Sample ID	*TSPY*	*ETSTY1*	*ETSTY2*	*ETSTY5*	*SRY*	*RBMY*	*HSFY*	*UBA1Y*
H787	5.6	1.9	3.8	2.6	0.8	1.9	0.9	3.1
23346	9.1	3.6	5.7	5.3	1	1.5	0.9	3.8
23348	8.2	7.4	5.7	4.8	1	2.3	0.9	3.5
70858	9.7	4.1	4.2	4.1	0.8	2	0.9	3.2
70980	10.8	3.9	5.1	4.4	0.9	1.7	0.9	3.8
70981	11.3	3.8	4.6	4.4	0.8	1.7	1.1	3
73901	8.6	3.9	4.1	3.9	1	2	1	3
74413	8.2	3.9	4.3	4.3	0.9	1.9	1	3.6
74836	8.2	3.7	4.6	4	0.8	1.7	1.2	3.3
74837	8.6	3.9	4.5	3.7	0.7	1.9	1	3.8
75052	9.2	3.6	4.6	4.2	0.9	2	1	3.7
TR007	9	4.2	3.8	4.1	1	1.9	1	3.9
TR008	9.5	4	3.7	4.1	0.9	2.1	1	4.3
TR009	8.2	3.9	3.8	3.8	0.9	2	1	4
H061	8	4.1	4.5	3.6	0.8	1.9	1.1	1.8
H294	13.3	3.8	5.4	4.4	0.8	1.7	1.1	3.3
Bravo	8.5	4.9	5.6	3.6	0.9	2	1.1	1.5
P-value	0.0249*	0.0293*	0.2494	0.0038**	0.4003	0.9756	0.2365	0.7583

CNs are rounded to the nearest tenth. CN values in red font indicate the lowest CN for each gene. *P*-values indicate whether CN in H787 was statistically significant; (*): $P < 0.05$; (**): $P < 0.01$. Sample ID *Bravo* - the DNA donor for MSY reference assembly. See Table 3 in (Castaneda *et al.* 2021b).

3.4.3. Discussion[*]

[*] Reprinted from MDPI-GENES under Creative Commons CC BY 4.0 license: Castaneda C., Ruiz A.J., Tibary A. & Raudsepp T. (2021b) Molecular Cytogenetic and Y Copy Number

Here we presented a detailed molecular cytogenetic analysis of a reciprocal translocation between ECAY and ECA13 in a Friesian stallion with azoospermia. This is the first and, to date, the only case of Y-autosome translocation, and one of the very few cytogenetically detectable Y chromosome structural rearrangements reported in horses (Herzog *et al.* 1989). Of the few cases with cytogenetically detectable Y chromosome abnormalities, there are two intersex horses with mosaicism for isochromosome Y (Raudsepp *et al.* 2010), a few cases of XY females with large Y chromosome deletions (Delobel *et al.* 1998) and a pony with abnormal external genitalia (no penis) and large deletion of the *ETSTY7* ampliconic array, also known as Y heterochromatin (Herzog *et al.* 1989). The low number of reported cases suggests that structural rearrangements of the Y chromosome are rare in horses. However, it is also possible that due to the small size of the Y chromosome, structural rearrangements easily remain undetected during conventional cytogenetic analysis. For example, in the present case of ECAY-ECA13 reciprocal translocation, the derivative chromosome 13q;Y was very similar in size, morphology and DAPI-banding to the normal ECA13 (Figure 12B) and the derivative chromosome Y;13p (Figure 12B) could have easily passed for a normal Y chromosome ((not present in this case; see (Ruiz *et al.* 2019)).

Analysis of a Reciprocal ECAY-ECA13 Translocation in a Stallion with Complete Meiotic Arrest. *Genes* **12**.

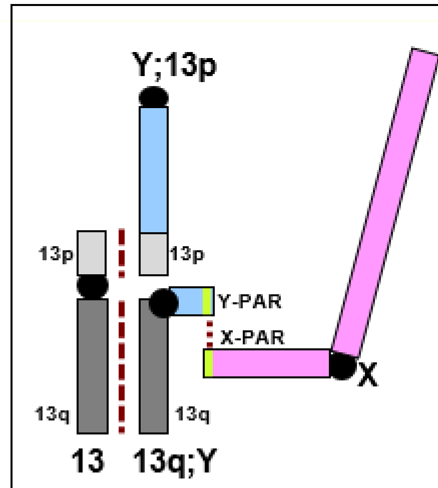


Figure 15. Schematic presentation of a hypothetical quadrivalent that may form in meiosis prophase of some primary spermatocytes of the Friesian stallion. ECA13p is shaded light grey; ECA13q is shaded dark grey; ECAY material is colored blue and ECAX is pink; PAR is light green; dotted maroon lines between homologous segments denote possible synapses.

However, regarding the phenotypic effect of Y-autosome translocations, it is perhaps even more important to determine the impact of meiotic configurations on MSCI, the extent of the sex body and functional regulation of genes residing in the chromosomes involved. Here again, we can only speculate based on elaborate immunogenetic studies of meiotic chromosomes and gene expression analyses of cases with Y-autosome translocations in other species. Both human (Skinner *et al.* 2016; Wang *et al.* 2017) and pig (Pinton *et al.* 2008; Barasc *et al.* 2012; Mary *et al.* 2018) studies show that Y-autosome translocations disturb the formation of sex body in meiosis prophase, and consequently, the MSCI process. Depending on the synaptic configurations formed in a particular cell, the sex body, which is immunogenetically visualized by the accumulation of histone variant γH2AX (Sciurano *et al.* 2007), can spread from sex chromosomes to the autosome and silence autosomal genes (Pinton *et al.* 2008; Mary *et al.* 2018). Alternatively, there may be cells with no sex body formation, resulting in no MSCI (Hsu

1994; Barasc *et al.* 2012; Wang *et al.* 2017; Mary *et al.* 2018). For example, in a study of Y-SSC13 translocation in an azoospermic boar, a sex body was found in only approximately 50% of cells (Mary *et al.* 2018). In either scenario, i.e., sex body spreading over autosomes vs. no sex body at all, Y-autosome translocations create a conflict during meiosis between the necessary transcriptional activity of autosomal genes and the obligatory silencing of the sex chromosomes. Therefore, one hypothesis trying to explain the meiotic arrest and azoospermia in these cases is that transcriptional silencing of certain regions in the autosomal genome leads to meiotic arrest (Bernasconi *et al.* 1999; Skinner *et al.* 2016; Wang *et al.* 2017). However, the autosomes involved in translocations with the Y chromosome in humans, pigs, cattle and in the present equine case are not completely comparable, as these translocations involve different genes. For example, reported Y-autosome translocations in humans involve all autosomes except HSA20 (Royo *et al.* 2010; Waters & Ruiz-Herrera 2020a), pig cases involve autosomes SSC1 (Barasc *et al.* 2012), SSC13 (Barasc *et al.* 2012; Mary *et al.* 2018) and SSC14 (Pinton *et al.* 2008), and cattle cases just two autosomes – BTA9 (Iannuzzi *et al.* 2001) and BTA21 (Switonski *et al.* 2011). Therefore, if autosomal factors are responsible for meiotic arrest in these cases, it must be due to transcriptional silencing of autosomal genes *per se* and not due to specific genes or regions.

On the other hand, while Y-autosome translocations involve non-homologous autosomes in different species, all cases involve the Y chromosome. Therefore, and as shown by several studies (Raudsepp & Chowdhary 2008a; Barasc *et al.* 2012; Cortez *et al.* 2014; Villagomez *et al.* 2017; Mary *et al.* 2018), a more plausible explanation for meiotic arrest is the failure to properly inactivate the Y chromosome. This is in line with a recently presented theory about “the persistent Y” and “meiotic executioner genes” (Vernet *et al.* 2014). The theory provides a

mechanistic explanation why the eutherian Y chromosome persists, despite millions of years of degeneration during evolution. The theory speculates that the Y-linked meiotic executioner genes play a critical role and are necessary for not only meiotic success but also to regulate their own silencing and therefore, must be subjected to MSCI. Abnormal expression of these executioner genes during the silencing window in cases of Y-autosome translocations, will result in fatal meiotic arrest (Raudsepp & Chowdhary 2008a; Cortez *et al.* 2014). In turn, meiotic arrest prevents the transmission of translocations and as a result, the Y chromosome persists (Vernet *et al.* 2014). The "persistent Y theory" proposes *ZFY* as a likely candidate for a "meiotic executioner" gene because, firstly, it is among the few genes found in all eutherian Y chromosomes (Bellot *et al.* 2014; Janecka *et al.* 2018; Hughes *et al.* 2020), and secondly, it is the only conserved eutherian Y gene where aberrant *ZFY* expression during MSCI is pachytene lethal (Skaletsky *et al.* 2003). In contrast, ectopic expression of other conserved eutherian Y genes such as *RBMY, UTY, DDX3Y* and *SRY* does not induce pachytene arrest (Cortez *et al.* 2014; Vernet *et al.* 2014). In the present equine case, there was no possibility to study meiosis cytogenetically or for gene expression. Therefore, we can only speculate based on the translocation breakpoints and the genetic content of the two derivative chromosomes (Figure 14). We propose that meiotic arrest and azoospermia in this Friesian stallion are due to the translocation of *ZFY*-containing portion of ECAY to ECA13q (Figure 14B), and that complete synapsis between normal ECA13q with the derivative 13q;Y (Figure 15) prevented silencing of the Y portion with the *ZFY* gene.

The theory about meiotic executioner genes with *ZFY* as the primary candidate (Vernet *et al.* 2014), also explains why some cases of Y-autosome translocations in animals (Switonski *et al.* 2011) and humans (Benitez *et al.* 1979; Sun *et al.* 2005) do not result in meiotic arrest. For

example, the two published Y-autosome translocations in cattle (Iannuzzi *et al.* 2001; Switonski *et al.* 2011) have different phenotypes. The case with azoospermia (Iannuzzi *et al.* 2001) had a reciprocal translocation between Y and BTA9, so that the two derivative chromosomes comprised of parts of the Y and large portions of BTA9. While the authors proposed that azoospermia was caused by the production of unbalanced gametes due to the formation of quadrivalent configurations in meiosis (Iannuzzi *et al.* 2001), an alternative explanation is the failure to silence the *ZFY* gene in BTAYp (Hamilton *et al.* 2012) which was translocated to the distal half of BTA9. On the other hand, the Y-autosome translocation in a reproductively normal bull (Switonski *et al.* 2011) was non-reciprocal, so that one derivative chromosome comprised of BTA21 and BTAYq, while BTAYp with *ZFY* and PAR (Hamilton *et al.* 2012) remained on a separate chromosome and could easily undergo MSCI. Likewise, in humans where *ZFY* is located in HSAYp (Krausz *et al.* 2010), non-reciprocal translocations of HSAYq to an autosome do not affect the phenotype or fertility (Benitez *et al.* 1979; Sun *et al.* 2005), while balanced reciprocal Y-autosome translocations involving euchromatic portions of the Y, including HSAYp, result in oligozoospermia or azoospermia (Benitez *et al.* 1979; Cribiu *et al.* 2001; Wang *et al.* 2017). Therefore, to evaluate the genetic and phenotypic consequences of Y-autosome translocations, it is necessary for not only proper analysis of the cytogenetic features but also to acquire knowledge about the organization of the Y chromosome of the species in question.

Lastly, we asked whether the presented case of a reciprocal translocation between ECAY and ECA13 was balanced. For this, we first tried to pinpoint the translocation breakpoint. This appeared to be difficult because the breakpoint was in the multi-copy region of Y and BAC-FISH produced hybridization signals in both derivative chromosomes (Figure 12C, Figure 13C, D), thus confounding the precise demarcation of the breakpoint. To obtain more information

about the breakpoint region, we evaluated copy numbers (CN) of 7 known ECAY multi-copy genes and the *SRY* as described in earlier sections. Analysis showed significant CN reduction for three genes/transcripts: *TSPY, ETSTY1* and *ETSTY7* (Table 15). These findings may suggest that the translocation was accompanied by the loss of some multi-copy sequences and was, thus, not truly balanced. Though, it is also possible that the observed CN variation was specific to the individual or the breed and needs further investigation. The *TSPY* gene has been associated with male fertility in cattle (Hamilton *et al.* 2012) and humans (Krausz *et al.* 2010), however, functional significance of *TSPY* copy number variation in stallions or other species is not known. Additionally, possible functions or protein coding potential of equine testis-specific transcripts, *ETSTY1* and *ETSTY5,* are not yet known (Das *et al.* 2012; Janecka *et al.* 2018).

In summary, molecular cytogenetic characterization, and copy number analysis of the first Y-autosome reciprocal translocation in horses adds a new case to equine clinical cytogenetics but also presents important information for better understanding the functions of Y chromosome genes and sex chromosome regulation in male meiosis.

4. CONCLUDING REMARKS

While it is well established that mammalian male fertility has a genetic component, genomics tools for evaluating fertility in stallions are limited. Therefore, there are great expectations on the horse genome reference, *EquCab3*, and the Y chromosome reference assembly, *eMSYv3*. Combined, these resources are critical for dissecting the genomic factors which affect stallion fertility. This study touches on a single autosomal factor, providing compelling evidence for a significant association between low fertility rates in Thoroughbred stallions and the combined A/A–A/A genotype of SNPs chr13:11 353 372G>A and chr13:11 353 436A>C in *FKBP6* exon 5, thus confirming and refining earlier findings. This is currently the most extensive population genetic analysis of *FKBP6*, the susceptibility gene for impaired acrosome reaction, in Thoroughbreds. However, despite the confirmed genotype–phenotype association, we still do not know the underlying molecular causes because the associated sequence variants are not causative. Therefore, the research has continued with detailed WGS-based analysis of Thoroughbreds and individuals of other breeds with the combined A/A–A/A genotype to identify breed-specific haplotype blocks or case-specific variants around the *FKBP6* locus and other genomic signatures unique to Thoroughbred stallions with confirmed IAR or idiopathic subfertility. Efforts are ongoing to include testis RNA sequencing data and gene expression levels to compare IAR Thoroughbreds and control males.

In addition to the autosomal factor *FKBP6*, the Y chromosome is enriched with genes important for spermatogenesis and sperm functions as evidenced by human and mouse studies which show causative links between Y chromosome rearrangements, mutations, CNVs and various male infertility phenotypes. Therefore, Y chromosome studies were initiated in the horse to identify important male-specific factors which result in subfertility/infertility. Analysis of

other species suggest the importance of male-specific, multi-copy genes as changes in the gene copy number can influence the fertility phenotype of an individual. Here we investigated CNVs using ddPCR technology, which is currently the best platform for absolute quantitation of gene copy numbers, in horse MSY multi-copy genes across a global cohort to lay a foundation for CNV studies in subfertile stallions.

We showed that droplet digital PCR (ddPCR) is a reliable approach for copy number analysis of horse MSY multi-copy genes and provides a more accurate CN evaluation compared to the current assembly of the ampliconic region in MSY reference eMSYv3 (Janecka *et al.* 2018). Gene CN analysis in a large multi-breed population of normal male horses showed that most multi-copy MSY genes are CN variable between individuals, breeds, but also in successive male generations and horses produced by SCNT. This suggests that MSY gene CNVs are caused by both meiotic and mitotic events and are mechanistically different from single nucleotide variants that are rare and determine Y chromosome haplotypes. Therefore, MSY CNV patterns are not correlated with haplogroups and haplotypes. Further studies are needed to determine selective constraints over horse MSY gene CN and how this relates to equine male development and fertility. For this and for the inclusion of the ampliconic genes that were missed in this study, the sequence assembly of the horse MSY ampliconic region must be improved. This will require a combined use of cutting-edge platforms for the assembly of complex genomic regions such as PacBio single-molecule, high-fidelity, long-read sequencing (Vollger *et al.* 2020) and Bionano optical mapping (Bocklandt *et al.* 2019). An improved MSY ampliconic assembly is also the prerequisite for functional annotation of these genes to determine their role in stallion reproduction and male biology.

While the current MSY reference is a valuable resource to the horse community, there are areas which need improvement. Here, efforts were made to improve multiple regions of the current horse Y chromosome assembly utilizing ultra-long read sequencing techniques and classical chromosome walking. This study identified novel Y chromosome BACs and added ~1Mbp of sequence to the MSY. Finally, this study describes the first-ever Y-autosome translocation in horses, suggesting that like in humans and mouse, the horse Y also carries sequences critical for normal spermatogenesis. On the other hand, it is an excellent proof-of-principle that the Y chromosome is critical for normal spermatogenesis in stallions, thus encouraging continuing research in the field. Most importantly, the available tools and recourses, together with the discovery of functional elements through collaborative efforts are dissecting the genomic factors which affect complex traits, such as stallion fertility. The improved understanding of the molecular underpinnings of these traits continue to benefit the horse community, and as a result, the equine industry.

REFERENCES

American Horse Council. URL https://www.horsecouncil.org/resources/economics/.

Amann R.P. & Veeramachaneni D.N.R. (2007) Cryptorchidism in common eutherian mammals. *Reproduction* **133**, 41-561.

Andersson L., Archbald A.L., Bottema C.D., Brauning R., Burgass S.C., Burt D.W., Casas E., Cheng H.H., Clarke L., Caouldrey C., Dalrymple B.P., Elsik C.G., Foissac S., Giuffra E., Groenen M.A., Hayes B.J., McEwan J.C., Moore S., Nanduri B., Notredame C., Palti Y., Plastow G.S., Reecy J.M., Rohner G.A., Sarripoulou E., Schmidt C.J., Silverstein J., Tellam R.L., Tixier-Boichard M., Tosser-kloop G., Tuggle C.K., BVilkki J., White S.N., Zhao S. & Zhou H. (2015) Coordinated international action to accelerate genome-to-phenome with FAANG, the functional annotation of animal genomics project. *Genome Biology and Evolution* **16**.

Barasc H., Mary N., Letron I.R., Calgaro A., Dudez A., Bonnet N., Lahbib-Mansais Y., Yerle M., Ducos A. & Pinton A. (2012) Y-autosome translocation interferes with meiotic sex inactivation and expression of autosomal genes: A case study in the pig. *Sexual Development* **6**, 143-50.

Bellot D.W., Cho T., Hughes J.F., Skaletsky H. & Page D.C. (2018) Cost-effective, high-throughput, single-haplotype iterative mapping ans sequencing for complex genomic structures. *Nature Protocols* **13**, 787-809.

Bellot D.W., Hughes J.F., Skaletsky H., Brown L.G., Pyntikova T., Cho T., Koutseva N., Zaghlul S., Graves T., Rock S., Kremitzki C., Fulton R., Dugan S., Ding Y., Morton D., Khan Z., Lewis L., Buhay C., Wang Q., Watt J., Holder M., Lee S., Nazareth L., Alfoldi J., Rozen S.,

Muzn D.M., Warren W.C., Gibbs R.A., Wilson R.K. & Page D.C. (2014) Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494-9.

Benitez J., Rivera L., Ramos C., Tejedor E. & Sanchez-Cascos A. (1979) Translocation of a supernumerary Y to a 15: Study of six cases (three males and three females) in three generations. *Quality of Life Research* **48**, 191-4.

Bernasconi P., Cavigliano P.M., Boni M., Malcovati L., Calatroni S., Astori C., Caresana M. & Bernasconi C. (1999) A novel t(Y;11) translocation with MLL gene rearrangement in a case of acute myelomonocytic leukemia (AML-M4). *Leukemia* **13**, 487-9.

Beruitex J., Rivera L., Ramos C., Tejedor E. & Sanchez-Cascos A. (1979) Translocation of a supernumerary Y to 15: study of six cases (three males and three females) in the generations. *Human Genetics* **48**, 191-4.

Bhatia S. & Shankar V. (1992) First report of a XX/XXY fertile goat buck. *Vet Records* **130**, 271-2.

Blanchard T.L. & Varner D.D. (1997) Evaluation breeding soundness in stallions. *Veterinary Medicine*, 815-8.

Bocklandt S., Hastie A. & Cao H. (2019) Bionano genome mapping: High-throughput, ultra-long molecular genome analysis system for precision genome assembly and haploid-resolved structural variation discovery. In: *Advances in experimental medicine and biology* (ed. by Suzuki Y). Springer, Singapore.

Bolzon C., Joone C.J., Schulman M.L., Harper C.K., Villagomez D.A.F., King W.A. & Revay T. (2016) Missense mutation in the ligan-binding domain of the horse androgen receptor gene in a Thoroughbred family with inherited 64,XY (SRY+) disorder of sex development. *Sexual Development* **10**, 37-44.

Bonomi M., Rochira V., Pasquali D., Balercia G., Jannini E.A., Ferlin A. & Italia N.G.K. (2017) Klinefelter syndrome (KS): Genetics, clinical phenotype and hypogonadism. *Journal of Endocrinology Investigations* **40**, 123-34.

Brinsko S.P., Love C.C., Bauer J.E., Macpherson M.L. & Varner D.D. (2007) Cholesterol-to-phospholipid ratio in whole sperm and seminal plasma from fertile stallions and stallions with unexplained subfertility. *Animal Reproduction Science* **99**, 65-71.

Brito L.F.C., Sertich P.L., Durkin K., Chowdhary B.P., Turner R.M., Greene L.M. & McDonnell S. (2008) Autosomic 27 trisomy in a standardbred colt. *Journal of Equine Veterinary Science* **28**, 431-6.

Bugno-Poniewierska M. & Raudsepp T. (2021) Horse Clinical Cytogenetics: Recurrent Themes and Novel Findings. *Animals (Basel)* **11**, 831.

Buoen L.C., Zhang T.Q., EWeber A.F., Turner T., Bellamy J. & Ruth G.R. (1997) Arthrogryposis in the foal and its possible relation to autosomal trisomy. *Equine Veterinary Journal* **29**, 60-2.

Cabanettes F. & Klopp C. (2018) D-GENIES: dot plot large genomes in an interactice, efficient and simple way. *PeerJ* **6**.

Carrell D.T. (2007) The genetics of male infertility in the era of genomics. In: *The Genetics of Male Infertility* (pp. 3-27. Humana press, Totowa, New Jersey.

Castaneda C., Hillhouse A., Janecka J., Juras R., Ruiz A.J., Tibary A., Love C.C., Varner D.D. & Raudsepp T. (2018) Contribution of the Y chromosome to stallion fertility. *Journal of Equine Veterinary Science* **66**, 30.

Castaneda C., Juras R., Khanshour A., Randlaht I., Wallner B., Rigler D., Lindgren G., Raudsepp T. & Cothran E.G. (2019) Population Genetic Analysis of the Estonian Native Horse

Suggests Diverse and Distinct Genetics, Ancient Origin and Contribution from Unique Patrilines. *Genes* **10**.

Castaneda C., Juras R., Kjollerstrom J., Hernandez Aviles C., Teague S.R., Love C.C., Cothran E.G., Varner D.D. & Raudsepp T. (2021a) Thoroughbred stallion fertility is significantly associated with FKBP6 genotype but not with inbreeding or the contribution of a leading sire. *Animal Genetics* **52**, 813-23.

Castaneda C., Ruiz A.J., Tibary A. & Raudsepp T. (2021b) Molecular Cytogenetic and Y Copy Number Analysis of a Reciprocal ECAY-ECA13 Translocation in a Stallion with Complete Meiotic Arrest. *Genes* **12**.

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

Colenbrander B., Gadella B.M. & Stout T.A.E. (2003) The predictive value of semen analysis in the evaluation of stallion fertility. *Reproduction in Domestic Animals* **38**, 305-11.

Consortium T.E.P. (2004) The ENCODE (ENCylopedia Of DNA Elements) Project. *Science* **306**, 363-640.

Consortium T.E.P., Moore J.E., Purcaro M.J., Pratt H.E., Epstein C.B., Shoresh N., Adrian J.K., T., CDavis C.A., Dobin A., Kaul R., Halow J., BVan Nostrand E.L., Freese P., Gorkin D.U., Shen Y., He Y., Mackiewicz M., Pauli-Behn F., Williams B.A., Mortazavi A., Keller C.A., Zhang X., Elhajjajy S.I., Huey J., Dickel D.E., Snetkova V., Wei X., Wang X., Rivera-Mulia J.C., Rozowsky J., Zhang J., Chhetri S.B., Zhang J., Victorsen A., White K.P., Visel A., Yeo G.W., Burge C.B., Lecuyer E., Gilbert D.M., Dekker J., Rinn J., Mendenhall E.M.,

Ecker J.R., Kellis M., Klein R.J., Noble W.S., Kundaje A., Guigo R., Farnham P.J., Cherry J.M., Myers R.M., Ren B., Graveley B.R., Gerstein M.B., Pennacchio L.A., Snyder M.P., Bernstein B.E., Wold B., Hardison R.C., Gingeras T.R., Stamatoyannopoulos J.A. & Weng Z. (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710.

Cortez D., Marin R., Toledo-Flores D., Froidevaux L., Liechti A., Waters P.D., Grutzner F. & Kaessman H. (2014) Origins and functional evolution of Y chromosomes across mammals. *Nature* **24**, 488-93.

Cribiu E.P., Di Berardino D., Di Meo G., Eggen A., Gallagher D.S., Gustavsson I., Hayes H., Iannuzzi L., Popescu C.P., Rubes J., Schmutz S., Stranzinger G., Vaiman A. & Womack J.E. (2001) International system for chromosome nomenclature of domestic bovids (ISCNDB 2000). *Cytogenetic Genome Research* **92**, 283-99.

Dain A.R. & Bridge P.S. (1978) A chimaeric calf with XY/XXY mosaicism and intersexuality. *Journal of Reproduction and Fertility* **54**, 197-201.

Danecek P., Auton A., Abecasis G.R., Albers C.A., Banks E., DePristo M.A., Handsaker R., Lunter G., Marth G., Sherry S.T., McVean G., Durbin R. & Group G.P.A. (2011) The variant call format and VCFTools. *Bioinformatics*.

Darling A.C., Mau B., Blattner F.R. & Perna N.T. (2004) Mauve: multiple alignment of conserved genomic sequences with rearrangements. *Genome Research* **14**, 1394-403.

Das P.J., Lyle S.K., Chowdhary B.P. & Raudsepp T. (2012) Cytogenetic and Molecular Characterization of Y isochromosome in a 63XO/64Xi(Yq) mosaic Karyotpe of an intersex horse. *Sexual Development* **6**, 117-27.

Delobel R., Djlelati R., Gabriel-Robez O., Croquette M.-F., Rousseaux-Prevost R., Rousseaux J., Rigot J.-M. & Rumpler Y. (1998) Y-autosome transclocation and infertility: usefulness of molecular, cytogenetic and meiotic studies. *Quality of Life Research* **102**, 98-102.

Donnelly C.G., Bellone R.R., Hales E.N., Nguyen A., Katzman S.A., Dujovne G.A., Knickelbein K.E., Avila F., Kalbfleisch T.S., Giulotto E., Kingsley N.B., Tanaka J., Esdaile E., Peng S., Dahlgren A., Fuller A., Mienaltowski M.J., Raudsepp T., Affolter V.K., Petersen J.L. & Finno C.J. (2021) Generation of a biobank from two adult Thoroughbred stallions for the Functional Annotation of Animal Genomes initiative. *Frontiers in Genetics*.

Durkin K., Raudsepp T. & Chowdhary B. (2011) Cytogenetic evaluation of the stallion. In: *Equine Reproduction* (eds. by Vaala WE & Varner DD), pp. 1462-8. Wiley Blackwell, Chichester, UK.

Durward-Akhurst S.A., Scaefer R.J., Grantham B., Carey W.K., Mickelson J.R. & McCue M.E. (2021) Genetic variation and the distribution of variant types in the horse. *Frontiers in Genetics*.

Edwards C.J., Ginja C., Kantanen J., Perez-Pardal L., Tresset A., Stock F., Consortium E.C.G.D., T. G.L., Penedo M.C.T., Bradley D., G,, Lenstra J.A. & Nijam I.J. (2011) Dual Origins of Dairy Cattle Farming – Evidence from a Comprehensive Survey of European Y-Chromosomal Variation. *PLOS One* **6**.

Ellis P.J.I., Clemente E.J., Ball P., Toure A., Ferguson L., Turner J.M.A., Loveland K.L., Affara N.A. & Burgoyne P.S. (2005) Deletions on mouse Yq lead to upregulation of multiple X- and Y-linked transcripts in spermatids. *Human Molecular Genetics* **14**, 2705-15.

Fawcett J.A., Sato F., Sakamoto T., Iwasaki W.M., Tozaki T. & Innan H. (2019) Genome-wide SNP analysis of Japanese Thoroughbred racehorses. *PLOS One* **14**, e0218407.

Felkel S., Vogl C., Rigler D., Dobretsberger V., Chowdhary B.P., Distl O., Fries R., Jagannathan V., Janecka J.E., Leeb T., Lindgren G., McCue M., Metzger J., Neuditschko M., Rattei T., Raudsepp T., Rieder S., Rubin C.J., Schaefer R., Schlotterer C., Thaller G., Tetens J., Velie B., Brem G. & Wallner B. (2019) The horse Y chromosome as an informative marker for tracing sire lines. *Scientific Reports* **9**.

Felkel S., Vogl C., Rigler D., Jagannathan V., Leeb T., Fries R., Neuditschko M., Rieder S., Velie B., Lindgren G., Rubin C.-J., Schlotterer C., Rattei T., Brem G. & Wallner B. (2018) Asian horses deepen the MSY phylogeny. *Animal Genetics* **49**, 90-3.

Fritch M.C., Mitsuhashi A. & Katoh K. (2021) lamassemble: Multiple alignment and consensus sequence of long reads. *Methods in Molecular Biology* **2231**.

Fujihara Y., YLu Y., Noda T., Oji A., Larasati T., Kojima-Kita K., Yu Z., Matzuk R.M., Matzuk M.M. & Ikawa M. (2020) Spermatozoa lacking fertilization influencing membrane protein (FIMP) fail to fuse with oocytes in mice. *PNAS* **117**, 9393-400.

Fukami K., Nakao K., Inoue T., Kataoka Y., Kurokawa M., DFissore R.A., Nakamura K., Katsuki M., Mikoshiba K., Yoshida N. & Takenawa T. (2001) Requirement of phospholipase cdelta4 for the zona pellucida-induced acrosome reaction. *Science* **292**, 920-3.

Garbe R.A. & Da Y. (2003) A software tool for the graphical visualization of large and complex populations. *Yi Chaun Xue Bao* **30**, 1193-5.

Gekas J., Thepot F., Turleau C., Siffroi J.P., Dadoune J.P., Briault S., Rio M., Bourouillou G., Caree-Pigeon F., Wasels R., Benzackn B. & Francaise A.d.C.d.L. (2001) Chromosomal factors of infertility in candidate ouples for ICSI: an equal risk of constitutional aberrations in women and men. *Human Reproduction* **16**, 82-90.

Ghenu A.H., Bolker B.M., Melnick D.J. & Evans B.J. (2016) Multi-copy gene family evolution on primate Y chrosomes. *BMC Genomics* **17**, 157.

Gibbs G.M., Orta G., Reddy T., Koppers A.J., Martinez-Lopez P., Luis de la Vega-Beltran J., Lo J.C., Veldhuis N., Jamsai D., McIntyre P., Darszon A. & O'Bryan M.K. (2011) Cysteine-rich secretory protein 4 is an inhibitor of transient receptor potential m8 with a role in establishing sperm function. In: *National Academy of Sciences of the United States of America* pp. 7034-9.

Giese A., Jude R., Kuiper H., Piumi F., Schambony A., Guerin G., Distl O., Topfer-Petersen E. & Leeb T. (2002) Molecular characterization of the equine AEG1 locus. *Gene* **292**, 65-72.

Giesecke K., Hamann H., Sieme H. & Distl O. (2010a) Evaluation of prolactin receptor (PRLR) as candidate gene for male fertility in Hanoverian warmblood horses. *Reproduction in Domestic Animals* **45**, e124-30.

Giesecke K., Hamann H., Stock K.F., Woehlke A., Sieme H. & Distl O. (2009) Evaluation of SPATA1-associated markers for stallion fertility. *Animals Genetics* **40**, 359-65.

Giesecke K., Sieme H. & Distl O. (2010b) Infertility and candidate gene markers for fertility in stallions. *Veterinary Journal* **185**.

Giuffra E., Tuggle C.K., Archibald A.L., Acloque H., Chamberlain A.J., Cochrane G., Daetwyler H.D., Djebali S., Eory L., Foissac S., Goddard M.E., Groenen M.A.M., Halstead M., Harrison P.W., Hayes B.J., Kern C., Khatib H., Kuo R.I., MacHugh D.E., Macleod I.M., Madsen O., Reecy J.M., Ross P.J., Watson M., Wells J.M. & Zhou H. (2019) Functional annotation of animal genomes (FAANG): Current achievements and roadmap. *Annual Review of Animal Biosciences* **7**, 65-88.

Glugovsci N., Birtriceanu M., Suciu A. & Bratu M. (1970) A case of intersexuakity in te horse with type 2A+XXXY chromosome formula. *British Veterinary Journal* **126**, 552-25.

Gosh S., Arnold C., Wade C.M., Leeb T., Distl O., Chowdhary B., Varner D.D. & Raudsepp T. (2016) AKR1C genes as candidate loci for equine cryptorchidism. *Journal of Equine Veterinary Science* **43**.

Gosh S., Carden C.F., Juras R., Mendoza M.N., Jevit M.J., Castaneda C., Phelops O., Dube J., Kelley D.E., Varner D.D., Love C.C. & Raudsepp T. (2020) Two novel cases of autosomal translocations in the horse: Warmblood family segregating t(4;30) and a cloned Arabian with a de novo t(12;25). *Cytogenetic Genome Research* **16**, 1-10.

Gosh S., Das P.J., Arnold C., Jaxheimer J., Varner D.D., Chowdhary B. & Raudsepp T. (2014a) Contribution of DNA copy number variants to equine cryptorchidism. *Journal of Equine Veterinary Science* **34**, 29.

Gosh S., Qu Z., Das P.J., Fang E., Juras R., Cothran E.G., McDonell S., Kenney D.G., Lear T.L., Adelson D.L., Chowdhary B. & Raudsepp T. (2014b) Copy number variation in the horse genome. *PLOS Genetics* **10**.

Graves J.A. (2006) Sex chromosome specialization and degeneration in mammals. *Cell* **124**, 901-14.

Grugni V., Raveane A., Ongaro L., Battaglia V., Trombetta B., Colombo G., Rosario Capodiferro M., Olivieri A., Achilli A., Perego U.A., Motta J., Tribaldos M., Woodward S.R., Ferretti L., Cruciani F., Torroni A. & Semino O. (2019) Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas. *BMC Biology* **17**.

Grzmil P., Golas A., Muller C. & Styrna J. (2007) The influence of the deletion on the long arm of the Y chromosome on sperm motility in mice. *Theriogenology* **67**, 760-6.

Guirao-Rico S., Ramirez O., Ojeda A., Amills M. & ERamos-Onsins S.E. (2018) Porcine Y-chromosome variation is consistent with the occurrence of paternal gene flow from non-Asian to Asian populations. *Heredity* **120**, 63-76.

Hamann H., Jude R., Sieme H., Mertens U., Topfer-Petersen E., Distl O. & Leeb T. (2014) A polymorphism within the equine CRISP3 gene is associated with stallion fertility in Hanoverian warmblood horses. *Animal Genetics* **38**.

Hamilton C., Verduzco-Gomez A., Favetta L.A., Blondin P. & King W.A. (2012) Testis-specific protein Y-encoded copy number is correlated to its expression and the field fertility of canadian Holstein bulls. *Sexual Development* **6**, 231-9.

Hamilton C.K., Favetta L.A., Di Meo G.P., Floriot S., Perucatti A., Peippo J., Kantanen J., Eggen A., Iannuzzi L. & King W.A. (2009) Copy number variation of testis-specific protein, Y encoded (TSPY) in 14 different breeds of cattle (Bos taurus). *Sexual Development* **3**, 205-13.

Han H., Zhao X., Xia X., Chen H., Lei C. & Dang R. (2017a) Copy number variations of five Y chromosome genes in donkeys. *Archives Animal Breeding* **60**.

Han H., Zhao X., Xia X., Chen H., Lei C. & Dang R. (2017b) Copy number variations of five Y chromosome genes in donkeys. *Archives Animal Breeding* **60**, 391–7.

Heck H. (1952) The breeding-back of the Tarpan. *Oryx*.

Herzog A., Hohn H., Klug E. & Hecht W. (1989) A sex chromosome mosaic in male pseudohermaphroditism in a horse. *Tierarztliche Praxis* **17**, 171-5.

Hoglund J., Rafati N., Rask-Andersen M., Enroth S., Karlsson T., Ek W.E. & Johansson A. (2019) Improved power and precision with whole genome sequencing data in genome-wide association studies of inflammatory biomarkers. *Scientific Reports* **9**.

Hohn H., Klug E. & Rieck G.W. (1980) A 63,XO/65,XYY mosaic in a case of questionable equine male pseudohermaphroditism. In: *4th European Colloquium on Cytogenetics of Domestic Animals*, pp. 82-92, Uppsala, Sweeden

Hsu L.Y.F. (1994) Phenotype/Karyotype correlations of Y chromosome aneuploidy with emphasis on structural aberrations in postnatally diagnosed cases. *American Journal of Human Genetics* **53**, 108-40.

Hu T., Chitnis N., Monos D. & Dinh A. (2021) Next-generation sequencing technolgies: An overview. *Human Immunology* **82**, 801-11.

Hughes J.F., Skaletsky H., Brown L.G., Pyntikova T., Graves T., Fulton R.S., Dugan S., Ding Y., Buhay C.J., Kremitzki C., Wang Q., Shen H., Holder M., Villasana D., Nazareth L.V., Cree A., Courtney L., Veizer J., Kotkiewicz H., Cho T., Koutseva N., Rozen S., Muzney D.M., Warren W.C., Gibbs R.A., Wilson R.K. & Page D.C. (2012) Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* **483**, 82-6.

Hughes J.F., Skaletsky H., Koutseva N., Pyntikova T. & Page D.C. (2015) Sex chromosome-to-autosome transposition events counter Y-chromosome gene loss in mammals. *Genome Biology Evolution* **16**.

Hughes J.F., Skaletsky H., Pyntikova T., Graves T.A., Van Daalen S.K., Minx P.J., Fulton R.S., McGrath S.D., Locke D.P., Friedman C., Trask B.J., Mardis E.R., Warren W.C., Repping S., Rozen S., K. W.R. & Page D.C. (2010) Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463**, 536-9.

Hughes J.F., Skaletsky H., Pyntikova T., Koutseva N., Raudsepp T., Brown L.G., Bellot D.W., Cho T., Graves-Lindsay T.A., Fulton L., Warren W.C., Wilson R.K., Owens E., Womack J.E., Murphy W.J., Muzny D.M., Worley K.C., Chowdhary B., Gibbs R.A. & Page D.C. (2020) Sequence analysis in Bos taurus reveals pervasiveness of X-Y arms race in mammalian lineages. *Genome Research* **30**, 1716-26.

Iannuzzi A., Parma P. & Iannuzzi L. (2021) Chromosome abnormalities and fertility in domestic bovids: a review. *Animals (Basel)* **11**, 802.

Iannuzzi L., Molteni L., Di Meo G., De Giovanni A., Perucatti A., Succi G., Incarnato D., Eggen A. & Cribiu E. (2001) A case of azzospermia in a bull carrying a Y-autosome reciprocal translocation. *Cytogenetic and Cell Genetics* **95**, 225-7.

Janecka J.E., Davis B.W., Gosh S., Paria N., Das P.J., Orlando L., Schubert M., Nielsen M.K., Stout T.A.E., Brashear W., Li G., Johnson C.D., Metz R.P., Al Zadjali A.M., Love C.C., Varner D.D., Bellot D.W., Murphy W.J., Chowdhary B.P. & Raudsepp T. (2018) Horse Y chromosome assembly displays unique evolutionary features and putative stallion fertility genes. *Nature Communication* **9**.

Jaszczak K., Parada R., Wardecka B. & Niemczewski C. (2003) A note on analysis of chromosome constitution in tissues, quality of semen and DNA microsatellite loci in bull with 60,XY/61,XXY karyotype. *Journal of Animal Feed Science* **12**, 521-7.

Jin J.L., O'Doherty A.M., Wang S., Zheng H., Sanders K.M. & Yan W. (2005) Catsper3 and catsper4 encode two cation channel-like proteins exclusively expressed in the testis. *Biology and Reproduction* **73**, 1235-42.

Jobling M.A. & Tyler-Smith C. (2003) The human Y chromosome: an evolutionary marker comes of age. *Nature Review Genetics* **4**, 598-612.

Jobling M.A. & Tyler-Smith C. (2017) Human Y-chromosome variation in the genome-sequencing era. *Nature Review Genetics* **18**, 485-97.

Kakoi H., Hirota K., Gawahara H., Kurosawa M. & Kuwajima M. (2005) Genetic diagnosis of sex chromosome abberations in horses based on parentage test by microsatellite DNA and analysis of X- and Y-linked markers. *Equine Veterinary Journal* **37**, 143-7.

Kalbfleisch T.S., Rice E.S., DePriest M.S., Walenz B.P., Hestand M.S., Vermeesch J.R., O'Connell B.L., Fiddas I.T., Vershinina A.O., Saremi N.F., Petersen J.L., Finno C.J., Bellone R.R., McCue M.E., Brooks S.A., Bailey E., Orlando L., Green R.E., Miller D.C., Antczak D.F. & MacLeod J.N. (2018) Improved reference genome for the domestic horse increases assembly contiguity and composition. *Communications Biology* **1**, 197.

Kanakis G.A. & Nischlag E. (2018) Klinefelter syndrome: more than hypogonadism. *Metabolism* **86**, 135-44.

Kenney R.M., Hurtgen J., Pierson R., Witherspoon D. & Simons J. (1983) Clinical fertility evaluation of the stallion. Society for Theriogenology manual. *Journal of The Society for Theriogenology* **IX**.

Khatchadourian K., Smith C.E., Metzler M., Gregory M., Hayden M.R., Cyr D.G. & Hermo L. (2006) Structural abnormalities in spermatids together with reduced sperm counts and motility underlie the reproductive defect in HIP1-/- mice. *Molecular Reproduction and Development* **74**, 341-59.

Kim D., Paggi J.M., Park C., Bennett C. & Salzberg S.L. (2019) Graph-based genome alignment and genotyping with HISAT@ and HISAT-genotype. *Nature Biotechnology* **37**, 907-15.

Kingsley N.B., Kern C., Creppe C., Hales E.N., Zhou H., Kalbfleisch T.S., MacLeod J.N., Petersen J.L., Finno C.J. & Bellone R.R. (2019) Functionally annotating regulatory elements in the equine genome using histone mark ChIP-Seq. *Gene* **11**, 3.

Kitamura K., Tanaka H. & Nishimune Y. (2005) The ring-finger protein haprin: Domains and function in the acrosome reaction. *Current Protein and Peptide Science* **6**, 567-74.

Klunder L.R. & McFeeley R.A. (1989) Chromosome analysis of 130 equine clinical cases. In: *6th North American Colloquium on Cytogenetics of Domestic Animals*, p. 8, West Lafayette, IN, USA.

Krausz C., Giachini C. & Forti G. (2010) TSPY and male fertility. *Genes* **1**, 308-16.

Krzeminska P., Nowacka-Woszuk J. & Switonski M. (2021) Copy number variation of the SRY gene showed an association with disorders of sex development in Yorkshire Terrier dogs. *Animal Genetics*.

Kubien E.M., Pozor M.A. & Tischner M. (1993) Clinical, cytogenetic and endocrine evaluation of a horse with 65,XXY karyotype. *Equine Veterinary Journal* **25**, 333-5.

Lahn B.T. & Page D.C. (1999) Four evolutionary strata on the human X chromsome. *Science* **286**, 964-7.

Lange J., Skaletsky H., Van Daalen S.K., Embry S.L., Korver C.M., Brown L.G., Oates R.D., Silber S., Repping S. & Page D.C. (2009) Isodicentric Y chromosomes and sex disorders as byproducts of homologous recombination that maintains palindromes *Cell* **138**.

Laseca N., Anaya G., Pena Z., Priosanto Y., Molina A. & Demyda Peyras S. (2021) Impaired reproductive function in equines: from genetics to genomics. *Animals* **11**.

Lear T.L., Cox J.H. & Kennedy G.A. (1999) Autosomal trisomy in a Thoroughbred colt: 65,XY,+31. *Equine Veterinary Journal* **31**, 85-8.

Leeb T. (2007) The horse genome project--sequence based insights into male reproductive mechanisms. . *Reproduction in Domestic Animals* **42**, 45-50.

Leeb T., Sieme H. & Topfer-Petersen E. (2005) Genetic markers for stallion fertility--lessons from humans and mice. *Animal Reproduction Science* **89**, 21-9.

Leeb T., Vogl C., de Jong P.J., Zhu B., Binns M.M., Chowdhary B.P., Nordsiek G., Scharfe M., Schrader F., Jarek M. & Bloecker H. (2006) A human-horse comparative map based on equine BAC end sequences. *Genomics* **87**, 772-6.

Li G., Davis B.W., Raudsepp T., Pearks Wikerson A.J., Mason V.C., Ferguson-Smith M., O'Brien P.C., Waters P.D. & Murphy W.J. (2013) Comparative analysis of mammalian Y chromosome illuminates ancestral structure and lineage-specific evolution. *Genome Research* **23**, 1486-95.

Li R., Yang P., Li M., Fang W., Yue X.P., Nanaei H.A., Gan S., Du D., Cai Y., Dai X., Yang Q., Cao C., Deng W., He S., Li W.H., Ma R., Liu M. & Jiang Y. (2020) A hu sheep genome with the first ovine Y chromosome reveal introgression history after sheep domestication. *Science China Life Sciences* **64**, 1116-30.

Lin Y.N., Roy A., Yan W., Burns K.H. & Matzuk M.M. (2007) Loss of zona pellucida binding proteins in the acrosomal matrix disrupts acrosome biogenesis and sperm morphogenesis. *Molecular and Cellular Biology* **27**, 6794-805.

Lindgren G., Backstrom N., Swinburne J., Hellborg L., Einarsson A., Sandberg K., Cothran G., Vila C., Binns M. & Ellegren H. (2004) Limited number of patrilines in horse domestication. *Nature Genetics* **36**, 335-6.

Liu S., Yang Y., Pan Q., Sun Y., Ma H., Liu Y., Wang M., Zhao C. & Wu C. (2020) Ancient Patrilineal Lines and Relatively High ECAY Diversity Preserved in Indigenous Horses Revealed With Novel Y-Chromosome Markers. *Frontiers in Genetics* **11**.

Long S.E. (1996) Tandem 1;30 translocation: a new structural abnormality in the horse (Equus caballus). *Cytogenetic and Cell Genetics* **72**, 162-3.

Los E. & Ford G.A. (2019) Klinefelter Syndrome. In: *StatPearls* (Treasure Island, FL.

Love C.C. (2011) Relationship between sperm motility, morphology and the fertility of stallions. *Theriogenology* **76**, 547-77.

Lucottee E.A., Skov L., Jensen J.M., Macia M.C., Munch K. & Schierup M.H. (2018) Dynamic Copy Number Evolution of X- and Y-Linked Ampliconic Genes in Human Populations. *Genetics* **209**, 907-20.

Mahmoud M., Gobet N., Cruz-Davalos D.I., Mounier N., Dessimoz C. & Sedlazak F.J. (2019) Structural variant calling: the long and the short of it. *Genome Biology and Evolution* **20**.

Mahyari E., Guo J., Lima A.C., Lewisohn D.P., Stendahl A.M., Vigh-Conrad K.A., Nie X., Nagirnaja L., Rockweiler N.B., Carrell D.T., Hotaling J.M., Aston K.I. & Conrad D.F. (2021) Comparative single-cell analysis of biopsies clarifies pathogenic mechanisms in Klinefelter syndrom. *American Journal of Human Genetics* **108**, 1924-45.

Makinen A., Katila T., Andersson M. & Gustavsson I. (2000) Two sterile stallions with XXY-syndrome. *Equine Veterinary Journal* **32**, 358-60.

Martinez-Pacheco M., Tenorio M., Almonte L., Fajardo V., Godinez A., Fernandez D., Cornejo-Paramo P., Diaz-Barba K., Halbert J., Liechti A., Szekely T., Urrtia A.O. & Cortez D. (2020) Expression evolution of ancestral XY gametologs across all major groups of placental mammals. *Genome Biology and Evolution* **12**, 2015-28.

Mary N., Villagomez D.A.F., Revay T., Rezaei S., Donaldson B., Pinton A. & King W.A. (2018) Meiotic synapsis and gene expression altered by a balanced Y-autosome reciprocal translocation in an azzospermic pig. *Sexual Development* **12**, 256-63.

Matzuk M.M. & Lamb D.J. (2008) The biology of infertility: research advances and clinical challenges. *Nature Medicine* **14**, 1197-213.

Mayernia K., Meinhardt A., Drabent B., Adham I.M., Muller C., Steckel M., Sancken U. & Engel W. (2003) Synergistic effects of germ cell expressed genes on male fertility in mice. *Cytogenetic Genome Research* **103**, 314-20.

McCue M., Bannasch D.L., Petersen J.L., Gurr J., Bailey E., Binns M., Distl O., Guerin G., Hasegawa T., Hill E.W., Leeb T., Lindgren G., Penedo M.C.T., Roed K.H., Ryder O.A., Swinburne J., Tozaki T., Valberg S.T., Vaudin M., Lindblad-To K., Wade C.M. & Mickelson J.R. (2012) A high density snp array for the domestic horse and extant perissodactyla: Utility for association mapping, genetic diversity, and phylogeny studies. *PLOS Genetics* **8**.

Meizel S. & Son J.H. (2005) Studies of sperm from mutant mice suggesting that two neurotransmitter receptors are important to the zona pellucida-initiated acrosome reaction. *Molecular Reproduction and Development* **72**, 250-8.

Meyers S.A., Liu I.K.M., Overstreet J.W., Vadas S. & Drobnis E.Z. (1996) Zona pellucida binding and zona-induced acrosome reactions in horse spermatozoa: comparisons between fertile and subfertile stallions. *Theriogenology* **46**, 1277-88.

Meyers S.A., Overstreet J.W., Liu I.K.M. & Drobnis E.Z. (1995) Capacitation in vitro of stallion spermatozoa: Comparison of progesterone-induced acrosome reactions in fertile and subfertile males. *Journal of Andrology* **16**, 47-54.

Miyake Y., Ishikawa T., Kanagawa H. & Sato K. (1981) A first case of XY/XYY mosaic bull. *Japanese Journal of Veterinary Research*.

Moraes J.C., Mattevi M.S. & Ferreira J.M. (1980) Chromosome studies in Brazilian rams. *Vet Records* **107**, 489-90.

Morel F., Dugueperoux I., McElreavey K., Le Bris M.J., Herry A., Parent P., Le Martelot M.T., Fellous M. & De Braekeleer M. (2002) Transmission of an unbalanced (Y;1) translocation in Brittany, France. *Journal of Medical Genetics* **39**.

Morel F., Roux C. & Bresson J.L. (1999) Sex chromosome aneuploidies in sperm of 47,XYY men. *Archives of Andrology* **43**, 27-36.

Morgan A.P. & Pardo-Manuel de Villena F. (2017) Sequence and structural diversity of mouse Y chromosomes *Molecular Biology and Evolution* **34**, 3186–204.

Morin S.J., Eccles J., Iturriaga A. & Zimmerman R.S. (2017) Translocations, inversions and other chromosome rearrangements. *Fertility and Sterility* **107**, 19-26.

Mu X.J., Lu Z.J., Kong Y., KLam H.Y.K. & Gerstein M.B. (2011) Analysis of genomic variation in non-coding elements using population-scale sequence data from the 1000 genomes project. *Nucleic Acids Research* **39**, 7058-76.

Mukherjee A., Dass G., Mohanarao J., Gohain M., Brahma B., Datta T.K. & De S. (2013) Absolute copy number differences of Y chromosomal genes between crossbred (Bos taurus × Bos indicus) and Indicine bulls. *Journal of Animal Science and Biotechnology* **4**.

Mukherjee A., Dass G., Mohanarao J., Katneni V.K., Banerjee D., Das T.K., Gohain M., Chakrabarty A.K., Datta T.K. & De S. (2015) Copy Number Differences of Y Chromosomal Genes Between Superior and Inferior Quality Semen Producing Crossbred (Bos taurus × Bos indicus) Bulls. *Animal Biotechnology* **26**, 65-72.

O'Brien K.L.O.F., Varghese A.C. & Agarwal A. (2010) The genetic causes of male factor infertility: a review. *Fertility and Sterility* **93**, 1-12.

Oetjens M.T., Shen F., Emery S.B., Zou Z. & Kidd J.M. (2016) Y-chromosome structural diversity in the bonobo and chimpanzee lineages. *Genome Biology Evolution* **8**, 2231–40.

Oluwole O.A., Mahboubi K., Favetta L.A., Revay T., Kroetsch T. & King W.A. (2017) Highly dynamic temporal changes of TSPY gene copy number in aging bulls. *PLOS One* **12**.

Paget S., Ducos A., Mignotte F., Pinton A., Sequela A., Berland H.M., Brun-Baronnat C., Darre R., Raymond I., Tamzali Y., VBergonier D., Berthelot X. & Darre A. (2001) 63,XO/65,XYY mosaicism in a case of equine male pseudohermaphroditism. *Vet Records* **148**, 24-5.

Paria N., Raudsepp T., Pearks Wikerson A.J., O'Brien P.C.M., Ferguson-Smith M., Love C.C., Arnold C., Rakestraw P., Murphy W.J. & Chowdhary B. (2011) A gene catalogue of the euchromatic male-specific region of the horse Y chromosome: Comparison with human and other mammals. *PLOS One*.

Pedersen A.S., Berg L.C., Almstrup K. & Thomsen P.D. (2014) A tortoiseshell male cat: Chromosome analysis and histologic examination of the testis. *Cytogenetic Genome Research* **142**.

Pei S.W., Qin F., Li W.H., Li F.D. & Yue X.P. (2019) Copy number variation of ZNF280AY across 21 cattle breeds and its association with the reproductive traits of Holstein and Simmental bulls. *Journal of Dairy Science* **102**, 7226-36.

Peng S., Bellone R.R., Petersen J.F., Kalbfleisch T.S. & Finno C.J. (2021a) Successful ATAC-Seq from snap-frozen equine tissue. *Frontiers in Genetics* **12**.

Peng S., Petersen J.L., Bellone R.R., Kalbfleisch T.S., Kingsley N.B., Barber A.M., Cappellatti E., Giulotto E. & Finno C.J. (2021b) Decoding the Equine Genome: Lessons from ENCODE. *Genes* **12**.

Pinton A., Letron I.R., Berland H., Bonnet N., Calgaro A., Garnier-Bonnet A., Yerle M. & CDucos A. (2008) Meiotic studies in an azoospermic boar carrying a Y;14 translocation. *Cytogenetic Genome Research* **120**, 106-11.

Poplin R., Ruano-Rubio V., DePristo M.A., Fennell T.J., Carneiro M.O., Van der Auwera G.A., Kling D.E., Gauthier L.D., Levy-Moonshine A., Roazen D., Shakir K., Thibault J., Chandra S., Whelan C., Lek M., Gabriel S., Daly M.J., Neale B., MacArthur D.G. & Banks E. (2017) Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*.

Power M.M. (1987) Equine half sibs with an unbalanced X;15 translocation or trisomy 28. *Cytogenetic and Cell Genetics* **45**, 163-8.

Power M.M., Gustavsson I., Switonski M. & Ploen L. (1992) Synaptonemal complex analysis of an autosomal trisomy in the horse. *Cytogenetic and Cell Genetics* **61**, 202-7.

Poznik G.D., Xue Y., Mendez F.L., Willems T.F., Massaia A., Wilson Sayres M.A., Ayub Q., McCarthy S.A., Narechania A., Kashin S., Chen Y., Banerjee R., Rodriquez-Flores J.L., Cerezo M., Shao H., Gymrek M., Malhotra A., Louzada S., Desalle R., Ritchie G.R.S., Cerveira E., Fitzgerald T.W., Garrison E., Marcketta A., Mittelman D., Romanovitch M., Zhang C., Zheng-Bradley X., Abecasis G.R., McCaroll S.A., Flicek P., Underhil P.A., Coin L., Zerbino D.R., Yang F., Lee C., Clarke L., Auton A., Erlich Y., Handsaker R., Consortium T.G.P., Bustamante C.D. & Tyler-Smith C. (2016) Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nature Genetics* **48**, 593-99.

Preston C.C., Storm E.C., Leonard R.J. & Faustino R.S. (2019) Emerging roles for nucleoporins in reproductive cellular physiology. *Canadian Journal of Physiology* **97**, 257-64.

Quach A.T., Oluwole O., King W.A. & Revay T. (2015) The porcine TSPY gene is tricopy, but not a copy number variant. *PLOS One*.

Rao D.S., Chang J.C., Kumar, P. D., Mizukami I., Smithson G.M., Bradley S.V., Parlow A.F. & Ross T.S. (2001) Huntingtin interacting protein 1 is a clathrin coat binding protein required for differentiation of late spermatogenic progenitors. *Molecular and Cellular Biology* **21**, 7796-806.

Raudsepp T. (2020) Genetics of Equine Reproductive Diseases. *Veterinary Clinics of North America-Equine Practices* **36**.

Raudsepp T. & Chowdhary B.P. (1999) Construction of chromosome-specific paints for meta- and submetacentric autosomes and the sex chromosomes in the horse and their use to detect homologous chromosomal segments in the donkey. *Chromosome Research* **7**, 103-14.

Raudsepp T. & Chowdhary B.P. (2008a) FISH for mapping single copy genes. *Molecular Biology and Evolution* **422**, 31-49.

Raudsepp T. & Chowdhary B.P. (2008b) The horse pseudoautosomal region (PAR): Characterization and comparison with the human, chimp, and mouse PARs. *Cytogenetic Genome Research* **121**, 102-9.

Raudsepp T. & Chowdhary B.P. (2016) Chromosome Aberrations and Fertility Disorders in Domestic Animals. *Annual Review of Animal Biosciences* **4**, 15-43.

Raudsepp T., Durkin K., Lear T.L., Das P.J., Avila F., Kachroo P. & Chowdhary B.P. (2010) Molecular heterogeneity of XY sex reversal in horses. *Animal Genetics* **41**, 41-52.

Raudsepp T., Finno C.J., Bellone R.R. & Petersen J.L. (2019) Ten years of the horse reference genome: insights into equine biology, domestication and population dynamics in the post-genome era. *Animal Genetics*.

Raudsepp T., Gustafson-Seabury A., Durkin K., Wagner M.L., Goh G., Seabury C.M., Brinkmeyer-Langford C., Lee E.-J., Agarwala R., Rice E.S., Schaffer A.A., Skow L.C., Tozaki T., Yasue H., Penedo M.C.T., Lyons L.A., Khazanehdar K.A., Binns M.M., MacLeod J.N., Distl O., Guerin G., Leeb T., Mickelson J.R. & Chowdhary B.P. (2008) A 4,103 marker integrated physical and comparative map of the horse genome. *Cytogenetic Genome Research* **122**, 28-36.

Raudsepp T., McCue M., Das P.J., Dobson L., Vishnoi M., Fritz K.L., Rendahl A.K., Derr J.N., Love C.C., Varner D.D. & Chowdhary B.P. (2012) Genome-wide association study implicates testis-sperm specific FKBP6 as a susceptibility locus for impaired acrosome reaction in stallions. *PLOS Genetics* **8**, e1003139

Reijo R., Alagappan R.K., Patrizio P. & Page D.C. (1996) Severe oligozoospermia resulting from deletions of azoospermia factor gene on Y chromosome. *The Lancet* **347**, 1290-3.

Repping S., Van Daalen S.K., Brown L.G., Korver C.M., Lange J., Marszalek J.D., Pyntikova T., Van Der Veen F., Skaletsky H., Page D.C. & Rozen S. (2006) High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nature Genetics* **38**, 463-67.

Revay T., Villagomez D.A.F., Brewer D., Chenier T. & King W.A. (2012) GTG mutation in the start codon of the androgen receptor gene in a family of horses with 64,XY disorder of sex development. *Sexual Development* **6**, 108-16.

Rhead B., Karolchik D., Kuhn R.M., Hinrichs A.S., Zweig A.S., Fujita P.A., Diekhans M., Smith K.E., Rosenbloom K.R., Raney B.J., Pohl A., Pheasant M., Meyer L.R., Learned K.,

Hsu F., Hillman-Jackson J., Harte R.A., Giardine B., Dreszer T.R., Clawson H., Barber G.P., Haussler D. & Kent W.J. (2010) The UCSC Genome Browser database: update 2010. *Nucleic Acids Research* **38**, D613-D9.

Rieck G.W., Hohn H. & Herzog A. (1969) Hypogonadismus, intermittierender Kryptorchismus und segmentäre Aplasie der Ductus Wolffii bei einem männlichen Rind mit XXY--Gonosomen-Konstellation bzw. XXY--XX--XY-Gonosomen-Mosaik. *Dtsch. Tierarztl. Wochenschr.* **76**, 133-8.

Rieck G.W., Hohn H. & Schmidt I. (1982) Vulvaaplasie und Urethra masculina: Maskulinisierungseffekte des Sinus urogenitalis bei genetisch weiblichen Rindern durch Chimärismen mit dem XXY-Gonosomen-Komplement. *Berl. Munch. Tierarztl. Wochenschr* **95**, 181/5.

Rives N., Milazzo J.P., Miraux L., North M.-O., Sibert L. & Mace B. (2005) From spermatocytes to spermatozoa in an infertile XYY male. *International Journal of Andrology* **28**, 304-10.

Robinson J.T., Thorvaldsdottir H., Winckler W., Guttman M., Lander E.S., Getz G. & Mesirov J.P. (2011) Integrative Genomics Viewer. *Nature Biotechnology* **29**, 24-6.

Rogers M.J. (2021) Y chromosome copy number variation and its effects on fertility and other health factors: a review. *Translational Andrology and Urology* **10**, 1373-82.

Royo H., Polikieqics G., Mahadevaiah S.K., Prosser H., Mitchell M., Bradley A., de Rooji D., Burgoyne P.S. & Turner J.M.A. (2010) Evidence that meiotic sex chromosome inactivation is essential for male fertility. *Current Biology* **20**, 2117-23.

Ruiz A.J., Castaneda C., Raudsepp T. & Tibary A. (2019) Azoospermia and Y chromosome-autosome translocation in a Friesian Stallion. *Journal of Equine Veterinary Science* **82**, e102781.

Sabeur K., Ball B.A., Corbin C.J. & Conley A. (2008) Characterization of a novel, testis-speific equine serine/threonine kinase. *Molecular Reproduction and Development* **75**, 867-73.

Sato Y., Son J.H., Tucker R.P. & Meizel S. (2000) The zona pellucida-initiated acrosome reaction: Defect due to mutations in the sperm glycine receptor/cl(-) channel. *Developmental Biology* **227**, 211-8.

Satouh Y., Inoue N., Ikawa M. & Okabe M. (2012) Visualization of the moment of mouse sperm-egg fusion and dynamic localization of IZUMO1. *Journal of Cell Science* **125**, 4985-90.

Saxena R., Brown L.G., Hawkins T., Alagappan R.K., Skaletsky H., Reeve M.P., Reijo R., Rozen S., Dinulos M.B. & 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. *Nature Genetics* **14**, 292-9.

Schmutz S.M., Barth A.D. & Moker J.S. (1994) A Klinefelter bull with a 1;29 translocation born to a fertile 61,XXX cow. *Canadian Veterinary Journal* **35**, 182-4.

Schrimpf R., Gottschalk M., Metzger J., Martinsson G., Sieme H. & Distl O. (2016) Screening of whole genome sequences identified high-impact variants for stallion fertility. *BMC Genomics* **17**, 288.

Schrimpf R., Metzger J., Martinsson G., Sieme H. & Distl O. (2015) Implication of FKBP6 for male fertility in horses. *Reproduction in Domestic Animals* **50**, 195-9.

Sciurano R., Rahn M., Rey-Valzacchi G. & Solari A.J. (2007) The asynaptic chromatin in spermatocytes of translocation carriers contains the histone variant y-H2AX and associates with the XY body. *Human Reproduction* **22**, 142-50.

Shi Q. & Martin R.H. (2001) Aneuplody in human spermatozoa: FISH analysis in men with constitutioal chromosomal abnormalities, and in infertile men. *Reproduction* **121**, 655-66.

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., Latrielle 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., Waterson R.H., Wilson R.K., Rozen S. & Page D.C. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825-37.

Skinner B.M., Lachani K., Sargent C.A., Yang F., Ellis P., Hunt T., Fu B., Louzada S., Churcher C., Tyler-Smith C. & Affara N.A. (2015) Expansion of the HSFY gene family in pig lineages. *BMC Genomics* **16**.

Skinner B.M., Sargent C.A., Churcher C., Hunt T., Herror J., Loveland J.E., Dunn M., Louzada S., Fu B., Chow W., Gilbert J., Austin-Guest S., Beal K., Carvalho-Silve D., Cheng W., Gordon D., Gradtham D., Hardy M., Harley J., Hauser H., Howden P., Howe K., Lachani K., Ellis P.J.I., Kelly D., Kerry G., Kerwin J., Ling Ng B., Threadgold G., Wileman T., Wood T.M.D., Yang F., Harrow J., Affara N.A. & Tyler-Smith C. (2016) The pig X and Y Chromosomes: structure, sequence, and evolution. *Genome Research* **26**, 130-9.

Skov L., Consortium T.D.P.G. & Schierp M.H. (2017) Analysis of 62 hybrid assembled human Y chromosomes exposes rapid structural changes and high rates of gene conversion. *PLOS Genetics* **13**, e1006834–20.

Soh Y.Q.S., Alfoldi J., Pyntikova T., Brown L.G., Graves T., Minx P.J., Fulton R.S., Kremitzki C., Koutseva N., Mueller J.L., Rozen S., Hughes J.F., Owens E., Womack J.E., Murphy W.J., Cao Q., de Jong P., Warren W.C., Wilson R.K., Skaletsky H. & Page D.C. (2014) Sequencing the mouse Y chromosome reveals convergent gene acquistion and amplification on both sex chromosomes. *Cell* **6**, 800-13.

Sood B. & Clemente Fuentes R.W. (2022) *Jacobs Syndrome*. StatPearls Publishing, Treasure Island, FL.

Strogantsev R., Krueger F., Yamazawa K., Shi H., Gould P., Goldman-Roberts M., McEwen K., Sun B., Pedersen R. & Ferguson-Smith A.C. (2015) Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biology* **16**.

Sun F., Olver-Bonet M., Turek P., Ko E. & Martin R. (2005) Meiotic studies in an azoospermic human translocation (Y;1) carrier. *Molecular Human Reproduction* **11**, 361-4.

Switonski M., Szczerbal I., Krumrych W. & Nowacka-Woszuk J. (2011) A case of Y-autosome reciprocal translocation in a Holstein-Friesian bull. *Cytogenetic Genome Research* **132**, 22-5.

Szczerbal I., Krzeminska P., Dzimira S., Tamminen T.M., Saari S., Nizanski W., Gogulski M., Nowacka-Woszuk J. & Switonski M. (2018) Disorders of sex development in cats with different completments of sex chromosomes. *Reproduction in Domestic Animals*.

Szczerbal I. & Switonski M. (2020) Genetic disorders of sex development in cats: an update. *Animal Reproduction Science* **216**.

Szczerbal I. & Switonski M. (2021) Clinical cytogenetics of the dog: A review. *Animals (Basel)* **11**, 947.

Takebayashi S.G. & Jorg W. (1986) Testicular hypoplasia in a horned goat with 61, XXY/60,XY karyotype. *Japan Journal of Genetics* **61**, 177-81.

Tanigawa M., Miyamoto K., Kobayashi S., Sat M., Akutsu H., Okabe M., Mekada E., Sakakibara K., Miyado M., Umezawa A. & Miyado K. (2008) Possible involvement of cd81 in acrosome reaction of sperm in mice. *Molecular Reproduction and Development* **75**.

Teitz L.S., Pyntikova T., Skaletsky H. & Page D.C. (2018) Selection Has Countered High Mutability to Preserve the Ancestral Copy Number of Y Chromosome Amplicons in Diverse Human Lineages. *American Journal of Human Genetics* **103**, 261-75.

Todd E.T., Hamilton N.A., Velie B.D. & Thomson P.C. (2020) The effects of inbreeding on covering success, gestation length and foal sex ration in Australian thoroughbred horses. *BMC Genetics* **21**.

Tomaszkiewicz M., Rangavittal S., Cechova M., Campos Sanchez R., Fescemyer H.W., Harris R., Ye D., O'Brien P.C.M., Chikhi R., Ryder O.A., Ferguson-Smith M.A., Medvedev P. & Makova K.D. (2016) A time- and cost-effective strategy to sequence mammalian Y Chromosomes: an application to the de novo assembly of gorilla Y. *Genome Research* **26**, 530-40.

Torresen O.K., Star B., Mier P., Andrade-Navarro M.A., CBateman A., Jarnot P., Gruca A., Grynberg M., Kajava A.V., Promponas V.J., Anisimova M., Jakobsen K. & Linke D. (2019)

Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. *Nucleic Acids Research* **47**, 10994-1006.

Toure A., Clemente E.J., Ellis P.J.I., Mahadevaiah S.K., Ojarikre O.A., Ball P.A., Reynard L., Loveland K.L., Burgoyne P.S. & Affara N.A. (2005) Identification of novel Y chromosome encoded transcripts by testis transcriptome analysis of mice with deletions of the Y chromosome long arm. *Genome Biology and Evolution* **6**, R102.

Tozaki T., Ohnuma A., Kikuchi M., Ishige T., Kakoi H., Hirota K., Kusano K. & Nagata S. (2021) Rare and common variant discovery by whole genome sequencing of 101 Thoroughbred racehorses. *Scientific Reports* **11**.

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

Varner D.D., Blanchard T.L., Brinsko S.P., Love C.C., Taylor T.S. & Johnson L. (2000) Techniques for evaluating selected reproductive disorders of stallions. *Animal Reproduction Science* **60-61**, 493-509.

Varner D.D., Brinsko S.P., Blanchard T.L., Love C.C., Macpherson M.L., Heck R.S. & Johnson L. (2001) Subfertility in stallions associated with spermatozoal acrosome dysfunction. *The American Association of Equine Practitioners* **47**, 227-8.

Vegesna R., Tomaszkiewicz M., Medvedev P. & Makova K.D. (2019) Dosage regulation, and variation in gene expression and copy number of human Y chromosome ampliconic genes. *PLOS Genetics* **15**.

Vegesna R., Tomaszkiewicz M., Ryder O.A., Campos Sanchez R., Medvedev P., DeGiorgio M. & Makova K.D. (2020) Ampliconic Genes on the Great Ape Y Chromosomes:

Rapid Evolution of Copy Number but Conservation of Expression Levels. *Genome Biology Evolution* **12**, 842-59.

Vernet N., Mahadevaiah S.K., Yamauchi Y., Decarpentrie F., Mitchell M.J., Ward M.A. & Burgoyne P.S. (2014) Mouse Y-linked ZFY1 and ZFY2 are expressed during the male-specific interphase betweein meiosis I and meiosis II and promote the 2nd meiotic division. *PLOS Genetics* **10**.

Vidal O., Drogemuller C., Obexer-Ruff G., Reber I., Jordana J., Martinez A., Balteanu V.A., Delgado J.V., Eghbalsaied S., Landi V., Goyache F., Traore A., Pazzola M., Vacca G.M., Badaoui B., Pilla F., D'Andrea M., Alvarez I., Capote J., Sharaf A., Pons A. & Amills M. (2017) Differential distribution of Y-chromosome haplotypes in Swiss and Southern European goat breeds. *Scientific Reports* **7**.

Villagomez D.A.F., Revay T., Donaldson B., Rezaei S., Pinton A., Palomino M., Junaidi A.H.A., Honaramooz A. & King W.A. (2017) Azoospermia and testicular hypoplasia in a boar carrier of a novel Y-autosome translocation. *Sexual Development* **11**, 46-51.

Villagomez D.A.F., Welsford G.E., King W.A. & Revay T. (2020) Androgen receptor gene variants in new cases of equine androgen insensisitivty syndrome. *Genes* **11**.

Vogt P.H., Edelmann A., Kirsch S., Henegariu O., Hirschemann P., Kiesewetter F., Kohn F.M., Schill W.B., Farah S., Ramos C., Hartmann M., Hartschuh W., Meschede D., Behre H.M., Castel A., Nieschiag E., Weidner W., Grone H.-J., Jung A., W. E. & Haidl G. (1996) Human Y chromosome azoospermia fac- tors (AZF) mapped to different subregions in Yq11. *Human Molecular Genetics* **5**, 933-43.

Vollger M.R., Logsdon G.A., Audano P.A., Sulovari A., Porubsky D., Peluso P., Wenger A.M., Concepcion G.T., Kronenberg Z.N., Munson K.M., Baker C., aners A.D., Spierings

D.C.J., Lansdorp P.M., Surti U., Hunkapiller M.W. & Eichler E.E. (2020) Improved assembly and variant detection of a haploid human genome using single-molecular, high fidelity long reads. *Annals of human genetics* **84**, 125-40.

Wallner B., Brem G., Muller M. & Achmann R. (2003) Fixed nucleotide difference on the Y chromosome indicate clear divergences between Equus przewalski and Equus caballus. *Animal Genetics* **34**, 453-6.

Wallner B., Palmieri N., Vogl C., Rigler D., Bozlak E., Druml T., Jagannathan V., Leeb T., Fries R., Tetens J., Thaller G., Metzger J., Distl O., Lindgren G., Rubin C.J., Andersson L., Schaefer R., McCue M., Neuditschko M., Rieder S., Schlotterer C. & Brem G. (2017) Y Chromosome Uncovers the Recent Oriental Origin of Modern Stallions. *Current Biology* **13**.

Wallner B., Piumi F., Brem G., Muller M. & Achmann R. (2004) Isolation of Y chromosome-specific microsatellites in the Horse and Cross-species amplification in the genus Equus. *Journal of Heredity* **95**, 158-64.

Wang D., Chen R., Kong S., Pan Q.-Y., Zheng Y.-H., Qiu W.-J., Fan Y. & Sun X.-F. (2017) Cytogenetic and molecular studies of male infertility in cases of Y chromosome balanced reciprocal translocation. *Molecular Medicine Reports* **16**, 2051-4.

Waters P.D. & Ruiz-Herrera A. (2020a) Meiotic executioner genes protect the Y from extinction. *Trends in Genetics* **36**, 728-38.

Waters P.D. & Ruiz-Herrera A. (2020b) Meiotic executioner genes protet the Y from extinction. *Trends in Genetics* **36**, 728-38.

Welsford G.E., Munk R., Villagomez D.A.F., Hyttel P., King W.A. & Revay T. (2017) Androgen insensitivity syndrome in a family of warmblood horses caused by a 25-bp deletion of the DNA-binding domain of the androgen receptor gene. *Sexual Development* **11**, 40-5.

Whiting J.R. (2022) JimWhiting91/genotype_plot: Genotype Plot.

Woods J., Rigby S., Brinsko S., Stephens R., Varner D.D. & Blanchard T.L. (2000) Effect of intrauterine treatment with prostaglandin E2 prior to insemination of mares in the uterine horn or body. *Theriogenology* **53**, 1827-36.

Wu C., Wang L., Iqbal F., Jiang X., Bukhari I., Guo T., Yin G., Cooke H.J., Cao Z., Jiang H. & Shi Q. (2016) Preferential Y-Y pairing and synapsis and abnormal meiotic recombination in a 47,XYY man with no obstructive azoospermia. *Molecular Cytogenetics* **9**.

Wu P.H., Fu Y., Cecchini K., Ozata D.M., Arif A.Y.T., Colpan C., Gainetdinov I., Weng Z. & Zamore P.D. (2020) The evolutionarily conserved piRNA-producing locus pi6 is required for male mouse fertility. *Nature Genetics* **52**, 728-39.

Wutke S., Sandoval-Castellanos E., Benecke N., Dohle H., friederich S., Gonzalez J., Hofreiter M., Lougas L., Magnell O., Malaspinas A., Morales-Muniz A., Orlando L., Reissmann M., Trinks A. & Ludwig A. (2018) Decline of genetic diversity in ancient domestic stallions in Europe. *Science Advances* **4**.

Xiao C., Li J., Xie T., Chen J., Zhang S., Hassan Elaksher S., Jiang F., Jiang Y.Z., L., Zhang W., Xiang Y., Wu Z., Zhao S. & Du X. (2021) The assembly of caprine Y chromosome sequence reveals a unique paternal phylogenetic pattern and improves our understanding of the origin of domestic goat. *Ecology and Evolution* **11**, 7779-95.

Yan J., Qiu Y., Ribeiro dos Santos A.M., Yin Y., Li Y.E., Vinckler N., Narial N., Benaglio P., Raman A., Li X., Fan S., Chiou J., Chen F., Frazer K.A., HGaulton K.J., Sander M., Taipale J. & Ren B. (2021) Systematic analysis of binding of transcription factors to noncoding variants. *Nature* **591**, 147-51.

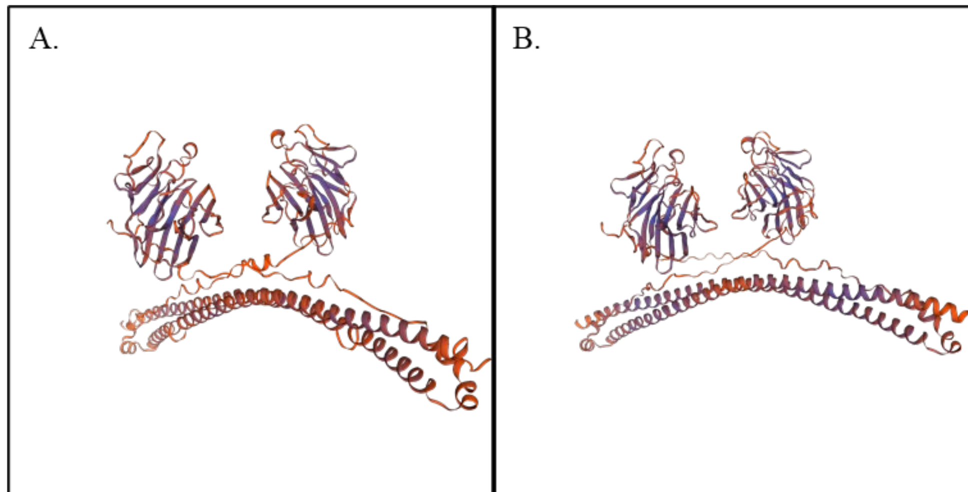
Ye D., Zaidi A.A., Tomaszkiewicz M., Anthony K., Liebowitz C., DeGiorgio M., Shriber M.D. & Makova K.D. (2018) High Levels of Copy Number Variation of Ampliconic Genes across Major Human Y Haplogroups. *Genome Biology Evolution* **10**, 1333-50.

Zhang G.W., Wu Y., Luo Z., Guan J., Wang L., Luo X. & Zu F. (2019a) Comparison of Y-chromosome-linked TSPY, TSPY2, and PRAMEY genes in Taurus cattle, yaks, and interspecific hybrid bulls. *Journal of Dairy Science* **102**.

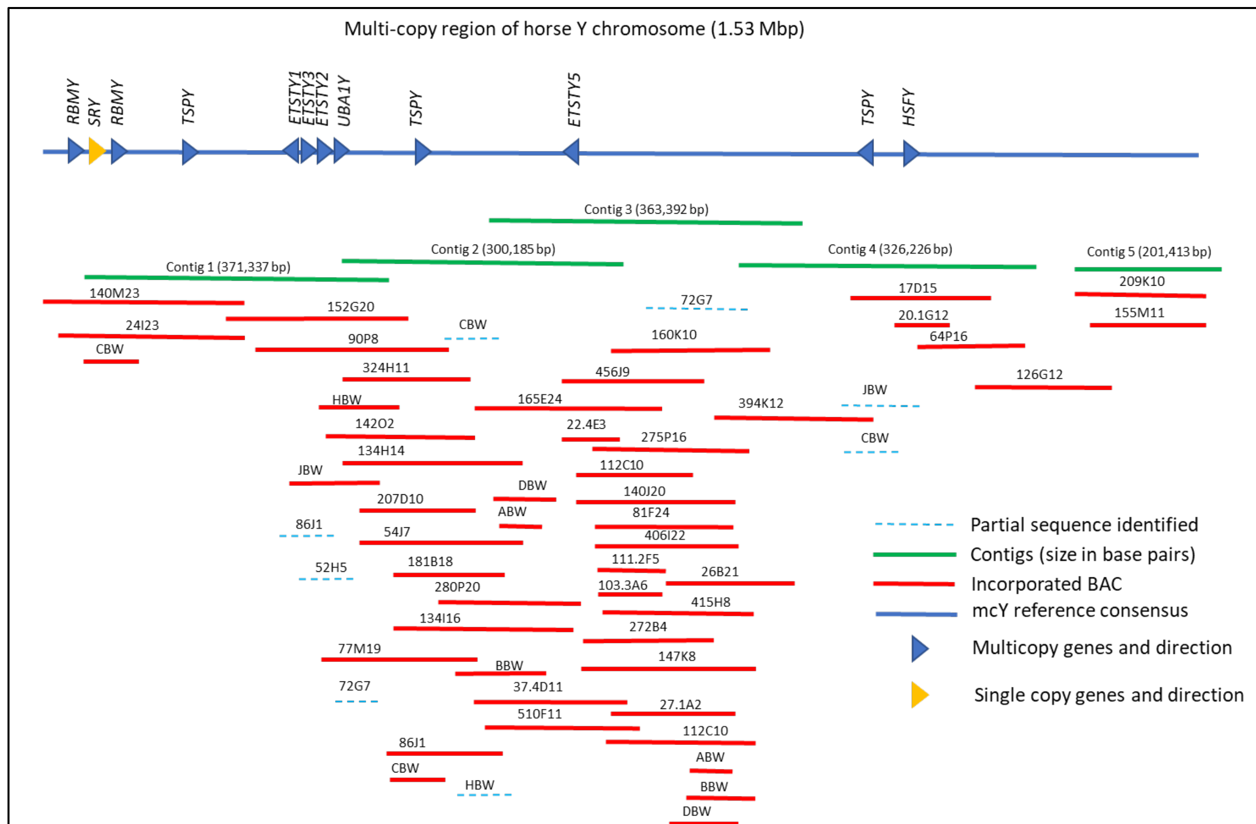
Zhang P., Wanjun J., Lua N., Zhu W. & Fan L. (2019b) Corrigendum to: IQ motif containing D (IQCD), a new acrosomal protein involved in the acrosome reaction and fertilization. . *Reproduction, Fertility, and Development* **31**.

Zhang T.Q., Bellamy J., Buwn L.C., Weber A.F. & Ruth G.R. (1992) Autosomal trisomy in a foal with contracted tendon syndrome. In: *10th European Colloqium on Cytogenetics of Domestic Animals*, pp. 281-4, Utrecht, The Netherlands.

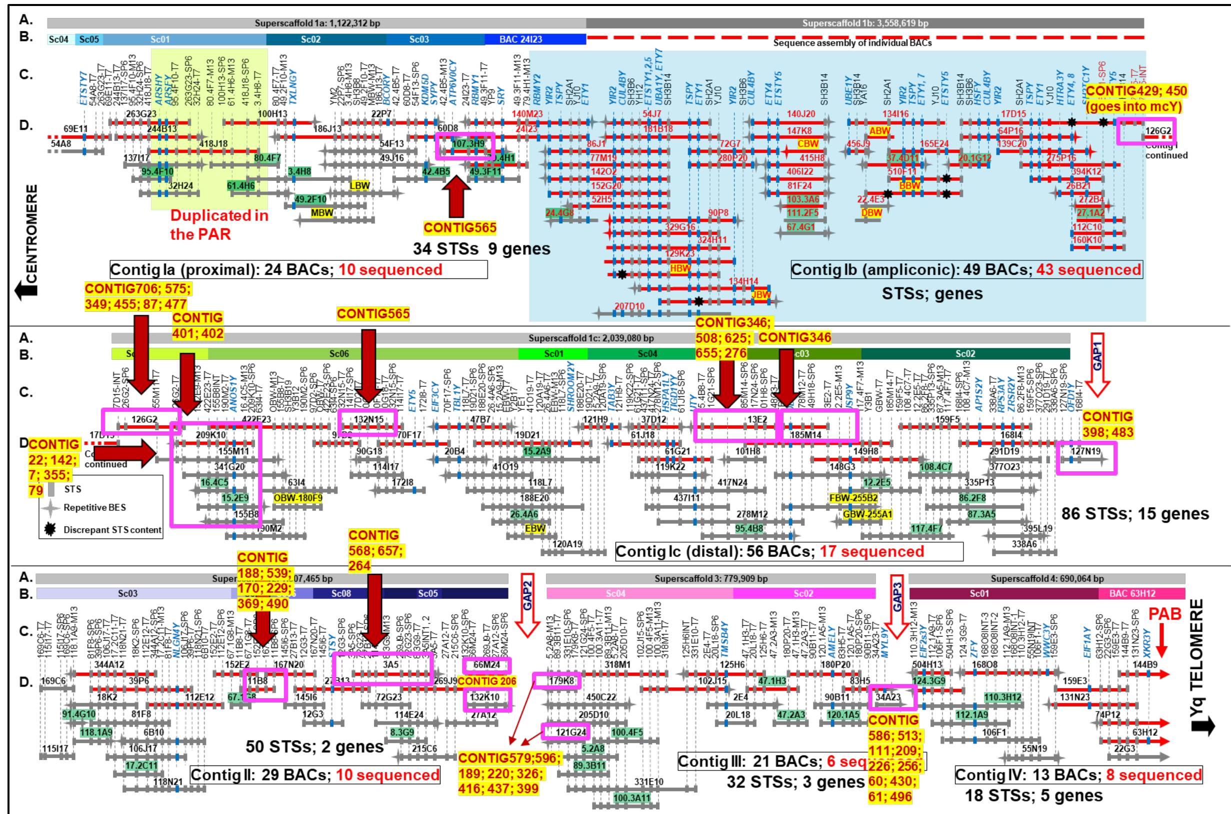
APPENDIX A: SUPPLEMENTARY FIGURES



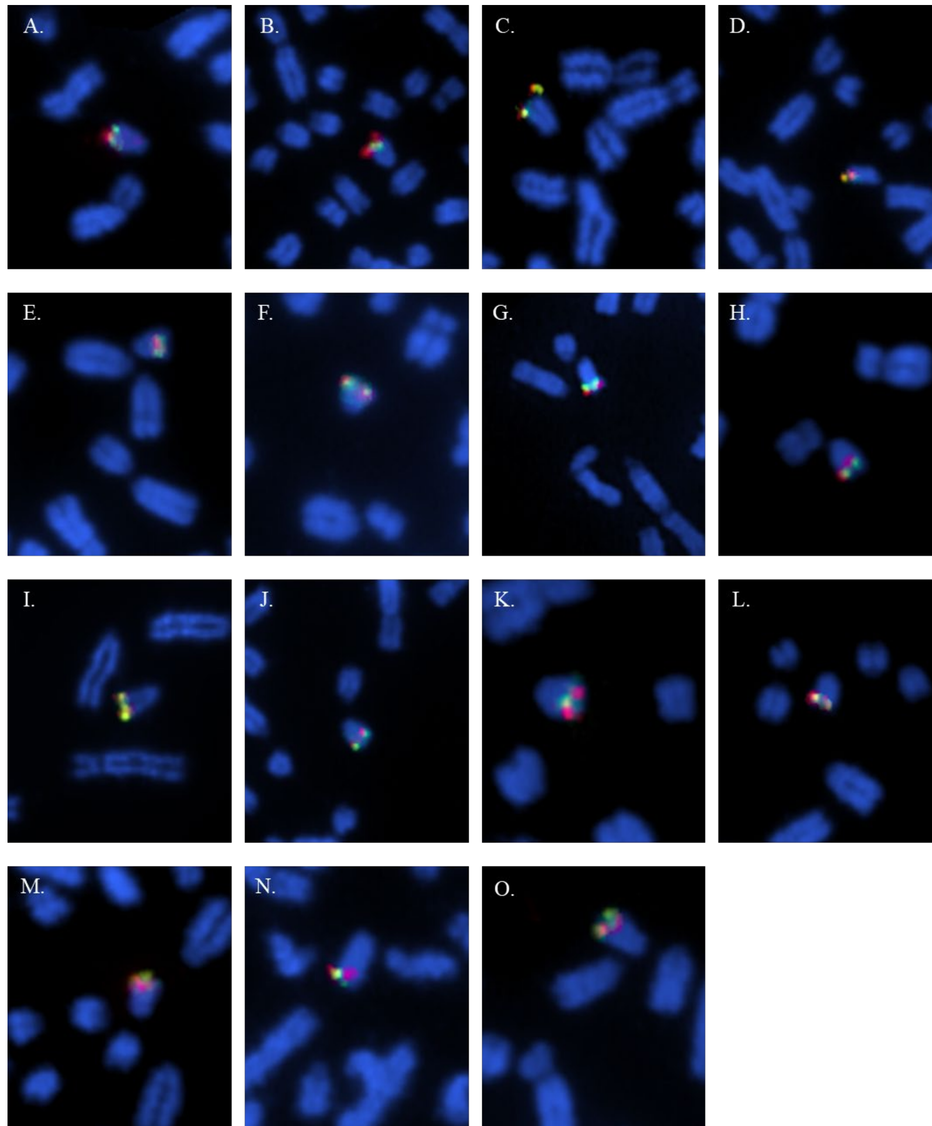
Supplementary Figure S1. Predicted protein structures of TRIM50 using SWISS-MODEL. A) Wild-type TRIM50 predicted structure without missense variant. B) Mutant TRIM50 protein predicted. Shading corresponds to the amino acid placement confidence levels when compared to the template used for predicative modeling.



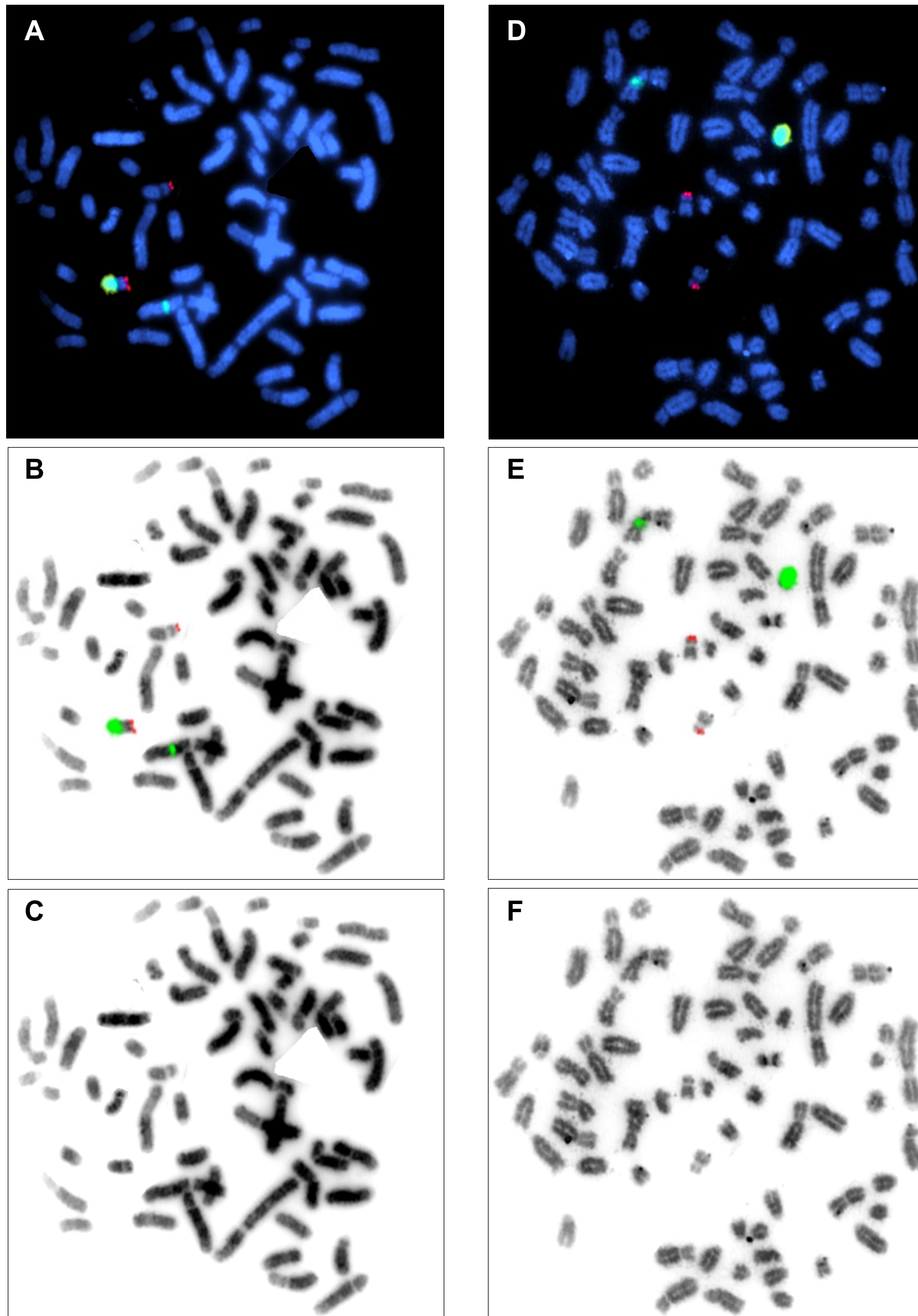
Supplementary Figure S2. A Detailed BAC clones map of the ampliconic sequence generated in this study using known multi-copy Y BACs. Dotted blue lines represent a partial BAC sequence that corresponds to the 1.53Mbp consensus sequence (solid navy blue). Red solid lines indicate the sequenced clones used in this study, and their corresponding location to the contigs and final sequence. Green solid lines represent contigs which were generated by aligning clones with similar sequences. Blue arrows indicate the mcY genes used to annotate the new assembly, and their direction corresponds to the gene direction. Yellow arrow denotes the single copy *SRY* gene.



Supplementary Figure S3. Modified detailed Y chromosome BAC map published by Janecka et al. 2018. Red arrows and pink boxes indicate the location where scY contigs amplify PCR product when screened against 94-BAC tiling path. Contig IDs are in red and highlighted in yellow.



Supplementary Figure S4. FISH validation of novel Y chromosome BAC clones located within Gap 1 using a control Y chromosome probe. Bold BAC IDs represent the new Y BAC probe, in some cases, the Y control BAC is a previously identified new Y chromosome BAC (A,C,D,J,L). **A.) 22E24** red + 121H7 green; **B.)** 125H6 red +**28P13** green; **C.) 31I9** green + 183I5 red; **D.) 43B24** green + 183I5 red; **E.) 56J9** green +167N20 red; **F.)** 125H6 red + **79D11** green; **G.)** 125H6 red + **88F22** green; **H.)** 39P6 green + **106F15** red **I.) 121H7 green** + 22P7 red; **J.)** 31I9 green + **133E10** red; **K.) 142D23** red + 159F5 green; **L.)** 43B21 green + **144G20** red; **M.) 148N16** red + 54J7 green; **N.) 172K19** red + 159F5 green; **O.)** 39P6 green + **183I5** red



Supplementary Figure S5. Three versions of the same **m**etaphase spread of the Friesian stallion (A-C) and a normal control male horse (D-F) showing FISH results with probes Y-1 (green) and 13-1 (red). Images A and D show FISH signals as green and red, and chromosomes as blue (DAPI); images B and E show FISH signals as green and red, and chromosomes as inverted DAPI, and images C and F show only chromosomes as inverted DAPI. Images A-C correspond to the partial metaphase in Figure 11A

APPENDIX B: SUPPLEMENTARY TABLES

Supplementary Table S1. Summary information, FKBP6 exon 5 SNP genotypes and fertility status (if available) for the 518 Thoroughbred horses used in this study; the 150 stallions of the select cohort with detailed fertility data are in blue font; male horses denoted with an asterisk (*) were subject for chromosome analysis and/or FKBP6 genotyping due to subfertility; animals denoted with double asterisk (**) are the 7 Thoroughbred stallions with confirmed IAR (Brinsko et al. 2007) that were used for GWAS and the discovery of FKBP6 as a susceptibility locus for IAR (Raudsepp et al. 2012); individuals with A/A-A/A genotype are in bold font.

Horse ID	Sex	Geographical Location	SNP1 (G>A)	SNP2 (A>C)	Combined genotype	Fertility status
FK195*	**Male**	**Australia**	**AA**	**AA**	**AAAA**	**idiopathic subfertility**
H402**	**Male**	**Ireland**	**AA**	**AA**	**AAAA**	**subfertile due to IAR**
H698*	**Male**	**Ireland**	**AA**	**AA**	**AAAA**	**subfertile**
H650*	**Male**	**United Kingdom**	**AA**	**AA**	**AAAA**	**subfertile**
H652*	**Male**	**United Kingdom**	**AA**	**AA**	**AAAA**	**subfertile**
H302	Male	United States	AA	AA	AAAA	low fertility
H815	Male	United States	AA	AA	AAAA	low fertility
FK143_JM10	Male	United States	AA	AA	AAAA	low fertility
FK085_DV142	Male	United States	AA	AA	AAAA	low fertility
H058**	**Male**	**United States**	**AA**	**AA**	**AAAA**	**subfertile due to IAR**
H144**	**Male**	**United States**	**AA**	**AA**	**AAAA**	**subfertile due to IAR**
H165**	**Male**	**United States**	**AA**	**AA**	**AAAA**	**subfertile due to IAR**
H166**	**Male**	**United States**	**AA**	**AA**	**AAAA**	**subfertile due to IAR**
H403**	**Male**	**United States**	**AA**	**AA**	**AAAA**	**subfertile due to IAR**
H404**	**Male**	**United States**	**AA**	**AA**	**AAAA**	**subfertile due to IAR**
FK167	**Male**	**United States**	**AA**	**AA**	**AAAA**	**n/a**
H367	**Male**	**United States**	**AA**	**AA**	**AAAA**	**n/a**
49629	**Female**	**Venezuela**	**AA**	**AA**	**AAAA**	**n/a**
54753	**Male**	**Venezuela**	**AA**	**AA**	**AAAA**	**n/a**

57262	**Female**	**Venezuela**	**AA**	**AA**	**AAAA**	**n/a**
60101	**Female**	**Venezuela**	**AA**	**AA**	**AAAA**	**n/a**
FK048_DV097	Male	United States	AA	AC	AAAC	normal fertile
FK091_DV204	Male	United States	AA	AC	AAAC	normal fertile
FK002_DV022	Male	United States	AA	AC	AAAC	normal fertile
FK145_JM12	Male	United States	AA	AC	AAAC	normal fertile
FK061_DV112	Male	United States	AA	AC	AAAC	normal fertile
FK036_DV083	Male	United States	AA	AC	AAAC	normal fertile
FK028_DV075	Male	United States	AA	AC	AAAC	normal fertile
FK149_JM16	Male	United States	AA	AC	AAAC	normal fertile
H083	Male	United States	AA	AC	AAAC	n/a
H094	Male	United States	AA	AC	AAAC	n/a
H592	Male	United States	AA	AC	AAAC	n/a
H691	Male	United States	AA	AC	AAAC	n/a
FK130	Male	United States	AA	AC	AAAC	n/a
H888	Male	United States	AA	AC	AAAC	n/a
FK212	Male	United States	AA	AC	AAAC	n/a
50237	Male	Venezuela	AA	AC	AAAC	n/a
71170	Female	Venezuela	AA	AC	AAAC	n/a
71182	Male	Venezuela	AA	AC	AAAC	n/a
71219	Male	Venezuela	AA	AC	AAAC	n/a
71230	Male	Venezuela	AA	AC	AAAC	n/a
71249	Female	Venezuela	AA	AC	AAAC	n/a
71305	Male	Venezuela	AA	AC	AAAC	n/a
73050	Female	Venezuela	AA	AC	AAAC	n/a
73133	Female	Venezuela	AA	AC	AAAC	n/a
76345	Male	Venezuela	AA	AC	AAAC	n/a
FK199	Male	Ireland	AA	CC	AACC	n/a
FK207	Male	Ireland	AA	CC	AACC	n/a
FK065_DV117	Male	United States	AA	CC	AACC	normal fertile
FK101_DV216	Male	United States	AA	CC	AACC	normal fertile
FK060_DV111	Male	United States	AA	CC	AACC	normal fertile
FK139_JM06	Male	United States	AA	CC	AACC	normal fertile
FK117_DV239	Male	United States	AA	CC	AACC	normal fertile
FK210	Male	United States	AA	CC	AACC	n/a
50817	Male	Venezuela	AA	CC	AACC	n/a
53624	Female	Venezuela	AA	CC	AACC	n/a
57294	Male	Venezuela	AA	CC	AACC	n/a
72936	Male	Venezuela	AA	CC	AACC	n/a
73107	Male	Venezuela	AA	CC	AACC	n/a
73122	Male	Venezuela	AA	CC	AACC	n/a

H265	Male	United Kingdom	AG	AA	AGAA	n/a
FK029_DV076	Male	United States	AG	AA	AGAA	normal fertile
FK031_DV078	Male	United States	AG	AA	AGAA	normal fertile
FK051_DV101	Male	United States	AG	AA	AGAA	normal fertile
FK118_DV_A	Male	United States	AG	AA	AGAA	normal fertile
FK054_DV104	Male	United States	AG	AA	AGAA	normal fertile
FK069_DV123	Male	United States	AG	AA	AGAA	normal fertile
FK081_DV137	Male	United States	AG	AA	AGAA	normal fertile
FK161_JM	Male	United States	AG	AA	AGAA	normal fertile
FK018_DV040	Male	United States	AG	AA	AGAA	normal fertile
FK121_DV_D	Male	United States	AG	AA	AGAA	normal fertile
FK124_DV_G	Male	United States	AG	AA	AGAA	normal fertile
FK142_JM09	Male	United States	AG	AA	AGAA	normal fertile
FK106_DV229	Male	United States	AG	AA	AGAA	n/a
FK115_DV237	Male	United States	AG	AA	AGAA	n/a
H096	Male	United States	AG	AA	AGAA	n/a
H114	Male	United States	AG	AA	AGAA	n/a
H155	Male	United States	AG	AA	AGAA	n/a
H161	Male	United States	AG	AA	AGAA	n/a
H542	Male	United States	AG	AA	AGAA	n/a
FK174	Male	United States	AG	AA	AGAA	n/a
FK185	Male	United States	AG	AA	AGAA	n/a
FK190	Male	United States	AG	AA	AGAA	n/a
FK208	Male	United States	AG	AA	AGAA	n/a
49695	Female	Venezuela	AG	AA	AGAA	n/a
54852	Male	Venezuela	AG	AA	AGAA	n/a
50797	Male	Venezuela	AG	AA	AGAA	n/a
53066	Male	Venezuela	AG	AA	AGAA	n/a
53092	Male	Venezuela	AG	AA	AGAA	n/a
53324	Male	Venezuela	AG	AA	AGAA	n/a
53407	Female	Venezuela	AG	AA	AGAA	n/a
53566	Male	Venezuela	AG	AA	AGAA	n/a
54629	Male	Venezuela	AG	AA	AGAA	n/a
54705	Male	Venezuela	AG	AA	AGAA	n/a
57327	Female	Venezuela	AG	AA	AGAA	n/a
57556	Male	Venezuela	AG	AA	AGAA	n/a
58047	Male	Venezuela	AG	AA	AGAA	n/a
58052	Female	Venezuela	AG	AA	AGAA	n/a
59318	Female	Venezuela	AG	AA	AGAA	n/a
59545	Male	Venezuela	AG	AA	AGAA	n/a

59546	Male	Venezuela	AG	AA	AGAA	n/a
59927	Female	Venezuela	AG	AA	AGAA	n/a
61416	Male	Venezuela	AG	AA	AGAA	n/a
61546	Female	Venezuela	AG	AA	AGAA	n/a
61601	Female	Venezuela	AG	AA	AGAA	n/a
61606	Male	Venezuela	AG	AA	AGAA	n/a
71185	Male	Venezuela	AG	AA	AGAA	n/a
71231	Female	Venezuela	AG	AA	AGAA	n/a
71280	Male	Venezuela	AG	AA	AGAA	n/a
72516	Male	Venezuela	AG	AA	AGAA	n/a
72601	Female	Venezuela	AG	AA	AGAA	n/a
72603	Female	Venezuela	AG	AA	AGAA	n/a
72604	Male	Venezuela	AG	AA	AGAA	n/a
72968	Female	Venezuela	AG	AA	AGAA	n/a
73078	Male	Venezuela	AG	AA	AGAA	n/a
73678	Female	Venezuela	AG	AA	AGAA	n/a
73738	Male	Venezuela	AG	AA	AGAA	n/a
76231	Female	Venezuela	AG	AA	AGAA	n/a
76289	Female	Venezuela	AG	AA	AGAA	n/a
76341	Male	Venezuela	AG	AA	AGAA	n/a
76421	Female	Venezuela	AG	AA	AGAA	n/a
85395	Female	Venezuela	AG	AA	AGAA	n/a
85396	Female	Venezuela	AG	AA	AGAA	n/a
85486	Male	Venezuela	AG	AA	AGAA	n/a
FK186	Male	Australia	AG	AC	AGAC	n/a
H113	Male	Ireland	AG	AC	AGAC	n/a
H667	Male	Ireland	AG	AC	AGAC	n/a
FK037_DV084	Male	United States	AG	AC	AGAC	low fertility
FK057_DV108	Male	United States	AG	AC	AGAC	low fertility
FK110_DV233	Male	United States	AG	AC	AGAC	low fertility
FK126_DV_I	Male	United States	AG	AC	AGAC	low fertility
FK015_DV035	Male	United States	AG	AC	AGAC	low fertility
FK049_DV098	Male	United States	AG	AC	AGAC	normal fertile
FK070_DV125	Male	United States	AG	AC	AGAC	normal fertile
FK067_DV120	Male	United States	AG	AC	AGAC	normal fertile
FK033_DV080	Male	United States	AG	AC	AGAC	normal fertile
FK152_JM19	Male	United States	AG	AC	AGAC	normal fertile
FK086_DV174	Male	United States	AG	AC	AGAC	normal fertile
FK093_DV206	Male	United States	AG	AC	AGAC	normal fertile
FK006_DV026	Male	United States	AG	AC	AGAC	normal fertile
FK038_DV085	Male	United States	AG	AC	AGAC	normal fertile

FK045_DV094	Male	United States	AG	AC	AGAC	normal fertile
FK137_JM04	Male	United States	AG	AC	AGAC	normal fertile
FK141_JM08	Male	United States	AG	AC	AGAC	normal fertile
FK016_DV036	Male	United States	AG	AC	AGAC	normal fertile
FK050_DV100	Male	United States	AG	AC	AGAC	normal fertile
FK035_DV082	Male	United States	AG	AC	AGAC	normal fertile
FK055_DV105	Male	United States	AG	AC	AGAC	normal fertile
FK039_DV086	Male	United States	AG	AC	AGAC	normal fertile
FK146_JM13	Male	United States	AG	AC	AGAC	normal fertile
FK113_DV235	Male	United States	AG	AC	AGAC	normal fertile
FK163_DV	Male	United States	AG	AC	AGAC	normal fertile
FK090_DV203	Male	United States	AG	AC	AGAC	normal fertile
FK138_JM05	Male	United States	AG	AC	AGAC	normal fertile
FK153_JM20	Male	United States	AG	AC	AGAC	normal fertile
FK105_DV221	Male	United States	AG	AC	AGAC	normal fertile
FK063_DV115	Male	United States	AG	AC	AGAC	n/a
FK116_DV238	Male	United States	AG	AC	AGAC	normal fertile
FK132_JM37	Male	United States	AG	AC	AGAC	n/a
FK022_DV069	Male	United States	AG	AC	AGAC	n/a
FK129	Male	United States	AG	AC	AGAC	n/a
FK111_DV234 A	Male	United States	AG	AC	AGAC	n/a
H024	Male	United States	AG	AC	AGAC	n/a
H097	Male	United States	AG	AC	AGAC	n/a
H109	Male	United States	AG	AC	AGAC	n/a
H110	Male	United States	AG	AC	AGAC	n/a
H112	Male	United States	AG	AC	AGAC	n/a
H167	Male	United States	AG	AC	AGAC	n/a
H183	Male	United States	AG	AC	AGAC	n/a
H262	Male	United States	AG	AC	AGAC	n/a
H845	Male	United States	AG	AC	AGAC	n/a
H851	Male	United States	AG	AC	AGAC	n/a
H858	Male	United States	AG	AC	AGAC	n/a
FK168	Male	United States	AG	AC	AGAC	n/a
FK179	Male	United States	AG	AC	AGAC	n/a
FK181	Male	United States	AG	AC	AGAC	n/a
FK182	Male	United States	AG	AC	AGAC	n/a
FK184	Male	United States	AG	AC	AGAC	n/a
FK188	Male	United States	AG	AC	AGAC	n/a
FK191	Male	United States	AG	AC	AGAC	n/a
FK196	Male	United States	AG	AC	AGAC	n/a

FK201	Male	United States	AG	AC	AGAC	n/a
FK202	Male	United States	AG	AC	AGAC	n/a
FK204	Male	United States	AG	AC	AGAC	n/a
FK206	Male	United States	AG	AC	AGAC	n/a
FK213	Male	United States	AG	AC	AGAC	n/a
H281	Male	Unknown	AG	AC	AGAC	n/a
H303	Male	Unknown	AG	AC	AGAC	n/a
H304	Male	Unknown	AG	AC	AGAC	n/a
H307	Male	Unknown	AG	AC	AGAC	n/a
H309	Male	Unknown	AG	AC	AGAC	n/a
H310	Male	Unknown	AG	AC	AGAC	n/a
H316	Male	Unknown	AG	AC	AGAC	n/a
H317	Male	Unknown	AG	AC	AGAC	n/a
47828	Male	Venezuela	AG	AC	AGAC	n/a
50127	Male	Venezuela	AG	AC	AGAC	n/a
50177	Female	Venezuela	AG	AC	AGAC	n/a
50628	Female	Venezuela	AG	AC	AGAC	n/a
50657	Male	Venezuela	AG	AC	AGAC	n/a
50671	Male	Venezuela	AG	AC	AGAC	n/a
53051	Female	Venezuela	AG	AC	AGAC	n/a
53168	Female	Venezuela	AG	AC	AGAC	n/a
54851	Male	Venezuela	AG	AC	AGAC	n/a
57341	Male	Venezuela	AG	AC	AGAC	n/a
59212	Male	Venezuela	AG	AC	AGAC	n/a
59382	Male	Venezuela	AG	AC	AGAC	n/a
59408	Female	Venezuela	AG	AC	AGAC	n/a
60011	Female	Venezuela	AG	AC	AGAC	n/a
60038	Female	Venezuela	AG	AC	AGAC	n/a
61275	Female	Venezuela	AG	AC	AGAC	n/a
61567	Male	Venezuela	AG	AC	AGAC	n/a
61648	Male	Venezuela	AG	AC	AGAC	n/a
61682	Female	Venezuela	AG	AC	AGAC	n/a
71163	Female	Venezuela	AG	AC	AGAC	n/a
71193	Male	Venezuela	AG	AC	AGAC	n/a
71216	Male	Venezuela	AG	AC	AGAC	n/a
71228	Female	Venezuela	AG	AC	AGAC	n/a
71241	Female	Venezuela	AG	AC	AGAC	n/a
71263	Male	Venezuela	AG	AC	AGAC	n/a
71269	Female	Venezuela	AG	AC	AGAC	n/a
71288	Female	Venezuela	AG	AC	AGAC	n/a
71328	Male	Venezuela	AG	AC	AGAC	n/a

72404	Male	Venezuela	AG	AC	AGAC	n/a
72523	Female	Venezuela	AG	AC	AGAC	n/a
72557	Female	Venezuela	AG	AC	AGAC	n/a
72583	Male	Venezuela	AG	AC	AGAC	n/a
72708	Female	Venezuela	AG	AC	AGAC	n/a
73158	Female	Venezuela	AG	AC	AGAC	n/a
73219	Female	Venezuela	AG	AC	AGAC	n/a
73716	Female	Venezuela	AG	AC	AGAC	n/a
76340	Female	Venezuela	AG	AC	AGAC	n/a
76359	Male	Venezuela	AG	AC	AGAC	n/a
76410	Female	Venezuela	AG	AC	AGAC	n/a
76454	Female	Venezuela	AG	AC	AGAC	n/a
85374	Male	Venezuela	AG	AC	AGAC	n/a
85409	Male	Venezuela	AG	AC	AGAC	n/a
85529	Female	Venezuela	AG	AC	AGAC	n/a
85534	Male	Venezuela	AG	AC	AGAC	n/a
85550	Female	Venezuela	AG	AC	AGAC	n/a
H162	Male	Brazil	AG	CC	AGCC	n/a
H664	Male	United Kingdom	AG	CC	AGCC	n/a
FK197	Male	United Kingdom	AG	CC	AGCC	n/a
FK123_DV_F	Male	United States	AG	CC	AGCC	normal fertile
FK014_DV034	Male	United States	AG	CC	AGCC	normal fertile
FK021_DV045	Male	United States	AG	CC	AGCC	normal fertile
FK083_DV139	Male	United States	AG	CC	AGCC	normal fertile
FK004_DV024	Male	United States	AG	CC	AGCC	normal fertile
FK157_DV	Male	United States	AG	CC	AGCC	normal fertile
FK098_DV213	Male	United States	AG	CC	AGCC	normal fertile
FK020_DV043	Male	United States	AG	CC	AGCC	normal fertile
FK159_DV	Male	United States	AG	CC	AGCC	normal fertile
FK047_DV096	Male	United States	AG	CC	AGCC	normal fertile
FK107_DV230	Male	United States	AG	CC	AGCC	normal fertile
FK059_DV110	Male	United States	AG	CC	AGCC	normal fertile
FK003_DV023	Male	United States	AG	CC	AGCC	normal fertile
FK162_JM	Male	United States	AG	CC	AGCC	normal fertile
FK147_JM14	Male	United States	AG	CC	AGCC	normal fertile
FK112_DV234 B	Male	United States	AG	CC	AGCC	n/a
H898	Male	United States	AG	CC	AGCC	n/a
FK172_DV	Male	United States	AG	CC	AGCC	n/a
FK214	Male	United States	AG	CC	AGCC	n/a

50220	Male	Venezuela	AG	CC	AGCC	n/a
50278	Female	Venezuela	AG	CC	AGCC	n/a
50359	Female	Venezuela	AG	CC	AGCC	n/a
54793	Male	Venezuela	AG	CC	AGCC	n/a
57543	Female	Venezuela	AG	CC	AGCC	n/a
59201	Female	Venezuela	AG	CC	AGCC	n/a
71168	Male	Venezuela	AG	CC	AGCC	n/a
71295	Male	Venezuela	AG	CC	AGCC	n/a
71364	Female	Venezuela	AG	CC	AGCC	n/a
72384	Female	Venezuela	AG	CC	AGCC	n/a
73272	Male	Venezuela	AG	CC	AGCC	n/a
73343	Female	Venezuela	AG	CC	AGCC	n/a
FK192	Male	Australia	GG	AA	GGAA	n/a
FK193	Male	Australia	GG	AA	GGAA	n/a
H689	Male	Canada	GG	AA	GGAA	n/a
H023	Male	Ireland	GG	AA	GGAA	n/a
H658	Male	United Kingdom	GG	AA	GGAA	n/a
FK200	Male	United Kingdom	GG	AA	GGAA	n/a
FK151_JM18	Male	United States	GG	AA	GGAA	low fertility
FK114_DV236	Male	United States	GG	AA	GGAA	low fertility
FK102_DV217	Male	United States	GG	AA	GGAA	low fertility
FK068_DV121	Male	United States	GG	AA	GGAA	normal fertile
FK017_DV038	Male	United States	GG	AA	GGAA	normal fertile
FK158_DV	Male	United States	GG	AA	GGAA	normal fertile
FK052_DV102	Male	United States	GG	AA	GGAA	normal fertile
FK109_DV232	Male	United States	GG	AA	GGAA	normal fertile
FK044_DV093	Male	United States	GG	AA	GGAA	normal fertile
FK013_DV033	Male	United States	GG	AA	GGAA	normal fertile
FK095_DV208	Male	United States	GG	AA	GGAA	normal fertile
FK077_DV133	Male	United States	GG	AA	GGAA	normal fertile
FK165_DV	Male	United States	GG	AA	GGAA	normal fertile
FK076_DV131	Male	United States	GG	AA	GGAA	normal fertile
FK160_JM	Male	United States	GG	AA	GGAA	normal fertile
FK164_DV	Male	United States	GG	AA	GGAA	normal fertile
FK119_DV_B	Male	United States	GG	AA	GGAA	normal fertile
FK154_JM21	Male	United States	GG	AA	GGAA	normal fertile
FK166_DV	Male	United States	GG	AA	GGAA	normal fertile
FK144_JM11	Male	United States	GG	AA	GGAA	normal fertile
FK148_JM15	Male	United States	GG	AA	GGAA	normal fertile
FK135_JM02	Male	United States	GG	AA	GGAA	normal fertile

FK026_DV073	Male	United States	GG	AA	GGAA	normal fertile
FK030_DV077	Male	United States	GG	AA	GGAA	normal fertile
FK062_DV114	Male	United States	GG	AA	GGAA	normal fertile
FK155_JM22	Male	United States	GG	AA	GGAA	normal fertile
FK131_JM36	Male	United States	GG	AA	GGAA	n/a
FK133_JM38	Male	United States	GG	AA	GGAA	n/a
H111	Male	United States	GG	AA	GGAA	n/a
H115	Male	United States	GG	AA	GGAA	n/a
H118	Male	United States	GG	AA	GGAA	n/a
H180	Male	United States	GG	AA	GGAA	n/a
H541	Male	United States	GG	AA	GGAA	n/a
FK169	Female	United States	GG	AA	GGAA	n/a
FK203	Male	United States	GG	AA	GGAA	n/a
FK209	Male	United States	GG	AA	GGAA	n/a
FK170	Male	Unknown	GG	AA	GGAA	n/a
49653	Female	Venezuela	GG	AA	GGAA	n/a
49808	Male	Venezuela	GG	AA	GGAA	n/a
50354	Male	Venezuela	GG	AA	GGAA	n/a
50396	Female	Venezuela	GG	AA	GGAA	n/a
50564	Female	Venezuela	GG	AA	GGAA	n/a
50641	Male	Venezuela	GG	AA	GGAA	n/a
50758	Male	Venezuela	GG	AA	GGAA	n/a
50800	Female	Venezuela	GG	AA	GGAA	n/a
50818	Female	Venezuela	GG	AA	GGAA	n/a
53086	Male	Venezuela	GG	AA	GGAA	n/a
53096	Female	Venezuela	GG	AA	GGAA	n/a
53197	Female	Venezuela	GG	AA	GGAA	n/a
53327	Female	Venezuela	GG	AA	GGAA	n/a
53388	Female	Venezuela	GG	AA	GGAA	n/a
53390	Female	Venezuela	GG	AA	GGAA	n/a
53522	Male	Venezuela	GG	AA	GGAA	n/a
53559	Male	Venezuela	GG	AA	GGAA	n/a
54637	Female	Venezuela	GG	AA	GGAA	n/a
57446	Female	Venezuela	GG	AA	GGAA	n/a
58077	Female	Venezuela	GG	AA	GGAA	n/a
59333	Female	Venezuela	GG	AA	GGAA	n/a
59337	Female	Venezuela	GG	AA	GGAA	n/a
59340	Male	Venezuela	GG	AA	GGAA	n/a
59612	Female	Venezuela	GG	AA	GGAA	n/a
59941	Male	Venezuela	GG	AA	GGAA	n/a
59951	Male	Venezuela	GG	AA	GGAA	n/a

59978	Female	Venezuela	GG	AA	GGAA	n/a
59990	Male	Venezuela	GG	AA	GGAA	n/a
61404	Female	Venezuela	GG	AA	GGAA	n/a
61406	Male	Venezuela	GG	AA	GGAA	n/a
61408	Male	Venezuela	GG	AA	GGAA	n/a
61517	Female	Venezuela	GG	AA	GGAA	n/a
71197	Female	Venezuela	GG	AA	GGAA	n/a
71210	Male	Venezuela	GG	AA	GGAA	n/a
71218	Female	Venezuela	GG	AA	GGAA	n/a
71220	Female	Venezuela	GG	AA	GGAA	n/a
71240	Male	Venezuela	GG	AA	GGAA	n/a
71283	Male	Venezuela	GG	AA	GGAA	n/a
71292	Female	Venezuela	GG	AA	GGAA	n/a
71373	Male	Venezuela	GG	AA	GGAA	n/a
71440	Female	Venezuela	GG	AA	GGAA	n/a
72438	Male	Venezuela	GG	AA	GGAA	n/a
72625	Female	Venezuela	GG	AA	GGAA	n/a
72689	Male	Venezuela	GG	AA	GGAA	n/a
73143	Female	Venezuela	GG	AA	GGAA	n/a
76226	Male	Venezuela	GG	AA	GGAA	n/a
76253	Female	Venezuela	GG	AA	GGAA	n/a
76271	Male	Venezuela	GG	AA	GGAA	n/a
76281	Male	Venezuela	GG	AA	GGAA	n/a
76288	Male	Venezuela	GG	AA	GGAA	n/a
76291	Female	Venezuela	GG	AA	GGAA	n/a
76380	Female	Venezuela	GG	AA	GGAA	n/a
76390	Male	Venezuela	GG	AA	GGAA	n/a
76467	Male	Venezuela	GG	AA	GGAA	n/a
85416	Male	Venezuela	GG	AA	GGAA	n/a
85466	Male	Venezuela	GG	AA	GGAA	n/a
85523	Female	Venezuela	GG	AA	GGAA	n/a
85530	Male	Venezuela	GG	AA	GGAA	n/a
85535	Male	Venezuela	GG	AA	GGAA	n/a
H098	Male	United Kingdom	GG	AC	GGAC	n/a
FK024_DV071	Male	United States	GG	AC	GGAC	low fertility
FK074_DV129	Male	United States	GG	AC	GGAC	low fertility
FK066_DV119	Male	United States	GG	AC	GGAC	low fertility
FK075_DV130	Male	United States	GG	AC	GGAC	low fertility
FK034_DV081	Male	United States	GG	AC	GGAC	low fertility
FK009_DV029	Male	United States	GG	AC	GGAC	low fertility

FK032_DV079	Male	United States	GG	AC	GGAC	low fertility
FK046_DV095	Male	United States	GG	AC	GGAC	low fertility
FK078_DV134	Male	United States	GG	AC	GGAC	normal fertile
FK104_DV219	Male	United States	GG	AC	GGAC	normal fertile
FK096_DV210	Male	United States	GG	AC	GGAC	normal fertile
FK125_DV_H	Male	United States	GG	AC	GGAC	normal fertile
FK099_DV214	Male	United States	GG	AC	GGAC	normal fertile
FK042_DV089	Male	United States	GG	AC	GGAC	normal fertile
FK010_DV030	Male	United States	GG	AC	GGAC	normal fertile
FK156_DV	Male	United States	GG	AC	GGAC	normal fertile
FK092_DV205	Male	United States	GG	AC	GGAC	normal fertile
FK019_DV041	Male	United States	GG	AC	GGAC	normal fertile
FK103_DV218	Male	United States	GG	AC	GGAC	normal fertile
FK011_DV031	Male	United States	GG	AC	GGAC	normal fertile
FK007_DV027	Male	United States	GG	AC	GGAC	normal fertile
FK012_DV032	Male	United States	GG	AC	GGAC	normal fertile
FK072_DV127	Male	United States	GG	AC	GGAC	normal fertile
FK087_DV175	Male	United States	GG	AC	GGAC	normal fertile
FK064_DV116	Male	United States	GG	AC	GGAC	normal fertile
FK053_DV103	Male	United States	GG	AC	GGAC	normal fertile
FK150_JM17	Male	United States	GG	AC	GGAC	normal fertile
FK071_DV126	Male	United States	GG	AC	GGAC	normal fertile
FK140_JM07	Male	United States	GG	AC	GGAC	normal fertile
FK100_DV215	Male	United States	GG	AC	GGAC	normal fertile
FK005_DV025	Male	United States	GG	AC	GGAC	normal fertile
FK084_DV140	Male	United States	GG	AC	GGAC	normal fertile
FK080_DV136	Male	United States	GG	AC	GGAC	normal fertile
FK043_DV091	Male	United States	GG	AC	GGAC	normal fertile
FK073_DV128	Male	United States	GG	AC	GGAC	normal fertile
FK040_DV087	Male	United States	GG	AC	GGAC	normal fertile
FK025_DV072	Male	United States	GG	AC	GGAC	normal fertile
FK082_DV138	Male	United States	GG	AC	GGAC	normal fertile
FK134_JM01	Male	United States	GG	AC	GGAC	normal fertile
FK120_DV_C	Male	United States	GG	AC	GGAC	normal fertile
FK058_DV109	Male	United States	GG	AC	GGAC	normal fertile
FK136_JM03	Male	United States	GG	AC	GGAC	normal fertile
FK122_DV_E	Male	United States	GG	AC	GGAC	normal fertile
FK089_DV201	Male	United States	GG	AC	GGAC	n/a
FK027_DV074	Male	United States	GG	AC	GGAC	n/a
FK108_DV231	Male	United States	GG	AC	GGAC	n/a
FK194	Male	United States	GG	AC	GGAC	n/a

H099	Male	United States	GG	AC	GGAC	n/a
H178	Male	United States	GG	AC	GGAC	n/a
H181	Male	United States	GG	AC	GGAC	n/a
H182	Male	United States	GG	AC	GGAC	n/a
H250	Male	United States	GG	AC	GGAC	n/a
H251	Male	United States	GG	AC	GGAC	n/a
H277	Male	United States	GG	AC	GGAC	n/a
H538	Male	United States	GG	AC	GGAC	n/a
H575	Male	United States	GG	AC	GGAC	n/a
FK180	Male	United States	GG	AC	GGAC	n/a
FK187	Male	United States	GG	AC	GGAC	n/a
FK198	Female	United States	GG	AC	GGAC	n/a
H906	Male	United States	GG	AC	GGAC	n/a
FK211	Male	United States	GG	AC	GGAC	n/a
H306	Male	Unknown	GG	AC	GGAC	n/a
H308	Male	Unknown	GG	AC	GGAC	n/a
H312	Male	Unknown	GG	AC	GGAC	n/a
H315	Male	Unknown	GG	AC	GGAC	n/a
H902	Male	Unknown	GG	AC	GGAC	n/a
49672	Male	Venezuela	GG	AC	GGAC	n/a
49735	Female	Venezuela	GG	AC	GGAC	n/a
49742	Male	Venezuela	GG	AC	GGAC	n/a
49812	Male	Venezuela	GG	AC	GGAC	n/a
49862	Male	Venezuela	GG	AC	GGAC	n/a
50259	Female	Venezuela	GG	AC	GGAC	n/a
50685	Female	Venezuela	GG	AC	GGAC	n/a
53460	Male	Venezuela	GG	AC	GGAC	n/a
53536	Female	Venezuela	GG	AC	GGAC	n/a
54717	Female	Venezuela	GG	AC	GGAC	n/a
54750	Male	Venezuela	GG	AC	GGAC	n/a
54759	Male	Venezuela	GG	AC	GGAC	n/a
54766	Female	Venezuela	GG	AC	GGAC	n/a
54852	Male	Venezuela	GG	AC	GGAC	n/a
57367	Female	Venezuela	GG	AC	GGAC	n/a
57398	Female	Venezuela	GG	AC	GGAC	n/a
57539	Female	Venezuela	GG	AC	GGAC	n/a
59270	Male	Venezuela	GG	AC	GGAC	n/a
59302	Male	Venezuela	GG	AC	GGAC	n/a
59480	Male	Venezuela	GG	AC	GGAC	n/a
59514	Female	Venezuela	GG	AC	GGAC	n/a
59571	Male	Venezuela	GG	AC	GGAC	n/a

59945	Male	Venezuela	GG	AC	GGAC	n/a
61366	Female	Venezuela	GG	AC	GGAC	n/a
61375	Female	Venezuela	GG	AC	GGAC	n/a
61481	Female	Venezuela	GG	AC	GGAC	n/a
71160	Female	Venezuela	GG	AC	GGAC	n/a
71209	Female	Venezuela	GG	AC	GGAC	n/a
71222	Male	Venezuela	GG	AC	GGAC	n/a
71232	Female	Venezuela	GG	AC	GGAC	n/a
71235	Female	Venezuela	GG	AC	GGAC	n/a
71242	Male	Venezuela	GG	AC	GGAC	n/a
71253	Male	Venezuela	GG	AC	GGAC	n/a
71272	Male	Venezuela	GG	AC	GGAC	n/a
71278	Female	Venezuela	GG	AC	GGAC	n/a
71307	Male	Venezuela	GG	AC	GGAC	n/a
72421	Male	Venezuela	GG	AC	GGAC	n/a
72458	Male	Venezuela	GG	AC	GGAC	n/a
72607	Female	Venezuela	GG	AC	GGAC	n/a
72676	Male	Venezuela	GG	AC	GGAC	n/a
72977	Male	Venezuela	GG	AC	GGAC	n/a
73096	Female	Venezuela	GG	AC	GGAC	n/a
73179	Female	Venezuela	GG	AC	GGAC	n/a
73195	Male	Venezuela	GG	AC	GGAC	n/a
73304	Male	Venezuela	GG	AC	GGAC	n/a
76218	Female	Venezuela	GG	AC	GGAC	n/a
76269	Female	Venezuela	GG	AC	GGAC	n/a
76383	Female	Venezuela	GG	AC	GGAC	n/a
85394	Male	Venezuela	GG	AC	GGAC	n/a
85410	Female	Venezuela	GG	AC	GGAC	n/a
85427	Female	Venezuela	GG	AC	GGAC	n/a
85451	Female	Venezuela	GG	AC	GGAC	n/a
85493	Male	Venezuela	GG	AC	GGAC	n/a
FK008_DV028	Male	United States	GG	CC	GGCC	normal fertile
FK056_DV107	Male	United States	GG	CC	GGCC	normal fertile
FK094_DV207	Male	United States	GG	CC	GGCC	normal fertile
FK023_DV070	Male	United States	GG	CC	GGCC	normal fertile
FK127_DV_J	Male	United States	GG	CC	GGCC	normal fertile
FK041_DV088	Male	United States	GG	CC	GGCC	normal fertile
FK079_DV135	Male	United States	GG	CC	GGCC	normal fertile
FK088_DV177	Male	United States	GG	CC	GGCC	n/a
FK097_DV211	Male	United States	GG	CC	GGCC	n/a
H276	Male	United States	GG	CC	GGCC	n/a

FK178	Male	United States	GG	CC	GGCC	n/a
FK205	Male	United States	GG	CC	GGCC	n/a
50441	Female	Venezuela	GG	CC	GGCC	n/a
50765	Female	Venezuela	GG	CC	GGCC	n/a
53453	Male	Venezuela	GG	CC	GGCC	n/a
59192	Male	Venezuela	GG	CC	GGCC	n/a
59627	Male	Venezuela	GG	CC	GGCC	n/a
61554	Male	Venezuela	GG	CC	GGCC	n/a
71410	Female	Venezuela	GG	CC	GGCC	n/a
71415	Male	Venezuela	GG	CC	GGCC	n/a
72443	Female	Venezuela	GG	CC	GGCC	n/a
85536	Male	Venezuela	GG	CC	GGCC	n/a
85973	Female	Venezuela	GG	CC	GGCC	n/a

Supplementary Table S2. Summary information for the select cohort of 150 Thoroughbred breeding stallions from United States including FKBP6 exon 5 SNP genotypes, percent per-cycle pregnancy rate (PCPR), percent per-seasonal pregnancy rate (PSPR), Mare book (MB), fertility status, percent pedigree-based inbreeding coefficient (PBIC), and inbred status.

Horse ID	**Sex**	**Geographical Location**	**SNP1 (G>A)**	**SNP2 (A>C)**	**Combined genotype**	**PCPR (%)**	**PSPR (%)**	**MB**	**Fertility status**	**PBIC (%)**	**Inbred Status**
H302	Male	United States	AA	AA	AAAA	2	2	51	low fertility	0	NO
H815	Male	United States	AA	AA	AAAA	6	8	61	subfertile	0	NO
FK143_JM10	Male	United States	AA	AA	AAAA	29	52	48	subfertile	0.39	NO
FK085_DV142	Male	United States	AA	AA	AAAA	39	61	23	subfertile	0	NO
FK048_DV097	Male	United States	AA	AC	AAAC	51	81	103	normal fertile	0.2	NO
FK091_DV204	Male	United States	AA	AC	AAAC	52	81	156	normal fertile	0.98	NO
FK002_DV022	Male	United States	AA	AC	AAAC	55	89	57	normal fertile	0	NO
FK145_JM12	Male	United States	AA	AC	AAAC	56	85	62	normal fertile	0.59	NO
FK061_DV112	Male	United States	AA	AC	AAAC	62	91	163	normal fertile	2.15	NO
FK036_DV083	Male	United States	AA	AC	AAAC	64	92	222	normal fertile	1.17	NO
FK028_DV075	Male	United States	AA	AC	AAAC	64	86	161	normal fertile	1.96	NO
FK149_JM16	Male	United States	AA	AC	AAAC	74	92	101	normal fertile	0.39	NO
FK065_DV117	Male	United States	AA	CC	AACC	60	90	146	normal fertile	1.17	NO
FK101_DV216	Male	United States	AA	CC	AACC	66	90	123	normal fertile	0	NO
FK060_DV111	Male	United States	AA	CC	AACC	70	93	145	normal fertile	0.2	NO
FK139_JM06	Male	United States	AA	CC	AACC	71	93	74	normal fertile	1.76	NO
FK117_DV239	Male	United States	AA	CC	AACC	100	100	5	normal fertile	1.96	NO
FK029_DV076	Male	United States	AG	AA	AGAA	52	85	164	normal fertile	1.76	NO
FK031_DV078	Male	United States	AG	AA	AGAA	53	83	93	normal fertile	0	NO

FK051_DV101	Male	United States	AG	AA	AGAA	58	87	154	normal fertile	1.37	NO
FK118_DV_A	Male	United States	AG	AA	AGAA	60	90	31	normal fertile	1.37	NO
FK054_DV104	Male	United States	AG	AA	AGAA	62	93	152	normal fertile	0.98	NO
FK069_DV123	Male	United States	AG	AA	AGAA	63	89	129	normal fertile	0.39	NO
FK081_DV137	Male	United States	AG	AA	AGAA	66	89	28	normal fertile	3.32	NO
FK018_DV040	Male	United States	AG	AA	AGAA	66	94	143	normal fertile	0.59	NO
FK121_DV_D	Male	United States	AG	AA	AGAA	70	96	24	normal fertile	0	NO
FK124_DV_G	Male	United States	AG	AA	AGAA	70	88	96	normal fertile	1.56	NO
FK142_JM09	Male	United States	AG	AA	AGAA	71	93	29	normal fertile	0.78	NO
FK037_DV084	Male	United States	AG	AC	AGAC	35	61	33	low fertility	0.78	NO
FK057_DV108	Male	United States	AG	AC	AGAC	42	72	72	low fertility	0	NO
FK110_DV233	Male	United States	AG	AC	AGAC	42	76	70	low fertility	1.76	NO
FK126_DV_I	Male	United States	AG	AC	AGAC	43	81	58	low fertility	0	NO
FK015_DV035	Male	United States	AG	AC	AGAC	46	70	21	normal fertile	2.15	NO
FK049_DV098	Male	United States	AG	AC	AGAC	51	82	54	normal fertile	1.95	NO
FK070_DV125	Male	United States	AG	AC	AGAC	54	80	86	normal fertile	0	NO
FK067_DV120	Male	United States	AG	AC	AGAC	54	86	119	normal fertile	0.2	NO
FK033_DV080	Male	United States	AG	AC	AGAC	56	85	27	normal fertile	0.57	NO
FK152_JM19	Male	United States	AG	AC	AGAC	60	86	140	normal fertile	0.2	NO
FK086_DV174	Male	United States	AG	AC	AGAC	61	82	60	normal fertile	0.57	NO
FK093_DV206	Male	United States	AG	AC	AGAC	61	88	171	normal fertile	1.96	NO
FK006_DV026	Male	United States	AG	AC	AGAC	62	91	110	normal fertile	1.17	NO
FK038_DV085	Male	United States	AG	AC	AGAC	62	96	91	normal fertile	1.56	NO
FK045_DV094	Male	United States	AG	AC	AGAC	63	92	154	normal fertile	1.76	NO
FK137_JM04	Male	United States	AG	AC	AGAC	64	97	60	normal fertile	0.39	NO
FK141_JM08	Male	United States	AG	AC	AGAC	64	86	42	normal fertile	0.98	NO
FK016_DV036	Male	United States	AG	AC	AGAC	65	87	101	normal fertile	0.78	NO
FK050_DV100	Male	United States	AG	AC	AGAC	65	90	99	normal fertile	2.15	NO

FK035_DV082	Male	United States	AG	AC	AGAC	66	95	177	normal fertile	0.98	NO
FK055_DV105	Male	United States	AG	AC	AGAC	66	94	157	normal fertile	3.32	NO
FK039_DV086	Male	United States	AG	AC	AGAC	67	97	109	normal fertile	0	NO
FK146_JM13	Male	United States	AG	AC	AGAC	68	89	141	normal fertile	0.39	NO
FK113_DV235	Male	United States	AG	AC	AGAC	68	87	47	normal fertile	0.39	NO
FK163_DV	Male	United States	AG	AC	AGAC	69	97	143	normal fertile	0.59	NO
FK090_DV203	Male	United States	AG	AC	AGAC	70	91	161	normal fertile	0.39	NO
FK138_JM05	Male	United States	AG	AC	AGAC	70	91	82	normal fertile	0	NO
FK153_JM20	Male	United States	AG	AC	AGAC	74	94	67	normal fertile	0.78	NO
FK105_DV221	Male	United States	AG	AC	AGAC	76	91	126	normal fertile	0.39	NO
FK116_DV238	Male	United States	AG	AC	AGAC	78	93	59	normal fertile	1.76	NO
FK123_DV_F	Male	United States	AG	CC	AGCC	54	82	130	normal fertile	0	NO
FK014_DV034	Male	United States	AG	CC	AGCC	56	81	65	normal fertile	0.78	NO
FK021_DV045	Male	United States	AG	CC	AGCC	56	86	140	normal fertile	0.57	NO
FK083_DV139	Male	United States	AG	CC	AGCC	58	77	71	normal fertile	0	NO
FK004_DV024	Male	United States	AG	CC	AGCC	62	91	171	normal fertile	1.17	NO
FK157_DV	Male	United States	AG	CC	AGCC	64	89	152	normal fertile	1.17	NO
FK098_DV213	Male	United States	AG	CC	AGCC	65	89	137	normal fertile	0.2	NO
FK020_DV043	Male	United States	AG	CC	AGCC	65	92	153	normal fertile	0.2	NO
FK159_DV	Male	United States	AG	CC	AGCC	66	91	109	normal fertile	0	NO
FK047_DV096	Male	United States	AG	CC	AGCC	66	92	114	normal fertile	0.59	NO
FK107_DV230	Male	United States	AG	CC	AGCC	67	91	105	normal fertile	1.95	NO
FK059_DV110	Male	United States	AG	CC	AGCC	70	95	149	normal fertile	0.2	NO
FK003_DV023	Male	United States	AG	CC	AGCC	71	95	120	normal fertile	0	NO
FK162_JM	Male	United States	AG	CC	AGCC	71	93	190	normal fertile	0	NO
FK147_JM14	Male	United States	AG	CC	AGCC	73	89	18	normal fertile	0	NO
FK151_JM18	Male	United States	GG	AA	GGAA	44	74	128	low fertility	3.52	NO
FK114_DV236	Male	United States	GG	AA	GGAA	48	69	16	subfertile	3.13	NO

FK102_DV217	Male	United States	GG	AA	GGAA	49	76	148	subfertile	1.76	NO
FK068_DV121	Male	United States	GG	AA	GGAA	52	81	73	normal fertile	0.78	NO
FK017_DV038	Male	United States	GG	AA	GGAA	53	86	132	normal fertile	0.59	NO
FK158_DV	Male	United States	GG	AA	GGAA	53	84	110	normal fertile	0	NO
FK052_DV102	Male	United States	GG	AA	GGAA	54	88	124	normal fertile	0.39	NO
FK109_DV232	Male	United States	GG	AA	GGAA	59	88	25	normal fertile	1.17	NO
FK044_DV093	Male	United States	GG	AA	GGAA	60	90	163	normal fertile	0.39	NO
FK013_DV033	Male	United States	GG	AA	GGAA	61	84	68	normal fertile	2.15	NO
FK095_DV208	Male	United States	GG	AA	GGAA	62	91	117	normal fertile	0.39	NO
FK077_DV133	Male	United States	GG	AA	GGAA	62	85	46	normal fertile	0.78	NO
FK165_DV	Male	United States	GG	AA	GGAA	63	88	165	normal fertile	1.17	NO
FK076_DV131	Male	United States	GG	AA	GGAA	64	92	61	normal fertile	1.96	NO
FK160_JM	Male	United States	GG	AA	GGAA	67	91	165	normal fertile	0.78	NO
FK164_DV	Male	United States	GG	AA	GGAA	67	91	163	normal fertile	0.59	NO
FK119_DV_B	Male	United States	GG	AA	GGAA	68	84	50	normal fertile	0	NO
FK154_JM21	Male	United States	GG	AA	GGAA	69	96	28	normal fertile	0.78	NO
FK166_DV	Male	United States	GG	AA	GGAA	69	91	177	normal fertile	0.78	NO
FK144_JM11	Male	United States	GG	AA	GGAA	70	88	84	normal fertile	1.56	NO
FK148_JM15	Male	United States	GG	AA	GGAA	70	95	110	normal fertile	1.95	NO
FK135_JM02	Male	United States	GG	AA	GGAA	71	92	180	normal fertile	1.95	NO
FK026_DV073	Male	United States	GG	AA	GGAA	71	92	89	normal fertile	2.73	NO
FK030_DV077	Male	United States	GG	AA	GGAA	74	96	146	normal fertile	2.34	NO
FK062_DV114	Male	United States	GG	AA	GGAA	81	89	38	normal fertile	1.95	NO
FK155_JM22	Male	United States	GG	AA	GGAA	57	55	89	normal fertile	1.56	NO
FK024_DV071	Male	United States	GG	AC	GGAC	34	64	106	low fertility	0.2	NO
FK074_DV129	Male	United States	GG	AC	GGAC	36	57	68	low fertility	1.56	NO
FK066_DV119	Male	United States	GG	AC	GGAC	41	69	35	low fertility	0.78	NO
FK075_DV130	Male	United States	GG	AC	GGAC	42	75	53	low fertility	1.56	NO

FK034_DV081	Male	United States	GG	AC	GGAC	47	77	214	normal fertile	0.39	NO
FK009_DV029	Male	United States	GG	AC	GGAC	48	75	55	normal fertile	2.15	NO
FK032_DV079	Male	United States	GG	AC	GGAC	50	86	28	normal fertile	0.78	NO
FK046_DV095	Male	United States	GG	AC	GGAC	50	82	77	normal fertile	0	NO
FK078_DV134	Male	United States	GG	AC	GGAC	53	75	12	normal fertile	0	NO
FK104_DV219	Male	United States	GG	AC	GGAC	54	86	132	normal fertile	2.73	NO
FK096_DV210	Male	United States	GG	AC	GGAC	56	87	145	normal fertile	0.39	NO
FK125_DV_H	Male	United States	GG	AC	GGAC	57	82	105	normal fertile	2.15	NO
FK099_DV214	Male	United States	GG	AC	GGAC	57	82	123	normal fertile	2.54	NO
FK042_DV089	Male	United States	GG	AC	GGAC	59	91	184	normal fertile	0.2	NO
FK010_DV030	Male	United States	GG	AC	GGAC	59	86	95	normal fertile	1.37	NO
FK156_DV	Male	United States	GG	AC	GGAC	59	87	116	normal fertile	0.78	NO
FK092_DV205	Male	United States	GG	AC	GGAC	60	89	106	normal fertile	0.39	NO
FK019_DV041	Male	United States	GG	AC	GGAC	61	90	125	normal fertile	0	NO
FK103_DV218	Male	United States	GG	AC	GGAC	62	94	150	normal fertile	0.98	NO
FK011_DV031	Male	United States	GG	AC	GGAC	62	81	83	normal fertile	0.98	NO
FK007_DV027	Male	United States	GG	AC	GGAC	62	91	110	normal fertile	0.78	NO
FK012_DV032	Male	United States	GG	AC	GGAC	62	88	59	normal fertile	1.95	NO
FK087_DV175	Male	United States	GG	AC	GGAC	63	89	73	normal fertile	0.59	NO
FK064_DV116	Male	United States	GG	AC	GGAC	64	91	119	normal fertile	1.95	NO
FK053_DV103	Male	United States	GG	AC	GGAC	64	89	141	normal fertile	0.78	NO
FK150_JM17	Male	United States	GG	AC	GGAC	65	88	245	normal fertile	0.78	NO
FK071_DV126	Male	United States	GG	AC	GGAC	66	91	78	normal fertile	0.78	NO
FK140_JM07	Male	United States	GG	AC	GGAC	66	91	116	normal fertile	1.76	NO
FK100_DV215	Male	United States	GG	AC	GGAC	66	89	158	normal fertile	0.78	NO
FK005_DV025	Male	United States	GG	AC	GGAC	67	93	158	normal fertile	0.2	NO
FK084_DV140	Male	United States	GG	AC	GGAC	68	91	21	normal fertile	1.56	NO
FK080_DV136	Male	United States	GG	AC	GGAC	68	84	31	normal fertile	0	NO

FK043_DV091	Male	United States	GG	AC	GGAC	69	93	144	normal fertile	0	NO
FK073_DV128	Male	United States	GG	AC	GGAC	69	94	125	normal fertile	0.78	NO
FK040_DV087	Male	United States	GG	AC	GGAC	70	97	38	normal fertile	2.15	NO
FK025_DV072	Male	United States	GG	AC	GGAC	70	94	139	normal fertile	0	NO
FK082_DV138	Male	United States	GG	AC	GGAC	71	83	12	normal fertile	1.95	NO
FK134_JM01	Male	United States	GG	AC	GGAC	73	100	27	normal fertile	1.76	NO
FK120_DV_C	Male	United States	GG	AC	GGAC	73	95	57	normal fertile	0.59	NO
FK058_DV109	Male	United States	GG	AC	GGAC	76	95	152	normal fertile	0	NO
FK136_JM03	Male	United States	GG	AC	GGAC	81	96	27	normal fertile	0	NO
FK122_DV_E	Male	United States	GG	AC	GGAC	86	93	82	normal fertile	1.17	NO
FK008_DV028	Male	United States	GG	CC	GGCC	55	88	155	normal fertile	1.56	NO
FK056_DV107	Male	United States	GG	CC	GGCC	56	86	143	normal fertile	0.98	NO
FK094_DV207	Male	United States	GG	CC	GGCC	56	81	126	normal fertile	3.13	NO
FK023_DV070	Male	United States	GG	CC	GGCC	56	95	19	normal fertile	0.2	NO
FK127_DV_J	Male	United States	GG	CC	GGCC	57	82	71	normal fertile	0.4	NO
FK041_DV088	Male	United States	GG	CC	GGCC	64	94	47	normal fertile	3.13	NO
FK079_DV135	Male	United States	GG	CC	GGCC	66	92	25	normal fertile	0	NO
FK161_JM	Male	United States	AG	AA	AGAA	66	92	49	normal fertile	5.08	YES
FK072_DV127	Male	United States	GG	AC	GGAC	62	90	176	normal fertile	4.88	YES

Supplementary Table S3. Contingency analysis of low fertility (< 46% per-cycle pregnancy rate; PCPR) by genotype of individual SNPs and a combined genotype.

Contingency Table of low fertility by genotype SNP G>A			
	Low fertility (< 46% PCPR)		
	NO	YES	TOTAL
AA	13	4	17
AG	53	4	57
GG	71	5	76
TOTAL	137	13	150

Test	P-Value
Fischer's exact	0.0971

Contingency Table of low fertility by genotype SNP A>C			
	Low fertility (< 46% PCPR)		
	NO	YES	TOTAL
AA	36	5	42
AC	73	8	81
CC	27	0	27
TOTAL	136	13	150

Test	P-Value
Fischer's exact	0.1995

Contingency Table of low fertility by combined SNP genotype			
	Low fertility (< 46% PCPR)		
	NO	YES	TOTAL
A/A-A/A	0	4	4
A/A-A/C	8	0	8
A/A-C/C	5	0	5
A/G-A/A	12	0	12
A/G-A/C	26	4	30
A/G-C/C	15	0	15
G/G-A/A	25	1	26
G/G-A/C	39	4	43
G/G-C/C	7	0	7
TOTAL	136	13	150

Test	P-Value
Fischer's exact	0.0007

Supplementary Table S4. Individual copy numbers (CN) of the 216 male cohort used as a normal control. Individual information includes Y haplogroup (HG), Y haplotype (HT), and breed. CNs in green are the largest CN and CNs in red are the lowest CN per gene. Individuals in red are missing CNs. Missing CNs are denoted by blacked out cells.

Sample ID	Y HG	Y HT	Breed	*ETSTY1*	*ETSTY2*	*ETSTY5*	*HSFY*	*RBMY*	*SRY*	*TSPY*	*UBA1Y*
23346	Crown	Ad	Dales Pony	3.6	5.7	5.3	0.88	1.46	1	9.1	3.8
23347	Crown	Ad	Dales Pony	4.9	6.7	4.1	0.87	1.34	0.65	7.7	2.9
23348	Crown	Ad	Dales Pony	7.4	5.7	4.8	0.94	2.3	0.96	8.2	3.5
24109	Crown	Ad	Dales Pony	5.3	7.3	5	0.47	1.2	0.94	8.1	3.7
70858	Crown	Ad	Suffolk Punch	4.15	4.2	4.05	0.91	1.99	0.748	9.7	3.17
70980	Crown	Ad	Suffolk Punch	3.92	5.105	4.4	0.88	1.74	0.85	10.8	3.8
70981	Crown	Ad	Suffolk Punch	3.78	4.6	4.37	1.14	1.73	0.81	11.3	3.07
73901	Crown	Ad	Suffolk Punch	3.96	4.09	3.94	1.04	1.99	0.95	8.6	3.07
74413	Crown	Ad	Suffolk Punch	3.98	4.3	4.31	1.09	1.93	0.9	8.17	3.63
74836	Crown	Ad	Suffolk Punch	3.74	4.6	4.04	1.18	1.67	0.739	8.2	3.3
74837	Crown	Ad	Suffolk Punch	3.96	4.5	3.71	1	1.92	0.695	8.6	3.8
75052	Crown	Ad	Suffolk Punch	3.59	4.6	4.17	0.94	1.95	0.87	9.2	3.7
BP280	Crown	Ad	Estonian Native	4.2	4.7	4.6	0.89	1.78	0.86	11.4	3.35
BP282	Crown	Ad	Estonian Native	4	5.4	4.6	0.88	1.92	0.79	11.7	2.9
BP314	Crown	Ad	Estonian Native	0.98	7.3	22.9	2.6	1.83	0.79	27	12.3
BP315	Crown	Ad	Estonian Native	4.5	4.4	4.09	0.96	2	1.06	6.98	3.6
BP388	Crown	Ad	Estonian Native	5.6	4.5	4	0.99	2.85	1.55	12.9	3.07
BP389	Crown	Ad	Estonian Native	4	4.4	4.5	0.68	2	0.9	10.5	3.69
H061	Crown	Ad	Quarter Horse	4.07	4.47	3.57	1.14	1.85	0.84	8	3.76
H294	Crown	Ad	Quarter Horse	3.77	5.37	4.44	1.08	1.67	0.78	13.3	3.3

H787	Crown	Ad	Friesian	1.96	3.78	2.62	0.85	1.91	0.78	5.55	3.13
TR007	Crown	Ad	Haflinger	4.21	3.77	4.08	1	1.88	1.01	9	3.94
TR008	Crown	Ad	Haflinger	4.02	3.68	4.05	0.97	2.1	0.91	9.5	4.27
TR009	Crown	Ad	Noriker	3.95	3.84	3.84	1.07	2.04	0.93	8.2	4.04
H334	Crown	Am	Caspian	4.15	5.49	4.23	0.98	1.77	0.793	8.6	3.21
22443	Crown	Ao	Arabian	5.3	6.4	4.5	1.12	1.56	0.8	15.4	3.1
22456	Crown	Ao	Arabian	5.9	6.7	5	1.25	1.49	0.79	18.7	3.5
H172	Crown	Ao	Arabian	3.96	3.74	3.72	1.11	2.01	0.71	8.61	3.35
H206	Crown	Ao	Arabian	5.03	4.05	3.93	0.98	2.07	0.966	8.8	2.58
H255	Crown	Ao	Arabian	4.9	3.82	3.96	0.98	1.87	0.953	9	4.01
H284	Crown	Ao	Arabian	5	4.16	3.96	0.97	1.87	0.94	7	1.42
H469	Crown	Ao	Arabian	3.57	3.71	3.88	0.53	2.01	0.991	8.4	2.79
H634	Crown	Ao	Arabian	4.55	4.43	4.28	1.12	1.92	0.9	8.5	1.12
H653	Crown	Ao	Arabian	3.76	4.41	3.96	1.03	1.87	0.83	10.07	3.38
H674	Crown	Ao	Arabian	4.66	3.67	3.71	1.07	1.57	0.96	8.14	2.97
H693	Crown	Ao	Arabian	3.86	3.57	4.03	1.09	1.97	0.972	8.79	3.47
H728	Crown	Ao	Arabian	3.42	3.35	4.27	1.07	2.04	1	8.93	3.56
TR001	Crown	Ao	Lipizzan	5.42	3.86	3.95	0.88	1.93	0.99	8.5	4.01
TR002	Crown	Ao	Lipizzan	5.24	4.01	4	0.92	2.13	0.97	8.8	3.92
TR003	Crown	Ao	Lipizzan	5.28	4.01	3.95	0.78	1.83	0.96	8.7	4.09
TR004	Crown	Ao	Noriker	5.34	3.86	3.94	0.99	1.95	1.03	8.4	4.08
TR005	Crown	Ao	Noriker	5.08	4.09	3.86	0.94	1.89	0.93	8.8	4
TR006	Crown	Ao	Noriker	5.27	3.89	3.92	0.91	1.78	1.04	9.3	3.89
BP306	Non-Crown	Domwest1	Mongolian Horse	4.9	4.4	3.9	0.87	1.83	0.92	8.44	3.4
BP287	Non-Crown	Domwest3	Mongolian Horse	3.5	3.8	3.48	0.96	1.81	0.81	9.4	3.8
BP298	Non-Crown	Domwest3	Mongolian Horse	4.37	4.31	3.88	0.98	2.785	1.99	9.5	3.14
BP304	Non-Crown	Domwest3	Mongolian Horse	4.5	4.4	3.59	0.93	1.69	0.99	6.53	3.6

BP309	Non-Crown	Domwest3	Mongolian Horse	3.7	4.4	3.61	0.97	1.88	0.91	8.2	4.1
BP310	Non-Crown	Domwest3	Mongolian Horse	4.13	4.5	3.7	0.89	1.63	0.92	8.3	3.5
121576	Non-Crown	Domwest4	Zemaitukai	5.3	5.3	4.69	0.87	2.8	1.7	11.6	3.91
121579	Non-Crown	Domwest4	Zemaitukai	5.8	5.1	4.8	0.74	2.84	1.38	12.7	3.7
121581	Non-Crown	Domwest4	Zemaitukai	6.1	6.6	5.3	0.91	2.77	1.66	13.3	3.7
121587	Non-Crown	Domwest4	Zemaitukai	5	4.2	4.8	0.95	2.79	1.58	11.9	3.6
121589	Non-Crown	Domwest4	Zemaitukai	5.8	5.2	4.5	0.98	2.93	1.89	10.6	3.6
BP364	Non-Crown	Domwest4	Estonian Native	3.8	3.7			2.81	1.96	10.1	
BP378	Non-Crown	Domwest4	Estonian Native	6.8	4.9	4	0.79	3	2	14.7	3.18
BP383	Non-Crown	Domwest4	Estonian Native	3.35	4.39	3.62	1.09	2.92	1.88	8.9	3.36
BP384	Non-Crown	Domwest4	Estonian Native	5.4	4	4	0.99	2.66	1.49	9.9	4
BP385	Non-Crown	Domwest4	Estonian Native					2.5	1.9		
BP386	Non-Crown	Domwest4	Estonian Native	4.3	4.4	3.8	0.99	2.35	1.58	9.4	3.1
BP387	Non-Crown	Domwest4	Estonian Native	5.3	6	3.7	0.94	2.6	1.77	15.8	3.5
BP394	Non-Crown	Domwest4	Estonian Native	3.81	4.7	3.9	0.95	1.96	0.85	8.4	3.08
BP398	Non-Crown	Domwest4	Estonian Native	4.6	4.6	4.7	1.42	1.68	0.99	11.9	3.4
BP399	Non-Crown	Domwest4	Estonian Native	5	5	3.3	1.23	2.53	1.76	10.4	3.06
BP400	Non-Crown	Domwest4	Estonian Native	4.4	4.2	3.7	0.94	2.61	1.98	9.2	3.7
102328	Crown	Hs	Lipizzan	4.93	4.01	3.87	1.02	1.85	0.9658	8.25	4.02
TR013	Crown	Hs	Lipizzan	5.19	4.05	3.89	1.02	1.79	0.91	8.8	4.33
TR014	Crown	Hs	Lipizzan	5.16	4.16	3.81	0.89	1.6	0.96	9.1	3.86
TR015	Crown	Hs	Lipizzan	4.2	3.22	3.77	1	2.14	0.99	8.2	3.49

BP373	Non-Crown	I	Icelandic Horse	4.97	4	4.3	0.92	1.47	0.813	9.5	3.3
BP376	Non-Crown	I	Icelandic Horse	4.84	4.2	4.14	1.1	1.48	0.643	10	3.47
BP377	Non-Crown	I	Icelandic Horse	4.88	4.5	4.25	1.09	1.87	1.03	10.9	3.47
TR025	Non-Crown	I	Icelandic Horse	5.2	3.62	4.09	1.02	2.04	1.01	8.6	4.33
TR026	Non-Crown	I	Icelandic Horse	4.47	4.12	3.78	1	2	1	9.1	3.82
TR027	Non-Crown	I	Icelandic Horse	4.63	3.19	3.76	0.92	1.77	0.93	8.8	3.86
TR019	Non-Crown	M	Mongolian Horse	4.76	3.87	3.89	0.99	2.02	0.98	8.4	4.14
25802	Non-Crown	N	Shetlandpony	5.2	5	4.39	0.91	1.16	0.56	7.85	3.6
45379	Non-Crown	N	Shetlandpony	5.2	5	4.9	1.03	1.38	0.462	8.02	3.4
TR016	Non-Crown	N	Shetlandpony	4.87	3.64	3.97	0.96	2.03	1.03	9.1	3.81
TR017	Non-Crown	N	Shetlandpony	4.89	3.92	3.89	0.91	2.16	1.01	8.3	4.09
TR018	Non-Crown	N	Shetlandpony	5.32	3.68	3.9	0.91	1.91	0.94	8.6	4.09
BP360	Non-Crown	O	Mongolian Horse	4.86	4.26	3.6	1.24	0.995	n/a	8.64	3.85
TR020	Non-Crown	O	Mongolian Horse	4.7	3.65	4.19	0.94	1.06	1.87	8.14	4.14
TR021	Non-Crown	O	Mongolian Horse	5.17	4.01	4.02	0.84	1.01	1.95	9.1	3.69
TR028	Non-Crown	O	Yakutian	10.8	14	10.7	2.2	1.37	2.7	38	3.7
KULAN1	Non-Crown	outgroup	Kulan	2.47	3.5	2.9	1.71	0.75	0.84	n/a	n/a
KULAN2	Non-Crown	outgroup	Kulan	2.11	4.7	4.1	1.67	0.88	0.66	n/a	n/a
15758	Non-Crown	Pa	Heck horse	7.3	7.2	5.4	1.17	0.6	0.68	17.8	3.6
21150	Non-Crown	Pa	Heck horse	7.7	6.6	5.3	1.05	0.79	0.73	18.7	3.4
TR022	Non-Crown	Pa	Przewalski's horse	6.95	3.81	4.04	0.94	1.5	0.98	8.7	3.88
14855	Non-Crown	Pb	Przewalski's horse	2.83	2.98	4.65	1.16	1.04	1.06	8.1	3.19
NO003	Non-Crown	Pb	Przewalski's horse	2.71	3.34	3.98	1.21	0.95	0.71	9.6	2.26

TR023	Non-Crown	Pb	Przewalski's horse	3	2.99	3.98	0.9	0.95	1	7.69	2.91
TR024	Non-Crown	Pb	Przewalski's horse	2.78	2.8	3.88	0.9	1.05	0.89	7.92	2.96
121580	Crown	Ta	Zemaitukai	5.4	5.2	4.9	0.95	1.63	0.89	11.8	3.7
76621	Crown	Tb	Caspian	4.14	4.6	3.8	0.92	1.81	0.82	10.3	2.67
76640	Crown	Tb-1	Caspian	5.14	3.9	4.35	0.91	1.67	0.65	10.5	4
51091	Crown	Tb-d	Tennessee Walking	4.2	5.7	4.1	1.08	1.47	0.57	16.3	2.56
51094	Crown	Tb-d	Tennessee Walking	6.1	6.8	4.44	1.2	1.4	0.53	20.5	3.34
51100	Crown	Tb-d	Tennessee Walking	7	5.1	5	0.84	1.8	0.92	17.6	4.2
51163	Crown	Tb-d	Tennessee Walking	5.9	5.1	4.9	1.01	1.78	0.97	13.3	3.79
51164	Crown	Tb-d	Tennessee Walking	4.8	7.4	4.6	1.2	1.65	0.94	16.9	3.92
67394	Crown	Tb-d	Standardbred	7	4.9	3.8	0.76	2	0.5	10.8	3.5
67395	Crown	Tb-d	Standardbred	4.6	5.2	3.6	1.05	1.22	0.69	10.4	3.13
67396	Crown	Tb-d	Standardbred	5.2	6.1	4.8	1.36	1.46	0.65	7.1	2.9
67398	Crown	Tb-d	Standardbred	3.29	3.5	4.6	1.17	1.345	0.54	8.8	3
67399	Crown	Tb-d	Standardbred	4.6	5.2	4.3	1.11	1.27	0.57	8.8	2.79
67400	Crown	Tb-d	Standardbred	5.3	3.9	6.7	1.17	1.155	0.53	8.5	2.79
67401	Crown	Tb-d	Standardbred	5.39	6.3	4.61	1.06	1.53	0.705	9.1	2.95
67402	Crown	Tb-d	Standardbred	5.22	5.8	4.52	1.19	1.33	0.66	8.5	2.79
BP390	Crown	Tb-d	Estonian Native	3.6	4.3	3.8	0.98	1.8	0.71	9.2	3
BRAVO	Crown	Tb-d	Thoroughbred	4.916	4.5825	3.56	1.035	1.99	0.9	8.48	3.47
FK003_DV023	Crown	Tb-d	Thoroughbred	5.67	4.4	4.18	1.05	1.76	0.79	9.8	4.4
FK009_DV029	Crown	Tb-d	Thoroughbred	5.44	4.3	4.1	1.05	1.93	0.94	9.1	3.7
FK015_DV035	Crown	Tb-d	Thoroughbred	5.24	4.3	4.03	0.98	1.91	0.96	10.1	3.4
FK024_DV071	Crown	Tb-d	Thoroughbred	5.46	4.7	4.22	0.9	2.11	0.88	10	3.65
FK026_DV073	Crown	Tb-d	Thoroughbred	5.18	3.7	4.36	0.87	1.97	0.87	9.5	4.6

FK030_DV077	Crown	Tb-d	Thoroughbred	5.14	5.2	3.74	1.09	0.9	0.92	9.6	3.3
FK034_DV081	Crown	Tb-d	Thoroughbred	5.3	3.9	4.25	1.01	1.98	1.05	9	4
FK037_DV084	Crown	Tb-d	Thoroughbred	5.1	4.2	4.29	0.91	1.92	0.88	9.9	3.7
FK057_DV108	Crown	Tb-d	Thoroughbred	5.3	3.4	3.8	1	1.84	0.96	9.7	4.3
FK058_DV109	Crown	Tb-d	Thoroughbred	5.51	4.1	3.79	0.89	1.92	0.929	9.21	3.49
FK062_DV114	Crown	Tb-d	Thoroughbred	4.92	4.01	4.03	0.95	1.66	0.98	8.7	3.7
FK066_DV119	Crown	Tb-d	Thoroughbred	4.55	4.2	3.78	1.09	1.64	0.93	9.1	4
FK074_DV129	Crown	Tb-d	Thoroughbred	5.11	4.6	4.09	0.95	1.83	0.98	8.8	3.96
FK075_DV130	Crown	Tb-d	Thoroughbred	5.37	4.2	3.73	1.39	2	0.95	9.3	4.3
FK082_DV138	Crown	Tb-d	Thoroughbred	5	4.6	4.02	1.03	2.11	0.95	9.5	3.6
FK085_DV142	Crown	Tb-d	Thoroughbred	4.71	4.2	4.01	0.93	2.07	0.94	9.4	4.3
FK102_DV217	Crown	Tb-d	Thoroughbred	4.98	3.9	4.18	0.96	2.08	0.98	9.4	3.7
FK114_DV236	Crown	Tb-d	Thoroughbred	4.94	3.5	4.21	1.02	1.84	1.04	10.2	4.2
FK120_DV_C	Crown	Tb-d	Thoroughbred	4.98	4.1	4.12	1.08	1.67	1.07	9.6	2.8
FK122_DV_E	Crown	Tb-d	Thoroughbred	4.78	3.97	4.13	0.95	1.73	1.12	8.8	4.1
FK126_DV_I	Crown	Tb-d	Thoroughbred	5.1	3.8	3.62	0.71	1.78	0.94	9.2	3.8
FK189	Crown	Tb-d	Thoroughbred	3.57	3.82	3.34	1.09	1.96	0.88	8.3	3.17
H016	Crown	Tb-d	Caspian	5.23	4.1	4.09	0.957	2.26	0.974	8.74	3.94
H017	Crown	Tb-d	Caspian	4.93	3.88	4.14	0.96	1.99	0.945	8.6	4.09
H023	Crown	Tb-d	Quarter Horse	5.11	3.93	3.71	1	1.95	0.97	8.66	3.77
H042	Crown	Tb-d	Quarter Horse	4.04	4.65	4.47	1.04	1.79	0.85	11.2	3.77
H043	Crown	Tb-d	Quarter Horse	4.51	4.84	4.43	1.03	1.74	0.844	12.5	3.66
H044	Crown	Tb-d	Quarter Horse	5.1	6	4.6	1.06	1.36	0.72	15.4	3.08
H045	Crown	Tb-d	Quarter Horse	4.81	5.31	4.68	0.97	1.54	0.66	12.4	3.57
H046	Crown	Tb-d	Quarter Horse	4.29	4.66	4.32	1	1.82	0.86	10.8	3.83
H047	Crown	Tb-d	Quarter Horse	4.01	4.8	4.54	1.03	1.78	0.78	10.9	3.59
H048	Crown	Tb-d	Quarter Horse	4.6	4.65	4	0.98	1.84	0.81	9.7	3.91
H049	Crown	Tb-d	Paint	3.6	4.48	4.26	1.04	1.77	0.9	9.8	3.48

H050	Crown	Tb-d	Quarter-Morgan mix	3.59	4.5	4.12	0.98	1.79	0.93	10	3.51
H070	Crown	Tb-d	Quarter Horse	3.39	4	3.36	1.05	2.16	0.76	6.7	3.11
H292	Crown	Tb-d	Quarter Horse	4.96	4.01	3.9	0.93	1.82	0.97	8.5	3.79
H293	Crown	Tb-d	Quarter Horse	3.89	4.06	3.98	0.98	1.86	0.9	9.33	3.68
H295	Crown	Tb-d	Quarter Horse	5.07	6.3	5.3	0.99	1.7	0.68	17.1	3.29
H296	Crown	Tb-d	Quarter Horse	4.94	5.54	4.78	1.2	1.73	0.73	13.7	3.26
H297	Crown	Tb-d	Quarter Horse	6.1	6.11	5.32	1.16	1.56	0.74	15.7	3.76
H335	Crown	Tb-d	Caspian	4.79	5.21	3.86	1.1	1.79	0.85	11.3	3.5
H337	Crown	Tb-d	Quarter Horse	4.09	4.24	3.78	1.09	2.1	1.03	8.9	3.45
H343	Crown	Tb-d	Quarter Horse	3.82	3.9	3.91	1.03	2.23	0.92	9.5	2.47
H346	Crown	Tb-d	Quarter Horse	4.58	5.2	3.92	0.675	1.46	0.64	7.85	2.06
H400	Crown	Tb-d	Quarter Horse	4.64	3.92	3.985	1.02	2	1.014	9.2	3.7
H401	Crown	Tb-d	Quarter Horse	7.2	4.6	4.41	0.65	1.22	0.71	9.2	3.9
H410	Crown	Tb-d	Quarter Horse	4.89	2.9	3.67	0.72	1.62	0.78	9.42	1.88
H446	Crown	Tb-d	Quarter Horse	4.92	3.75	3.93	0.95	1.87	0.916	8.7	3.42
H593	Crown	Tb-d	Quarter Horse	4.26	4.08	4.24	1.15	2.08	0.94	8.3	3.45
H907	Crown	Tb-d	Quarter Horse	4.35	3.94	3.78	0.82	1.95	0.99	9.3	3.43
STO-0015	Crown	Tb-d	Thoroughbred	3.9	4.03	3.84	1.26	1.97	0.85	12.4	3.72
STO-0021	Crown	Tb-d	Thoroughbred	4.6	4.06	4.31	0.93	1.72	0.8	10.2	3.71
STO-0023	Crown	Tb-d	Thoroughbred	4.1	4.02	4.03	0.79	1.65	0.84	9.1	3.33
STO-0041	Crown	Tb-d	Thoroughbred	4.81	4.14	3.81	0.93	1.9	0.96	10.4	4.11
STO-0103	Crown	Tb-d	Thoroughbred	2.34	3.54	4.22	1.57	2.08	0.76	7.3	2.92
STO-0173	Crown	Tb-d	Thoroughbred	5.41	4.58	4.47	1.6	1.52	0.77	12.7	2.95
STO-0177	Crown	Tb-d	Thoroughbred	3.63	4.65	4.7	0.56	1.74	0.68	9.4	2.63
STO-0181	Crown	Tb-d	Thoroughbred	2.81	4.64	4.23	1.15	1.78	0.75	7.8	3.77
STO-0183	Crown	Tb-d	Thoroughbred	3.75	5.15	3.02	0.52	1.66	0.62	8.1	2.49
STO-0185	Crown	Tb-d	Thoroughbred	2.84	7.7	2.21	0.22	1.6	0.46	7.7	1.81
STO-0187	Crown	Tb-d	Thoroughbred	3.03	4.17	2.49	0.65	1.59	0.79	9.6	3.07

STO-0191	Crown	Tb-d	Thoroughbred	3.8	4.2	3.43	0.56	1.8	0.75	8.9	2.9
STO-0193	Crown	Tb-d	Thoroughbred	4.21	4.21	3.45	0.75	1.64	0.83	12.2	3.16
STO-0195	Crown	Tb-d	Thoroughbred	3.56	4.55	3.15	0.62	1.53	0.79	9.9	2.98
STO-0203	Crown	Tb-d	Thoroughbred	2.49	6.9	1.88	0.31	1.49	0.52	7.3	1.85
STO-0209	Crown	Tb-d	Thoroughbred	3.02	5.25	2.5	0.34	1.64	0.62	6.7	2.43
STO-0217	Crown	Tb-d	Thoroughbred	5.22	4.13	3.29	0.86	1.82	0.94	9.6	3.9
STO-0219	Crown	Tb-d	Thoroughbred	4.84	4.29	3.74	0.99	1.92	1.03	8.7	4.24
STO-0221	Crown	Tb-d	Thoroughbred	5.12	4.17	4.16	0.85	1.77	0.93	10	3.71
STO-0227	Crown	Tb-d	Thoroughbred	3.8	4.12	3	0.83	1.69	0.86	10.9	3.42
STO-0235	Crown	Tb-d	Thoroughbred	4.87	4.14	4.05	0.96	1.88	0.96	9.25	3.69
STO-0239	Crown	Tb-d	Thoroughbred	3.2	3.98	2.68	0.7	1.63	0.83	7.9	3.09
21151	Crown	Tb-o	Heck horse	6.2	4.9	5	1.06	1.54	0.95	11.4	4.2
21152	Crown	Tb-o	Heck horse	5.2	6.7	5.2	1	1.74	0.81	14.7	3.4
59872	Crown	Tb-o	Heck horse	5.3	6.5	5.1	1.08	1.71	0.69	14.4	3.5
64064	Crown	Tb-o	Heck horse	5	5.9	4.3	1.18	1.57	0.59	17.2	2.8
68983	Crown	Tb-o	Heck horse	4.3	5.4	4.6	1.19	1.44	0.76	13.3	4
121577	Crown	Tb-o	Zemaitukai	5	4.32	4.35	0.96	1.97	0.84	10.9	3.16
121578	Crown	Tb-o	Zemaitukai	5.1	5.6	4.8	0.88	1.8	0.85	12.8	4.1
121583	Crown	Tb-o	Zemaitukai	5.5	4.6	4.7	0.97	1.82	0.92	10.6	3.3
121584	Crown	Tb-o	Zemaitukai	4.85	4.11	4.15	1.03	1.85	0.94	10	3.39
BP001	Crown	Tb-o	Miniature Horse	3.65	3.91	3.7	0.815	1.51	0.92	8.6	3.48
BP002	Crown	Tb-o	Miniature Horse	3.76	3.9	4.18	0.828	1.75	0.8	9.3	3.41
BP372	Crown	Tb-o	Estonian Native	4	5.1	3.8	1.14	1.97	0.86	11.6	3.21
BP395	Crown	Tb-o	Estonian Native	5.3	4.6	4	1.05	2.52	1.6	10.9	3.7
BP396	Crown	Tb-o	Estonian Native	4.8	8.1	3.8	0.94	1.81	0.99	8.1	3.4
H051	Crown	Tb-o	Quarter Horse	4.58	4.58	4.6	0.96	1.71	0.722	11	3.73

H052	Crown	Tb-o	Quarter Horse	4.79	3.14	3.66	1.29	1.94	0.893	9.58	3.89
H300	Crown	Tb-o	Quarter Horse	5.01	5.49	4.6	0.96	1.76	0.71	13.6	3.47
STO-0095	Crown	Tb-o	Thoroughbred	4.81	3.93	3.68	0.84	1.75	0.9	9.3	3.6
TR010	Crown	Tb-o	Lipizzan	5.39	3.79	4	0.97	1.88	0.9	9	3.95
TR011	Crown	Tb-o	Lipizzan	5.23	3.8	3.76	0.91	1.85	0.98	8.9	4.15
TR012	Crown	Tb-o	Lipizzan	4.93	3.54	3.95	1.08	2.05	0.98	9	4.06
BP276	Crown	Tu	Estonian Native	2.19	8	3.57	2.6	1.76	0.82	7.5	5
BP277	Crown	Tu	Estonian Native	5.03	5.6	4.7	0.94	1.73	0.82	13	3.7
BP312	Crown	Tu	Estonian Native	2.5	9.8	29	2	1.76	0.78	30	12.2
BP313	Crown	Tu	Estonian Native	4.9	4.9	3.9	1.02	1.64	0.75	6.85	3
BP365	Crown	Tu	Estonian Native	5.05	4.4	4.22	1.08	2.19	0.9	11.6	3.92
BP379	Crown	Tu	Estonian Native	6.3	12.6	5.5	1.06	2.685	1.74	36	3.5
BP380	Crown	Tu	Estonian Native	5.9	12.6	5.5	1.25	1.53	0.9	20	3.7
BP392	Crown	Tu	Estonian Native	6.1	9.4	4.4	0.73	1.87	0.99	9.4	2.8
YT007	Non-Crown	Y	Yakutian	5.22	4.12	3.95	0.89	1.88	0.87	9.23	4.2
YT024	Non-Crown	Y	Yakutian	2.96	3.6	4.11	0.92	2.12	1.08	7.04	2.95
YT032	Non-Crown	Y	Yakutian	4.88	3.9	3.55	0.88	1.74	0.95	8.61	3.5

Supplementary Table S5. Information about 69 abnormal male horses.

Animal ID	Breed	Karyotype	*SRY* PCR	Phenotype
H217	American Quarter Horse	64,XY	positive	Unilateral Cryptorchid
H221	American Quarter Horse	64,XY	positive	Unilateral Cryptorchid
H222	American Quarter Horse	64,XY	positive	Unilateral Cryptorchid
H335	American Quarter Horse	n/a	positive	Unilateral Cryptorchid
H373	American Quarter Horse	n/a	positive	Unilateral Cryptorchid
H390	American Quarter Horse	n/a	positive	Bilateral Cryptorchid
H441	American Quarter Horse	n/a	positive	Bilateral Cryptorchid
H447	American Quarter Horse	n/a	positive	Unilateral Cryptorchid
H451	American Quarter Horse	n/a	positive	Bilateral Cryptorchid
H487	American Quarter Horse	n/a	positive	Unilateral Cryptorchid
H513	American Quarter Horse	n/a	positive	Unilateral Cryptorchid
H528	American Quarter Horse	64,XY	positive	Unilateral Cryptorchid
H571	American Quarter Horse	64,XY	positive	Unilateral Cryptorchid
H572	American Quarter Horse	64,XY	positive	Unilateral Cryptorchid
H573	American Quarter Horse	64,XY	positive	Unilateral Cryptorchid
H582	American Quarter Horse	n/a	positive	Unilateral Cryptorchid
H596	American Quarter Horse	64,XY	positive	Unilateral Cryptorchid
H614	American Quarter Horse	64,XY	positive	Bilateral Cryptorchid
H615	American Quarter Horse	64,XY	positive	Unilateral Cryptorchid
H616	American Quarter Horse	64,XY	positive	Unilateral Cryptorchid
H638	American Quarter Horse	64,XY	positive	Bilateral Cryptorchid
H640	American Quarter Horse	64,XY	positive	Unilateral Cryptorchid
H643	American Quarter Horse	64,XY	positive	Unilateral Cryptorchid
H646	American Quarter Horse	64,XY	positive	Bilateral Cryptorchid
H949	Appaloosa	64,XY	negative	female, SRY-negative sex reversal
H089	Arabian	64,XY	negative	female, SRY-negative sex reversal
H266	Arabian	64,XY	negative	female, SRY-negative sex reversal
H184	American Quarter Horse	64,XY	negative	female, SRY-negative sex reversal
H909	American Quarter Horse	64,XY	negative	female, SRY-negative sex reversal
H360	Standardbred	64,XY	negative	female, SRY-negative sex reversal

H947	Tennessee Walking Horse	64,XY	negative	female, SRY-negative sex reversal
H199	Thoroughbred	64,XY	negative	female, SRY-negative sex reversal
H227	Thoroughbred	64,XY	negative	female, SRY-negative sex reversal
H748	Thoroughbred	64,XY	negative	female, SRY-negative sex reversal
H844	Thoroughbred	64,XY	negative	female, SRY-negative sex reversal
H921	Thoroughbred	64,XY	negative	female, SRY-negative sex reversal
H806	Appaloosa	64,XY	positive	female-like, SRY-positive sex reversal
H169	Appaloosa	64,XY	positive	female-like, SRY-positive sex reversal
H252	Mixed Warmblood	64,XY	positive	female-like, SRY-positive sex reversal
H038	American Quarter Horse	64,XY	positive	female-like, SRY-positive sex reversal
H348	Standardbred	64,XY	positive	female-like, SRY-positive sex reversal
H369	Standardbred	64,XY	positive	female-like, SRY-positive sex reversal
H224	Thoroughbred	64,XYdel	negative	female-like, SRY-positive sex reversal
H737	American Quarter Horse	64,XY	positive	intersex
H826	American Quarter Horse	64,XY	positive	intersex
H145	American Quarter Horse	64,XY	positive	intersex
H423	American Quarter Horse	64,XY	positive	intersex
H690	American Quarter Horse	64,XYdel	negative	intersex
H543	Standardbred	64,XY	positive	intersex
H545	Standardbred	64,XY	positive	intersex
H544	Tennessee Walking Horse	64,XY	positive	intersex
H141	Thoroughbred	64,XYdel	negative	intersex
H546	Thoroughbred	64,XYdel	negative	intersex
H172	Arabian	64, XY	positive	subfertile
H206	Arabian	64, XY	positive	sterile
H255	Arabian	64, XY	positive	subfertile
H284	Arabian	64, XY	positive	subfertile
H320	Connemera	64,XY	positive	sterile
H343	American Quarter horse	64, XY	positive	subfertile
H620	Appaloosa	64,XY	positive	sterile

H653	Arabian	64, XY	positive	subfertile
H693	Arabian	64, XY	positive	subfertile
H726	Standardbred	64,XY	positive	subfertile
H787	Fresian	64, XY;t(Y;13p)	positive	sterile
H962	Arabian	64,XY	positive	sterile
H963	Arabian	64,XY;t(12q; 25q), der(12p)	positive	subfertile
H004	American Quarter horse	64,XY	positive	subfertile

Supplementary Table S6. Variant information used for MSY genotyping to determine Y chromosome haplotypes.

Variant ID	Position on LipY764 reference	Ancestral Allele	Derived Allele	Reference	Study first detected	Variant type	Determined HT (HG) in genotyping
fFQ	LipY764_contig226:3883	C	A	C	Felkel et al., 2019	SNV	P
fJI	LipY764_contig323:8668	A	T	A	Felkel et al., 2019	SNV	Pa
fU	LipY764_contig81:19955	G	A	G	Felkel et al., 2019	SNV	Pb
fAR	LipY764_contig197:1982	T	A	A	Felkel et al., 2019	SNV	Dom_All
sCO	LipY764_contig178:7326	A	C	A	Felkel et al., 2018	SNV	O
sES	LipY764_contig259:10235	C	A	A	Felkel et al., 2018	SNV	DomWest1
rBA	LipY764_contig10:19042	T	C	T	Wallner et al., 2017	SNV	N
rBG	LipY764_contig4:19212	CTT	CT	CT	Wallner et al., 2017	indel	DomWest2
sAA	LipY764_contig312:5283	G	C	G	Felkel et al., 2018	SNV	M
rBF	LipY764_contig26:10753	C	T	T	Wallner et al., 2017	SNV	DomWest3
rAW	LipY764_contig20:37008	T	C	T	Wallner et al., 2017	SNV	I
sAZ	LipY764_contig158:6198	C	T	C	Felkel et al., 2018	SNV	J
sQ	LipY764_contig90:21274	A	G	A	Felkel et al., 2018	SNV	Y
rAY	LipY764_contig60:5826	G	A	A	Wallner et al., 2017	SNV	DomWest4
rAX	LipY764_contig359:7646	T	C	C	Wallner et al., 2017	SNV	Crown
fYR	LipY764_contig312:3646	T	A	A	Felkel et al., 2019	SNV	H
fSQ	LipY764_contig49:30278	C	T	T	Felkel et al., 2019	SNV	Hs
sQB	LipY764_contig9:14318	T	A	T	Felkel et al., 2018	SNV	Hc
rW	LipY764_contig79:13451	G	A	G	Wallner et al., 2017	SNV	A
rX	LipY764_contig296:1984	G	T	G	Wallner et al., 2017	SNV	Ao
sQF	LipY764_contig4:24776	G	T	G	Felkel et al., 2018	SNV	Am
rAF	LipY764_contig112:2021	G	A	G	Wallner et al., 2017	SNV	Ad
rA	LipY764_contig54:27348	T	A	T	Wallner et al., 2017	SNV	T
fVZ	LipY764_contig169:2523	G	C	G	Felkel et al., 2019	SNV	T2

rT	LipY764_contig111:2628	C	T	C	Wallner et al., 2017	SNV	Tu
sPZ	LipY764_contig4:39228	A	T	A	Felkel et al., 2018	SNV	Ta
rB	LipY764_contig233:6238	C	G	C	Wallner et al., 2017	SNV	Tb
fWU	LipY764_contig198:17295	T	A	T	Felkel et al., 2019	SNV	Tb-o
fAAC	LipY764_contig468:3932	A	T	A	Felkel et al., 2019	SNV	Tb-1
rC	LipY764_contig213:4369	T	C	T	Wallner et al., 2017	SNV	Tb-d

Supplementary Table S7. Haplotype variant information for the 216-cohort used in this study. Brd – breed; HG – haplogroup; HT – haplotype; S-HT – sub-haplotype; HT-V – haplotype determining variant; C – crown; NC – non-crown; OG- outgroup; DP – Dales Pony; EN – Estonian native horse; FR – Friesian; HF – Haflinger; NR – Noriker; AR – Arabian; CS– Caspian; HH – Heck horse; IC – Icelandic horse; EHK – Kulan; LI – Lipizzan; MH – Miniature horse; PH – Przewalski's horse; PN – Paint; WH – Quarter horse; MO – Mongolian horse; QM – Quarter-morgan mix; SP – Shetland pony; ST – Standardbred; SP – Suffolk Punch; TW – Tennessee Walking; TB – Thoroughbred; YK – Yakutian; ZM - Zemaitukai

ID	Brd	HG	HT	Sub-HT	HT-V	fFQ	fJI	fU	fAR	sCO	sES	rBA	rBG	sAA	rBF	rAW	sAZ	sQ	rAY	rAX	fYR	fSQ	sQB	rW	rX	sQF	rAF	rA	fVZ	rT	sPZ	rB	fWU	fAAC	rC
23346	DP	C	Ad	Ad-b*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
23347	DP	C	Ad	Ad-b*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
23348	DP	C	Ad	Ad-b*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
24109	DP	C	Ad	Ad-b*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
BP280	EN	C	Ad	Ad-b*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
BP282	EN	C	Ad	Ad-b*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
BP388	EN	C	Ad	Ad-b*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
BP389	EN	C	Ad	Ad-b*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
BP314	EN	C	Ad	Ad-b*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
BP315	EN	C	Ad	Ad-b*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
H787	FR	C	Ad	Ad-b*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
TR7	HF	C	Ad	Ad-b*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
TR8	HF	C	Ad	Ad-h*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
TR9	NR	C	Ad	Ad-h*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
H061	QH	C	Ad	Ad-hA1	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
H294	QH	C	Ad	Ad-hA1	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
70858	SP	C	Ad	Ad-h*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
70980	SP	C	Ad	Ad-h*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
70981	SP	C	Ad	Ad-h*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
73901	SP	C	Ad	Ad-h*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
74413	SP	C	Ad	Ad-bN	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
74836	SP	C	Ad	Ad-h*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
74837	SP	C	Ad	Ad-h*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
75052	SP	C	Ad	Ad-h*	rAF[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	G	A	T	G	C	A	C	T	A	T
H334	CS	C	Am	Am-a*	sQF[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	G	T	G	T	G	C	A	C	T	A	T
22443	AR	C	Ao	Ao-aA1a*	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
H172	AR	C	Ao	Ao-aD	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
H206	AR	C	Ao	Ao-aA1a*	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
H255	AR	C	Ao	Ao-aA1a*	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
H284	AR	C	Ao	Ao-aA1a*	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
H634	AR	C	Ao	Ao-aA1a*	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
H653	AR	C	Ao	Ao-aD	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
H674	AR	C	Ao	Ao-aA1a*	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
H693	AR	C	Ao	Ao-aA1a*	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
H728	AR	C	Ao	Ao-aA1a*	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
22456	AR	C	Ao	Ao-aA1a*	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
H469	AR	C	Ao	Ao-aA1a*	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
TR1	LI	C	Ao	Ao-aA1a*	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
TR2	LI	C	Ao	Ao-aA1a*	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
TR3	LI	C	Ao	Ao-aA1a*	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
TR4	NR	C	Ao	Ao-nM1b	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T

TR5	NR	C	Ao	Ao-nM1b	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
TR6	NR	C	Ao	Ao-nM1b	rX[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	A	T	G	G	T	G	C	A	C	T	A	T
BP306	MO	NC	DomWest1	Domwest1	sES[A]	C	A	G	A	A	A	T	CTT	G		T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP287	MO	NC	DomWest3	Domwest3	rBF[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP298	MO	NC	DomWest3	Domwest3	rBF[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP304	MO	NC	DomWest3	Domwest3	rBF[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP309	MO	NC	DomWest3	Domwest3	rBF[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP310	MO	NC	DomWest3	Domwest3	rBF[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP383	EN	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP364	EN	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP378	EN	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP384	EN	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP385	EN	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP386	EN	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP387	EN	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP394	EN	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP398	EN	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP399	EN	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP400	EN	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
121576	ZM	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
121579	ZM	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
121581	ZM	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
121587	ZM	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
121589	ZM	NC	DomWest4	Domwest4	rAY[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
TR13	LI	C	Hs	Hs-aA	fSQ[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	A	T	T	G	G	G	G	T	G	C	A	C	T	A	T
TR14	LI	C	Hs	Hs-aA	fSQ[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	A	T	T	G	G	G	G	T	G	C	A	C	T	A	T
TR15	LI	C	Hs	Hs-aA	fSQ[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	A	T	T	G	G	G	G	T	G	C	A	C	T	A	T
102328	LI	C	Hs	Hs-aA	fSQ[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	A	T	T	G	G	G	G	T	G	C	A	C	T	A	T
TR25	IC	NC	I	I*	rAW[C]	C	A	G	A	A	A	T	CT	G	T	C	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
TR26	IC	NC	I	I*	rAW[C]	C	A	G	A	A	A	T	CT	G	T	C	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
TR27	IC	NC	I	I*	rAW[C]	C	A	G	A	A	A	T	CT	G	T	C	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP373	IC	NC	I	I*	rAW[C]	C	A	G	A	A	A	T	CT	G	T	C	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP376	IC	NC	I	I*	rAW[C]	C	A	G	A	A	A	T	CT	G	T	C	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP377	IC	NC	I	I*	rAW[C]	C	A	G	A	A	A	T	CT	G	T	C	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
TR19	MO	NC	M	M	sAA[C]	C	A	G	A	A	A	T	CT	C	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
TR16	SP	NC	N	Ns	rBA[C]	C	A	G	A	A	A	C	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
TR17	SP	NC	N	Ns	rBA[C]	C	A	G	A	A	A	C	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
TR18	SP	NC	N	Ns	rBA[C]	C	A	G	A	A	A	C	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
25802	SP	NC	N	Ns	rBA[C]	C	A	G	A	A	A	C	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
45379	SP	NC	N	Ns	rBA[C]	C	A	G	A	A	A	C	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
BP360	MO	NC	O	Oa	sCO[C]	C	A	G	A	C	C	T	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
TR20	MO	NC	O	Oa-b	sCO[C]	C	A	G	A	C	C	T	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T

TR21	MO	NC	O	Oa-b	sCO[C]	C	A	G	A	C	C	T	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
TR28	YK	NC	O	Oa-b	sCO[C]	C	A	G	A	C	C	T	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
KULAN1	EHK	OG	Outgroup	outgroup	fAR[A], sCO[C]	C	A	G	T	A	A	T	CTT	G	C	T	C	A	G	T	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
KULAN2	EHK	OG	Outgroup	outgroup	fAR[A], sCO[C]	C	A	G	T	A	A	T	CTT	G	C	T	C	A	G	T	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
15758	HH	NC	Pa	PRZ-a	fJI[T]	A	T	G	T	A	C	T	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
21150	HH	NC	Pa	PRZ-a	fJI[T]	A	T	G	T	A	C	T	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
TR22	PH	NC	Pa	PRZ-a	fJI[T]	A	T	G	T	A	C	T	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
14855	PH	NC	Pb	PRZ-b	fU[A]	A	A	A	T	A	C	T	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
NO3	PH	NC	Pb	PRZ-b	fU[A]	A	A	A	T	A	C	T	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
TR23	PH	NC	Pb	PRZ-b	fU[A]	A	A	A	T	A	C	T	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
TR24	PH	NC	Pb	PRZ-b	fU[A]	A	A	A	T	A	C	T	CTT	G	C	T	C	A	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
121580	ZM	C	Ta	Ta	sPZ[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	T	G	T	A	T
76621	CS	C	Tb	Tb*	rB[G]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	A	T
76640	CS	C	Tb-1	Tb-1*	fAAC[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	T
H16	CS	C	Tb-d	Tb-dM	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H17	CS	C	Tb-d	Tb-dM	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H335	CS	C	Tb-d	Tb-dM	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
BP390	EN	C	Tb-d	Tb-dM	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H049	PN	C	Tb-d	Tb-dM	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H023	QH	C	Tb-d	Tb-dM	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H042	QH	C	Tb-d	Tb-dM	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H043	QH	C	Tb-d	Tb-dM	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H044	QH	C	Tb-d	Tb-dM	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H045	QH	C	Tb-d	Tb-dM	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H046	QH	C	Tb-d	Tb-dM	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H047	QH	C	Tb-d	Tb-dM	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H048	QH	C	Tb-d	Tb-dM	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H070	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H292	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H293	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H295	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H296	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H297	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H337	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H343	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H346	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H400	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H401	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H410	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H446	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H593	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C

H907	QH	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
H050	QM	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
67394	ST	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
67395	ST	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
67396	ST	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
67398	ST	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
67399	ST	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
67400	ST	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
67401	ST	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
67402	ST	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
51091	TW	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
51094	TW	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
51100	TW	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
51163	TW	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
51164	TW	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
BRAVO	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK003_DV023	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK009_DV029	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK015_DV035	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK024_DV071	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK026_DV073	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK030_DV077	TB	C	Tb-d	Tb-d	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK034_DV081	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK037_DV084	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK057_DV108	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK058_DV109	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK062_DV114	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK066_DV119	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK074_DV129	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C

FK074_DV129	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK075_DV130	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK082_DV138	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK085_DV142	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK102_DV217	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK114_DV236	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK120_DV_C	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK122_DV_E	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK126_DV_I	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
FK189	TB	C	Tb-d	Tb-dW4	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0015	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0021	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0023	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0041	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0103	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0173	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0177	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0181	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0183	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0185	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0187	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0191	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0193	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0195	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0203	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0209	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0217	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0219	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0221	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0227	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0235	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
STO-0239	TB	C	Tb-d	Tb-dW1	rC[C]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	T	T	C
BP372	EN	C	Tb-o	Tb-o	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
BP395	EN	C	Tb-o	Tb-o	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T

BP396	EN	C	Tb-o	Tb-o	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
21151	HH	C	Tb-o	Tb-oB4	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
21152	HH	C	Tb-o	Tb-oB4	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
59872	HH	C	Tb-o	Tb-oB4	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
64064	HH	C	Tb-o	Tb-oB3b	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
68983	HH	C	Tb-o	Tb-oB3b1b	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
TR10	LI	C	Tb-o	Tb-o	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
TR12	LI	C	Tb-o	Tb-o	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
TR11	LI	C	Tb-o	Tb-o	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
BP1	MH	C	Tb-o	Tb-o	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
BP2	MH	C	Tb-o	Tb-o	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
H051	QH	C	Tb-o	Tb-o	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
H300	QH	C	Tb-o	Tb-oB3b1a	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
H052	QH	C	Tb-o	Tb-oB3b1a	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
STO-0095	TB	C	Tb-o	Tb-oB3b	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
121577	ZM	C	Tb-o	Tb-oB3b1b	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
121578	ZM	C	Tb-o	Tb-oL	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
121583	ZM	C	Tb-o	Tb-oL1	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
121584	ZM	C	Tb-o	Tb-oL	fWU[A]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	C	A	G	A	A	T
BP276	EN	C	Tu	Tu*	rT[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	T	A	C	T	A	T
BP277	EN	C	Tu	Tu*	rT[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	T	A	C	T	A	T
BP312	EN	C	Tu	Tu*	rT[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	T	A	C	T	A	T
BP313	EN	C	Tu	Tu*	rT[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	T	A	C	T	A	T
BP365	EN	C	Tu	Tu*	rT[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	T	A	C	T	A	T
BP379	EN	C	Tu	Tu*	rT[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	T	A	C	T	A	T
BP380	EN	C	Tu	Tu*	rT[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	T	A	C	T	A	T
BP392	EN	C	Tu	Tu*	rT[T]	C	A	G	A	A	A	T	CT	G	T	T	C	A	A	T	T	C	T	G	G	G	G	A	C	T	A	C	T	A	T
YT24	YK	NC	Y	Y	sQ[G]	C	A	G	A	A	A	T	CT	G	T	T	C	G	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
YT32	YK	NC	Y	Y	sQ[G]	C	A	G	A	A	A	T	CT	G	T	T	C	G	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T
YT7	YK	NC	Y	Y	sQ[G]	C	A	G	A	A	A	T	CT	G	T	T	C	G	G	C	T	C	T	G	G	G	G	T	G	C	A	C	T	A	T

Supplementary Table S8. Average copy number (CN) and corresponding standard deviation (SD) of 7 ampliconic genes and SRY by Y haplotype (HT). Green numbers denote the largest CN value and red numbers denote the lowest CN value per gene.

Y HG	Y HT	N	TSPY CN	TSPY SD	ETSTY1 CN	ETSTY1 SD	ETSTY2 CN	ETSTY2 SD	ETSTY5 CNV	ETSTY5 SD	SRY CN	SRY SD	RBMY CN	RBMY SD	HSFY CN	HSFY SD	UBA1Y CN	UBA1Y SD
Crown	Ad	24	10.07	4.05	4.06	1.15	4.88	1.03	4.98	3.85	0.89	0.17	1.88	0.32	1.02	0.37	3.87	1.83
Crown	Am	1	8.60	n/a	4.15	n/a	5.49	n/a	4.23	n/a	0.79	n/a	1.77	n/a	0.98	n/a	3.21	n/a
Crown	Ao	18	9.60	2.83	4.75	0.73	4.21	0.90	4.05	0.31	0.93	0.09	1.88	0.18	0.99	0.16	3.29	0.87
Non-Crown	Domwest1	1	8.44	n/a	4.90	n/a	4.40	n/a	3.90	n/a	0.92	n/a	1.83	n/a	0.87	n/a	3.40	n/a
Non-Crown	Domwest3	5	8.39	1.20	4.04	0.43	4.28	0.28	3.65	0.15	1.12	0.49	1.96	0.47	0.95	0.04	3.63	0.36
Non-Crown	Domwest4	16	11.25	2.16	4.98	0.95	4.82	0.76	4.20	0.59	1.62	0.34	2.58	0.35	0.99	0.17	3.49	0.31
Crown	Hs	4	8.59	0.44	4.87	0.46	3.86	0.43	3.84	0.06	0.96	0.03	1.85	0.22	0.98	0.06	3.93	0.35
Non-Crown	I	6	9.48	0.86	4.83	0.26	3.94	0.47	4.05	0.23	0.90	0.15	1.77	0.25	1.01	0.08	3.71	0.37
Non-Crown	M	1	8.40	n/a	4.76	n/a	3.87	n/a	3.89	n/a	0.98	n/a	2.02	n/a	0.99	n/a	4.14	n/a
Non-Crown	N	5	8.37	0.50	5.10	0.20	4.25	0.69	4.21	0.44	0.80	0.27	1.73	0.43	0.94	0.05	3.80	0.30
Non-Crown	O	4	15.97	14.69	6.38	2.95	6.48	5.02	5.63	3.39	2.17	0.46	1.11	0.18	1.31	0.62	3.85	0.21
Non-Crown	outgroup	2	n/a	n/a	2.29	0.25	4.10	0.85	3.50	0.85	0.75	0.13	0.82	0.09	1.69	0.03	n/a	n/a
Non-Crown	Pa	3	15.07	5.53	7.32	0.38	5.87	1.81	4.91	0.76	0.80	0.16	0.96	0.47	1.05	0.12	3.63	0.24
Non-Crown	Pb	4	8.33	0.86	2.83	0.12	3.03	0.23	4.12	0.35	0.92	0.15	1.00	0.06	1.04	0.17	2.83	0.40
Crown	Ta	1	11.80	n/a	5.40	n/a	5.20	n/a	4.90	n/a	0.89	n/a	1.63	n/a	0.95	n/a	3.70	n/a
Crown	Tb	1	10.50	n/a	5.14	n/a	3.90	n/a	4.35	n/a	0.65	n/a	1.67	n/a	0.91	n/a	4.00	n/a
Crown	Tb-1	1	10.30	n/a	4.14	n/a	4.60	n/a	3.80	n/a	0.82	n/a	1.81	n/a	0.92	n/a	2.67	n/a
Crown	Tb-d	87	10.09	2.51	4.64	0.93	4.59	0.89	4.02	0.67	0.83	0.15	1.75	0.25	0.96	0.23	3.43	0.59
Crown	Tb-o	21	11.15	2.39	4.89	0.59	4.85	1.22	4.28	0.50	0.89	0.20	1.81	0.23	1.01	0.13	3.61	0.37
Crown	Tu	8	16.79	10.92	4.75	1.57	8.41	3.26	7.60	8.67	0.96	0.32	1.90	0.37	1.34	0.63	4.73	3.09
Non-Crown	Y	3	8.29	1.13	4.35	1.22	3.87	0.26	3.87	0.29	0.97	0.11	1.91	0.19	0.90	0.02	3.55	0.63
F-statistic P-value		**216**	**P=0.0014*** **		**P<0.0001*** **		**P<0.0001*** **		**P=0.2436**		**P<0.0001*** **		**P<0.0001*** **		**P=0.0183** **		**P=0.366**	

Supplementary Table S9. Individual copy number (CN) of cryptorchid Quarter Horse males of 7 ampliconic genes and SRY. Green numbers denote the largest CN value and red letters denote the lowest CN value.

Sample ID	**Cryporchid Status**	**Phenotype**	**ETSTY1**	**ETSTY2**	**ETSTY5**	**HSFY**	**RBMY**	**SRY**	**TSPY**	**UBA1Y**
H023	NO	Normal Control	5.11	3.93	3.71	1	1.95	0.97	8.66	3.77
H042	NO	Normal Control	4.04	4.65	4.47	1.04	1.79	0.85	11.2	3.77
H043	NO	Normal Control	4.51	4.84	4.43	1.03	1.74	0.844	12.5	3.66
H044	NO	Normal Control	5.1	6	4.6	1.06	1.36	0.72	15.4	3.08
H045	NO	Normal Control	4.81	5.31	4.68	0.97	1.54	0.66	12.4	3.57
H046	NO	Normal Control	4.29	4.66	4.32	1	1.82	0.86	10.8	3.83
H047	NO	Normal Control	4.01	4.8	4.54	1.03	1.78	0.78	10.9	3.59
H048	NO	Normal Control	4.6	4.65	4	0.98	1.84	0.81	9.7	3.91
H051	NO	Normal Control	4.58	4.58	4.6	0.96	1.71	0.722	11	3.73
H052	NO	Normal Control	5.07	3.14	3.66	1.29	1.94	0.893	9.58	3.89
H061	NO	Normal Control	4.07	4.47	3.57	1.14	1.85	0.84	8	3.76
H070	NO	Normal Control	3.39	4	3.36	1.05	2.16	0.76	6.7	3.11
H292	NO	Normal Control	4.96	4.01	3.9	0.93	1.82	0.97	8.5	3.79
H293	NO	Normal Control	3.89	4.06	3.98	0.98	1.86	0.9	9.33	3.68
H294	NO	Normal Control	3.77	5.37	4.44	1.08	1.67	0.78	13.3	3.3
H295	NO	Normal Control	5.07	6.3	5.3	0.99	1.7	0.68	17.1	3.29
H296	NO	Normal Control	4.94	5.54	4.78	1.2	1.73	0.73	13.7	3.26
H297	NO	Normal Control	6.1	6.11	5.32	1.16	1.56	0.74	15.7	3.76
H300	NO	Normal Control	5.01	5.49	4.6	0.96	1.76	0.71	13.6	3.47
H337	NO	Normal Control	4.09	4.24	3.78	1.09	2.1	1.03	8.9	3.45
H343	NO	Normal Control	3.82	3.9	3.91	1.03	2.23	0.92	9.5	2.47
H346	NO	Normal Control	4.58	5.2	3.92	0.675	1.46	0.64	7.85	2.06
H400	NO	Normal Control	4.64	3.92	3.985	1.02	2	1.014	9.2	3.7

H401	NO	Normal Control	7.2	4.6	4.41	0.65	1.22	0.71	9.2	3.9
H410	NO	Normal Control	4.89	2.9	3.67	0.72	1.62	0.78	9.42	1.88
H446	NO	Normal Control	4.92	3.75	3.93	0.95	1.87	0.916	8.7	3.42
H593	NO	Normal Control	4.94	4.08	4.24	1.15	2.08	0.94	8.3	3.45
H907	NO	Normal Control	4.35	3.94	3.78	0.82	1.95	0.99	9.3	3.43
H217	YES	Unilateral Cryp.	10.5	4.91	3.51	0.58	1.3	0.84	13	4.88
H221	YES	Unilateral Cryp.	4.1	3.7	5.2	1.23	2.49	1.08	8.1	4.2
H222	YES	Unilateral Cryp.	4.9	4.21	6.8	1.43	2.48	1.52	9.3	4.04
H335	YES	Unilateral Cryp.	4.79	5.21	3.86	1.1	1.79	0.85	11.3	3.5
H373	YES	Unilateral Cryp.	4.8	4	3.22	1.17	1.72	0.76	8.9	3.8
H390	YES	Bilateral Cryp.	5	4.17	4.4	0.99	2.1	1	9.5	4.03
H441	YES	Bilateral Cryp.	4.78	4.26	3.82	1.04	2.11	0.9	10.4	3.85
H447	YES	Unilateral Cryp.	3.78	3.73	3.6	1.19	2.02	0.91	7.9	3.09
H451	YES	Bilateral Cryp.	3.49	3.88	3.52	0.97	1.96	0.93	8.2	3.43
H487	YES	Unilateral Cryp.	3.55	4.03	3.83	0.98	1.9	1.11	7.6	3.59
H513	YES	Unilateral Cryp.	4.19	3.96	3.66	1.17	2.12	0.94	7.73	3.57
H528	YES	Unilateral Cryp.	5.2	4.3	4.2	1.01	1.88	0.97	9.6	3.6
H571	YES	Unilateral Cryp.	4.21	3.83	3.63	1.05	1.96	0.93	7.38	3.6
H572	YES	Unilateral Cryp.	4.76	3.71	3.8	0.91	2.33	0.96	8.5	3.82
H573	YES	Unilateral Cryp.	2.99	4.09	5.5	1.06	2.06	0.98	8.1	4
H582	YES	Unilateral Cryp.	4.03	3.49	4.17	1.19	1.89	1.05	8.7	3.81
H596	YES	Unilateral Cryp.	3.86	4.42	3.85	1	1.86	0.79	9.2	3.46
H614	YES	Bilateral Cryp.	4.39	4.21	4.23	1.11	1.81	0.83	8.8	3.24
H615	YES	Unilateral Cryp.	7	5.17	4.47	1.02	1.78	0.9	10.6	3.29
H616	YES	Unilateral Cryp.	4.37	3.81	4.25	0.87	1.92	0.98	9	3.82
H638	YES	Bilateral Cryp.	3.3	4.51	4.44	0.99	1.73	0.87	8.78	3.49
H640	YES	Unilateral Cryp.	3.19	4.27	4.24	1.08	1.91	0.87	12.9	2.88
H643	YES	Unilateral Cryp.	4.12	4.18	3.9	1	1.64	0.83	8.16	3.6

H646	YES	Bilateral Cryp.	3.98	4.3	3.64	1.06	1.83	0.86	9	3.73

Supplementary table S10. Individual MSY gene CNs of horses with confirmed 64,XY disorders of sex development (XY DSDs). Numbers in green font denote the largest CN value and numbers in red font denote the lowest CN value. (-) means that CNs were unable to be generated due to missing content.

Breed	ID	Karyotype	*SRY* PCR	Sex phenotype	ETSTY1	ETSTY2	ETSTY5	HSFY	RBMY	SRY	TSPY	UBA1Y
Appaloosa	H949	64,XY	negative	female, SRY-negative sex reversal	4.39	3.26	4.6	1.18	1.04	(-)	8.1	3.8
Arabian	H089	64,XY	negative	female, SRY-negative sex reversal	4.18	3.95	3.72	1.12	0.96	(-)	8.9	4.9
Arabian	H266	64,XY	negative	female, SRY-negative sex reversal	4.79	3.96	3.96	0.99	0.94	(-)	9	3.38
American Quarter Horse	H184	64,XY	negative	female, SRY-negative sex reversal	4.72	3.83	3.57	1.03	0.96	(-)	9.3	3.2
American Quarter Horse	H909	64,XY	negative	female, SRY-negative sex reversal	4.32	3.75	4.2	1.21	1.06	(-)	8.53	3.73
Standardbred	H360	64,XY	negative	female, SRY-negative sex reversal	5.16	3.9	3.71	1.01	1	(-)	9.8	2.97
Tennessee Walking Horse	H947	64,XY	negative	female, SRY-negative sex reversal	4.14	3.63	4.7	0.98	0.99	(-)	8.53	3.45
Thoroughbred	H199	64,XY	negative	female, SRY-negative sex reversal	3.64	4.01	3.36	0.99	0.93	(-)	8.5	3.6
Thoroughbred	H227	64,XY	negative	female, SRY-negative sex reversal	5.32	4.24	3.96	1.12	0.94	(-)	9.3	1.42
Thoroughbred	H748	64,XY	negative	female, SRY-negative sex reversal	4.07	3.52	4.72	1.23	1.04	(-)	9.2	2.88

Thoroughbred	H844	64,XY	negative	female, SRY-negative sex reversal	4.28	4.05	4.34	1.18	1.03	(-)	8.17	3.49
Thoroughbred	H921	64,XY	negative	female, SRY-negative sex reversal	4.05	3.93	5.5	1.3	1.04	(-)	9.2	3.5
Appaloosa	H806	64,XY	positive	female-like, SRY-positive sex reversal	4.35	3.44	3.9	1.08	1.99	1.06	8.8	3.82
Appaloosa	H169	64,XY	positive	female-like, SRY-positive sex reversal	4.82	3.99	3.66	1.08	1.83	1.02	8.1	4.4
Mixed Warmblood	H252	64,XY	positive	female-like, SRY-positive sex reversal	3.42	3.95	3.96	1.07	1.89	0.98	8.9	4.01
American Quarter Horse	H038	64,XY	positive	female-like, SRY-positive sex reversal	5.33	4.01	3.91	1.11	1.88	0.84	8.1	3.88
Standardbred	H348	64,XY	positive	female-like, SRY-positive sex reversal	4.26	3.93	3.72	1.24	2.01	0.88	8.5	3.35
Standardbred	H369	64,XY	positive	female-like, SRY-positive sex reversal	4.4	4.1	4.35	0.98	1.91	1.14	8.6	4
Thoroughbred	H224	64,XYdel	negative	female-like, SRY-positive sex reversal	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
American Quarter Horse	H737	64,XY	positive	intersex	4.14	3.87	4.31	1.25	1.88	1.01	8.7	3.63
American Quarter Horse	H826	64,XY	positive	intersex	4.12	3.41	3.99	0.98	1.71	0.83	8.17	3.42
American Quarter Horse	H145	64,XY	positive	intersex	4.63	3.86	3.25	1.47	1.92	0.86	8.3	3.67
American Quarter Horse	H423	64,XY	positive	intersex	5.15	4.1	3.75	1.03	1.91	0.98	9.6	4.16
American Quarter Horse	H690	64,XYdel	negative	intersex	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
Standardbred	H543	64,XY	positive	intersex	4.79	3.95	3.86	0.94	1.88	1.02	8.5	2.98
Standardbred	H545	64,XY	positive	intersex	5.07	3.92	5.3	0.97	1.91	1.04	9	2.67

Tennessee Walking Horse	H544	64,XY	positive	intersex	4.86	4.02	3.8	0.9	2.01	0.94	9.5	3.12
Thoroughbred	H141	64,XYdel	negative	intersex	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
Thoroughbred	H546	64,XYdel	negative	intersex	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)

Supplementary Table S11. Individual copy number (CN) of MSY genes in subfertile males. Numbers in green font denote the largest CN value and numbers in red font denote the lowest CN value per gene

Breed	Sample ID	Reproductive phenotype (detailed)	Reproductive phenotype (broad)	Karyotype	ETSTY1	ETSTY2	ETSTY5	HSFY	RBMY	SRY	TSPY	UBA1Y
Arabian	H172	idiopathic	subfertile	64, XY	3.96	3.74	3.72	1.11	2.01	0.93	8.61	3.35
Arabian	H206	normal sperm count but 97% of sperm abnormal. Low motility, abnormal morphology	sterile	64, XY	5.03	4.05	3.93	0.682	2.07	0.966	8.8	2.58
Arabian	H255	idiopathic	subfertile	64, XY	2.49	3.82	3.96	0.48	1.87	0.953	9	4.01
Arabian	H284	idiopathic	subfertile	64, XY	5	4.16	3.96	0.71	1.87	0.94	7	1.42
Connemera	H320	azoospermia	sterile	64,XY	3.2	4.02	2.9	1.01	2.2	0.85	9	3.2
American Quarter horse	H343	low sperm count	subfertile	64, XY	3.82	3.9	3.91	1.03	2.23	0.92	9.5	3.76
Appaloosa	H620	azoospermia	sterile	64,XY	4.12	3.975	4.32	0.92	1.84	0.91	9.7	3.52
Arabian	H653	abnormal sperm motility and morphology	subfertile	64, XY	3.76	4.41	3.96	1.03	1.87	0.83	10	3.38
Arabian	H693	idiopathic	subfertile	64, XY	3.86	3.57	4.03	1.09	1.97	0.972	8.79	3.47
Standardbred	H726	idiopathic	subfertile	64,XY	3.48	3.65	4.35	0.8	2.05	0.7	12	2.85
Fresian	H787	azoospermia; chromosomal abnormality	sterile	64, XY; t(Y;13p)	1.96	3.78	2.62	0.85	1.91	0.78	5.55	3.13
Arabian	H962	azoospermia	sterile	64,XY	4.37	4.62	4.73	1.07	1.74	0.99	11.9	3.11
Arabian	H963	clinically idiopathic; chromosomal abnormality	subfertile	64,XY; t(12q;25q), der(12p)	4.27	4.87	5.86	0.99	1.7	0.86	14.2	3.1

American Quarter horse	H004	segmental aplasia, partially missing epididymis	subfertile	64,XY	3.15	4.02	3.81	1.02	1.97	0.92	7.91	3.47

Supplementary Table S12. All multicopy Y BACs and select contig I BACs used in this study. All BACs are known and published in Janecka et al. 2018.

BAC location	BAC ID	Sequenced 2018 study	Sequenced this study
Multicopy	017D15	Yes	Yes
Multicopy	020.1G12	Yes	Yes
Multicopy	022.4 E3	No	Yes
Multicopy	024.4G8	No	No
Multicopy	024I23	Yes	Yes
Multicopy	026B21	Yes	Yes
Multicopy	027.1A2	Yes	Yes
Multicopy	037.4D11	Yes	Yes
Multicopy	052H5	Yes	Yes
Multicopy	054J7	Yes	Yes
Multicopy	064P16	Yes	Yes
Multicopy	067.4G1	No	No
Multicopy	072G7	Yes	Yes
Multicopy	077M19	Yes	Yes
Multicopy	081F24	Yes	Yes
Multicopy	086J1	Yes	Yes
Multicopy	090P8	Yes	Yes
Multicopy	103.3A6	Yes	Yes
Multicopy	111.2F5	No	Yes
Multicopy	112C10	Yes	Yes
Contig I	126G2	Yes	Yes
Multicopy	129K23	Yes	Yes
Multicopy	134H14	Yes	Yes
Multicopy	134I16	Yes	Yes
Multicopy	139C20	Yes	Yes
Multicopy	140J20	Yes	Yes
Multicopy	140M23	Yes	Yes
Multicopy	142O2	Yes	Yes
Multicopy	147K8	Yes	Yes
Multicopy	152G20	Yes	Yes
Contig I	155M11	No	Yes
Multicopy	160K10	Yes	Yes
Multicopy	165 E24	Yes	Yes
Multicopy	181B18	Yes	Yes

Multicopy	207D10	No	Yes
Contig I	209K10	Yes	Yes
Multicopy	272B4	Yes	Yes
Multicopy	275P16	Yes	Yes
Multicopy	280P20	Yes	Yes
Multicopy	324H11	Yes	Yes
Multicopy	329G16	Yes	Yes
Multicopy	394K12	Yes	Yes
Multicopy	406I22	Yes	Yes
Multicopy	415H8	Yes	Yes
Multicopy	456J9	Yes	Yes
Multicopy	510F11	Yes	Yes
Multicopy	ABW	Yes	Yes
Multicopy	BBW	Yes	Yes
Multicopy	CBW	Yes	Yes
Multicopy	DBW	No	Yes
Multicopy	HBW	Yes	Yes
Multicopy	JBW	Yes	Yes

Supplementary Table 13. Details for sequence tagged sites (STSs) in CHORI-241 BAC library including base pair (bp) product size and GenBank Accession number.

BAC ID STS Marker	Forward primer 5'-3'	Reverse primer 5' to 3'	bp	GenBank Accession
CH241-121H7-SP6	TTGGGGTTCCCATAAAACTG	CCAAAAGACTATGCCCCGTA	245	CU028708.1
CH241-148N16-SP6	TGCAACAGCTCTGACCATCT	CCGTTTTGGAGAGGAATGTG	233	CU049035.1
CH241-56J9-T7	TCCAACTCAGAATACTGCAGCT	GTGTTTCTAAGTCAAACTGATGGT	503	CT974159.1
CH241-56J9-SP6	TAGGTACTTGCAGCACAGCC	TCTTAGATTCCTGGGAGGGCA	571	CT973767.1
CH241-142D23-SP6	AGCTGCAGAGAGACAATGTTGA	TCCGTCTCCTTTCTGAGCTTC	212	CU044388.1
CH241-142D23-T7	AGGATCTTCATAGGGAATTCTAATGCA	ACAGCCTGAAGAATCCAGTACT	409	CU044389.1
CH241-172K19-SP6	AGCCAGGTCTCTATTGCACAG	AGATAGCTGCCACATCTCAACT	500	CU080056.1
CH241-172K19-T7	CCATAAGCACAGTGTACATGGC	ACAGAGCCTTTTTCTGCATCC	204	CU080057.1
CH241-28P13-SP6	TCTCCCACAGAAGCAACAGA	TGAGGAGAAAGCTGGCTTAG	322	CT954182.1
CH241-28P13-T7	TCACTCCACCAATTTTCCCCA	CAGGTTTCTGCCGTCAAGGA	264	CT954183.1
CH241-79D11-T7	AGGGAATTCATAGCTCAAAGGCA	GCCATGCAATTCCTCCCTGA	213	CU008703.1
CH241-88F22-SP6	ATCGGGTAGAAACTAGTGT	GTGCCAAAAGTTCTGGA	n/a	CU014509.1
CH241-88F22-T7	TGCAAAGGCTGTACCCGT	ACCCAATTGTATCATGGATA	n/a	CU014509.1
CH241-106F15-SP6	TCAGCATGGTCATATACTATGTGT	AGCATGCACAGATCTGGTGT	512	CT008000.1
CH241-106F15-T7	TCTGCAATCTGGACTCAGCC	GTGCTTGCCTCAATCTCCCT	256	CT008000.1
CH241-127N19-SP6	TTGGGGTTCCCATAAAACTG	CCAAAAGACTATGCCCCGTA	358	CU032020.1
CH241-183I5-SP6	GTGTGTAGCTGTGGTGAGGG	TGGATCTCGTCTTGTTCTCAGG	203	CU086943.1
CH241-183I5-T7	TGGTGTACGTTGTGGTGATGA	CTCCTTCAACCCCTCCACAC	462	CU086713.1
CH241-31I9-T7	CAGGCAGAGGAGGAGAGAGA	TGTGGCAATGAACTCCCTCA	521	CT962322.1
CH241-31I9-SP6	ACCTTCTGTTCTTGGTTTGCT	GGGCAAATGAGCGAATTCCC	577	CT961960.1
CH241-43B24-T7	AATCACCCCTTTTGCCCTCC	GGGTAGCTTCTCAGGCAAACT	264	CT966686.1
CH241-43B24-SP6	TGGCCTATTAATTGCTCTAAGGA	TCTGCTACTGATTTTTACCTCTGA	405	CT966685.1
CH241-133E10-T7	ACGACTCACTATAGGGAGAGGA	TACCAGAGCCAGAGCAGAGT	468	CU035722.1

CH241-133E10-SP6	AGTCTCTGCAGAAGTCTTTGC	ACTCCCCTGGTAAGGCCTC	301	CU035959.1
CH241-144G20-T7	CATCCCAGGACAGCGTATCC	TCCCCACAGAACCACAATGG	562	CU045921.1
CH241-144G20-SP6	ACACTAGCCTAAGCAGGTTCT	AGGTTTACAATGCTTTGAGAAAGGA	235	CU045920.1

Supplementary Table S14. Information regarding the contigs not present in the published eMSYv3. Information includes the contig size in base pairs (bp), the Y chromosome BACs which correspond to the contig location, and the Y chromosome location in eMSYv3. Red font denotes full repeat sequences identified through RepeatMasker, or contigs not able to be PCR screened.

Contig size (bp)	Contig ID	Tiling Path BAC amplified	eMSYv3 location
9293	LipY764_contig100	56J9; 148N16; 172K19	Gap 1
19018	LipY764_contig101	148N16; 56J9; 142D23; 172K19; 28P13	Gap 1
27363	LipY764_contig111	34A23	III near Gap 3
13602	LipY764_contig118	28P13; 142D23; 72K19	Gap 1
24689	LipY764_contig124	22E24; 121H7	Gap 1
1283	LipY764_contig127	121H7; 127N19; 183I5	Gap 1
7132	LipY764_contig129	79D11; 88F22; 106F15	Gap 1
9958	LipY764_contig130	79D11; 88F22; 106F15	Gap 1
5747	LipY764_contig131	79D11; 88F22; 106F15	Gap 1
23831	LipY764_contig133	179K8; 121G24	III near Gap 2
15472	LipY764_contig141	56J9; 148N16; 172K19	Gap 1
23572	LipY764_contig142	209K10; 155M11; 341G20	I Distal
2028	LipY764_contig145	79D11	Gap 1
13608	LipY764_contig146	79D11; 28P13	Gap 1
5637	LipY764_contig147	28P13; 79D11	Gap 1
21192	LipY764_contig165	22E24; 121H7	Gap 1
20773	LipY764_contig170	11B8	II
20122	LipY764_contig174	79D11	Gap 1
19851	LipY764_contig176	121H7	Gap 1
19813	LipY764_contig177	121H7; 127N19; 183I5	Gap 1
6004	LipY764_contig188	11B8	II

18731	LipY764_contig189	179K8; 121G24	III near Gap 2
18616	LipY764_contig192	28P13; 79D11	Gap 1
16389	LipY764_contig199	121H7; 127N19; 183I5	Gap 1
REPEAT	LipY764_contig200	n/a	n/a
17045	LipY764_contig202	121H7	Gap 1
16614	LipY764_contig206	132K10; 66M24	II near Gap 2
6518	LipY764_contig207	22E24; 121H7	Gap 1
9358	LipY764_contig208	22E24; 121H7	Gap 1
16254	LipY764_contig209	34A23	III near Gap 3
16021	LipY764_contig213	22E24; 121H7	Gap 1
24212	LipY764_contig22	209K10; 155M11; 341G20; 16.4C5; 15.2E9; 155B8	I distal
15478	LipY764_contig220	179K8; 121G24	III near Gap 2
15129	LipY764_contig226	34A23	III near Gap 3
14860	LipY764_contig229	11B8	II
REPEAT	LipY764_contig23	n/a	n/a
14724	LipY764_contig234	121H7; 127N19; 183I5	Gap 1
13701	LipY764_contig249	88F22; 106F15	Gap 1
13535	LipY764_contig250	56J9; 148N16	Gap 1
13457	LipY764_contig251	Could not get amplification; contig not screened	n/a
13225	LipY764_contig255	88F22; 106F15	Gap 1
10964	LipY764_contig256	34A23	III near Gap 3
1572	LipY764_contig257	Could not get amplification; contig not screened	n/a
12858	LipY764_contig258	107.3H9; 79.4H1	I proximal

12858	LipY764_contig264	3A5; 72G23; 114E24; 8.3G9; 215C6	II
2932	LipY764_contig274	28P13; 79D11; 142D23	Gap 1
8025	LipY764_contig275	28P13; 79D11; 142D23	Gap 1
11678	LipY764_contig276	13E2; 417N24; 437I11; 278M12; 95.4B8; 101H8	I distal
10973	LipY764_contig292	28P13; 142D23; 172K19	Gap 1
10686	LipY764_contig296	121H7; 127N19; 183I5	Gap 1
7798	LipY764_contig298	121H7	Gap 1
2006	LipY764_contig299	22E24; 121H7	Gap 1
10414	LipY764_contig301	22E24; 121H7	Gap 1
2690	LipY764_contig303	121H7	Gap 1
6927	LipY764_contig304	121H7	Gap 1
10317	LipY764_contig308	121H7; 127N19; 183I5	Gap 1
9968	LipY764_contig313	22 E24; 121H7	Gap 1
1341	LipY764_contig317	88F22; 106F15	Gap 1
8010	LipY764_contig318	88F22; 106F15	Gap 1
9715	LipY764_contig322	148N16; 56J9	Gap 1
9537	LipY764_contig326	179K8; 121G24	III near Gap 2
9392	LipY764_contig328	148N16	Gap 1
9745	LipY764_contig329	22 E24; 121H7	Gap 1
7829	LipY764_contig330	22 E24; 121H7	Gap 1
9030	LipY764_contig333	88F22; 106F15	Gap 1
8768	LipY764_contig335	209K10; 155M11	I distal
REPEAT	LipY764_contig340	n/a	n/a
5834	LipY764_contig342	22E24; 121H7	Gap 1
REPEAT	LipY764_contig343	n/a	n/a

8292	LipY764_contig346	13E2; 148G3; 417N24; FBW; 278M12	I distal
8107	LipY764_contig349	126G2	I distal
8046	LipY764_contig350	121H7; 127N19	Gap 1
2692	LipY764_contig354	28P13; 79D11; 142D23; 172K19	Gap 1
4598	LipY764_contig355	28P13; 79D11; 142D23	Gap 1
7891	LipY764_contig356	121H7; 127N19	Gap 1
7872	LipY764_contig358	56J9; 148N16; 172K19	Gap 1
7752	LipY764_contig361	28P13; 142D23; 172K19	Gap 1
7725	LipY764_contig362	79D11	Gap 1
7674	LipY764_contig366	22E24; 121H7	Gap 1
7628	LipY764_contig368	28P13; 142D23; 172K19	Gap 1
7613	LipY764_contig369	11B8	I distal
7533	LipY764_contig374	28P13; 79D11	Gap 1
6965	LipY764_contig388	121H7	Gap 1
1138	LipY764_contig393	28P13; 79D11; 142D23	Gap 1
5005	LipY764_contig394	28P13; 79D11; 142D23	Gap 1
6659	LipY764_contig398	127N19; 183I5	Gap 1
6600	LipY764_contig399	179K8; 121G24	III near Gap 2
28846	LipY764_contig40	79D11	Gap 1
	LipY764_contig401	209K10	I distal
	LipY764_contig402	209K10	I distal
12161	LipY764_contig41	79D11	Gap 1
6255	LipY764_contig413	Could not get amplification; contig not screened	n/a
6171	LipY764_contig416	179K8; 121G24	III near Gap 2
41683	LipY764_contig42	121H7; 127N19; 183I5	Gap 1
2906	LipY764_contig426	148N16	Gap 1

2045	LipY764_contig427	148N16	Gap 1
REPEAT	LipY764_contig428	n/a	n/a
3106	LipY764_contig429	126G2; 112C10; 129K23; 132K10; 134H14; 168i4	I distal; I multicopy
5775	LipY764_contig430	34A23	III near Gap 3
5625	LipY764_contig435	88F22; 106F15	Gap 1
5523	LipY764_contig437	179K8; 121G24	III near Gap 2
5112	LipY764_contig450	126G2; 112C10; 129K23; 132K10; 134H14; 168i4	I distal; I multicopy
4970	LipY764_contig455	126G2	I distal
4696	LipY764_contig467	79D11	Gap 1
4510	LipY764_contig477	126G2	I distal
1381	LipY764_contig483	127N19; 183I5	Gap 1
4345	LipY764_contig490	11B8	I distal
REPEAT	LipY764_contig494	n/a	n/a
4186	LipY764_contig496	34A23	III near Gap 3
39418	LipY764_contig50	28P13; 172K19; 56J9; 22E24; 28P13; 31I9; 43B24; 79D11; 88F22; 106F15; 116I17; 133E10; 148N16	Gap 1
1661	LipY764_contig501	28P13; 79D11	Gap 1
1881	LipY764_contig502	28P13; 79D11	Gap 1
4113	LipY764_contig503	148N16; 56J9; 179K19	Gap 1
3982	LipY764_contig508	13 E2; 19D21; 417N24; 437I11; 278M12; 95.4B8	I distal
3780	LipY764_contig513	34A23	III near Gap 3
REPEAT	LipY764_contig514	n/a	n/a
3620	LipY764_contig519	121H7; 127N19	Gap 1

27777	LipY764_contig52	56J9; 148N16; 172K19	Gap 1
REPEAT	LipY764_contig525	n/a	n/a
10448	LipY764_contig53	56J9; 148N16; 172K19; 79D11	Gap 1
REPEAT	LipY764_contig534	n/a	n/a
3328	LipY764_contig539	11B8	II
3301	LipY764_contig541	22 E24; 121H7	Gap 1
3103	LipY764_contig547	88F22; 106F15	Gap 1
2789	LipY764_contig564	REPEATS-bad primers	n/a
2773	LipY764_contig565	132N15	I distal
	LipY764_contig568	3A5; 72G23	II
3216	LipY764_contig569	148N16	Gap 1
2614	LipY764_contig575	126G2	I distal
REPEAT	LipY764_contig576	n/a	n/a
2553	LipY764_contig579	179K8; 121G24	III near Gap 2
2332	LipY764_contig586	34A23	III near Gap 3
2270	LipY764_contig591	148N16; 56J9	Gap 1
2144	LipY764_contig596	179K8; 121G24	III near Gap 2
18371	LipY764_contig60	34A23	III near Gap 3
REPEAT	LipY764_contig600	n/a	n/a
REPEAT	LipY764_contig602	n/a	n/a
1818	LipY764_contig608	148N16	Gap 1
11032	LipY764_contig61	34A23	III near Gap 3
1661	LipY764_contig618	28P13	Gap 1
REPEAT	LipY764_contig619	n/a	n/a
1554	LipY764_contig625	13 E2; 19D21; 417N24; 437I11; 95.4B8; 278M2	I distal
1546	LipY764_contig626	121H7	Gap 1

770	LipY764_contig63	22E24; 121H7	Gap 1
REPEAT	LipY764_contig630	n/a	n/a
REPEAT	LipY764_contig631	n/a	n/a
REPEAT	LipY764_contig633	n/a	n/a
1451	LipY764_contig634	148N16; 56J9	Gap 1
REPEAT	LipY764_contig640	n/a	n/a
REPEAT	LipY764_contig641	n/a	n/a
REPEAT	LipY764_contig645	n/a	n/a
REPEAT	LipY764_contig647	n/a	n/a
	LipY764_contig655	13 E2; 101H8; 417N24; 278M12; 95.4B8	I distal
1248	LipY764_contig657	3A5; 72G3; 114E24; 8.3G9	II
REPEAT	LipY764_contig668	n/a	n/a
36205	LipY764_contig67	56J9; 148N16; 172K19	Gap 1
REPEAT	LipY764_contig672	n/a	n/a
REPEAT	LipY764_contig673	n/a	n/a
REPEAT	LipY764_contig674	n/a	n/a
874	LipY764_contig691	148N16; 56J9	Gap 1
REPEAT	LipY764_contig693	n/a	n/a
REPEAT	LipY764_contig698	n/a	n/a
5351	LipY764_contig7	418J18; 155M11	I distal; I proximal
771	LipY764_contig703	REPEATS-bad primers	n/a
REPEAT	LipY764_contig705	n/a	n/a
	LipY764_contig706	126G2	I distal
REPEAT	LipY764_contig709	n/a	n/a
REPEAT	LipY764_contig725	n/a	n/a
REPEAT	LipY764_contig727	n/a	n/a

REPEAT	LipY764_contig737	n/a	n/a
REPEAT	LipY764_contig738	n/a	n/a
REPEAT	LipY764_contig747	n/a	n/a
REPEAT	LipY764_contig750	n/a	n/a
REPEAT	LipY764_contig763	n/a	n/a
31984	LipY764_contig79	209K10; 155M11	I distal
31580	LipY764_contig80	121H7	Gap 1
13110	LipY764_contig87	126G2	I distal
30832	LipY764_contig90	121H7; 127N19	Gap 1
30145	LipY764_contig95	121H7; 127N19; 28P13?	Gap 1