Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Veterinarian Medicine and Animal Science
Department of Animal Breeding and Genetics

Exploration of the hereditary cause of sex ratio distortion in horses

Susan E. Ott

Master Thesis Biology • 30 ECTS
Swedish University of Agricultural Sciences (SLU)
Animal Breeding and Genetics, SLU
MSc Biology - Cellular and Molecular biology, Wageningen University
Uppsala – 2021

Exploration of the hereditary cause of sex ratio distortion in horses

Susan E. Ott

Supervisor:	**Gabriella Lindgren, Swedish University of Agricultural Sciences (SLU), Animal Breeding and Genetics**
Assistant supervisor:	Ole Madsen, Wageningen University (WUR), Animal Breeding and Genomics
Assistant supervisor:	Terje Raudsepp, Texas A&M University, Veterinary Integrative Bioscience
Examiner:	Susanne Eriksson, SLU, Animal Breeding and Genetics

Credits:	30
Level:	Advanced, A2E
Course title:	Independent Project in Biology
Course code:	EX0871
Programme/education:	MSc. Biology, WUR
Course coordinating dept:	Department of Animal Environment and Health
Place of publication:	Uppsala, Sweden
Year of publication:	2021
Cover picture:	Susan E. Ott

Keywords: Sex ratio, horse, selfish genetic elements, meiotic drive, Icelandic horse, Standardbred horse, Coldblooded trotters, heritability, Trivers-Willard hypothesis, pedigree data

Swedish University of Agricultural Sciences
Faculty of Veterinarian Medicine and Animal Science
Department Animal Breeding and Genetics

Publishing and archiving

Approved students' theses at SLU are published electronically. As a student, you have the copyright to your own work and need to approve the electronic publishing. If you check the box for **YES**, the full text (pdf file) and metadata will be visible and searchable online. If you check the box for **NO**, only the metadata and the abstract will be visible and searchable online. Nevertheless, when the document is uploaded it will still be archived as a digital file.

If you are more than one author you all need to agree on a decision. Read about SLU's publishing agreement here: https://www.slu.se/en/subweb/library/publish-and-analyse/register-and-publish/agreement-for-publishing/.

☒ YES, I/we hereby give permission to publish the present thesis in accordance with the SLU agreement regarding the transfer of the right to publish a work.

☐ NO, I/we do not give permission to publish the present work. The work will still be archived and its metadata and abstract will be visible and searchable.

Abstract

Even sex ratios are seen as evolutionarily stable and are maintained by a selection against a skewed sex ratio. However, sometimes sex ratios at birth are skewed towards one sex. Reasons for this phenomenon are often unknown. The aim of this project was to explore distorted sex ratios in Icelandic horses, Standardbred trotters, and Coldblooded trotters, to estimate the heritability of distorted sex ratios in Icelandic horses and explore families with distorted sex ratios, and set the outline for future research.

To investigate the aims, datasets were extracted from the Icelandic studbook, WorldFengur, and the online studbook information for Standardbreds and Coldblooded trotters from the Swedish Trotting association. The datasets included 509 008 Icelandic horses (born between 1860-2020), 504 Standardbred stallions, and 125 Coldblooded trotter stallions (both born between 1990-2020). In the breeds, individuals with skewing sex ratios among their offspring (skewing horses) were identified using chi-square tests. For Icelandic horses the heritability was estimated with a parent-offspring regression, using the proportion of male offspring of fathers and sons. Families were identified using the hundred most skewing Icelandic stallions and as preparation for future research a new protocol for real-time PCR (qPCR) analysis was developed and tested, to identify sex ratios in horse semen.

The proportion of males registered in the complete Icelandic horse dataset was 43.7%, whereas it was 49.0% for the last 10-year period studied. For Standardbreds and Coldblooded trotter stallions, the proportion of male offspring was 48.9% and 50.5%, respectively. Multiple skewing horses were identified in all breeds, 16.0% in Icelandic stallions, 5.8% in Icelandic mares, 11.5% in Standardbred stallions, and 6.4% in Coldblooded trotter stallions. In Icelandic horses and Standardbreds, female skewing was more common than male skewing. The heritability of the proportion of male offspring in Icelandic horses was estimated at 0.29 (SE: 0.08), but the sensitivity of this estimate to the number of offspring of father and sons required was high. Five female-skewing families were identified containing between 31 and 513 significantly skewing horses. Lastly, qPCR experimentation was performed using four primers. The Y-primer did not result in an amplification product, however the X and two autosomal controls were successful and usable on purified (without somatic cells) and raw horse semen.

The difference between sex ratio and proportion of female-skewing horses in Icelandic horses and Standardbreds might be explained by incomplete recordings especially in the past, and the existence of meat horses in the Icelandic horse pedigree. These meat horses cause incomplete registration of offspring, most often males. However, the extent of this bias to the analyses is still unknown. In further analyses, meat horses need to be excluded and some more data editing performed. The heritability estimates may be improved by using a better fitting model. In future, it would be interesting to investigate sex ratios in horse semen, to compare female-skewing and male-skewing stallions to a control, to investigate where in development skewing of sex ratios start. Lastly, whole genome sequencing of skewing individuals can give us more information on the genetic cause, with the aim to identify genetic elements causing distorted sex ratios.

Keywords: Sex ratios, horse, selfish genetic elements, meiotic drive, sex chromosome drive, Icelandic horse, Standardbred horse, Coldblooded trotters, heritability, Trivers-Willard hypothesis, pedigree data

Table of contents

List of tables ..... **v**

List of figures ..... **vi**

1. Introduction **1**

1.1. Selfish genetic elements 1

1.1.1. T-haplotype in mice 2

1.2. Sex chromosome drive 4

1.2.1. Sex reversal – Wood lemming 4

1.2.2. Male-drive 4

1.2.3. Arms race 5

1.3. Sex ratios in equines 10

1.4. Sex chromosomal abnormalities 11

1.5. Rationale and Aims 12

2. Materials and Methods **13**

2.1. Datasets 13

2.1.1. Icelandic horse dataset 13

2.1.2. Standardbred horse and Coldblooded trotter dataset 13

2.2. Statistical analyses 14

2.3. Heritability 15

2.4. Families with distorted sex ratios 16

2.5. A pilot experiment to investigate sex ratios in semen 19

3. Results **21**

3.1. Data characteristics 21

3.2. Heritability 26

3.3. Families with distorted sex ratios 27

3.4. qPCR pilot experimentation 32

4. Discussion **34**

4.1. Datasets 34

4.2. Heritability 38

4.3. Families 40

4.4. qPCR pilot experimentation 42

5. Conclusions and Future implications **44**

Acknowledgment **45**

6. References **46**

7. Appendix 53

7.1. Extended data characteristics 53
7.2. R-script to create the main dataset of Icelandic horses 56
7.3. Statistical analyses, influence of number of offspring 57
7.4. Pedigree identifier 58
7.5. Heritability Analyses 60
7.5.1. Matching sons to fathers and maternal grandfathers 60
7.5.2. Heritability calculations and extra figures 61
7.6. Families with distorted sex ratios 64
7.6.1. Identification of families with distorted sex ratios 64
7.6.2. Small fraction of Family 2 68
7.7. qPCR protocol for semen 69
7.7.1. Semen purification 69
7.7.2. QIASymphony DNA extraction 69
7.7.3. Simplistic DNA extraction 69
7.7.4. Plates - qPCR 70
7.7.5. Taqman qPCR settings 71
7.7.6. Fold gene expression calculation 71
7.7.7. qPCR results 72
7.8. Scientific Poster – Master thesis day 76

List of tables

Table 1. An extended list of genes which have coamplications on both sex chromosomes. 8

Table 2. Equine primers used for qPCR reaction. 20

Table 3. Descriptive values Icelandic dataset. 22

Table 4. Chi-square characteristics in Icelandic horses. 24

Table 5. Descriptive values Swedish trotter dataset. 25

Table 6. Chi-square characteristics in Coldblooded trotters and Standardbreds. 25

Table 7. Increasing heritability by increasing number of offspring 26

Table 8. Descriptive statistics per identified family 29

Table 9. Overlap between the families using different thresholds. 30

Table 10. Proportion of significant progeny. 31

Table 11a. Country of birth and sex ratios of Icelandic horses. 53

Table 11b. Country of birth and sex ratios of Standardbreds and Coldblooded trotters. 54

Table 12. Number of Icelandic horses registered per decade 55

Table 13. How significance and power is influenced by number of offspring. 57

Table 14. Plate format used during qPCR experiment. 70

List of figures

Figure 1. The t-haplotype in mice. *3*

Figure 2. The wood lemming (Myopus schisticolor). *4*

Figure 3. The mosquito Aedes aegypti. *4*

Figure 4. A visualisation of the arms race between sex chromosomes. *6*

Figure 5. Elaboration on the identification of families using a threshold system. *18*

Figure 6. Distribution of the birth registration of Icelandic horses over time. *21*

Figure 7. Distribution of offspring in stallions and mares. *22*

Figure 8. Distribution of sex ratios in significantly skewing Icelandic horses. *23*

Figure 9. Distribution of sex ratios in significantly skewing Standardbred horses. *26*

Figure 10. Father-offspring regression to estimate heritability of sex ratios. *27*

Figure 11. The primer melting curves of the four tested genes and the plate setup. *32*

Figure 12. Relative gene expression comparing purified and raw semen. *33*

Figure 13. Distribution of birth of year in Coldblooded trotters and Standarbreds. *55*

Figure 14. Heritability estimate of father-son and grandfater-grandson regressions. *63*

Figure 15. Heritability estimates of father-son regression, displaying variable nr. offspring *63*

Figure 16. A small part of the family tree of family 2. ..... *68*

Figure 17. qPCR reaction conditions. *71*

Figure 18. Melting curve of the AR primer. *72*

Figure 19. Melting curve of the DDX3Y primer. *73*

Figure 20. Melting curve of the GADPH primer. *74*

Figure 21. Melting curve of the Actin-β primer. *75*

1. Introduction

Even sex ratios are seen as evolutionarily stable and are maintained by a selection against a skewed sex ratio. If one sex becomes more frequent, it will create a reproductive advantage for the other sex. This frequency-dependent selection favours equal proportion of males and females. However, sometimes sex ratios at birth are skewed towards one sex. Some skewing can be explained by, for example, that one sex is more costly to produce, which favours the ratio towards the cheaper sex, or the mortality for one sex is higher (Gellatly, 2019; Uller et al., 2007). However, the causes of skewing are often unknown, which is the reason behind this project. The interest for this project started after horse breeders noticed a distorted sex ratio in offspring of one trotter stallion. Preliminary studies in Standardbred horses indicated that, along with this stallion, there may be more stallions with distorted sex ratios among their offspring, skewing both towards more male or female offspring than would be expected from equal sex ratios. The aim of this project is to explore such distorted sex ratios more into detail.

1.1. Selfish genetic elements

One of the mechanisms which could explain distorted sex ratios in animals are selfish genetic elements. Selfish genetic elements are known to exist in numerous species, including fungi, plants, insects and mammals (Montchamp-Moreau et al., 2006). They distort Mendelian segregation in favour of its own transmission to the gametes (Courret et al., 2019; Helleu et al., 2015; Lindholm et al., 2016). This causes the majority of the offspring to inherit this element, sometimes even up to 90% (Bauer et al., 2005). In some cases, selfish genetic elements reduce the fitness of a whole population and can affect fertility of the carriers (Helleu et al., 2015; Zanders & Unckless, 2019).

Selfish genetic elements operate differently in females and males. In females, the elements use gonotaxis, which means that there is a preferential transmission to the ovule (Courret et al., 2019). Females have an asymmetric meiosis in which only one of the four meiosis products becomes a gamete (Courret et al., 2019; Lindholm et al., 2016). In meiosis, centromeres direct the segregation, which makes them the best target for female selfish genetic elements (Akera et al., 2019). There are true centromere drivers and mutations outside the native centromeres (Courret et al., 2019). Female selfish genetic elements may expand the centromeric satellite, or move faster to the spindle pole than their homologue, which both cause a preferential segregation into the gamete (Akera et al., 2019; Courret et al., 2019).

In males, selfish genetic elements obstruct or kill the opposite homologue during spermatogenesis. Male selfish genetic elements include at least two alleles, a driver and a target or a driver and an antidote. In the first scenario, the driver prevents the formation of functional gametes or destroys target-bearing sperm. To prevent self-destruction, the driver is closely linked to an insensitive target allele (Courret et al., 2019). In case of a driver and an antidote, the driver poisons the surrounding of the gametes, which kills the gametes without a linked antidote allele (Bravo Núñez et al., 2018). The linkage between the two alleles is critical for the success of the selfish genetic element. The importance of linkage cause them to accumulate mainly in low- or non-recombining regions (Cocquet et al., 2012; Courret et al., 2019). One prominent and well-studied example of an autosomal selfish genetic element is the t-haplotype in mice.

Heteromorphic sex chromosomes have the largest non-recombining regions, which increases the chance of selfish genetic elements to arise there (Cocquet et al., 2012). When the selfish genetic elements are on the sex chromosomes (sex chromosome drive), it typically leads to skewing sex ratios in offspring (Helleu et al., 2015). Selfish genetic elements have previously not been investigated in horses, but we hypothesize that the observed distorted sex ratios may be induced by such elements.

1.1.1. T-haplotype in mice

The t-haplotype is an example of a well-studied selfish genetic element. It was discovered in 1927 and is visible because the t-haplotype carries an allele causing taillessness in mice (Bravo Núñez et al., 2018; Lyon, 1984; Schimenti, 2000). It is a 40 Mb variant of the proximal part of chromosome 17 of the *Mus musculus* and is present in multiple subspecies (Kelemen & Vicoso, 2018). Resembling most selfish genetic elements, the *t*-haplotype has a low recombination-rate with the wildtype chromosome 17. The low recombination-rate of the *t*-haplotype is caused by at least four inversions in the haplotype (Kruger & Mueller, 2021). For females, the t-haplotype has a normal inheritance pattern, however, for males the driver locus appears in 99% of the functional sperm and has a transmission rate of over 90% (Bravo Núñez et al., 2018; Kelemen & Vicoso, 2018; Leitschuh et al., 2018).

This selfish genetic element uses a driver-antidote system, where the haplotype contains several alleles which encode for factors disrupting sperm motility and has an allele rescuing the motility in the *t*-haplotype (Zanders & Unckless, 2019). In non-driver gonads, the sperm motility kinase 1 (SMOK1) is overactivated causing impaired sperm motility (Kelemen & Vicoso, 2018; Kruger & Mueller, 2021). The overactivation in the wild-type chromosome is probably caused by at least three poisonous driver genes in the t-haplotype, identified as *Tagap1*, *Fgd2*, and *Nme3*. Both *Tagap1* and *Fgd2* have a distinctive pathway in which it is hypothesized that *Tagap1* inhibits SMOK1 while *Fgd2* over activates SMOK1 (Bravo Núñez et al., 2018; Kruger & Mueller, 2021). *Nme3* is an allele which phosphorylates GDP to GTP, however it is not clear how it interacts with the SMOK1 pathway. The antidote gene on the haplotype, which rescues the flagellar function, is a dominant negative version of *Smok1* (Bravo Núñez et al., 2018). This rescue is, however, is not complete, causing motility defects and decreasing fertility in carrier males (Zanders & Unckless, 2019).

Because of the low recombination rate, the t-haplotype accumulates recessive mutations (Braidotti & Barlow, 1997), which makes homozygous t-haplotypes often lethal or semi-lethal (Leitschuh et al., 2018; Zanders & Unckless, 2019). Animals that do survive a homozygous genotype are normal-tailed and males are sterile (Lyon, 1984). Sterility is probably caused by the combined partial motility defects of sperm (Zanders & Unckless, 2019). The effects of high lethality of homozygous animals and decreasing fertility are the reasons why despite the strong driver capacity, the frequency of the t-haplotype in wild mice is only 10-25% (Figure 1) (Braidotti & Barlow, 1997; Kelemen & Vicoso, 2018).

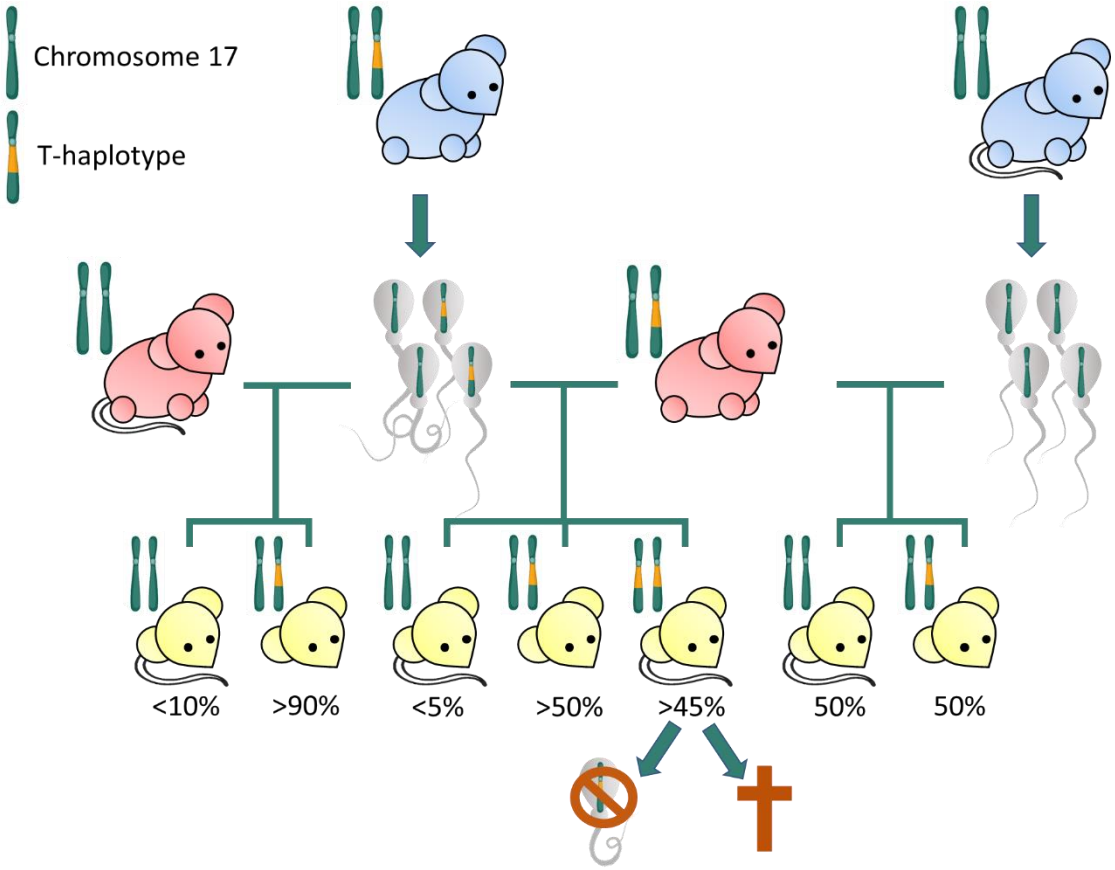


Figure 1. The t-haplotype is located on chromosome 17. If a male carries this haplotype, the motility of the non-driver gamete will be disrupted. For this reason over 90% of the offspring will inherit the t-haplotype. Homozygous genotypes are often lethal or semi-lethal, however, if males survive, they will be sterile. Female carriers have a normal Mendelian inheritance pattern.

1.2. Sex chromosome drive

There are several reports about sex chromosome driven sex bias of which the majority have been described in Drosophila species (Courret et al., 2019; Helleu et al., 2015). The first sex biased offspring was observed in 1925 in *Drosophila affinis* (Helleu et al., 2015). Three years later, this research was followed by the discovery of the first X-linked selfish genetic element in *D. obscura*, causing female-biased offspring (Gershenson, 1928). Sex-chromosome drivers are inherited similarly to autosomal genetic elements, and most have a driver and a target/antidote system. The following two sections will present some extraordinary cases of sex chromosome drivers in more detail.

1.2.1. Sex reversal – Wood lemming

Figure 2. The wood lemming (Myopus schisticolor), by Nutmeg's Wildlife Photography.

The wood lemming (*Myopus schisticolor*) is an herbivore living in pine forests from southern Norway to eastern Siberia (Figure 2) (Vuorinen & Eskelinen, 2005). Special for this species is that it is known to have three to four times more females than males, both in captivity and in the wild (Fredga et al., 1977). The disrupted sex ratio is caused by a mutated X-chromosome, which is refered to as X*. The X* silences the Y-chromosome and causes sex reversal in X*Y individuals (Vuorinen & Eskelinen, 2005). As a result, the wood lemming has three female genotypes: XX, XX* and X*Y, where XX* carriers produce only 25% sons and X*Y only daughters (Fredga et al., 1977). Both fitness and fertility are similar within all genotypes (Fredga et al., 1977; Vuorinen & Eskelinen, 2005). Since fertility is the same, the sex ratio in X*Y females is reached by a driver system where the X* wins over Y. So far, it is not known what causes the sex reversal in X*Y females, however there are abnormalities of the X* chromosome, including a deletion in Xp21-p23 which causes a rearrangement in the *Ccth* gene. The murine homologue of this gene is mainly expressed in testis, which may explain sex reversal of X*Y individuals (Liu et al., 2001).

1.2.2. Male-drive

Figure 3. The mosquito Aedes aegypti, by James Gathany.

X-linked sex chromosome drive is much more common than Y-linked drive (Helleu et al., 2015). This has multiple reasons. Firstly, because the Y-linked driver element is hemizygous and expressed in every generation. Secondly, Y-linked drivers are spreading faster through the populations, which may lead to a higher risk of extinction, because of the lack of females. Thirdly, there is less chance for a driver to evolve, since the Y chromosome usually has fewer genes. There are only a few cases of Y-linked drivers, two mechanisms will be discussed below; one in mosquitos and another in wasps (Helleu et al., 2015).

The first mechanism has been described in both in the *Aedes aegypti* and the *Culex pipiens quinquefasciatus*, two mosquito species, in which a killer-target system is used (Figure 3). In both species, a sex-determining locus on chromosome 1 determines the sex. A heterozygous genotype of the sex-determining locus becomes a male and a homozygous genotype a female. In the male driver system, the driver locus is closely linked to the sex-determining site. The driver targets a locus in the female homologue causing chromosome breakage during male meiosis. In this way the majority of the offspring becomes male. The target sensitivity of the driver gene varies considerately between haplotypes, which causes a wide range of sex distortion ratios between strains (Mori et al., 2004).

Similarly to mosquitos, the drive in *Nasonia vitripennis*, the jewel wasp, works with a killer-target system. However, differently from mosquitos, the jewel wasp is a hymenopteran insect (like ants and bees) where males are haploid and develop from unfertilized eggs, while fertilized eggs become diploid females. The driver can be seen as a whole chromosome, which eliminates all essential hereditary material of the sperm during fertilisation, excluding the chromosome itself. Thus, converting all diploid female-intended eggs into haploid males. One of the genes identified in the driver chromosome is *haploidizer*. The molecular mechanism, as yet, is unknown, although it is thought that *haploidizer* disrupts paternal chromatin by altering enzymes and replacing sperm chromatin protamines with histones (Benetta et al., 2020). During normal meiosis, sperm chromatin histones are removed and temporary repackaged with protamines – chromatin-binding proteins that condense and organize paternal DNA in sperm heads. After fertilisation, chromatin is remodelled again and protamines are replaced by histones (Kanippayoor et al., 2013). In jewel wasps carrying the driver, protamines are successful removed from the DNA, however the driver disrupts histone repackaging, probably causing elimination of the chromosomes (Aldrich et al., 2017).

1.2.3. Arms race

Recent publications present theories of a whole different paternal driving mechanism, known as the arms race. This mechanism has been proposed for multiple species, including drosophila, mice, cats, cattle and possibly even humans (Brashear et al., 2018; Cocquet et al., 2012; Hughes et al., 2020), where the selfish genetic elements are located on the Male Specific Region of the Y chromosome (MSY). The MSY is the part of the Y-chromosome that does not participate in crossing over because of a lacking recombination partner during male meiosis and harbours mostly male benefit genes (Brashear et al., 2018; Hughes et al., 2020; Janečka et al., 2018)). The MSY also differs drastically between species, both in gene content and in copy number variants (Brashear et al., 2018; Janečka et al., 2018) . For example, ampliconic sequences make up 98% of the mouse MSY, 96% of bull MSY, but only 54% of horse MSY (Hughes et al., 2020; Janečka et al., 2018). Further, between closely related species the proportion of ampliconic sequences can differ greatly, for example 45% in human, 57% in chimpanzees, and 5% in rhesus macaque (Hughes et al., 2020). Some of the amplicons in the MSY seem to be co-amplified in the X chromosome (Brashear et al., 2018; Hughes et al., 2020; Soh et al., 2014), and in mice this is seen for multiple gene families (Cocquet et al., 2012) .

The expansion of these multicopy gene families may have been a result of a dosage specific sex-chromosome driver and suppressor, where the proportion of male or female offspring is influenced by the number of amplified drivers and suppressors on both sex chromosomes (Courret et al., 2019; Hughes et al., 2020; Soh et al., 2014). When both driver and suppressor are dosage sensitive, it will cause iterated cycles of expansion (Soh et al., 2014), which causes the occurrence of the antagonistic relationship between the two loci and can be seen as an evolutionary arms race between sex chromosomes (Figure 4) (Courret et al., 2019; Hughes et al., 2020).

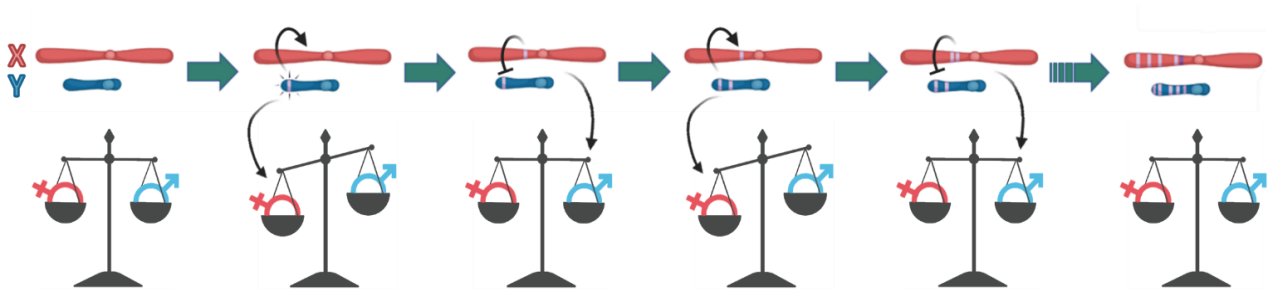


Figure 4. A visualisation of the arms race between sex chromosomes. The gene on one of the sex chromosome (in this example the Y chromosome) causes skewing sex ratios. Since uneven sex ratios are evolutionarily not preferred, at some point the other sex chromosome receives an amplification as well, to suppress or work antagonistic against the first gene. When again an amplification appears on Y chromosome, in time a second amplification will appear again at the X chromosome. Over time this will cause iterated cycles of expansions. (Based on figure 1, Bachtrog (2014).)

For most of the species described above, this theory is only hypothesized and not tested. However, there are strong indications about the presence of an arms race mechanism. There are three species in which this arms race mechanism is investigated in more detail: *D. melanogaster*, *D. pseudoobscura*, and *M. musculus*. In *D. melanogaster*, the genes *Stellate* (*Ste*) and *Suppressor of Stellate* (*Su(Ste)*) work antagonistically (Hurst, 1996). *Ste* is a multicopy X-linked gene that leads to defective sperm by forming crystal-like structures in premeiotic gametes (Courret et al., 2019). *Su(Ste)* is located on the Y chromosome and is also present in multiple copies, supressing *Ste* (Courret et al., 2019; Hurst, 1996). It is proposed that *Su(Ste)* evolved from *Ste*, because of sequence similarity. Later on, there appeared to be high evolutionary pressure for amplification since both X and Y have high copy numbers of *Ste* and *Su(Ste)* (Malone et al., 2015). Depending on the copy number of *Ste*, males with deletions in the *Su(Ste)* region lead to a low fertility or sterility. When fertile, deletions in *Su(Ste)* skew sex ratios towards more female offspring. It is, however, unknown if the meiotic drive is a result of a decrease in copy number of *Su(Ste)* or because of overexpression of *Ste* (Courret et al., 2019).

The driver mechanism in the *D. pseudoobscura* is much less known compared to the *Ste*/*Su(Ste)* mechanism. The *S-Lap1* gene is coamplified on both the X and Y chromosome, and codes for a protein component used for sperm. There is no knowledge about intracellular relation between X and Y-linked *S-Lap1*, and some of the copies are truncated, suggesting different functions (Ellison & Bachtrog, 2019). Although there is no known antagonistic effect, males with an RNA-interference (RNAi) knockout of Y-linked *S-Lap1* produce offspring with sex ratios skewing towards females, suggesting the presence of an arms race (Ellison et al., 2018).

Similarly to the *Ste*/*Su(Ste)* in *D. melanogaster*, in mice there are gene families with intragenomic conflicts, controlling offspring sex ratios (Cocquet et al., 2012; Rathje et al., 2019). The *Sly* (*Sycp3-like Y*-linked) is located on the MSY and represses postmeiotic expression of the sex chromosomes. Evolutionarily coamplified with *Sly*, are the multicopy *Slx* and *Slxl1* in the X-chromosome, which are essential for normal sperm differentiation. *Slxl1* and *Sly* both interact with the protein DKKL1, and it has been shown that *Slx/Slxl1* and *Sly* have an antagonistic effect on gene expression (Cocquet et al., 2012). Likewise, *Slx*/*Slxl1* and *Sly* also interact with the spindling gene family, especially with SPIN1 and SSTY1/2 (Kruger et al., 2019). SPIN1 is a histone effector, which is able to recognize two histones, activates the Wnt-signaling, and regulates RNA stability (Kruger et al., 2019; Su et al., 2014). SSTY1/2 has the same histone binding domain as SPIN1, and is proposed be involved in the control of XY gene expression during spermatogenesis (Comptour et al., 2014; Kruger et al., 2019). When *Slx*/*Slxl1* and *Sly* are dysregulated, interactions with the spindlin gene family may cause a difference in X and Y sperm fitness (Kruger et al., 2019). Partial deletions in the Y chromosome that reduce the *Sly* copy number, lead to an overexpression of several sex-linked genes and cause skewed sex ratios towards females, or male sterility in case of large deletions (Good, 2012; Rathje et al., 2019). In knockout studies for *Slx/Slxl1* or *Sly*, the sex ratios changed and spermatogenesis was impaired. In case of *Slx/Slxl1* knockout, the sex ratio skewed towards males, while *Sly* knockouts had an excess of female offspring. Mice deficient for both *Slx/Slxl1* and *Sly* did have an even sex ratio and normal fertility (Cocquet et al., 2012). Distorted sex ratios and affected fertility is not only seen in knockouts, but also in hybrid species. Closely related mice species all have different number of gene copies of *Slx/Slxl1* and *Sly*, which is for example the case between the *M. m. musculus* and the *M. m. domesticus* (Table 1). Male hybrids with a *M. m. domesticus* as father are sterile, while males with a *M. m. musculus* father have normal fertility (Good, 2012). Infertile hybrids have an excess of *Slx/Slxl1* copies and show a similar phenotype as the *Sly* knockout mice (Bachtrog, 2014). Skewed sex ratios, fertility loss and massive co-amplifications suggest the presence of a dosage sensitive driver and target system (Soh et al., 2014).

Table 1. Extent of coamplification and function of different genes likely involved in a chromosomal arms race. Copy numbers in the X, Y or autosome (A) are shown.

Species	X, Y or autosomal coamplified gene families		Known copy number of genes		Gene function or expression
	X/A	Y	X/A	Y	
Drosophila melanogaster	*Ste*	*Su(Ste)*	10-400	~80	*Ste* forms crystal-like structures in premeiotic gametes. *Su(Ste)* suppresses *Ste* (Malone et al., 2015).
D. pseudoobscura	*S-Lap 1*		2	61	*S-Lap 1* codes for protein component of sperm (Ellison et al., 2018).
Mus musculus domesticus	*Slx/ Slxl1*	*Sly*	50	50	*Slx/Slxl1* are essential for sperm differentiation and are regulated by *Sly* (Cocquet et al., 2012; Good, 2012). *Slx/Slxl1* and *Sly* interact with the spindlin gene family (Kruger et al., 2019).
M. m. musculus			100	80	
Panthera leo	*CCD71L*	*CCD71LY*	~27	~118	*CCD71L* and *CCD17LY* both are expressed in testes, which may suggest a functional relationship (Brashear et al., 2018).
P. tigris			15	12	
Bos taurus	*HSFX*	*HSFY*	11	79	*HSFX* and *HSFY* are expressed exclusively in the testes and male gametes (Hughes et al., 2020).
Homo sapiens	*VCX*	*VCY*	~12	2	*VCX* and *VCY* are expressed in testes, where *VCX* is negatively charged and *VCY* positively charged (Hughes et al., 2020; Lahn & Page, 2000).
Equus ferus caballus	*ETSTY7*		~9	>15	Testes expressed transcript (Janečka et al., 2018).

In cats, cattle and humans, the mechanism of the arms race is proposed as well, although not yet proven. A summary of these species and their genes is presented in Table 1. In cat families, similar to mice hybrids, hybrid backcrosses also give sex-biased litters and infertile males. This is, for example, the case in tiger – lion hybrids. Male offspring are infertile, but if a female liger (offspring of male lion and female tiger) is backcrossed to a lion, the sex ratios skew towards females. If this is due to a similar mechanism as described in the mouse, the autosomal gene *CCD71L* and *CCD71LY* may explain the distorted sex ratios, since the lion and tiger have extensive differences in the number of amplicons of both genes. The *CCD71L* has the highest expression in both the fallopian tubes and testes and *CCD71LY* mainly in testes, which makes a functional relationship between the genes plausible (Brashear et al., 2018).

In bovine, like in cats, also two genes are identified: *HSFY* has 79 copies in the Y chromosome while *HSFX* has three variants in the X chromosome. Of the latter, two variants are single copy and one has nine copies. Similarly to cats, the exact function of *HSFX* and *HSFY* is not known, although it is known that they are expressed exclusively in testes and male gametes. The *HSFX* and *HSFY* genes show a high intra-family nucleotide identity, but the proteins show only 34% amino acid identity. This indicates that *HSFX* and *HSFY* are independently co-amplified and is not a result of X-Y recombination (Hughes et al., 2020).

Lastly there are suggestions that the arms race mechanism may also occur in humans. In humans the *Variable Charge X (VCX)* and *VCY* are coamplified. There are around 12 copies of *VCX* and two copies of *VCY* (Hughes et al., 2020; Zanders & Unckless, 2019). The *VCX* and *VCY* genes are very similar (~96% identical) and they have, similarly to the previously described genes, an expected antagonistic function (Hughes et al., 2020; Zanders & Unckless, 2019). Differences between these genes include high polymorphism of the *VCX* genes, which is probably due to uneven crossing over, and a 30 base-pair sequence that is present once in *VCY*, but has 2-30 tandem repeats in *VCX* genes (Lahn & Page, 2000; Zanders & Unckless, 2019). *VCX* and *VCY* are exclusively active in the male germline, although they may have other functions since *VCX* deletions are associated with intellectual disability. Also, *VCX* encodes for a highly negatively charged protein and *VCY* for a highly positively charged protein, which makes their antagonistic function very plausible (Hughes et al., 2020; Lahn & Page, 2000). Although meiotic drive of the *VCX* and *VCY* genes has not been tested, there is a correlation between *VCX* copy number and both female and male fertility (Zanders & Unckless, 2019). Considering the fact that there are slightly more males conceived in humans makes it very possible that *VCX* and *VCY* play a role in the sex ratio distortion (Lahn & Page, 2000).

Similarly to the species described above, the horse MSY contains 29 X-Y homologs (gametologs), which is so far the highest number of gametologs found in eutherians. Most of them are expressed in testes. The gametologs show some amplification, but this is moderate compared to the other investigated animals. However, very interestingly, there is a testis-specific expressed transcript in the MSY, *ETSTY7*, which is highly amplified in both sex chromosomes and has additional autosomal copies in other equids (Janečka et al., 2018). The origin and functions of *ETSTY7* amplicons are not known, nor are its possible roles in sex chromosome arms race or sex ratio distortion.

1.3. Sex ratios in equines

Although for horses, evidence of an arms race are very hypothetical, equine hybrids have, similarly to felines, significant differences in sex ratios. Normal sex ratios in horses differ slightly per study. The first record of sex ratios in horses was by Charles Darwin in 1874. He obtained 25 560 records of racing Thoroughbreds with a sex ratio of 49.92% males and 50.08% females (Darwin, 1871). More reports of sex ratios vary between 49.3% to 52.5% males (Craft, 1938). One of the recent studies of sex ratios in Australian Thoroughbreds reported 49.95% stallions born over 50.05% mares in 7 578 live foal births (Todd et al., 2020). Significantly different sex ratios have been reported for equid hybrids: of 1 416 mules 44.28 ± 0.89 percent were male compared to 52.52 ± 0.95 percent of males among 1 263 horses (Craft, 1938). The sex ratio of mules was according to Craft (1938), similar to other *Equidae* hybrids, however it should be noted that the sample size of these hybrids were rather small (Craft, 1938). In interspecific hybrids the heterogametic sex (XY or ZW) are often less well represented and more often sterile, because of incompatibility between the sex chromosomes (Haldane, 1922; Hurst & Pomiankowski, 1991). For this reason, the disrupted sex ratios and sterility in the mules are probably not due to an arms race. Although in cats, skewing sex ratios in hybrid back-crossings supported the arms race hypothesis, in horse hybrids sex ratio in back-crossings is difficult to test. Fertile female mules do exist but are very rare because of an uneven number of chromosomes (Johnsen, 1985). The rare occasions of fertile mules make studies of sex ratios in backcrosses difficult to find. Although sex ratios in mules do not support or contradict an arms race mechanism, the X-Y ampliconic regions in the horse genome still makes the hypothesis for the existence of selfish genetic elements plausible in horses.

However, although the hypothesis of a selfish genetic element is a very interesting to potentially explain deviating sex ratios in horses, there may be other explanations. One of the theories, already investigated in horses, is the Trivers-Willard hypothesis (Cameron et al., 1999). This hypothesis suggests that mothers in good conditions would benefit more from producing reproductively variating sons, while mothers in poor conditions benefit more from reproductively stable daughters. This hypothesis is based on the fact that a mother in good condition has more resources to bear and nurse her offspring, and in this way can produce stronger offspring compared to mothers in a poor condition. If the effect of the maternal condition is maintained in the offspring during adulthood, it may affect the offspring's reproductive success. In most species, male reproductive success is more affected by its size, condition and health than female reproductive success. This makes it for a female in good conditions more desirable to produce a son and for females in a poor condition to produce a daughter (Trivers & Willard, 1973). In case of horses, stallions produced by mothers in a good condition may have an advantage over other stallions, and in this way increase their reproductive success. For reproductive stable mare offspring the condition of the mother does impact the reproductive success, which makes them preferable for mothers in poorer conditions.

The model was tested in wild Kaimanawa horses and showed, indeed, that mothers of females were generally in a poorer condition than mothers of males (Cameron et al., 1999). Another study in Camargue horses also reported more female offspring in an environmentally poor year, compared to a rich year (Monard et al., 1997). Furthermore, in support of this hypothesis, older mares sire more female offspring than younger mares (Santos et al., 2015). To explain this biologically is very difficult, although it may have something to do with uterine glucose levels. High energy intake may cause glucose enrichment of the uterine environment, which is unfavourable for female embryos because they fail to develop in high glucose concentrations (Beckelmann et al., 2013).

In addition to physical conditions, sex ratios may also be influenced by psychological stressors. For example in humans, anxiety and depression in mothers are correlated with increasing number of daughters and the same effect was observed under economic stress in East Germany (R. Catalano et al., 2005; Catalano, 2003). It was argued that stress may cause spontaneous abortions and low birth weight, which may influence the sex ratio (Catalano, 2003).

Known paternal effects on the sex ratios are very limited, although there is an effect of the stallions' genetic variation on the sex ratio (Todd et al., 2020). While the Trivers-Willard hypothesis can be applied for horses, it does not explain skewed sex ratios observed in some stallions.

1.4. Sex chromosomal abnormalities

A last explanation for skewed sex ratios in horses may be XY females. After X-monosomy, XY male-to-female sex reversal is the second most common chromosomal abnormality in sterile mares (Mäkinen et al., 2001). Individuals with this genotype have the opposite phenotype, which can be seen as sex reversal. Normally the male phenotype is caused by the sex-determining gene (*SRY*). In horses, this gene is single copy and located in the Y chromosome between two ampliconic sequences, surrounded by repeats. This facilitates the chance of deletions of the *SRY* gene because of intra-chromosomal recombination. The high chance of deletion is well seen in XY females, since around 70% have lost the *SRY* gene compared to 10-20% of XY females in humans (Janečka et al., 2018). XY female horses have phenotypes ranging from feminine to very masculine. The feminine XY females have a very similar phenotype to females with X-monosomy (Raudsepp et al., 2010).

XY sex reversal may be in some occasions hereditary. In humans for example there are several familial forms of XY females (Brauner et al., 2016). However in horses familial forms of XY females are very rare and have only been reported in a few cases so far (Terje Raudsepp, 2020; Villagómez et al., 2011). In one of these studies three XY females were found in one family, two full siblings and one maternal sibling. In this case it was hypothesized to be a X-linked mutation on the androgen receptor gene, causing androgen insensitivity syndrome (Villagómez et al., 2011). Although hereditary effects of XY females are not common, it would be interesting to take familial sex developmental disorders in consideration. Of XY sex reversal females the genotype is often not known and for this reason they are registered as female horses. Because of the unknown genotype, these horses could contribute to skewing sex ratios in studbooks.

1.5. Rationale and Aims

As described above, there are multiple mechanisms leading to distorted sex ratios in animals. These include selfish genetic elements, the arms race, the Trivers-Willard hypothesis, and sex chromosomal abnormalities. This project is very relevant since horses provide a compelling model to investigate selfish genetic elements. So far, selfish genetic elements are rarely investigated in mammals, other than the wood lemming and mice. This research can give some insight in the evolutionary competition between the X and Y chromosome as well as sex-chromosomal abnormalities in horses, and the possibility of a still undescribed mechanism.

This project may not only be important for scientific reasons, but it is as well of interest for horse breeders. Horse breeders often have a preference for an increased number of offspring of one sex. For example, in Polo sport females are preferred, while in racing, males are favoured. The same preference for a sex is visible in other sports; females are more often selected as cutting horses and males are desired for reining. Also, prices paid for the different sexes can differ drastically. For these reasons there are multiple ways to influence sex ratios. Examples of less accurate methods are the nutritional status of the mother, based on the Trivers-Wilard hypothesis, or time of mating. More accurate ways are sex determination during embryo transfer or in utero embryos. Sexing of embryos before insemination is done using male specific fluorescing antibodies, which is non-invasive and determines sex with 82% accuracy. After sexing in utero, pregnancies of the undesired sex are in some countries even terminated (Aurich & Schneider, 2014). The importance of a desired sex is as well seen in cattle. In dairy cows, for example, female offspring is preferred. To accomplish this, in cattle, sex-sorted semen is used, which can result in offspring with over 90% with the desired sex. However, in horses this method is so far not widely used because of several difficulties. Firstly, the best results are obtained by using fresh semen, which means that the mare needs to be close. In contrast to cattle, freezing of horse semen after sorting results in a low fertility. Another reason is that in cattle the number of spermatozoa needed for insemination is only 4 million, while in horses 50-100 million spermatozoa are needed. For these reasons sex-sorted semen is not yet widely implemented in horses (Squires, 2019). Investigating hereditary mechanisms of distorted sex ratio may provide other ways to influence sex ratios in horses. For example, by selecting on possible genetic variances causing distorted sex ratios. However, it is questionable if horse breeders would prioritize on such trait, compared to performance or conformation.

During my master thesis I lay the foundations of a new fascinating project, which will continue after I finish. During my research I had the following aims:

- To explore sex ratios in horses, focussing on Icelandic horses, Standardbred horses, and Coldblooded trotters.
- To estimate the heritability of distorted sex ratios in Icelandic horses.
- To identify families with distorted sex ratios in Icelandic horses.
- To provide a solid basis to continue this research.

2. Materials and Methods

2.1. Datasets

2.1.1. Icelandic horse dataset

WorldFengur (WF) is an international database containing the studbook of all Icelandic horses. WF provides information of Icelandic horses all over the world, such as pedigrees, if horses are dead or alive, offspring, breeding assessment, sport competitions, owners, and breeders (Lorange, 2011). WF provided this study with a transcript containing all Icelandic horses included in the studbook, which was a total of 509 008 horses. Horses were born in 31 different countries, between at least 1860 and 2020, including some unknown birth years. Records included in this transcripts were: horse id, horse name, horse origin, father's id, father's name, father's origin, mother's id, mother's name and mother's origin (WorldFengur, 2021). The WF's id-numbers provide extra information about the horse. The id-number starts with two letters, representing the country of birth, followed by year of birth (four digits), sex (1 = male, 2 = female, and 3 = unknown), area code (two digits), and individuals identifier (three digits) (Hreidarsdóttir et al., 2014). Using Microsoft Excel, the names and origin were removed from the dataset, to reduce the data volume making analysis more feasible. All horses in this dataset were used to generate our main dataset including horse id, father's id, mother's id, sex, number (n) of total offspring, n male offspring, n female offspring, n offspring with an unknown sex, and n offspring with a known sex. To generate the dataset R-studio was used with the openxlsx package (RStudio Team, 2021; Schauberger & Walker, 2020) (Appendix 7.2).

For the downstream analyses additional information was received in a second transcript including; the last known country of residence, life alive/dead status (alive, dead, stillborn, slaughtered and unknown), and date of castration in case of a gelding (WorldFengur, 2021). This transcript was used to prune out horses of interest, for analyses planned in the future, including selecting horses for semen collection or whole genome sequencing.

2.1.2. Standardbred horse and Coldblooded trotter dataset

For the trotters, pedigrees for Coldblooded trotters and Standardbreds were found at the Swedish Trotting association website. These pedigrees were used to generate manually the second dataset. Similar to WF this studbook includes information such as name, id, sex, pedigrees, life status, offspring, breeding assessment, sport competitions, owners and breeders (Sportapp Svensk Travsport, 2021). For this dataset, both Coldblooded trotters and Standardbreds were included. In the selection procedure, stallions born between 1990 and 2020 with at least 10 offspring in Sweden were incorporated, resulting in 504 Standardbreds born in 10 different countries and 125 Coldblooded trotters, born Sweden or Norway. The lower limit of 10 offspring born in Sweden was chosen to make offspring more accessible to enable testing in future analyses (i.e. alive). This dataset included name, year of birth, father's name, n total offspring (also including international offspring), n male offspring, and n female offspring.

2.2. Statistical analyses

In Icelandic horses sex ratios were calculated including all individuals in the dataset over decades (1950 – 2019). Of Standardbreds and Coldblooded trotters only stallions were included in the analyses and for this reason the overall sex ratio was calculated using the offspring of included stallions. In all breeds significantly skewing horses (horses with significantly more offspring of one sex than would be expected from equal sex ratios) were identified using a Pearson's chi-square tests with one degree of freedom and a 1:1 sex ratio as the null hypothesis. For example, if a horse had 24 offspring the expected number of males would be 12. The Chi-square tests where done using Excel and RStudio (RStudio Team, 2021) (Appendix 7.2).

Chi-squared tests on its own only explain if the observed data is significantly different compared to an expectation, a method was search to differentiate more between the tests and discriminate more and less reliable results. Sample size (number of offspring) and proportion of one sex both influence the reliability of the result. To be more aware of type 2 errors, post-hoc power calculations were performed. The calculations were accomplished using the Real Statistics Resource Pack for Excel (Zaiontz, 2021).

In the dataset, including all Icelandic horses, a very large number of chi-square tests were performed, thus a Bonferroni correction was applied to correct for multiply testing. Without adjustment of the *p-values* the probability of a false positive results can be even higher than 0.05. The Bonferroni correction is very straightforward and multiplies the *p-value* by the number of tests (Aickin & Gensler, 1996). One downside of the Bonferroni correction is that the test is harsh and may exclude more significant tests than necessary. For that reason, the correction was not only done on 195 600 tests (number of horses with at least one offspring), but as well differentiated between males and females with eight or more offspring (8787 for males and 13 703 for females). Eight or more offspring was chosen for Icelandic horses since that is the smallest number of offspring for which a chi-square test can be significant (Appendix 7.3, Table 13).

2.3. Heritability

To investigate if distorted sex ratios are heritable, the heritability was estimated in the Icelandic dataset. In this analysis, only horses, both male and female with eight or more offspring were included. To estimate the heritability a parent offspring regression was used, one including fathers and one including maternal grandfathers. Maternal grandfathers were included to investigate if the mother of the horses as well had an effect on the proportion of male offspring. These maternal grandfathers were identified by Python, using a script written in Pycharm to generate a pedigree for "*x"* number of generations (pedigree identifier script) (Appendix 7.4) (Jetbrains, 2020; Python Software Foundation, 2020). The proportion of male offspring was calculated for each horse, males, females, fathers, and maternal grandfathers, using Excel. A second python script was used to match the proportions of fathers and maternal grandfathers with their offspring and grand-offspring, creating a file including id, n male offspring, n offspring with a known sex, proportion male offspring (0 – 1), id-father, n male offspring father, n offspring with a known sex father, proportion of male offspring father, id-maternal grandfather, n male offspring of each maternal grandfather, n offspring with a known sex of each maternal grandfather (Appendix 7.5.1). Rstudio was used to plot correlations between parents and offspring and estimate the heritability using a parent-offspring regression. Some of the plots were made using the ggplot package (Wickham, 2016) (Appendix 7.5.2). The regression coefficients were multiplied by two and four for father and maternal grandfathers respectively, since only one parent was included per regression (Pierce, 2017).

After the primary heritability calculations, the heritability was estimated using different minimal numbers of offspring in fathers and sons. Eight offspring is a small number of offspring, where the sex of one offspring has a large effect on the proportion of male offspring. For this reason estimating the heritability with a larger number of offspring may decrease bias. During these analyses we used four additional thresholds, of 15, 30, 60, and 100 and more offspring.

2.4. Families with distorted sex ratios

After heritability calculations, Icelandic horse families in which offspring's sex ratios are frequently distorted were investigated. Family investigations are important, since the mutations or genetic variations that cause the distorted sex ratios in horses are not known. If distorted sex ratios are caused by multiple mutations, it is important that genetically tested horses have the same mutations or mutation pattern to be able to find the causative mutation(s). The chance to include only horses with the same genetic mutation(s) increases when we focus the gene mapping in one family. Also, looking at family relationships may give insight into inheritance patterns. In the family identifications a slightly different approach was used compared to the heritability analyses. During the heritability calculations a proportion of male offspring was used, as a continuous scale. However, horses with a high or low proportion of male offspring are not always significantly skewing. This could for example be the case in animals with a low number of offspring. For this reason, the trait was defined as horses with eight or more offspring, with significantly more offspring of one sex than would be expected from an even sex ratio.

The investigation of families was started with the hundred horses with the lowest p-values, these horses were as well the only significant skewing horses after the Bonferroni correction, for the test of 8787 stallions. All of the hundred horses were skewing towards females (had significantly more female offspring). These horses were the starting population to identify families. Using the pedigree identifier script, parents and grandparents were identified (Appendix 7.4). In Excel, individuals appearing multiple times in the pedigree of the 100 horses were searched, to identify families in this subset of the data. A total of 13 families were identified, which included at least three horses. Starting with the largest identified family, of 16 individuals within the 100 horses, the oldest known male ancestor was identified by passing through the studbook. A python script was created to find all significantly skewing offspring of this ancestor. The data resulting from this script also included the n sons with eight or more offspring, n significant skewing sons, proportion of significant skewing sons, n maternal grandsons with eight or more offspring, n significant skewing maternal grandsons, and proportion of significant skewing maternal grandsons of each individual (Appendix 7.6.1). The proportions of significant skewing sons and maternal grandsons were used to tell something about inheritance patterns, by investigating the proportions of skewing individuals.

When only significant skewing horses were included, some stallions which may be prone to have more offspring of one sex were missing. This may happen because of chance. For example, slightly skewing horses are not always significant, but offspring of these horses may have a larger chance to skew their sex ratio. To solve this problem, the script was adjusted to include non-significant skewing horses as well. It was however difficult to choose were to draw a line with the inclusion of non-significant skewing horses. A too strict threshold may exclude horses with genetic variants of interest and a too lenient threshold may include horses which do not have the genetic variants of interest. Also a prominent problem caused by a too lenient threshold may be an increasing overlap between families. For this reason, different thresholds were tried. The thresholds were based on the proportion of significant skewing sons or grandsons because horses passing skewing sex ratios down to their progeny have a larger chance to have the genetic mutation(s) causing the skewing sex ratios, even if they do not show significant skewing sex ratios in their own offspring (Appendix 7.6.1). For example, if the threshold was 0.8, horses with 80% or more significant skewing sons or grandsons were included. The following thresholds were tested: 0 (including all male progeny), 0.3, 0.5 and 0.8. A more elaborate explanation is presented in Figure 5.

During the exploration of the different thresholds, the proportion of sons with significantly more female offspring and significant skewing maternal grandsons were calculated per threshold and family, to make an indication of the inheritance pattern. This proportion was calculated by dividing the number of significant skewing sons by the total number of sons with eight or more offspring. The same calculation was done with the number significantly skewing maternal grandsons and the total number of maternal grandsons with eight or more offspring.

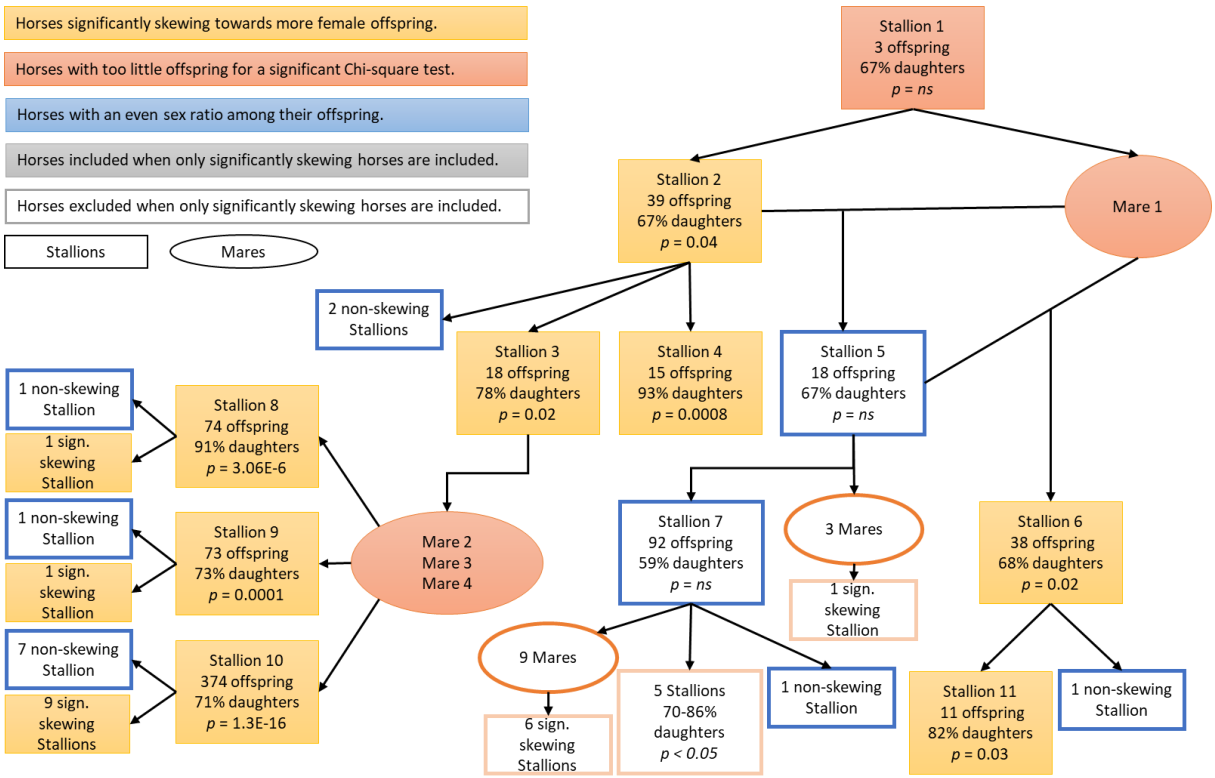


Figure 5. This is an example of a possible family tree to elaborate more on the reasoning behind the threshold system. In this figure yellow boxes show stallions skewing towards more female offspring, red boxes with female horses or horses with too little offspring for a significant test result, and blue boxes are stallions which do not significantly skew. Filled boxes indicate that horses are included in the initial script and open boxes indicates that the horses were not included. Stallion 1 is the identified oldest known male ancestor and has a son and a daughter. The script will identify male progeny which skews significantly towards more female offspring. In this way Stallions 2, 3, 4, and 6 will be successfully identified. Stallions 2, 3 and 4 via their father and Stallion 6 via its mother. However, the other son of Stallion 5, Stallion 7, does not skew significantly and initially the script would exclude him from the family because the inclusion of progeny stops by a stallion with an even sex ratio (Stallion 5). Looking at the offspring of Stallion 7 may suggest that it does have the genetic variation causing skewing sex ratios, since it has many progeny skewing towards more female offspring. With the threshold system these non-skewing stallions can be included on different levels. For example, Stallion 5 has two sons and three maternal grandsons with enough offspring for a significant chi-square (≥8 offspring). The proportion of sons with skewing sex ratios would be 50% and the proportion of grandsons with skewing sex ratios 33%. So in case of a threshold of 0.8 this horse will be excluded with all its progeny. However, with a threshold of 0.5 this horse and its progeny will be included in the family.

2.5. A pilot experiment to investigate sex ratios in semen

To continue this project in the future, it would be very interesting to investigate sex ratios in semen of stallions with a disrupted sex ratio among their offspring. Sex ratios in semen would give information about in which stage of conception the sex ratio skews. A selection of skewing horses was made of non-gelding horses living in Sweden or Iceland. This selection was ranked based on high significance level, high proportion of offspring in one sex, and number of offspring, with a higher number preferable. The focus was both on female skewing horses and male skewing horses, because comparison between these two groups and a control group would be very interesting. Family connections between the horses were checked, because in this part of the study unrelated individuals would be preferable. Unrelated individuals are favoured because if there are multiple genetic mechanisms skewing sex ratios, the chance to detect them is larger in unrelated individuals compared to family members. At the moment the selection of stallions to investigate is still under development and consideration. For example for so far, no decision has been made about how far back the horses need to be unrelated. Going back in the family tree it may be possible to make educated estimates were a new branch of a family start with possible a different genetic variation. When the selection of stallions is more finalized, collection of semen of the selected horses will follow.

To determine sex ratios in horse semen, X and Y proportions can be assessed using quantitative real-time PCR (qPCR). Although sex ratios in semen are investigated using qPCR for example in cattle, in horses it is not common (Resende et al., 2011). For this reason, a new protocol was tested on horse semen to investigate which treatment of semen would work the best. We tested three groups, purified semen, raw semen, and cooked semen. Semen was purified with one-layer of 80% Equipure™ gradient, discarding somatic cells and underdeveloped sperm. Following, DNA from the purified semen and raw semen were extracted with a QiaSymphony DNA extraction using a sperm protocol (adjusted from QIAGEN, 2013). In raw and purified semen the DNA concentration was measured using nanodrop, after which purified semen was diluted to a comparable concentration as the raw semen. To look if a very simplistic DNA extraction would give a result as well, diluted semen was incubated for 20 minutes at 98 °C (cooked semen). Nuclease-free water was used as an control. In total four primers were used, including the sex-chromosomal primers of the horse *AR* gene, for the X chromosome, and *DDX3Y*, for the Y chromosome. *GAPDH* and *actin-β* were used as an autosomal controls (Table 2) (Bogaert et al., 2006; Paria, 2009; Raudsepp et al., 2004). The primers were dyed using PowerUp™ SYBR™ green master mix and the fold change between the groups was calculated using the 2^{-ddct} methods, with *actin-β* to normalize and raw semen as a control (Kholghi et al., 2020). For a more detailed protocol, see Appendix 7.7.

Table 2. Equine primers used for qPCR reaction.

Chromo-some	Gene symbol	Gene name	Primer sequence (5'-3')		Product size	Annealing Temperature (°C)
X	*AR*	*androgen receptor*	Forward:	CATGCTCTGCCCATTGACTA	107	54.9
			Reverse:	TTGCAGCTTCCACAAGTGAG		55.7
Y	*DDX3Y*	*DEAD box helicase 3*	Forward:	CTCGAGATCCAAAACTGCTG	181	53.5
			Reverse:	GCTGGTCTGGACCTGAACTC		57.6
Autosome	*GAPDH*	*GAPDH*	Forward:	CCTTCTCTTGCTGGGTGATTG	103	55.9
			Reverse:	GACAATGAATTTGGCTACAGCA		53.8
Autosome	*ACTB*	*actin-β*	Forward:	CCAGCACGATGAAGATCAAG	88	53.5
			Reverse:	GTGGACAATGAGGCCAGAAT		55.1

3. Results

3.1. Data characteristics

The transcript of the studbook of Icelandic horses consisted of 509 008 individuals, born in 31 countries. Most registered horses were born in Iceland, 331 806 (65.2%), followed by Germany, 54 996 (10.8%). The sex ratios ranged from 42.7% - 60.6% males and 47.9% - 57.3% females in countries with at least 30 registered horses (Appendix 7.1 Table 11a). The Icelandic horse dataset included in total 222 356 (43.7%) males, 283 414 (55.7%) females, and 3239 (0.6%) horses with an unknown sex, born between 1860 and 2020. The distribution of horses born each year is visible in Figure 6, showing an exponential increase in number of registrations around 1950 with the largest number of registrations around 2010. The sex ratios differed between decades, starting with 30.9% males in 1950-1959 and 49.0% in 2010-2019 (Appendix 7.1, Table 12). Horses with an unknown sex did not have any offspring, while 18.3% of all males had at least one offspring, as did 54.7% of all females. In the studbook only 357 stillborn foals were registered, of which most were born after 2009. Of these stillborn foals 48.7% were male, 43.7% female and 7.6% had an unknown sex. The mean incidence based on the reported stillborn foals was 0.25% (95% Confidence Interval: 0.11 - 0.40) over 2009 – 2020. The distribution of number of offspring of all Icelandic horses with offspring is presented in Figure 7. Of horses with eight or more offspring the mean number of offspring was 44 in males and 10.4 in females. The median was 18 offspring in males and 10 offspring in females. The maximum number of offspring was 1347 in males and 21 in females (Table 3).

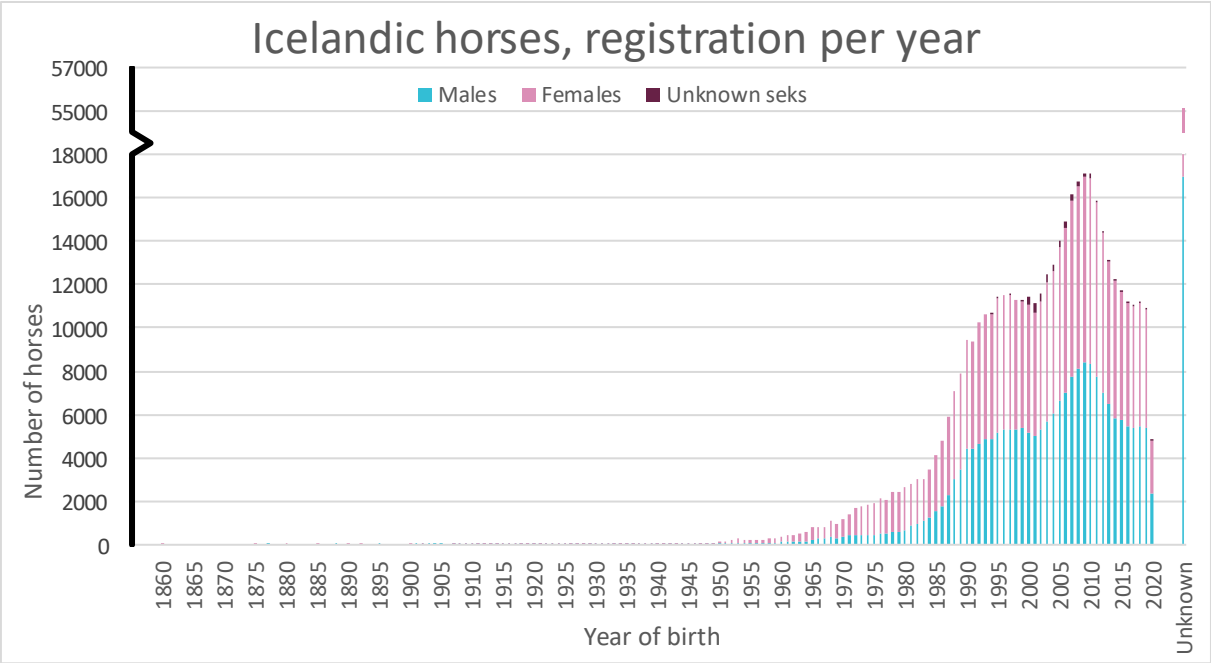


Figure 6. A distribution figure to give an indication about the birth registration of Icelandic horses over time.

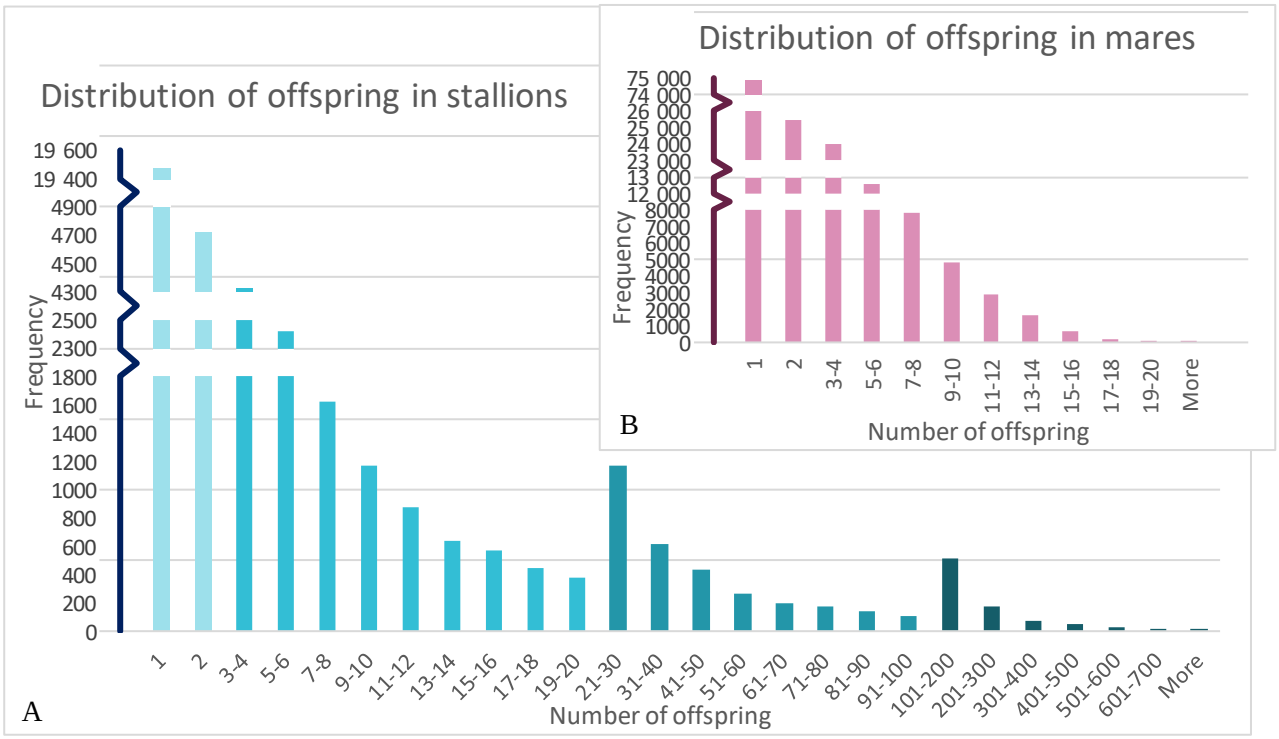


Figure 7. The distribution of offspring in stallions (A) and mares (B). The different colours in A indicate different X-axis intervals.

Table 3. Descriptive values Icelandic horse dataset, including the number of horses with at least one offspring and at least eight offspring. The mean number of offspring and median number of offspring were calculated only using horses with eight or more offspring.

	Male (%)	Female (%)	Total
Horses with one or more offspring	40 627 (18.3)	154 975 (54.7)	195 600
Horses with eight or more offspring	8787 (4.0)	13703 (4.8)	22 490
The mean and median are calculated with horses with eight or more offspring.			
Mean number of offspring	44.5	10.4	
Median number offspring	18	10	
Maximum number of offspring	1347	21	

Of the 8787 Icelandic stallions 1404 did significantly skew to more female or more male offspring. Females had similar to males a larger proportion of horses skewing towards more female offspring, however, the proportional difference between female-skewing and male-skewing horses was much smaller. In females, 3.5% skewed towards more female offspring and 2.3% skewed towards more male offspring. Compared to 13.9% female-skewing stallions and 2.1% male-skewing stallions. Of the skewing horses, most females had a sex ratio of 85-90% towards one sex in their offspring. In case of males, most female-skewing horses had a sex ratio of 85% female offspring. Both, male- and female-skewing horses did not show a clear normal distribution, although female-skewing horses were closer to a normal distribution (Figure 8). Two Bonferroni corrections were computed, for all horses with at least one offspring, and separate for males and females (for 8787 and 13 703 tests). After the corrections only female-skewing males were still significant (Table 4).

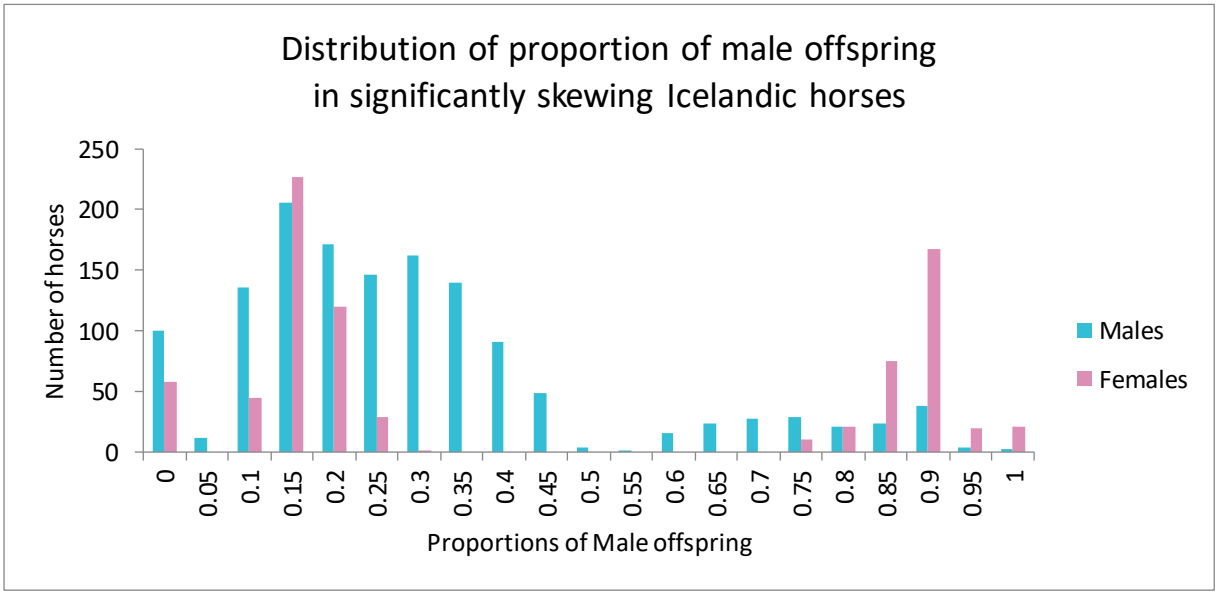


Figure 8. The distribution of the proportion of male offspring in significantly skewing Icelandic horses, both in males and females.

Table 4. Chi-square characteristics in Icelandic horses. Included were only horses with eight or more offspring, which were 8787 males and 13 703 females. Of the significantly skewing horses ($\chi 2$: <0.05), some horses skewed to more female offspring (Female-skewing) and some to more male offspring (Male-skewing). The Bonferroni correction was performed for 195 600 tests, which is the number of horses with at least 1 offspring and for 8787 tests, which is the number of males with eight or more offspring. All horses significant after the Bonferroni corrections were males skewing towards more female offspring.

	Males (%)	Females (%)	Total (%)
Horses with eight or more offspring	8787	13 703	22 490
Significantly skewing horses (p<0.05)	1404 (16.0)	794 (5.8)	2198 (9.8)
Female-skewing	1218 (13.9)	480 (3.5)	1698 (7.6)
Male-skewing	186 (2.1)	314 (2.3)	500 (2.2)
Sign. after Bonferroni correction (195 600)	53 (0.60)		
Sign. after Bonferroni correction (8787)	100 (1.14)		

The dataset generated from the Swedish trotter association consisted of 629 stallions, 125 Coldblooded trotters and 504 Standardbreds. The Coldblooded trotters were born only in two countries, Norway, 78 (62.4%) and Sweden, 47 (37.6%). Standardbreds were born in 10 different countries, with the highest number of horses born in the United States of America, 192 (37.9%) and Sweden, 163 (32.2%) (Appendix 7.1, Table 11). The sex ratio was calculated using the sex of the offspring of the stallions which resulted in overall 50.5% males in Coldblooded trotters and 48.9% males in Standardbreds. In Coldblooded trotters the sex ratios did not differ drastically between Norway and Sweden, 50.4% and 50.7% male offspring respectively. In Standardbreds, in countries with at least ten horses, the proportion of males ranged from 46.6% and 50.7% (Appendix 7.1, Table 11b). Although the inclusion criteria included horses born between 1990 and 2020, no horses born after 2015 had at least 10 offspring born in Sweden. Coldblooded trotters and Standardbreds had more offspring than Icelandic horses, on average 101.4 and 153.4 respectively. The median was 55 offspring in Coldblooded trotters and 134 offspring in Standardbreds. The maximum number of offspring was 990 in Coldblooded trotters and 1517 in Standardbreds (Table 5).

Table 5. Descriptive values Swedish trotters dataset, including the number of horses, the number of offspring, the mean, the median and maximum number of offspring.

	Coldblooded trotters (%)	Standardbred horses (%)	Total (%)
Stallions included	125	504	629
Number of offspring	12 669	77 639	90 308
Number of male offspring	6363 (50.5)	37 939 (48.9)	44 332 (50.1)
Mean number of offspring	101.4	153.4	143.1
Median number offspring	55	134	66
Maximum number of offspring	990	1517	

Of the 125 Coldblooded trotters, 8 (6.4%) did significantly skew to more female or more male offspring. The 504 Standardbreds had a larger proportion of skewing sex ratios, 58 (11.5%). Standardbreds had similar to Icelandic horses a larger proportion of horses skewing towards more female offspring (11.5%) than skewing towards males (2.6%). Coldblooded trotters had a larger proportion of horses skewing towards more male offspring (4% male-skewing versus 2.4% female-skewing) (Table 6). Noticeable, of the 5 significant male-skewing Coldblooded trotters included two half-brothers. The Coldblooded trotters did not have enough significantly skewing individuals to include them in a distribution graph. However, in Standardbreds skewing horses had mostly a sex ratio around 60% of their offspring skewing towards one sex (Figure 9).

Table 6. Chi-square characteristics in trotters. Of the significantly skewing horses ($\chi 2$: <0.05), some horses skewed to more female offspring (Female-skewing) and some to more male offspring (Male-skewing).

	Coldblooded trotters (%)	Standardbreds (%)	Total (%)
Stallions Included	125	504	629
Significantly skewing horses (p<0.05)	8 (6.4)	58 (11.5)	66(10.5)
Female-skewing	3 (2.4)	45 (8.9)	48(7.6)
Male-skewing	5 (4.0)	13 (2.6)	18(2.9)

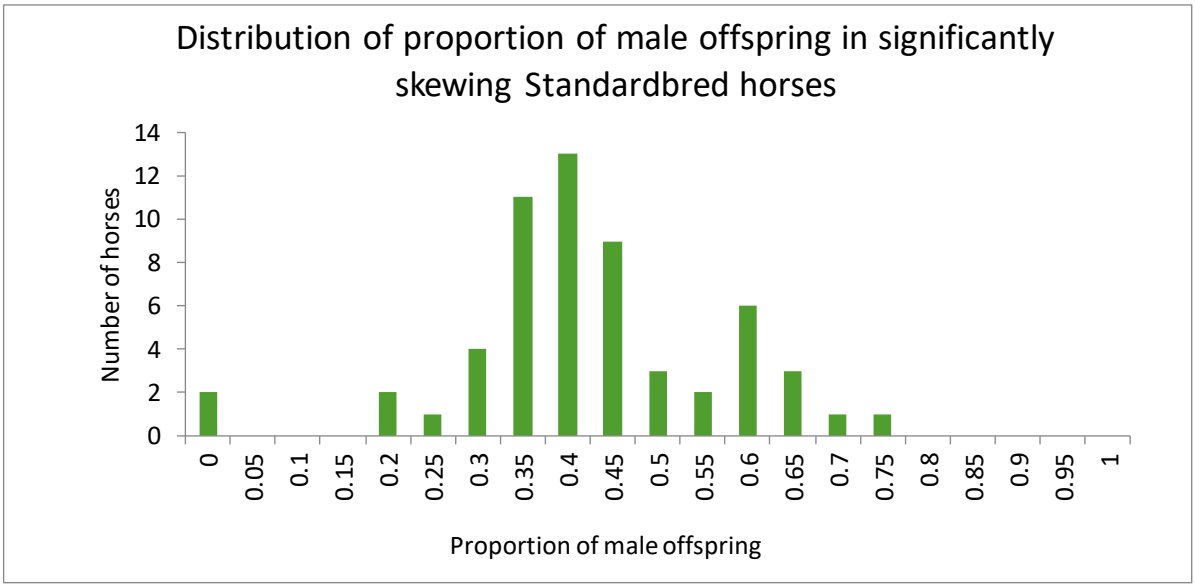


Figure 9. The distribution of the proportion of male horses in significantly skewing Standardbred horses.

3.2. Heritability

To estimate heritability, regression analyses were made using the proportion of male offspring of father and sons and daughters, and maternal grandfather and grandsons and granddaughters. There was no clear relation between the sex ratios in father and daughters, and maternal grandfathers and granddaughters (estimated heritability of 0.08 and 0.13, respectively). However, larger regression coefficients were estimated between fathers and sons and maternal grandfathers and grandsons, with estimated heritabilities of 0.29 (Standard Error regression coefficient (SE): 0.08) and 0.56 (SE: 0.12), respectively (Appendix 7.5.2, Figure 14), using the limit of at least 8 offspring.

The heritability increased with number of offspring up to very high estimates (Table 7 and Figure 10). However, it should be noted that the number of skewing horses decreased when the number of offspring increased, which means that horses with an even sex ratio became an increasing majority (Appendix 7.5.2, Figure 15).

Table 7. Increasing heritability when father and son have an increasing number of offspring, including the number of parent-offspring pairs (n), the regression coefficient, heritability and the standard error.

	n	regression line	h^2	Standard Error co-efficient
Father and sons with eight or more offspring	6935	0.4068 + 0.1453x	0.2906	0.07811
Father and sons with 15 or more offspring	5209	0.3724 + 0.2121x	0.4242	0.07258
Father and sons with 30 or more offspring	2714	0.3190 + 0.3214x	0.6428	0.06485
Father and sons with 60 or more offspring	1402	0.2724 + 0.4154x	0.8308	0.06036
Father and sons with 100 or more offspring	807	0.2558 + 0.4503x	0.9006	0.05518

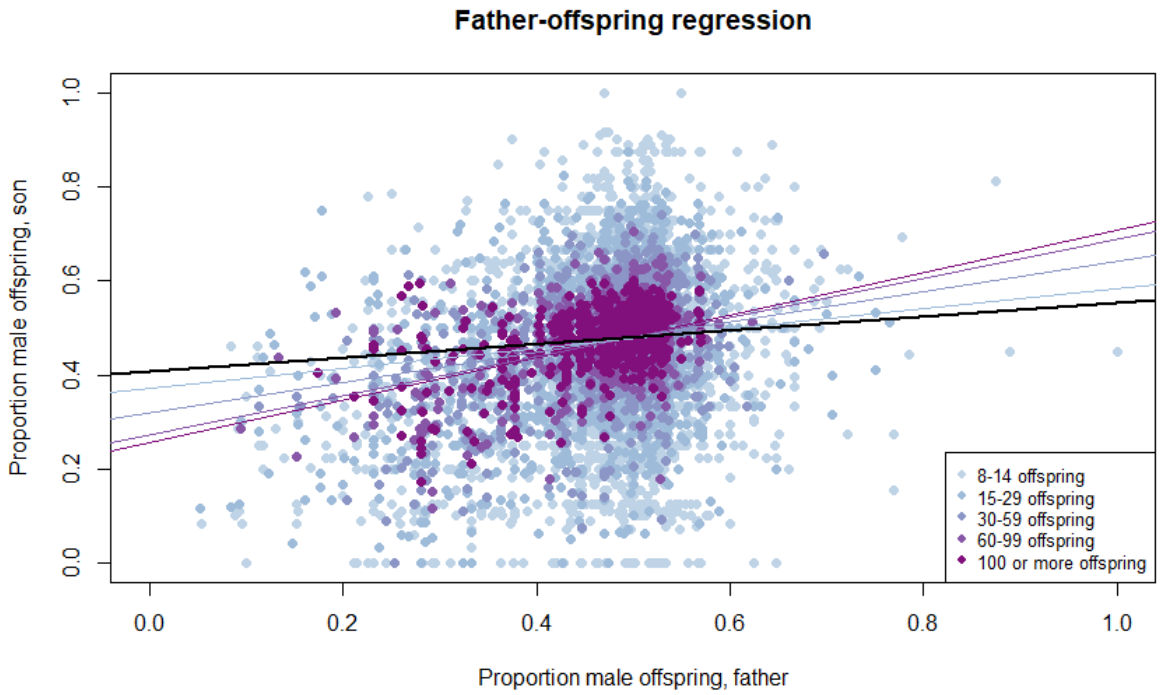


Figure 10. Estimation of heritability from father-son, including 6935 father-son pairs. The colours represent the number of offspring in both fathers and sons, starting with 8 offspring until 100 or more offspring. The black line represents the regression coefficient over the whole sample (h^2 = 0.2906 (SE: 0.07811)), parent offspring pairs with eight or more offspring. The coloured lines represent the regression coefficient of the different groups with corresponding colours to the groups, starting with 15 or more offspring, 30 or more offspring, 60 or more offspring, and ending with 100 or more offspring.

3.3. Families with distorted sex ratios

The 100 most significantly skewing stallions were used to identify families. These were as well the only horses still significantly skewing after a Bonferroni correction for 8787 horses. Of these horses, their uncorrected *p*-value ranged from 5.22E-6 to 9.40E-29 and skewed all towards more female offspring (proportion between 62-97% female offspring). Of the 100 most significant skewing horses, 13 families were identified, including three or more horses. When running the last known male ancestor of several families through the script more distantly related individuals were found and multiple of the original 13 families were dissolved in the new families. Meaning that at the end of the analysis five families included, almost all individuals of the original 100 were assigned to a family. Family 2 was considered the largest family with 2739 individuals. Of the original 100 most significant skewing horses, only 6 were not found in any family. For these six stallions the last known male ancestor was found and 6 separate small families were identified, only including a few individuals skewing significantly towards more female offspring (Table 8).

There were different thresholds of the inclusion of non-significant skewing horses tested on the five identified families, starting with only significant skewing horses, horses with 30%, 50%, and 80%, or more of their sons or maternal grandsons skewing towards more female offspring and lastly families with all existing male descendants. With every threshold, the number of horses per family increased, especially at the threshold including all male descendants. The goal of this analyses is to determine the best threshold to include significantly skewing horses with possibly the same genetic variation without creating too much overlap between the different families. With the inclusion of only significant skewing progeny the families included between 27 and 429 significant skewing horses. This number increased when the threshold criteria changed with at largest 6429 individuals in family 2 with all male descendants included. Between the 0.8 and 0.5 threshold was the least difference, family 1 and 3 were defined exactly the same when using either of those two thresholds (Table 8).

There were multiple horses overlapping between different families, with some horses even between multiple families (Table 9). For the first threshold, including only significantly skewing horses, all families combined included 733 significantly skewing horses, of which 466 (63.6%) did only appear in a single family (unique horses). The 0.8, 0.5, 0.3, and all male descendants thresholds included respectively 826, 849, 923, and 1134 significant skewing horses, of which 485 (59.2%), 464 (54.7%), 461 (49.9%), and 399 (35.1%) were unique horses. The number significant skewing horses between the first and second threshold (0.8) increases with 93 horses. The unique horse change is manly visible in Family 1, where 29 unique horses are added. Between the 0.8 and 0.5 thresholds the number of significant skewing horses does not increase drastically. Families 1, 3 and 4 remain the same during this threshold, although there are less unique horses in Family 1. Despite that the number of significant skewing horses become larger in the last two thresholds, the number of unique horses decreases. For this reason the 0.8 threshold may best applicable to identify the families. Appendix 7.6.2 includes a small fragment of the family tree of Family 2. This tree was produced for a future project, where the largest family was chosen as family of interest.

Table 8. Descriptive statistics per family including different thresholds. For each threshold the number of horses is included (n), the number of horses with 8 or more offspring (n ≥8 offsp.), and the number of significant skewing horses (n sign.) are included. Horses 1 – 6 were not assigned to any family.

	Only significant skewing horses included		>80% of sons/maternal grandsons skewing			>50% of sons/maternal grandsons skewing			>30% of sons/maternal grandsons skewing			All sons/maternal grandsons included		
Families	n	n sign.	n	n ≥8 offsp.	n sign.	n	n ≥8 offsp.	n sign.	n	n ≥8 offsp.	n sign.	n	n ≥8 offsp.	n sign.
Family 1	1952	299	2890	441	413	2890	441	413	3167	496	443	15 674	5448	691
Family 2	2739	429	3058	513	485	3604	574	536	5278	724	646	18 884	6429	875
Family 3	196	27	201	31	29	201	31	29	202	31	29	800	269	46
Family 4	356	48	420	61	57	420	61	57	470	73	65	998	269	78
Family 5	1696	247	1962	294	281	1991	304	288	2206	347	314	12 154	4291	547

	Only significant skewing horses included		All sons/maternal grandsons included		
Individually found families	n	n sign.	n	n ≥8 offsp.	n sign.
Family horse 1	47	12	52	17	12
Family horse 2	39	9	96	28	10
Family horse 3	85	7	120	30	7
Family horse 4	5	3	5	3	3
Family horse 5	4	1	7	3	2
Family horse 6	6	2	13	3	2

Table 9. Number of significantly skewing overlapping individuals between different families. The coloured boxes are unique individuals who do not overlap with any other family. Some horses overlap with multiple families.

Only significant skewing horses included.

	Family 1	Family 2	Family 3	Family 4	Family 5	n sign.
Family 1	127	134	6	11	60	299
Family 2	134	209	1	12	113	429
Family 3	6	1	13	0	8	27
Family 4	11	12	0	18	15	48
Family 5	60	113	8	15	99	247

Threshold of 0.8, includes horses with skewing sex ratios in at least 80% of their sons or maternal grandsons.

	Family 1	Family 2	Family 3	Family 4	Family 5	n sign.
Family 1	156	169	12	20	93	413
Family 2	169	206	4	19	124	485
Family 3	12	4	12	0	9	29
Family 4	20	19	0	17	16	57
Family 5	93	124	9	16	98	281

Threshold of 0.5, includes horses with skewing sex ratios in at least 50% of their sons or maternal grandsons.

	Family 1	Family 2	Family 3	Family 4	Family 5	n sign.
Family 1	128	241	12	20	95	413
Family 2	241	206	5	19	147	536
Family 3	12	5	12	0	9	29
Family 4	20	19	0	17	16	57
Family 5	95	147	9	16	98	288

Threshold of 0.3, includes horses with skewing sex ratios in at least 30% of their sons or maternal grandsons.

	Family 1	Family 2	Family 3	Family 4	Family 5	n sign.
Family 1	116	288	12	26	111	443
Family 2	288	239	9	27	193	646
Family 3	12	9	10	0	9	29
Family 4	26	27	0	15	20	65
Family 5	111	193	9	20	84	314

All sons and maternal grandsons included.

	Family 1	Family 2	Family 3	Family 4	Family 5	n sign.
Family 1	118	531	27	36	339	691
Family 2	531	194	26	45	442	875
Family 3	27	26	9	4	22	46
Family 4	36	45	4	13	28	78
Family 5	339	442	22	28	65	547

When looking at the proportions of skewing offspring and grand offspring, these decreased with lower thresholds. At the first threshold, only including significantly skewing progeny. These significantly skewing horses had between 29% - 39% skewing sons within the different families. Skewing maternal grandparents had between 13% - 18% of their grandsons skewing. With the less strict 0.5 threshold the proportions did not change extensively, however with all male progeny included the proportions became much lower, 10% – 18% of the sons were skewing and 8% - 13% of maternal grandsons (Table 10).

Table 10. For three different thresholds, the number of stallions (n stallions), number of sons with eight or more offspring (n ≥8 sons), number of significantly skewing sons (n sign. sons), proportion of significantly skewing sons, number of maternal grandsons with eight or more offspring (n ≥8 m. grandsons), number of significantly maternal grandsons (n sign. m. grandsons), the proportion of significantly skewing maternal grandsons.

Only significant skewing horses included	n stallions	n ≥8 sons	n sign. sons	Proportion sign. sons	n ≥8 m. grandsons	n sign. m. grandsons	Proportion sign. m. grandsons
Family 1	1799	452	175	0.39	955	161	0.17
Family 2	2513	877	269	0.31	1605	242	0.15
Family 3	183	38	14	0.36	94	15	0.16
Family 4	331	65	22	0.33	145	26	0.18
Family 5	1588	520	152	0.29	862	110	0.13
All families combined	6414	1952	632	0.32	3661	554	0.15
Threshold: >50% of sons/maternal grandsons skewing							
Family 1	2678	837	254	0.30	1545	218	0.14
Family 2	3313	1235	358	0.29	2139	308	0.14
Family 3	188	41	16	0.39	96	15	0.16
Family 4	390	74	28	0.38	170	29	0.17
Family 5	1863	589	184	0.31	977	128	0.13
All families combined	8432	2776	840	0.30	4927	698	0.14
All sons/maternal grandsons included							
Family 1	15 213	5309	532	0.10	4107	338	0.08
Family 2	18 226	6318	697	0.11	5476	543	0.10
Family 3	778	254	31	0.12	171	16	0.09
Family 4	951	207	38	0.18	328	43	0.13
Family 5	11 840	4183	410	0.10	2933	235	0.08
All families combined	1729	16271	1708	0.10	13015	1175	0.09

3.4. qPCR pilot experimentation

The concentration of 93.13 ng/µl DNA in purified semen was diluted with nuclease-free water to 39.16 ng/µl which was the same concentration as found in the raw semen. The qPCR assays of purified semen and raw semen were successful. The cooked semen had melting curves with the *GAPDH* and *Actin-β* primer, however the Ct values were much higher. The melting curves of *AR* and *GAPDH* gave both a clean single peak, while *Actin-β* included a smaller second peak. *DDX3Y* did give a single peak, although it was only detected in the purified semen and the water-control with similar Ct values, which made the primer unreliable. *GAPDH* was as well detected in the water-control, however with an alternate peak (Tm of 75 °C compared to 83 °C) with a much higher number of PCR cycles (Figure 11). Unfortunately the sex ratios in horse semen could not be determined because of the malfunction of the *DDX3Y* primer. However, the relative gene copy number was assessed between purified and raw semen. The relative gene copy number of *AR* and *GDAPH* did not much differ much between purified and raw semen (Figure 12).

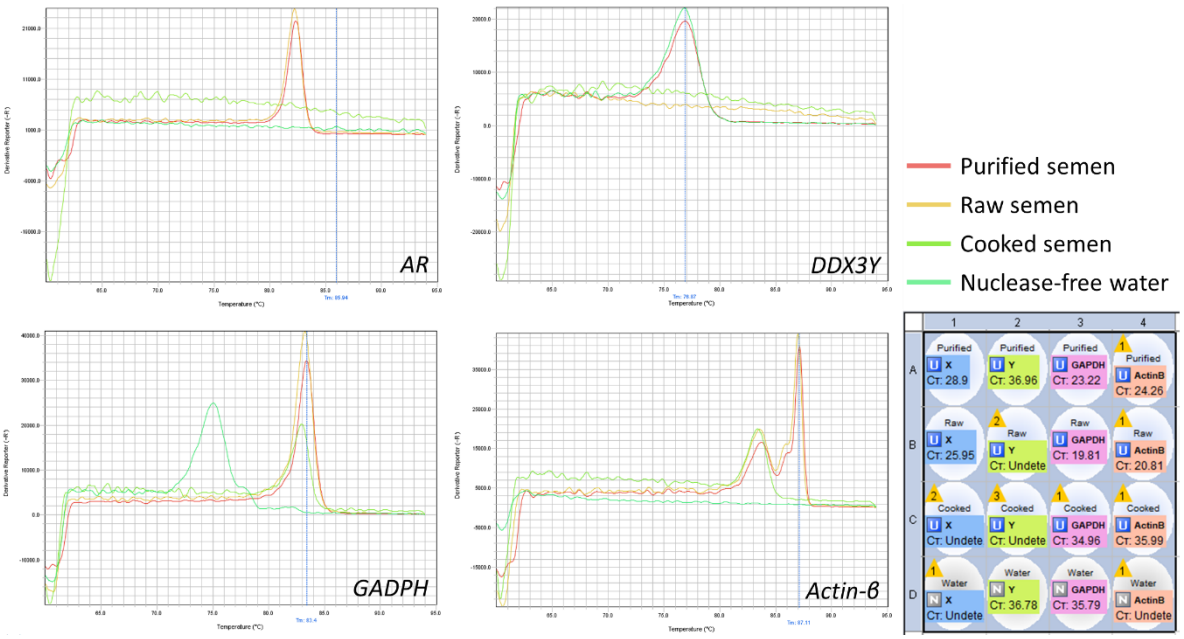


Figure 11. The melting curves of the four tested genes and the plate setup. For enlarged versions of these figures see Appendix 7.7 (Figures 18-21).

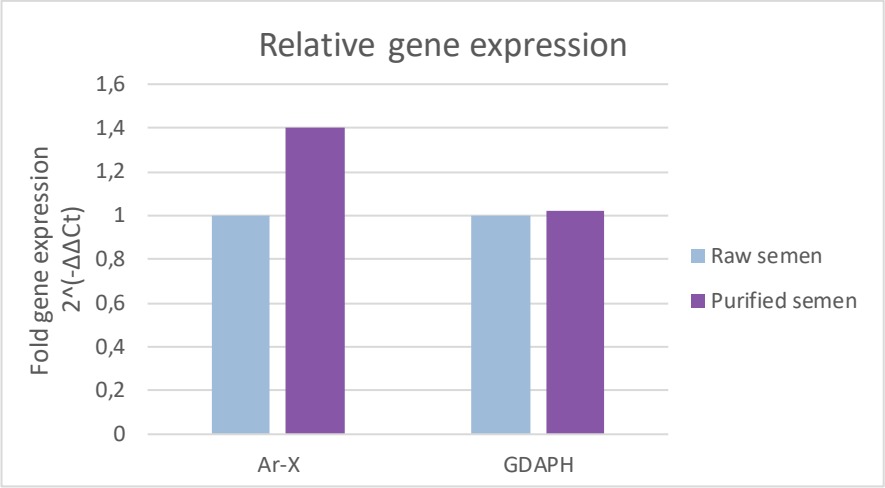


Figure 12. Relative gene expression comparing purified and raw semen.

4. Discussion

The aim of this study was to explore distorted sex ratios in horses, by identifying skewing individuals of the Icelandic horse and Standardbred horse studbooks, estimating heritability values, and investigating female-skewing families. Multiple horses with significantly more offspring of one sex were identified and in both Icelandic horses as Standardbreds and Coldblooded trotters the proportion of skewing individuals was between 6.4% - 16.0%. In both Icelandic horses and Standardbreds, female skewing occurred more than male skewing. The heritability for proportion of male offspring in Icelandic horses was estimated at 0.29, which became larger when the number of offspring of father and sons increased. In total, five female-skewing families were identified containing between 31-513 skewing individuals. Lastly, some extra pilot project work was done, including qPCR experimentation for future investigation of sex ratios in horse semen. Here we found that the *AR, GADPH,* and *Actin-β* primers worked on purified and raw semen and that the relative gene expression between raw and purified semen was not much different.

4.1. Datasets

For this study, data was drawn from the Icelandic horse studbook, World Fengur (WF) and the Swedish trotter association. Especially the Icelandic horse dataset was extensive, with a large amount of information available of each horse. The Icelandic horse dataset included horses from many countries, born over a broad period of time. In the earlier years, only a small number of horses were registered each year. The sex ratio of all registered individuals skewed towards less males (43.7% males), which has a smaller proportion of males compared to earlier research (Craft, 1938; Todd et al., 2020). However, when looking at sex ratios of every ten years, starting from 1950 the sex ratio proceeded more towards an even distribution over the course of time. In Coldblooded trotters and Standardbreds the sex ratio of registered offspring was as well more similar to sex ratios found in literature (50.5% and 48.9% males respectively), as was the results for Icelandic horses for the last studied time period (49.0% males).

Distorted sex ratios were found to be more common in stallions compared to mares. In Icelandic horse mares, only 5.8% of the horses was skewing, while in stallions this proportion was 16.0%. When comparing the different breeds, the proportion of significantly skewing stallions was higher in Icelandic horses compared to Coldblooded trotters and Standardbreds. However, the proportion of skewing Icelandic horses seems to be less in recent years compared to the past: when looking only from 2000 – 2016, only 8.5% of the stallions where skewing, with 5.8% towards females and 2.8% towards males, which is closer to the proportions of horses with skewing offspring in Coldblooded trotters and Standardbreds. When looking at the distribution in proportion of skewing horses, a difference was seen between Icelandic horses and Standardbreds (Figures 8-9). In Icelandic horses the skewing seems to be more extreme than in Standardbreds. This may be explained by the larger number of offspring in Standardbreds compared to Icelandic horses. Not only larger number of offspring may have influenced the proportion of male offspring in skewing horses, also genetic differences between the breeds may play a role. The different genetics in breeds may also explain why Coldblooded trotters had a larger proportion male-skewing horses compared to female-skewing horses. However, it should be noted that the sample size of Coldblooded trotters might have been too small.

In the Icelandic horse studbook there are several reasons which can cause a bias in the overall or individual sex ratios. These biases include horses with an unknown sex or misregistration which can among others be caused by Icelandic horses from farms producing horse meat. In WF some horses are registered with an unknown sex (0.6%). These horses do not have any offspring, and are often dead. Most horses with an unknown sex are born in herds used for bloodmaring or medication production. Bloodmaring is an industry collecting blood from pregnant horses for the synthesis of the equine hormone chorionic gonadotropin (eCG). This hormone is used to induce early reproduction in a variety of livestock. Mares used for bloodmaring are impregnated for this purpose and the foals born are slaughtered after summer (Lally, 2018). Fathers of these foals are often young and breeders do not want their breeding values to be influenced. To prevent the breeding value to be influenced the foals are registered with an unknown sex, so they will not be included in Best Linear Unbiased Prediction (BLUP) calculations. Since there is no preference of any sex in bloodmaring or production of medication, the actual sex is probably random (Personal communication Dr. Elsa Albertsdóttir, 2021).

Iceland is one of the greatest horse meat consumer and producer within the European Union and it produces around 2.19 kg per capita per year (Balji et al., 2020; Belaunzaran et al., 2015). Only very few farms keep Icelandic horses solely for meat production, however, slaughtering horses is seen as a part of the breeding process, and between 4000 - 6000 foals and 3000 - 5000 adult horses are slaughtered each year. Many farmers have herds of feral horses, and gather slaughter animals at the end of the autumn (Thorkelsson, 2000). As the herds are wild, offspring of meat horses may not be completely registered, excluding mostly male offspring. Male offspring may more often be excluded from registration because they are slaughtered at a young age, while some females may continue as a new member of the herd (Personal communication Dr. Elsa Albertsdóttir, 2021). The incomplete registration can explain skewing of sex ratios, as well as the high percentage of significantly skewing horses. Unfortunately, there was no time and means available to correct the results for missing registrations caused by the horse meat industry in Iceland. In future studies, it would be very important to investigate this matter in more detail and exclude horses used for the meat industry.

Missing or incorrect data is always possible in datasets, which includes all studbooks. The data included in this study from the Coldblooded trotters and Standardbreds is quite resent which may indicate that it is more reliable than older data. Especially at the start of the Icelandic studbook many horses are missing or not registered. Also digitalization of old records may have caused some errors. One of the errors found in the Icelandic studbook included 3 mares which were their own mother. However, most of the older horses were not used for any analyses since they did not have eight or more offspring. Another expected missing values are stillborn foals, which may not be totally included in the studbook. In the Icelandic studbook only 357 stillborn foals were reported, which should be much higher since in Thoroughbreds an incidence of a stillborn foals was 1.4% (95% CI 1.1 – 1.9) (Roach et al., 2020). In the Standardbreds and Coldblooded trotter it is as well expected that there is a misregistration of stillborn foals, although not proven. In humans the sex ratio of stillborn babies skew more towards males, which seems as well the case in other species, including buffalos, cows, and elephants (Ghavi Hossein-Zadeh et al., 2008, 2012; Jakobovits et al., 1987; Saragusty et al., 2009). For so far there is no evidence found that the sex ratio in stillborn foals is deviating.

Correcting for multiple testing was performed because of the extensive number of tests used during the data analyses. However, by doing so many results were discarded, including all females and male skewing horses. In case of female horses, it was simply not possible to maintain a significant *p*-value after correction. For a Bonferroni correction for 13 703 tests, a significant test result requires at least 26 offspring (Appendix 7.3, Table 13). In our dataset the maximum number of offspring of a mare was 21, which makes it impossible for a female to have a significant test result after correction for multiple testing. For this research the Bonferroni correction may be too strict, so an improvement for further studies could be the use of another method to correct for multiple testing.

Another way to correct for multiple testing is to control for false discovery rate. In this method a threshold is selected using a funnel plot and the threshold will be adjusted depending on the number extreme outcomes (Jones et al., 2008). Or the use of post-hoc power calculations could be used to select reliable test results. The power can be interpreted as the probability to reject the alternative hypotheses and in multiple testing (Zehetmayer & Posch, 2010). The post-hoc power was calculated for this study, and we may argue only to include horses which show a sufficient power. For a power higher than 0.8 a minimum of 12 offspring is needed, which makes it possible for female horses to maintain significant (Appendix 7.3, Table 13).

For this study a threshold of eight or more offspring was chosen as inclusion criteria of Icelandic horses. However, in retrospective this number is quite small to base all analyses on. In total 1404 males and 794 females were skewing towards one direction. In females this number would be more justified, since females are not able to produce that many offspring and their median number of offspring was 10. Although increasing the number to 10 or more offspring would still give 381 skewing females. However, in our case most analyses were focussed on males. Males had many more offspring (mean: 44.5, median: 18) and also more skewing animals. For this reason 20 offspring (779 significantly skewing horses) or even more would probably have been more appropriate. In future this would be a better inclusion criteria.

4.2. Heritability

The heritability of proportion of male offspring in Icelandic horses was estimated at 0.29 (SE: 0.08) , which means that around 29% of the variation is determined by additive genetic variance (Pierce, 2017). A heritability of 0.56 (SE: 0.12) was estimated based on the proportion of male offspring in maternal grandfathers and their grandson. So far, heritability of distorted sex ratios is not well investigated. In humans, the heritability of sex ratios at birth was estimated at 0.00058, which was not considered heritable (Orzack & Hardy, 2021; Zietsch et al., 2020). Other species in which heritability of sex ratios is investigated are the song sparrow (*Melospiza melodia*) and the *Drosophila mediopunctata*. The sparrow had a heritability estimated at $6.03x10^{-5}$, which had a variance expected under a random Mendelian segregation (Postma et al., 2011). In the *D. mediopunctata* a sex chromosome drive causes progeny to bias towards female offspring, however the researchers argued that there might be an autosomal suppressor to correct the skewing sex ratios. Sex chromosome drive is expected to be sex linked, and the heritability was measured using both a maternal and paternal component concluding with a heritability of 0.41 (Varandas et al., 1997). It may be possible that there are similarities in the mechanisms, however, we need to keep in mind that *Drosophilas* and horses are very different animals.

Striking in our analyses is the increasing heritability when the number of offspring in father and sons increase, from 0.29 up to 0.90. Increasing number of offspring decreases the effect of chance. For example, in case of eight offspring the effect of one offspring on the proportion of males is much larger compared to the effect of one offspring with 49 siblings (1/8 compared to 1/50). Estimating a heritability of 0.90 is very high, however we need to keep in mind that the number of skewing horses in this group is very small and this group only contains a very small part of the starting population. The small number of skewing horses causes the slope of the regression line to be formed by the few present extreme horses. Reasons why there are not that many skewing horses in this subgroup might be because sex ratio levels out to the true sex distribution. It might be that extreme skewness appears more in horses with a low number of offspring compared to horses with a high number of offspring. Possibly skewing horses in general do not have that many offspring, for example because of fertility issues which might be present with a selfish genetic element (Helleu et al., 2015).

During this study a parent-offspring regression was used to estimate the heritability, using both fathers and maternal grandfathers. In case of the fathers-son regression, the regression coefficient was multiplied by two and the maternal grandfather-grandson model the heritability was multiplied by four. Possibly this method of multiplication might not be ideal since there is a possibility for the heritability to become larger than one. Especially with the multiplication of four in a grandparent-offspring regression. Probably, another mathematical method should be developed to estimate heritability for more distant relatives, including grandparents.

However, it could also be that the parent-offspring regression has not been the preferable method to estimate the heritability in this situation. The parent-offspring regression is a very basic model to estimate heritability and many variables are not included in the analyses. For example, a parent-offspring regression does only take one variable into account. Measurements of parents and offspring are mostly determined over different time periods, which may influence the outcome (Behera & Paul, 2007). This could also have been the case in our calculations, since fathers and sons probably had their offspring over different time periods, causing different environmental conditions. Since the Trivers-Willard hypothesis applies to horses, it is possible that in different breeding seasons conditions were poorer than in other years (Cameron et al., 1999). This makes it possible that the sex ratios of the offspring were influenced by time periods stallions were breeding. Not only the time period may have influenced the conditions of mares, but as well locations. The Icelandic horses in the studbook live in 31 different countries, which have different weather conditions and cultural differences in human-horse interactions (van Dierendonck & Goodwin, 2005). Lastly parental age may have played a role in the sex ratios. Although stallions age does not influence the sex ratio of the offspring, mares do in that older mares have more female offspring (Santos et al., 2015). Variables including ages of the horses, years of births, and countries of origin were all known but were not incorporated in the present analyses. This causes our analyses to lack an abundance of information which could have been included.

One major way to improve the current analyses would be to estimate the heritability using another method. The method used to estimate heritability of sex ratios in humans was a logistic regression, in which sex of an individual's children was compared to the sex of the individuals siblings' children, followed by tetrachoric correlations to measure agreements of the binary data (Zietsch et al., 2020). Although this methods only included two variables in the logistic regression, it might be possible to add more. However, there are aspects which makes it more difficult to adjust this model to horses. Firstly, horses have mainly half brothers and sisters. In the case of humans the heritability was calculated as the tetrachoric correlation multiplied by two, since full-siblings share 50% of their co-segregating alleles (Zietsch et al., 2020). Half siblings only share 25% of their alleles, which may make the analysis less reliable. However, it might be possible to perform the analyses on halfsiblings by multiplying the tetrachoric correlation by four. A second problem appearing is the difference in number of offspring between individuals. The analyses were performed on pairs of cousins, resulting some cousins to appear multiple times in the analyses. Humans in Sweden had a total fertility rate (average number of children per women over her reproductive life) between 1.66 – 2.17 in 1960 – 2020, while stallions can have over 1000 offspring (SCB, 2021). When a stallion with 1000 offspring would have a sibling with only one offspring, the single offspring will occur 1000 times in the analyses while the 1000 offspring would only occur once individually. The difference of occurrence of one offspring may bias the results. In humans this difference in occurrence would be much smaller since large number of children do not occur that often.

Another method to improve the heritability calculations would be to use of a generalized linear mixed model analyses, which was used to investigate the heritability of sex ratios in sparrows. The mixed model used was an animal model which uses a pedigree to assess the additive genetic variance components (Postma et al., 2011). The animal model has several advantages over a parent-offspring regression. Advantages include that an animal model takes multiple relations between individuals into account, which maximizes the statistical power, it is more robust for inbreeding and selection, and the model can account for confounding effects. Despite these advantages, without any specified environmental effects the parent-offspring regression performs better than the animal model (de Villemereuil et al., 2013). In our case however, the extensive pedigree was available, including much more extra information, for this reason an animal model might have been better analysis to estimate the heritability. It is possible, however that the data structure with rather low number of offspring would have created problems using such a method for this type of trait.

Our current heritability analyses can be improved substantially, and might at the moment be a very crude estimate from the real heritability, which indicates only that there is a genetic component causing skewing sex ratios. A different method to estimate heritability could improve the quality of our heritability analyses and make the heritability more reliable.

4.3. Families

The goal of this analysis was to identify families of horses with skewing sex ratios, to exclude the chance of horses with different genetic variants or mutations causing skewing sex ratios in future gene mapping experiments. Different thresholds were tried, to estimate for which threshold the most complete and reliable families would be identified. During the pedigree analyses five families were identified, ranging from 27 to 429 significantly skewing horses. With easing the thresholds families became larger and the overlap between families increased. With the last threshold including all male progeny, the number of significant skewing individuals ranged from 46 to 875 in the families.

The Icelandic studbook goes back over hundred years, and the families contain between 8 - 14 generations, which makes overlap between the families is unavoidable. Already at the most strict threshold including only significant skewing progeny 36.4% of the horses were assigned to multiple families. Although the number of significant skewing horses increased, the proportion of number of horses appearing only in one family decreased. The threshold with the most unique individuals included horses with over 80% of their sons or maternal grandsons skewing towards more female offspring. This threshold would be suggested as preferable for future family analyses.

The most eased threshold including all male progeny, was not sufficient to identify horses with similar genetic variances, since it increased the overlap between families drastically. Also looking into skewing sex ratios, families consisting all male progeny have many horses without a disrupted sex ratio. The many non-significant skewing horses on top of the overlap between different families, may cause genetic variance to be traced back to an incorrect source.

When looking for a sex chromosome drive, one would expect the driving/(selfish genetic) element to be present on the X or Y chromosome (Helleu et al., 2015). In case of a Y-linked inheritance pattern one would expect that 100% of the sons of a skewing stallion would skew as well, while if the trait would be X-linked, we would expect 100% of the daughters to carry the trait, and 50% of their maternal grandsons (Shetty, 2018). In our families we see neither a Y-linked or an X-linked inheritance pattern. The proportion of significant skewing sons was on average 32% and the proportion of significant skewing maternal grandsons was on average 15%. These proportions correspond more to an autosomal or multifactorial trait and may as well support an autosomal suppression of a sex chromosome drive, as suggested above (Varandas et al., 1997). This means genetic cause of sex skewing in horses might be very complex and may be influenced by non-genetic factors.

During the analyses we started with the 100 most significantly skewing horses. We assigned many horses to a family and a total 788 individuals with significantly skewing offspring were identified. In the whole Icelandic studbook 1218 stallions are significantly skewing towards more female offspring. The difference between the identified significantly skewing horses and the skewing horses in the studbook shows that 430 skewing horses were not assigned to a family. It is possible that those horses are single cases or do only have a very small family, similar to the six unassigned horses in our original 100. However, it may as well be possible that there are more families present in the dataset, still unfound.

4.4. qPCR pilot experimentation

The qPCR was performed to test four primers and compare different prepared horse semen. Three of four primers were considered successful because of clean melting curves. The *DDX3Y* primer was however not successful. The fairly high Ct values of *DDX3Y* were only detected in purified semen and in the water control. Reasons for amplification in water could be contamination or forming of primer dimers. Since the peaks of *DDX3Y* in purified semen and water are at the same temperature and around the same Ct value, the *DDX3Y* primers cannot be trusted for purified semen. One of the reasons for malfunctioning could be an insufficient annealing temperature. During our experiments, an annealing temperature of 58°C was used, however, the optimal annealing temperature was not calculated. For this reason, we could improve the experiment by calculating the optimal annealing temperature.

Of the three conditions tested, both purified semen and raw semen were successfully amplified. Cooked semen did only detect the *GADPH* primer correctly, however with a higher Ct value. For *Actin-β*, a Ct value was measured in cooked semen as well. Looking at the melting curves, it only appeared at the first peak, while purified and raw semen included both peaks. With this analysis we can conclude that DNA extraction with cooked semen was not sufficient.

The relative gene copy number between X and Y could not be calculated because of the malfunction of *DDX3Y* primer. The relative gene copy number of *AR* and *GDAPH* was not much different between raw and purified semen, which is what one would expect, since the same semen was used for both analyses. Nonetheless, we need to keep in mind that when investigating sex ratios, the change between X and Y copy numbers may be small and not noticeable. Since the Y primer did not work sufficiently it was not possible to investigate sex ratios during this experiment. For this reason, we should continue the analysis with purified semen until it is certain that raw semen can detect distorted sex ratios.

Information of this pilot will eventually be used to investigate sex ratio in horse semen, including three groups, female-skewing horses, male-skewing horses and non-skewing horses. If distorted sex ratios are caused by a meiotic drive system it might be possible that semen of one sex is impaired like in *Slx/Slxl1* and *Sly mice* (Cocquet et al., 2012). For so far sex ratios in horse semen is, to the current knowledge, not investigated, which makes it a very interesting addition for future studies.

There are some animals were sex ratios in semen compared to offspring are investigated. An example is the captive population of the pygmy hippopotamus (*Choeropsis liberiensis*). In these hippopotamus the sex ratio of offspring is skewing towards more females (proportion male offspring: 0.43). When investigating the sex ratio of semen a similar proportion of Y-bearing semen was found, 0.43 ±0.0094. Reasons and causes for the skewing are so far unknown, however, it is theorized that if the habitat is saturated with occupied territories, sons will likely compete with their fathers for territory and mating success. For this reason it is in fathers interest to produce daughters, which increase their reproductive success and avoid competition with their sons. In captivity the solitary hippopotamus live often in the same enclosure, which may cause a sense of population density, resulting in the skewing sex ratios. For so far there is no data available on wild populations. However the proportion of wild-caught hippopotamuses was equal to 0.5 (Saragusty et al., 2012).

Matching sex ratios between offspring and semen is however not always the case. In humans the sex ratio of birth is around 51.4% males, while the sex ratio in semen is reverse with 52.0% X-bearing semen (Chadhary et al., 2014). Also contradicting results, or indifferent sex ratios between skewing horses would be very interesting to investigate, since it will assume that there are other mechanisms in place, which causes skewing sex ratios.

5. Conclusions and Future implications

- Distorted sex ratios were proportionally more common in stallions compared to mares in Icelandic horses.
- Female-skewing stallions were more common than male-skewing stallions in Icelandic horses and Standardbreds in our datasets.
- The heritability for the proportion of male offspring was estimated at 0.29, which indicate a genetic influence.
- The heritability for the proportion of male offspring increased when fathers and sons had more offspring.
- Five families with distorting sex ratios were identified.
- Sex ratio analyses of horse semen were tested, using qPCR, however, some improvements are needed for the Y chromosome assay.

Meat horses in Iceland may have influenced the analyses drastically, however, it is not clear to which extent. For this reason, I would recommend to exclude these horses and repeat the analyses. However, identification of these horses may be very hard. Although meat horses may influence the results there are strong indications for familial sex ratio distortion, since sex ratio distortion was present in Standardbreds and Coldblooded trotters as well. To improve the foundation for this claim, the heritability estimations should be repeated with a better fitting model, taking the whole pedigree into account and including environmental factors. A next step would be to increase the number of Standardbreds and Coldblooded trotters and investigate their families as well. It may also be interesting to investigate additional horse breeds, especially breeds not affected by meat production. At the moment only Icelandic horses, Coldblooded trotters, and Standardbred horses were investigated, but doing similar analyses in more breeds would improve this research.

In the future, we want to investigate sex ratios in horse semen, including female-skewing and male-skewing stallions. We would also like to perform whole genome sequencing on skewing individuals, to study the genetic cause, with the aim to identify genetic elements causing distorted sex ratios.

Acknowledgment

I had a very great time during my stay in Sweden and I very much enjoyed this research topic! During my stay I met so many interesting people, which made it an awesome time. For this time, I have many people to thank. First, of course, my very enthusiastic and nice supervisor Gabriella! You are great to work with and always very happy and motivating! Secondly, my supervisor Ole, who always supported my decisions and was very easy to talk to, with sometimes ideas I did not think about. Terje, I always loved the stories about hermaphrodite mares, and I was looking forward to see you ever week. I read many of your very interesting papers for this project. For my programming, I need to thank Rakan, for is very useful input, especially for helping me out with Rstudio. Elsa for the fantastic dataset, I was able to work with and Tytti for supervising and helping me in the lab.

Lastly, BioRender.com, which I partly used to design Figure 1 and 2.

6. References

Aickin, M., & Gensler, H. (1996). Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm methods. *American Journal of Public Health*, *86*(5). https://doi.org/10.2105/AJPH.86.5.726

Akera, T., Trimm, E., & Lampson, M. A. (2019). Molecular Strategies of Meiotic Cheating by Selfish Centromeres. *Cell*, *178*(5). https://doi.org/10.1016/j.cell.2019.07.001

Aldrich, J. C., Leibholz, A., Cheema, M. S., Ausió, J., & Ferree, P. M. (2017). A "selfish" B chromosome induces genome elimination by disrupting the histone code in the jewel wasp Nasonia vitripennis. *Scientific Reports*, *7*. https://doi.org/10.1038/srep42551

Aurich, C., & Schneider, J. (2014). Sex determination in horses-Current status and future perspectives. In *Animal Reproduction Science* (Vol. 146, Issues 1–2). https://doi.org/10.1016/j.anireprosci.2014.01.014

Bachtrog, D. (2014). Signs of genomic battles in mouse sex chromosomes. In *Cell* (Vol. 159, Issue 4). https://doi.org/10.1016/j.cell.2014.10.036

Balji, Y., Knicky, M., & Zamaratskaia, G. (2020). Perspectives and safety of horsemeat consumption. In *International Journal of Food Science and Technology* (Vol. 55, Issue 3). https://doi.org/10.1111/ijfs.14390

Bauer, H., Willert, J., Koschorz, B., & Herrmann, B. G. (2005). The t complex-encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. *Nature Genetics*, *37*(9). https://doi.org/10.1038/ng1617

Beckelmann, J., Budik, S., Helmreich, M., Palm, F., Walter, I., & Aurich, C. (2013). Sex-dependent insulin like growth factor-1 expression in preattachment equine embryos. *Theriogenology*, *79*(1). https://doi.org/10.1016/j.theriogenology.2012.10.004

Behera, S. K., & Paul, A. K. (2007). *Estimation of heritability*.

Belaunzaran, X., Bessa, R. J. B., Lavín, P., Mantecón, A. R., Kramer, J. K. G., & Aldai, N. (2015). Horse-meat for human consumption - Current research and future opportunities. In *Meat Science* (Vol. 108). https://doi.org/10.1016/j.meatsci.2015.05.006

Benetta, E. D., Antoshechkin, I., Yang, T., My Nguyen, H. Q., Ferree, P. M., & Akbari, O. S. (2020). Genome elimination mediated by gene expression from a selfish chromosome. *Science Advances*, *6*(14). https://doi.org/10.1126/sciadv.aaz9808

Bogaert, L., Van Poucke, M., De Baere, C., Peelman, L., Gasthuys, F., & Martens, A. (2006). Selection of a set of reliable reference genes for quantitative real-time PCR in normal equine skin and in equine sarcoids. *BMC Biotechnology*, *6*. https://doi.org/10.1186/1472-6750-6-24

Braidotti, G., & Barlow, D. P. (1997). Identification of a male meiosis-specific gene, Tcte2, which is differentially spliced in species that form sterile hybrids with laboratory mice and deleted in t chromosomes showing meiotic drive. *Developmental Biology*, *186*(1). https://doi.org/10.1006/dbio.1997.8574

Brashear, W. A., Raudsepp, T., & Murphy, W. J. (2018). Evolutionary conservation of Y Chromosome ampliconic gene families despite extensive structural variation. *Genome Research*, *28*(12). https://doi.org/10.1101/gr.237586.118

Brauner, R., Picard-Dieval, F., Lottmann, H., Rouget, S., Bignon-Topalovic, J., Bashamboo, A., & McElreavey, K. (2016). Familial forms of disorders of sex development may be common if infertility is considered a comorbidity. *BMC Pediatrics*, *16*(1). https://doi.org/10.1186/s12887-016-0737-0

Bravo Núñez, M. A., Nuckolls, N. L., & Zanders, S. E. (2018). Genetic Villains: Killer Meiotic Drivers. In *Trends in Genetics* (Vol. 34, Issue 6). https://doi.org/10.1016/j.tig.2018.02.003

Cameron, E. Z., Linklater, W. L., Stafford, K. J., & Veltman, C. J. (1999). Birth sex ratios relate to mare condition at conception in Kaimanawa horses. *Behavioral Ecology*, *10*(5). https://doi.org/10.1093/beheco/10.5.472

Catalano, R. A. (2003). Sex ratios in the two Germanies: A test of the economic stress hypothesis. *Human Reproduction*, *18*(9). https://doi.org/10.1093/humrep/deg370

Catalano, R., Bruckner, T., Hartig, T., & Ong, M. (2005). Population stress and the Swedish sex ratio. *Paediatric and Perinatal Epidemiology*, *19*(6). https://doi.org/10.1111/j.1365-3016.2005.00677.x

Chadhary, I., Jain, M., & Halder, A. (2014). Sperm Sex Ratio (X: Y Ratio) and its Variations. *Austin J. Reprod. Med. Infertil*, *1*(1).

Cocquet, J., Ellis, P. J. I., Mahadevaiah, S. K., Affara, N. A., Vaiman, D., & Burgoyne, P. S. (2012). A Genetic Basis for a Postmeiotic X Versus Y Chromosome Intragenomic Conflict in the Mouse. *PLoS Genetics*, *8*(9). https://doi.org/10.1371/journal.pgen.1002900

Comptour, A., Moretti, C., Serrentino, M. E., Auer, J., Ialy-Radio, C., Ward, M. A., Touré, A., Vaiman, D., & Cocquet, J. (2014). SSTY proteins co-localize with the post-meiotic sex chromatin and interact with regulators of its expression. *FEBS Journal*, *281*(6). https://doi.org/10.1111/febs.12724

Courret, C., Chang, C. H., Wei, K. H. C., Montchamp-Moreau, C., & Larracuente, A. M. (2019). Meiotic drive mechanisms: Lessons from Drosophila. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1913). https://doi.org/10.1098/rspb.2019.1430

Craft, W. A. (1938). The Sex Ratio in Mules and Other Hybrid Mammals. *The Quarterly Review of Biology*, *13*(1). https://doi.org/10.1086/394547

Darwin, C. (1871). The Descent of Man, and Selection in relation to Sex. *The Descent of Man, and Selection in Relation to Sex*.

de Villemereuil, P., Gimenez, O., & Doligez, B. (2013). Comparing parent-offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: A simulation study for Gaussian and binary traits. *Methods in Ecology and Evolution*, *4*(3). https://doi.org/10.1111/2041-210X.12011

Ellison, C., Leonard, C., Landeen, E., Gibilisco, L., Phadnis, N., & Bachtrog, D. (2018). Rampant cryptic sex chromosome drive in Drosophila. *BioRxiv*.

Ellison, Christopher, & Bachtrog, D. (2019). Recurrent gene co-amplification on Drosophila X and Y chromosomes. *PLoS Genetics*, *15*(7). https://doi.org/10.1371/journal.pgen.1008251

Fredga, K., Gropp, A., Winking, H., & Frank, F. (1977). A hypothesis explaining the exceptional sex ratio in the wood lemming (Myopus schisticolor). *Hereditas*, *85*(1). https://doi.org/10.1111/j.1601-5223.1977.tb00956.x

Gellatly, C. (2019). The global male-bias in sex ratio at birth is sustained by the sex ratio genotypes of replacement offspring. *Genetica*, *147*(3–4). https://doi.org/10.1007/s10709-019-00074-2

Gershenson, S. (1928). A NEW SEX-RATIO ABNORMALITY IN DROSOPHILA OBSCURA . *Genetics*, *13*(6). https://doi.org/10.1093/genetics/13.6.488

Ghavi Hossein-Zadeh, N., Madad, M., Shadparvar, A. A., & Kianzad, D. (2012). An observational analysis of secondary sex ratio, stillbirth and birth weight in iranian

buffaloes (bubalus bubalis). *Journal of Agricultural Science and Technology*, *14*(SUPPL.).

Ghavi Hossein-Zadeh, N., Nejati-Javaremi, A., Miraei-Ashtiani, S. R., & Kohram, H. (2008). An observational analysis of twin births, calf stillbirth, calf sex ratio, and abortion in Iranian Holsteins. *Journal of Dairy Science*, *91*(11). https://doi.org/10.3168/jds.2008-1079

Good, J. M. (2012). The Conflict within and the Escalating War between the Sex Chromosomes. *PLoS Genetics*, *8*(9). https://doi.org/10.1371/journal.pgen.1002955

Haldane, J. B. S. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, *12*(2). https://doi.org/10.1007/BF02983075

Helleu, Q., GÉrard, P. R., & Montchamp-Moreau, C. (2015). Sex chromosome drive. *Cold Spring Harbor Perspectives in Biology*, *7*(2). https://doi.org/10.1101/cshperspect.a017616

Hreidarsdóttir, G. E., árnason, T., Svansson, V., & Hallsson, J. H. (2014). Analysis of the history and population structure of the Icelandic horse using pedigree data and DNA analyses. *Icelandic Agricultural Sciences*, *27*(1).

Hughes, J. F., Skaletsky, H., Pyntikova, T., Koutseva, N., Raudsepp, T., Brown, L. G., Bellott, D. W., Cho, T.-J., Dugan-Rocha, S., Khan, Z., Kremitzki, C., Fronick, C., Graves-Lindsay, T. A., Fulton, L., Warren, W. C., Wilson, R. K., Owens, E., Womack, J. E., Murphy, W. J., … Page, D. C. (2020). Sequence analysis in Bos taurus reveals pervasiveness of X–Y arms races in mammalian lineages. *Genome Research*, *30*(12), 1716–1726. https://doi.org/10.1101/gr.269902.120

Hurst, L. D. (1996). Further evidence consistent with Stellate's involvement in meiotic drive. In *Genetics* (Vol. 142, Issue 2). https://doi.org/10.1093/genetics/142.2.641

Hurst, L. D., & Pomiankowski, A. (1991). Causes of sex ratio bias may account for unisexual sterility in hybrids: A new explanation of Haldane's rule and related phenomena. *Genetics*, *128*(4). https://doi.org/10.1093/genetics/128.4.841

Jakobovits, A., Jakobovits, Á. A., & Viski, A. (1987). Sex ratio of the stillborn fetuses and neonates dying in the first week. *Early Human Development*, *15*(3). https://doi.org/10.1016/0378-3782(87)90001-6

Janečka, J. E., Davis, B. W., Ghosh, 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., Zadjali, A. M. Al, Love, C. C., Varner, D. D., Bellott, 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 Communications*, *9*(1). https://doi.org/10.1038/s41467-018-05290-6

Jetbrains. (2020). *Pycharm Community* (2020.3.2). www.jetbrains.com/pycharm

Johnsen, D. W. (1985). A fertile female mule. *Journal of Equine Veterinary Science*, *5*(2). https://doi.org/10.1016/S0737-0806(85)80054-X

Jones, H. E., Ohlssen, D. I., & Spiegelhalter, D. J. (2008). Use of the false discovery rate when comparing multiple health care providers. In *Journal of Clinical Epidemiology* (Vol. 61, Issue 3). https://doi.org/10.1016/j.jclinepi.2007.04.017

Kanippayoor, R. L., Alpern, J. H. M., & Moehring, A. J. (2013). Protamines and spermatogenesis in Drosophila and Homo sapiens . *Spermatogenesis*, *3*(2). https://doi.org/10.4161/spmg.24376

Kelemen, R. K., & Vicoso, B. (2018). Complex history and differentiation patterns of the t-haplotype, a mouse meiotic driver. *Genetics*, *208*(1). https://doi.org/10.1534/genetics.117.300513

Kholghi, M., Rostamzadeh, J., Razmkabir, M., & Heidari, F. (2020). Blood Testosterone Level Affects Sex Ratio of Bull Semen. *Concepts of Dairy and Veterinary Science*, *4*(1).

Kruger, A. N., Brogley, M. A., Huizinga, J. L., Kidd, J. M., de Rooij, D. G., Hu, Y. C., & Mueller, J. L. (2019). A Neofunctionalized X-Linked Ampliconic Gene Family Is Essential for Male Fertility and Equal Sex Ratio in Mice. *Current Biology*, *29*(21). https://doi.org/10.1016/j.cub.2019.08.057

Kruger, A. N., & Mueller, J. L. (2021). Mechanisms of meiotic drive in symmetric and asymmetric meiosis. In *Cellular and Molecular Life Sciences* (Vol. 78, Issue 7). https://doi.org/10.1007/s00018-020-03735-0

Lahn, B. T., & Page, D. C. (2000). A human sex-chromosomal gene family expressed in male germ cells and encoding variably charged proteins. *Human Molecular Genetics*, *9*(2). https://doi.org/10.1093/hmg/9.2.311

Lally, A. E. (2018). Circulatory [food] systems: Icelandic bloodmaring, reproductive technologies, and the porosity of female agricultural bodies. *Cuizine*, *9*(1). https://doi.org/10.7202/1052115ar

Leitschuh, C. M., Kanavy, D., Backus, G. A., Valdez, R. X., Serr, M., Pitts, E. A., Threadgill, D., & Godwin, J. (2018). Developing gene drive technologies to eradicate invasive rodents from islands. *Journal of Responsible Innovation*, *5*. https://doi.org/10.1080/23299460.2017.1365232

Lindholm, A. K., Dyer, K. A., Firman, R. C., Fishman, L., Forstmeier, W., Holman, L., Johannesson, H., Knief, U., Kokko, H., Larracuente, A. M., Manser, A., Montchamp-Moreau, C., Petrosyan, V. G., Pomiankowski, A., Presgraves, D. C., Safronova, L. D., Sutter, A., Unckless, R. L., Verspoor, R. L., … Price, T. A. R. (2016). The Ecology and Evolutionary Dynamics of Meiotic Drive. In *Trends in Ecology and Evolution* (Vol. 31, Issue 4). https://doi.org/10.1016/j.tree.2016.02.001

Liu, W. S., Nordqvist, K., Lau, Y. F. C., & Fredga, K. (2001). Characterization of the Xp21-23 region in the wood lemming, a region involved in XY sex reversal. *Journal of Experimental Zoology*, *290*(6). https://doi.org/10.1002/jez.1105

Lorange, J. B. (2011). WorldFengur - the studbook of origin for the Icelandic horse. *Acta Veterinaria Scandinavica*, *53*(SUPPL. 1). https://doi.org/10.1186/1751-0147-53-S1-S5

Lyon, M. F. (1984). Transmission ratio distortion in mouse t-haplotypes is due to multiple distorter genes acting on a responder locus. *Cell*, *37*(2). https://doi.org/10.1016/0092-8674(84)90393-3

Mäkinen, A., Suojala, L., Niini, J., Katila, T., Tozak, T., Miyake, Y. I., & Hasegawa, T. (2001). X chromosome detection in an XO mare using a human X paint probe, and PCR detection of SRY and amelogenin genes in 3 XY mares. *Equine Veterinary Journal*, *33*(5). https://doi.org/10.2746/042516401776254844

Malone, C. D., Lehmann, R., & Teixeira, F. K. (2015). The cellular basis of hybrid dysgenesis and Stellate regulation in Drosophila. In *Current Opinion in Genetics and Development* (Vol. 34). https://doi.org/10.1016/j.gde.2015.09.003

Monard, A. M., Duncan, P., Fritz, H., & Feh, C. (1997). Variations in the birth sex ratio and neonatal mortality in a natural herd of horses. *Behavioral Ecology and Sociobiology*, *41*(4). https://doi.org/10.1007/s002650050385

Montchamp-Moreau, C., Ogereau, D., Chaminade, N., Colard, A., & Aulard, S. (2006). Organization of the sex-ratio meiotic drive region in Drosophila simulans. *Genetics*, *174*(3). https://doi.org/10.1534/genetics.105.051755

Mori, A., Chadee, D. D., Graham, D. H., & Severson, D. W. (2004). Reinvestigation of an endogenous meiotic drive system in the mosquito, Aedes aegypti (Diptera: Culicidae). *Journal of Medical Entomology*, *41*(6). https://doi.org/10.1603/0022-2585-41.6.1027

Orzack, S. H., & Hardy, I. C. W. (2021). Does the lack of heritability of human sex ratios require a rethink of sex ratio theory? No: A Comment on Zietsch et al. 2020. In *Proceedings of the Royal Society B: Biological Sciences* (Vol. 288, Issue 1947). https://doi.org/10.1098/rspb.2020.2638

Paria, N. (2009). *Discovery of candidate genes for stallion fertility from the horse Y chromosome*. Texas A&M University.

Pierce, B. (2017). Quantitative genetics. In *Genetics a conceptual approach* (Sixth). W. H. Freeman and Company.

Postma, E., Heinrich, F., Koller, U., Sardell, R. J., Reid, J. M., Arcese, P., & Keller, L. F. (2011). Disentangling the effect of genes, the environment and chance on sex ratio variation in a wild bird population. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1720). https://doi.org/10.1098/rspb.2010.2763

Python Software Foundation. (2020). *Python v3.8.5*. http://www.python.org

QIAGEN. (2013). *Rapid extraction of DNA from sperm using the QIAsymphony® DNA Investigator® Kit*.

Rathje, C. C., Johnson, E. E. P., Drage, D., Patinioti, C., Silvestri, G., Affara, N. A., Ialy-Radio, C., Cocquet, J., Skinner, B. M., & Ellis, P. J. I. (2019). Differential Sperm Motility Mediates the Sex Ratio Drive Shaping Mouse Sex Chromosome Evolution. *Current Biology*, *29*(21). https://doi.org/10.1016/j.cub.2019.09.031

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*(SUPPL. 2). https://doi.org/10.1111/j.1365-2052.2010.02101.x

Raudsepp, Terje. (2020). Genetics of Equine Reproductive Diseases. In *Veterinary Clinics of North America - Equine Practice* (Vol. 36, Issue 2). https://doi.org/10.1016/j.cveq.2020.03.013

Raudsepp, Terje, Lee, E. J., Kata, S. R., Brinkmeyer, C., Mickelson, J. R., Skow, L. C., Womack, J. E., & Chowdhary, B. P. (2004). Exceptional conservation of horse-human gene order on X chromosome revealed by high-resolution radiation hybrid mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(8). https://doi.org/10.1073/pnas.0308513100

Resende, M. V., Lucio, A. C., Perini, A. P., Oliveira, L. Z., Almeida, A. O., Alves, B. C. A., Moreira-Filho, C. A., Santos, I. W., & Hossepian de Lima, V. F. M. (2011). Comparative validation using quantitative real-time PCR (qPCR) and conventional PCR of bovine semen centrifuged in continuous density gradient. *Arquivo Brasileiro de Medicina Veterinaria e Zootecnia*, *63*(3). https://doi.org/10.1590/S0102-09352011000300002

Roach, J. M., Foote, A. K., Smith, K. C., Verheyen, K. L., & de Mestre, A. M. (2020). Incidence and causes of pregnancy loss after Day 70 of gestation in Thoroughbreds. *Equine Veterinary Journal*. https://doi.org/10.1111/evj.13386

RStudio Team. (2021). *RStudio: Integrated Development Environment for R*. RStudio, PBC. http://www.rstudio.com/

Santos, M. M., Maia, L. L., Nobre, D. M., Oliveira Neto, J. F., Garcia, T. R., Lage, M. C. G. R., de Melo, M. I. V., Viana, W. S., Palhares, M. S., da Silva Filho, J. M., Santos, R. L., & Valle, G. R. (2015). Sex ratio of equine offspring is affected by the ages of the mare and stallion. *Theriogenology*, *84*(7).

https://doi.org/10.1016/j.theriogenology.2015.07.001
Saragusty, J., Hermes, R., Göritz, F., Schmitt, D. L., & Hildebrandt, T. B. (2009). Skewed birth sex ratio and premature mortality in elephants. *Animal Reproduction Science*, *115*(1–4). https://doi.org/10.1016/j.anireprosci.2008.10.019
Saragusty, J., Hermes, R., Hofer, H., Bouts, T., Göritz, F., & Hildebrandt, T. B. (2012). Male pygmy hippopotamus influence offspring sex ratio. *Nature Communications*, *3*. https://doi.org/10.1038/ncomms1700
SCB. (2021, March 18). *Summary of Population Statistics 1960–2020*.
Schauberger, P., & Walker, A. (2020). *openxlsx: Read, Write and Edit xlsx Files*. (R package 4.2.3). https://cran.r-project.org/package=openxlsx
Schimenti, J. (2000). Segregation distortion of mouse t haplotypes - The molecular basis emerges. In *Trends in Genetics* (Vol. 16, Issue 6). https://doi.org/10.1016/S0168-9525(00)02020-5
Shetty, N. K. (2018). Inheritance of Chromosomes, Sex Determination, and the Human Genome. *Gender and the Genome*, *2*(1). https://doi.org/10.1177/2470289718787131
Soh, Y. Q. S., Alföldi, 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., … Page, D. C. (2014). Sequencing the mouse y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell*, *159*(4). https://doi.org/10.1016/j.cell.2014.09.052
Sportapp Svensk Travsport. (2021, February). *https://sportapp.travsport.se/*.
Squires, E. L. (2019). Perspectives on the development and incorporation of assisted reproduction in the equine industry. *Reproduction, Fertility and Development*, *31*(12). https://doi.org/10.1071/RD19365
Su, X., Zhu, G., Ding, X., Lee, S. Y., Dou, Y., Zhu, B., Wu, W., & Li, H. (2014). Molecular basis underlying histone H3 lysine-arginine methylation pattern readout by Spin/Ssty repeats of Spindlin1. *Genes and Development*, *28*(6). https://doi.org/10.1101/gad.233239.113
Thorkelsson, G. (2000). Meat Producers LTD Information on Icelandic horse meat. In *Icelandic Fisheries Laboratories and University of Iceland*.
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 ratio in Australian thoroughbred horses. *BMC Genetics*, *21*(1). https://doi.org/10.1186/s12863-020-00847-1
Trivers, R. L., & Willard, D. E. (1973). Natural selection of parental ability to vary the sex ratio of offspring. *Science*, *179*(4068). https://doi.org/10.1126/science.179.4068.90
Uller, T., Pen, I., Wapstra, E., Beukeboom, L. W., & Komdeur, J. (2007). The evolution of sex ratios and sex-determining systems. *Trends in Ecology and Evolution*, *22*(6). https://doi.org/10.1016/j.tree.2007.03.008
van Dierendonck, M., & Goodwin, D. (2005). Social contact in horses: implications for human-horse interactions. *The Human-Animal Relationship*.
Varandas, F. R., Sampaio, M. C., & Carvalho, A. B. (1997). Heritability of sexual proportion in experimental sex-ratio populations of Drosophila mediopunctata. *Heredity*, *79*(1). https://doi.org/10.1038/hdy.1997.128
Villagómez, D. A. F., Lear, T. L., Chenier, T., Lee, S., McGee, R. B., Cahill, J., Foster, R. A., Reyes, E., St John, E., & King, W. A. (2011). Equine disorders of sexual development in 17 mares including XX, SRY-negative, XY, SRY-negative and XY,

SRY-positive genotypes. *Sexual Development*, *5*(1). https://doi.org/10.1159/000322811

Vuorinen, J. A., & Eskelinen, O. (2005). Long-term stability of allozyme frequencies in a wood lemming, Myopus schisticolor, population with a biased sex ratio and density fluctuations. *Heredity*, *94*(4). https://doi.org/10.1038/sj.hdy.6800639

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-verlag.

WorldFengur. (2021, February). *https://www.worldfengur.com/*.

Zaiontz, C. (2021). *the Real Statistics Resource Pack software* (Release 7.6). www.real-statistics.com

Zanders, S. E., & Unckless, R. L. (2019). Fertility Costs of Meiotic Drivers. In *Current Biology* (Vol. 29, Issue 11). https://doi.org/10.1016/j.cub.2019.03.046

Zehetmayer, S., & Posch, M. (2010). Post hoc power estimation in large-scale multiple testing problems. *Bioinformatics*, *26*(8). https://doi.org/10.1093/bioinformatics/btq085

Zietsch, B. P., Walum, H., Lichtenstein, P., Verweij, K. J. H., & Kuja-Halkola, R. (2020). No genetic contribution to variation in human offspring sex ratio: A total population study of 4.7 million births. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1921). https://doi.org/10.1098/rspb.2019.2849

7. Appendix

7.1. Extended data characteristics

Icelandic horses were born in 31 different countries, Coldblooded trotters in 2 and Standardbreds in 10. In Table the different countries and distributions are shown.

Table 11a. The country of birth and sex ratio of Icelandic horses.

Country	Icelandic horses	(%)	Males	(%)	Females	(%)	Country	Icelandic horses	(%)	Males	(%)	Females	(%)
AT	5442	(1.1)	2345	(43.1)	3097	(56.9)	IS	331 806	(65.2)	139 149	(41.9)	189 444	(57.1)
AU	304	(0.1)	151	(49.7)	153	(50.3)	IT	302	(0.1)	152	(50.3)	150	(49.7)
BE	1619	(0.3)	808	(49.9)	811	(50.1)	LI	3	(0.0)	3		0	
CA	2042	(0.4)	1013	(49.6)	1029	(50.4)	LT	3	(0.0)	1	(33.3)	2	(66.7)
CH	1766	(0.3)	825	(46.7)	941	(53.3)	LU	189	(0.0)	85	(45.0)	104	(55.0)
CZ	2	(0.0)	1		1		NL	9260	(1.8)	4645	(50.2)	4615	(49.8)
DE	54 996	(10.8)	23534	(42.8)	31462	(57.2)	NO	12 469	(2.4)	6170	(49.5)	6299	(50.5)
DK	49 107	(9.6)	23760	(48.4)	25322	(51.6)	NZ	159	(0.0)	72	(45.3)	87	(54.7)
EE	2	(0.0)	1		1		PL	320	(0.1)	166	(51.9)	154	(48.1)
ES	2	(0.0)	1		1		PT	3	(0.0)	0		3	
FI	2384	(0.5)	1160	(48.7)	1224	(51.3)	RO	33	(0.0)	20	(60.6)	13	(39.4)
FO	96	(0.0)	50	(52.1)	46	(47.9)	SE	27 593	(5.4)	13 901	(50.4)	13 692	(49.6)
FR	4467	(0.9)	2122	(47.5)	2345	(52.5)	SI	323	(0.1)	138	(42.7)	185	(57.3)
GB	791	(0.2)	392	(49.6)	399	(50.4)	US	3439	(0.7)	1651	(48.0)	1788	(52.0)
HU	27	(0.0)	14	(51.9)	13	(48.1)	Unknown	3	(0.0)	0		3	
IE	56	(0.0)	26	(46.4)	30	(53.6)							

Total Icelandic horses: 509 008

Table 12b. The country of birth and sex ratio of Standardbreds and Coldblooded trotters.

Country	Standardbreds (%)	Males offspring (%)	Female offspring (%)
CA	15 (2.96)	665 (46.6)	762 (53.4)
DE	9 (1.78)	501 (51.2)	477 (48.8)
DK	3 (0.59)	71 (45.5)	85 (54.5)
FI	4 (0.79)	509 (51.3)	483 (48.7)
FR	84 (16.60)	5946 (50.7)	5775 (49.3)
IT	29 (5.73)	1861 (48.3)	1992 (51.7)
NL	1 (0.20)	26 (48.1)	28 (51.9)
NO	6 (1.19)	507 (51.5)	478 (48.5)
Total Standardbred horses: 506			

Country	Standardbreds (%)	Males offspring (%)	Female offspring (%)
SE	163 (32.21)	8105 (49.4)	8318 (50.6)
US	192 (37.94)	19 748 (48.1)	21 302 (51.9)

Country	Coldblooded trotter (%)	Male offspring (%)	Female offspring (%)
NO	78 (62.4)	4329 (50.4)	4265 (49.6)
SE	47 (37.6)	2064 (50.7)	2011 (49.3)
Total Coldblooded trotters: 125			

Table 13. Number of Icelandic horses registered every ten year including both male and female proportions. starting in 1950.

Years of registration	Male (%)	Female (%)	Total
1950 – 1959	709 (31.0)	1581 (69.0)	2290
1960 – 1969	2238 (32.3)	4683 (67.7)	6921
1970 – 1979	4909 (26.0)	13 944 (74.0)	18 853
1980 – 1989	16 891 (37.7)	27 936 (62.3)	44 827
1990 – 1999	49 744 (46.4)	57 496 (53.6)	107 266
2000 – 2009	65 157 (47.1)	70 254 (50.8)	138 360
2010 – 2019	62 918 (49.0)	65 269 (50.8)	128 448

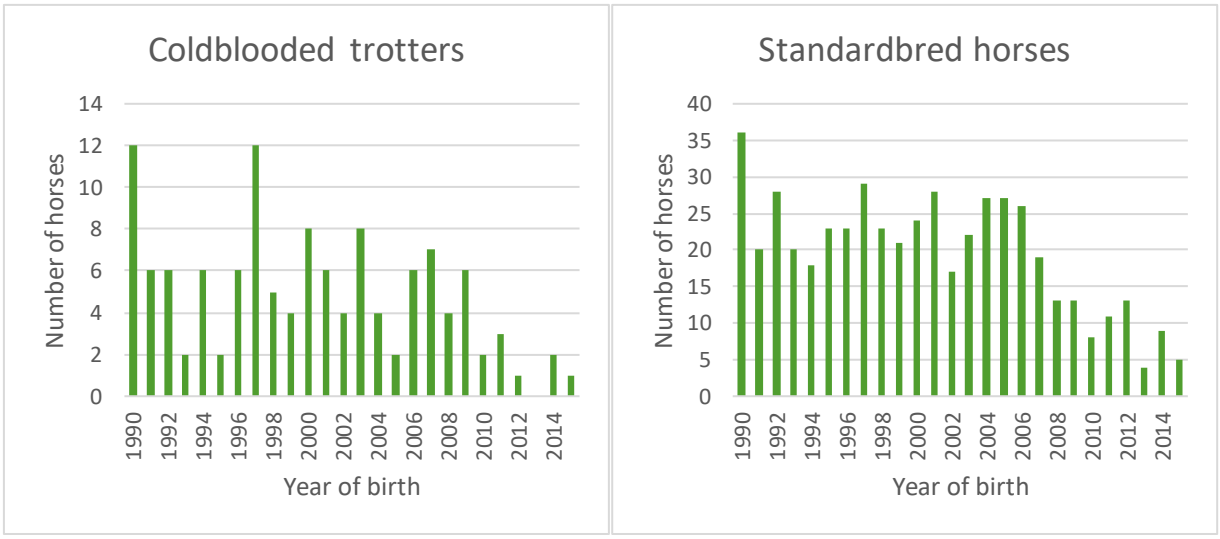


Figure 13. The distribution of birth of year in Coldblooded trotters and Standarbred horses.

7.2. R-script to create the main dataset of Icelandic horses

This R-script was used to create the main dataset, using a transcript of all Icelandic horses from WorldFengur (WorldFengur, 2021).

```
#Cleaned R-script to receive the start-dataset of Icelandic horses
#Information from hrossFadirModir.txt:
#Before starting the script, paste to excel and remove name and origin.
horses <- openxlsx::read.xlsx("~/location_of_dataset/name_of_dataset.xlsx",
                colNames = FALSE)
horses <- horses[,c(1:3)]
head(horses)
colnames(horses) <- c("id", "id_father", "id_mother")

#Assign the sex to the horses
horses$sex <- 0
horses[(substr(horses$id, 7, 7) == "1"),"sex"] <- "m"
horses[(substr(horses$id, 7, 7) == "2"),"sex"] <- "f"
horses[(substr(horses$id, 7, 7) == "3"),"sex"] <- "unknown"

#Count the number of offspring per horse
horses$offspring <- 0
a=1
while (a<=nrow(horses)) {
 id_tmp <- horses[a,"id"]
 sex_tmp <- horses[a,"sex"]
 if (sex_tmp == "f") {
  horses[a,"offspring"] <- length(which(horses$id_mother == id_tmp))
 }
 if (sex_tmp == "m") {
  horses[a,"offspring"] <- length(which(horses$id_father == id_tmp))
 }
 #head(horses)
 if (a %in% seq(1,509008,10000)) {
  print(a)
 }
 a=a+1
}

#Count the number of male, female or unkown offspring per horse
horses$male_offspring <- 0
horses$female_offspring <- 0
horses$unknown_offspring <- 0
a=1
while (a<=nrow(horses)) {
 if (horses[a,"offspring"] > 0) {
  id_tmp <- horses[a,"id"]
  list_off_sexes <- horses[ c(which(horses$id_mother == id_tmp), which(horses$id_father ==
id_tmp)),"sex"]
  #
  horses[a,"male_offspring"] <- length(which(list_off_sexes == "m"))
  horses[a,"female_offspring"] <- length(which(list_off_sexes == "f"))
  horses[a,"unknown_offspring"] <- length(which(list_off_sexes == "unknown"))
 }
 if (a %in% seq(1,509008,10000)) {
  print(a)
 }
 a=a+1
}

#Calculated the number of offspring with a known gender
horses$known_offspring <- horses$male_offspring + horses$female_offspring
```

7.3. Statistical analyses, influence of number of offspring

Table 14. This table shows how significance and power is influenced by the number of offspring. The colours emphasize the number of offspring needed for a significant p-value or a power greater than 0.8.

Number offspring	Sex A	Sex B	Proportion Sex A	*p-value*	Power	Bon-ferroni (8787)	Bon-feroni (13 703)	Bon-ferroni (195 600)
1	1	0						
2	1	1	0.50	1	0.05	1	1	1
3	2	1	0.67	0.564	0.09	1	1	1
4	3	1	0.75	0.317	0.17	1	1	1
5	4	1	0.80	0.180	0.27	1	1	1
6	5	1	0.83	0.102	0.37	1	1	1
7	6	1	0.86	0.059	0.47	1	1	1
8	7	1	0.88	0.034	0.56	1	1	1
9	8	1	0.89	0.020	0.65	1	1	1
10	9	1	0.90	0.011	0.72	1	1	1
11	10	1	0.91	0.007	0.77	1	1	1
12	11	1	0.92	0.004	0.82	1	1	1
13	12	1	0.92	0.002	0.86	1	1	1
14	13	1	0.93	0.001	0.89	1	1	1
15	14	1	0.93	7.89E-4	0.92	1	1	1
16	15	1	0.94	4.65E-4	0.94	1	1	1
17	16	1	0.94	2.75E-4	0.95	1	1	1
18	17	1	0.94	1.62E-4	0.96	1	1	1
19	18	1	0.95	9.62E-5	0.97	0.845	1	1
20	19	1	0.95	5.70E-5	0.98	0.501	0.781	1
21	20	1	0.95	3.38E-5	0.99	0.297	0.463	1
22	21	1	0.95	2.01E-5	0.99	0.176	0.275	1
23	22	1	0.96	1.19E-5	0.99	0.105	0.164	1
24	23	1	0.96	7.10E-6	0.99	0.062	0.097	1
25	24	1	0.96	4.22E-6	1.00	0.037	0.058	0.826
26	25	1	0.96	2.52E-6	1.00	0.022	0.034	0.492
27	26	1	0.96	1.50E-6	1.00	0.013	0.021	0.293
28	27	1	0.96	8.94E-7	1.00	0.008	0.012	0.175
29	28	1	0.97	5.34E-7	1.00	0.005	0.007	0.104
30	29	1	0.97	3.19E-7	1.00	0.003	0.004	0.062
31	30	1	0.97	1.90E-7	1.00	0.002	0.003	0.037
32	31	1	0.97	1.14E-7	1.00	9.99E-4	0.002	0.022
33	32	1	0.97	6.80E-8	1.00	5.97E-4	9.32E-4	0.013
34	33	1	0.97	4.07E-8	1.00	3.57E-4	5.57E-4	0.008
35	34	1	0.97	2.43E-8	1.00	2.14E-4	3.33E-4	0.005

7.4. Pedigree identifier

This python script makes pedigrees with any chosen number of generations. It was used to identify parents, grandparents and great-grandparents of the Icelandic horses.

```
#This program makes pedigrees for selected individuals, using a reference dataset.
#The number of generations need to be specified.
import pandas as pd

def Family_tree(dataset, ref_dataset, generations):
  ref_data = pd.read_csv(ref_dataset, sep=";", header=None, encoding="ISO-8859-1")
#Opens the table with all the data, this file includes all Icelandic horses.
  ref_data.columns = ['ID', 'name', 'origin', 'ID_sire', 'name_sire', 'origin_sire',
'ID_dam', 'name_dam','origin_dam'] # Assigns names to columns
  new_data = open('result_pedigree.txt', 'w') #Open the new file with the results.
  parental_search = (2**generations)-2 #The number of ids to look up is 2^generations - 2.

  for i in range(len(dataset)):
    line = '' #Is for later, to make a string of the list with results.
    #if i%1000 ==0: #Counter for a large dataset.
    if i%25 == 0: #Counter for a small dataset
      print(i)
    for j in range(len(ref_data)):
      pedigree = []
      if dataset[i] == ref_data['ID'][j]: #Find the ID in the data
        pa = [ref_data['ID'][j], ref_data['ID_sire'][j], ref_data['ID_dam'][j]]
        pedigree = pedigree + pa #Include info in the pedigree list.

        for l in range(parental_search):
          pedigree = pedigree #Adjust the pedigree, we added variables.
          var1 = pedigree[1+l] #Select the ids of which we want to find parents.
          fa = '' #The string in which the father is included.
          for m in range(len(ref_data)):
            if var1 == ' ' or var1 == 'nan':
#If no id is found, there are also no parents.
              fa = ' '
              ma = ' ' #mothers ID
            elif var1 == ref_data['ID'][m]: #Find the variable in the dataset.
              fa = ref_data['ID_sire'][m] #Account father.
              ma = ref_data['ID_dam'][m] #Account mother.
          if fa == '': #if the ID is not found both IDs stay empty as well.
            fa = ' '
            ma = ' '
          pedigree = pedigree + [fa] + [ma] #Add the variables to the list.

        for k in range(len(pedigree)): #Appoint every variable in the list.
          lst = pedigree[k]
          line = line + str(lst) + ';'#';' is used to separate the columns in the table.
        line = line + '\n' #Add an enter between the different lists.
    new_data.write(line)

  new_data.close()
  return
```

```
def Reading_str(file_name, ref_dataset,generations): #To obtain a id-list I like to
search.
  B= open(file_name, 'r') #Open the file.
  lines = B.readlines()

  List = [] #Make a list for the Find_ID_dadmum formula.

  for i in range(len(lines)): #Go over the string.
    str = ''

    for j in range(len(lines[i])): #Go over the characters.
      if lines[i][j:j+1] != '\n': #When there is a enter start a new variable.
        str = str + lines[i][j] #Include all characters in one string until the enter.

    List = List + [str] #the list which include all ID numbers.

  print(List) #To see the individuals.
  Family_tree(List, ref_dataset, generations)
  B.close()
  return


#You can use a txt file or a Python list.
dataset = 'IS_horses_included.txt' #The list of individuals you want to look up, txt
format.
#dataset = ['IS19902xxxxx','IS19941xxxxx'] #An example of a Python list
ref_dataset = 'hrossFadirModir.txt' #The reference dataset, including individuals and
parents.
generations = 2 #Number of generations, in this case parents and grandparents.
Reading_str(dataset, ref_dataset, generations) #In case your dataset is txt file.
#Family_tree(dataset, ref_dataset, generations) #In case your dataset is a list.
```

7.5. Heritability Analyses

7.5.1. Matching sons to fathers and maternal grandfathers

To estimate the heritability the proportion of male offspring between fathers and sons, and maternal grandfathers and grandsons were made. To perform this analyse the proportion of male offspring had to be matched between individuals, fathers and maternal grandfathers. This python script match the number of offspring in the three generations.

```
#This script makes a file which match offspring with their father and grandfathers.
#Their proportions of male offspring can be used to do a correlation analyses, which is
needed to calculate heritability
import pandas as pd

def Proportion_match(test_dataset, ref_dataset):
  data = pd.read_excel(test_dataset) #Open the dataset you are interested in.
  data.columns = ['ID', 'offspring', 'Male offspring', 'significance', 'ID_sire', 'ID_dam-
sire'] #Assigns names to columns.
  ref_data = pd.read_excel(ref_dataset) #Open your reference dataset.
  ref_data.columns = ['ID', 'offspring', 'Male offspring', 'significance']
                                    #Assign names to columns.
  new_data = open('Proportion_result.txt', 'w') #Open the new file with the results.

  for i in range(len(data)): #Go through every row.
    p_sire = [] #Data father.
    p_d_sire = [] #Data maternal grandfather.
    proportion = [data['ID'][i], data['offspring'][i], data['Male offspring'][i],
data['significance'][i]] #Starting string.
    line = ''
    if i%100 ==0: #Counter if needed.
      print(i)
    for j in range(len(ref_data)):
      if data['ID_sire'][i] == ref_data['ID'][j]:#Match father with father in reference.
        p_sire = [ref_data['ID'][j], ref_data['offspring'][j], ref_data['Male
offspring'][j], ref_data['significance'][j]] #Include in the father list.
      if data['ID_dam-sire'][i] == ref_data['ID'][j]: #Match maternal grandfather with
maternal grandfather in reference.
        p_d_sire = [ref_data['ID'][j], ref_data['offspring'][j], ref_data['Male
offspring'][j], ref_data['significance'][j]]
    if p_sire == [] or p_sire == '' or p_sire == 'nan':
      p_sire = [data['ID_sire'][i], ' ', ' ', ' ']
    if p_d_sire == [] or p_d_sire == '' or p_d_sire == 'nan':
      p_sire = [data['ID_dam-sire'][i], ' ', ' ', ' ']
    proportion = proportion + p_sire + p_d_sire

    for k in range(len(proportion)): #Appoint every list in the dictionary.
      lst = proportion[k]
      line = line + str(lst) + ';'#';' is used to separate the columns in the table.
    line = line + '\n' #Add a enter between the different individuals.
    new_data.write(line) #Write in the datafile.

  new_data.close()
  return

test_dataset = 'Correlations_significant.xlsx'
ref_dataset = 'Correlations_significant_ref.xlsx'
Proportion_match(test_dataset, ref_dataset)
```

7.5.2. Heritability calculations and extra figures

This R-script estimates the heritability. All figures related to heritability in this thesis were made using this script.

```
#Cleaned R-script to calculated the heritability of distorted sex ratios of Icelandic
horses
#Using the python-script I made a file which includes offspring, fathers, and maternal
grandfathers variables.
######Importing the datasets######
library("readxl")
her <- read_excel("Master Thesis
Uppsala/Data/horses_total_sex_skew_analysis_heritability.xlsx")

#Split males and females
her_m <- her[her$sex == "m",]
her_f <- her[her$sex == "f",]

######Correlation plots######
#total plots, including both female and male individuals
plot(her$Prop_m_s, her$Prop_m)#prop. = proportion, m = male & s = sire
plot(her$Prop_m_ds, her$Prop_m) #ds = dam-sire

#For females the there is almost no correlation, with father and maternal grandfather:
plot(her_f$Prop_m_s, her_f$Prop_m)
plot(her_f$Prop_m_ds, her_f$Prop_m)

#For males:
plot(her_m$Prop_m_s, her_m$Prop_m, xlab = "Proportion male offspring, father",
   ylab = "Proportion male offspring, son", col = "blue", pch = 19,
   title("Father - offspring regression"),xlim = c(0,1), ylim = c(0,1))
abline(lm(her_m$Prop_m_s~her_m$Prop_m), col=1, lty =1, lwd=2)
title(sub="n = 6935", adj=1, line=3, font=2)
title(sub="0.0.4068 + 0.1453x", adj=1, line=4, font=2)

plot(her_m$Prop_m_ds, her_m$Prop_m, xlab = "Proportion male offspring, maternal
grandfather",
   ylab = "Proportion male offspring, son", col = "red", pch = 19,
   title("Maternal-grandfater - grandoffspring regression"),xlim = c(0,1), ylim = c(0,1))
abline(lm(her_m$Prop_m_ds~her_m$Prop_m), col=1, lty =1, lwd=2)
title(sub="n = 7156", adj=1, line=3, font=2)
title(sub="0.3395 + 0.1402x", adj=1, line=4, font=2)

#Sample size per plot
length(which(nchar(her_m$Prop_m_s)!='' & nchar(her_m$Prop_m)!= ''))
length(which(nchar(her_m$Prop_m_ds)!='' & nchar(her_m$Prop_m)!= ''))

######Calculate the heritability by the regression coefficient######
lm(formula = her_m$Prop_m_s~her_m$Prop_m) # h^2 = 2b
lm(her_m$Prop_m_ds~her_m$Prop_m)

#In case of females heritability is very small
lm(her_f$Prop_m_s~her_f$Prop_m)
lm(her_f$Prop_m_ds~her_f$Prop_m)


#Standard error's
summary(lm(her_m$Prop_m_s~her_m$Prop_m))
summary(lm(her_m$Prop_m_ds~her_m$Prop_m))

######Number of offspring and heritability######
#Make different datasets with X number of offspring
her_m_15 <- her_m[her_m$known_offspring >14 & her_m$known_offspring_s >14,]
her_m_30 <- her_m[her_m$known_offspring >29 & her_m$known_offspring_s >29,]
her_m_60 <- her_m[her_m$known_offspring >59 & her_m$known_offspring_s >59,]
her_m_100 <- her_m[her_m$known_offspring >99 & her_m$known_offspring_s >99,]
```

```
#calculate the heritability
lm(her_m_15$Prop_m_s~her_m_15$Prop_m)
lm(her_m_30$Prop_m_s~her_m_30$Prop_m)
lm(her_m_60$Prop_m_s~her_m_60$Prop_m)
lm(her_m_100$Prop_m_s~her_m_100$Prop_m)

#Standard errors
summary(lm(her_m_15$Prop_m_s~her_m_15$Prop_m))
summary(lm(her_m_30$Prop_m_s~her_m_30$Prop_m))
summary(lm(her_m_60$Prop_m_s~her_m_60$Prop_m))
summary(lm(her_m_100$Prop_m_s~her_m_100$Prop_m))

#Make one correlation graph of all different numbers of offspring together
plot(her_m$Prop_m_s, her_m$Prop_m, xlab = "Proportion male offspring, father",
   ylab = "Proportion male offspring, son", col = "#bfd3e6", pch = 19,
   title("Father-offspring regression"),xlim = c(0,1), ylim = c(0,1))
points(her_m_15$Prop_m_s, her_m_15$Prop_m, col = "#9ebcda", pch = 19)
points(her_m_30$Prop_m_s, her_m_30$Prop_m, col = "#8c96c6", pch = 19)
points(her_m_60$Prop_m_s, her_m_60$Prop_m, col = "#8856a7", pch = 19)
points(her_m_100$Prop_m_s, her_m_100$Prop_m, col = "#810f7c", pch = 19)
legend("bottomright", c("8-14 offspring", "15-29 offspring", "30-59 offspring", "60-99
offspring", "100 or more offspring"),
    cex=.8, col=c("#bfd3e6", "#9ebcda", "#8c96c6", "#8856a7","#810f7c"), pch=c(19, 19, 19,
19, 19))
abline(lm(her_m_100$Prop_m_s~her_m_100$Prop_m), col="#810f7c", lty =1, lwd=1)
abline(lm(her_m_60$Prop_m_s~her_m_60$Prop_m), col="#8856a7", lty =1, lwd=1)
abline(lm(her_m_30$Prop_m_s~her_m_30$Prop_m), col="#8c96c6", lty =1, lwd=1)
abline(lm(her_m_15$Prop_m_s~her_m_15$Prop_m), col="#9ebcda", lty =1, lwd=1)
abline(lm(her_m$Prop_m_s~her_m$Prop_m), col="black", lty =1, lwd=2)

#Make four correlation graphs in one frame
attach(mtcars)
par(mfrow=c(2,2))
plot(her_m_15$Prop_m_s, her_m_15$Prop_m, xlab = "Proportion male offspring, father",
   ylab = "Proportion male offspring, sons", col = "#b3cde3", pch = 19,
   title("Father-offspring regression, both with at least 15 offspring"),
   xlim = c(0,0.8), ylim = c(0,1))
abline(lm(her_m_15$Prop_m_s~her_m_15$Prop_m), col=1, lty =1, lwd=1)
title(sub="n = 5209", adj=1, line=3, font=2)
title(sub="0.3724 + 0.2121x", adj=1, line=4, font=2)
plot(her_m_30$Prop_m_s, her_m_30$Prop_m, xlab = "Proportion male offspring, father",
   ylab = "Proportion male offspring, sons", col = "#8c96c6", pch = 19,
   title("Father-offspring regression, both with at least 30 offspring"),
   xlim = c(0,0.8), ylim = c(0,1))
abline(lm(her_m_30$Prop_m_s~her_m_30$Prop_m), col=1, lty =1, lwd=1)
title(sub="n = 2714", adj=1, line=3, font=2)
title(sub="0.3190 + 0.3214x", adj=1, line=4, font=2)
plot(her_m_60$Prop_m_s, her_m_60$Prop_m, xlab = "Proportion male offspring, father",
   ylab = "Proportion male offspring, sons", col = "#8856a7", pch = 19,
   title("Father-offspring regression, both with at least 60 offspring"),
   xlim = c(0,0.8), ylim = c(0,1))
abline(lm(her_m_60$Prop_m_s~her_m_60$Prop_m), col=1, lty =1, lwd=1)
title(sub="n = 1402", adj=1, line=3, font=2)
title(sub="0.2724 + 0.4154x", adj=1, line=4, font=2)
plot(her_m_100$Prop_m_s, her_m_100$Prop_m, xlab = "Proportion male offspring, father",
   ylab = "Proportion male offspring, sons", col = "#810f7c", pch = 19,
   title("Father-offspring regression, both with at least 100 offspring"),
   xlim = c(0,0.8), ylim = c(0,1))
abline(lm(her_m_100$Prop_m_s~her_m_100$Prop_m), col=1, lty =1, lwd=1)
title(sub="n = 807", adj=1, line=3, font=2)
title(sub="0.2558 + 0.4503x", adj=1, line=4, font=2)

detach()
```

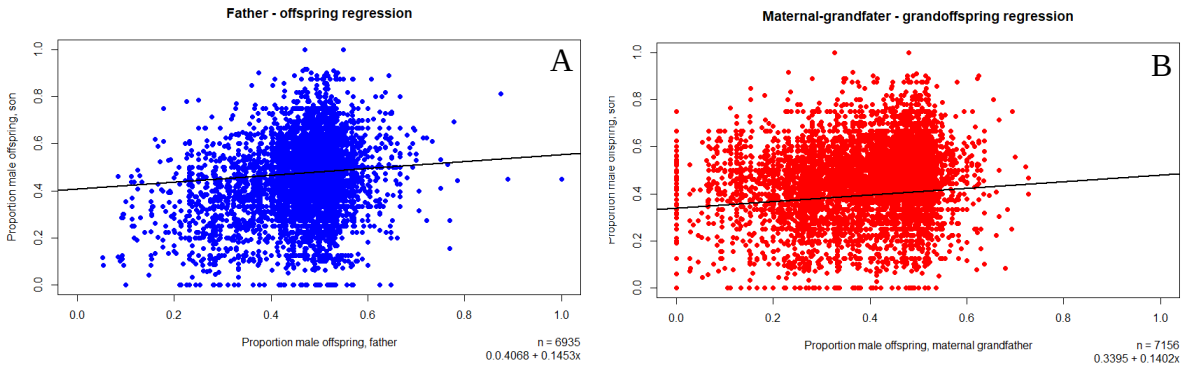


Figure 14. Estimation of heritability from the regression of proportion of male offspring in sons, fathers, and maternal grandfathers. The heritability is calculated by multiplying the regression coefficient by two for father and sons and by four for maternal grandfathers and grandsons. Graph A shows the father-son regression, including 6935 father-son pairs. The heritability was estimated at 0.2906 (SE: 0.07811). Graph B shows the relation between maternal grandfathers and grandsons, including 7136 maternal grandfathers-grandson pairs. Here the heritability was estimated at 0.5608 (SE: 0.1169).

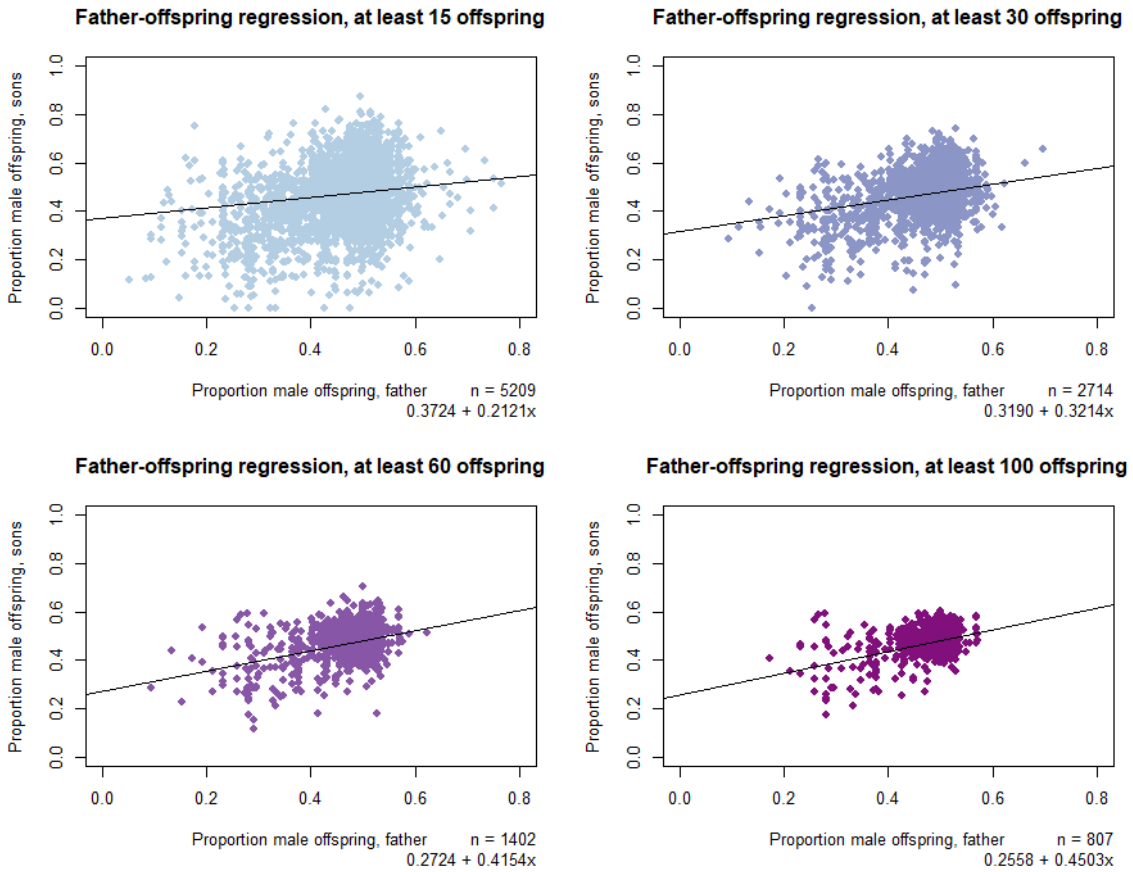


Figure 15. Estimation of heritability from father-son, displayed in separate graphs showing the change in heritability when the number of offspring in fathers and sons increase. Visible is that the number of skewing individual decreases when the number of offspring increases.

7.6. Families with distorted sex ratios

7.6.1. Identification of families with distorted sex ratios

This python script was used to identify families, using one progenitor.

With a threshold system non-significant skewing individuals can be included.

```
#with this script you can insert 1 individual.
#Using pedigree information you can find all offspring significant skewing offspring of
this individual.
#And the offspring of the offspring of the individual. And so on.
#There are two outcome data, the first one gives you all significant individuals of the
family.
#The second one gives you all individuals in this family, and gives four values.
#The first value is the amount of male offspring with 8 or more offspring.
#The second value is the amount of significant skewing male offspring
#The third value is the amount of daughters male grandchildren with 8 or more offspring
#The forth value is the amount of significant skewing maternal grandchildren.

import pandas as pd # is not needed, but sure.
import csv

def insign_offspring(id, ref_data): #Threshold system, to include non-signifcant
individuals.
  ref = csv.reader(open(ref_data, encoding='utf8'), delimiter=';')#opens the reference
data.
  header = next(ref)
  offspring = 0
  sign_offspring = 0
  grandoffspring = 0
  sign_grndoffspring = 0
  ratio_offspring = 0
  ratio_grndoffspring = 0
  included = 'no'
  if header != None:
    for row in ref:
      id_fa = row[1] # example --> this is collom 2.
      id_ma_fa = row[5]
      offspr = row[12]
      pval = float(row[14])
      sex = row[7]
      offspring_f = int(row[10])
      if sex =='m':
        if int(offspr) >7:
          if id == id_fa:
            offspring = offspring + 1
            if pval < 0.05 and offspring_f > (float(offspr)/2):
              sign_offspring = sign_offspring + 1
          if id == id_ma_fa:
            grandoffspring = grandoffspring + 1
            if pval < 0.05:
              sign_grndoffspring = sign_grndoffspring +1
  if offspring >0:
    ratio_offspring = sign_offspring / offspring
  if grandoffspring >0:
    ratio_grndoffspring = sign_grndoffspring / grandoffspring
  if ratio_offspring > 1 or ratio_grndoffspring > 1: #Thresshold for your inclusion.
    included = 'yes'

  return included

def family_finder(Id_pa, ref_data):
```

```
  ref = csv.reader(open(ref_data, encoding = 'utf8'), delimiter=';')#Opens the data file.
  header = next(ref) #To delete the first line of the dataset.
  Family = open("Family.txt", 'w')#This file will include all significant skewing family
members.
  Parents = open("Parents.txt", 'w')#This file includes the significant skewing offspring
list.
  line = ''

  #set parameters on 0 and [empty]
  offspring = 0
  sign_offspring = 0
  grandoffspring = 0
  sign_grndoffspring = 0
  nsign_incl_s = 0
  nsign_incl_g = 0
  offspring_list = []
  Fam = []
  if Id_pa[6] == '1': #So far the script is designed for a founder male.
    if header != None: #Exclude the header of the file from the analyses.
      for row in ref: # You can search in certain columns.
        id_fa = row[1] #example --> this is column 2.
        id_ma = row[2]
        id_ma_fa = row[5]
        offspr = row[12]
        id = row[0]
        pval = float(row[14])
        sex = row[7]
        offspring_m = int(row[9])
        offspring_f = int(row[10])
        #print(row)

        if sex == 'm': # Include male offspring.
                              #Females are included using maternal grandsires.

          if Id_pa == id_fa: #Look if founder is father of some offspring.
            if int(offspr) <8: #Less than 8 offspring -> the trait could be
                                    hidden, so we will include these.
              offspring_list = offspring_list + [id]#Add to the search list.
            else:
              offspring = offspring + 1#Count sons with 8 or more offspring.
              if pval <= 0.05: #If pval = significant
                if offspring_f > offspring_m:#Only include female skewing.
                  sign_offspring = sign_offspring + 1#Count the significant skewing
offspring.
                  offspring_list = offspring_list + [id]#Add to the search list.
                      Fam = Fam + [row] #This variable will be written in the individuals
                      file.
              else:
                included = insign_offspring(id, ref_data)#Look to include.
                if included == 'yes':
                  offspring_list = offspring_list + [id]#Add to search list.
                  Fam = Fam + [row]
                  nsign_incl_s = nsign_incl_s +1

          if Id_pa == id_ma_fa: #significant maternal grandsons search.
            if int(offspr) <8: #Again less than 8 offspring will be included.
              offspring_list = offspring_list + [id_ma]#add mother to avoid
                                                                    doubles.
            else:
              grandoffspring = grandoffspring + 1 #Count number of daughters
                                                          children.
              if pval <= 0.05:
                if offspring_f > offspring_m: #Only female skewing.
                  offspring_list = offspring_list + [id_ma]#add to the list.
                  sign_grndoffspring = sign_grndoffspring + 1 #Count the
significant skewing maternal grandsonsn.
              else:
                included = insign_offspring(id, ref_data) #Again inclusion?
                if included == 'yes':
                  offspring_list = offspring_list + [id]
                  Fam = Fam + [row]
                  nsign_incl_g = nsign_incl_g +1
    if offspring != 0: #Looks if the stallion has sons with 8 or more offspring.
      Prop_offsp = sign_offspring / offspring #Calculate proportion skewing offspring.
```

```
    else:
      Prop_offsp = ''

    if grandoffspring != 0 and grandoffspring != '':
      Prop_grand = sign_grndoffspring / grandoffspring #Proportion of skewing
                                                          grandoffspring.
    else:
      Prop_grand = ''
    sign_tot = sign_grndoffspring + sign_offspring
    Pa_str = Id_pa, sex, offspring, sign_offspring, nsign_incl_s, Prop_offsp,
               grandoffspring, sign_grndoffspring, nsign_incl_g, Prop_grand, sign_tot
             #Make a list of our counting's including the Id of the father.

    for k in range(len(Pa_str)): # Appoint every variable to the list.
      lst = Pa_str[k]
      line = line + str(lst) + ';'#';' is used to separate the columns in the table.
    line = line + '\n' #Ad a enter between the different lists.
    Parents.write(line) #Write it in the file.

gen = 2
  print(len(offspring_list))#Print the number of individuals the program need to go
through
  print(gen)#Print the generation we are at.

  while offspring_list !=[] and gen <25:#I took 25 generations as maximum, to stop the
loop.
  #while offspring_list !=[]: #Here is a while loop which causes the loop to continue
until there are no relatives anymore. This loop is dangerous, because it may go on
forever.
    gen = gen +1 #Generation counter.
    Fam = Fam + [str(gen)] #Include generation per horse in the file, nice for order.
    print(gen) #In the studbook most horses go back 12 generations.
                This gives an indication how many generation are still needed.
    offspring_temp = [] #This is a temporal file, to make a new offspring list.
                         It will replace the old one in the while loop.
    for i in offspring_list: #for all individuals walk through the loop
      offspring = 0 #Recet the variables
      sign_offspring = 0
      grandoffspring = 0
      sign_grndoffspring = 0
      nsign_incl_s = 0
      nsign_incl_g = 0
      ref = csv.reader(open(ref_data, encoding='utf8'), delimiter=';')
      header = next(ref)
      if header != None: #Again exclude the heather.
        for raw in ref: #Start the loop, row became raw.
          id_fa = raw[1]
          id_ma = raw[2]
          id_ma_fa = raw[5]
          offspr = raw[12]
          id = raw[0]
          pval = float(raw[14])
          sex = raw[7]
          offspring_m = int(raw[9])
          offspring_f = int(raw[10])

          if sex == 'm': #Include male offspring.

            if i[6] == '1': #If the offspring is a male.
              if i == id_fa:#'i' is the new id_pa

                if int(offspr) < 8: #From here it repeats script above.
                  offspring_temp = offspring_temp + [id]#New offspring_list.
                else:
                  offspring = offspring + 1
                  if pval <= 0.05:
                    if offspring_f > offspring_m:
                      sign_offspring = sign_offspring + 1
                      if raw not in Fam: #Reduce the doubles in the file.
                        offspring_temp = offspring_temp + [id]
                        Fam = Fam + [raw]
                  else:
                    included = insign_offspring(id, ref_data)
                    if included == 'yes':
```

```
                    if raw not in Fam:
                      offspring_temp = offspring_temp + [id]
                      Fam = Fam + [raw]
                    nsign_incl_s = nsign_incl_s + 1

              if i == id_ma_fa:
                if int(offspr) < 8:
                  offspring_temp = offspring_temp + [id_ma]
                else:
                  grandoffspring = grandoffspring + 1
                  if pval <= 0.05:
                    if offspring_f > offspring_m:
                      offspring_temp = offspring_temp + [id_ma]
                      sign_grndoffspring = sign_grndoffspring + 1
                    else:
                      included = insign_offspring(id, ref_data)
                      if included == 'yes':
                        offspring_temp = offspring_temp + [id_ma]
                        nsign_incl_g = nsign_incl_g + 1
            else: #In case of female individuals.
              grandoffspring = '' #For females grandchildren are excluded.
              sign_grndoffspring = ''
              nsign_incl_g = ''
              sex = 'f'
              if i == id_ma:#The same idea as for males.
                if int(offspr) < 8:
                  offspring_temp = offspring_temp + [id]
                else:
                  offspring = offspring + 1
                  if pval <= 0.05:
                    if offspring_f > offspring_m:
                      sign_offspring = sign_offspring + 1
                      if raw not in Fam:
                        offspring_temp = offspring_temp + [id]
                        Fam = Fam + [raw]
                    else:
                      included = insign_offspring(id, ref_data)
                      if included == 'yes':
                        if id not in offspring_temp:
                          if raw not in Fam:
                            offspring_temp = offspring_temp + [id]
                            Fam = Fam + [raw]
                        nsign_incl_s = nsign_incl_s + 1


    line = ''#Appoint.
    if offspring != 0:
      Prop_offsp = sign_offspring / offspring #Same proportion calculations.
    else:
      Prop_offsp = ''

    if grandoffspring != 0 and grandoffspring != '':
      Prop_grand = sign_grndoffspring / grandoffspring
    else:
      Prop_grand = ''
    if sign_grndoffspring != '':
      sign_tot = sign_offspring + sign_grndoffspring
    else:
      sign_tot = ''
    Pa_str = i, sex, offspring, sign_offspring,nsign_incl_s,Prop_offsp, grandoffspring,
             sign_grndoffspring, nsign_incl_g, Prop_grand, sign_tot
             #Again! Females will not have a number for the last two variables.
    for l in range(len(Pa_str)): #Appoint every list in the dictionary
      lst = Pa_str[l]
      line = line + str(lst) + ';' #';' is used to separate columns in the table.
    line = line + '\n' #Add a enter between the different lists
    Parents.write(line)


  offspring_list = offspring_temp #Very important step. Appoint a new offspring list.
  print(len(offspring_list)) #Print length of the offspring list, for time indications.
line2 = '' #Add all individuals to the family file.
for n in Fam: #For all family member.
  for o in n:#For all variable per family member.
```

```
            lst = o
            line2 = line2 + str(lst) + ';'#Add variables to the list.
        line2 = line2 + '\n' #An enter between each family member.
    Family.write(line2) #Write it in the file.

    Family.close() #Close the directories.
    Parents.close()
    return Fam #If you want an extencive list, print Fam.

Id_pa = "IS19841xxxxx" #Example of founder father.
ref_data = "Icelandic horses_met offspring.csv" #Reference file, I only included stallions
with at least 1 offspring.
family_finder(Id_pa, ref_data)
```

7.6.2. Small fraction of Family 2

The family tree in Figure 16 was made for future studies, demonstrating the relation between some horses in Family 2. The Swedish University of Agricultural Science (SLU) has collected blood samples of almost 3000 Icelandic horses and in future some of these samples could be possibly used for whole genome sequencing, to explore the genetic cause of skewing sex ratios.

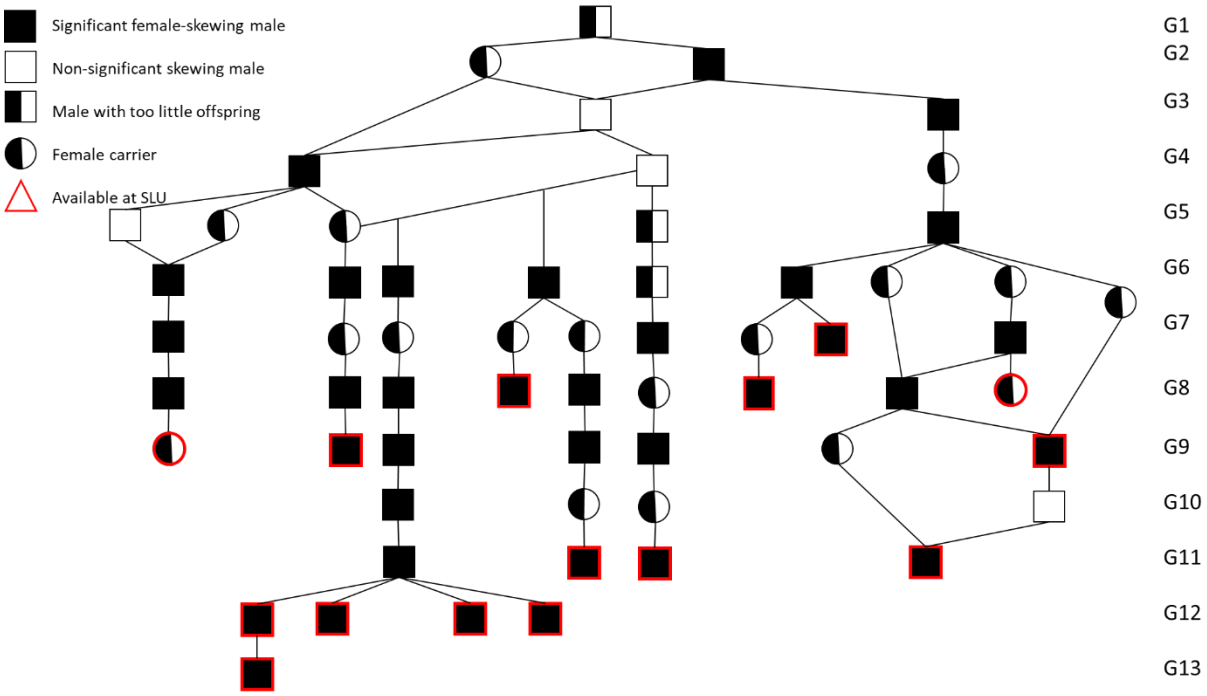


Figure 16. A small part of family 2, showing the relation between the 15 horses of which a blood sample is available at SLU.

7.7. qPCR protocol for semen

7.7.1. Semen purification

1 ml frozen semen was melted at room temperature.
1 ml of 80% Equipure^{MT} was pipetted in a 2 ml Eppendorf tube.
The 1 ml melted semen was added on top of the Equipure^{MT}.
The tube was centrifuged at 400 x g for 20 minutes in a balanced swing-out rotor centrifuge (without break).
The supernatant was discarded and the pellet was washed once with 1 ml 1% Phosphate-buffer saline (PBS).

7.7.2. QIASymphony DNA extraction

This protocol was used for both the pellet purified semen and 50 µl raw semen, both in a two ml Eppendorf tube. For DNA extraction, an adjusted protocol from QIAGEN, 2013 was used

160 µl ATL Buffer was added to the tubes.
20 µl proteinase K of the QiaS Minikit was added to the tubes.
20 µl DTT 1M was added to the tubes.
The tubes were vortexed and incubated at 56°C for around 1,5 hour.

The tubes went in the QIA S, with the issue high content protocol (Tisue_HC_200).
Afterwards concentration was measured using nanodrop.

The following concentration were measured:
Purified semen: 93.13 ng/µl
Raw semen: 39.16 ng/µl

7.7.3. Simplistic DNA extraction

20 µl of raw semen was vortexed with 30 µl PBS and incubated at 98°C for 21 minutes.
This semen will be called "cooked" in future references.

7.7.4. Plates - qPCR

The concentration of purified semen was diluted to 39.16 ng/µl.
4.2 µl purified semen was added to 5.8 µl H_2O and vortexed in a 2 ml Eppendorf tube.
10 µl of raw semen and cooked semen were extracted in two Eppendorf tubes.

The primers were suspended in the recommended amount of H_2O to create a 100 µM solution.
For the SYBR-Green PCR reaction 300-800 nM primer was needed.
We diluted the stock with the following calculation:

0.5 µM / 100 µM * 10 µl = 0.05 µl.
Which means that 10 µl dilution (500nM) exist of 0.05 µl stock solution and 9.95 µl H_2O.

For each primer we made a master mix, enough for five reaction.
In the end only four reactions per primer are needed for the purified semen, raw semen, cooked semen, and a water control. The plate format is presented in Table 14.

qPCR master mix
25 µl SYBR Green Master Mix
5 µl Primer Forwards (500 nM)
5 µl Primer Revers (500 nM)
10 µl H_2O

A 96 well plate was filled according to the format shown in Table 14.
Each well consisted of 9 µl qPCR master mix and 1 µl of each sample.
The plate was sealed with an optical adhesive cover, and centrifuged to eliminate air bubbles.

Table 15. Plate format used during the experiment.

Plate		X-*AR*	Y-*DDX3Y*	*GAPDH*	*Actin-β*
		1	2	3	4
Purified	A	P-X	P-Y	P-*GAPDH*	P-*Actin*
Raw	B	R-X	R-Y	R-*GAPDH*	R-*Actin*
Cooked	C	C-X	C-Y	C-*GAPDH*	C-*Actin*
Control	D	Water-X	Water-Y	Water-*GAPDH*	Water-*Actin*

7.7.5. Taqman qPCR settings

The qPCR assay was performed on the StepOnePlusTM Real-Time PCR system, 96 wells (Applied Biosystems, Foster city, CA, USA). The quantitation – standard curve experiment was chosen as setup and SYBR® Green Reagents as detection methods. The ramp speed of the instrument was installed as standard. The reaction conditions started with a holding stage, in which Uracil-DNA glycosylases was activated for 2 minutes at 50°C. This was followed by 2 min at 95 °C to dual-lock DNA polymerase and
40 cycles of 15 seconds at 95 °C, 15 seconds at the annealing temperature of 58 °C, and 1 minute at 72 °C. Lastly in the melting curve stage the plate was incubated for 15 seconds at 95 °C, 1 minute at 60 °C , and 15 seconds at 95 °C (Figure 17). Threshold cycles (Cts) and melting curves were measured by StepOnePlus version 2.3 software (Applied Biosystems).

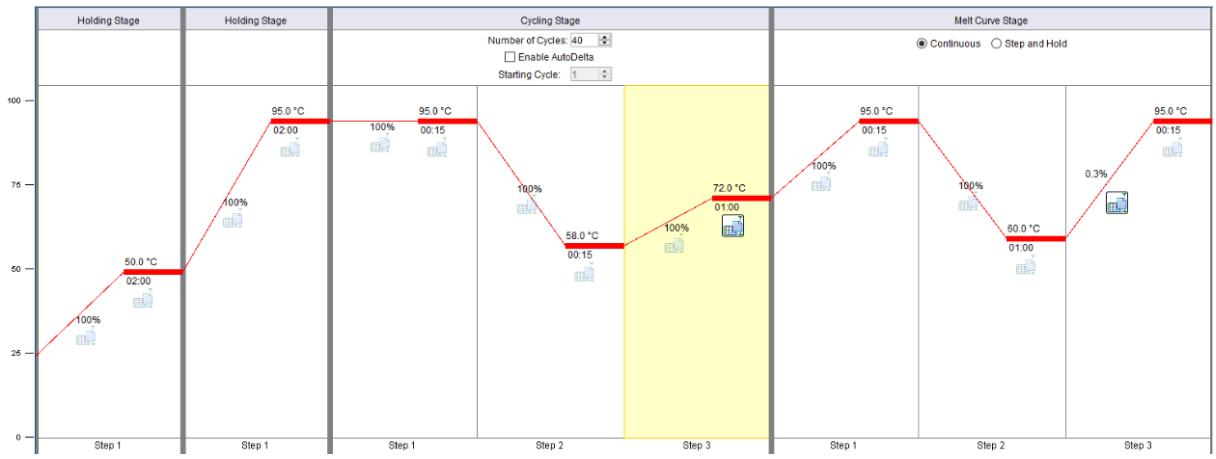


Figure 17. qPCR reaction conditions

7.7.6. Fold gene expression calculation

The fold change was calculated using the 2^{-ddct} method.
First the ΔCt was calculated for the *AR* and *GADPH* (target genes), *actin-β* served as the reference gene.

ΔCt(target gene) = Ct(target gene) – Ct(reference gene)

After ΔCt the expression status of the target genes were calculated, relative to the reference gene were calculated using the raw semen as a control. Followed by the calculation of the fold gene expression.

ΔΔCt = ΔCt(target gene) – ΔCt(target gene, control sample)
Fold gene expression = 2 ^ (-ΔΔCt)

7.7.7. qPCR results

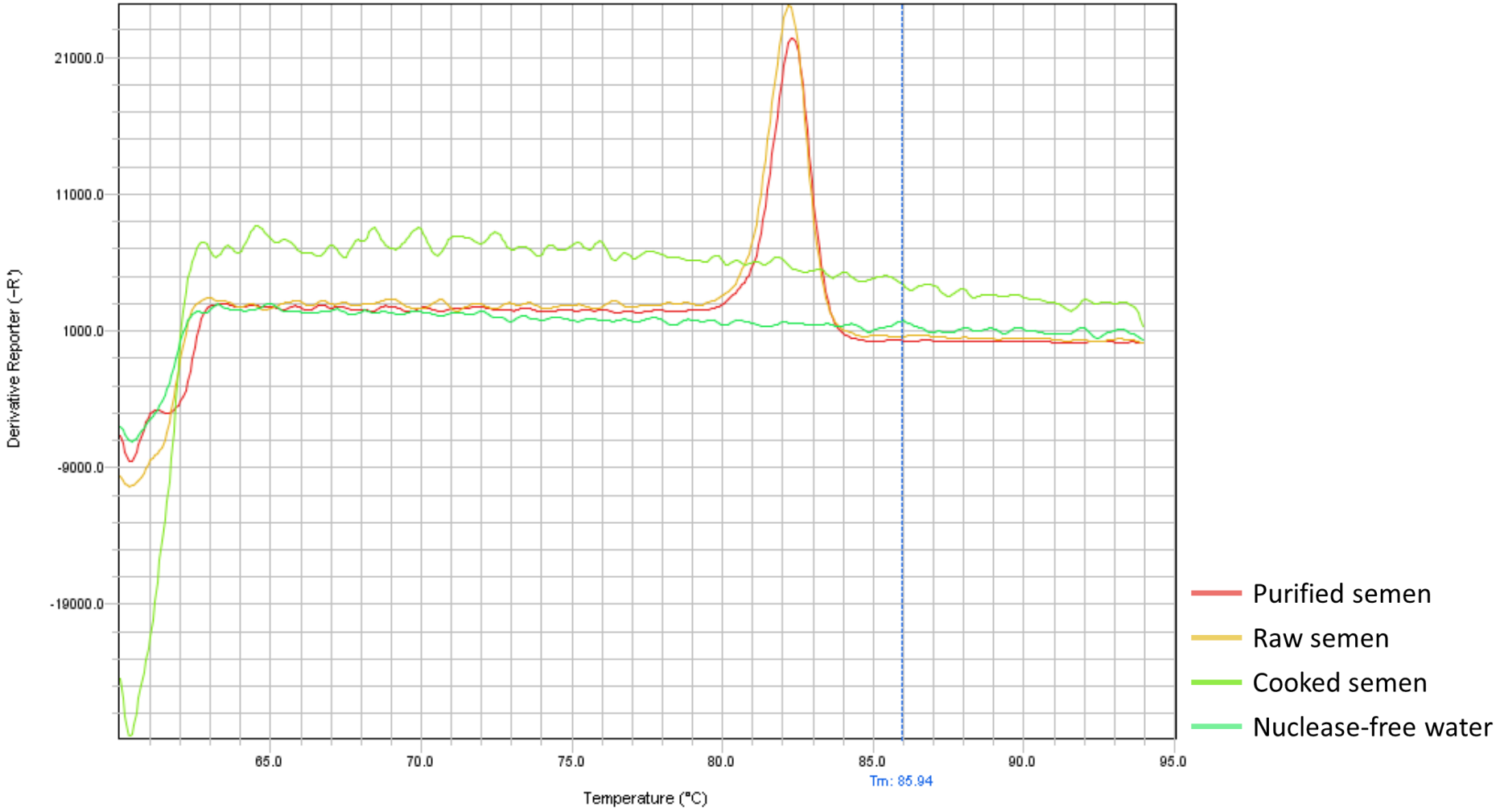


Figure 18. Melting curve of the AR primer, enlargement of figures 11 (AR).

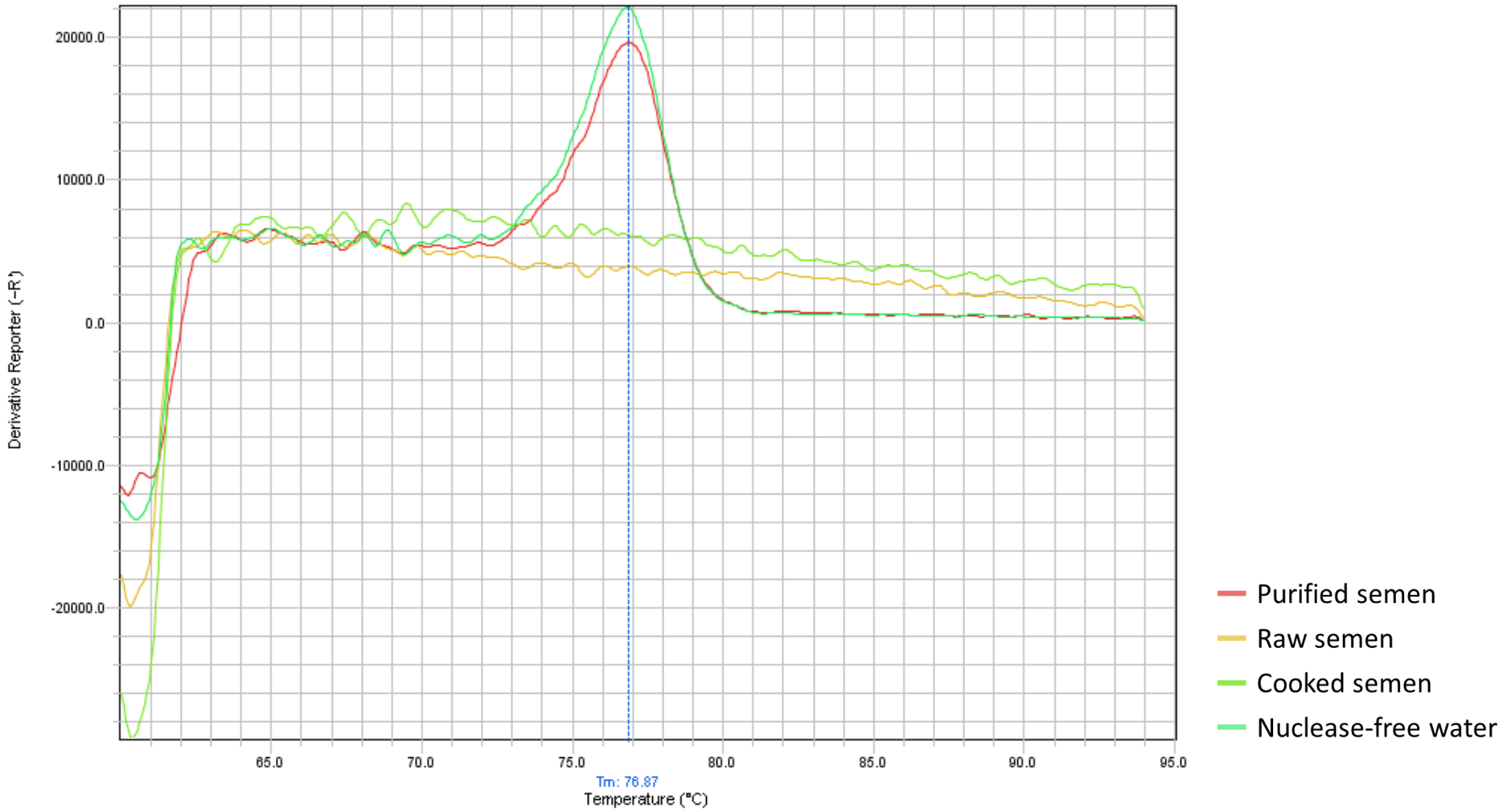


Figure 19. Melting curve of the DDX3Y primer, enlargement of figures 11 (DDX3Y).

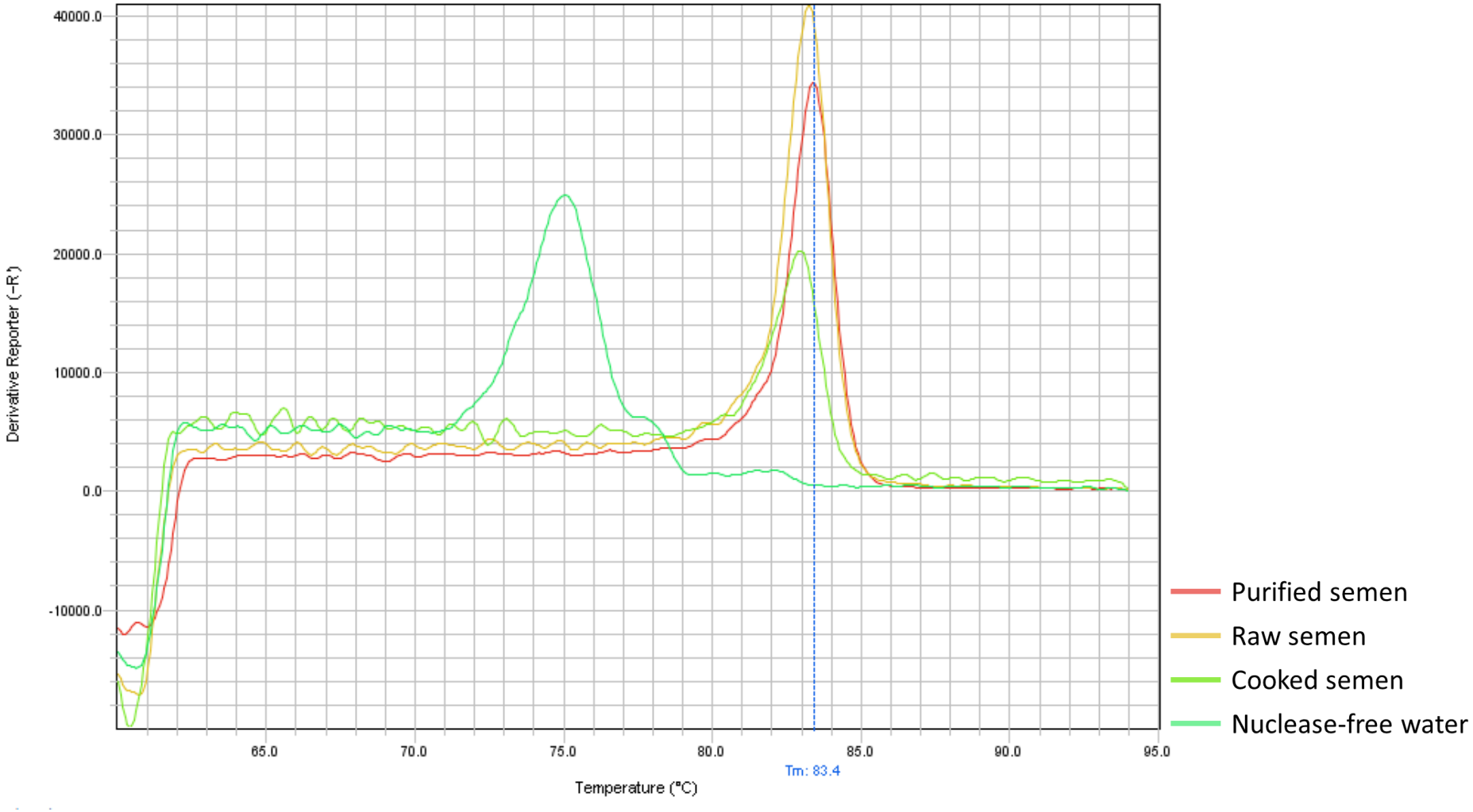


Figure 20. Melting curve of the GADPH primer, enlargement of figures 11 (GADPH).

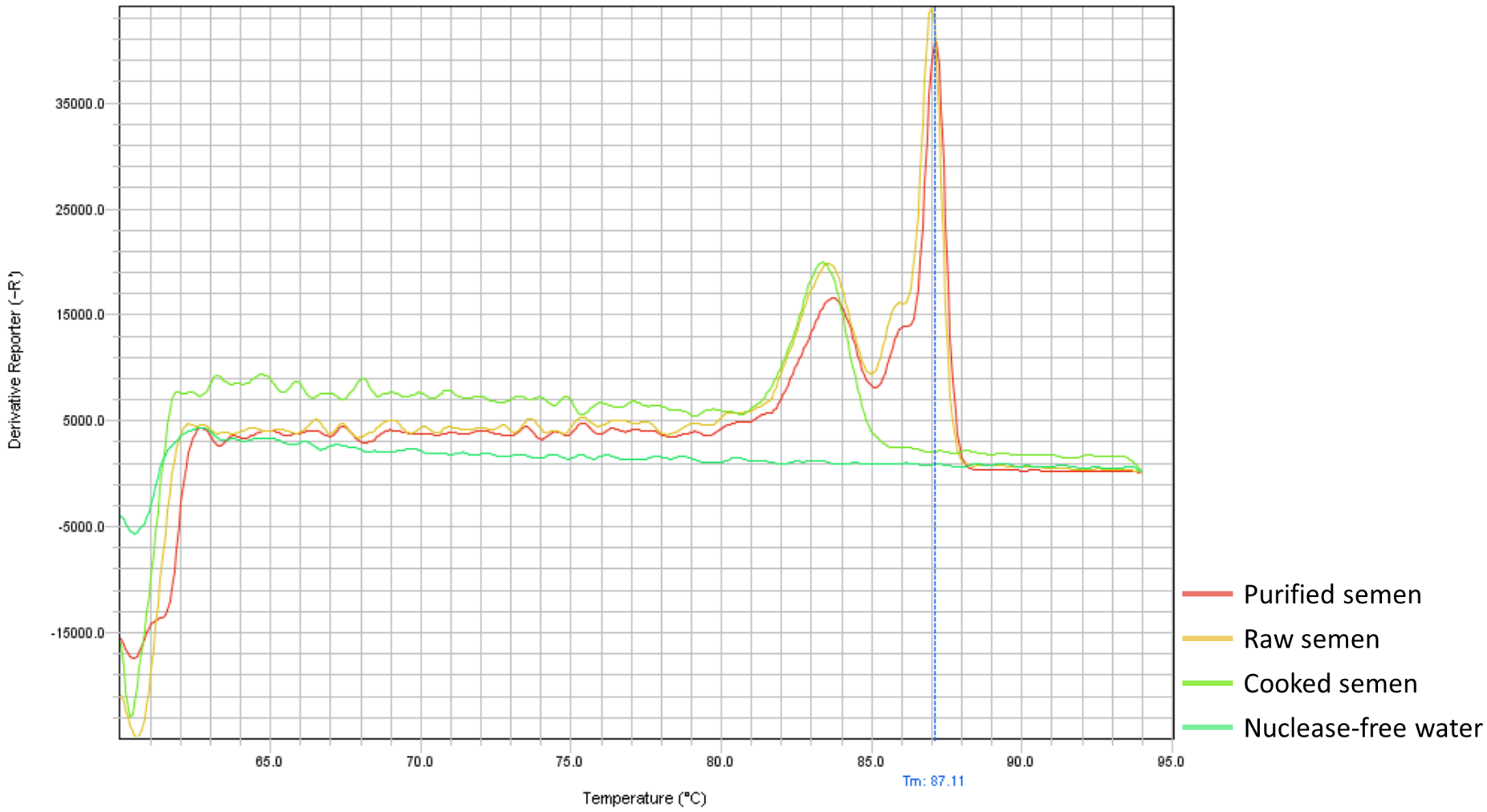


Figure 21. Melting curve of the Actin-β primer, enlargement of figures 11 (Actin-β).

7.8. Scientific Poster – Master thesis day

Exploring Heredity of Distorted Sex-ratios in Icelandic Horses

Introduction

Preliminary analyses have demonstrated that some stallions have significantly more male offspring, while others sire more female offspring than expected from an equal sex ratio.

Aim

- To investigate distorted sex-ratios in Icelandic horses.
- To assess the heritability of distorted sex-ratios in Icelandic horses

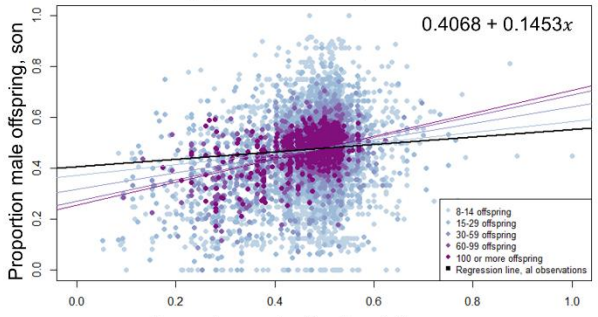
Materials & Methods

- All 509.008 horses of the Icelandic studbook were included (43.7% male, 55.7% female, 0.6% unknown sex)
- Significant individuals were identified using Chi-square tests. (Only individuals with at least 8 offspring were included in the analyses.)
- Heritability was calculated using a regression analysis

Results

8 or more offspring	Male	Female	Total
Number	8787	13703	22490
Sign. Chi-square	1404 (16.0%)	794 (5.8%)	2198 (9.8%)
Female-skewing	1218 (13.9%)	480 (3.5%)	1698 (7.6%)
Male-skewing	186 (2.1%)	314 (2.3%)	500 (2.2%)

Correlation proportion male offspring fathers vs sons



- 6935 observations
- Heritability = 0.29 (= 2 * 0.1453)

Conclusions

- Distorted sex-ratios are more common in stallions compared to mares.
- Female-skewing stallions are more common than male-skewing stallions.
- The heritability for the proportion of male offspring is 0.29 in Icelandic horses.
- The heritability increases when fathers and sons have more offspring.

Susan E. Ott
MSc Biology student
susan.ott@wur.nl

Supervised by:

Gabriella Lindgren
Daily Supervisor
Uppsala (SE)

Ole Madsen
WUR Supervisor
Wageningen (NL)

Terje Raudsepp
Texas Supervisor
Texas (US)

Elsa Albertsdóttir
Hvanneyri (IS)

Illustrations:Susan Ott

WorldFengur
The Studbook of Origin for the Icelandic horse

WAGENINGEN UNIVERSITY & RESEARCH